Handbook of Experimental Pharmacology 260

James E. Barrett
Clive P. Page
Martin C. Michel *Editors*

Concepts and Principles of Pharmacology

100 Years of the Handbook of Experimental Pharmacology

Springer

Handbook of Experimental Pharmacology

Volume 260

The *Handbook of Experimental Pharmacology* is one of the most authoritative and influential book series in pharmacology. It provides critical and comprehensive discussions of the most significant areas of pharmacological research, written by leading international authorities. Each volume in the series represents the most informative and contemporary account of its subject available, making it an unrivalled reference source.

HEP is indexed in PubMed and Scopus.

More information about this series at http://www.springer.com/series/164

James E. Barrett • Clive P. Page •
Martin C. Michel
Editors

Concepts and Principles of Pharmacology

100 Years of the Handbook of Experimental Pharmacology

Springer

Editors
James E. Barrett
Center for Substance Abuse Research
Lewis Katz School of Medicine at Temple University
Philadelphia, PA, USA

Clive P. Page
Sackler Institute of Pulmonary Pharmacology,
Institute of Pharmaceutical Science
King's College London
London, UK

Martin C. Michel
Department of Pharmacology
Johannes Gutenberg University
Mainz, Rheinland-Pfalz, Germany

ISSN 0171-2004 ISSN 1865-0325 (electronic)
Handbook of Experimental Pharmacology
ISBN 978-3-030-35361-2 ISBN 978-3-030-35362-9 (eBook)
https://doi.org/10.1007/978-3-030-35362-9

This Springer imprint is published by the registered company Springer Nature Switzerland AG.
The registered company address is: Gewerbestrasse 11, 6330 Cham, Switzerland

Preface

This volume celebrating the 100th anniversary of the *Handbook of Experimental Pharmacology* acknowledges several significant milestones in the discipline of Pharmacology. Moreover, it is a testimony to the evolution of the scientific scope and breadth and of the important role the *Handbook* has had in capturing the many advances over this 100-year period. The *Handbook* was founded in 1919 by Arthur Heffter as the "Handbuch der Experimentellen Pharmakologie." As of Volume 50 in 1978, the series was renamed to its current title. The first volume published in English was in 1937, titled "General Pharmacology" and the entire volume was written by Alfred Joseph Clark from the University of Edinburgh. The series has published more than 250 volumes, representing a continuing tradition and commitment to the evolution of pharmacology as a vibrant discipline in generating innovative basic and clinical research that have facilitated the development of safe and effective therapeutics. While originally designed as a handbook, each chapter and the entire volumes are now available electronically and primarily used in this way.

The collection of volumes of the *Handbook of Experimental Pharmacology* over the past 100 years reflects the tremendous growth of the discipline of pharmacology and provides tangible stepping-stones to the many significant advances in the discovery and development of new drugs for a wide range of diseases. The *Handbook* has consistently provided critical and comprehensive reviews of the most significant areas of pharmacological research, written by leading international authorities, and contributing significantly to the tremendous progress evidenced over the past 100 years.

The *Handbook* has captured and disseminated the dynamic nature of the discipline of pharmacology, celebrating and publishing new discoveries from basic understanding of mechanisms of drug action to the delivery of new, safe, and effective therapeutics. Writing in the *Handbook* over 50 years ago, the Nobel Laureate Sir Henry Dale, commented that "Heffter's great Handbuch der Experimentellen Pharmakologie may be regarded, perhaps, as giving some measure of the prodigious growth, during the past half-century, of those areas of scientific knowledge which can properly be regarded now as belonging to the domain of Pharmacology. And to make its full contribution to experimental progress on these lines, pharmacology would now have to work in an increasing intimacy of cooperation, with the accelerating growth of relevant knowledge in physiology,

biochemistry, pathology, and immunology, and, indeed, in any of the more fundamental scientific disciplines." That growth and the "intimacy" predicted by Dale have continued as pharmacology has evolved and embraced those interactions. The imprint of the evolution of pharmacology is strongly reflected in the series, which includes contributions from over 20 Nobel Laureates. These fundamental advances have generated newer and deeper insights into signaling pathways, elucidated our understanding of molecular mechanisms of drug action, while also witnessing remarkable advances in Quantitative Systems and Computational Pharmacology, as well as in enabling technologies such as Pharmacogenomics, Metabolomics, Natural Products, and Drug Delivery Systems, to name just a few.

While it is difficult to cover all the developments in pharmacology over the 100-year period, we hope that this volume will capture much of the progress in pharmacology, hoping to provide a window to some of the past achievements as well as an anticipation of future progress. Perhaps even more importantly, several chapters provide visions for the future of pharmacology. We thank the authors for their contributions to this important volume in the history of this prestigious series, while also expressing our deep appreciation to the many scientists whose passion and commitment to pharmacology have made it a vibrant discipline, translating advances in basic science to safe and effective therapeutics.

We would also like to express our sincere appreciation to Susanne Dathe, Springer Editor for Neurosciences/Pharmaceutical Sciences/Protocols, whose commitment and competence have helped to continue the tradition of this remarkable series, and to the past and current editorial board members who have dedicated time and effort into establishing this series as one of the most recognized publications in pharmacology.

Philadelphia, PA, USA — James E. Barrett
London, UK — Clive P. Page
Mainz, Rheinland-Pfalz, Germany — Martin C. Michel

Contents

Part I A Century of Advances in Pharmacology

Perspectives of Pharmacology over the Past 100 Years 3
James E. Barrett, Clive Page, and Martin C. Michel

Emergent Concepts of Receptor Pharmacology 17
Terry Kenakin

The Evolving Landscape of Cancer Therapeutics 43
Madeha Khan and James Spicer

Monoclonal Antibodies: Past, Present and Future 81
J. Posner, P. Barrington, T. Brier, and A. Datta-Mannan

100 Years of Drug Delivery to the Lungs 143
Federico Lavorini, Francesca Buttini, and Omar S. Usmani

Ion Channel Pharmacology for Pain Modulation 161
Francesco De Logu and Pierangelo Geppetti

Exploiting the Diversity of Ion Channels: Modulation of Ion Channels for Therapeutic Indications 187
Yani Liu and KeWei Wang

Part II Recent Developments and Enabling Technologies in Pharmacology

Genetically Encoded Fluorescent Calcium and Voltage Indicators 209
Irene Mollinedo-Gajate, Chenchen Song, and Thomas Knöpfel

Mechanistic Image-Based Modelling: Concepts and Applications 231
Denis Menshykau and Simon Tanaka

Pharmacometabonomics: The Prediction of Drug Effects Using Metabolic Profiling 263
Jeremy R. Everett

The Microbiome and Its Potential for Pharmacology 301
Aries Chavira, Pedro Belda-Ferre, Tomasz Kosciolek, Farhana Ali, Pieter C. Dorrestein, and Rob Knight

Harnessing Human Microphysiology Systems as Key Experimental Models for Quantitative Systems Pharmacology 327
D. Lansing Taylor, Albert Gough, Mark E. Schurdak, Lawrence Vernetti, Chakra S. Chennubhotla, Daniel Lefever, Fen Pei, James R. Faeder, Timothy R. Lezon, Andrew M. Stern, and Ivet Bahar

Part III The Future of Pharmacology

The Future of Clinical Trial Design: The Transition from Hard Endpoints to Value-Based Endpoints 371
Matthijs D. Kruizinga, Frederik E. Stuurman, Geert J. Groeneveld, and Adam F. Cohen

Placebos and the Placebo Effect in Drug Trials 399
Paul Enck and Sibylle Klosterhalfen

Pharmacoepidemiology 433
Nicholas Moore, Patrick Blin, and Cécile Droz

Why Are New Drugs Expensive and How Can They Stay Affordable? . . . 453
Basma Hammel and Martin C. Michel

Perspectives of Pharmacology over the Past 100 Years

James E. Barrett, Clive Page, and Martin C. Michel

Contents

1 Introduction 5
2 The Emergence of Pharmacology in Germany: Rudolf Buchheim 5
2.1 Schmiedeberg's Contribution to the Development of Pharmacology 7
3 The Spread of Pharmacology 8
3.1 Otto Krayer and the Origins of Behavioral Pharmacology 10
4 Pharmacology Through 100 Years and a Future Perspective 12
5 Conclusions 14
References 14

Abstract

It is fitting that the 100th anniversary of the *Handbook of Experimental Pharmacology* celebrates not only its founding but also the founding of experimental pharmacology as both had their beginnings in Germany. Founded in 1919 by Arthur Heffter (1859–1925) as the "Handbuch der Experimentellen Pharmakologie" and renamed to its current title in 1937, the Handbook has continued to capture the emergence and developments of experimental pharmacology since the initial systematic work of Rudolf Buchheim and his student Oswald Schmiedeberg. Heffter, the first Chairman of the German Society of Pharmacology, was also responsible for isolating mescaline as the active

J. E. Barrett (✉)
Center for Substance Abuse Research, Lewis Katz School of Medicine at Temple University, Philadelphia, PA, USA
e-mail: jeb92@drexel.edu

C. Page
Sackler Institute of Pulmonary Pharmacology, Institute of Pharmaceutical Science, King's College London, London, UK

M. C. Michel
Department of Pharmacology, Johannes Gutenberg University, Mainz, Germany

J. E. Barrett et al. (eds.), *Concepts and Principles of Pharmacology*,
Handbook of Experimental Pharmacology 260, https://doi.org/10.1007/164_2019_334

psychedelic component from the peyote cactus, thereby initiating a series of studies along with an Institute that, much like the Handbook and the discipline of pharmacology, continues to discover and disseminate new findings to this day. These early endeavors to establish pharmacology as a viable and valuable contributor to the medical sciences met with considerable resistance and challenges. However, the persistence and dedication of these early pharmacologists placed pharmacology on a firm foundation from which to spread this discipline globally, leading ultimately to our current understanding of the principles of drug action and with an impact likely unanticipated by these founding scientists. Summarizing the beginnings of these efforts and their early spread to other countries provides an appropriate context in which to document the many contributions pharmacological research has made over the past 100 years and provide an opportunity to anticipate expectations around its future developments.

Keywords
Buchheim · Heffter · History of German pharmacology · Krayer · Schmiedeberg

1 Introduction

Pharmacology is a discipline with a rich history that, since its emergence during some rather challenging times, has exerted a tremendous impact as a vibrant science with an exciting and promising future. Some have argued that pharmacology is the oldest discipline in the health sciences (Norton 2005), with medicines derived from plants used since prehistoric times. In China, the use of various plant sources such as herbs and herbal remedies dates back some 3,000 years, imbuing traditional Chinese medicine with an interesting and extensive history that continues with great momentum with the current interest in natural products. Despite the very early realization that compounds derived from natural sources could have therapeutic value, the term "pharmacology" was not used in print until the seventeenth century (Norton 2005). Early practitioners such as Theophrastus Bombastus zu Hohenheim (a.k.a. Paracelsus) attempted to determine the active ingredients of these early preparations and formulated the earliest concept of dose-response functions by suggesting that the dose determines potential therapeutic value or toxicity. Writing the "Third Defense Pertaining to the Description of the New Prescriptions" in 1564, Paracelsus asked: "What is there that is not poison, all things are poison and nothing is without poison. Solely the dose determines that a thing is not a poison" (Deichmann et al. 1986). Although Paracelsus is acknowledged for this remarkable insight on the importance of dose, it is not often recognized that he also described the first recognition of "species specificity" in the context of his general principle of toxicology, also described by Deichmann et al.:

> For instance, the food placed on the table: if eaten by a man, it becomes the flesh of man; if ingested by a dog, it is converted into dog flesh and in the cat, to cat flesh. With medicine, it is the same, its fate depends on the species or what you do with it. It is possible that something good will cause harm, just as it is possible that something harmful may become beneficial.

2 The Emergence of Pharmacology in Germany: Rudolf Buchheim

The research on curare and carbon monoxide poisoning in the early part of the nineteenth century helped to establish Francois Magendie and his student, Claude Bernard, as important precedents to what was to become pharmacology. Although both were physiologists, their techniques and some of the principles stemming from their studies were useful in the development of pharmacology (Parascandola 1980). These early precedents of pharmacology reached fruition in the mid-nineteenth century, with several significant developments in Germany that included the founding of the world's first Institute of Pharmacology ("Pharmakologisches Institut") established at the University of Dorpat (now Tartu) by Rudolf Buchheim in 1847. Buchheim was a professor of Materia Medica, Dietetics, and History of Medicine. As Trendelenburg (1998) has stated in his review of "Pharmacology in

Germany," "the cradle for our hobby was located in the very improbable site of a university in Tsarist Russia, situated in a then largely German-speaking town of today's Estonia." For several years, the laboratories of this "institute" were housed in the basement of Buchheim's home (Koch-Weser and Schechter 1978; Rang 2006). In 1860 the Institute, with its focus on experimental pharmacology, was moved into a large building specifically constructed to include the Institute of Pharmacology. It was this stimulus, initiated by Buchheim and developed by his students, that fostered the emergence of pharmacology as a well-defined discipline.

At the time of the founding of the first Institute of Pharmacology, the contribution that pharmacology had to make to medicine was increasingly questioned. Pharmacology had been dropped from the final medical examinations in most German states, and physicians were being counseled to forget as soon as possible what they might remember about drugs from their lectures or books (Meyer 1922). The development of pharmacology as a scientific discipline comparable to physiology or pathology received little or no support from the medical establishment and only gradually gained acceptance and a firm foundation in medical school education and research. The attitude of most clinicians was expressed by the famous Viennese surgeon T. Billroth:

> Considering how little he has to teach and that half of what he teaches is superfluous, it is difficult to keep a professor of pharmacology busy in a full-time teaching position. What is needed is merely a short review of the most important drug groups and an experimental demonstration of the most intensive poisons. This can be done in 3–4 h. The students should not be burdened with more lectures. How to use drugs one can only learn in the clinics anyway (cited in Koch-Weser and Schechter 1978).

Even when Buchheim convincingly and compellingly defended pharmacology, he concluded with this statement: "What can a man who gives his whole strength to pharmacologic research achieve today under the best of circumstances? A professorship with a minimum salary and an empty auditorium."

Rudolf Buchheim was born into a time when an appreciation of scientific methods and thinking were growing and some of the fundamentals of modern medicine were being established. These included the work of Pasteur, who opened the path to microbiology, and Darwin who developed the theory of evolution, while Virchow published on cellular pathology, and Helmholtz and Mayer formulated the law of conservation of energy along with other principles of vision and perception. Chemistry, physics, and physiology had advanced relatively rapidly, and there was a growing need for a scientific foundation of therapeutics (Habermann 1974). Buchheim was convinced that pharmacology should progress beyond the descriptive approach of materia medica and, by experimentation, explore how drugs cause their effects, regardless of whether those effects are beneficial or undesirable. Buchheim introduced a principle that would have a considerable bearing on the discipline of pharmacology – the "Natural System of Drugs" – which was the formulation of a system for the classification of drugs based on elucidating their mode of action. Buchheim essentially envisioned and postulated a new, independent science, formulating objectives and establishing a methodological approach while also

expressing a reaction to apparent objections from other disciplines surrounding the existing skepticism of pharmacology as a science when he wrote:

> If we translate our often obscure ideas about drug actions into an exact physiological language: this should, without doubt, be a considerable achievement. However, scientific cognition of the action of a given drug would imply our ability to deduce each of its actions from its chemical formula. The new era of pharmacology will bear its date not from the discovery of chloral hydrate, but from that time when pharmacology will cease to ornate itself by the waste of other disciplines; when pharmacology with its own area and aided by related sciences, will become equivalent to its sisters, chemistry and physiology (cited in Habermann 1974).

2.1 Schmiedeberg's Contribution to the Development of Pharmacology

An early student of Buchheim was Oswald Schmiedeberg (1838–1921) who was strongly influenced by Buchheim's teaching and with whom Schmiedeberg obtained his Dr. of Medicine degree in 1866 for his work determining chloroform in the blood. Schmiedeberg was largely responsible for pursuing and disseminating Buchheim's concepts and methods into pharmacological research and continuing the movement initiated by Buchheim. In the words of Koch-Weser and Schechter, writing in 1978:

> One century ago pharmacology was an antiquated, denigrated and waning discipline content with transmitting impressionistic and largely erroneous dictums. In one generation one man in one city redefined its tasks, demonstrated its experimental methods and trained its work force … Schmiedeberg brought scientific pharmacology into being.

Schmiedeberg's profound influence on pharmacology was accomplished through his excellence as a researcher, as a passionate and dedicated teacher, and as a prolific writer. His experimental scope was extensive and included work on muscarine, isolating the alkaloid from the mushroom *Amanita muscaria*. Its analysis through the experimental study of its actions remains a model of the pharmacological investigation of natural substances. Schmiedeberg's interest in muscarine led to investigations into the action of nicotine on the heart and to later studies on cardiac innervation and on the whole pharmacology of the autonomic nervous system. Schmiedeberg also studied the actions of caffeine and related xanthines on striated muscles as well as their diuretic effects. Schmiedeberg and his Strassburg group of researchers and students performed the first experimental studies on the pharmacology of most of the important drugs and poisons known at that time. Much of their work was published in the *Archiv für Experimentelle Pathologie und Pharmakologie*, founded in 1873 by Schmiedeberg, B. Naunyn, and E. Klebs (this journal is now the *Naunyn-Schmiedeberg's Archives of Pharmacology*). In this way these individuals succeeded in giving pharmacology a solid scientific foundation and assured its position alongside the other biomedical sciences (Starke 1998).

Schmiedeberg's work in this institute during the next five decades was largely responsible for the initial rise of pharmacology to a respected scientific discipline and as an indispensable foundation for medical practice. Schmiedeberg unquestionably has had a profound impact in forming and cultivating the discipline of pharmacology, not only due to his expansive and pioneering experimental work but also for his training of over approximately 200 pharmacologists from over 20 countries during his 46 years at the University of Strassburg. At the time of his death in 1921, an astonishing number of over 40 chairs of pharmacology were held by his students around the world (Koch-Weser and Schechter 1978).

3 The Spread of Pharmacology

The seeds of experimental pharmacology as a distinct discipline with a different focus from that of physiology were founded and developed in Germany, but the growth and continuing shaping of pharmacology rapidly became global. Many of the discoveries over the past hundred-year period have been documented in articles which, themselves, provide testimony to the seminal discoveries made by the growing number of academic pharmacologists. For example, Rubin (2007) has written a comprehensive review on the history of great discoveries in pharmacology to celebrate the centennial anniversary of the founding of the American Society of Pharmacology and Experimental Therapeutics (ASPET), while Cuthbert (2006) traced the history of the British Pharmacological Society to celebrate its 75th anniversary. One of Schmiedeberg's students was J.J. Abel who brought pharmacology to the United States. Abel occupied the first professorship of experimental pharmacology in the United States at the University of Michigan in 1891 and subsequently moved to Johns Hopkins University 1893 to hold the first Chair of Pharmacology in the United States. Johns Hopkins was established in 1876, modeled after the German research universities, and one of the missions for medicine was to focus on the physiological action of drugs. The physiologist Henry Newell Martin, a member of the Johns Hopkins faculty, in a lecture to the Medical and Chirurgical Faculty of Maryland in 1885, selected pharmacology as his topic "because I believe that it is destined in the near future to acquire an importance in regard to therapeutics which is not properly appreciated" (Parascandola 1992). Abel had spent the 1883 academic year working in Martin's laboratory prior to going to Strassburg to work under Schmiedeberg. The contributions of Abel to the development of pharmacology in the United States were substantial as his efforts greatly expanded pharmacology research and education in several universities; pharmacology also spread and established a firm footing with the movement of pharmacologists into the growing pharmaceutical industry and into the federal government with the founding of the Food and Drug Administration (FDA) in 1930. Abel contributed significantly to the development of pharmacology in the United States, principally by his founding of the American Society of Pharmacology and Experimental Therapeutics (ASPET) and the *Journal of Pharmacology and Experimental Therapeutics*. These many initiatives are thoroughly researched and

documented in the volume *The Development of American pharmacology: John J. Abel and the Shaping of a Discipline* by Parascandola (1992). Among his many other contributions, Abel introduced significant participation in laboratory work as part of the medical school curriculum but, quite ironically, was opposed to the creation of a Ph.D. in pharmacology which, at Johns Hopkins, was not established until 1969, more than 30 years following his retirement (Parascandola 1992).

The launching of the *Journal of Pharmacology and Experimental Therapeutics* by Abel provided a much-needed outlet for pharmacological papers in the English language, with the earlier volumes of that journal offering effective hospitality to papers by British scientists pursuing pharmacology (*British Medical Journal*, "A Milestone in Pharmacology" 1946). In a remarkable parallel to the early difficulties of establishing pharmacology in Germany, there was hesitation in England for according pharmacology full academic recognition and the opportunities for publication had been limited to weekly medical journals and *The Journal of Physiology*. The editor of that journal, J. N. Langley, was showing a steadily increasing reluctance to accept papers which could be regarded as pharmacology based on the view that pharmacology was encroaching unduly on his space (Cuthbert 2006; Forward, Dale 1946). The result of this apparent impasse for Abel was to provide joint editorial control of the *Journal of Pharmacology and Experimental Therapeutics* in 1912 to both himself and to A.R. Cushny with an editorial board comprised of a number of British Pharmacologists.

In Britain, pharmacology also had some difficulties getting established, not just with regard to publishing. Those difficulties were attributed to a number of factors including the fact that trainee doctors were taught little about drugs in their medical school curriculum and, what was taught, was more aligned with pharmacy and prescription writing than pharmacology (Cuthbert 2006). A few chairs of Materia Medica and Therapeutics existed in universities in Scotland, but it was not until 1905 that such a position was established in England when a chair in pharmacology was established for Arthur Cushny at University College London. Cushny, originally from Scotland and one of the pioneer pharmacologists of England, also a pupil of Schmiedeberg, had been Abel's successor at Ann Arbor, Michigan. After approximately 12 years at the University of Michigan, Cushny returned to England in 1905 to be the first holder of a new Chair of Materia Medica and Pharmacology at University College, London (Abel 1926; Parascandola 1992). Abel's (1926) tribute to Cushny on the occasion of his premature death by a cerebral hemorrhage is a remarkable testimony to his life and provides a detailed account of Cushny's prolific contributions to the science of pharmacology and medicine, including a contribution to Heffter's *Handbuch der Experimentellen Pharmakologie* on the nitrites, the members of the atropine group, and ergot. Pharmacology rapidly established a foothold in England, led in large part by both Cushny and W.E. Dixon who held a part-time chair of Materia Medica and Pharmacology at King's College London, and moved subsequently to the position of Reader in Pharmacology at Cambridge in 1919 (Cuthbert 2006).

One of the more dominant figures in this early period of pharmacology in Britain was H.H. Dale. Dale, the recipient of a Nobel Prize in 1936, accepted a position at

the Wellcome Physiological Research Laboratories in 1904 where his research involved substances found in ergot and also included the study of the role of histamine in allergic reactions (Cuthbert 2006). A third individual working around this time was J.A. Gunn who became a professor of pharmacology at Oxford in 1912. These three individuals – Dixon, Dale, and Gunn – were responsible for undertaking the formation of the British Pharmacological Society (BPS) that held its first formal meeting in 1931 (Cuthbert 2006). Writing in the Forward of the inaugural issue of the *British Journal of Pharmacology and Chemotherapy*, Sir Henry Dale (1946) commented on the "growing volume and importance of pharmacological publication … seen throughout the scientific world … [adding that] pharmacology has rapidly risen to a major rank among the group of scientific disciplines which come from within the scope of experimental medicine" (p.1).

3.1 Otto Krayer and the Origins of Behavioral Pharmacology

One other German pharmacologist warrants recognition among those that fostered and further developed pharmacology and added branches to the foundation established in Germany that were expanded to the United States and Great Britain. Otto Krayer (1899–1982) had embarked on a promising career in pharmacology that was initiated when he was a medical student working with Paul Trendelenburg at the University of Freiburg. When Trendelenburg became Chair of Pharmacology at the highly prestigious University of Berlin in 1927, Krayer moved with him and continued to work in his laboratory. Unfortunately, Trendelenburg became ill with tuberculosis in 1930 and passed away the following year. The Chair at Berlin could only be filled by someone who held an existing chair, and when the new Chair, Professor Heubner, from the University of Göttingen arrived, he appointed Krayer Professor Extraordinarius of Pharmacology and Toxicology (Anderson 2005). In 1933, Krayer was offered, but turned down the offer to become the Chair of Pharmacology and Toxicology at the Medical Academy of Düsseldorf with the reason that the Jewish incumbent Chair, Philipp Ellinger, had been removed according to Nazi law (Anderson 2005). The decision by Krayer, based on his personal convictions that this was an injustice, resulted in Krayer subsequently being banned from teaching, using university or state libraries, while also being forbidden to enter any government academic institution or scientific facility (Trendelenburg 1978). This compassionate decision effectively ended Krayer's career in Germany (Anderson 2005; Goldstein 1987), but he was to continue to have a significant long-term impact on pharmacology. At the time of these events, Krayer was completing Volume 2 of Trendelenburg's *Die Hormone* that was unfinished when Trendelenburg passed away. Friends clandestinely brought him books and journals from the library for him to finish his work which, when completed, resulted in Krayer's providing the proofs to Springer-Verlag and his almost immediate departure from Germany (Dews 1983).

Initially, Krayer took temporary positions at University College London and the American University in Beirut. In 1936 Krayer was offered and accepted an

appointment in the Pharmacology Department at Harvard Medical School. He spent considerable time developing the curriculum and teaching until he was offered a position of Chair of Pharmacology at the Peiping Union Medical College for Columbia University in China. His temptation to accept this position, however, was countered when there was a petition by the medical students to the administration that Krayer remain at Harvard. Krayer was then offered and accepted the Chair of the Pharmacology Department with tenure (Dews 1983; Anderson 2005).

Krayer was interested primarily in molecular mechanisms of drug action and sought to evaluate further those findings in integrated physiological systems. In the 1950s Krayer was able to recruit a large number of faculty, one of whom was U. Trendelenburg, the son of Krayer's first mentor. Another member of the Department who was recruited during this time was Peter B. Dews. As Dews has written (Dews 1978), Krayer, in his effort to embrace diverse approaches to pharmacological research, sent him to meet with B.F. Skinner, the well-known behaviorist in the Psychology Department at Harvard who was conducting experiments with pigeons. Skinner and his students and colleagues were using automated experimental equipment and recording behavior on "cumulative recorders" in real time. Skinner had conducted a few prior experiments in the late 1930s with caffeine and benzedrine but had not further pursued those studies. At the time of Dews' visit to the Harvard Pigeon Lab, Skinner and his colleague, C.B. Ferster, were recording responding of pigeons pecking an illuminated plexiglass key. The key pecking behavior was controlled by a "schedule of reinforcement" that provided access to grain depending on prearranged contingencies such as the passage of time or the number of key pecks. The ability to record behavior objectively and quantitatively over lengthy time periods captured Dews interest and set the stage for the founding of the discipline of behavioral pharmacology.

Pharmacology has traditionally focused on and found order in the study of drug effects on relatively isolated pieces of tissue or organ systems. Progress in behavioral pharmacology, as in other areas of pharmacology, had to await the development and useful integration of suitable techniques that not only permitted but promoted the intensive study of the behavioral effects of drugs. The operant conditioning techniques that employed various schedules of reinforcement to control behavior that were so thoroughly and extensively explored by Ferster and Skinner (1957) provided the impetus for Dews to embrace those techniques and incorporate them into research to characterize the effects of drugs on behavior. Research in behavioral pharmacology has demonstrated that comparable order also exists when drugs are examined at the level of integrated behavior (Barrett 1980). Behavioral pharmacology began as a formal scientific discipline with the finding by P. B. Dews (1955) that the behavioral effects of drugs depended on the specific ways in which behavior was controlled by the schedule of reinforcement. The foresight by Krayer and the implementation by Dews and his colleagues launched an entirely new discipline within the field of pharmacology that coincided with the discovery and introduction of several new psychotherapeutic drugs to treat anxiety, depression, and schizophrenia – developments that necessitated an expansion of pharmacologists into the

pharmaceutical industry and fostered a multitude of new behavioral assays for drug discovery.

4 Pharmacology Through 100 Years and a Future Perspective

The discipline of pharmacology has made many significant advances over the past 100 years and will undoubtedly continue to do so in the future. It is now the "parent" of many subdisciplines that include behavioral pharmacology but also biochemical pharmacology, neuropharmacology, molecular pharmacology, pulmonary pharmacology, immunopharmacology, cardiovascular pharmacology, and other areas of importance such as pharmacoepidemiology, pharmacogenetics, and quantitative and systems pharmacology to name just a few. These areas of pharmacology and pharmacological research are an enduring testimony to the visionary and pioneering efforts of the early pharmacologists Buchheim and Schmiedeberg and their students who persevered in their efforts to establish pharmacology as a scientific endeavor within the field of medicine despite some formidable challenges.

Tremendous advances have been made in the quantitative analysis of drug-receptor interactions, in identifying mechanisms of signaling, and in our ability to predict how drug molecules bind to their protein target (Rachman et al. 2018; Wootten et al. 2018), areas of research whose history is rich in detail beginning with the work of A.V. Hill with subsequent contributions by other legendary figures in pharmacology including A.J. Clark, J.H. Gaddum, and H.O. Schild working at the University College London and the University of Edinburg (Colquhoun 2008; Rang 2006; Vallance and Smart 2006). A brief summary of the first 50 years of pharmacological research starting at the turn of the twentieth century yields evidence of remarkable advances in quantitative pharmacology with the clarification of agonist-antagonist relationships and the existence of partial agonists, along with distinctions between affinity and efficacy. The continued elaboration of these concepts, together with advances that were to follow, such as those of radioligand binding, allosteric and orthosteric modulation, inverse agonists, biased agonism at G protein-coupled receptors (GPCRs), and second messengers – all within the context of receptor theory – has provided the foundation for the study of receptor function at the biochemical and molecular level and the ability to further pursue what Kenakin (2019) has termed "analytical pharmacology," based on initial formulations of the Nobel Laureate pharmacologist Sir James Black. These developments only partially populate the expansive domain of pharmacological research and are complemented by the emergence of other areas that include groundbreaking work in crystallography that enables the stabilization of GPCRs in active conformations, thereby enabling the understanding of binding and aiding in drug development. The complexity of the signaling mechanisms activated by GPCRs and the interaction with cellular proteins represent new challenges to better understand the pharmacology of this important class of proteins that have yielded the rich pharmacology of drugs to treat a wide variety of disorders ranging from hypertension, schizophrenia, and depression. A

detailed review of significant developments within GPCRs including oligomerization and accessory proteins and the processes of desensitization, internalization, and trafficking has been provided by Hill (2006), with suggestions on the further challenges and opportunities existing for this receptor family.

Pharmacology has long been a discipline working closely with, contributing to, and benefiting from other areas in the biomedical sciences, such as physiology, biochemistry, and cell biology. As Lohse (1998) has commented, "while other biomedical sciences are trying to explain the world (of the human body), pharmacologists try to alter it," a statement somewhat similar to that of Kenakin (this volume) that "pharmacology is the chemical control of physiology." Many therapeutic drugs that have a pronounced physiological impact, such as the psychotherapeutic agents, were discovered before their true mechanism of action – the "target" – was identified. Over the past few decades, the pharmaceutical industry has placed considerable emphasis on "target identification" and validation based, in part, on the belief that focusing on a single selective target, presumably associated with the disease, would minimize "off-target" side effects. This approach has been balanced, more recently by the return to phenotypic screening of cells or whole animals (Moffat et al. 2017) and by the focus on multitarget approaches, recognizing that most diseases are complex and involve multiple targets, necessitating "polypharmacological" or multidrug combinations (Weiss and Nowak-Sliwinska 2017). This is certainly true of the heterogeneously complex disorders such as autism and schizophrenia, as well as diseases of the immune system and the targeting of tumors in oncology. Presumed pharmacological selectivity often vanishes under the scrutiny of new techniques and deeper knowledge of other potential targets. Even when a drug has been on the market for some time, it is not uncommon to identify a new mechanism and/or a new therapeutic utility for that compound. Such is the case, for example, with the anthelminthic drug mebendazole, discovered in the 1970s. Recently, using a combination of novel, rapid computational proteochemometric methods coupled to cell-based assays, it was shown that mebendazole should be considered for use in combination with trametinib as a therapeutic option for patients with specific types of melanomas (Simbulan-Rosenthal et al. 2017). This "repurposing" or "repositioning" of existing drugs has captured a great deal of attention to reposition drugs either as combination therapeutics or as stand-alone drugs for other disorders (Frail et al. 2015; Kumar et al. 2019; Polamreddy and Gattu 2019). These efforts clearly require the integration of other disciplines from computational pharmacology and other specialties, further emphasizing the alignment and integration of other techniques and methodologies into pharmacological research.

Significant advances in pharmacology have been made in other areas such as drug metabolism with the focus on metabolomics, as well as pharmacogenomics and pharmacoepidemiology (see chapters in this volume by Chavira et al. 2019; Everett 2019; Moore et al. 2019). The development of in silico models in drug development, coupled to the use of microfluidic devices such as "organs-on-chips," and the incorporation of induced pluripotent stem cells have been enthusiastically integrated into drug discovery and development (Borenstein 2016; Esch et al. 2015; Piñero et al. 2018; Suh 2016) providing relatively rapid assessments of the actions of

various drugs to predict efficacy as well as toxicological actions. The area of quantitative and systems pharmacology incorporates these areas with the perspective of developing more personalized or precision medicine and the integration of artificial intelligence (AI) to deal with complex data sets (see Taylor et al. this volume and Wishart 2016).

5 Conclusions

Progress in pharmacology over the past 100 years has been monumental. However, significant challenges remain in a number of therapeutic areas with significant unmet medical need, not the least of which is in the need for new antibiotics, as resistance to existing drug classes increases. Other areas of unmet need include those of identifying medications to treat substance abuse disorders and pain where these conditions may, under some conditions, be related. The great promise offered by many of the more recent developments suggests that pharmacology will become even more important over the next 100 years. Many of these developments have been spawned in academic research laboratories which will likely to continue to play an important role for pharmacologists and the pharmaceutical industry. As the pharmaceutical industry has consolidated and discontinued a focus on certain therapeutic areas, particularly in the area of neuroscience, much of the effort has shifted to academic research aligned with government funding to fill the gap for new targets and potential therapeutics. Finally, as the discipline of pharmacology has become more diverse and technologically rich, there is a continuing need to emphasize the importance of training, education, and research in the fundamental principles of pharmacology, an area that was emphasized in the early days of academic laboratories in Germany and which continues to be an important contemporary priority.

References

A Milestone in Pharmacology (1946) Br Med J 1:727–728

Abel JJ (1926) Arthur Robertson Cushny and pharmacology. J Pharmacol Exp Ther 27:265–286

Anderson R (2005) The singular moral compass of Otto Krayer. Mol Interv 5:324–329

Barrett JE (1980) Behavioral pharmacology: recent developments and new trends. Trends Pharmacol Sci 1:215–218

Borenstein JT (2016) Organs-on-chips: how microsystems technology can transform the drug development process. IEEE Pulse 7:22–26

Chavira A, Belda-Ferre P, Kosciolek T, Farhana A, Dorrestein C, Knight R (2019) The microbiome and its potential for pharmacology. Handb Exp Pharmacol. https://doi.org/10.1007/164_2019_317

Colquhoun D (2008) The quantitative analysis of drug-receptor interactions: a short history. Trends Pharmacol Sci 27:149–157

Cuthbert AW (2006) A brief history of the British Pharmacology Society. Br J Pharmacol 147: S2–S8

Deichmann WB, Henschler D, Holmstedt B, Keil G (1986) What is there that is not poison? A study of the Third Defense by Paracelsus. Arch Toxicol 58:207–213
Dews PB (1978) Origins of behavioral pharmacology. Life Sci 22:1115–1122
Dews PB (1955) Studies on behavior. I. Differential sensitivity to pentobarbital of pecking performance in pigeons depending on the schedule of reward. J Pharmacol Exp Ther 113:393–401
Dews PB (1983) Otto Krayer 1899-1982. Trends Pharmacol Sci 4:143–146
Esch EW, Bahinski A, Huh D (2015) Organs-on-chips at the frontiers of drug discovery. Nat Rev Drug Discov 14:248–260
Everett JR (2019) Pharmacometabonomics: the prediction of drug effects using metabolic profiling. Handb Exp Pharmacol. https://doi.org/10.1007/164_2019_316
Ferster CB, Skinner BF (1957) Schedules of reinforcement. Appleton-Century-Crofts, New York
Foreword (1946) Br J Pharmacol Chemother 1:1–3
Frail DE, Brady M, Escott KJ, Holt A, Sanganee HJ, Pangalos MN, Watkins C, Wegner CD (2015) Pioneering government-sponsored drug repositioning collaborations: progress and learning. Nat Rev Drug Discov 14:833–841
Goldstein A (1987) Otto Krayer: a biographical memoir. Biogr Mem Natl Acad Sci 57:11–225
Habermann ER (1974) Rudolf Buchheim and the beginning of pharmacology as a science. Annu Rev Pharmacol 14:1–9
Hill SJ (2006) G-protein-coupled receptors: past, present and future. Br J Pharmacol 147:S27–S37
Kenakin T (2019) Analytical pharmacology: how numbers can guide drug discovery. ACS Pharmacol Transl Sci 2:9–17
Koch-Weser J, Schechter PJ (1978) Schmiedeberg in Strassburg 1872-1918: the making of modern pharmacology. Life Sci 22:1361–1372
Kumar R, Harilal S, Gupta SV, Jose J, Thomas Parambi DG, Uddin MS, Shah MA, Mathew B (2019) Exploring the new horizons of drug repurposing: a vital tool for turning hard work into smart work. Eur J Med Chem 182:111602
Lohse MJ (1998) The future of pharmacology. Trends Pharmacol Sci 19:198–200
Meyer HH (1922) Oswald Schmiedeberg. Naturwissenschaften 10:105–107
Moffat JG, Vincent F, Lee JA, Eder J, Prunotto M (2017) Opportunities and challenges in phenotypic drug discovery: an industry perspective. Nat Rev Drug Discov 16:531–543
Moore N, Blin P, Droz C (2019) Pharmacoepidemiology. Handb Exp Pharmacol. https://doi.org/10.1007/164_2019_318
Norton S (2005) Origins of pharmacology in the 16th century. Mol Interv 5:144–148
Parascandola J (1980) Reflections on the history of pharmacology: the 1980 Kremers Award Address. Pharm Hist 22:131–140
Parascandola J (1992) The development of American pharmacology: John J. Abel and the shaping of a discipline. Johns Hopkins, Baltimore
Piñero J, Furlong LI, Sanz F (2018) In silico models in drug development: where we are. Curr Opin Pharmacol 42:111–121
Polamreddy P, Gattu N (2019) The drug repurposing landscape from 2012-2017: evolution, challenges and possible solutions. Drug Discov Today 24:789–795
Rachman M, Barril X, Hubbard RE (2018) Predicting how drug molecules bind to their protein targets. Curr Opin Pharmacol 42:34–39
Rang HP (2006) The receptor concept: pharmacology's big idea. Br J Pharmacol 147:S9–S16
Rubin RP (2007) A brief history of great discoveries in pharmacology: in celebration of the Centennial Anniversary of the American Society of Pharmacology and Experimental Therapeutics. Pharmacol Rev 59:289–359
Simbulan-Rosenthal CM, Dakshanamurthy S, Gaur A, Chen Y-S, Fang HB, Abdussamad M, Zhou H, Zapas J, Calvert V, Petricoin EF, Atkins MB, Byers SW, Rosenthal DS (2017) The repurposed anthelmintic mebendazole in combination with trametinib suppresses refractory $NRAS^{Q61K}$ melanoma. Oncotarget 8:12576–12595

Starke K (1998) A history of Naunyn-Schmiedeberg's archives of pharmacology. Naunyn Schmiedebergs Arch Pharmacol 358:1–109
Suh W (2016) A new era of disease modeling and drug discovery using induced pluripotent stem cells. Arch Pharm Res 40:1–12
Trendelenburg U (1978) Remembrances. Life Sci 22:113–114
Trendelenburg U (1998) Pharmacology in Germany. Trends Pharmacol Sci 19:196–198
Vallance P, Smart TG (2006) The future of pharmacology. Br J Pharmacol 147:S304–S307
Weiss A, Nowak-Sliwinska P (2017) Current trends in multidrug optimization: an alley of future treatment of complex disorders. SLAS Technol 22:254–275
Wishart DS (2016) Emerging applications of metabolomics in drug discovery and precision medicine. Nat Rev Drug Discov 15:473–484
Wootten D, Christopoulos A, Marti-Solano M, Babu NM, Sexton PM (2018) Mechanisms of signaling and biased agonism in G protein coupled receptors. Nat Rev Mol Cell Biol 19:638–653

Emergent Concepts of Receptor Pharmacology

Terry Kenakin

Contents

1 Introduction 18
2 Shots in the Dark: Null Methods in Pharmacology 18
3 Recombinant Systems Redefine Receptor Pharmacology 19
4 Binding Gives Way to Functional Experiments 19
5 Understanding Agonism: The Black/Leff Operational Model 20
6 The Shift from Orthosteric to Allosteric Drug Action 22
7 The Move from Parsimonious Models to Dynamic Models of Receptor Conformation ... 25
8 Allosteric Probe Dependence: Biased Receptor Signaling 28
9 Structure: Receptors Show Themselves 31
10 Genetics and Computer Science Impact Pharmacology 32
11 Conclusion 35
References 35

Abstract

Pharmacology, the chemical control of physiology, emerged as an offshoot of physiology when the physiologists using chemicals to probe physiological systems became more interested in the probes than the systems. Pharmacologists were always, and in many ways still are, bound to study drugs in systems they do not fully understand. Under these circumstances, null methods were the main ways in which conclusions about biologically active molecules were made. However, as understanding of the basic mechanisms of cellular function and biochemical systems were elucidated, so too did the understanding of how drugs affected these systems. Over the past 20 years, new ideas have emerged in the field that have completely changed and revitalized it; these are described herein. It will be seen how null methods in isolated tissues gave way to, first

T. Kenakin (✉)
Department of Pharmacology, University of North Carolina School of Medicine, Chapel Hill, NC, USA
e-mail: kenakin@email.unc.edu

J. E. Barrett et al. (eds.), *Concepts and Principles of Pharmacology*,
Handbook of Experimental Pharmacology 260, https://doi.org/10.1007/164_2019_297

biochemical radioligand binding studies, and then to a wide array of functional assay technologies that can measure the effects of molecules on drug targets. In addition, the introduction of molecular dynamics, the appreciation of the allosteric nature of receptors, protein X-ray crystal structures, genetic manipulations in the form of knock-out and knock-in systems and Designer Receptors Exclusively Activated by Designer Drugs have revolutionized pharmacology.

Keywords
Drug discovery · Pharmacodynamics · Pharmacology history

1 Introduction

Pharmacology is a unique scientific discipline in that it incorporates many 'pure' scientific disciplines such as genetics, chemistry, biochemistry and physiology. Actually, to be more specific, not so much incorporates as 'borrows from'. Therefore pharmacology has evolved as a uniquely separate melding of chemistry and physiology. Since this is the case, advances in any of these contributing sciences, either from the point of technology or understanding of the basic science, necessarily impacts pharmacology and thus changes the discipline. Pharmacology originated from a basic practical need, i.e. humans require chemicals to prevent, diagnose and cure disease and improve health. Therefore a system had to evolve to promote this process. This chapter will discuss the various influences that have shaped receptor pharmacology to where it is at this present time.

2 Shots in the Dark: Null Methods in Pharmacology

Pharmacology began, and in some ways still is, a science operating in a sea of uncertainty, i.e. pharmacologists often do not fully understand the systems they study. Thus from the early days of receptor pharmacology, isolated tissue preparations from animals were used to characterize what then was only a concept, i.e. the 'receptor' was not biochemically characterized nor available for physical study but simply was a unifying concept for pharmacologists and medicinal chemists used to order and modify physiology (Rang 2009). The basic idea behind the use of isolated tissues was that drugs interact with receptors to induce a visible response and the tissue simply amplifies this initial signal (defined by Stephenson (1956) as 'stimulus') to allow quantification of drug response through an undefined biochemical cascade referred to as the 'stimulus-response' mechanism of the cell. The tacit assumption in this process is that the amplification process is uniform and thus ratios of activity seen through the amplified signal accurately reflected ratios of pharmacological effect at the receptor. Such concepts led to some extremely useful tools in drug discovery such as the agonist potency ratio (PR), which allows quantification of relative agonist activity in test systems that ostensibly allowed prediction of

similar activity in all systems. The underlying assumption in this scheme is that stimulus-response cascades are monotonic (only one '*y*', tissue response, for every '*x*', drug concentration); only such systems ensure the accurate translation of receptor events to observed tissue response in a predictable manner.

There are two inherent weaknesses in this historical scenario. The first is obvious in that animal receptors and cells are different from human receptors and cells; thus errors in activity translation occur. The second was not made evident until the emergence of recombinant systems in pharmacology, namely that stimulus response mechanisms are not routinely monotonic in nature. In fact, it was the discovery of this fact that directly led to the discovery of biased receptor signaling (vide infra). Thus the introduction of recombinant technology the 1980s ushered in a new era in receptor pharmacology that essentially overturned many cornerstone assumptions of the previous 60 years.

3 Recombinant Systems Redefine Receptor Pharmacology

To a large extent, the necessarily limited application of functional pharmacological systems (common isolated tissues used by most researchers) led to the apparently harmonious concepts unifying receptor pharmacology at the time. However, the definition of the human genome and the introduction of recombinant systems into pharmacology allowed the study of receptors in a wealth of different interacting systems and this, in turn, unveiled the greater complexity inherent in receptor-response element interaction. These systems also obviated one of the underlying weaknesses of isolated tissue systems in that human receptors now could routinely be used in drug discovery in functional systems in vitro. However, the second weakness in isolated tissue pharmacology was then exposed as numerous instances of failure to adhere to standard PR predictions began to be observed and reported in the literature. Before discussing this latter idea, the emergence of another trend in pharmacology should be considered as it was instrumental in the exploration and exploitation of recombinant systems, namely the introduction of more functional pharmacological assay formats and the de-emphasis of binding as the primary tool for definition of receptor activity.

4 Binding Gives Way to Functional Experiments

Radioligand binding experiments can be extremely valuable to detect and measure direct interaction of molecules with receptor proteins. Such technology became important in pharmacology because it is amenable to robotic high throughput procedures required for high volume screening in drug discovery. On the other hand, the drug response that is most relevant to therapy is cellular function and there are numerous instances where binding falls short of predicting the functional effects of molecules. One reason this is the case is the fact that the two experimental formats measure different protein species; binding captures the protein interacting

with the radioligand while function captures the protein sending the signals to the cells (Kenakin 2009). Theoretically, functional experiments also are superior to binding formats because there are more interrogators of receptor conformation in function (namely the signaling proteins that interact with the receptor); binding relies simply on the interference of the radioligand-receptor interaction. While this has always been known to be true, technological advances were required to bring the state of the art of functional measurement of cellular response to the level of binding technology in terms of high throughput screening and the measurement of drug functional response in lead optimization assays. From the 1990s on, functional assay technology increased tremendously to the point where it could be used for high throughput screening and also for molecule characterization. An added bonus to this change in emphasis is the increase in the capability of detection and characterization assays to discover allosteric action of new ligands (Rees et al. 2002).

5 Understanding Agonism: The Black/Leff Operational Model

Concomitant with the introduction of more extensive functional assay technology was the introduction of arguably the most important model to describe drug agonism, specifically the operational model described by Black and Leff (1983) and Black et al. (1985). The chemical production of cellular response (agonism) had long been a mysterious property of drugs. Pioneers of receptor pharmacology such as Ariens (1954, 1964), Ariens and Van Rossum (1957), Furchgott et al. (1966) and Stephenson (1956) produced useful models to accommodate this intriguing property of some ligands. These approaches led to the insertion of a calibration factor to accommodate the receptor occupancy of agonists with the observed responses produced by agonists; this factor is referred to as efficacy (Stephenson 1956) or intrinsic efficacy (Furchgott et al. 1966). The ad hoc nature of this approach was seen as a shortcoming of pharmacology by Sir James Black and Paul Leff and they formulated a physiologically based model to describe agonism. This model was fundamentally unappreciated when published but subsequently has now become the state of the art approach to the handling of agonism in pharmacology.

The Black/Leff operational model defines the response to an agonist [*A*] as (Black and Leff 1983):

$$\text{Response} = \frac{[A]^n \tau_A^n E_m}{[A]^n \tau_A^n + ([A] + K_A)^n} \tag{1}$$

where agonist affinity is given by the equilibrium-dissociation constant of the agonist-receptor complex (K_A) and the agonist efficacy is denoted τ_A. The slope of the concentration response curve is n and E_m is the maximal window of response for the assay. For partial agonists, unique values for both K_A and τ_A can be derived but for full agonists, an infinite combination of K_A and τ_A values can fit the curve. However, the unique identifier for full agonists then becomes $\log(\tau_A/K_A)$ (Kenakin et al. 2012).

One of the most valuable applications of the Black/Leff operational model is that it gives the capability to compare the activity of full to partial agonists. While potency ratios are theoretically sound for the comparison of full agonists to each other, the location parameter of partial agonist concentration-response curves changes differently from the location parameter of full agonists with changing assay sensitivity. Figure 1 shows the ratio of pEC_{50}'s (−log of molar EC_{50} concentrations) as ΔpEC_{50} as a function of the receptor density in the functional assay. It can be seen that as one of the agonists becomes a partial agonist at lower receptor densities, the linear relationship with receptor density becomes non-linear and not comparable to values found at higher tissue sensitivities. In contrast, the index $\Delta Log(\tau/K_A)$ remains linear for full and partial agonists thus providing a useful scale for comparison that is independent of tissue sensitivity (Kenakin 2017).

The operational model provides a physiologically plausible description of agonist efficacy and a structure within which the affinity and efficacy of an agonist can be quantified in a system independent manner. This, in turn, can be used to predict agonism in all therapeutically relevant systems. The acceptance of this model in the pharmacological approaches to receptor pharmacology was critical to the advancement of the functional systems to the development of new drug candidates in the process of drug discovery.

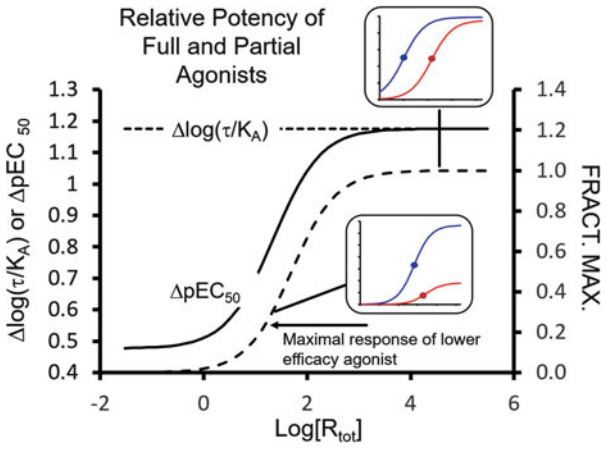


Fig. 1 Changes in relative potency of agonists with changing receptor density (tissue sensitivity). Left ordinate axis is $\Delta Log(\tau/K_A)$ or ΔpEC_{50} of two agonists with a relative efficacy of 0.2 (the more powerful agonist has 5-times the efficacy of the weaker agonist). Right ordinate axis is the maximal response (as a fraction of maximal assay window) of the weaker agonist. Abscissae is the log of the receptor density in the assay system. It can be seen that the relative potency as measured by the ΔpEC_{50} varies with system sensitivity until both agonists produce the full maximal response (both are full agonists). If the weaker agonist is a partial agonist, then pEC_{50} is variable. In contrast, the $\Delta Log(\tau/K_A)$ remains constant throughout the complete range of assay sensitivities being constant whether both agonists are full agonists or if one or both are partial agonists

6 The Shift from Orthosteric to Allosteric Drug Action

Seven transmembrane receptors are nature's prototypical allosteric protein, i.e. these are proteins designed to bind molecules in the extracellular space, change their conformation accordingly, and present different tertiary conformations toward signaling proteins in the cytosol. This is the very definition of allosteric function (etymology 'allo' = other, 'steric'-arrangement of atoms in space) which connotes binding at one site to induce an effect in another. Everything seven transmembrane receptors do is allosteric and can be described by an allosteric vector of 'modulator', conduit (the receptor protein) and guest molecule. Within this context, the natural binding site for the natural agonist (neurotransmitter, hormone) can be considered the 'orthosteric' site and the different site binding the modulator the 'allosteric' site. Thus, an orientation of the vector from the outside to the inside of the cell is agonism with the modulator binding as the agonist, conduit the receptor and guest the signaling protein (Kenakin 2012). A vector along the plane of the membrane describes receptor oligomerization with a modulator being either another receptor or Receptor Activity Modifying Proteins (RAMPs), conduit the receptor and guest the receptor itself. A vector between two drugs binding on the receptor is the conventional 'guest allostery' where one ligand affects the binding and function of another both binding to the receptor protein.

Models of allosteric binding and function involve the interaction of a probe ligand (agonist or radioligand denoted [*A*]) with a protein (receptor *R*) and a guest molecule (modulator denoted [*B*]). Figure 2a shows the allosteric binding model for binding

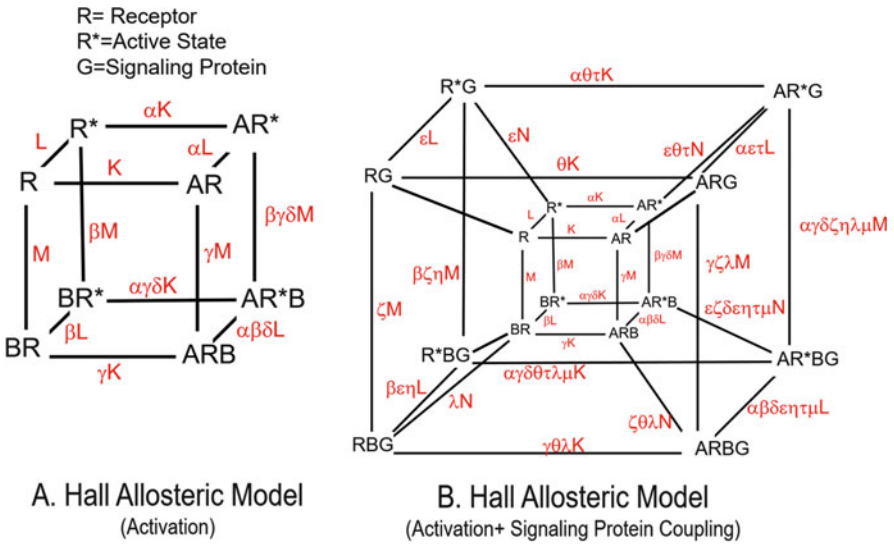


Fig. 2 Allosteric models for an Agonist *A*, allosteric modulator *B*, receptor is *R* and a signaling protein *G*. The model defines the receptor species present in system through formation of an active receptor state (R∗)-Panel **a**, Hall Allosteric model (Hall 2000) or through formation of an active state and allowing the receptor to couple to signaling proteins (Panel **b**) (Hall 2006)

and receptor activation in a relatively simple scheme; Fig. 2b shows how these models rapidly become more complex with the introduction of other receptor behaviors, in this case the interaction of the receptor species with signaling proteins. The problem with these more inclusive models is that they are heuristic and contain a larger number of independently unverifiable parameters. This also belies the notion that an advantage of binding formats is they are 'simple' (Christopoulos and Kenakin 2002).

Allosteric effects can be very complex and can involve changes in the affinity and/or efficacy of the probe molecule. This being the case, not all allosteric effects are detected or can be studied through radioligand binding; functional assays are a much better format for the study of allosteric effects. The lack of functional assays was a hindrance to the effective study of allosteric receptor behavior but as more functional assays became available through technological advances, the more prevalent in the literature allosteric effects became. Thus, a near exponential increase in the prevalence of scientific papers citing the words allosteric or allosterism can be seen from 1990 up to the present day; presumably some of this increased trend is due to increased availability of simple pharmacological functional assays (Rees et al. 2002). Thus in the 2000s, increased emphasis on allosteric mechanisms was evident in pharmacological receptor literature.

There are theoretical reasons why functional response is a more predictive and useful measure of drug activity than binding. A barrier to the creation of a functional allosteric model was that there was no plausible means to process cellular response emanating from the active receptor species. This problem was solved by the melding of the Black/Leff operational model to allosteric binding models – see Fig. 3. Thus,

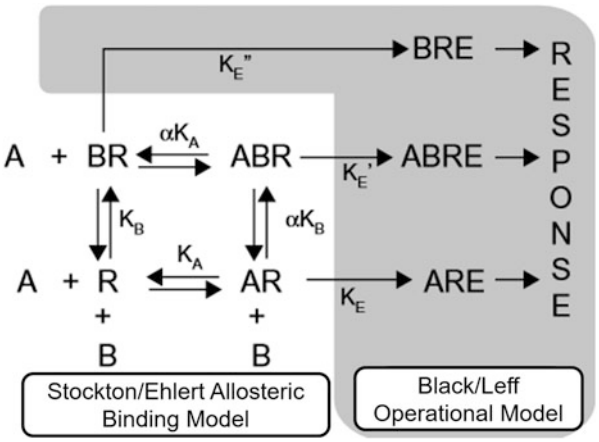


Fig. 3 Model for functional receptor allosterism [46]. A probe ligand [A] (agonist) binds to the receptor and the resulting complex (ARE) can produce response. Similarly, the allosteric modulator B can simultaneously bind to the receptor and produce response through the complex BRE and can modify the agonist response through the species ABRE. Binding to the receptor is described by the allosteric binding model (Stockton et al. 1983; Ehlert 1988) and response is described by the Black/Leff operational model (Black and Leff 1983)

the current functional allosteric model (Kenakin 2005; Ehlert 2005; Price et al. 2005) joins the Stockton/Ehlert allosteric binding model (Stockton et al. 1983; Ehlert 1988) to the Black/Leff operational model (Black and Leff 1983) to yield:

$$\text{Response} = \frac{([A]/K_A\tau_A(1+\alpha\beta[B]/K_B)+\tau_B[B]/K_B)E_m}{[A]/K_A(1+\alpha[B]/K_B+\tau_A(1+\alpha\beta[B]/K_B))+[B]/K_B(1+\tau_B)+1} \quad (2)$$

where the allosteric modulator [*B*] has a receptor-ligand equilibrium dissociation constant K_B and a possible direct efficacy τ_B. The modulator changes the affinity of the agonist by a cooperativity factor α and changes the efficacy of the agonist by a cooperativity factor β. Thus Eq. (2) can be fit to concentration-response data for agonists in the absence and presence of modulators to yield constants that characterize modulator activity in the form of α, β, K_B and τ_B. Figure 4 shows how Eq. (2) can be recast with either agonists or signaling proteins (i.e. G proteins) as the allosteric modulator (of the other). The emergence of allosteric pharmacology brought with it a whole new collection of drug candidates with unique properties that promise to revolutionize drug therapy. Specifically, while the previous orthosterically based discovery efforts yielded copies of neurotransmitters and/or hormones or steric antagonists that block physiological effect, allosteric mechanisms promise a new array of possible interactions. This new array of behaviors stems from the unique properties of allosteric molecules. Firstly, because modulators bind to their own site on the receptor, their effects are saturable. This allows them to modify natural signaling without completely overwhelming the system. Orthosteric drugs hijack the receptor to yield pharmacology linked to the orthosteric ligand. Allosteric modulators work in partnership with existing physiology to modify

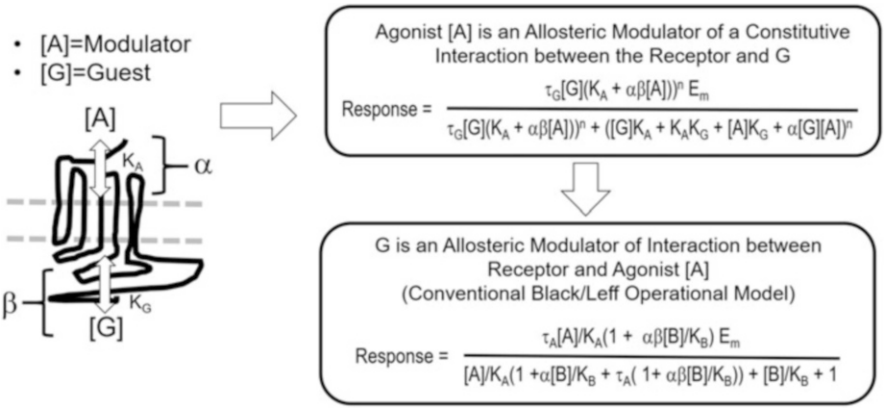


Fig. 4 Equation for the facilitation of spontaneous interaction between receptors ([R]) and signaling proteins (*G*) by an allosteric modulator (*A*) from the functional allosteric model. As an example for $n = 1$, this equation is identical to the functional allosteric model equation for agonist response from the Black/Leff operational model for direct agonism

signaling; the fact that they can simply reset ligand affinity and efficacy allows them to modify response (slightly reduce or increase signaling) with the modulator influence ending when the allosteric site is saturated. Secondly, allosteric effects are probe dependent thus a given modulator may change the response to one agonist/radioligand without affecting the response to another. This can be tremendously useful therapeutically where it may be beneficial to change the interaction of receptors with one protein species but not another. For example, HIV-1 utilizes the chemokine receptor CCR5 to cause cell infection yet there is evidence to suggest that the natural chemokine function of CCR5 is protective after HIV-1 infection with respect to progression to AIDs (Gonzalez et al. 2005). Therefore, an allosteric modulator that blocks the interaction of CCR5 with HIV-1 but otherwise permits the receptor to interact with natural chemokine would be preferable to a simple blocker of the CCR5 receptor (Muniz-Medina et al. 2009). There are data to show that such allosteric advantage can be found in allosteric molecules such as TAK 652 (Muniz-Medina et al. 2009).

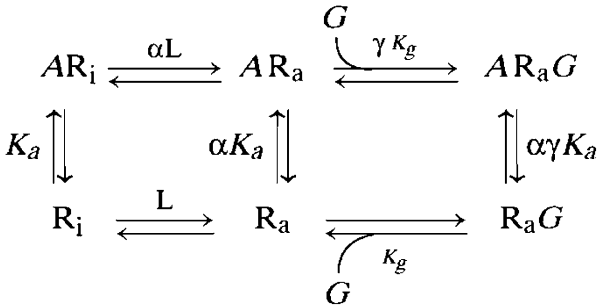
The new array of possible therapeutics from allosteric mechanisms range from antagonists (negative allosteric modulators referred to as NAMs) to potentiators of physiological response (positive allosteric modulators referred to as PAMs). Obviously PAMs are only possible through an allosteric mechanism since they require the co-binding of the natural agonist to produce their effect. PAMs can produce potentiation either through increasing the affinity of the receptor for the agonist (through an α-effect – see Eq. 2) or increasing the efficacy of the agonist (through a β-effect see Eq. 2) or both. PAMs are unique for possible rejuvenation of failing systems in disease. Allosteric modulators also can produce direct agonism and, unlike standard orthosteric agonists that preclude the natural agonist when they bind, allosteric agonists can produce an additive agonism with variable effects on the natural system. Thus, allosteric agonists may block the natural agonist (NAM agonists), potentiate the natural agonist (PAM agonists) or not affect the natural agonist at all (allosteric agonist).

Another impact on the drug discovery process brought on by the advent of allosteric ligands is it has created an expansion of opportunity for intractable targets and peptide receptors (Class B receptors). Specifically, finding druglike non-peptide ligands for receptors that have peptides as their natural agonist has been difficult. However, allosteric ligands for these receptors can readily be found when the target is screened as an allosteric target (Kenakin 2010; Burford et al. 2014; Alt 2016).

7 The Move from Parsimonious Models to Dynamic Models of Receptor Conformation

In lieu of further information, the models defining receptor function were parsimonious, i.e. based on the premise that one should not propose a receptor state for which there is no evidence supporting it's existence. This led to minimal models of receptor function such as the extended tertiary complex model of receptor function proposing an inactive ([Ri]) and active ([Ra]) receptor active state interacting with

a single signaling protein to form a ternary complex species mediating cellular response (Samama et al. 1993):



(3)

Within this scheme, a ligand [A] has an equilibrium association constant for the receptor-ligand complex of K_a for Ri and αK_a for Ra. The equilibrium between Ri and Ra is given by L where L = [Ra]/[Ri]. The signaling protein (in this case denoted G) has an equilibrium association constant for the unliganded receptor of K_g and for the ligand-bound receptor of γK_g. The response is produced by the activated receptor complex coupling to the signaling protein. This appears to be a parsimonious model of receptor activation in that the only two receptor species present appear to be the inactive state ([Ri]) and the active state ([Ra]) Ostensibly this model proposes the existence of only two receptor states, Ri and Ra but this is an illusion. In fact, an infinite number of receptor states is predicted by this model by variation of the γ term with different agonists. Specifically, the value of γ determines the agonism and it is not specifically linked to Ra but rather can be different for different agonists (Kenakin 2012). This actually is in keeping with the basic tenets of allosterism. Specifically, allosterism predicts that the effects of allosteric modulator, which change the tertiary conformation of the receptor, can be unique to the modulator. In this case, the modulator is the agonist therefore different agonists can produce receptor species with differing affinities for the signaling protein; this is encompassed in the term γ of the extended ternary complex model. It is this property of allostery that confers selectivity of the agonist-bound receptor species of affinity for signaling coupling proteins. The selectivity in the allosteric effect of modulators for different guests is given the name 'probe dependence' and it governs all receptor-ligand and receptor-protein interactions (vide infra).

The extended ternary complex model is widely regarded as being a 'two state' model mainly because of the pre-existence of the two states Ri and Ra but it can be seen from the previous discussion of the γ term, in actuality it is a multistate model that can be used to describe a myriad of ligand-receptor signaling options. However, it is still limited in that it is, as were all models up to this point in time, a linkage model comprised of various defined protein species linked together such that the transitions between them are energy neutral, i.e. energy is conserved. These types of models necessarily have a linearity built into them, i.e. a strict linkage system may require a receptor species to be made as a pre-requisite to another one thus linking

the two. For example, with no explicit knowledge of the intervening pharmacological processes involved, the disappearance of agonism usually was taken as a surrogate reading for receptor internalization. The underlying assumption then is that receptor activation is a required pre-requisite for receptor internalization thus assuming a linearity linking activation with internalization. The emergence of new functional assays that directly measure receptor internalization through imaging revealed that this linearity is incorrect as many ligands (i.e. receptor antagonists i.e. Willins et al. (1999) actively produce receptor internalization without producing any activation of the receptor. In fact, many such dissociations between receptor functions are noted in the literature thus illustrating that efficacy is collateral, i.e. not only can ligands have many efficacies but there may be no connection between these efficacies (Kenakin 2002) – see Fig. 5. The standard linkage models of receptor function cannot accommodate these parallel but unrelated efficacies (unless a separate linkage model scheme were devised for each one). While multi-state linkage models were devised, they were unsatisfactory ad hoc approaches to an increasingly complex array of the receptor behaviors revealed by new functional pharmacological assays. One of the main shortcomings of these models is the need to predefine the receptor protein species, a random process at best in light of the unknown conformations possibly stabilized by signaling proteins and/or ligands. This disadvantage was eliminated by adopting dynamic probability models of receptor function; the first to be applied to pharmacology was given by Onaran and Costa (1997) and Onaran et al. (2000).

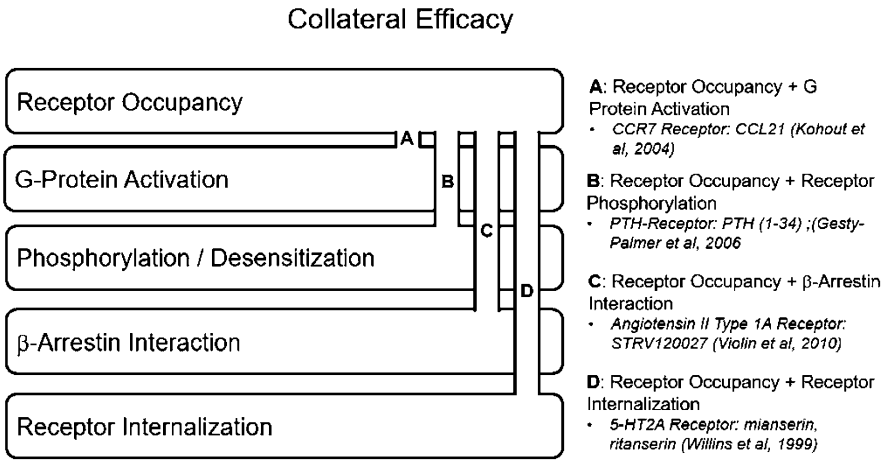


Fig. 5 The conventional linear stimulus-response chain of receptor occupancy by an agonist (activation), followed by G protein activation, receptor phosphorylation and desensitization, β-arrestin interaction and receptor internalization can be sampled by different agonists without activating the entire sequence. Thus agonists such as the chemokine CCL21 activates CCR7 receptors without activing β-arrestin and causing receptor internalization (Kohout et al. 2004), the PTH receptor agonist analog [Tyr34]PTH-(7–34) selectively activates Gs/PKA-mediated ERK1/2 activation (Gesty-Palmer et al. 2006)

The introduction of molecular dynamics into pharmacology has provided a radically different view of receptor function namely that agonism occurs because of the selection of receptor conformations from a pre-existing ensemble of similar (in terms of free energy) but different conformations (Park 2012; Nygaard et al. 2013; Motlagh et al. 2014; Boehr et al. 2009; Dror et al. 2010, 2011). These ensembles of receptor conformations form a dynamic system (Vardy and Roth 2013; Manglik and Kobilka 2014; Manglik et al. 2015) which combines with signaling systems through a full range of allosteric linkages (Monod et al. 1965; Changeux and Edelstein 2005). These concepts have been described in terms of oscillating dynamic systems of multiple conformations (Cui and Karplus 2008; Changeux and Edelstein 2011) constituting 'fluctuating networks' operating on a real time scale of microseconds (Ichikawa et al. 2016).

In dynamic schemes of receptor function, the conformational state of the receptor is not pre-defined but rather it is recognized that the receptor may exist in a large number of co-existing conformations (termed 'ensembles'). Thus it is the probability of changing a receptor state that is the discerning property of an agonist in terms of producing cellular response (Onaran and Costa 1997; Onaran et al. 2000). The important advance in thinking with these types of models is that they remove the linearity from the process of signal transduction, i.e. the production of a given receptor active conformation is governed by a free movement of the receptor on an 'energy landscape' (Frauenfelder et al. 1988, 1991; Woodward 1993; Dill and Chan 1997; Hilser and Freire 1997; Miller and Dill 1997; Hilser et al. 1998, 2006; Freire 1998) where there is no need to transition through one conformation to get to another. In this scheme, receptors can 'skip' efficacies and practice 'collateral' efficacy-see Fig. 5 (Kenakin 2002). The serious expansion of protein dynamics into receptor pharmacology has greatly increased the vistas open to the prediction of what ligands can do when they interact with receptors.

8 Allosteric Probe Dependence: Biased Receptor Signaling

Deviations from the predictions of the simple model of a single receptor active state and conservation of agonist potency ratios compelled a re-thinking of the parsimonious receptor active state models put forth in the 1970s. and 1980s. Specifically, prior to that time, agonist potency ratios were found to be uniform for a given receptor and set of agonists thus reinforcing the assumption of a monotonic linkage between receptor stimulus and tissue response. One of the main reasons for this congruency is the fact that a limited number of functional responses from receptors could be measured. To detect selective signaling pathway requires the ability to measure two signals emanating from the same receptor and this capability was lacking at that time. With the increase in the number of possible functional readouts from receptors in new assays applied to recombinant assays where the same receptor can be studied as it couples to different pathways led to a change in the ideas describing agonism.

As more data accumulated, instances appeared where agonist potency ratios were not constant and by the 1980s, these reports laid the foundation for ideas around alternative schemes for agonism. As put by Roth and Chuang (1987), '*. . .the possibility is raised that selective agonists and antagonists might be developed which have specific effects on a particular receptor-linked effector system. . .*'. In general, deviations from monotonic stimulus-response coupling were seen to increase with recombinant systems when receptors were placed into different environments to interact with different signaling proteins. In 1989, a theoretical study predicted that if agonists stabilized unique active state conformations then differences in the relative agonism of these agonists would be seen (Kenakin and Morgan 1989); this offered a plausible mechanism for the discordant potency ratios seen experimentally. By 1993, a body of data had been published to show that agonist potency ratios can vary between functional systems routinely; a definitive study by Spengler et al. (1993) showed that two natural agonists for the pituitary adenylate cyclase-activating polypeptide (PACAP) receptor, $PACAP_{1–17}$ and $PACAP_{1–35}$ showed opposite potency ratios for the receptor depending on whether cyclic AMP or inositol phosphate metabolism was used for the functional readout. This result is absolutely incompatible with a monotonic stimulus-response coupling mechanism.

In 1995, a mechanism was proposed for the non-congruent potency ratios, namely that it is the agonist bound receptor active state and not the receptor that is the minimal unit of control for agonism and that the stabilization of different receptor active states leads to the production of what was then referred to as 'stimulus-trafficking' (Kenakin 1995). Numerous experimental efforts noted the same phenomenon and a number of names were given to the effect: stimulus trafficking (Kenakin 1995), functional dissociation (Whistler et al. 2002), biased agonism (Jarpe et al. 1998), biased inhibition (Kudlacek et al. 2002), differential engagement (Manning 2002), functional selectivity (Lawler et al. 1999; Kilts et al. 2002; Shapiro et al. 2003), and ligand directed signaling (Michel and Alewijnse 2007). The term 'bias' emerged as a general label for the agonist production of discrete signaling from receptors, i.e. a given agonist selectively produces one signal from pleiotropically coupled receptor at the expense of others. A great deal of literature has amassed around this effect and when first presented, it appeared to be a unique property of receptors. However, clearly this simply falls within the repertoire of allosteric proteins where the allosteric vector aimed toward the cell cytosol (modulator as agonist, signaling protein as guest) practices probe dependence. Thus the allosteric effect resulting in agonism is different for one probe-guest pair (eg. G protein signaling) over another guest-receptor probe (eg. β-arrestin).

Signaling bias, while being a limitation to the prediction of agonist response in vivo, also was recognized as a potentially useful therapeutic extension of pharmacological agonism. Biased signaling can: (1) emphasize beneficial signaling pathways, i.e., parathyroid hormone (PTH)–mediated bone building for osteoporosis (Gesty-Palmer and Luttrell 2011; Gesty-Palmer et al. 2013); (2) de-emphasize harmful signaling pathways, i.e., respiratory depression for opioid analgesics (Raehal et al. 2005; Kelly 2013; Koblish et al. 2017); (3) de-emphasize harmful

pathways and prevent the natural agonist from activating these pathways, i.e., biased angiotensin blockers for heart failure (Violin et al. 2006, 2010); and (4) allow pursuit of previously forbidden drug targets due to side effects, i.e., κ-opioid receptor analgesics (White et al. 2014; Brust et al. 2016). As this phenomenon was widely explored it basically revitalized seven transmembrane receptors as a therapeutic drug target (Kenakin 2015). Biased signaling allows the fine tuning of agonist response and is used in many natural systems to differentiate subtleties of signaling due to natural endogenous agonists. Thus whereas previously classified redundant systems such as the chemokine receptor system (19 receptors are activated by 47 chemokines) which has a large amount of natural agonist cross-over upon further study has determined that biased signaling can fine tune natural agonist response. For instance, the chemokines CCL19 and CCL21 are both natural agonists for the CCR7 receptor yet while both activate G proteins, only one (CCL19) also causes receptor agonist-dependent phosphorylation and recruitment of β-arrestin to terminate the G protein stimulus (Kohout et al. 2004; Byers et al. 2008; Hauser and Legler 2016).

One of the main lessons learned from the study of biased signaling is that not all agonists are created equal. There have been various methods proposed to quantify bias (see Onaran et al. 2017 for review) and various ways to display biased arrays of signaling. Figure 6 shows the two most popular methods; radar plots (Fig. 6a) and cluster analysis of heatmaps (Fig. 6b). Therefore, in addition to the ability to create

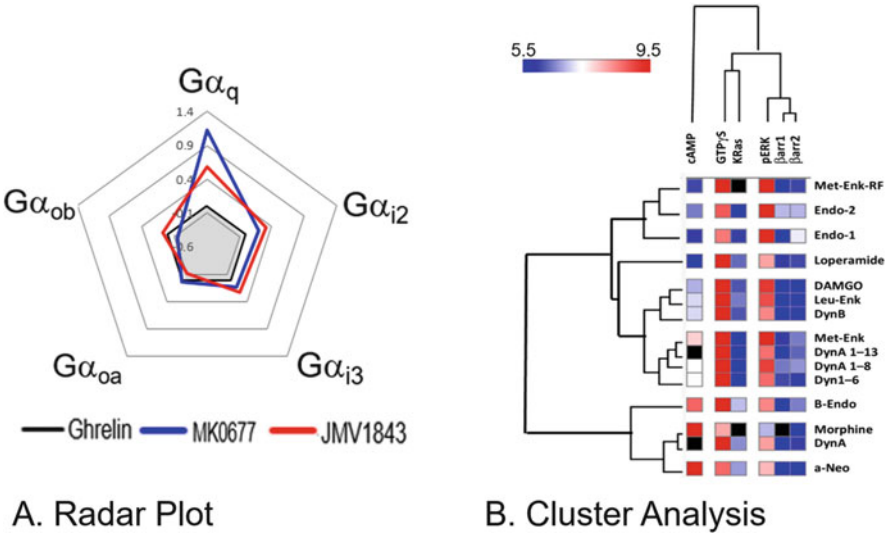


Fig. 6 Two common methods of displaying agonist biased signaling. (**a**) Depiction of G protein selectivity with a radar plot based on G protein type showing $\Delta\text{Log}(\tau/K_A)$ values (with ghrelin as the reference agonist); filled pentagon shows activity equal to ghrelin. Data shown for agonists MK0677 and JMV1843. Redrawn from M'Kadmi et al. (2015). (**b**) Cluster analysis of 15 μ-opioid agonists in six different functional assays (data from Thompson et al. 2016). The gene cluster program GENE-E groups the agonists according to their $\text{Log}(\text{max}/\text{EC}_{50})$ values in each assay. Analysis and data redrawn from Kenakin (2015)

agonists that preserve beneficial signaling and diminish harmful signaling, medicinal chemists now have detailed scaled pathway-dependent activity that can be used in structure activity relationships to optimize therapeutically useful activity. Another lesson learned however is that biased stimuli from receptors are processed differently by cells to yield cell dependent response thereby making predictions from simple in vitro functional systems to in vivo systems (where a great many different cell types are encountered) difficult. This underscores the failure of in vitro potency ratios as successful predictors of in vivo activity not surprising since this incongruence was what led to the discovery of biased signaling in the first place.

9 Structure: Receptors Show Themselves

For a predominant period in Pharmacology, the 'receptor' was only a concept with no physical representation. The difficulties in crystalizing receptor structures were consistent with the known flexibility of allosteric proteins (Liapakis et al. 2012). However, through a series of ingenious experimental techniques and through persistence, the first crystal structure for the β-adrenoceptor was achieved and published in 2011 (Rasmussen et al. 2011). Shortly thereafter, the Nobel Prize in Chemistry was awarded for this great accomplishment to Brian Kobilka and Robert J Lefkowitz (Kenakin 2013). This achievement ushered in a new era into Pharmacology whereby the true nature of pharmacological drug discovery, which is the interaction of biologists with medicinal chemists to modify physiology, could finally be realized. Now chemists have a physical docking point for their scaffolds to inform their ideas about how structure modifies biological activity. This was the aim of Lefkowitz and Kobilka and Lefkowitz and the other scientists striving to determine the crystal structure of seven transmembrane receptors. As they wrote as early as 1987: '*Our proposed model for the structure of the β-adrenergic receptor and its interaction with pharmacologically important ligands should, together with the biochemical and genetic studies now possible, provide a rational basis for a new approach to the development of more selective drugs*' (Kobilka et al. 1987).

In the ensuing years, a number of X-ray crystal structures of seven transmembrane receptors have been published – see Fig. 7. These structures have been

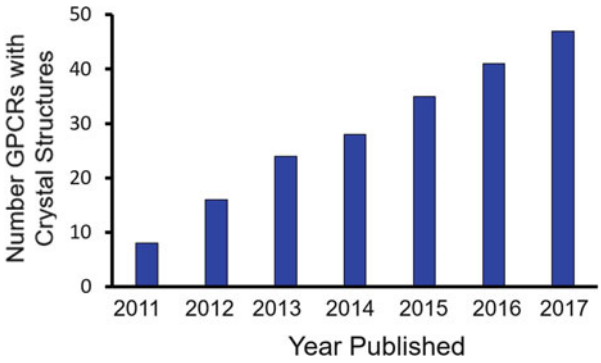


Fig. 7 Bar graph showing the number of GPCR crystal structures reported in the literature for the years 2011–2017

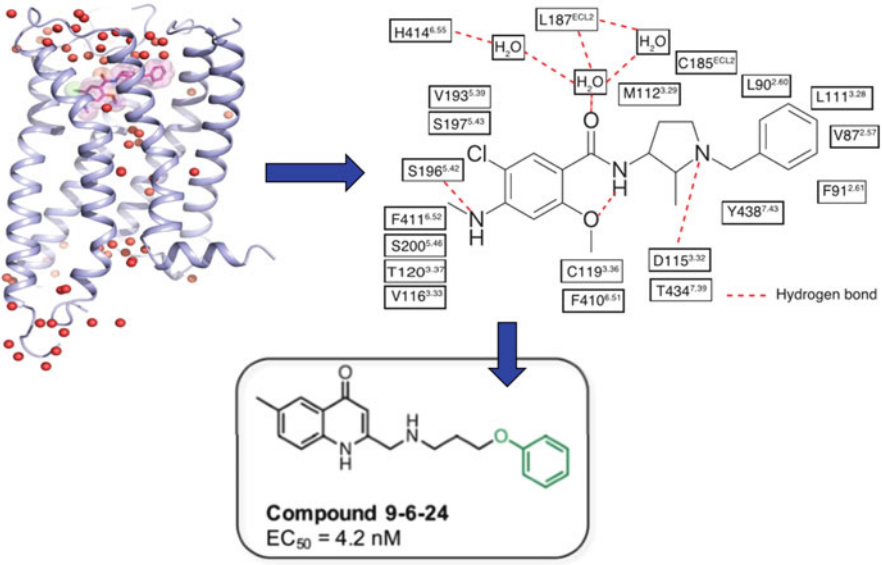


Fig. 8 The crystal structure of the dopamine D4 receptor was utilized to design molecules that culminated in the discovery and development of Compound 9-6-24, a ligand with nanomolar affinity for the receptor. From Wang et al. (2017)

extremely valuable in designing matching structural drug moieties through complimentary binding motifs (Hauser et al. 2017). These have furnished a solid basis for lead screening and optimizaton programs in many drug discovery efforts. For example, the crystal structure of the dopamine D4 receptor was instrumental in the design and discovery of a nanomolar agonist for the receptor. –see Fig. 8, (Wang et al. 2017).

10 Genetics and Computer Science Impact Pharmacology

The impact of the elucidation of the human genome on pharmacology is of course immense. Besides allowing the cloning and expression of receptors in most cells (recombinant pharmacology), genetic knockout animals have been extremely useful in the linking of receptor effects to physiology. These allow the study of drug effect without selected components in studies to determine the importance of those components to the physiology. For instance, the reduction in morphine-mediated respiratory depression in β-arrestin knockout mice (Raehal et al. 2005; Kelly 2013; Koblish et al. 2017) and the reduction in the bone building effects of PTH in β-arrestin knockout mice (Gesty-Palmer and Luttrell 2011; Gesty-Palmer et al. 2013) suggest the relevance of this signaling pathway to these drug effects. Knock out animals have been important to the elucidation of pathway signaling for dopamine D_1 receptors (Xu et al. 1994), 5-HT_{2B} receptors (Saudou et al. 1994),

angiotensin 1A receptors (Ito et al. 1995; Coffman 1997), metabotropic $mGLu_1$ receptors (Aiba et al. 1994), α_{2b}-adrencoceptors, (Link et al. 1996; MacMillan et al. 1996), μ-opioid receptors (Sora et al. 1997), α_{1b}-adrenoceptors (Cavalli et al. 1997), muscarinic M_3 receptors (Duttaroy et al. 2004), β_1/β_2 adrenoceptors (Rohrer et al. 1999) and β_3-adrenoceptors (Susulic et al. 1995). A recent new technology that promises to be extremely useful in this regard is through application of RNA-guided clustered regularly interspaced short palindromic repeat (CRISPR)/Cas9 endonucleases (Naylor et al. 2016).

Knock in studies where the endogenous GPCR gene is replaced with a gene for a mutant receptor and the expression of that mutant is driven by the wild type promoter have also added value to the determination of receptor signaling in physiology. This furnishes the mutant receptor in the same tissue types and at the same receptor levels as the wild type receptor. For example, the relative contributions of δ and μ opioid receptors in the sensation of heat and mechanical pain with eGFP δ-opioid receptors (Scherrer et al. 2006, 2009; Shenoy and Lefkowitz 2011; Pradhan et al. 2009; Faget et al. 2012) has been delineated with this technology. Genetically modified receptors such as phosphorylation-deficient muscarinic M3 receptors have been used to study receptor internalization and phospholipase coupling (Kong et al. 2010; Poulin et al. 2010; Torrecilla et al. 2007; Budd et al. 2001).

Genetically modified receptors also have become an important element in pharmacological drug discovery and the elucidation of receptor signaling in vivo. In the late 1990s experiments whereby a coding sequence of a mutant receptor that is only activated by a synthetic ligand is inserted into the cellular genome were used to explore the relevance the receptor signaling (Peng et al. 2008); these receptors were given the name RASSL for 'Receptor Activated Solely by a Synthetic Ligand'. The first application of this approach was made for the κ-opioid receptor modified to contain the second extracellular loop of the δ-opioid receptor (Coward et al. 1998) to yield a receptor with a 200-fold reduction in the binding of the endogenous opioid agonist dynorphin (and reductions in the binding of 21 other opioid peptides) but a maintained binding and activation for the synthetic agonist spiradoline. The problem with RASSLs was that high activity of the synthetic ligands for native receptors caused concomitant activation of native receptors in the transgenically modified animals (Redfern et al. 1999). These disadvantages were eliminated in second generation genetically modified receptors named DREADs for 'Designer Receptors Exclusively Activated by Designer Drugs' (Urban and Roth 2015; Conklin et al. 2008); these utilized a synthetic agonist clozapine-N-oxide that does not activate native receptors (Conklin et al. 2008; Armbruster et al. 2007; Urban and Roth 2015; Giguere et al. 2014). An example of where this technique has been used is in the elucidation of M_3 receptor signaling (Armbruster et al. 2007; Dong et al. 2010). Specifically, the use of DREADs for the evaluation of the importance of biased signaling as in the cell type specific expression of a muscarinic M_3 receptor DREAD mutationally modified to not interact with β-arrestin but rather only $G_{q/11}$ proteins (Hu et al. 2016). In general, the impact of genetics on pharmacological drug discovery has led to advances far beyond what the initial technology promised and looks to be a huge vista for increased value in pharmacology.

Another extremely important technology for new drug discovery entered the field in the form of virtual screening. This technique docks millions of molecules into target binding sites to optimize fits and generate chemical structures that fit into drug binding sites. For example, this procedure led to the discovery of the positive allosteric modulator of the proton sensor GPR68 (Huang et al. 2015) – see Fig. 9. The starting point for this process was the initial discovery of lorazepam activation of GPR68 in yeast. A starting template from the CXCR4 receptor (29% homology with GPR68 and related receptors GPR4, GPR65 and GPR132) led to the creation of 3,307 homology models through sampling of backbone and loop conformations. Computational docking with benzodiazepines and other inactive compounds furnished 622 decoy molecules leading to a stable lorazepam docking pose. The roles of critical amino acid residues were explored with mutation and 3.1 million lead-like molecules were computationally docked; from 3.3 trillion calculated complexes the top 0.1% docking-ranked were obtained for testing. This provided hundreds of analogues which were then docked against the GPR68 model to generate 25 key molecules for testing. This lead to the discovery of ZINC67740571 (subsequently called ogerin (Huang et al. 2015)), a potent PAM

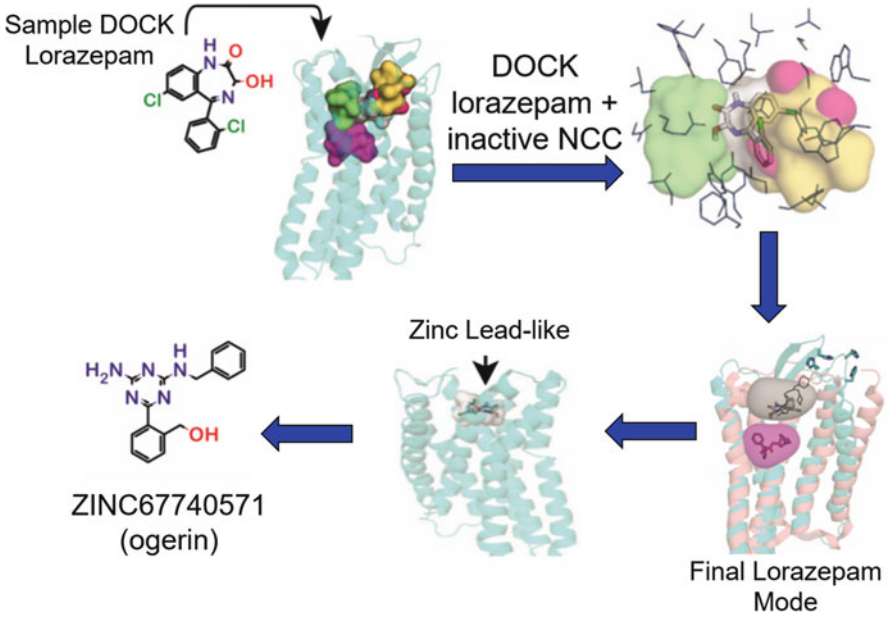


Fig. 9 The discovery of lorazepam activation of GPR68 in yeast led to the creation of a template from the CXCR4 receptor (29% homology with GPR68 and related receptors GPR4, GPR65 and GPR132) and creation of 3,307 homology models for docking with benzodiazepines and other inactive compounds. This process led to establishment of a stable lorazepam docking pose for virtual docking with one million lead-like molecules (3.3 trillion calculated complexes). The result led to identification of analogues which were docked against the GPR68 model to generate 25 key molecules for testing and the eventual discovery of ZINC67740571 (subsequently called ogerin (Huang et al. 2015)), a potent PAM of GPR68 hydrogen sensing activity

of GPR68 hydrogen sensing activity. As demonstrated with this example and others, virtual screening and docking has become an effective means of generating new lead molecules.

11 Conclusion

Pharmacology was formed from a subset of physiology, a discipline many hundreds of years old. At some point approximately 150 years ago, a cadre of physiologists, trained to question the mechanism of physiological systems and using drugs to probe these systems, became more interested in the probes than the system and thus became pharmacologists. This melded the disciplines of medicinal chemistry and physiology and formed pharmacology, the chemical control of physiology. A unique feature of pharmacology is that it is an amalgam of many biological and chemical disciplines and as advances in each of these disciplines occurs, it is reflected as an advance in pharmacology. Thus pharmacology is ever forming and evolving as these emergent ideas enter the field.

References

Aiba A, Chen C, Herrup K, Rosenmund C, Stevens CF, Tonegawa S (1994) Reduced hippocampal long-term potentiation and context-specific deficit in associateive learning in mGluR1 mutant mice. Cell 79:365–375

Alt A (2016) Overview of critical parameters for the design and execution of a high-throughput screen for allosteric ligands. Curr Protoc Pharmacol 74:9.20.1–9.20.23

Ariens EJ (1954) Affinity and intrinsic activity in the theory of competitive inhibition. Arch Int Pharmacodyn Ther 99:32–49

Ariens EJ (1964) Molecular Pharmacology: V1. Academic Press, New York

Ariens EJ, Van Rossum JM (1957) pDx, pAx and pDx values in the analysis of pharmacodynamics. Arch Int Pharmacodyn Ther 110:275–299

Armbruster BN, Li X, Pausch MH, Herlitze S, Roth BL (2007) Evolving the lock to fit the key to create a family of G protein–coupled receptors potently activated by an inert ligand. Proc Natl Acad Sci U S A 104:5163–5166

Black JW, Leff P (1983) Operational models of pharmacological agonism. Proc R Soc Lond B Biol Sci 220:141–162

Black JW, Leff P, Shankley NP, Wood J (1985) An operational model of pharmacological agonism: the effect of E/[A] curve shape on agonist dissociation constant estimation. Br J Pharmacol 84:561–571

Boehr DD, Nussinov R, Wright PE (2009) The role of dynamic conformational ensembles in biomolecular recognition. Nat Chem Biol 5:789–796

Brust TF, Morgenweck J, Kim SA, Rose JH, Locke JL, Schmid CL, Zhou L, Stahl EL, Cameron MD, Scarry SM et al (2016) Biased agonists of the kappa opioid receptor suppress pain and itch without causing sedation or dysphoria. Sci Signal 9:ra11

Budd DC, Willars GB, McDonald JE, Tobin AB (2001) Phosphorylation of the Gq/11-coupled M3 muscarinic rec3ptor is involved in receptor activation of the ERK1/2 mitogen-activated protein kinase pathway. J Biol Chem 276:4581–4587

Burford NT, Wehrman T, Bassoni D, O'Connell J, Banks M, Zhang L, Alt A (2014) Identification of selective agonists and positive allosteric modulators for μ- and δ-opioid receptors from a single high-throughput screen. J Biomol Screen 19:1255–1265

Byers MA, Calloway PA, Shannon L, Cunningham HD, Smith S, Li F, Fassold BC, Vines CM (2008) Arrestin 3 mediates endocytosis of CCR7 following ligation of CCL19 but not CCL21. J Immunol 181:4723–4732

Cavalli A, Lattion AL, Hummler E, Nenniger M, Pedrazzini T et al (1997) Decreased blood pressure response in mice deficient of the α1b-adrenergic receptor. Proc Natl Acad Sci U S A 94:11589–11594

Changeux J-P, Edelstein SJ (2005) Allosteric mechanisms of transduction. Science 308:1424–1428

Changeux JP, Edelstein S (2011) Conformational selection or induced fit? 50 years of debate resolved. F1000 Biol Rep 3:19

Christopoulos A, Kenakin T (2002) G protein-coupled receptor allosterism and complexing. Pharmacol Rev 54:323–374

Coffman TM (1997) A genetic approach for studying the physiology of the type1A (AT1A) angiotensin receptor. Semin Nephrol 17:404–411

Conklin BR, Hsiao EC, Claeysen S, Dunuis A, Srinivasan S, Srinivasan S, Forsayeth JR, Guettier JM, Chang WC, Pei Y, McCarthy KD, Nissenson RA, Wess J, Bockaert J, Roth BL (2008) Engineering GPCR sigbnaling pathways with RASSLs. Nat Methods 5:673–678

Coward P, Wada HG, Falk MS, Chan SD, Meng F, Akil H, Conklin BR (1998) Controlling signaling with a specifically designed Gi-coupled receptor in transgenic mice. Nat Biotechnol 17:165–169

Cui Q, Karplus M (2008) Allostery and cooperativity revisited. Protein Sci 17:1295–1307

Dill KA, Chan HS (1997) From Levinthal to pathways to funnels. Nat Struct Biol 4:10–19

Dong S, Rogan SC, Roth BL (2010) Directed molecular evolution of DREADDs: a generic approach to creating next-generation RASSLs. Nat Protoc 5:561–573

Dror RO, Jensen MO, Borhani DW, Shaw DE (2010) Exploring atomic resolution physiology on a femtosecond to millisecond timescale using molecular dynamics simulations. J Gen Physiol 135:555–562

Dror RO, Arlow DH, Shaw DE et al (2011) Activation mechanism of the β2-adrenergic receptor. Proc Natl Acad Sci U S A 108:18684–18689

Duttaroy A, Zimliki CL, Gautam D, Cui Y, Mears D, Wess J (2004) Muscarinic stimulation of pancreatic insulin and glucagon release is abolished in M3 muscarinic acetylcholine receptor–deficient mice. Diabetes 53:1714–1720

Ehlert FJ (1988) Estimation of the affinities of allosteric ligands using radioligand binding and pharmacological null methods. Mol Pharmacol 33:187–194

Ehlert FJ (2005) Analysis of allosterism in functional assays. J Pharmacol Exp Ther 315:740–754

Faget L, Erbs E, Le Merrer J, Scherrer G, Matifas A, Benturquia N, Noble F, Decossas M, Koch M, Kessler P, Vonesch JL, Schwab Y, Kieffer BL, Massotte D (2012) In vivo visualization of delta opioid receptors upon physiological activation uncovers a distinct internalization profile. J Neurosci 32:7301–7310

Frauenfelder H, Parak F, Young RD (1988) Conformational substates in proteins. Annu Rev Biophys Biophys Chem 17:451–479

Frauenfelder H, Sligar SG, Wolynes PG (1991) The energy landscapes and motions of proteins. Science 254:1598–1603

Freire E (1998) Statistical thermodynamic linkage between conformational and binding equilibria. Adv Protein Chem 51:255–279

Furchgott RF, Harper NJ, Simmonds AB (1966) The use of β-haloalkylamines in the differentiation of receptors and in the determination of dissociation constants of receptor-agonist complexes. In: Harper NJ, Simmonds AB (eds) Advances in drug research. Academic Press, New York, pp 21–55

Gesty-Palmer D, Luttrell LM (2011) 'Biasing' the parathyroid hormone receptor: a novel anabolic approach to increasing bone mass? Br J Pharmacol 164:59–67

Gesty-Palmer D, Chen M, Reiter E, Ahn S, Nelson CD, Wang S, Eckhard AE, Cowan CL, Spurney RF, Luttrell LM, Lefkowitz RJ (2006) Distinct β-arrestin-and G protein-dependent pathways for parathyroid hormone receptor-stimulated ERK1/2 activation. J Biol Chem 281:10856–10864

Gesty-Palmer D, Yuan L, Martin B, Wood WH III, Lee MH, Janech MG, Tsoi LC, Zheng WJ, Luttrell LM, Maudsley S (2013) β-Arrestin-selective G proteincoupled receptor agonists engender unique biological efficacy in vivo. Mol Endocrinol 27:296–314

Giguere PM, Kroeze WK, Roth BL (2014) Tuning up the right signal: chemical and genetic approaches to study GPCR function. Curr Opin Cell Biol 27:51–55

Gonzalez E, Kulkarni H, Bolivar H, Mangano A, Sanchez R, Catano G, Nibbs RJ, Freedman BI, Quinones MP, Bamshad MJ et al (2005) The influence of CCL3L1 gene-containing segmental duplications on HIV-1/AIDS susceptibility. Science 307:1434–1440

Hall DA (2000) Modeling the functional effects of allosteric modulators at pharmacological receptors: an extension of the two-state model of receptor activation. Mol Pharmacol 58:1412–1423

Hall DA (2006) Predicting dose-response curve behavior: mathematical models of allosteric-ligand interactions. In: Bowery NG (ed) Allosteric receptor modulation in drug targeting. Taylor and Francis, New York, pp 39–78

Hauser MA, Legler DF (2016) Common and biased signaling pathways of the chemokine receptor CCR7 elicited by its ligands CCL19 and CCL21 in leukocytes. J Leukoc Biol 99:869–882

Hauser AS, Attwood MM, Rask-Andersen M, Schiöth HB, Gloriam DE (2017) Trends in GPCR drug discovery: new agents, targets and indications. Nat Rev Drug Disc 16:829–842

Hilser J, Freire E (1997) Predicting the equilibrium protein folding pathway: structure-based analysis of staphylococcal nuclease. Proteins 27:171–183

Hilser VJ, Dowdy D, Oas TG, Freire E (1998) The structural distribution of cooperative interactions in proteins: analysis of the native state ensemble. Proc Natl Acad Sci U S A 95:9903–9908

Hilser VJ, García-Moreno EB, Oas TG, Kapp G, Whitten ST (2006) A statistical thermodynamic model of the protein ensemble. Chem Rev 106:1545–1558

Hu J, Stern M, Gimenez LE, Wanka L, Zhu L, Rossi M, Meister J, Inoue A, Beck-Sickinger AG, Gurevich VV, Wess J (2016) A G prptein designer G protein coupled receptor useful for studying the physiological relevance of the Gq/11-dependent signaling pathways. J Biol Chem 291:7809–7820

Huang XP, Karpiak J, Kroeze WK, Zhu H, Chen X, Moy SS, Saddoris KA, Nikolova VD, Farrell MS, Mangano SW, Deshpande DA, Jiang A, Penn RB, Jin J, Koller BH, Kenakin T, Shoichet BK, Roth BL (2015) Allosteric ligands for the pharmacologically dark receptors GPR68 and GPR65. Nature 527:477–483

Ichikawa O, Fujimoto K, Yamada A, Okazaki S, Yamazaki K (2016) G-protein/β-arrestin-linked fluctuating network of G-protein-coupled receptors for predicting drug efficacy and bias using short-term molecular dynamics simulation. PLoS One 11:e0155516

Ito M, Oliverio MI, Mannon PJ, Best CF, Maeda N et al (1995) Regulation of blood pressure by the type 1A angiotensin II receptor gene. Proc Natl Acad Sci U S A 92:3521–3585

Jarpe MB, Knall C, Mitchell FM, Buhl AM, Duzic E, Johnson GL (1998) [D-Arg1, D-Phe5,D-Trp7,9,Leu11]substance P acts as a biased agonist toward neuropeptide and chemokine receptors. J Biol Chem 273:3097–3104

Kelly E (2013) Efficacy and ligand bias at the μ-opioid receptor. Br J Pharmacol 169:1430–1446

Kenakin T (1995) Agonist-receptor efficacy. II. Agonist trafficking of receptor signals. Trends Pharmacol Sci 16:232–238

Kenakin T (2002) Drug efficacy at G protein-coupled receptors. Annu Rev Pharmacol Toxicol 42:349–379

Kenakin T (2005) New concepts in drug discovery: collateral efficacy and permissive antagonism. Nat Rev Drug Discov 4:919–927

Kenakin TP (2009) Cellular assays as portals to seven-transmembrane receptor-based drug discovery. Nat Rev Drug Disc 8:617–626

Kenakin TP (2010) Ligand detection in the allosteric world. J Biomol Screen 15:119–130

Kenakin TP (2012) Biased signalling and allosteric machines: new vistas and challenges for drug discovery. Br J Pharmacol 165:1659–1669
Kenakin TP (2013) Making receptors a reality: the 2012 Nobel Prize in Chemistry. Trends Pharmacol Sci 34:2–5
Kenakin T (2015) New lives for seven transmembrane receptors as drug targets. Trends Pharmacol Sci 36:705–706
Kenakin T (2017) A system-independent scale of agonism and allosteric modulation for assessment of selectivity, bias, and receptor mutation. Mol Pharmacol 92:1–11
Kenakin TP, Morgan PH (1989) Theoretical effects of single and multiple transducer receptor coupling proteins on estimates of the relative potency of agonists. Mol Pharmacol 35:214–222
Kenakin T, Watson C, Muniz-Medina V, Christopoulos A, Novick S (2012) A simple method for quantifying functional selectivity and agonist bias. ACS Chem Neurosci 3:193–203
Kilts JD, Connery HS, Arrington EG, Lewis MM, Lawler CP, Oxford GS, O'Malley KL, Todd RD, Blake BL, Nichols DE et al (2002) Functional selectivity of dopamine receptor agonists. II. Actions of dihydrexidine in D2L receptor-transfected MN9D cells and pituitary lactotrophs. J Pharmacol Exp Ther 301:1179–1189
Kobilka BK, Frielle T, Collins S, Yang-Feng T, Kobilka TS, Francke U, Lefkowitz RJ, Caron MG (1987) An intronless gene encoding a potential member of the family of receptors coupled to guanine nucleotide regulatory proteins. Nature 329:75–79
Koblish M, Carr R III, Siuda ER, Rominger DH, Gowen-MacDonald W, Cowan CL, Crombie AL, Violin JD, Lark MW (2017) TRV0109101, a G protein-biased agonist of the m-opioid receptor, does not promote opioid-induced mechanical allodynia following chronic administration. J Pharmacol Exp Ther 362:254–262
Kohout TA, Nicholas SL, Perry SJ, Reinhart G, Junger S, Struthers RS (2004) Differential desensitization, receptor phosphorylation, beta-arrestin recruitment, and ERK1/2 activation by the two endogenous ligands for the CC chemokine receptor 7. J Biol Chem 279:23214–23222
Kong KC, Butcher AJ, McWilliams P, Jones D, Wess J, Hamdan FF, Werry T, Rosethorne EM, Charlton SJ, Munson SE, Cragg HA, Smart AD, Tobin AB (2010) M3-muscarinic receptor promotes insulin release via receptor phosphorylation/arrestin-dependent activation of protein kinase D1. Proc Natl Acad Sci U S A 107(21181–86):157
Kudlacek O, Waldhoer M, Kassack MU, Nickel P, Salmi JI, Freissmuth M, Nanoff C (2002) Biased inhibition by a suramin analogue of A1-adenosine receptor/G protein coupling in fused receptor/G protein tandems: the A1-adenosine receptor is predominantly coupled to Goa in human brain. Naunyn Schmiedebergs Arch Pharmacol 365:8–16
Lawler CP, Prioleau C, Lewis MM, Mak C, Jiang D, Schetz JA, Gonzalez AM, Sibley DR, Mailman RB (1999) Interactions of the novel antipsychotic aripiprazole (OPC-14597) with dopamine and serotonin receptor subtypes. Neuropsychopharmacology 20:612–627
Liapakis G, Cordomi A, Pardo L (2012) The G protein-coupled receptor family: actors with many faces. Curr Pharm Des 18:175–185
Link RE, Desai K, Hein L, Stevens ME, Chruscinski A, Bernstein D, Barsh GS, Kobilka BK (1996) Cardiovascular regulation in mice lacking α2-adrenergic receptor subtypes b and c. Science 273 (803–5):88
M'Kadmi C, Lyris J-P, Onfroy L, Gales C, Sauliere A, Gagne D, Damian M, Mary S, Maingot M, Denoyelle S, Verdie P, Fehrentz J-A, Martinez J, Baneres J-L, Marie J (2015) Angonism, antagonism, and inverse agonism bias at the ghrelin receptor signaling. J Biol Chem 290:27021–27039
MacMillan LB, Hein L, Smith MS, Piascik MT, Limbird LE (1996) Central hypotensive effects of the α2a-adrenergic receptor subtype. Science 273:801–803
Manglik A, Kobilka B (2014) The role of protein dynamics in GPCR function: insights from the β2AR and rhodopsin. Curr Opin Cell Biol 27:136–143
Manglik A, Kim TH, Masureel M, Altenbach C, Yang Z, Hilger D, Lerch MT, Kobilka TS, Thian FS, Hubbell WL, Prosser RS, Kobilka BK (2015) Structural insights into the dynamic process of beta-2 adrenergic receptor signaling. Cell 161:1101–1111

Manning DR (2002) Measures of efficacy using G proteins as endpoints: differential engagement of G proteins through single receptors. Mol Pharmacol 62:451–452
Michel MC, Alewijnse AE (2007) Ligand-directed signaling: 50 ways to find a lover. Mol Pharmacol 72:1097–1099
Miller DW, Dill KA (1997) Ligand binding to proteins: the binding landscape model. Protein Sci 6:2166–2179
Monod J, Wyman J, Changeux J-P (1965) On the nature of allosteric transitions; a plausible model. J Mol Biol 12:88–118
Motlagh HN, Wrabl JO, Hilser VJ (2014) The ensemble nature of allostery. Nature 508:331–339
Muniz-Medina VM, Jones S, Maglich JM, Galardi C, Hollingsworth RE, Kazmierski WM, Ferris RG, Edelstein MP, Chiswell KE, Kenakin TP (2009) The relative activity of "function sparing" HIV-1 entry inhibitors on viral entry and CCR5 internalization: is allosteric functional selectivity a valuable therapeutic property? Mol Pharmacol 75:490–501
Naylor J, Suckow AT, Seth A, Baker DJ, Sermadiras I, Ravn P, Howers R, Li J, Snaith MR, Coghlan MP, Hornigold DC (2016) Use of CRISPR/Cas9-engineered INS-1 pancreatic β-cells to define the pharmacology of dual GIPR/GLP-1R agonists. Biochem J 473:2881–2891
Nygaard R, Zou Y, Dror RO, Mildorf TJ, Arlow DH, Manglik A et al (2013) The dynamic process of beta(2)-adrenergic receptor activation. Cell 152:532–542
Onaran HO, Costa T (1997) Agonist efficacy and allosteric models of receptor action. Ann N Y Acad Sci 812:98–115
Onaran HO, Scheer A, Cotecchia S, Costa T (2000) A look at receptor efficacy. Fromthesignalingnetworkofthecelltotheintramolecularmotionofthereceptor. In: Kenakin TP, Angus JA (eds) The pharmacology of functional, biochemical, and recombinant systems, Handbook of experimental pharmacology, vol 148. Springer, Heidelberg, pp 217–280
Onaran HO, Ambrosio C, Ugur Ö, Madaras Koncz E, Grò MC, Vezzi V, Rajagopal S, Costa T (2017) Systematic errors in detecting biased agonism: analysis of current methods and development of a new model-free approach. Sci Rep 7:44247
Park PS (2012) Ensemble of G protein-coupled receptor active states. Curr Med Chem 19:1146–1154
Peng J, Bencsik M, Louie A, Lu W, Millard S, Nguyen P, Burghardt A, Majumdar S, Wronski TJ, Halloran B, Conklin BR, Nissenson RA (2008) Conditional expression of a Gi-coupled receptor in osteoblasts results in trabecular osteopenia. Endocrinology 149:1329–1337
Poulin B, Butcher A, McWilliams P, Bourgognon JM, Pawlak R, Kong KC, Bottrill A, Mistry S, Rosethorne EM, Charlton SJ, Tobin AB (2010) The M3-muscarinic receptor regulates learning and memory in a receptor phosphorylation/arrestin-dependent manner. Proc Natl Acad Sci U S A 107:9440–9445
Pradhan AA, Becker JA, Scherrer G, Tryoen-Toth P, Filliol D, Matifas A, Massotte D, Gavériaux-Ruff C, Kieffer BL (2009) In vivo delta opioid receptor internalization controls behavioral effects of agonists. PLoS One 4:e5425
Price MR, Baillie GL, Thomas A, Stevenson LA, Easson M, Goodwin R, McLean A, McIntosh L, Goodwin G, Walker G et al (2005) Allosteric modulation of the cannabinoid CB1 receptor. Mol Pharmacol 68:1484–1495
Raehal KM, Walker JK, Bohn LM (2005) Morphine side effects in b-arrestin 2 knockout mice. J Pharmacol Exp Ther 314:1195–1201
Rang HP (2009) The receptor concept: pharmacology's big idea. Br J Pharmacol 147:S9–S16
Rasmussen SG, Choi HJ, Fung JJ, Paron L, Casarosa P, Chae PS, Devree BT, Rosenbaum DM, Thian FS, Kobilka TS, Schnapp A, Konetzki I, Sunahar RK, Gellman SH, Pautsch A, Steyaert J, Weis WI, Kobilka BK (2011) Structure of a nanobody-stabilized active state of the beta (2) adrenoceptor. Nature 469:175–180
Redfern CH, Coward P, Degtyarev MY, Lee EK, Kwa AT, Salomonis N, Cotte N, Nanevicz T, Fidelman N, Desai K, Vranizan K, Lee EK, Coward P, Shah N, Warrington JA, Fishman GI, Bernstein D, Baker AJ, Conklin BR (1999) Conditional expression and signaling of a specifically designed Gi-coupled receptor in transgenic mice. Nat Biotechnol 17:165–169
Rees S, Morrow D, Kenakin T (2002) GPCR drug discovery through the exploitation of allosteric drug binding sites. Receptors Channels 8:261–268

Rohrer DK, Chruscinski A, Schauble EH, Bernstein D, Kobilka BK (1999) Cardiovascular and metabolic alterations in mice lacking both β1- and β2-adrenergic receptors. J Biol Chem 274:16701–16708

Roth BL, Chuang DM (1987) Multiple mechanisms of serotonergic signal transduction. Life Sci 41:1051–1064

Samama P, Cotecchia S, Costa T, Lefkowitz RJ (1993) A mutation-induced activated state of the β2-adrenergic receptor: extending the ternary complex model. J Biol Chem 268:4625–4636

Saudou F, Amara DA, Dierich A, LeMeur M, Ramboz S, Segu L, Buhot MC, Hen R (1994) Enhanced aggressive behavior in mice lacking 5-HT1B receptor. Science 265:1875–1878

Scherrer G, Tryoen-Toth P, Filliol D, Matifas A, Laustriat D, Cao YQ, Basbaum AI, Dierich A, Vonesh J-L, Gavériaux-Ruff C, Kieffer BL (2006) Knockin mice expressing fluorescent δ-opioid receptors uncover Gprotein–coupled receptor dynamics in vivo. Proc Natl Acad Sci U S A 103:9691–9696

Scherrer G, Imamachi N, Cao YQ, Contet C, Mennicken F, O'Donnell D, Kieffer BL, Basbaum AI (2009) Dissociation of the opioid receptor mechanisms that control mechanical and heat pain. Cell 137:1148–1159

Shapiro DA, Renock S, Arrington E, Chiodo LA, Liu LX, Sibley DR, Roth BL, Mailman R (2003) Aripiprazole, a novel atypical antipsychotic drug with a unique and robust pharmacology. Neuropsychopharmacology 28:1400–1411

Shenoy SK, Lefkowitz RJ (2011) β-Arrestin-mediated receptor trafficking and signal transduction. Trends Pharmacol Sci 32:521–533

Sora I, Takahashi N, Funada M, Ujike H, Revay RS, Donovan DM, Miner LL, Uhl GR (1997) Opiate receptor knockout mice define μ receptor roles in endogenous nociceptive responses and morphine-induced analgesia. Proc Natl Acad Sci U S A 94:1544–1549

Spengler D, Waeber C, Pantaloni C, Holsboer F, Bockaert J, Seeburg PH, Journot L (1993) Differential signal transduction by five splice variants of the PACAP receptor. Nature 365:170–175

Stephenson RP (1956) A modification of receptor theory. Br J Pharmacol Chemother 11:379–393

Stockton JM, Birdsall NJ, Burgen AS, Hulme EC (1983) Modification of the binding properties of muscarinic receptors by gallamine. Mol Pharmacol 23:551–557

Susulic VS, Frederich RC, Lawitts J, Tozzo E, Kahn BB, Harper M-E, Himms-Hagen J, Flier JS, Lowell BB (1995) Targeted disruption of the β3adrenergic receptor gene. J Biol Chem 270:29483–29492

Thompson GL, Lane JR, Coudrat T, Sexton PM, Chistopoulos A, Canals M (2016) Systematic analysis of factors influencing observations of biased agonism at the mu-opioid receptor. Biochem Pharmacol 113:70–87

Torrecilla I, Spragg FJ, Poulin B, McWilliams PJ, Mistry SC et al (2007) Phosphorylation and regulation of a G protein-coupled receptor by protein kinase CK2. J Cell Biol 177:127–137

Urban DJ, Roth BL (2015) DREADDs (designer receptors exclusively activated by designer drugs): chemogenetic tools with therapeutic utility. Annu Rev Pharmacol Toxicol 55:399–417

Vardy E, Roth BL (2013) Conformational ensembles in GPCR activation. Cell 152:385–386

Violin JD, Dewire SM, Barnes WG, Lefkowitz RJ (2006) G protein-coupled receptor kinase and b-arrestin-mediated desensitization of the angiotensin II type 1A receptor elucidated by diacylglycerol dynamics. J Biol Chem 281:36411–36419

Violin JD, DeWire SM, Yamashita D, Rominger DH, Nguyen L, Schiller K, Whalen EJ, Gowen M, Lark MW (2010) Selectively engaging b-arrestins at the angiotensin II type 1 receptor reduces blood pressure and increases cardiac performance. J Pharmacol Exp Ther 335:572–579

Wang S, Wacker D, Anat Levit A, Che T, Betz RM, McCorvy JD, Venkatakrishnan AJ, Huang XP, Dror RO, Shoichet BK, Roth BL (2017) D4 dopamine receptor high-resolution structures enable the discovery of selective agonists. Science 358:381–386

Whistler JL, Enquist J, Marley A, Fong J, Gladher F, Tsuruda P, Murray SR, von Zastrow M (2002) Modulation of postendocytic sorting of G protein-coupled receptors. Science 297:615–620

White KL, Scopton AP, Rives ML, Bikbulatov RV, Polepally PR, Brown PJ, Kenakin T, Javitch JA, Zjawiony JK, Roth BL (2014) Identification of novel functionally selective k-opioid receptor scaffolds. Mol Pharmacol 85:83–90

Willins DL, Berry SA, Alsayegh L, Backstrom JR, Sanders-Bush E, Friedman L, Roth BL (1999) Clozapine and other 5-hydroxytryptamine-2A receptor antagonists alter the subcellular distribution of 5-hydroxytryptamine-2A receptors in vitro and in vivo. Neuroscience 91:599–606

Woodward C (1993) Is the slow exchange core the protein folding core? Trends Biochem Sci 18:359–360

Xu M, Moratalla R, Gold LH, Hiroi N, Koob GF et al (1994) Dopamine D1 receptor mutant mice are deficient in striatal expression of dynorphin and in dopamine-mediated behavioral responses. Cell 79:729–742

The Evolving Landscape of Cancer Therapeutics

Madeha Khan and James Spicer

Contents

1 Biology of Cancer 44
1.1 Cancer as a Genetic Disease 44
1.2 Signalling in Health and Malignant Disease 45
1.3 Tumour Microenvironment and Host Immunity 46
2 Introduction 47
2.1 History 47
2.2 Roles for Systemic Therapies 48
2.3 Cytotoxic Drugs 49
2.4 Targeted Therapies 50
2.5 Immunotherapy 51
3 Clinical Trials in Oncology 53
3.1 Phase 1 Trials 53
3.2 Phase 2 Trials 54
3.3 Phase 3 Trials 55
4 Rationally Designed Therapies 57
4.1 Ligands as a Target 58
4.2 Targeting Receptors 59
4.3 Other Targets 62
4.4 Targeting the T-Cell 63
5 Nuclear Medicine Therapies 64
5.1 Radio-Iodine in Thyroid Cancer 64
5.2 Somatostatin Analogues 65
5.3 Radium-223 65
5.4 PSMA Ligand: Lutetium 66

M. Khan
Guy's Hospital, London, UK

J. Spicer (✉)
King's Health Partners at Guy's Hospital, London, UK
e-mail: james.spicer@kcl.ac.uk

J. E. Barrett et al. (eds.), *Concepts and Principles of Pharmacology*,
Handbook of Experimental Pharmacology 260, https://doi.org/10.1007/164_2019_312

6 Pharmacogenetics, Pharmacogenomics and Patient Selection for Treatment 66
6.1 Lung Cancer and *EGFR* Mutations 66
6.2 *BRCA1*, *BRCA2* Mutation and PARP Inhibition 67
6.3 Prediction of Toxicity 68
7 Resistance Mechanisms 68
7.1 *CML BCR-ABL* Mutations and Resistance to Imatinib 69
7.2 Molecular Markers of Resistance to EGFR Inhibition 70
7.3 ALK Resistance 72
8 Cancer Drug Discovery and Preclinical Development 72
9 Current Issues in the Development of Drugs in Oncology 73
9.1 Improving the Odds of Success of Phase 3 Trials 73
9.2 Regulation of Cancer Drug Development 74
10 Conclusion and Future Perspectives 75
References 76

Abstract

The last 100 years have seen a dramatic alteration in the treatment of cancer. Aside from small molecule inhibitors of protein tyrosine kinases, monoclonal antibodies have also been found to provide valuable therapeutic approaches for modulating tumour pathophysiology. As our knowledge of cancer biology improves, the specificity of this new generation of drugs is generally delivering an improved therapeutic ratio compared to traditional cytotoxic agents. However, patient selection through the use of biomarkers is key in optimising efficacy and improving cost-effectiveness. The most recent wave of revolutionary new systemic therapy approaches to cancer has arrived in recent years in the form of immune checkpoint inhibitors, now clinically validated as modulators of immune-regulatory pathways. The future of oncology therapeutics includes a combination of cytotoxic agents, targeted therapies and immunotherapy.

Keywords

Checkpoint inhibition · Cytotoxic · Drug resistance · Immunotherapy · Monoclonal antibody · Oncogene · Signalling · Tyrosine kinase

1 Biology of Cancer

1.1 Cancer as a Genetic Disease

The malignant phenotype of cancer is driven by a series of genetic aberrations in a cell clone that evolves in Darwinian fashion to form a tumour. The first recognition of a cellular origin for viral oncogenes was made in 1970. Proto-oncogenes exist in the normal genome and generally encode proteins that have an important role to play in regulating normal cell growth, proliferation and development. If these genes are dysregulated, they have the potential to contribute to tumourigenesis (Croce 2008). Another important group of genetic contributors to initiating and maintaining malignancy are tumour suppressor genes, with an anti-proliferative effect in the normal cell.

Tumour suppressor genes undergo genetic change such as deletion and missense mutation, resulting in loss of function in cancer cells. MicroRNAs are the products of genes that do not encode any protein. These short RNA sequences contribute to tumourigenesis by complementing the sequence of specific mRNAs and so preventing their translation.

The hallmarks of cancer can be considered within a conceptual framework entailing the fundamental aspects of neoplastic biology. Genetic insults to oncogenes and tumour suppressor genes contribute to tumour formation by affecting key aspects of this biology (Fig. 2) (Hanahan and Weinberg 2011). For example, the action of oncogenes may lead to abnormal growth and proliferation in the absence of appropriate signals, failure of programmed cell death, upregulated angiogenesis or unconstrained replication potential. Loss of tumour suppressor gene function can result in absence of normal signals controlling cell division.

There may be hundreds of genetic changes to the germline genome in a single cancer cell. The mechanisms giving rise to these mutations are only partially understood. In some cancers there is clearly a role for chemical carcinogens (cigarette smoke in lung cancer). In others, oncogenic viral genes act to inactivate tumour suppressors in infected cells, exemplified by E6 antigen expressed by human papilloma virus. E6 inactivates p53 and contributes to the increased incidence of cervical carcinoma in individuals infected by this virus.

In some cases, there is inherited susceptibility to genetic events in families. For example, a defective allele of the tumour suppressor genes *BRCA1* or *BRCA2* is inherited by some patients with breast, ovarian or prostate cancer. More commonly, these events occur in the somatic genome. Multistep tumourigenesis refers to a serial accumulation of insults and partly accounts for the fact that cancer is more common in a large long-lived organism such as man. Between 1,000 and 10,000 mutations have been implicated as contributing to a single human cancer, the majority of which affect dominantly acting oncogenes (Stratton 2011). Defects in a critical subset of genes, known as driver mutations, must arise in a single cell in order to give rise to a malignant phenotype. These are thought to be critical to providing a survival advantage to the cancer clone, whilst a larger number are likely to be passenger molecular events arising in an increasingly unstable genome. This genetic instability gives rise to the multiple and heterogeneous clone characteristic of a mature cancer. An exponential expansion in understanding of these aspects of cancer biology has defined potential targets for new therapies, a number of which has been approved in the past two decades.

1.2 Signalling in Health and Malignant Disease

Cellular signalling pathways composed of extracellular soluble ligands, transmembrane receptors and intricate intracellular kinase cascades are ubiquitous in nature. The ErbB receptor family has been more extensively studied than any other signal transduction network. EGFR is a receptor tyrosine kinase in this family, which consists of four members: EGFR (HER1/ErbB1), HER2 (ErbB2), HER3 (ErbB3) and HER4 (ErbB4) (Salomon et al. 1995). Ligand binding results in rapid receptor dimerisation, phosphorylation and activation of intracellular signalling pathways, which in turn leads to cell growth, proliferation and differentiation (Yarden and Sliwkowski 2001). ErbB receptors undergo various types of alteration and

dysregulation in human tumours including gene amplification, receptor overexpression, activating mutations, overexpression of receptor ligands and/or loss of negative regulatory controls (Fig. 3) (Krause and Van Etten 2005). These tyrosine kinases can be targeted both by inhibitors of the intracellular signalling domain and by monoclonal antibodies specific for the extracellular ligand-binding domain.

1.3 Tumour Microenvironment and Host Immunity

Tumours evolve through multistep tumourigenesis to form complex 'organs'. It is not only the individual carcinogenic cells that define its properties but also the microenvironment which nurtures its development and progression. The tumour microenvironment is composed of multiple cell types including lymphoid, myeloid, stromal and endothelial cells. The microenvironment is a hostile environment therapeutically due to low pH, necrosis, hypoxia, shortage of nutrient and the presence of immunosuppressive host immune components (Riviere and Sadelain 2017; Sadelain et al. 2017). Between and within patients, there is tumour heterogeneity which is reflected in histopathology showing varying degrees of differentiation, invasion, inflammation and vascularity. During tumour evolution there are progressive changes within the microenvironment (Hanahan and Weinberg 2011). Advances in immunotherapy have arisen from the current understanding of the ability of a tumour to circumvent host immunity.

Genetic alterations within a tumour cell result in the expression of neoantigens, which are processed into peptides that can bind to the major histocompatibility class I (MCHI) molecules on the surface of cancer cells, differentiating them from normal cells. These cancer-specific peptide-MHCI complexes can be recognised by host CD8+ T-cells. However, through evolutionary deletion of this complex, known as immune editing, cancer cells can avoid host attack (Chen and Mellman 2013). Expression on the tumour cell surface of ligands for inhibitory T-cell receptors such as PD1 provides another mechanism for immune evasion. Immune-mediated tumour cell death releases neoantigens, to be captured by dendritic cells in regional lymph nodes, which in turn prime and activate effector T-cells by presenting antigens in the context of MHCI and MHC II molecules. T-cells may subsequently traffic to and infiltrate the tumour, where they recognise and bind to cancer cells through interaction with the T-cell receptor and co-stimulatory pathways to cause further tumour cell death, completing the so-called cancer-immunity cycle (Chen and Mellman 2013).

The tumour microenvironment plays a critical role in the modulation of immune activity. In a microenvironment which is otherwise not responsive to immunotherapy, chemotherapeutics can be used to sensitise the tumour to become immunogenic. In vitro studies have demonstrated that chemotherapeutics induce a systemic host response including adaptive immunity. They influence tumour-host interactions, stimulating CD8+ T-cell activation and infiltration into the tumour microenvironment in otherwise T-cell-naive tumours. Immunogenic chemotherapeutics also have direct actions on the tumour bed. Collectively, these processes can sensitise tumours to immunotherapy. Together, chemotherapeutics

and checkpoint inhibitors can provide a synergistic treatment option for tumours resistant to checkpoint blockade therapy alone (Chen and Mellman 2013).

2 Introduction

2.1 History

The introduction of both surgery and radiotherapy for the treatment of cancer predate the advent of drug therapy by many years. First used clinically more than a century ago, radiotherapy was used either alone or in conjunction with surgery as the only available treatment modality until the first trials with cancer-targeting drugs in the 1940s. Approximately two thirds of patients will require radiotherapy during their cancer treatment. X-rays were first used diagnostically by Wilhelm Conrad Rontgen in 1895. Subsequently, skin cancers were treated with x-rays due to low tissue penetration. In the early days dosimetry was unsophisticated, and toxicities often outweighed the benefits of treatment. However, by the 1920s the radiobiological properties of electromagnetic radiation (x-rays and gamma rays), particles (electrons, protons and neutrons) and radioactive isotopes (particularly radium) were better understood. Radiotherapy either directly damages cellular DNA or causes indirect damage through the production of free radicals, thereby damaging the genome of clonogenic tumour cells. This in turn leads to mitotic arrest and cell death when the cells enter mitosis without repair of this DNA damage. However, normal cells, particularly those that are rapidly dividing, may also be damaged. Therefore, radiation oncology clinicians and physicists have to plan accurately focused radiation beams, with fractionation of the total treatment dose to allow for maximum dose delivery to the tumour, whilst sparing normal tissue and allowing sufficient repair and recovery of non-tumour regions unavoidably included in the treatment field.

By the 1980s devices used to deliver proton beams were established, particularly to treat benign diseases such as keloid scarring. The following decades saw the marriage of machines delivering x-rays and advanced computer software allowing for three-dimensional conformal imaging. As computer systems became more sophisticated, intensity-modulated radiation therapy and stereotactic radiotherapy were introduced. In many centres, these techniques form the mainstay of radical treatment. In addition, we are now able to add a fourth dimension of time to accommodate for real-time motion as a result of the breathing cycle for treatment of tumours in the lungs and upper abdomen. Currently under trial is ‘adaptive radiotherapy’ that allows for repeating imaging in between fractions to account for alterations in the size and motion of tumours during radiotherapy, which is particularly useful for rapidly responding tumours. Radiotherapy continues to remain an exceptionally important mode of treatment in both a radical and palliative context (Gianfaldoni et al. 2017). Coupled with surgery, radiotherapy remains the mainstay of treatment for tumours localised at the time of diagnosis. However, in many cases tumours are metastatic at the time of presentation. A detailed discussion of surgery and radiotherapy in the treatment of cancer is beyond the scope of this chapter which focusses on the diverse systemic therapies developed since the dawn of medical oncology 75 years ago.

During the First World War, nitrogen mustards were deployed as a chemical weapon. Soldiers that survived exposure to nitrogen mustards were noted to have reversible leucopenia and mucocutaneous blistering. In the 1940s, drugs in the same class were first used in clinical trials for the treatment of haematological malignancies (Gilman and Philips 1946) with promising outcomes. By the 1960s newer cytotoxic drugs were made available so that diseases such as leukaemia and some solid organ tumours, most notably germ cell malignancies, could be controlled by halting the dividing cell and, in cases such as testicular cancer, cured.

Over the next decades, the spectrum of cytotoxic agents expanded further with candidate drugs exhibiting antimitotic activity through a variety of mechanisms. The landmark discovery of platinum conjugates, particularly cisplatin (Rosenberg et al. 1969), allowed the first curative treatment for patients even with advanced testicular cancer. The phenomenon of tumour resistance to anticancer therapeutics was overcome in some contexts with the ability to safely combine multiple cytotoxic agents. Drug combinations have been particularly effective in haematological malignancies, especially aggressive lymphomas and acute lymphoblastic leukaemia. The potential for the more common tumours of epithelial origin such as breast, colorectal and lung cancer to benefit from cytotoxic drugs led to the development of further drug classes including taxanes and antimetabolites in the last two decades of the twentieth century. However, although these drugs demonstrate useful palliative and adjuvant efficacy in various settings, they failed to deliver the hoped-for outcome of cure in common advanced-stage solid tumours, partly due to evolving tumour cell resistance.

In response to a perceived stalling of progress with newer cytotoxics used in more complex and toxic combinations, drug discovery and clinical development in oncology began to focus at the end of the twentieth century on new classes of drugs, driven by progress in the molecular understanding of tumour biology. For example, transtuzumab, a monoclonal antibody targeting the oncogenic HER2 receptor, was licensed for the treatment of breast cancer in 1998. More recently, research into the host immune system's response to cancer has led to the development of immune checkpoint inhibitors and adoptive T-cell therapy.

2.2 Roles for Systemic Therapies

A common characteristic of cancers is a high mitotic index reflecting rapid cell proliferation. Antimitotic drugs, targeted agents and immunotherapies are each now widely used in the treatment of cancer patients, with either radical or palliative intent. In the radical setting, they are often used following surgery, commonly known as adjuvant therapy. In some rapidly proliferating lymphomas, acute leukaemia and germ cell tumours, chemotherapy alone may be curative. However, in more common metastatic tumours of epithelial origin such as breast, lung and colorectal, cure is almost never achieved. In these situations, the aim is improvement in quality of life through symptom control and extension of survival time.

With the use of novel targeted therapies, combination cytotoxic regimens and immune checkpoint inhibitors, a pivotal decision is often the selection of optimal therapy and sequence of treatment for each individual. Even where cure cannot be

offered with any certainty, cancer may now be considered a chronic disease in some tumour types. Throughout treatment the aim remains to prolong life and maintain a reasonable quality of life.

Systemic therapies are also used in the neoadjuvant and adjuvant setting. In the former, the goal is to reduce the size of the tumour, facilitating a successful outcome with radical surgery or radiotherapy. Adjuvant therapies make a meaningful reduction in the risk of relapse after radical treatment, thereby extending overall survival following surgical treatment in many common cancers including colorectal, breast and lung. The rationale for adjuvant therapies is that despite locally confined disease macroscopically, and using the most sensitive imaging techniques, it is apparent in retrospect that many patients had micrometastases at the time of radical local treatment. Studies showing circulating tumour cells and epithelial cells in the bone marrow even in patients with early stage cancers support this hypothesis.

Systemic therapies are used concurrently or sequentially with radiotherapy, particularly for locally advanced head and neck, lung, breast and some gastrointestinal cancers. Together, both modalities provide an efficacy not achievable by either modality alone.

2.3 Cytotoxic Drugs

Cytotoxic agents interact with the cellular machinery of mitosis, the cellular process critical to malignant proliferation. This machinery includes DNA synthesis, DNA structure, and tubulin-based cytoskeletal mitotic structures. Box 1 summarises the broad classes of cytotoxic agents. The inevitable consequences of targeting proliferating cells is that all organs with rapid rates of healthy cell turnover, such as hair follicles, mucosal surfaces in the gastrointestinal tract and bone marrow, are potentially affected. Haematopoietic lineages are affected in chronological sequences dictated by their normal circulating half-life. In this way, leucocytes are depopulated first, followed by platelets and then red cells. After the administration of chemotherapy, leucocytes take approximately 3 weeks to recover in the absence of exogenous growth factors. Therefore, chemotherapy is generally administered every 3 weeks.

Box 1 Summary of Broad Classes of Cytotoxic Drugs

Agents targeting DNA structure	
Alkylation	Cyclophosphamide CCNU Melphalan
Platinum coordination cross-linking	Cisplatin
Double-stranded cleavage via topoisomerase 2 antibiotics	Doxorubicin Daunorubicin Podophyllotoxins Etoposide Teniposide

(continued)

Box 1 (continued)

Single-stranded cleavage via topoisomerase 1	Topotecan Irinotecan
Intercalation blocking RNA synthesis	Dactinomycin
Uncertain mechanism	Bleomycin
Agents targeting DNA synthesis: antimetabolites	
Pyrimidine analogues	5-flurouracil Capecitabine (fluoropyrimidines)
Antifolates	Methotrexate (DHFR) Pemetrexed (TS;DHFR)
Agents targeting tubulin	
Taxanes (stabilise microtubules)	Paclitaxel Docetaxel Novel taxanes
Vinca alkaloids (inhibit tubular polymerisation)	Vinorelbine Vincristine Vinblastine

DHFR dihydrofolate reductase, *TS* thymidylate synthase

The primary dose-limiting factor for cytotoxic treatments is unwanted toxicities on normal tissues. It is regarded desirable to administer as high a dose as tolerable, both to maximise anticancer efficacy, but to limit toxicities affecting quality of life. Phase I trials are traditionally designed to reach a maximum tolerated dose, with the result that at approved doses, cytotoxic drugs and their combinations are usually associated with a narrow therapeutic index. In routine clinical practice, this translates to the necessity for a thorough assessment of fitness and comorbidities as an essential precursor to prescribing chemotherapy. In the context of clinical drug development, early-phase trials are generally conducted in populations of cancer patients, rather than in healthy volunteers, because some compensation for toxicity by clinical benefit is at least a possibility.

2.4 Targeted Therapies

By the first decade of the twenty-first century, a new era of cancer therapeutics was born with the exponential advent of many classes of drugs that mediated anticancer effects through targets other than the mitotic machinery. These have become generally known as ‘targeted therapies’.

Targeted therapies can be divided into two broad categories: therapeutic monoclonal antibodies or small molecules. The latter penetrate the cell membrane and can interact with their cellular targets. Unlike cytotoxics, these agents tend to have more intrinsic specificity for cancer cells, and so they are associated with a higher therapeutic ratio than cytotoxic drugs. Unlike cytotoxic therapies they are seldom associated with significant myelosuppression. However, on-target toxicities are still

observed because the targets for these agents often have a physiological role to play, in addition to their aberrant function in the cancer cell. In many cases significant efficacy has been observed at well-tolerated doses. Biological markers such as cell surface expression of antigens or hormone receptors, or molecular features in the cancer genome, are used to select patients who would benefit from these agents, thereby personalising the approach to cancer treatment.

2.5 Immunotherapy

Clinical validation of immunotherapy in the treatment of cancer has only recently been achieved, but the proposed concept that the patient's immune system might be induced to control tumour cells is over a century old. Neoantigens, resulting from tumourigenic mutations in the cancer genome, are expressed or presented on the cancer cell surface, potentially rendering these cells recognisable as non-self by host cytotoxic CD8+ T-cells. However, cancer cells can evade immune surveillance by expressing proteins such as PDL1, which can engage with its receptor, the inhibitory molecule PD1, on the T-cell surface. Inhibiting the PDL1/PD-1 interaction can

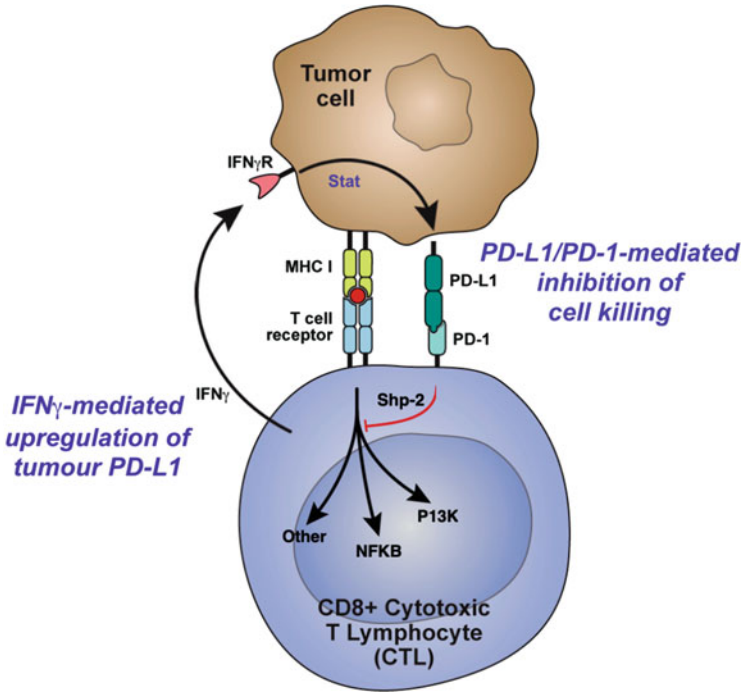


Fig. 1 Targeting the immune checkpoint. Neoantigens in cancer cells potentially render them recognisable as non-self. However, cancer cells can evade immune surveillance by expressing proteins such as PD-L1 recognised by the negative regulatory T-cell receptor PD1. Inhibiting the PD-L1/PD1 interaction can restore T-cell cytotoxic activity

restore antitumour T-cell activity (Fig. 1). The concept of immune evasion has been established as a biological hallmark of tumour capabilities (Fig. 2) (Hanahan and Weinberg 2011).

So-called checkpoint inhibitor monoclonal antibodies target the suppressive mechanisms at the interface between T-cell and tumour or between T-cells and antigen-presenting cells. Aside from PD1 and PDL1, these targets include CTLA-4, and the first of these checkpoint inhibitors gained marketing approval in 2013. This approach has shown unprecedented clinical benefit across multiple tumour groups, and in many indications is better tolerated than cytotoxic treatment, or offers increased efficacy alone or in combination. However, there remains a large proportion of patients that fail to respond to the current early generation checkpoint inhibitors. Measurement of tumour PDL1 expression has been used as a clinical biomarker to enrich the patient population for those that may respond to treatment. However, in practice, PD1 and PDL1 expression correlates poorly with clinical response. Laboratory studies show that the proportion of cancer cells responding to checkpoint inhibitors may be increased by combining them with immunogenic drugs which alter cell expression of proteins and/or the tumour microenvironment (Havel et al. 2019).

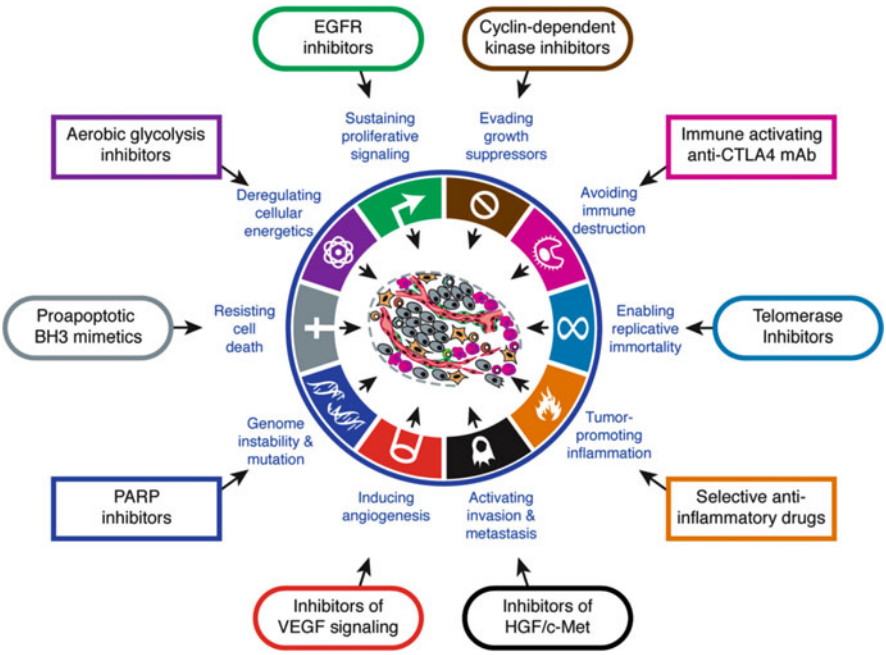


Fig. 2 Therapeutic targeting of the Hallmarks of Cancer. *Extracted with permission* (from (Hanahan and Weinberg 2011)) Properties recognised to be responsible for tumour evolution, and how respective cancer therapeutics are developed to target tumours, are depicted below

3 Clinical Trials in Oncology

3.1 Phase 1 Trials

Traditionally, Phase 1 oncology trials are conducted in patients with advanced cancer as opposed to healthy volunteers. This has been because of the low therapeutic ratio expected for cytotoxic drugs. More recently, with the advent of better-tolerated targeted agents, initial dosing in healthy volunteers is more commonly undertaken, especially where the toxicity profile is predictable from other drugs in the same class. For example, single-dose administration to explore initial pharmacokinetics can often be best conducted in this way. This strategy reduces the risk of exposing cancer patients to subtherapeutic doses.

The commonest dose escalation scheme remains the traditional '3 + 3' design, in which three patients are first treated at a given dose, with a further three added if a single dose-limiting toxicity (DLT; as predefined for each study) occurs. If no DLT is seen in the first three patients, or in no more than one of six patients in an expanded cohort, then dose may be escalated for the next cohort. A dose at which two or more DLTs occur is regarded as intolerable, and a dose level below this is likely to be explored as the maximum tolerated dose (MTD). There is no statistical basis to this trial design, but it has proved practical and informative in the development of countless drugs for the treatment of cancer. Nevertheless, newer strategies including accelerated dose escalation and the continuous reassessment method (CRM) are becoming more widely adopted (Piantadosi et al. 1998). CRM, based on Bayesian statistics, aims to adjust dose increments and cohort sizes by taking into account emergent toxicity data and has the potential for reducing the time and sample number required in a dose escalation trial.

The usual pharmacokinetic parameters such as C_{max}, $t_{1/2}$ and AUC are collected in cancer Phase 1 studies and may inform key decisions including dose escalation and dosing schedule. These data answer the fundamental question of whether drug can circulate at adequate concentrations to allow therapeutic effect, as predicted from target plasma concentration established in preclinical models.

Whether drug delivered to tumour can accumulate there and exert a biological effect requires on-treatment tumour biopsy, allowing for improved pharmacodynamic analysis. As an example, western blotting of phosphoproteins in a study of a kinase inhibitor may help to demonstrate successful target modulation and corresponding downstream effects. Serial sampling of tumours is preferred to collection of surrogate tissues such as skin or peripheral blood mononuclear cells. This is because drug penetration of normal tissues may differ from that in tumour, where abnormal vasculature, altered pH and hypoxia may give rise to very different localised effects. In general, patients participating in Phase 1 cancer trials are perhaps surprisingly willing to undergo these procedures. Further insight into the effect of a drug on tumours may be provided by functional imaging, for example, evaluating tumour perfusion or vessel permeability using MRI techniques. However, few such imaging endpoints are validated for making go/no-go drug development decisions.

As in other areas of experimental therapeutics, the primary objectives of a Phase 1 oncology trial are to study the safety profile of the drug and to establish a recommended Phase 2 dose (RP2D). Traditionally, dose is escalated to the maximum tolerated dose (MTD), but for drugs with known mechanism of action and available assays to demonstrate target inhibition, an alternative is to determine an optimal biological dose (Adjei 2006). This endpoint may appear better suited to trials of rationally designed targeted therapies where activity might be expected below the MTD, but nevertheless it has not become established as a standard (Parulekar and Eisenhauer 2004). This is because of concern that a reliance on the demonstration of PD effect in tumours may increase the risk of selecting a Phase 2 dose below maximal clinical activity, which might occur if the drug's mechanism of action is incompletely understood, or biomarkers of efficacy are misleading in samples from heterogeneous tumours (Yap et al. 2010). Ideally, randomised Phase 2 trials should compare biologically active doses with the higher MTD. If escalation to MTD is not possible, as with some new well-tolerated non-cytotoxic agents, RP2D may be selected based on PK and PD parameters.

The likelihood of efficacy is clearly an important metric when discussing trial participation with patients. Assessment of response is never a primary objective in a Phase 1 trial, but where seen this is of course extremely encouraging. There is evidence, in the current era of rationally designed drugs and with many trials combining novel agents with more established therapies, that on average 11% of Phase 1 patients experience a partial beneficial response (Horstmann et al. 2005). Protocols should preferably be written to allow enrichment with patients whose tumours express biomarkers believed to be predictive of response, and this can accelerate progression into late-phase clinical trials in an appropriately targeted population (Kwak et al. 2010).

3.2 Phase 2 Trials

Compared with other disciplines, Phase 2 trials in oncology have often been conducted as single-arm (non-randomised) studies with response rate as the primary endpoint. Two-stage designs incorporating early stopping rules in the event of lack of efficacy are widely used for ethical purposes to minimise the numbers of patients treated on an ineffective agent (Simon 1989). Lack of randomisation in cancer studies may have arisen from a reluctance to allocate patients with a life-threatening diagnosis to placebo, or to no treatment, in indications where there is no standard of care. Response rate (RR) was an obvious endpoint to focus on when most agents studied were cytotoxic and, if active, were expected to shrink tumours. RR in these Phase 2 trials is usually compared to historical controls, if available. However, multiple experiences of promising Phase 2 activity followed by a negative Phase 3 trial, as well as a shift to studies of targeted agents, have led to a renewed emphasis on randomisation in Phase 2 trials (Eisenhauer et al. 2009). Evidence-based treatment options in many cancers have expanded over recent years, so comparators for control arms are more likely to exist, although in some settings

a placebo control arm may still be appropriate. Strategies for reducing exposure to placebo are discussed below.

Another endpoint commonly used for efficacy assessment in Phase 2 trials of cancer drugs is progression-free survival (PFS, time from randomisation to disease progression), which may be a more meaningful surrogate of clinical benefit. PFS is also likely to allow a more appropriate assessment of efficacy of newer drug classes with mechanisms of action likely to block proliferation rather than induce apoptosis and tumour shrinkage.

Assessment of disease status in solid tumours is generally performed using CT scanning, and reproducible quantification of this is essential for determination of RR and PFS. An arbitrary but widely accepted technique for evaluating disease status is provided by the response evaluation criteria in solid tumours (RECIST) in which the long axis of selected target lesions is summed to provide a total measurement (Eisenhauer et al. 2009). Progression of disease is defined as an enlargement of the RECIST measurement by more than 20%, and conversely reduction by more than 30% represents a partial response (complete response if no assessable disease remains). RECIST disease assessment has, however, been widely criticised as being cumbersome and misleading, and some have argued for the use of RECIST to be 'resisted' (Ratain and Eckhardt 2004). Nevertheless, RECIST criteria have led to a useful international standard.

Additional imaging modalities such as PET and functional MRI are also frequently used in Phase 2 trials to assess efficacy and explore mechanism of action (Josephs et al. 2009). In some patients, measurable disease by radiological imaging is not present, and other measures of tumour burden are being evaluated as intermediate endpoints of clinical benefit, for example, circulating tumour-secreted proteins (tumour markers), tumour cell counts and circulating plasma nucleic acids. Prostate-specific antigen (PSA) is shed into the plasma in proportion to tumour burden, and criteria for PSA change in response to trial therapies have been agreed (Small and Roach 2002). Phase 2 trials provide an opportunity for development and initial validation of novel biomarkers to inform patient selection for future studies. This is especially important for therapies with a defined target where marketing approval may not be granted in the absence of a companion diagnostic to maximise efficacy in a defined patient population.

3.3 Phase 3 Trials

As in other disciplines, Phase 3 trials in oncology are randomised studies and wherever possible are designed so that both patient and investigator are blind to the treatment allocation (Booth and Tannock 2008). The control arm may be placebo, if there is no currently available evidence-based active treatment, or may be a standard treatment. Blinding may not be practical if, for example, an oral therapy is being compared to another administered parenterally. Key eligibility criteria include histological diagnosis, stage, prior therapies and performance status (an important measure of fitness and symptom burden) (Oken et al. 1982). Upper age

limits are rarely appropriate, but older patients have historically been significantly under-represented in Phase 3 oncology trials, clearly an undesirable situation when most common cancers are more common in older patients. An accepted primary endpoint for Phase 3 oncology trials is overall survival, which has the advantages of a lack of ambiguity or bias. However, as the treatment armamentarium expands in many tumour types, this endpoint is increasingly likely to be confounded by post-study therapies. As a result, PFS is increasingly accepted for registration trials. RECIST measurements in serial CT scans are generally used to assess this endpoint. PFS has clinical relevance in many cases because disease progression in metastatic cancer often causes worsening symptoms and deterioration in quality of life (QOL).

Prospective assessment of QOL is desirable in Phase 3 trials. This is especially the case in oncology where any improvements in symptoms, OS or PFS need to be counterbalanced by consideration of potentially considerable toxicity. The UK National Institute for Health and Care Excellence (NICE) analyses measures of efficacy including QOL and takes into account drug pricing when evaluating cost-effectiveness for use of new therapies in the UK National Health Service. NICE uses a measure of benefit that corrects survival improvement for QOL, called a QALY (quality-adjusted life year), so that greater value is attached to a year's extra survival at a perfect level of fitness than to the same period at an impaired level of function (Faden and Chalkidou 2011).

Phase 3 trials are large undertakings including sometimes thousands of patients treated at hundreds of centres and are therefore costly to conduct. It is self-evident that measures should be taken to maximise the chances of success, but in the era of targeted therapies, this has not always occurred. In fact oncology drugs are less likely to progress successfully through clinical development than most other clinical disciplines, with only 5% of cancer drugs awarded Investigational New Drug status going on to gain marketing approval (Adjei et al. 2009). By contrast, some of the most important Phase 3 results with novel agents have been obtained through careful selection of patients with tumours expressing the target, as in trials of trastuzumab in HER2+ breast cancer, or EGFR inhibition in *EGFR*-mutated non-small-cell lung cancer (Mok et al. 2009; Slamon and Pegram 2001). It is important to note that the predictive value of biomarkers such as *HER2* amplification or *EGFR* mutation can only be definitively confirmed in a randomised trial because this is the only way to exclude a purely prognostic effect of these markers.

The inclusion of a placebo arm in an oncology trial can be problematic and may impair recruitment because of patients' negative perceptions of this design. A number of strategies have been proposed for minimising exposure to placebo, including weighted randomisation and crossover design. Crossover is particularly suitable if OS is not the primary endpoint and allows patients on the placebo arm to receive experimental treatment upon progression. Interim analyses conducted by a robust data monitoring committee help to keep the sample number to a minimum and so minimise exposure to placebo in the control arm.

4 Rationally Designed Therapies

Advances in the understanding of cancer biology have led to the identification of new targets and driven the development of specific therapies directed against them. Examples include drug classes directed against ligands, receptors, intracellular signalling components and cellular machinery such as the proteasome and chromatin (Box 2).

Box 2 Targets for Anticancer Therapies: Examples of Approved and Investigational Drugs

Ligands	
Steroid Hormones	AIs Abiraterone Apalutamide Enzalutamide
VEGF	Bevacizumab
Receptors	
Oestrogen	Tamoxifen
erbB	Cetuximab Trastuzumab
Receptor tyrosine kinases	
erbB	Erlotinib Gefitinib Afatinib Osimertinib
VEGFR	Sunitinib Sorafenib Axitinib
MET	Hh
Intracellular kinases	
mTOR	Everolimus
BRAF	Vemurafenib
MEK	Trametinib Cobimetinib Binimetinib
bcr-abl	Imatinib
EML4-ALK	Crizotinib Brigatinib Alectinib
Proteasome	
	Bortezomib
Chromatin	
	HDACs Demethylase inhibitors PARP inhibitors – olaparib

(continued)

Box 2 (continued)

Immunomodulating antibodies	
PD1	Pembrolizumab Nivolumab
PDL1	Atezolizumab
CTLA4	Ipilimumab
Haematological targets	
CD20	Rituximab
CD52	Alemtuzumab
CD20	Ofatumumab

AI aromatase inhibitors, *HDAC* histone deacetylase, *Hh* hedgehog, *PARP* polyADPribose polymerase, *VRGF* vascular endothelial growth factor

4.1 Ligands as a Target

4.1.1 Oestrogen

Oestrogen is crucial for the growth and propagation of hormone-sensitive breast cancer. The circulating oestrogen ligand binds to and activates cytosolic receptors in tumour cells. These receptors are expressed in approximately 75% of all breast cancers, suggesting that these tumours may respond to oestrogen deprivation. Removing sites of oestrogen production (oophorectomy or irradiation), antagonising oestrogen activity or blocking oestrogen synthesis can all reduce available oestrogen. Tamoxifen is a selective oestrogen-receptor antagonist that blocks ligand binding, thereby blocking tumour cell proliferation. In premenopausal women with tumours strongly expressing oestrogen receptor (ER+), 5 years of adjuvant tamoxifen reduced the risk of recurrent disease and reduced the risk of death by 34% (Early Breast Cancer Trialists' Collaborative 2005). More recently a comprehensive statistical model, PREDICT 2.0, has been developed to assess the survival benefit of adjuvant hormone treatment over a 5- and 10-year period, based on numerous clinical factors. This model is widely used in clinical practice to select patients that are likely to benefit from adjuvant hormone, targeted or cytotoxic treatment (Punglia et al. 2018).

Tamoxifen, an oestrogen receptor antagonist, has long been the gold standard of endocrine treatment in ER+ breast cancer, but its use is associated with some significant (but uncommon) adverse effects including endometrial cancer and thromboembolism. Furthermore, a significant number of women receiving tamoxifen experience disease recurrence or progression, whether they are treated in the adjuvant or metastatic setting. There is therefore a need for further agents to treat tamoxifen-resistant disease.

Other drugs in the aromatase inhibitor category, such as letrozole and anastrozole, block the production of oestrogen by preventing the last step of oestrogen synthesis.

They are aromatase-specific and thus have little effect on the synthesis of other steroids or on the adrenal axis (Choueiri et al. 2004). Anastrozole, letrozole and exemestane have all been compared with tamoxifen in randomised studies in the metastatic setting. In a large Phase 3 randomised trial, letrozole demonstrated a superior outcome when compared with tamoxifen. Based upon these results, studies of aromatase inhibitors in the adjuvant setting were also undertaken comparing efficacy and toxicity with that of tamoxifen. In addition to replacing tamoxifen with an aromatase inhibitor as an initial adjuvant therapy, other strategies that have been investigated are switching between tamoxifen and an aromatase inhibitor or the addition of extended adjuvant aromatase inhibition after 5 years of tamoxifen. In practice, letrozole is used more commonly in women with postmenopausal status.

4.1.2 Androgen

In 1941 Charles Huggins demonstrated that at initial presentation, prostate cancer is an androgen-dependent cancer that responds to either surgical or hormonal withdrawal of circulating androgen (Huggins and Hodges 2002). Twenty-five years later he received the Nobel Prize for his observation. This led to the development of first-generation anti-androgens, for example, bicalutamide, a partial agonist of the androgen receptor. Inevitably, the disease enters a castration-refractory phase. There is good evidence that this phase is driven by upregulation of androgen receptors, leading to increased sensitivity to even low levels of circulating and intratumoural ligand. Second-generation anti-androgens, such as abiraterone, specifically inhibit the adrenal synthetic enzymes 17 alpha-hydroxylase and C17,20-lyase, significantly decreasing testosterone production in castration-refractory prostate cancer. Abiraterone is associated with marked progression-free and overall survival benefit (de Bono et al. 2011). Enzalutamide blocks testosterone from binding to cytosolic androgen receptor, which impedes receptor migration to the nucleus and thereby inhibits androgen-dependent gene expression. The SPARTAN trial showed that the third-generation anti-androgen, apalutamide, which has a high affinity for the androgen receptor, showed superior disease-free survival in cases of non-metastatic castrate-resistant prostate cancer (Smith et al. 2018). As a result, apalutamide has recently been licensed for this indication.

4.2 Targeting Receptors

4.2.1 Vascular Endothelial Growth Factor (VEGF) and Its Receptors

Vascular endothelial growth factor (VEGF) is a key component of a pathway regulating tumour angiogenesis. Tumour-derived VEGF is the ligand for a group of three receptors, VEGFR-1, 2 and 3 (also known as Flt-1, KDR and Flt-4, respectively) expressed on endothelial cells. Both the ligand and its receptor family are the target of rationally designed drugs (Ferrara 2005). Agents targeting VEGF include the monoclonal antibody bevacizumab and the soluble VEGF-binding protein aflibercept. Bevacizumab has been approved for the treatment of a range of solid tumours including ovarian and lung cancer. Aflibercept is an engineered

soluble receptor from extracellular domains of VEGFR-1 and VEGFR-2. It binds to all isoforms of VEGF and has a higher affinity than bevacizumab for VEGF A and B.

The VEGF receptor is a target for validated small molecular inhibitors (Rhee and Hoff 2005). These so-called multi-targeted inhibitors include vandetanib, which inhibits EGFR and VEGFR-2, as well as sorafenib, sunitinib and cediranib which have broad specificity for receptor tyrosine kinases including members of the VEGFR family. Vandetanib is licensed for use in medullary thyroid cancer. Sunitinib and sorafenib are both approved for use in advanced renal cell cancer, and sorafenib is also active in hepatocellular and thyroid carcinoma. Unsurprisingly, a prominent toxicity of these agents is hypertension, because of the involvement of VEGF in blood pressure homeostasis, although this toxicity is readily managed with careful blood pressure monitoring and early introduction of antihypertensives.

4.2.2 Epidermal Growth Factor Receptor

Epidermal growth factor receptor (EGFR) is a receptor tyrosine kinase (RTK) and a member of the ErbB receptor superfamily. Whilst it was first discovered in 1962, its role in tumourigenesis was only understood in the 1980s. *EGFR* overexpression is associated with poorer outcomes in various human malignancies; pathways involved in EGFR signal transduction therefore represent promising therapeutic targets. Binding of extracellular growth factor ligands to the ErbB receptor family causes dimerisation of the receptors, forming homo- or heterodimers that stimulate tyrosine kinase activity, initiating intracellular signalling cascades. The considerable clinical impact of therapies targeting EGFR and HER2 explains the central role these two receptors play in driving human cancer. The past three decades have seen the development of both monoclonal antibodies and small molecule tyrosine kinase inhibitors specific for ErbB family members.

Erlotinib and gefitinib are small molecule reversible inhibitors selective for the intracellular tyrosine kinase domain of EGFR (Baselga 2002). These orally bioavailable drugs prevent ATP binding and autophosphorylation of the EGFR tyrosine kinase. Trials in unselected patient populations resulted in response rates of 10–19% (reviewed in (Spicer and Rudman 2010)). Modest improvement in overall survival was observed in comparison with placebo in randomised trials with these agents. Further analyses from these studies reported variations in efficacy according to clinical characteristics and activating mutations in the *EGFR* gene were eventually identified as a potent predictive biomarker.

Second-generation tyrosine kinase inhibitors (TKIs) such as afatinib and dacomatinib irreversibly bind to EGFR and HER2. Dacomatinib showed a statistically superior overall survival compared with gefitinib (34 versus 27 months), in patients with an *EGFR* mutation in exon 19 or 21 (Wu et al. 2019). Osimertinib is a third-generation drug, less potently active against wild-type EGFR, and, importantly, able to bind avidly to the target even when bearing the T790M point mutation characteristic of resistance to earlier-generation inhibitors (Soria et al. 2018).

Cetuximab is a chimeric IgG monoclonal antibody against the extracellular domain of EGFR approved for use in colorectal, head and neck and lung cancers. The use of receptor-targeted antibody therapies offers two potential mechanisms of

action, adding the potential for activation of immune mechanisms to the signalling inhibition also seen with small molecules. All antibodies approved for cancer therapy belong to the IgG immunoglobulin subclass and as such are able to recruit cells such as NK expressing Fcγ receptor, which in turn can have cytotoxic or phagocytic effects on the tumour cell. Thus, therapeutic monoclonal antibodies may have antitumour effects mediated by both signalling inhibition and by antibody-directed cellular cytotoxicity. Several current and future strategies in the development of antibody therapies are directed at improving or broadening the affinity of these molecules for their receptors on immune effector cells.

Genomic analysis is routinely carried out on diagnostic samples of adenocarcinomas in lung and colorectal cancer. Patients with advanced-stage disease are selected to receive primary treatment with EGFR-targeted agents. TKIs are routinely used first line in the case of *EGFR*-mutated lung cancer, and monoclonal antibodies are selected in the case of colorectal cancer lacking activating mutation of *KRAS*, the product of which signals downstream of EGFR.

4.2.3 HER2

Binding of ligand to the extracellular domain of RTKs induces receptor dimerisation, both between the same and different (heterodimerisation) receptor subtypes. Heterodimerisation is assumed to be of particular significance for HER2 (Klapper et al. 1999), for which no endogenous ligand has been identified. *HER2* amplification can lead to constitutive proliferative signalling in the absence of ligand and has been detected in a wide range of tumour types including those originating from breast and stomach. The efficacy of the anti-HER2 monoclonal antibody trastuzumab appears to depend on HER2 overexpression in the targeted tumour, and this drug is approved in both these diseases where HER2 is upregulated. In patients with HER2-positive breast tumours, trastuzumab is associated with marked survival superiority in both the metastatic (Slamon et al. 2001) and adjuvant settings.

Even in the context of resistance to prior therapy with trastuzumab, the tumour can be effectively targeted, and systemic toxicities limited, using an antibody-drug conjugate (ADC). Trastuzumab emtansine (T-DM1) combines humanised antibody trastuzumab and the potent microtubule polymerisation inhibitor DM1, through a stable thioether linker. The latter component was found to have activity against breast cancer in the 1970s. However, as a single agent its toxicity far outweighed its benefits. Delivered via an ADC combination, the cytotoxic agent is internalised and delivered directly into the target cancer cells resulting in apoptosis. A randomised study demonstrated an impressive delay in progression-free survival in HER2-positive breast cancer (Hurvitz et al. 2013).

4.2.4 CD20

An early advance in the antibody therapy of human cancers was the development of rituximab, an IgG antibody specific for CD20. This target is ubiquitously expressed in lymphocytes of B-cell lineage. The addition of rituximab significantly improves the efficacy of chemotherapy of non-Hodgkin's lymphomas (Coiffier 2005) and has also found a role in the therapy of chronic lymphocytic leukaemia.

4.3 Other Targets

A growing knowledge of the diversity and complexity of signalling networks in malignant cells is reflected in the number of targeted therapies tested in clinical trials. In addition to the ligands and receptors outlined above, many others have been studied including inhibitors and antagonists of RTKs and other receptors such as MET, RET and FGFR (Jiang and Ji 2019). Targets of small molecules inhibiting intracellular kinases include mTOR, Akt, PI3K, BRAF, ALK and MEK (Fig. 3).

Translocations in the cancer genome can result in unique fusion kinases that can be the driver for some cancers. Some of these are the target for approved inhibitors. Examples include inhibition by imatinib and other drugs of the fusion kinase encoded by *BCR-ABL* on the Philadelphia chromosome resulting from a balanced translocation in most cases of chronic myeloid leukaemia (CML; see below) and targeting with crizotinib, alectinib and other molecules of the *EML4-ALK* fusion gene product present in about 5% of non-small-cell lung cancers.

Proteasomes which degrade tumour suppressor gene products are appealing targets for cancer therapy. Bortezomib has established activity in multiple myeloma (Richardson et al. 2005). The structure of chromatin, and hence the expression of genes controlling the malignant phenotype, can be modulated by histone deacetylase or demethylase inhibition (Piekarz and Bates 2009; Turner et al. 2004). Therapy-targeting mechanisms maintaining the integrity of the cancer

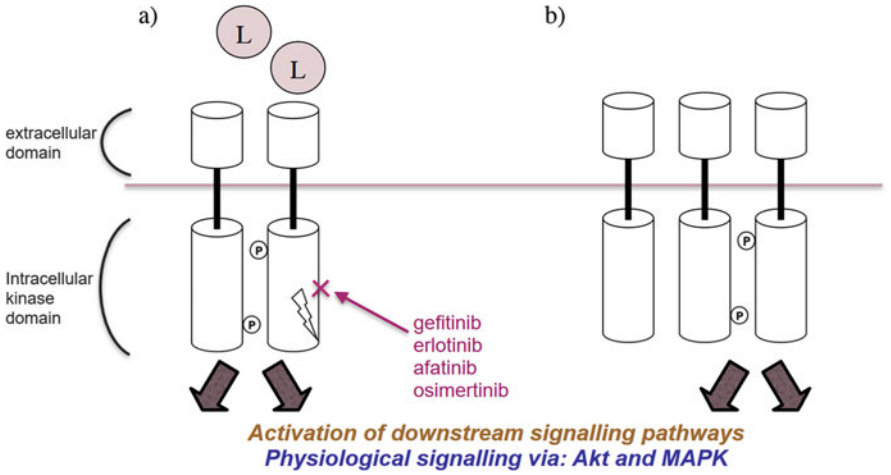


Fig. 3 Mechanisms of receptor tyrosine kinase activation in cancer. (**a**) Binding of upregulated ligand (L) to the extracellular domain, or presence of an activating kinase domain mutation (jagged arrow), leads to receptor dimerisation and autophosphorylation (P) of the intracellular kinase domain. Activation of downstream signalling events (open brown arrows) results in proliferation. (**b**) Overexpression of the receptor itself, for example, as a result of gene amplification in the tumour genome, results in inappropriate extracellular domain proximity and again activates downstream signalling. Erlotinib and gefitinib are small molecule reversible inhibitors of the intracellular tyrosine kinase domain of the EGFR receptor

genome itself, especially poly-ADP-ribose polymerase (PARP), have proven clinically effective especially in tumours occurring on a background of germline heterozygosity for DNA repair genes such as *BRCA1* and *BRCA2*, discussed further below (Fong et al. 2009). Such tumours become more dependent on parallel DNA repair pathways than are surrounding normal cells, allowing targeting of DNA repair that is highly selective for tumour cells through a process that has become known as synthetic lethality.

Other promising approaches include antisense technology, oncolytic viral therapy, vaccines, immune checkpoint inhibition and adoptive T-cell therapies.

4.4 Targeting the T-Cell

4.4.1 Checkpoint Inhibition

The introduction of immunotherapy using checkpoint inhibition has provided unprecedented improvement in outcomes for some cancers with largely manageable toxicities. Immune checkpoint inhibitors targeting molecules such as CTLA-4 and PD1 enhance the activity of endogenous T-cells against tumour antigens. However, a proportion of patients exhibit incomplete efficacy (intrinsic resistance), and others will experience loss of tumour control in due course (acquired resistance) (Havel et al. 2019).

Blockade of PD1 or its ligand PDL1, through use of antibodies including pembrolizumab, has been validated as a therapeutic strategy as outlined above. These drugs have demonstrated a survival benefit across several malignancies including lung, melanoma, lymphoma, renal cell carcinoma, head and neck squamous cell carcinoma and bladder cancer. PDL1 is often expressed on tumours or within the microenvironment. PD1 or PDL1 directed targeted therapeutics can stimulate exhausted T-cells by blocking this inhibitory T-cell signalling interaction.

Observations support the prediction that checkpoint blockade is only effective in tumours that are infiltrated with T-cells ('hot' tumours) and in those with a high burden of somatic mutations. Identifying biomarkers to select patients that are likely to respond to checkpoint inhibition has generally been unsuccessful in many tumour groups such as melanoma and renal cancer. In lung cancer, higher expression of PD1 expression is used to select patients that are more likely to benefit from immunotherapy, but even here the association between PD1 expression and activity is incomplete (Havel et al. 2019).

4.4.2 Adoptive T-Cell Therapy

T-cells are an essential part of adaptive immunity and central to pathogen clearance, and their physiological function also includes tumour surveillance and rejection. During the early stages of development in the thymus, T-cells mature the specificity of their T-cell receptor (TCR), which recognises processed antigens in an MHC context. Some tumours have significant populations of host T-cells present in their microenvironment, and in some circumstances, it is possible to extract this population of tumour-infiltrating lymphocytes (TILs) from a surgically resected tumour,

culture them ex vivo and reinfuse into the patient to generate a durable clinical response (Yang and Rosenberg 2016).

The artificial genetic transfer of TCR genes, or chimeric antigen receptors (CAR), to naive T-cells, which originally do not have any antitumour specificity, is a compelling concept which has shown promising results in several tumour types. This autologous approach entails T-cell extraction from patient blood and manipulating the expression of tumour-targeting receptors through genetic engineering. A period of cytotoxic conditioning which results in depletion of endogenous T-cells may be important prior to introducing primed T-cells, in part to address the population of inhibitory lymphocytes already present in the tumour microenvironment. Frequently observed toxicities following administration of adoptive T-cell therapies include cytokine release syndrome, central neurotoxicity and infections which can be fatal. However, major responses have been achieved with CAR-T-cell therapy in the treatment of acute leukaemia, although the treatment of solid tumours is proving more difficult, presumably because of the presence of a hostile microenvironment in solid, but not liquid tumours (Sadelain et al. 2017).

5 Nuclear Medicine Therapies

A further option for targeted treatment of metastatic disease is the use of radioactive pharmaceuticals to combine diagnostic imaging and therapy. A single agent is formed by combining a diagnostic and therapeutic radioisotope with a binding molecule to allow for diagnosis, drug delivery and treatment response monitoring. Broadly this is referred to as the theranostic approach. The radiopharmaceutical component is identical or a similar molecule that is radiolabelled differently or administered at different dosages. For example, iodine-123 is a gamma emitter, and iodine-131 is a gamma and beta emitter, both of which can be used for theranostic purposes (Yordanova et al. 2017) Desirable properties of therapeutic radionuclide include emission characteristics proportional to the tumour volume, minimising local toxicity. This is a particularly attractive option for patients with multiple comorbidities that are unfit for cytotoxic treatments.

5.1 Radio-Iodine in Thyroid Cancer

Iodine is used in the formation of the thyroid hormones thyroxine (T4) and triiodothyronine (T3), in the thyroid gland. Physiologically, these are vital in human development and metabolism. In 1946, the first radiopharmaceutical was developed from the neutron bombardment of tellurium-131 forming the radionuclide ^{131}I. ^{131}I combines beta and gamma emitters to irradiate cancerous cells, thyroid remnant or distant metastatic disease. It is licensed for particular cases of papillary and follicular carcinoma. Iodine is taken up by the follicular cells of the thyroid gland, whilst some is directly excreted renally. Beta emission, which penetrates up to 1 mm, is therapeutic, whilst simultaneous gamma emission can image the target lesion

using SPECT or a gamma camera, allowing for real-time visualisation. Physiologically the salivary glands take up some iodine, and therefore a common toxicity is sialadenitis. Strict precautions for patients receiving radioactive treatment, including isolation, should be adhered to. Successful ablation of the thyroid gland will often require subsequent long-term thyroxine replacement.

5.2 Somatostatin Analogues

Somatostatin receptors (SSTRs) have important physiological roles including inhibition of hormones secreted by the pituitary gland, inhibition of pancreatic exocrine hormones (insulin and glucagon), inhibition of motility and exocrine secretions in the GI tract and central nervous system regulation. They are overexpressed in neuroendocrine tumours where somatostatin plays a critical role in secretion and growth (Reubi and Schonbrunn 2013). Neuroendocrine tumours originate from endocrine organs or neuroendocrine cells within an organ such as the gastrointestinal tract. They often have a high density of somatostatin receptors, making diagnostic and targeted therapy an attractive treatment option.

Three somatostatin analog tracers known as DOTA-TATE, DOTA-TOC and DOTA-NOC are labelled with gamma-emitting gallium-68 to specifically target SSTRs for diagnostic purposes. By targeting tumours with alpha- or beta-emitting isotopes such as ^{90}Y or ^{177}Lu, selective radiotherapy can be delivered using these peptides to the primary tumour and metastatic sites with high specificity and sensitivity. This is known as peptide receptor radionuclide therapy (PRRT) (Yordanova et al. 2017).

Octreotide and lanreotide are synthetic somatostatin analogues, specific for SSTR2, one of the five known receptor subtypes, and can be used for the treatment of symptoms caused by hormone overproduction including non-malignant conditions such as acromegaly. Targeting the SSRTs inhibits downstream signalling, halting cell growth and stimulating apoptosis.

5.3 Radium-223

Radium-223 is an alpha-emitting isotope which selectively binds to areas of increased bone turnover, through its property of mimicking calcium. It is used in cases of metastatic castrate-resistant prostate cancer with bone-only metastasis, to prevent morbidity associated with skeletal metastasis. Alpha particles travel a short range, and radium-223 has low emission of gamma photons, optimising safety of administration and minimizing concerns about close patient contact afterwards. Studies have shown a delay in time to first symptomatic skeletal event and improved overall survival (Parker et al. 2018).

5.4 PSMA Ligand: Lutetium

Prostate-specific membrane antigen (PSMA) is a transmembrane protein which plays a critical role in cell migration, survival and proliferation through a receptor-internalisation process. It is overexpressed on the surface of prostate cancer cells, particularly in patients with high-grade and castrate-refractory disease, allowing for the development of specific diagnostic and therapeutic ligands. The latter often uses a small beta-emitting molecule, lutetium 177 (^{177}Lu), which binds to PSMA with high affinity.

Lutetium has a long half-life, and preliminary studies have shown promising outcomes (Yordanova et al. 2017). Whilst radionuclide therapy manipulates tumour-specific receptors, the radioactive component emits radiation which can disseminate causing local toxicity. As such, use of this therapy is dependent on the sites of metastases. PSMA has a physiological role in normal intestinal, renal and salivary gland, where it is expressed albeit to a lesser degree compared to cancer cells, driving the toxicity profile of this targeted treatment, with dry mouth being a common side effect. Radionuclides are often excreted by the kidneys, and therefore these remain as the most pertinent organs at risk. Overall, radionuclides are better tolerated than cytotoxics.

6 Pharmacogenetics, Pharmacogenomics and Patient Selection for Treatment

An understanding of somatic mutations in the cancer genome has led to the development of targeted therapies. These mutations can serve as biomarkers predicting clinical benefit. Retrospectively, they may seem predictable given a drug's mechanism of action, an example being the use of the HER2-specific trastuzumab only in those patients with *HER2* amplification on their tumour. Other predictive somatic genetic events include the *BCR-ABL* chromosomal translocation in CML sensitive to imatinib, *EGFR* mutations responding to EGFR inhibitors (erlotinib, gefitinib, afatinib, osimertinib; see elsewhere in this chapter) and ALK mutations in non-small-cell cancer responding to crizotinib, alectinib and other members of a growing class of tyrosine kinase inhibitors. Other biomarkers predictive of toxicity, rather than benefit, are polymorphisms in the patient's somatic genome (Wang et al. 2011). Predicting clinical benefit from immune checkpoint blockade appears to be more complex than simple reference to tumour PDL1 expression, and other factors such as tumour mutational burden are being investigated.

6.1 Lung Cancer and *EGFR* Mutations

Approximately 90% of non-small-cell lung carcinoma cases are associated with tobacco exposure. Cancers in the remaining 10% tend to occur with relatively greater frequency in younger, female, non-smoking patients, most often with a particular histology. It is now understood that these clinical characteristics correlate with a

discrete underlying biology that drives the malignant phenotype. Specifically, this is an upregulation of EGFR signalling and in particular mutations in the tyrosine kinase domain of this receptor. Sensitivity to EGFR inhibition with TKIs such as geftinib and erlotinib is associated with activating *EGFR* mutations (Lynch et al. 2004; Paez et al. 2004; Pao et al. 2004). NR*EGFR* kinase domain mutations are found in four exons (Klapper et al. 1999; Slamon et al. 2001; Coiffier 2005; Richardson et al. 2005) which are in close proximity to the ATP-binding pocket. In-frame deletions in exon 19, and an exon 21 substitution (L858R), are the most common mutations, together representing 85–90% of all *EGFR* mutations found in NSCLC. The location of these mutations leads to an alteration in the catalytic site, resulting in enhanced affinity for the competitive TKI relative to ATP substrate. Retrospective analyses show superior outcomes including response rates of up to 75% in patients with activating mutations treated with EGFR-specific TKIs.

Trials comparing first-line TKI treatment (gefitinib, erlotinib and afatinib) versus the previous gold standard of platinum-based chemotherapy in patients with *EGFR*-mutated lung adenocarcinoma showed a superior progression-free survival with TKIs (Mok et al. 2009; Rosell et al. 2012). Tailoring treatment of lung cancer according to mutation status has become the standard of care. Furthermore, 50% of patients that progress during or following first-line treatment have evidence of EGFR T790M point mutation (discussed further below). In these cases, osimertinib, an oral, third-generation, irreversible EGFR-TKI that selectively inhibits both EGFR-TKI–sensitizing and EGFR T790M resistance mutations, with lower activity against wild-type EGFR, is licensed globally. More recently, studies comparing first- and second-generation TKI versus third-generation TKIs in treatment-naive patients with *EGFR* mutations showed superior efficacy with the latter group (Soria et al. 2018).

6.2 *BRCA1*, *BRCA2* Mutation and PARP Inhibition

Poly(ADP-ribose)polymerase (PARP) is an enzyme activated by damage to the genome and involved in DNA repair. PARP1 acts at DNA single-strand breaks via the mechanism of base excision repair (BER). PARP synthesises ADP-ribose polymers, which protect the strand break and provide a scaffold for assembly of the DNA repair complex.

BRCA1 and *BRCA2* are tumour suppressor genes encoding proteins critical for DNA repair and genomic stability. BRCA-deficient cells are dependent on BER because alternative repair mechanisms are inactivated. PARP inhibition can induce synthetic lethality in cells where *BRCA1* or *BRCA2* are mutated, by inhibiting the alternative BER repair pathway. In patients with BRCA protein loss due to hereditary mutation of one *BRCA* allele, and somatic loss of the other allele in their tumour, inhibition of PARP function creates irreparable damage to tumour DNA. By contrast, normal tissues heterozygous for BRCA function are unaffected. The predicted combination of efficacy and tolerability has been confirmed in patients selected for genomic *BRCA* mutation.

PARP inhibition may also play a critical role in tumours presenting features of "BRCAness" (Turner et al. 2004), in which other genetic changes occur in sporadic tumours to create a phenotype similar to that of *BRCA* mutation carriers. These tumours may also be vulnerable to PARP inhibitors in combination with DNA-damaging agents. Biomarkers useful for patient selection in this setting are yet to be definitively identified.

6.3 Prediction of Toxicity

Anticancer drug therapy can be associated with significant toxicity. Variation in clinical outcome between individuals may partly be attributed to genomic polymorphism. For example, patients carrying one of four variants in the *DPYD* gene encoding dihydropyrimidine dehydrogenase, present in 5% of the UK population, experience significant toxicity to 5-fluoruracil (5-FU) (Diasio 2001). Similarly, the number of dinucleotide repeats in the promoter of *UGT1A1* is associated with increased toxicity of irinotecan because of reduced metabolism. Testing for these predictive mutations prior to therapy is often performed, to guide dose reduction or offer alternative therapy. Such personalised strategies are already being successfully used in other health disciplines, for example, to optimise the efficacy of azathioprine treatment of patients with inflammatory bowel disease. Therapeutic drug monitoring, as used in the routine prescribing of oral anticoagulation and post-transplant immunosuppression, has been relatively underused in oncology but may now be gaining some traction.

7 Resistance Mechanisms

The variation in efficacy seen between patients with the same histological diagnosis can partly be explained by heterogeneity in resistance mechanisms. Broadly, these mechanisms can be classified as genetic or pharmacokinetic. Whilst drug resistance maybe de novo, it may also be acquired as a result of the selection pressure of the therapy itself.

It is widely appreciated that alterations in tumour vascularity can be responsible for tumour resistance. These can be altered by altering the structure of the drug to enhance delivery, such as with the case of liposomal doxorubicin or albumin-bound paclitaxel. Secondly, a number of pharmacokinetic resistance mechanisms are driven by membrane transporter proteins, especially members of the multidrug resistance family such as MDR1, also known as P-glycoprotein (Pgp). These proteins can drive ATP-dependent efflux of drugs from cancer cells. This has been seen following treatment with platinum- and anthracycline-based chemotherapy and can be overcome by co-administering with either small molecules, non-competitive inhibitors or competitive inhibitors. The third principle involves drug inactivation through gamma-glutamyl-cysteine synthetase or gluteihione-based enzymes. Some other

resistance mechanisms arise from somatic genetic events in the cancer genome that alter the structure of the drug target or DNA damage and repair.

7.1 *CML BCR-ABL* Mutations and Resistance to Imatinib

CML pathogenesis is driven by a reciprocal translocation between chromosome 9 and 22 to give risk to a fusion protein kinase, BCR-ABL1. The first therapeutically successful small molecule tyrosine kinase inhibitor, imatinib, developed was against the BCR-ABL1 mutation. Imatinib has a high therapeutic ratio because the target is expressed in malignant cells only. Complete haematological response was seen in 53 of 54 patients treated at doses of >300 mg. The rational design and spectacular efficacy of this TKI ensured that imatinib became widely recognised as the paradigm for a new generation of targeted therapies.

Cytogenetic and/or molecular monitoring at 3, 6 and 12 months following initial TKI therapy can highlight patients in whom primary therapy with imatinib is likely to be ineffective. This is used to categorise the disease as being either BCR-ABL1 dependent versus independent. BCR-ABL1 dependence, whereby resistance occurs later in the course of the disease, suggests the mechanism of resistance is likely related to mutations in the kinase domain affecting the structure of BCR-ABL, which ultimately results in subtherapeutic delivery of drug to the target through impairment of binding or interference of biological and cellular processes. Over half of cases of BCR-ABL1 resistance are attributed to point mutations at the ATP-binding kinase domain (KD). In vitro studies show the T315I 'gatekeeper' point mutation causes steric hindrance preventing inhibitor binding and an active conformation of the fusion kinase, thereby promoting drug efflux. Second-generation inhibitors nilotinib and dasatinib retain activity against the majority of kinase domain mutations, aside from the T351I 'gatekeeper' mutation through tighter binding to a similar, inactive ABL1 conformation (Gorre et al. 2001). This allows for sequential treatment according to emergent resistance mutations. A further generation of agents with activity against T315I is in development.

Resistance due to dasantinib is a result of distinct mutations including VS299 and F317. All TKIs used in CML undergo hepatic first-pass metabolism via CYP3A4, strong inducers of which will lead to TKI resistance. In a fewer number of cases, intrinsic or primary resistance is observed and commonly associated with BCR-ABL1-independent mechanisms mediated through alternative survival pathways, of which numerous have been identified (Patel et al. 2017). For example, the activation of parallel integrin and/or growth factor receptor signalling pathways adds another dimension responsible for imatinib resistance. The latter pathway, named a dependence pathway, given its receptor dependence and ligand independence, is crucial for survival and anti-apoptotic signalling mediated via activation of the PI3K/Akt pathway in multiple cancers. Regardless of the numerous upstream pathways, they may converge onto the same downstream signals thereby allowing for therapeutic targets. Amongst the downstream pathways are STAT3, PIK3/AKT and RAF/MEK/ERK.

7.2 Molecular Markers of Resistance to EGFR Inhibition

Patients with activating EGFR mutations generally show an initial response to inhibition; however, inevitably acquired resistance develops to first- and second-generation EGFR-TKIs. Numerous mechanisms, including second *EGFR* mutations, are associated with the development of resistance to TKI therapy. Approximately 40–50% of acquired resistance to first-generation EGFR inhibitors can be accounted for by the T790M mutation, the commonest resistance event, in exon 20 of the *EGFR* kinase domain (Chen et al. 2008). This mutation results in the insertion of a bulky methionine residue at the active site and is analogous to the T315I gatekeeper mutations seen in CML. A molecular analysis of circulating tumour cells from TKI-naïve patients with metastatic NSCLC found the T790M mutation in 38% of patients. The presence of T790M even before patient exposure to TKI was associated with a significantly shorter progression-free survival compared with patients who did not have detectable levels of T790M (Maheswaran et al. 2008). Therefore, this resistance mechanism is naturally evolving and, in some cases, very likely to be propagated by TKI selection. More recently, the third-generation TKI osimertinib, with irreversible and covalent binding at the active site (cysteine-797 residue in the ATP-binding site), has been approved.

Other less common *EGFR* mutations can also lead to resistance. Additionally, alterations in parallel signalling pathways, such as *MET* amplification, may also overcome the effects of all three generations of TKI therapy (Sequist et al. 2011). The presence of mutations in other signalling components may be associated with intrinsic resistance and the lack of sensitivity to TKI therapy. Specifically, an activating *KRAS* mutation is present in 15–25% of lung adenocarcinomas and correlates with de novo lack of sensitivity to EGFR TKIs. Similarly, patient selection for treatment with monoclonal antibodies specific for EGFR, such as cetuximab, is informed by the presence of activating mutation in the *KRAS* oncogene, encoding a downstream signalling partner of EGFR, which perhaps unsurprisingly predicts lack of benefit in colorectal cancer (Karapetis et al. 2008).

In up to 10% of cases where response to EGFR targeted therapy has failed, there is histological transformation of the tumour to a small-cell morphology known as epithelial-mesenchymal transition (Druker et al. 2001) with reduction in EGFR expression and subsequent alteration in the biochemical behaviour of the disease. Repeat biopsies to clarify the histology are necessary to guide treatment (Westover et al. 2018)

7.2.1 Multiple Targets

Cancer development and progression is driven by an array of complex biological processes. The molecular signalling pathways in a tumour are adaptable and redundant (Yarden and Sliwkowski 2001), exemplified by the ErbB receptor family members. This allows HER2, which has no identified ligand, and HER3, which has no kinase activity, to become actively involved in signalling. This combination of complexity and redundancy in the cancer cell implies that therapy focusing on a single target may be unlikely to achieve adequate, long-term disease control for many patients. Less than half of acquired EGFR TKI resistance is attributed to

non-EGFR-centric adaptions, otherwise known as 'bypass' resistance mechanisms as they activate the same downstream signalling pathways resulting in tumour survival and growth. Most commonly, these pathways are related to the ErbB receptors through which IGF1 is activated and MET amplified (Fig. 4) (Westover et al. 2018; Ricordel et al. 2018). Perhaps, targeting multiple receptors with a single agent can potentially overcome resistance driven by molecular heterogeneity and hence improve efficacy. Lapatinib, a reversible inhibitor of both EGFR and HER2, is active in HER2-positive metastatic breast cancer, a disease in which inhibitors of EGFR alone are not active. It is approved for use in combination with capecitabine chemotherapy (Geyer et al. 2006). This approach is analogous to the class of multi-targeted 'dirty' drugs that owe efficacy in inhibiting angiogenesis to their ability to inhibit multiple pathways activated by VEGF and related ligands, including VEGFR-1 and -2, platelet-derived growth factor (PDGF)-α and -β, c-Kit and fms-like tyrosine kinase 3.

7.2.2 Irreversible Binding

The acquisition of resistance mutations in the *EGFR* gene, such as T790M described above, interferes with reversible erlotinib and gefitinib binding at the active site and suppresses the inhibition of EGFR signalling in non-small-cell lung cancer. An attractive feature of a number of second-generation inhibitors of ErbB receptors is irreversible binding to the target receptor. At least in preclinical studies, these irreversible inhibitors effectively inhibit EGFR signalling even in gefitinib-resistant

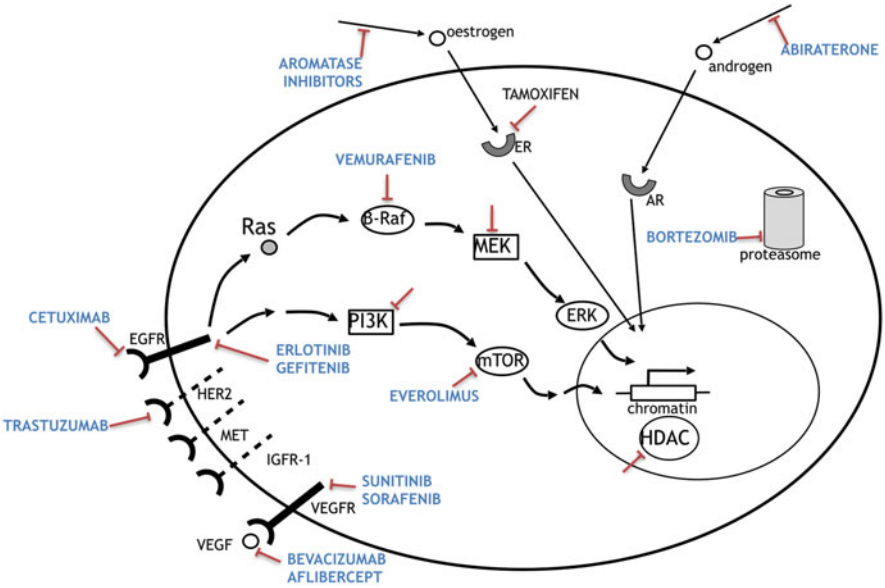


Fig. 4 Some targets for novel anticancer therapies. Inhibitory effects are indicated in red. Solid black arrows indicate activating effects

cell lines harbouring the T790M mutation. Prolonged suppression of EGFR kinase activity results from covalent elimination of kinase activity until the synthesis of new receptors. Third-generation EGFR inhibitors irreversibly bind to the ATP-binding site in the kinase domain of the receptor, with superior activity in the presence of T790M mutations compared to the earlier-generation drugs.

All patients eventually develop resistance to treatment with EGFR inhibitors. Preclinical studies and clinical evidence alike provide evidence that resistance mechanisms to first-, second- or third-generation treatment are similar. They may be due to a single or a combination category of resistance mechanism including: tertiary EGFR mutation, bypass signalling, downstream activation, or histological transformation (Ricordel et al. 2018; Spicer and Rudman 2010).

7.3 ALK Resistance

Small molecule anaplastic lymphoma kinase (ALK) inhibitors, such as crizotinib, alectinib and brigatinib, have good clinical effect in ALK-rearranged non-small-cell lung cancers. However, as with EGFR inhibition, secondary resistance inevitably develops. Similar mechanisms of resistance as those seen in EGFR-TKI resistance are seen in in vitro studies. In particular, inactivating kinase domain mutation is deemed responsible for some resistance to crizotonib, most commonly L1196M and G1269A mutations. Following treatment with first-generation ALK inhibitors, these tumours respond well to potent second-generation ALK inhibitors, alectinib and ceritinib. Alternatively, activation of separate oncogenes can override ALK to become the dominant driver mutation (Doebele et al. 2012). Newer ALK-targeting drugs show clinical benefit in treatment-naive ALK-positive cancers compared to crizotinib, suggesting that resistance develops due to inadequate suppression of ALK.

An isolated resistance mechanism is seldom the cause of loss of disease control but is rather a combination of mechanisms evolving to allow tumour survival. A complex combination of drug efflux, modulation of drug metabolism, secondary mutations of the target protein, induction of alternate signalling, induction of epigenetic mechanisms and selection of a drug-refractory cancer stem cell population may be involved.

8 Cancer Drug Discovery and Preclinical Development

Traditional cancer drug discovery has relied heavily on screening large compound libraries for activity against cancer cell lines in culture, initially murine leukaemias, and later human cancer cell lines. Many of these compounds were originally natural products, such as the vinca alkaloids (derived from the periwinkle) and taxanes (including paclitaxel, now synthetically manufactured, but originally available only by extraction from the pacific yew). More recently, small molecular weight drugs have been rationally designed with reference to target crystal structures and

optimised via in silico modelling and combinatorial chemistry. Whilst structural modelling techniques may be rational, they offer no guarantee that a resulting lead compound will have desirable pharmaceutical properties, which may still require optimisation using traditional preclinical pharmacology and medicinal chemistry approaches.

Since the mid-twentieth century, cancer cell lines have been derived from common human cancers and are used as an early step in the validation of novel targets and therapy combinations. A range of in vivo preclinical models in rodents include xenografts (immunocompromised animals bearing tumours derived from other species including human), and transgenic animals are also widely used in preclinical drug development of novel treatments for cancer. Transgenics null for tumour suppressor genes can in some cases spontaneously develop tumours; the use of conditional knockouts can allow tissue-specific gene targeting and tumourigenesis.

Preclinical toxicity testing for modern rationally designed targeted drugs has been the subject of much debate, in large part because of the species-specificity of many modern biotherapeutics. Non-human toxicity data might be informative for a new cytotoxic which is expected to target mitosis in any mammalian cell, or even a novel molecular therapy where a drug candidate binds to both human and animal target, but the exploration in animals of on-target toxicity for precisely targeted agents, especially monoclonal antibodies, may be falsely reassuring (Chapman et al. 2007).

9 Current Issues in the Development of Drugs in Oncology

9.1 Improving the Odds of Success of Phase 3 Trials

There is evidence that the clinical development of new agents for the treatment of cancer is less efficient than for many other diseases (Adjei et al. 2009). Over the years this has at least in part been due to the failure of many Phase 3 trials. This might seem paradoxical in the era of drugs that have been rationally designed to hit targets known to drive human cancers, but it is only recently that patient selection has been implemented to optimise efficacy in a subset of patients. In the last 5 years, open-label Phase 3 studies have shown impressive outcomes. Brigatinib compared to crizotinib in patients with ALK-positive advanced lung cancer improved progression-free survival with an impressive hazard ratio of 0.49 (Camidge et al. 2018). The preferable strategy is to adopt a patient selection approach from late Phase 1, assuming an appropriate biomarker is available. This optimises the chance of demonstrating efficacy (or lack of it) early on in clinical development and of beginning the validation of a companion diagnostic alongside the new therapy.

The explosion of new knowledge in cancer biology, and the associated plethora of new potential therapeutic targets, places an additional pressure on the clinical drug development community, namely, how to prioritise and pick winners early. Careful design and conduct of studies in appropriate patient populations allows the possibility of establishing proof of mechanism, and evidence of efficacy, prior

to initiation of costly Phase 3 trials. A successful early-phase trial is not just one which leads on to Phase 2 and 3 studies but also one allowing the early discontinuation of a development programme where an agent fails to meet early and rational go/no-go criteria.

9.2 Regulation of Cancer Drug Development

The standards of preclinical safety assessment required for the new generation of therapies are the subject of ongoing debate. Conventional toxicity studies in animals may provide limited information relevant to human use in the context of potent species-specific novel agents (Chapman et al. 2007). Indeed, preclinical data may be falsely reassuring as occurred in a Phase 1 study of an immunostimulatory anti-CD28 antibody agonist. Despite acceptable toxicity findings in non-human primates, the first six patients treated in a Phase 1 clinical trial all experienced a life-threatening cytokine storm resulting from uncontrolled T-cell activation (Suntharalingam et al. 2006). The best available preclinical exploration of safety should of course continue to be required, but for the development of highly specific agents such as antibodies, the use of conventional animal studies, especially those in non-human primates, should not be mandated where information relevant to human use is unlikely to be forthcoming.

Improvements in study design are required, as discussed above, to reduce late-stage failures of new drugs for cancer. There is also a pressing need to address the regulatory burden placed on those conducting clinical studies in cancer, as in other disciplines (Rawlins 2011). Such changes are likely to reduce delays and improve cost-effectiveness in the development of new treatments for diseases where a high level of unmet need remains, without materially compromising the safety of trial participants.

One aspect of new drug regulation that has evolved to optimise access to novel agents is provision of earlier patient access to new agents through accelerated approval (Kwak et al. 2010). Here, interim approval is granted on the basis of results of early clinical trials, on the understanding that post-approval studies will be completed. This approach allows for drug approval based on the use of surrogate endpoints 'reasonably likely to predict clinical benefit', response rate being a typical surrogate in oncology. Accelerated approval is especially appropriate where dramatic clinical benefit is demonstrated in early-phase trials conducted in patients selected for expression of the drug target, as has been the case for the ALK inhibitor crizotinib. Further work is required to encourage the development of new drugs to address unmet need in rare cancers, where small numbers of patients mean randomised trials are difficult or impossible to conduct and therefore any potential commercial market is limited. One approach has been the introduction of the orphan drug initiative which has relevance for the less common cancer diagnoses (Braun et al. 2010).

10 Conclusion and Future Perspectives

Continued progress in identifying the molecular targets that drive the biology of malignant cells, especially protein kinases which are readily druggable using small molecular inhibitors, represents a key trend in cancer drug development. These small molecules have contributed to the large class of target therapies also including an ever-expanding number of monoclonal antibody-based therapies. The utility of antibodies is driven not only by their intrinsic specificity for a single target but by the relatively recent focus on these agents as offering immune modulation in addition to signalling inhibition. Antibody conjugation, glyco-engineering and even class switching are ongoing and expected future trends in the development of antibody drugs.

Manipulation of the interplay between host immunity and tumourigenesis has affirmed its role as a therapeutic modality across several solid organ tumour groups. However, appropriate regulation of individual immunity to prevent self-reactive toxicity and autoimmunity is yet to be achieved. Understanding of the complex downstream signalling network, alteration in the tumour microenvironment and changes in the surface expression of HLA molecules will allow for more predictable use of highly specific immunomodulatory therapies. The future may lie in a selective combination of cancer vaccines, checkpoint inhibitors and immune cytokine modulation (Wraith 2017; de Aquino et al. 2015).

The efficacy of the newer agents, both targeted and immune therapies, appears in many cases to be optimal in combination with other drugs, like the longer-established cytotoxic agents. The favourable therapeutic ratio of these newer agents facilitates their combination either with each other or with chemotherapy drugs. Rational pairings of agents targeted at either multiple components of a single signalling pathway, or at parallel pathways, are, respectively, likely to increase the chance of delivering optimal efficacy and of countering the development of resistance. Indeed a growing number of studies are showing improved outcomes with a combination of cytotoxics and checkpoint inhibition (Lisberg and Garon 2019).

The parallel development of biomarkers to inform patient selection will continue to be vital for the optimum use of cancer drugs. Targeted therapies are clearly not expected to be effective in tumours lacking expression of the target, but in many cases the presence of protein may not be enough. Additional molecular features of target activation (such as gene mutation, amplification and translocation) may exert a strong influence on driving oncogene addiction. For an individual patient, identifying which of these is present in their tumour will make increasing demands on cancer diagnostic services. Multiplex analysis for the presence of predictive biomarkers have now become the norm, and it may be that next-generation sequencing of patients' cancer genome, or of a panel of selected genes, will some become routinely achievable as this technology becomes rapidly more cost-effective. Identification of the best biomarkers for selection of patients as candidates for checkpoint inhibitor immunotherapies remains work in progress. In some diseases and for some drugs, but not others, expression of the PDL1 protein in tumour tissue is associated

with clinical benefit, but this correlation is not perfect. For example, current evidence suggests that lung cancer and melanoma patients may benefit from treatment with some of these drugs irrespective of the expression level of this target. Other selection parameters, such as tumour mutational burden (TMB) and tumour-infiltrating lymphocytes (TILs), are being investigated.

For reasons associated with the practicalities of clinical drug development, new anticancer agents are studied first in patients with advanced disease, and some of the significant advances made in the treatment of many of these patient groups have been described above. Whilst many of the targeted therapies have succeeded in delivering periods of good quality of life, long-term survival is still not expected in most metastatic cases of the common malignant diseases. This is largely due to resistance mechanisms, some of which are themselves becoming better understood at a molecular level. Similarly, the use of systemic therapies in the adjuvant setting is already proven to increase the chances of cure in several common cancers if treated when still at an early stage, and the clinical benefits of newer therapies will be amplified in this setting.

The cost of cancer care continues to escalate, and it has been projected that by 2020 spending in the United States will have risen in real terms by 600% in 30 years (Sullivan et al. 2011). This problem is amplified in less-developed nations, where already limited resources are challenged by rapid increases in cancer incidence across aging populations acquiring the risk factors associated with economic development. Initiatives such as careful patient selection based on predictive biomarkers, and changes to clinical trial design to allow earlier go/no-go decision-making, will become increasingly important in controlling drug costs. Rational therapy design has led directly to the development of molecular targeted therapies and immune checkpoint inhibitors. Combined with careful biomarker-driven patient selection, these newer treatment approaches provide the opportunity to make step changes in clinical outcomes to contrast with the modest increments made with many previous advances (Sobrero and Bruzzi 2009).

References

Adjei AA (2006) What is the right dose? The elusive optimal biologic dose in phase I clinical trials. J Clin Oncol 24:4054–4055

Adjei AA, Christian M, Ivy P (2009) Novel designs and end points for phase II clinical trials. Clin Cancer Res 15:1866–1872

Baselga J (2002) Why the epidermal growth factor receptor? The rationale for cancer therapy. Oncologist 7(Suppl 4):2–8

Booth CM, Tannock I (2008) Reflections on medical oncology: 25 years of clinical trials – where have we come and where are we going? J Clin Oncol 26:6–8

Braun MM et al (2010) Emergence of orphan drugs in the United States: a quantitative assessment of the first 25 years. Nat Rev Drug Discov 9:519–522

Camidge DR et al (2018) Brigatinib versus crizotinib in ALK-positive non-small-cell lung Cancer. N Engl J Med 379:2027–2039

Chapman K et al (2007) Preclinical safety testing of monoclonal antibodies: the significance of species relevance. Nat Rev Drug Discov 6:120–126

Chen DS, Mellman I (2013) Oncology meets immunology: the cancer-immunity cycle. Immunity 39:1–10
Chen FL, Xia W, Spector NL (2008) Acquired resistance to small molecule ErbB2 tyrosine kinase inhibitors. Clin Cancer Res 14:6730–6734
Choueiri TK et al (2004) Role of aromatase inhibitors in the treatment of breast cancer. Clin Ther 26:1199–1214
Coiffier B (2005) State-of-the-art therapeutics: diffuse large B-cell lymphoma. J Clin Oncol 23:6387–6393
Croce CM (2008) Oncogenes and cancer. N Engl J Med 358:502–511
de Aquino MT et al (2015) Challenges and future perspectives of T cell immunotherapy in cancer. Immunol Lett 166:117–133
de Bono JS et al (2011) Abiraterone and increased survival in metastatic prostate cancer. N Engl J Med 364:1995–2005
Diasio RB (2001) Clinical implications of dihydropyrimidine dehydrogenase on 5-FU pharmacology. Oncology (Williston Park) 15:21–26; discussion 27
Doebele RC et al (2012) Mechanisms of resistance to crizotinib in patients with ALK gene rearranged non-small cell lung cancer. Clin Cancer Res 18:1472–1482
Druker BJ et al (2001) Efficacy and safety of a specific inhibitor of the BCR-ABL tyrosine kinase in chronic myeloid leukemia. N Engl J Med 344:1031–1037
Early Breast Cancer Trialists' Collaborative, G (2005) Effects of chemotherapy and hormonal therapy for early breast cancer on recurrence and 15-year survival: an overview of the randomised trials. Lancet 365:1687–1717
Eisenhauer EA et al (2009) New response evaluation criteria in solid tumours: revised RECIST guideline (version 1.1). Eur J Cancer 45:228–247
Faden RR, Chalkidou K (2011) Determining the value of drugs – the evolving British experience. N Engl J Med 364:1289–1291
Ferrara N (2005) The role of VEGF in the regulation of physiological and pathological angiogenesis. EXS:209–231
Fong PC et al (2009) Inhibition of poly(ADP-ribose) polymerase in tumors from BRCA mutation carriers. N Engl J Med 361:123–134
Geyer CE et al (2006) Lapatinib plus capecitabine for HER2-positive advanced breast cancer. N Engl J Med 355:2733–2743
Gianfaldoni S et al (2017) An overview on radiotherapy: from its history to its current applications in dermatology. Open Access Maced J Med Sci 5:521–525
Gilman A, Philips FS (1946) The biological actions and therapeutic applications of the B-chloroethyl amines and sulfides. Science 103:409–415
Gorre ME et al (2001) Clinical resistance to STI-571 cancer therapy caused by BCR-ABL gene mutation or amplification. Science 293:876–880
Hanahan D, Weinberg RA (2011) Hallmarks of cancer: the next generation. Cell 144:646–674
Havel JJ, Chowell D, Chan TA (2019) The evolving landscape of biomarkers for checkpoint inhibitor immunotherapy. Nat Rev Cancer 19:133–150
Horstmann E et al (2005) Risks and benefits of phase 1 oncology trials, 1991 through 2002. N Engl J Med 352:895–904
Huggins C, Hodges CV (2002) Studies on prostatic cancer. I. The effect of castration, of estrogen and of androgen injection on serum phosphatases in metastatic carcinoma of the prostate. 1941. J Urol 167:948–951; discussion 952
Hurvitz SA et al (2013) Phase II randomized study of trastuzumab emtansine versus trastuzumab plus docetaxel in patients with human epidermal growth factor receptor 2-positive metastatic breast cancer. J Clin Oncol 31:1157–1163
Jiang W, Ji M (2019) Receptor tyrosine kinases in PI3K signaling: the therapeutic targets in cancer. Semin Cancer Biol
Josephs D, Spicer J, O'Doherty M (2009) Molecular imaging in clinical trials. Target Oncol 4:151–168

Karapetis CS et al (2008) K-ras mutations and benefit from cetuximab in advanced colorectal cancer. N Engl J Med 359:1757–1765
Klapper LN et al (1999) The ErbB-2/HER2 oncoprotein of human carcinomas may function solely as a shared coreceptor for multiple stroma-derived growth factors. Proc Natl Acad Sci U S A 96:4995–5000
Krause DS, Van Etten RA (2005) Tyrosine kinases as targets for cancer therapy. N Engl J Med 353:172–187
Kwak EL et al (2010) Anaplastic lymphoma kinase inhibition in non-small-cell lung cancer. N Engl J Med 363:1693–1703
Lisberg A, Garon EB (2019) Does platinum-based chemotherapy still have a role in first-line treatment of advanced non-small-cell lung cancer? J Clin Oncol 37:529–536
Lynch TJ et al (2004) Activating mutations in the epidermal growth factor receptor underlying responsiveness of non-small-cell lung cancer to gefitinib. N Engl J Med 350:2129–2139
Maheswaran S et al (2008) Detection of mutations in EGFR in circulating lung-cancer cells. N Engl J Med 359:366–377
Mok TS et al (2009) Gefitinib or carboplatin-paclitaxel in pulmonary adenocarcinoma. N Engl J Med 361:947–957
Oken MM et al (1982) Toxicity and response criteria of the Eastern Cooperative Oncology Group. Am J Clin Oncol 5:649–655
Paez JG et al (2004) EGFR mutations in lung cancer: correlation with clinical response to gefitinib therapy. Science 304:1497–1500
Pao W et al (2004) EGF receptor gene mutations are common in lung cancers from "never smokers" and are associated with sensitivity of tumors to gefitinib and erlotinib. Proc Natl Acad Sci U S A 101:13306–13311
Parker C et al (2018) Current approaches to incorporation of radium-223 in clinical practice. Prostate Cancer Prostatic Dis 21:37–47
Parulekar WR, Eisenhauer EA (2004) Phase I trial design for solid tumor studies of targeted, non-cytotoxic agents: theory and practice. J Natl Cancer Inst 96:990–997
Patel AB, O'Hare T, Deininger MW (2017) Mechanisms of resistance to ABL kinase inhibition in chronic myeloid leukemia and the development of next generation ABL kinase inhibitors. Hematol Oncol Clin North Am 31:589–612
Piantadosi S, Fisher JD, Grossman S (1998) Practical implementation of a modified continual reassessment method for dose-finding trials. Cancer Chemother Pharmacol 41:429–436
Piekarz RL, Bates SE (2009) Epigenetic modifiers: basic understanding and clinical development. Clin Cancer Res 15:3918–3926
Punglia RS et al (2018) Clinical risk score to predict likelihood of recurrence after ductal carcinoma in situ treated with breast-conserving surgery. Breast Cancer Res Treat 167:751–759
Ratain MJ, Eckhardt SG (2004) Phase II studies of modern drugs directed against new targets: if you are fazed, too, then resist RECIST. J Clin Oncol 22:4442–4445
Rawlins M (2011) A new era for UK clinical research? Lancet 377:190–192
Reubi JC, Schonbrunn A (2013) Illuminating somatostatin analog action at neuroendocrine tumor receptors. Trends Pharmacol Sci 34:676–688
Rhee J, Hoff PM (2005) Angiogenesis inhibitors in the treatment of cancer. Expert Opin Pharmacother 6:1701–1711
Richardson PG et al (2005) Bortezomib or high-dose dexamethasone for relapsed multiple myeloma. N Engl J Med 352:2487–2498
Ricordel C et al (2018) Molecular mechanisms of acquired resistance to third-generation EGFR-TKIs in EGFR T790M-mutant lung cancer. Ann Oncol 29:i28–i37
Riviere I, Sadelain M (2017) Chimeric antigen receptors: a cell and gene therapy perspective. Mol Ther 25:1117–1124
Rosell R et al (2012) Erlotinib versus standard chemotherapy as first-line treatment for European patients with advanced EGFR mutation-positive non-small-cell lung cancer (EURTAC): a multicentre, open-label, randomised phase 3 trial. Lancet Oncol 13:239–246

Rosenberg B et al (1969) Platinum compounds: a new class of potent antitumour agents. Nature 222:385–386
Sadelain M, Riviere I, Riddell S (2017) Therapeutic T cell engineering. Nature 545:423–431
Salomon DS et al (1995) Epidermal growth factor-related peptides and their receptors in human malignancies. Crit Rev Oncol Hematol 19:183–232
Sequist LV et al (2011) Genotypic and histological evolution of lung cancers acquiring resistance to EGFR inhibitors. Sci Transl Med 3:75ra26
Simon R (1989) Optimal two-stage designs for phase II clinical trials. Control Clin Trials 10:1–10
Slamon D, Pegram M (2001) Rationale for trastuzumab (Herceptin) in adjuvant breast cancer trials. Semin Oncol 28:13–19
Slamon DJ et al (2001) Use of chemotherapy plus a monoclonal antibody against HER2 for metastatic breast cancer that overexpresses HER2. N Engl J Med 344:783–792
Small EJ, Roach M 3rd (2002) Prostate-specific antigen in prostate cancer: a case study in the development of a tumor marker to monitor recurrence and assess response. Semin Oncol 29:264–273
Smith MR et al (2018) Apalutamide treatment and metastasis-free survival in prostate cancer. N Engl J Med 378:1408–1418
Sobrero A, Bruzzi P (2009) Incremental advance or seismic shift? The need to raise the bar of efficacy for drug approval. J Clin Oncol 27:5868–5873
Soria JC et al (2018) Osimertinib in untreated EGFR-mutated advanced non-small-cell lung Cancer. N Engl J Med 378:113–125
Spicer JF, Rudman SM (2010) EGFR inhibitors in non-small cell lung cancer (NSCLC): the emerging role of the dual irreversible EGFR/HER2 inhibitor BIBW 2992. Target Oncol 5:245–255
Stratton MR (2011) Exploring the genomes of cancer cells: progress and promise. Science 331:1553–1558
Sullivan R et al (2011) Delivering affordable cancer care in high-income countries. Lancet Oncol 12:933–980
Suntharalingam G et al (2006) Cytokine storm in a phase 1 trial of the anti-CD28 monoclonal antibody TGN1412. N Engl J Med 355:1018–1028
Turner N, Tutt A, Ashworth A (2004) Hallmarks of 'BRCAness' in sporadic cancers. Nat Rev Cancer 4:814–819
Wang L, McLeod HL, Weinshilboum RM (2011) Genomics and drug response. N Engl J Med 364:1144–1153
Westover D et al (2018) Mechanisms of acquired resistance to first- and second-generation EGFR tyrosine kinase inhibitors. Ann Oncol 29:i10–i19
Wraith DC (2017) The future of immunotherapy: a 20-year perspective. Front Immunol 8:1668
Wu Z et al (2019) The international conference on intelligent biology and medicine (ICIBM) 2018: genomics with bigger data and wider applications. BMC Genomics 20:80
Yang JC, Rosenberg SA (2016) Adoptive T-cell therapy for cancer. Adv Immunol 130:279–294
Yap TA et al (2010) Envisioning the future of early anticancer drug development. Nat Rev Cancer 10:514–523
Yarden Y, Sliwkowski MX (2001) Untangling the ErbB signalling network. Nat Rev Mol Cell Biol 2:127–137
Yordanova A et al (2017) Theranostics in nuclear medicine practice. Onco Targets Ther 10:4821–4828

Monoclonal Antibodies: Past, Present and Future

J. Posner, P. Barrington, T. Brier, and A. Datta-Mannan

Contents

1 Introduction 83
2 Background 83
3 Monoclonal Antibody Structure Function and Production 86
3.1 Immunoglobulin Structural and Functional Components 86
3.2 Considerations for IgGs as Therapeutic mAbs 87
3.3 The Production of Therapeutic mAbs 88
4 Pharmacokinetics 90
4.1 Introduction 90
4.2 Route of Administration and Bioavailability 90
4.3 Distribution and Clearance 91
4.4 Metabolism, Stability and Elimination 92
5 Development 96
5.1 Preclinical Evaluation 96
5.2 The First-in-Human (FIH) Study 102
6 Monoclonal Antibodies for the Treatment of Inflammatory Conditions 106
6.1 Rheumatoid Arthritis 106
6.2 Anti-TNFs for Inflammatory Bowel Disease (IBD) 111
6.3 Monoclonal Antibodies for Spondyloarthropathies 113

J. Posner (✉)
JPC PharMed Ltd, Kent, UK

King's College London, London, UK

P. Barrington
Transcrip Partners, Reading, UK

T. Brier
Kings Health Partners, London, UK

Guy's and St Thomas' NHS Foundation Trust, London, UK

AstraZeneca, Cambridge, UK

A. Datta-Mannan
Eli-Lilly and Company, Indianapolis, IN, USA

J. E. Barrett et al. (eds.), *Concepts and Principles of Pharmacology*,
Handbook of Experimental Pharmacology 260, https://doi.org/10.1007/164_2019_323

7 Monoclonal Antibodies for the Treatment of Solid Tumours 115
7.1 Binding of Epidermal Growth Factors 115
7.2 Inhibition of Vascular Endothelial Growth Factor 117
7.3 CD20 Antigen and Fc-Mediated B-Cell Lysis 117
7.4 Immune Checkpoint Inhibition 118
7.5 Antigenic Targets 122
7.6 Immune Checkpoint Inhibitor Toxicity 122
8 Adverse Effects and Limitations of Monoclonal Antibodies 123
8.1 Infusion-Related Reactions Including Cytokine Release Syndrome 123
8.2 Antidrug Antibodies 125
8.3 Tumour Lysis Syndrome 125
8.4 Neurotoxicity 125
8.5 Other Limitations 126
9 The Future 127
9.1 Beyond mAbs: mAb-Based Biotherapeutics 127
9.2 Altering the Drug-Ability 128
9.3 Autoimmune Disease 128
9.4 Combination Therapies 129
9.5 Atopic Dermatitis 129
9.6 Bispecifics 129
9.7 Nanobodies 130
9.8 Intrabodies 131
9.9 Stereospecific and Catalytic mAbs 131
9.10 Local Administration of Monoclonal Antibodies 132
9.11 Monoclonal Antibodies for Less Common Diseases: Personalised Medicine 132
References 132

Abstract

Monoclonal antibodies (mAbs) are immunoglobulins designed to target a specific epitope on an antigen. Immunoglobulins of identical amino-acid sequence were originally produced by hybridomas grown in culture and, subsequently, by recombinant DNA technology using mammalian cell expression systems. The antigen-binding region of the mAb is formed by the variable domains of the heavy and light chains and contains the complementarity-determining region that imparts the high specificity for the target antigen. The pharmacokinetics of mAbs involves target-mediated and non-target-related factors that influence their disposition.

Preclinical safety evaluation of mAbs differs substantially from that of small molecular (chemical) entities. Immunogenicity of mAbs has implications for their pharmacokinetics and safety. Early studies of mAbs in humans require careful consideration of the most suitable study population, route/s of administration, starting dose, study design and the potential difference in pharmacokinetics in healthy subjects compared to patients expressing the target antigen.

Of the ever-increasing diversity of therapeutic indications for mAbs, we have concentrated on two that have proved dramatically successful. The contribution that mAbs have made to the treatment of inflammatory conditions, in particular arthritides and inflammatory bowel disease, has been nothing short of revolutionary. Their benefit has also been striking in the treatment of solid tumours and,

most recently, as immunotherapy for a wide variety of cancers. Finally, we speculate on the future with various new approaches to the development of therapeutic antibodies.

Keywords
Antibodies · Antidrug antibodies · Anti-TNFs · Biologics · Biotherapeutics · Epidermal growth factors · Immune checkpoint inhibition · Immunogenicity · Immunoglobulins · Inflammatory bowel disease · Infusion-related reactions · Monoclonal antibodies · Psoriasis · Rheumatoid arthritis · Spondyloarthropathies

1 Introduction

In this chapter, we describe the structure, production, pharmacological mechanism of action, pharmacokinetics and adverse effects of mAbs (monoclonal antibodies). Some important aspects of nonclinical and early clinical development are discussed along with challenges in this space. The enormous contributions that mAbs have made to two therapeutic areas, namely, inflammatory conditions and some cancers, are briefly described, and mention is made of some other conditions in which they have shown impressive efficacy in certain well-defined populations. Finally, we speculate on the future, pointing to various directions in which antibody-based biologics may offer further therapeutic opportunities.

2 Background

Biological therapies (biologics) encompass a large variety of medicines, and new technologies have enabled an exponential growth in these agents in the last 25 years. Nevertheless, it is worth recognising that some of the greatest advances in healthcare attributable to biological agents were made many years ago. The first, albeit indirect, use of antibody therapy was that of Jenner in 1796 when he administered cowpox virus as an immunisation against smallpox. In the early 1920s, vaccines against diphtheria, tetanus and pertussis were developed, followed by vaccines against polio in the 1950s and measles, mumps and rubella in the 1960s. The effectiveness and safety of these biologics remain superior to almost any other therapies ever discovered.

Antibodies that bind to ‘poisons’, the so-called antivenom or venom antisera, were first developed in the late nineteenth century and came into common use in the 1950s. They are still essential as life-saving therapies against venoms of various snakes, scorpions, spiders, box jellyfish and stonefish. Antivenoms can be effective against a single species (monovalent) or several species (polyvalent). As they are produced in animals, there is the potential for serious immune toxicity including anaphylaxis and serum sickness. However, the development of humanised and entirely human antibodies has reduced the frequency of such adverse effects and has contributed greatly to their improved ratio of benefit/risk and hence their use as therapeutics.

Antibodies are immunoglobulins produced by B lymphocytes, in particular differentiated as plasma cells, in response to exposure to antigens. The diversity of

antibody responses to antigens is attributable to the billions of recombinations that can be encoded by three types of genes entitled V, D and J, from which one allele of each type is combined in the hyper-variable regions of the kappa light and heavy chains of the antibodies.

In order to develop immunoglobulins targeting a specific membrane receptor or soluble ligand, a manufacturing process is required. The major breakthrough came in 1975 when César Milstein and Georges Köhler described a method for generating large amounts of monoclonal antibodies of a single predefined specificity. Their solution was to generate a hybridoma from fusion of mouse myeloma cell lines with mouse spleen B lymphocytes (Kohler and Milstein 1975). The hybridoma is a cell line that can be grown in culture and produces immunoglobulins that all have the same sequence of amino acids, i.e. a monoclonal antibody (mAb), and consequently the same affinity for a single chosen epitope on an antigen. They suggested that the ability to provide specific antibodies in massive in vitro cultures could be valuable for medical and industrial use, a discovery for which they shared the Nobel Prize in Physiology or Medicine in 1984. The application of hybridoma technology spread rapidly and widely throughout the world of biological research (Kennett 1981), and the development of monoclonal antibodies for medicinal use soon followed. Today, the large-scale manufacture and production of therapeutic mAbs for clinical application are mainly conducted by recombinant DNA (rDNA) technology using mammalian cell expression systems.

Another important milestone in the development of antibodies was the creation of a chimeric human antibody from murine variable region genes of a myeloma cell line with known antigen-binding specificity joined to a human immunoglobulin constant region gene using recombinant DNA methodology (Morrison et al. 1984). However, the first therapeutic mAb licensed for marketing in 1985 was not chimeric but was a murine orthoclone (OKT3) called muromonab-CD3 for treatment (not prevention) of allogeneic transplant rejection (Kung et al. 1979). Muromonab is an IgG2a directed against the CD3 epsilon chain of the CD3/TCR complex that characterises T lymphocytes and has been successfully used to treat allograft rejection in kidney, liver and heart transplantation (Hooks et al. 1991).

Since OKT3 is murine, it was, not surprisingly, extremely immunogenic in humans, producing a high titre of anti-mouse antibodies in most patients (Chatenoud et al. 1986; Lobo and Patel 1997). In addition, OKT3 is a potent mitogen, promoting T-cell proliferation and cytokine secretion and triggering a wide spectrum of adverse effects that include fever, chills, nausea, vomiting and headaches, summarised as ‘flu-like’, ‘cytokine release’ or ‘first-dose’ syndrome. Some patients suffer even more severe adverse effects such as cardiopulmonary distress, seizures, encephalopathy, meningitis, renal insufficiency and graft thrombosis (Sgro 1995).

The advances in genetic engineering in antibody structure permitted the shortcomings of OKT3 (immunogenicity and adverse effects) to be addressed. It was shown that the undesired effects provoked by that first-generation anti-CD3 mAb were caused by concomitant binding to the Fc receptors (FcR) on antigen-presenting cells (APCs) and to the CD3/TCR complex on T cells. The resulting strong T-cell activation produced a high transient release of pro-inflammatory

cytokines such as tumour necrosis factor-alpha (TNFα), interleukin 2 (IL-2), interleukin 6 (IL-6) and interferon-gamma (IFN-γ) by the targeted T cells after the first administration (Ferran et al. 1990). After it had been shown that non-FcR binding anti-CD3 mAb was still tolerogenic (Chatenoud et al. 1997), human anti-CD3 mAbs were rendered non-mitogenic by introducing mutations into the IgG backbone that led to highly decreased affinity to Fc receptors (Bolt et al. 1993; Alegre et al. 1994).

The advent of monoclonal antibodies (mAbs) has provided biological treatments that are effective in a very broad range of diseases. By the year 2000, about 15 mAbs had been registered, and in the last 20 years, the total has risen to about 100. The impact these molecules has had on patient care in the last 20 years or so is nothing short of revolutionary, and the pace of new developments shows no sign of slowing. In a period of 3 years from 2016 to 2018, 27 new antibodies were approved for marketing by the FDA, constituting over 20% of all approvals. In terms of the value of worldwide sales, seven of the top ten products in 2018 were biologics.

The number of new mAbs in development is increasing almost every year with about 50 currently in late-stage clinical development and over 500 in earlier phases. The vast majority of products have been targeted at two types of disease, namely, immune-inflammatory conditions and malignancies. The inflammatory conditions include rheumatoid arthritis, systemic lupus erythematosus, ankylosing spondylitis and other spondyloarthropathies, psoriasis and psoriatic arthritis and inflammatory bowel diseases such as ulcerative colitis and Crohn's disease. The malignancies included melanoma, breast cancer and various other solid tumours, certain lymphomas, leukaemias and multiple myeloma. A diverse range of mAbs for other indications have been developed including haemophilia A, thrombotic thrombocytopaenia, relapsing multiple sclerosis, atopic dermatitis, asthma, *Clostridium difficile* infection, migraine and prevention of osteoporosis.

Related biologics currently under development include bispecifics, fragments and antibody-drug conjugates. Some of the most recent developments include cytokines, stem cells and chimeric antigen receptor (CAR) T cells, but in this chapter we shall confine most of the discussion to mAbs and related molecules.

The variable region of the antibody determines the affinity for a given antigen, and many mAbs have very high affinities, frequently less than 1 nM. The constant region of the mAb induces the immune response after binding to the antigen. This specificity of mAbs makes them relatively devoid of undesired 'off-target' effects, a major difference from small chemical molecules. Furthermore, IgG molecules are metabolised by lysosomal proteolysis and not by cytochrome enzymes making them much less liable to drug interactions.

Their high specificity and low likelihood of interactions with other drugs led many in the scientific and medical communities to believe mistakenly that mAbs were much safer than chemical entities, and it took some time to recognise that mAbs have toxicities that can be serious and long-lasting. Immunogenicity is the most obvious property that can give rise to adverse effects. The first generation of mAbs was mostly murine, and administration to patients produced human anti-mouse antibodies (HAMAs) which caused hypersensitivity reactions including anaphylaxis and serum sickness and also resulted in rapid clearance of the mAb and

neutralisation of their effectiveness. To overcome these problems, genetic engineering and transgenic animals were developed to alter the amino acid composition of the antibody making it closer to that occurring in humans. Thus, a cell line could be transformed with alteration of specific residues by one or more replicable expression vectors including a suitable promoter linked to a DNA sequence which encodes at least part of an lgG heavy or light chain. The transformed cell line could then be cultured to produce the altered antibody. These techniques pioneered by scientists in Cambridge UK in the 1980s (Jones et al. 1986; Bruggemann et al. 1989) resulted in chimeric mAbs, e.g. infliximab and rituximab, in which the constant region of the antibody was human while the variable region remained murine.

Humanised antibodies, e.g. daclizumab and trastuzumab, represented the next step in which only the complimentary determining region was from a non-human source. Finally, entirely human mAbs, e.g. adalimumab, could be produced by recombinant DNA technology. The terminology should be noted as the suffixes -zumab, -ximab and -umab are used to denote murine, chimeric, and human mAbs, respectively. However, it must be recognised that, while human and humanised mAbs are less immunogenic and generally better tolerated than those from other species, they are still immunogenic. There is no evidence that the production of antidrug antibodies (ADAs) by fully human mAbs is less than with humanised and their administration to patients is by no means free from production of antibodies that may have important adverse effects (see Sect. 8).

3 Monoclonal Antibody Structure Function and Production

3.1 Immunoglobulin Structural and Functional Components

Therapeutic mAbs are immunoglobulin G (IgG) molecules that possess the same basic Y-shaped structure that consists of four polypeptide chains with two identical heavy (~50 kDa each) and light chain units each (~25 kDa each) (Fig. 1). The total molecular weight of mAbs is ~150 kDa. Disulphide bonds between the heavy and light chains provide covalent bond interactions that hold the polypeptide chains together. Each chain is comprised of constant (C_H and C_L) and variable domains (V_H and V_L). The antigen-binding region or Fv is formed by the variable domains of the heavy and light chains (V_H and V_L) and contains the complementarity-determining region (CDR), which imparts the high functional binding specificity of mAbs for the target antigen. The Fab or antigen-binding fragment is the 'V' part of the Y-shaped mAb structure and is composed of the V_H and V_L domains of the variable region and the C_{H1} domains of the heavy chain. The stem portion of the Y-shaped mAb structure is deemed the Fc region and comprises the C_{H2} and C_{H3} domains of the heavy chain. Functionally, the Fc region can interact with a diversity of cellular receptors, including (1) the neonatal Fc receptor (FcRn) that is responsible for the long circulating half-life of mAbs, their placental passage, and transport of IgG to and from mucosal surfaces and (2) components of complement C1q

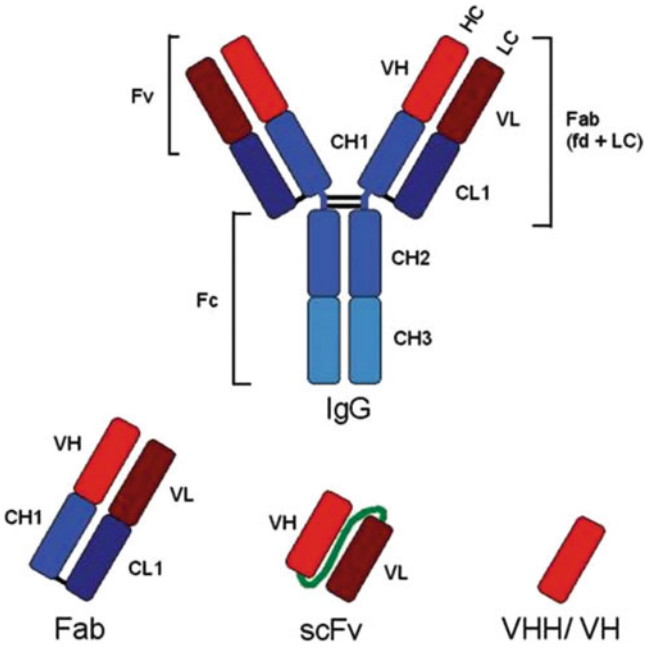


Fig. 1 Illustration of an IgG and the structural units of IgGs (Nelson 2010)

(i.e. complement system) and Fcγ receptors on the effector cells of the innate immune system (Vidarsson et al. 2014; Ryman and Meibohm 2017).

3.2 Considerations for IgGs as Therapeutic mAbs

There are four IgG subclasses IgG1, IgG2, IgG3 and IgG4 (Fig. 2). From a structural perspective, they are highly conserved but have differences in their constant region, most notably the hinges and upper CH2 domain. The increased structural complexity with multiple disulphide bonds of IgG3 molecules has made these difficult to develop as therapeutic mAbs. They have a much longer hinge region than the other three IgG subclasses making it more difficult to optimise target binding and drug-ability characteristics (pharmacokinetics, formulation and delivery) (Vidarsson et al. 2014; Ryman and Meibohm 2017). As a consequence, the majority of marketed mAb therapeutics are from the other IgG subclasses.

While the global structures of IgG1, IgG2 or IgG4 are similar, there are differences specific to each subclass that are important from a functional perspective. Most of these differences are in the hinge region and CH2 domain, as well as some more limited variations in Fc region. The space most proximal to the hinge region in the CH2 domain of the Fc component is responsible for effector functions of antibodies as it has largely overlapping binding sites for C1q (complement) and IgG-Fc receptors (FcγR) on

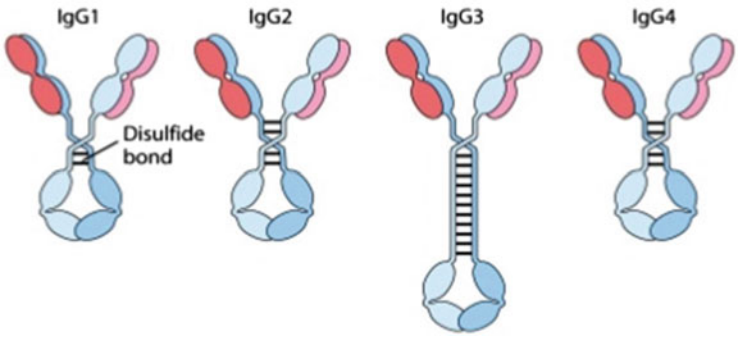


Fig. 2 Illustration of the four human IgG subclasses

effector cells of the innate immune system. Thus, the preference for one IgG class over the other is partly determined by factors such as effector function activity including antibody-dependent cellular cytotoxicity (ADCC) or complement-dependent cytotoxicity (CDC), which may be desired for the mAb pharmacodynamics. For example, IgG4 antibodies cannot fix complement or induce ADCC. It is also worth noting that some companies have engineered these regions to both reduce and enhance effector function as a means of harmonising therapeutic mAb platforms. Thus, therapeutic mAb isotype selection may also be a consequence of historical precedence, experience and availability of a particular IgG subclass in a company's discovery and development portfolio (Vidarsson et al. 2014; Ryman and Meibohm 2017).

In the development of biotherapeutics, it is critical to understand the target antigen. This includes the antigen's relevance in disease progression/dysregulation and its mechanism of action, site of activity, expression level, turnover kinetics and inherent nature, i.e. whether soluble, bound to or associated with membrane or shed from cell surface. This information about the target antigen impacts the design and selection of the properties of a therapeutic mAb such as the intended target binding affinity, need for ADCC and CDC effector functions, interaction with Fc receptors including FcRn and Fcγ and drug-ability properties, i.e. solubility, stability, off-target binding and potential to elicit antidrug antibody (ADA) response (Datta-Mannan 2019). The dissociation constant K_D of mAbs being generated today is usually in the pico-molar range, i.e. they have very high binding affinity.

3.3 The Production of Therapeutic mAbs

Every molecule in the mAb product is identical in its protein sequence and is thereby anticipated to have the same antigen-binding epitope, target binding affinity, biological interactions and downstream biologic effects. These attributes highlight the major distinguishing characteristics of mAbs from polyclonal antibodies, which are heterogeneous in protein sequence, and recognise various epitopes on an antigen which can lead to disparities in activity. The most common host cell lines for

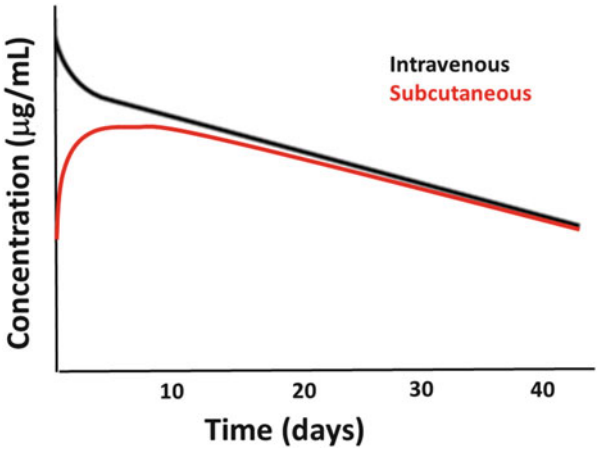


Fig. 3 The PK profile of a mAb following intravenous or subcutaneous administration

recombinant mAb expression are Chinese hamster ovary (CHO), used for about 70% of the pharmaceutical protein therapeutics (Fig. 3), Sp2/0 and NS0 mouse myeloma B-cell lines, HEK293 human embryonic kidney cells and the human cell line PER. C6. Considerations for each have been reviewed extensively (Kunert and Reinhart 2016).

Glycosylation is a common post-translational modification for IgG antibodies produced by mammalian cells, and recombinant mAbs are also glycosylated, though non-glycosylated mAbs are now being developed. Therapeutic mAbs have a consensus sequence (asparagine-X-serine/threonine where X is any residue except proline) for N-linked glycosylation at asparagine 297 in the CH2 domain of the heavy chain (Liu 2015). It is known that FcγRs interact with the carbohydrates on the CH2 domain and that the composition of these glycans has a substantial effect on effector function activity (Jefferis 2009).

Perhaps the best example of this is afucosylated (non-fucosylated) antibodies, which exhibit greatly enhanced ADCC activity through increased binding to FcγRIIIa (Yamane-Ohnuki and Satoh 2009). There are also some mAbs that are glycosylated in their Fab region, including the epidermal growth factor receptor (EGFR) mAb, cetuximab. It has been noted that change in glycan structure and composition leads to conformational alterations within the Fc domain and can result in variable effector function activities and pharmacokinetics; hence careful characterisation and process controls are an important part of therapeutic mAb production (Liu 2015). Given the complexities of production, validation of target engagement activity, comparability of post-translational modification and potential to elicit variable ADA, the development of biosimilar mAbs is an evolving space that is being evaluated by regulatory agencies on a case-by-case basis (Kaida-Yip et al. 2018).

4 Pharmacokinetics

4.1 Introduction

The majority of marketed mAb therapeutics are from the immunoglobulin G (IgG) subclasses consisting of IgG1, IgG2 or IgG4. The PK properties of these mAbs are generally characterised by limited distribution into tissues, slow peripheral clearance and long half-life in plasma. The slow clearance and long half-life facilitate less frequent administration than is required for small-molecule drugs. While this section focuses on the PK profile characteristics and factors influencing the PK and disposition of mAbs, the concepts also apply to other mAb-based biologics such as antibody drug conjugates (ADCs), Fc fusion proteins, bispecific antibodies and scFv domains, albeit with some unique PK differences as well as similarities. The reader is referred to articles considering these alternate modalities for further information (Coats et al. 2019; Datta-Mannan et al. 2016, 2018; Sedykh et al. 2018).

4.2 Route of Administration and Bioavailability

Most mAbs are administered by the intravenous (i.v.) or subcutaneous (s.c.) route; the intramuscular (i.m.) route is also sometimes used. Oral administration is precluded by their high molecular weight, poor solubility, proteolytic instability and poor permeability and by the fact that, being proteins, they are denatured in the gastrointestinal tract due to proteolysis. Figure 3 shows typical mAb serum or plasma PK profiles following i.v. or s.c. dosing. After i.v. administration, most mAbs display biphasic PK profiles with rapid distribution (α phase) and relatively slow elimination (β phase). After s.c. administration to humans, mAbs are absorbed slowly into the circulation facilitated by lymphatic drainage with time taken to reach maximal plasma concentrations (t_{max}) being several days. Bioavailability by the s.c. route typically ranges from ~50 to 100%.

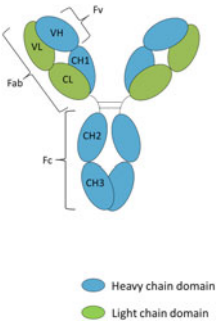
Subcutaneous administration has obvious advantages over the i.v. route. The patient does not have to attend a clinic or hospital for the injection to be delivered by a healthcare professional; rather it can be given by the patient themselves or care provider. However, the volume that can be delivered s.c. is limited, and injection at multiple sites is often required. Hyaluronidase has been included in some formulations to facilitate dispersion after s.c. administration, particularly when large volumes are required. Other problems with the s.c. route include increased likelihood of ADA formation compared with the i.v. route and a high incidence of local injection site reactions including pain, erythema, induration, oedema and pruritus, though they are generally tolerable.

Therapeutic mAbs have also been approved for local targeted delivery and other approaches to various tissues. For example, the locally active antibody fragment ranibizumab, which has been approved for treatment of age-related wet macular degeneration, is administered intravitreally (Bakri et al. 2007; Knodler et al. 2018). Direct pulmonary administration is being investigated for targeted delivery of mAbs for lung cancer (Cortez-Jugo et al. 2015).

4.3 Distribution and Clearance

Irrespective of the route of administration, the distribution and clearance of mAbs are related strongly to their structure and the targeted antigen. They can be broadly categorised into target-mediated drug disposition (TMDD) and non-target-related clearance mechanisms. Figure 4 shows the connectivity of the structural regions on mAbs and their PK profiles. Figure 5 shows representative mAb peripheral PK profiles affected by TMDD and non-target-related clearance mechanisms.

Due to their large size and physiochemical properties (charge and hydrophobicity), the distribution of mAbs is mainly limited to the vascular and interstitial spaces. The isoelectric point (pI) is the pH at which protein carries no net electric charge, and tissue uptake and clearance from the circulation are enhanced by higher pI values because of the tendency for these more basic mAbs to adhere to anionic sites of cell surfaces (Boswell et al. 2010). Tissue exposure is usually only about 5–15% of the total amount of mAb in the body, and the proportion able to enter the brain is generally about 0.1%. Distribution from the vascular space into the tissue interstitial space occurs via convection, and the IgG is taken up by endothelial cells or monocytes by endogenous fluid-phase pinocytosis or receptor-mediated endocytosis. When TMDD occurs, target-related factors including mAb binding affinity, target tissue expression, target turnover and mAb-target complex kinetics can impact distribution.



Structural Component	Functional Component	Connectivity to mAb PK
Fab	Antigen binding	TMDD mediates non- linear but saturable PK
Fab	Immunogenicity	
Fab	Physiochemical properties	PI, charge and hydrophobic interaction mediated off-target interactions that facilitate non-dose linear PK that is not saturable
CH2	Glycosylation	Glycan receptor mediated mAb elimination reduces $t_{1/2}$
Fc	FcRn interaction	Salvage of mAb in tissues and recycling mediates long $t_{1/2}$

Fig. 4 Relationship of the structural regions of mAbs on their PK profile

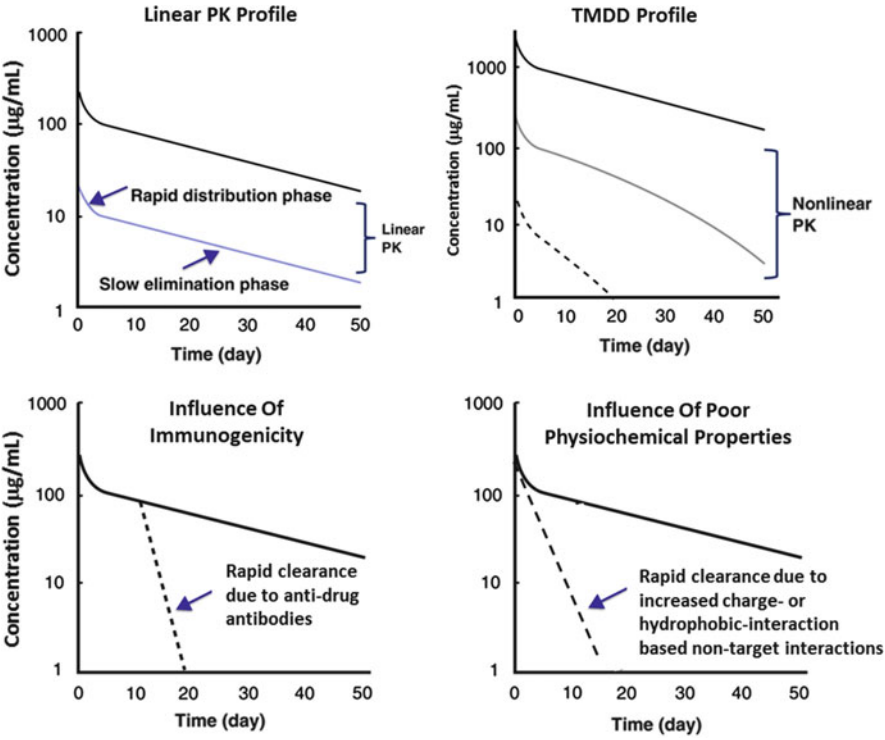


Fig. 5 Various PK profiles of a mAb affected by TMDD and non-target-related clearance mechanisms

4.4 Metabolism, Stability and Elimination

Unlike low molecular weight chemical entities, cytochrome enzymes are not involved in the clearance of mAbs. Their large molecular weight also precludes elimination of antibodies in urine though peptides and amino acids produced by their metabolism can be excreted by the kidney if they have not been reused for de novo synthesis of other proteins. The metabolism and elimination of therapeutic mAbs involve target and non-target-related mechanisms, and, as with distribution, target-related factors can impact metabolism and peripheral elimination when TMDD occurs.

Non-target binding-related metabolism and clearance of therapeutic mAbs is expected to be similar to those of endogenous IgGs and occurs in various body tissues as well as in plasma. The processes include fluid-phase pinocytosis or non-specific endocytosis and proteolysis by the tissues, recycling mediated by FcRn and the reticuloendothelial system (RES). The contribution of various organs to the elimination of endogenous IgG has been estimated by modelling to be 33%, 24%, 16% and 12% for the skin, muscle, liver and gut tissue, respectively (Garg and Balthasar 2007). Modifications and transformations of mAbs via post-translational

and product handling processes may affect the preponderance of the contributions of the target and non-target-related metabolism and clearance mechanisms. Compared with IgG1 and IgG2 molecules, IgG4s can exhibit instability due to exchanging antigen-binding parts between antibody molecules that results in new combinations in vivo, a process deemed 'Fab arm exchange' (Aalberse et al. 2009; Angal et al. 1993; Schuurman et al. 1999, 2001; Stubenrauch et al. 2010).

4.4.1 Target-Mediated Drug Disposition (TMDD)

TMDD occurs when a significant proportion of the administered mAb binds with high affinity to its intended pharmacological target and, as a result, the mAb displays exposures that are not proportional to dose (non-linear PK) at exposures that are lower than the concentrations of the target. TMDD is imparted by the mAb's complementary binding region (CDR) located within the Fab domain (Fig. 4). With increasing dose, the mAb exposures begin to exceed the target concentrations, and the binding sites become saturated. Accordingly, the mAb kinetics change and come to reflect non-target-related clearance mechanisms (Fig. 5) with the half-life often extending to that of a typical IgG of that class, e.g. 2–3 weeks for IgG1.

The non-linearity attributable to TMDD has implications for the safety and efficacy of mAbs. The effects may be minor or very pronounced, depending on both the nature of the target (or antigen) and the binding properties of the mAb to the target antigen (Samineni et al. 2016). Factors include turnover of the target, turnover of tissue(s) that express the target and whether the antigen is circulating as a soluble entity, present within cell membranes or is a shed extracellular domain (ECD) of membrane-bound targets. In the case of soluble or shed ECD targets, mAbs interact with the target in the circulation, thereby preventing the soluble target from interacting with its cognate receptor. The soluble and shed ECD targets can accumulate within the peripheral circulation bound to the mAb and follow the PK pattern of slowed clearance and long half-life of the mAb (PMID 24287601; 26265093). As a consequence, once bound, the clearance and elimination of these bound targets changes from their inherent mechanisms and follows the same pathway as the mAb. This phenomenon is particularly pronounced if the soluble antigen is present in high concentrations in the blood and/or has a high rate of turnover (Samineni et al. 2016; Talbot et al. 2015). In the case of membrane-bound targets, mAbs typically behave as agents that block the interaction of the target with their native ligand. There are mixed reports on the mechanisms related to the intracellular trafficking pathways membrane targets bound by mAbs, albeit the extent of partitioning of membrane targets to degradative pathways involving lysosomal compartments is clearly increased when the membrane target is mAb-bound (Liu 2018).

Once bound to its soluble/shed or membrane-bound target, variable effects have been noted. The mAb may be cleared via intracellular degradation as a complex bound to target so that higher mAb doses or more frequent administration is required to elicit its pharmacodynamic effects. Alternatively, intracellular release of the mAb from the target may be mediated by pH or target proteolysis with subsequent recycling of the non-target-bound mAb back into the peripheral circulation for additional target engagement. This can be beneficial by requiring administration of

reduced and less frequent dosing. In the case of proprotein convertase subtilisin/kexin type 9 (PCSK9) and members of the angiopoietin-like family (including ANGPTL-3, ANGPTL-4 and ANGPTL-8) being targeted for lipid-related cardiovascular disease, multiple investigated mAbs have displayed both of the aforementioned mechanisms influencing mAb PK (Lupo and Ferri 2018; Chaparro-Riggers et al. 2012; Henne et al. 2015). Similarly, with mAbs directed at membrane targets such as epidermal growth factor receptor (EGFR) and vascular endothelial growth factor receptor (VEGFR), differences in PK are due to differences in target-binding epitopes and binding properties (Wiley et al. 2003; Samineni et al. 2016). For therapeutic mAbs directed at interleukin 6 (IL-6), agents targeting the ligand or the IL-6 receptor have also shown variability in their PK and consequently in the doses and required dosing intervals in the treatment of rheumatoid arthritis (Kim et al. 2015; Narazaki et al. 2017). There remains an incomplete understanding of the implications for safety, efficacy, immunogenic response and tolerability of the differences in PK for mAbs directed at the same target and biological pathway. However, these examples illustrate that an understanding of the biology of the target and properties of the specific mAb in combination are not fully generalisable to all mAbs directed at the same target. The target and mAb binding properties including target affinity, intracellular trafficking and target epitope are critical for understanding the TMDD-based pharmacology and potential differences in patient responses to treatment.

4.4.2 Non-target-Related Factors Influencing mAb Disposition and PK

Many non-target factors intimately connected with the structure of mAbs impact their PK. These include physiological mechanisms such as the neonatal Fc receptor (FcRn). The hypothesis that IgG is protected from catabolism by specific receptors was first proposed by Roger Brambell in the 1960s (Brambell et al. 1964). The evidence for the existence of these receptors accumulated, and the 'Brambell receptor' was subsequently renamed the 'neonatal Fc receptor' (FcRn). FcRn is a major histocompatibility complex (MHC) class I-related receptor consisting of an α-FcRn chain and β2 microglobulin components. It is endogenously expressed in endothelial cells of most tissues in rodents, non-human primates (NHPs) and humans (Brambell 1966; Burmeister et al. 1994). FcRn mediates the long circulating half-life of mAbs via the Fc region (Fig. 2). Its role has been well established by both in vitro studies and in vivo murine knock-out systems which show rapid elimination of endogenous IgG. FcRn functions to salvage IgG, which is taken into cells by means of a pH-dependent binding mechanism within endosomes (Fig. 6). The high affinity of FcRn for IgGs and albumin within the low-pH environment of endosomes (pH ~5 to 6) facilitates binding, followed by recycling of the FcRn-IgG complex and release of bound species at the higher extracellular pH environment (pH ~7.4) where the FcRn affinity is markedly lower (Roopenian and Akilesh 2007) (Fig. 6). IgG that is not bound to FcRn within endosomes undergoes proteolytic degradation in lysosomes (Ward et al. 2003, 2005). Thus, the proportion of IgG processed through the recycling versus degradative pathways is critical in determining the clearance and half-life of IgG (and albumin) molecules in the circulation. Protein engineering is now being used to optimise the interactions of mAbs with FcRn with the aim of

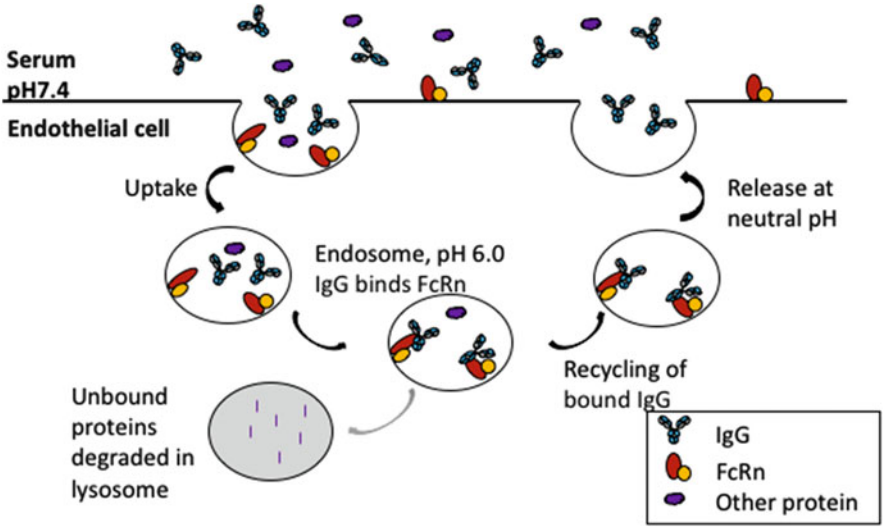


Fig. 6 FcRn-mediated pH-dependent salvage and recycling of IgGs (by permission of Absolute antibody.com. https://absoluteantibody.com/antibody-resources/antibody-overview/other-antibody-interactions/#fcrn)

improving their PK and disposition properties. Fc fusion molecules have shorter half-lives than whole IgG, e.g. etanercept half-life of around 3 days; this is partly the result of lower binding affinity to FcRn.

Other factors affecting PK include the composition of the parenteral injection and the site of administration, the presence of ADA and inherent physiochemical properties such as charge, hydrophobicity, molecular weight, secondary and tertiary structure and thermal and catabolic stability. Post-translational modifications including glycosylation, deamidation and methylation must also be taken into account.

4.4.3 Drug-Drug and Drug-Disease Interactions

Traditional drug-drug interactions (DDI) that affect the exposure of the administered drug or of concomitant medications that occur through time-dependent effects on enzymatic systems such as the cytochrome P450 (CYP) have not been observed for therapeutic mAbs. There have been no reports of an influence of mAbs on the metabolism or exposure of concomitant medications in trials specifically designed to evaluate these parameters (Ettlinger et al. 2006; Gaudreault et al. 2005; Xu et al. 2008; Zinner et al. 2004). However, some mAbs can have an indirect influence on hepatic clearance involved in the metabolism of small drugs. For example, by interfering with the activity of cytokines such as IL-6, tocilizumab indirectly affected simvastatin exposure in rheumatoid arthritis (RA) patients. The basis of this drug-disease interaction is that inflammation associated with increased cytokines reduces the activity of some CYP450 enzymes. In the case of simvastatin, reduced CYP3A4 activity slows its metabolism but suppression of inflammation by administration of

tocilizumab leads to enzyme activity similar to that found in the absence of inflammation. Hence, in cases where the recommended dosing regimen for an approved biologic is intended for long-term use in certain disease indications such as psoriasis and RA, a disease-drug interaction study should be considered (Wang et al. 2014).

There are also documented reports of potentially clinically relevant interactions between mAbs (such as infliximab, adalimumab and golimumab) and immunosuppressing drugs like methotrexate that may interfere with the clearance of mAbs by reducing the formation of antidrug antibodies (Zhou et al. 2007). Some investigators have also suggested small drugs could affect the expression of target that a mAb is directed towards and thereby alter TMDD. For example, the clearance of evolocumab and alirocumab was affected by statins and fibrates that induce PCSK9 expression though this was not considered clinically relevant (Ferri et al. 2016). Given the biological complexity and chemical heterogeneity of mAb therapeutics, it is entirely possible that these modalities will not show consistent DDI or drug-disease interactions based on similarities to the target or pathway towards which the mAb is directed. Considerable opportunities remain to continue to interrogate the clinical implications of potential DDI and drug-disease interactions with the long-term use of therapeutic mAbs.

5 Development

5.1 Preclinical Evaluation

5.1.1 Safety Evaluation

The ICH S6(R1) guideline is the primary guidance for the preclinical safety evaluation of biotechnology-derived pharmaceuticals (biopharmaceuticals). Examples of biopharmaceuticals include proteins and peptides, derived from cell cultures or produced using recombinant DNA technology including production by transgenic plants and animals. This guidance applies to monoclonal antibodies and fusion proteins.

The goals of preclinical safety evaluation are the same for both ‘small’ molecules and biopharmaceuticals and aim to:

- Characterise toxicity in nonclinical models
  - Identify potential target organs of toxicity
  - Determine whether such toxicity is reversible
  - Establish the no-observed-adverse-effect level (NOAEL)
- Assess and communicate risk (vs. benefit) for clinical use
  - Identify parameters that are predictive of toxicity and can be used clinically to monitor for potential toxicity in humans
- Identify an initial safe dose and dose escalation scheme in humans
  - Determine the exposure multiple at the NOAEL compared to the minimum anticipated biological effect level (MABEL) and proposed initial and maximum human dose

Table 1 Preclinical safety testing of small molecular entities and monoclonal antibodies

Study type or consideration	Small molecule entities	Monoclonal antibodies
General toxicology species selection and short-term dose range finding	Rodent and non-rodent with relevant metabolic profile	Rodent and non-rodent but often only non-human primate has pharmacological relevance
General toxicology	Typically 4 weeks progressing to 3 months and 6 months	Typically 6 months
Target binding assays	Broad panel of receptors	Target receptor/epitope tissue cross-reactivity by immunohistochemistry
Functional activity	Primary and secondary (safety) pharmacology in cardiovascular, respiratory and central nervous system and other systems as required	Primary and secondary (safety) pharmacology in cardiovascular, respiratory and central nervous system and other systems as required
Metabolism studies	Required	Not applicable
Immunogenicity	Not generally required	Required
hERG study	Required	Not generally required
Abuse liability studies	Often required	Not generally required
Developmental and reproductive toxicity	Required	Required
Genetic toxicology	Required	Not required
Carcinogenicity studies	Required	Often not justified or required

The ICH S6(R1) guideline promotes consistency without providing a standard list of nonclinical tests. A 'case-by-case' approach to preclinical safety evaluation is typical for mAbs for which the pharmacology of the agent should provide the scientific rationale and justification for the 'appropriate' nonclinical tests. Therefore, the preclinical studies and the endpoints investigated are determined by its mode of action, the nature of the target and the relevance of animal model(s).

Monoclonal antibodies have very high target specificity and selectivity binding to extracellular targets (soluble and membrane bound) with limited intracellular access. As a result of these properties, they have low off-target potential for adverse effects compared to small molecules but have the potential for ADCC via Fc receptor binding and CDC. Table 1 compares the preclinical assessment required for 'small' and 'large' molecules.

5.1.2 Choice of Animal Species for Toxicology

Due to the high target and species specificity of mAbs, there is often a lack of target binding in rodents, and the only pharmacologically relevant species may be a non-human primate (NHP) (e.g. cynomolgus or rhesus macaque monkey). Binding affinity should be assessed in rabbits as well as to human, rodent and non-human primate targets, to establish whether this species can be used in embryo-foetal developmental toxicity studies.

Table 2 Binding affinities and IC50 values across species

	Human	Cyno	Rat
K_D (pM)	2	78	3
IC_{50} (pM)	21	24	51

In addition to measuring target affinity, some form of functional assay should be performed as it is possible for the mAb to bind without engagement of the target. A real example of the data used to choose the appropriate species is shown in Table 2 below.

Cell-based data indicated similar neutralisation of species-specific target (IC_{50}), but the affinity data (K_D) for cynomolgus monkey suggested a significant species-specific difference. Due to these conflicting data, a study was performed to determine if the test article produced the expected PD effect in cynomolgus monkey with positive results. Based on these data, toxicity studies were then performed in rats (based K_D values) and in monkeys, as the test article produced the expected PD effect in the latter.

When there is no relevant preclinical species, one option with mAbs is to perform studies using a homologous protein, often referred to as a 'surrogate' antibody. For example, during the development of certolizumab pegol, a PEGylated Fab fragment of a humanised mAb which inhibits tumour necrosis factor-alpha (TNFα), it was recognised that its species specificity would not permit conventional reproductive toxicology in rodents or rabbits. Therefore, a PEGylated Fab anti-rat TNFα antibody was developed as a surrogate antibody in order to assess the potential reproductive toxicity (Wakefield et al. 2011). The importance of characterising the pharmacodynamic activity and toxicity of potential surrogate antibodies has been emphasised (Clarke et al. 2004). If the test article is capable of producing ADCC, and/or CDC, then the surrogate antibody should also have these functions. Surrogate antibodies are often used as models of efficacy to assess PK/PD models in animals, but much larger quantities are required for a toxicology programme.

A second option is to breed transgenic mice that express the human target using a 'gene knock-in' procedure. This technique can be expensive and time-consuming. If either a surrogate antibody or a transgenic animal approach is being considered, it is strongly recommended that advice from regulatory agencies is sought before toxicology studies are initiated.

Lack of a relevant preclinical species does not preclude administration of the mAb to man. Development programmes have proceeded in the absence of any toxicology and very limited in vivo data. In these circumstances, emphasis should be placed on the clinical plan in which the starting dose would be very low and the dose escalation scheme very cautious. Early discussion with regulatory authorities on the acceptability of individual programmes is recommended (Chapman et al. 2009).

As specified in the ICH S6 guideline, tissue cross-reactivity (TCR) studies are required to assess the potential binding of mAbs and related products to the target epitope in various tissues. This is an ex vivo assessment using immunohistochemical (IHC) techniques to evaluate potential binding of the mAb to a standardised panel of

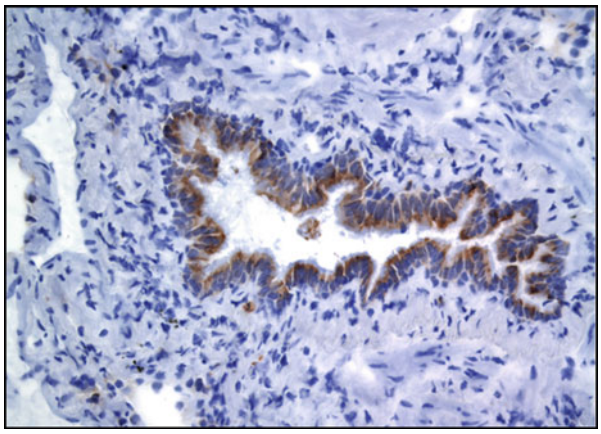


Fig. 7 Antibody binding indicated by brown staining in this tissue section

32 different tissues from laboratory species and human tissue bank. The test antibody is applied to tissue sections, and binding is detected using a second antibody that includes a chromophore or 'colouring agent'. Any tissue that has bound the test antibody is demonstrated by the affected cells showing brown pigmentation (see Fig. 7).

The data obtained from the TCR supplement the knowledge of target distribution. Staining may be present in tissues that were not expected (unintentional cross-reactivity) and differences in binding profile/pattern across and within tissues from human compared to the animal species selected for toxicology studies.

The antigen exposure in a TCR assay does not mimic the exposure that may occur in vivo. Freezing, cutting and fixation of the tissues disrupt cells and expose intracellular epitopes that are not normally accessible in vivo (Leach et al. 2010). Therefore off-target binding may occur as a result of the procedure.

5.1.3 T-Cell-Dependent Antibody Response (TDAR)

Before plasma cells can produce antibodies to an antigen, a complex series of events must occur, including antigen uptake and presentation, T-cell help and B-cell activation. If the mAb being tested could have immunomodulatory effects (e.g. anti-cytokine antibodies), assessment of the T-cell-dependent antibody response should be performed. Subcutaneous (SC) injections of a suitable antigen, e.g. keyhole limpet haemocyanin (KLH), are administered as an 'immunisation' dose followed approximately 14 days later, by a 'challenge' dose. Serial blood samples are taken for measurement of anti-IgM and IgG antibodies to KLH. The antibody responses can be compared to placebo-treated animals, and if reduced in animals receiving the test article, any potential dose response can be assessed.

5.1.4 Duration of Toxicology

For the main toxicology programme, two studies are performed, with the first typically of 2–8 weeks duration. The chosen duration is determined by several factors such as clinical indication/development plan and the perceived risk of immunogenicity of the

test article occurring in the toxicology species. The second study is of approximately 6 months in duration. If the T-cell-dependent antibody response is incorporated into the first toxicology study, the study should be of at least 6 weeks duration in order to administer the antigen and obtain the appropriate serial blood samples.

In addition to study duration, the need for recovery groups needs to be assessed. As discussed earlier, mAbs have a low potential to produce off-target effects; the need for recovery groups is likely to be driven by pharmacology. For example, a mAb targeting CD20 will cause profound B-cell depletion, and therefore a recovery group (normally the highest dose group) would be required to demonstrate B-cell recovery.

5.1.5 Dosing in Toxicology

Normally three dose levels will be required in the first toxicology study. The low and mid doses would be based on the in vivo animal PK and PD data and would use both the dose expected to saturate target and information regarding the steepness of the dose-response curve for the pharmacodynamic endpoint(s). The low dose should be anticipated to have no or minimal PD effects. The highest dose would be no greater than ten times the anticipated therapeutic dose. In practice, doses of 10, 30 and 100 mg/kg are often used in NHP studies. As mAbs are often given by the subcutaneous route for patient convenience, dosing by this route would normally be included at one of two lower dose levels in the toxicology studies. The highest dose level is administered intravenously as the total injection volume would prohibit the subcutaneous route. Doses are usually administered once or twice weekly, although if the compound is anticipated to be highly immunogenic in animals, larger and more frequent doses (three times weekly) may be administered. This is referred to as a 'dosing through' strategy.

In addition to standard toxicological endpoints (body weight, food consumption, clinical signs, clinical pathology, anatomic pathology), repeat-dose studies for a mAb often include the following safety pharmacology endpoints:

- Cardiovascular: ECGs, heart rate, blood pressure
- Respiratory: rate (minute volume, tidal volume, oxygen saturation)
- Central nervous system: neurological examination, coordination, sensory/motor reflexes, core temperature, motor activity, behavioural changes

5.1.6 Immunogenicity

As discussed in Sect. 2, ADA can develop both in animals and humans. Immunogenicity is directly assessed by determining ADA titres in serum using a drug-tolerant assay (tolerance threshold typically 100 μg/mL). Immunogenicity may be indirectly assessed using PK/PD data. Reductions in exposure using an antigen-capture assay or a reduced pharmacodynamic effect are highly suggestive of the formation of neutralising ADA.

In toxicology studies there are two main consequences of ADA formation. The first is that neutralising antibodies develop reducing the exposure in the animals. In some cases, the lowest dose group will have no exposure for a significant portion of the dosing period and therefore cannot be used in the determination of a no

observable adverse effect level (NOAEL). Secondly, ADA formation can result in the formation of immune complexes. These become deposited in blood vessels and can result in inflammation, e.g. vasculitis, myositis and/or glomerulonephritis. The development of immune complex disease does not necessarily preclude further development of the mAb as immunogenicity risk is higher in animals as they have received humanised or fully human mAb.

Immunomodulation can cause immune suppression or enhancement. The effect may be desired, e.g. suppression of a targeted pathway or process involved in autoimmune disease such as RA or psoriasis, but excessive immune suppression can increase the rate of infection and, in particular, opportunistic infections.

Immunotoxicity results in an immune-mediated adverse event/s, e.g. injection site reactions, anaphylactic reactions and unintended immunosuppression. If the mAb being tested (e.g. anti-cytokine antibodies) could affect humoral immunity, assessment of the T-cell-dependent antibody response should be performed (See Sect. 5.1.3).

Effects on cellular immunity are studied by measurement of white blood cell counts. A typical safety panel would include total B cells, total T cells, $CD4^+$ and $CD8^+$ T cells and natural killer (NK) cells. Cytotoxicity assays for T and NK cells may also be performed to determine if the test article is affecting cell number, cell function or both. Clearly if a specific cell type is being targeted, e.g. memory B cell, these would be measured during the study.

5.1.7 Carcinogenicity

The assessment of carcinogenic potential or the ability to promote tumour growth is among the most challenging areas in the nonclinical assessment of biotherapeutics. In the initial development of these therapies, there was a perception that biotherapeutics appeared to be exempt from carcinogenicity concerns. This perception was largely based on the fact that 2-year rodent studies were often not possible and genotoxicity concerns typically do not exist for biologics (Vahle et al. 2010).

However, biotherapeutics may increase the incidence of existing neoplasms by secondary mechanisms related to their pharmacology, e.g., promotion of growth and cell differentiation or proliferation. In addition, effects on immunomodulation could result in neoplasia. Interestingly chronic immune activation (inflammation) enhances the risk of neoplastic progression, but several anti-TNF monoclonal antibodies (that reduce inflammation) have a black box warning for 'lymphoma and other malignancies'. Suppression of anti-tumour immune responses can foster carcinogenicity via, for example, activation of latent oncogenic viruses such as HPV or EBV. In the case of ustekinumab (anti-IL-12/IL-23), a large post-marketing study was performed to assess this risk (see Sect. 6.3.1).

When considering the need for a carcinogenicity study, the potential/theoretical concerns should be identified based on the target biology, mode of action, published nonclinical evidence and available clinical experience. All available pharmacology/toxicology data relating to the monoclonal antibody should be assessed to determine if there are any signals of concern. Finally, discussion with regulatory agencies to agree a carcinogenicity assessment strategy is strongly encouraged.

5.2 The First-in-Human (FIH) Study

In this section, the particular considerations relating to the first administration of a mAb and how such studies differ from those with small molecular entities (SMEs) will be briefly discussed.

5.2.1 Study Population

Traditionally the FIH of SMEs in single ascending doses (SAD) followed by multiple ascending doses (MAD) is conducted in healthy subjects unless there are particular concerns relating to toxicity, e.g. chemotherapy and anaesthetics. Though dosing of patients with the target disease may be incorporated within the same protocol as the SAD/MAD studies, this transition often follows pharmacokinetic studies of concomitant food ingestion, the elderly and critical drug-drug interactions, as well as pharmacodynamics, to arrive at the range of doses to be studied in the target patient population. In the case of mAbs, the first-in-human study may often be conducted in patients. A number of factors are used to determine the most appropriate population on a case-by-case basis. These include safety, mechanism of action, target, the possibility of immunogenicity, potential for local or systemic infusion reactions and cytokine release. It should also be remembered that in contrast to SMEs which typically have half-lives of a few hours or possibly a day or two, the typical half-life of an IgG is 2–3 weeks and biological effects may last far longer.

There is some evidence that the immunogenicity findings in healthy subjects may be poorly predictive of the response in a target patient population due to the confounding effects of disease and concomitant medication. For example, patients with an autoimmune disease such as systemic lupus erythematosus (SLE) may have a higher incidence of ADAs than healthy subjects. By contrast, patients receiving immunosuppressants such as methotrexate or azathioprine may have a lower incidence of ADAs. The development of ADAs is dependent on several factors, a major one being duration of dosing. A low incidence of ADAs following single doses may underestimate the true value that occurs following multiple dosing.

There are also positive reasons for conducting the FIH in patients. For some diseases, responses to mAbs will be seen after a single dose so that valuable dose-response information can be obtained using clinical endpoints rather than biomarkers or, as is often the case, simply plasma concentration data. An example of a clinical response being evident after a single dose is that of the FIH conducted with of ustekinumab in patients with psoriasis in which doses as low as 0.1 mg/kg showed a significant reduction in skin lesions 16 weeks after dosing in some patients (Kauffman et al. 2004).

Furthermore, in contrast to the PK of SMEs in healthy subjects, which usually are predictive of those in patients, the disposition of mAbs in patients may be quite different from those in healthy subjects. As explained in Sect. 4, target-mediated disposition is often present, the PK will not be linear with dose and will be dependent on expression of the target. For example, the clearance of an anti-IL-23 mAb (anrukinzumab) was faster in patients with ulcerative colitis compared to both healthy volunteers and patients with asthma (Hua et al. 2015). By contrast, healthy

subjects may not express the target at all, and the pharmacokinetics will simply be those of an IgG.

Therefore, the decision whether to dose healthy subjects or patients must be carefully considered. Administration of single doses of a mAb to healthy subjects can be useful to assess bioavailability by the s.c. route and to obtain information on pharmacodynamic effects if a suitable marker is available. An example of a target engagement assay is the measurement of the concentration of IL-17 after the administration of an anti-IL-17 antibody. The concentrations of IL-17 are low in healthy subjects, but when this cytokine binds to the mAb, it acquires the half-life of the antibody, and therefore total IL-17 concentrations increase, generally, in a dose- and time-dependent manner. It must be remembered that target engagement does not automatically lead to pharmacodynamic effects, and therefore PD marker/s should always be sought.

5.2.2 Route of Administration and Dosing Cohorts

The route of administration in the FIH study of a novel mAb is usually intravenous (i.v.). Unlike the swallowing of a tablet from which there is no going back, drug can be administered i.v. slowly over at least an hour but often for a longer period, e.g. 4 h. This means that, should there be an infusion reaction, cytokine release or some other clinically significant adverse event, the infusion can be slowed or stopped. Another advantage of administration by the i.v. route is that bioavailability is 100% so that, in contrast to oral administration of an SME, it is known precisely what dose entered the systemic circulation.

Sentinel dosing should be employed. This means that on the first day of administration of the single ascending dose and subsequent dose levels, only two subjects should be dosed with one receiving the mAb and the other placebo under double-blind conditions. At least 24 h but preferably 2 or more days should be allowed for observation of these two subjects, and all safety data must be reviewed before dosing of further subjects should be undertaken. For some mAbs it may be appropriate to dose a total of 6–8 subjects with at least two of these (including one of the sentinels) receiving placebo.

In view of the long half-life of mAbs and potential immunogenicity, the initial investigation will only involve one dose for each subject, and follow-up visits for safety should continue for 8–12 weeks. Safety permitting, the dose will normally be escalated for the second cohort, and once again, the initial assessment of safety should be conducted in two sentinel subjects. All of the doses may be administered i.v., but if it is intended to administer the mAb subcutaneously (s.c.) in subsequent clinical trials, it is often useful to include a cohort to receive drug by the s.c. route in the FIH study. This will provide an estimate of bioavailability from the ratio of AUC by the s.c. and i.v. routes and the time taken to reach maximum concentrations (t_{max}) which is typically several days. t_{max} is usually long because of slow transportation of the material from the site of injection and surrounding tissue via lymphatic drainage. Because of the time taken to reach the systemic circulation, it may be necessary to keep subjects under observation in the study unit for up to a week. It may not be practical to administer high doses by the s.c. route due to the volume required even if

it is split into two or three injections, but at least one dose level should be given by this route for the reasons stated above.

For studies with anticancer mAbs in which it is not usual to administer a placebo, a typical design will be 3 + 3 in which three subjects are dosed at a particular dose level with just one or maybe two subjects dosed on 1 day. If there are no dose-limiting or clinically significant adverse events, the dose may be escalated for the next cohort of three, but if such adverse events are evident, an additional three subjects will be dosed at the same dose level. Depending on the outcomes in the total of six subjects, a decision will be made to proceed with escalation in the next cohort or to stop escalation and proceed to expand the cohort at the same dose level or possibly to expand at a lower dose.

Complications of treatment with mAbs in patients with malignancies include cytokine release syndrome (CRS) and tumour lysis syndrome (TLS). CRS is characterised by fever, nausea, headache, rash, tachycardia, hypotension, dyspnoea, arthralgia and myalgia. With increased severity hypotension progressing to circulatory shock and multi-organ failure with acute respiratory distress syndrome and coagulopathy may occur.

TLS comprising hyperuricemia, hyperkalemia, hyperphosphatemia and hypocalcemia is much less common but can be life-threatening and can occur several days after treatment. Patients with a large tumour burden may be at greatest risk, and TLS syndrome can occur in patients with solid tumours treated with checkpoint inhibitors as well as in patients with liquid tumours (leukaemias and lymphomas).

5.2.3 Selection of Starting Dose

Selection of the starting dose will take into account both toxicology and pharmacology. Toxicology will provide the no observable adverse effect level (NOAEL), and the nature of the findings may be critical. However, the minimum anticipated biological effect level (MABEL) can be estimated from the primary animal pharmacology with PK/PD modelling. Most importantly, it should be recognised that there are serious limitations when making predictions for the active dose in humans from animal data. The calculations and decisions should recognise the assumptions being made, and a 'case-by-case' approach is essential, as summarised by J. Sims on behalf of the Association of the British Pharmaceutical Industry (ABPI). While the MABEL should underlie the calculation, it is often appropriate to add a further margin of safety of, for example, tenfold or more, to arrive at a safe starting dose.

Expert advice may be needed and is available from regulatory authorities such as the Medicines and Healthcare Regulatory Agency (MHRA) in the UK https://www.gov.uk/guidance/clinical-trials-for-medicines-apply-for-authorisation-in-the-uk#clinical-trial-phases. Extra caution is required if the mAb has the following characteristics often described as 'higher-risk' molecules:

- Mode of action involves a target that is connected to multiple signaling pathways (target with pleiotropic effects), e.g. leading to various physiological effects or targets that are ubiquitously expressed.

- Compound acts (directly or indirectly) via a cascade system where there may be an amplification effect which might not be sufficiently controlled by a physiological feedback mechanism.
- Compound acts (directly or indirectly) via the immune system with a target or mechanism of action which is novel or currently not well characterised.
- Novelty in the structure of the active substance, e.g. a new type of engineered structural format such as those with enhanced receptor interaction as compared with the parent compound.
- Level of expression and biological function of the target receptor may differ between healthy individuals and patients with the relevant disease.
- There is insufficient available knowledge of the structure, tissue distribution, cell specificity, disease specificity, regulation, level of expression and biological function of the human target, including downstream effects.
- Compound acts via a possible or likely species-specific mechanism or where animal data are unlikely to be predictive of activity in humans.

One of the early decisions to be made is whether the dose will be administered on an mg/kg (or body surface area) basis or as a dose based on a fixed body weight (60 kg based on FDA guidance). A pragmatic approach is to administer a fixed dose and subsequently analyse the PK data to determine if body weight or body surface area appears to affect exposure. This approach is supported by Wang et al. (2009).

5.2.4 Study Design for Single Ascending Doses

A typical design of the FIH study is shown in Table 3 which shows an example of dose levels, administration routes and treatment allocation. It should be noted that only six subjects receive the mAb via the s.c. route making any conclusion regarding the tolerability (injection site pain) or occurrence of injection site reactions (swelling, induration, erythema, etc.) difficult to assess. However, the formulation may not be the same as that which is finally marketed.

The dose escalations shown in Table 3 (up to 300 mg) are based on approximately a half log unit, and this may need to be reduced for higher-risk mAbs. The dosing interval between cohorts should be at least 14 days, and a formal data review meeting should be held to review all available data and determine if dose escalation is appropriate. Blood samples for safety, PK, PD biomarkers and immunogenicity (IgM and IgG ADAs) should be taken at the appropriate intervals over a 12-week period.

If the study is being conducted in patients and depending on the target disease, it may be considered inappropriate to dose eight subjects (including two placebo) at doses that are likely to be ‘subtherapeutic’. Therefore, consideration should be given to dosing fewer subjects at the lower end of the dose range and then increasing the numbers as the dose is increased using clinical responses (desired and undesired), biomarkers and plasma concentration data on which to base a decision at which dose level to enlarge the cohorts.

Regarding repeat dosing, this will begin with a lower dose than has been reached in the single ascending dose and with an interval of 2–4 weeks between doses. The

Table 3 Typical study design for single ascending doses of a mAb

Cohort	Dose and route									
1	10 mg i.v.	Subject no.	01[a]	02[a]	03	04	05	06	07	08
		Treatment	P	A	A	A	P	A	A	A
2	30 mg i.v.	Subject no.	09[a]	10[a]	11	12	13	14	15	16
		Treatment	A	P	A	A	A	P	A	A
3A	100 mg i.v.	Subject no.	17[a]	18[a]	19	20	21	22	23	24
		Treatment	P	A	A	A	A	A	P	A
3B	100 mg s.c.	Subject no.	25[a]	26[a]	27	28	29	30	31	32
		Treatment	A	P	A	P	A	A	A	A
4	300 mg i.v.	Subject no.	33[a]	34[a]	35	36	37	38	39	40
		Treatment	A	P	P	A	A	A	A	A
5	600 mg i.v.	Subject no.	41[a]	42[a]	43	44	45	46	47	48
		Treatment	P	A	A	A	A	A	A	P

A active, *P* placebo, *i.v.* intravenous, *s.c.* subcutaneous
[a]Sentinel subjects

response of the disease and associated biomarkers are of course critically important in establishing the effective dose range. The ability to demonstrate efficacy with multiple ascending doses is limited by the number of patients and treatment duration, and of course the lowest dose administered may be subtherapeutic. However, changes or trends in disease biomarkers may be observed.

6 Monoclonal Antibodies for the Treatment of Inflammatory Conditions

One of the most important disease areas treated with biologics and mAbs in particular has been the inflammatory arthritides including rheumatoid arthritis (RA), juvenile rheumatoid/idiopathic arthritis, ankylosing spondylitis and psoriatic arthritis. Many patients with inflammatory bowel diseases such as Crohn's disease and ulcerative colitis have also benefited enormously from the same group of mAbs.

6.1 Rheumatoid Arthritis

RA is an autoimmune, inflammatory condition affecting about 1% of the adult population. Rheumatoid factor is an autoantibody produced by plasma cells. One clinical measure of the severity of inflammation is C-reactive protein, an acute-phase reactant. RA primarily attacks the synovial membrane of multiple joints causing inflammation with symmetrical synovitis which typically affects the meta-carpophalangeal and the proximal interphalangeal joints of the hands and feet (Lee and Weinblatt 2001). Progressive painful joint damage with destruction of cartilage and bone is accompanied by systemic inflammation with many complications

affecting cardiovascular and respiratory systems, the kidney, eyes, bones and other organs.

Until almost the early years of the twenty-first century, treatment comprised symptomatic relief with nonsteroidal anti-inflammatory drugs and immunosuppression with methotrexate, azathioprine, corticosteroids, disease-modifying anti-rheumatic drugs (DMARDs) such as gold and penicillamine and the antimalarials chloroquine and hydroxychloroquine. Despite the variety of drugs available, treatment was often ineffective and associated with toxicity. Patients typically suffered from painful, progressively deforming arthritis with many of the extraarticular complications, disability and reduced life expectancy. The advent of biological therapies such as cytokine modulators has changed the outlook and quality of life dramatically for the majority of patients with severe disease. The clinical benefit to patients of these biological DMARDS listed in Table 4 has been demonstrated in randomised, controlled clinical trials, and their early introduction to the therapy of patients with active disease can have a remarkably beneficial effect on the outcomes (Smolen et al. 2016).

Table 4 Monoclonal antibodies for treatment of rheumatoid arthritis

Agent	Structure	Target and mechanism of action
Infliximab	Chimeric mAb	Blockade of soluble and transmembrane cell surface TNFα receptors
Etanercept	Dimeric p75 Fc fusion protein of TNFR2 with an IgG1 Fc region	Blockade of transmembrane TNFα cell surface receptors
Adalimumab	Human mAb	Blockade of soluble and transmembrane TNFα cell surface receptors
Golimumab	Human mAb	Blockade of soluble and transmembrane TNFα cell surface receptors TNFα
Certolizumab pegol	Humanised PEGylated fab fragment conjugated to polyethylene glycol (No Fc)	Blockade of soluble and transmembrane TNFα cell surface receptors TNFα
Abatacept	Fusion protein of IgG1 Fc region fused to extracellular domain of CTLA-4	Binds to CD80 and CD86. Prevents T-cell activation by antigen presenting cells delivering co-stimulatory signal
Rituximab[a]	Chimeric IgG1 mAb	Fab binds to transmembrane phosphoprotein CD20 on pre-B and mature B lymphocytes. Fc recruits immune effector functions to mediate B-cell lysis
Tocilizumab[b]	Humanised IgG1 mAb	Soluble and membrane-bound IL-6 receptors
Sarilumab	Human IgG1 mAb	Soluble and membrane-bound IL-6 receptors

[a]Also indicated for NHL, CLL, Wegener's granulomatosis and microscopic polyangiitis, pemphigus vulgaris

[b]Also indicated for treatment of severe cytokine release syndrome

6.1.1 Anti-TNFα Cytokine Modulators

Tumour necrosis factor-alpha (TNFα) is a cytokine synthesised by macrophages as a transmembrane protein which is cleaved into a soluble form. The cytokine has inflammatory properties, and antagonism of both the transmembrane and soluble forms has proved to form the basis of highly effective therapies for inflammatory arthritides and bowel disease. TNFα is so called because it was found to cause necrosis of subcutaneous tumours in mice but was subsequently found to contribute to growth of a variety of malignant tumours including ovary, breast, colon and skin. Targeting by TNFα of its R1 receptor present on all tissues results in cross-linking which results in a pro-inflammatory response with production of cytokines such as IL-1, IL-6 and granulocyte-macrophage colony-stimulating factor (GM-CSF) (Buchan et al. 1988; Feldmann et al. 1996). Increased concentrations of TNFα are present in synovial biopsies of the inflamed joints and in the synovial fluid of patients with RA; concentrations correlate with disease activity. The TNFα is produced by cultured mononuclear cells from the joints of patients with RA, and immunohistochemistry showed that TNFα is present in the synovium including in the cells in the cartilage pannus junction (Haworth et al. 1991). This dysregulation with persistent overexpression of several cytokines was found to be blocked by administering an anti-TNF which suggested that TNF was responsible for the overexpression of pro-inflammatory cytokines (Brennan et al. 1989). The biological activity of TNFα is dependent on cross-linking of cell surface TNF receptors. Studies in animal models of arthritis showed the anti-mouse-TNF mAbs to be effective in suppressing collagen-induced arthritis (Williams et al. 1992). Chronic exposure to TNF in vitro was shown to impair T-cell activation, and this could be reversed in vivo by anti-TNF antibodies in RA patients (Cope et al. 1994).

Proof of principle of the anti-TNFα chimeric IgG mAb known as cA2, subsequently called infliximab, was then established in an open clinical trial conducted in just ten patients (Elliott et al. 1993). All patients showed clinical and biochemical responses. It soon became apparent that the effect of treatment wore off after about 12 weeks. However, further infusions were administered to eight of the original ten patients when they relapsed, and four received four cycles of treatment with benefit seen as reduction of swollen joint counts and CRP, though the duration of benefit tended to decline with repeated infusions. Human anti-chimeric A2 antibody responses developed in some patients but were not necessarily associated with symptoms or a decline in response (Elliott et al. 1994).

The open label study was followed by a randomised, double-blind clinical trial conducted in four centres with single infusions of 1 or 10 mg/kg cA2 compared with placebo in 73 patients with active RA (Elliot et al. 1994). The clinical response at that time was based on a composite index developed by Paulus et al. (1990). A Paulus 20% response was based on an improvement of 20% or greater for morning stiffness, erythrocyte sedimentation rate, joint pain/tenderness score and joint swelling score and physician's overall assessments of current disease severity. Only 2 of 24 patients who received placebo showed a Paulus 20% response at week 4. By contrast, of 25 patients treated with 1 mg/kg cA2, 11 responded, and of 24 patients treated with 10 mg/kg cA2, 19 showed a Paulus 20% response, and over

half showed a Paulus 50% response. The magnitude of these responses was impressive, with maximum mean improvements in individual disease activity assessments, such as tender or swollen joint counts, and in serum CRP, exceeding 60% for patients on high-dose treatment. As stated in the publication (Elliot et al. 1994), the results provided 'the first good evidence that specific cytokine blockade can be effective in human inflammatory disease and define a new direction for the treatment of rheumatoid arthritis'.

Infliximab binds soluble and transmembrane TNFα resulting in a reduction of inflammatory cells, cellular adhesion, chemoattraction and tissue degradation associated with reduced concentrations of IL-6 and CRP in patients with RA. The reduction of IL-6 which follows rapidly after administration of infliximab was evidence that TNF gives rise to a cascade of cytokines. Larger clinical trials in patients with active RA on treatment with stable doses of methotrexate showed dose-related increases in ACR20 and ACR50, reduction of structural joint damage and improved physical function which were significantly superior to control subjects on methotrexate alone (Lipsky et al. 2000). To optimise the rate of attainment of therapeutic concentrations, doses were administered at 0, 2 and 6 weeks, and then maintenance was administered every 4 or 8 weeks.

In another placebo-controlled trial in methotrexate-naïve patients, infliximab in combination with methotrexate was shown to be superior to methotrexate alone. The benefit of adding the anti-TNF to methotrexate therapy in patients with persistently active RA was demonstrated in early controlled clinical trials of anti-TNFs with improvement of disability and arrest of progression of joint damage.

A review of TNF and the early work with infliximab are to be found in Monaco et al. (2014). In the years that followed the demonstrated efficacy of infliximab in RA, several other anti-TNFα biologics were developed (see Table 4). All of these agents have been shown to reduce joint destruction associated with downregulation of pro-inflammatory cytokines, reduced leukocyte trafficking, normalised T-cell function, decreased angiogenesis, improved blood counts, reduced platelets and fibrinogen and reduced cardiovascular risk associated with RA.

Etanercept, a human TNFα receptor p75 (R2) fusion protein, is produced by recombinant DNA technology in a Chinese hamster ovary (CHO) mammalian expression system. It is a dimer of a chimeric protein produced by fusion of the extracellular ligand binding domain of human TNF R2 to the Fc domain of human IgG1. Etanercept acts as a competitive inhibitor of TNF binding to its cell surface receptors. Clinical trials established the improvement in patients with RA compared with methotrexate alone (Bathon et al. 2000). It was in fact the first anti-TNF to be licensed for treatment of RA in the USA in 1998 and for polyarticular juvenile idiopathic arthritis the following year.

Adalimumab, the first fully human mAb approved by the FDA, was licensed for RA in 2002, and it has for several years been the most successful drug of all time in terms of sales worldwide. It is composed of two light kappa chains and two heavy IgG1 chains which bind to both TNF receptors p55 and p75 (R1 and R2). Its pharmacodynamic effects include reduction in the acute-phase reactants CRP and fibrinogen, ESR and concentrations of IL-1, IL-6, IL-8 and GM-CSF.

In addition to the effects on acute-phase reactants and cytokines, downstream effects of the TNF inhibitors include neutralisation of TNF-induced cell surface expression of various adhesion molecules such as E-selectin, vascular cell adhesion molecule (VCAM)-1 and intercellular adhesion molecule (ICAM)-1 by human endothelial cells. These effects result in inhibition of leukocyte migration. Golimumab has also been shown to inhibit matrix-metalloproteinase (MMP)-3 and vascular endothelial growth factor (VEGF).

Though the target of soluble and transmembrane cell surface TNF receptors is common to a number of these mAbs, it is important to recognise that there are significant differences between them with regard to structure and hence action. Thus, infliximab is chimeric, whereas adalimumab and golimumab are human which has implications for immunogenicity. Certolizumab pegol does not contain an Fc region and therefore does not fix complement or cause antibody-dependent cell-mediated cytotoxicity in vitro. It also does not induce apoptosis in human peripheral blood-derived monocytes or lymphocytes or neutrophil degranulation in vitro.

It is not appropriate to discuss the details of dosage here, but it is important to point out that there are some major differences between frequency and routes of administration. Infliximab is generally administered as an intravenous infusion, and, after loading doses, frequency of administration for maintenance is typically once every 8 weeks. Etanercept is administered once or twice weekly by subcutaneous injection. Adalimumab and certolizumab pegol are usually administered subcutaneously every 2 weeks and golimumab as monthly subcutaneous injections. To some extent the differences in frequency administration reflect the half-life of the mAb, but the activity of the disease, severity of adverse reactions, co-administered medications and increased clearance due to presence of antidrug antibodies (ADAs) are also factors. The production of ADAs is to some extent inhibited by immunosuppressants such as methotrexate, and it is usual for patients with RA to continue receiving methotrexate with the anti-TNF.

Adverse effects of anti-TNFs include opportunistic infections, sepsis, fungal infections and hepatitis B. Reactivation of TB is an interesting example of how a drug target (in this case TNF) can produce both adverse and beneficial effects. TNF is responsible for maintaining the integrity of TB granulomas and thereby preventing the escape of bacteria. Treatment with anti-TNFs appears to result in disruption of the granuloma and subsequent release of bacteria (Keane 2005). Melanoma and non-melanoma skin cancers have been reported in patients receiving anti-TNFs. However, skin cancers are common in the whole population, and it is not clear whether there is an excess of cases among patients treated with anti-TNFs.

6.1.2 Anti-IL-6 Cytokine Modulators

While the benefit of anti-TNFs has transformed the prognosis of RA for the majority of patients, it should be recognised that about 30% of RA patients do not respond, and some have less than satisfactory responses and/or relapse after a short time. This means there is an important role for other approaches including the targeting of other cytokines. IL-6 is a pleiotropic pro-inflammatory cytokine produced by a variety of cell types including T and B cells, monocytes and fibroblasts and is a product of

activated synovial macrophages and fibroblast-like synoviocytes in RA. The cytokine is involved in T-cell activation, induction of immunoglobulin secretion, induction of hepatic acute-phase protein synthesis and stimulation of haemopoiesis. IL-6-mediated signaling involves signal-transducing glycoprotein 130 and the signal transducer and activator of transcription-3 (STAT-3).

Two mAbs are currently available which target IL-6, tocilizumab and more recently sarilumab. Tocilizumab binds to both soluble and membrane-bound IL-6 receptors and inhibits signaling mediated by these receptors. Decreases in platelet count and neutrophils follow rapidly after dosing and, depending on dose, recover within a few days. The time to onset of the beneficial effects in RA is rapid with clear evidence of response frequently within 2 weeks. It is usually administered by intravenous infusion once every 4 weeks.

6.1.3 CD80 and CD86

Abatacept is a fusion protein of the extracellular domain of human CTLA4 linked to a modified Fc portion of IgG1. It acts as a selective immunosuppressant by inhibiting a key costimulatory signal required for full activation of T lymphocytes expressing CD28 by binding specifically to CD80 and CD86 molecules on the surface of antigen-presenting cells (APCs). Studies in vitro and in animal models demonstrate that abatacept modulates T-lymphocyte-dependent antibody responses and inflammation. T-lymphocyte activation as measured by decreased proliferation and cytokine production is attenuated with reduced production of TNFα, interferon-γ and IL-2.

Following abatacept treatment of patients with active RA, serum levels of soluble IL-2 receptor, a marker of T-lymphocyte activation, and IL-6, a product of activated synovial macrophages and fibroblast-like synoviocytes, were reduced (Weisman et al. 2006). Soluble E-selectin and intercellular adhesion molecule-1 were also significantly lower. In other studies serum levels of matrix metalloproteinase-3, which produces cartilage destruction and tissue remodelling, were decreased. Reductions in serum TNFα were also observed. The efficacy of abatacept has been demonstrated in patients with RA who have an inadequate response to anti-TNFα therapy (Genovese et al. 2005).

6.2 Anti-TNFs for Inflammatory Bowel Disease (IBD)

Crohn's disease is an IBD characterised by transmural inflammation that can affect any part of the bowel. The importance of TNF in contributing to the inflammatory process in Crohn's disease was recognised before anti-TNFα therapies became available. As early as 1992, it had been shown that there were dramatic increases in the excretion of TNF from the bowel in the stool of paediatric patients suffering exacerbation of symptoms of Crohn's disease and ulcerative colitis, compared with children with quiescent inflammatory bowel disease, diarrhoea from other causes or healthy controls (Braegger et al. 1992). The anti-TNFs are thought to be effective in Crohn's disease by inducing lysis of cells that express TNF in the presence of

complement. In animals, adalimumab has been shown to produce apoptosis of monocytes via caspase pathway activation.

One of the first reports of a mAb anti-TNF being used to treat Crohn's disease was with cA2, subsequently named infliximab. In 1997 a 12-week trial was conducted in 108 patients with moderate-to-severe Crohn's disease randomised to receive a single intravenous infusion of 5, 10 or 20 mg/kg cA2 or placebo (Targan et al. 1997). Response rates at 4 weeks ranged from 50 to 81% in the active treatment groups and 17% in the placebo group, and 33% went into remission compared a response rate of 17% in the patients who received placebo of whom just 4% went into remission. In a subsequent trial in 94 adult Crohn's disease patients with draining abdominal or perianal fistulae, infliximab 5 mg/kg at weeks 0, 2 and 6 produced a response rate of 68% (Present et al. 1999).

The benefit of using an induction regimen to suppress the disease followed by a maintenance regimen was subsequently confirmed (Sandborn and Hanauer 2002). An aggressive induction regimen is used initially to suppress the disease followed by a maintenance regimen. This is logical in light of the fact that exacerbations of IBD can be serious and debilitating with abdominal pain, chronic diarrhoea, and intestinal or rectal bleeding and may even be life-threatening. Hence there is a need to achieve therapeutic concentrations rapidly. The time taken to achieve steady state of any drug is determined by its half-life irrespective of the size or frequency of dosing, and the half-life of adalimumab is approximately 2 weeks; therefore it would take approximately 10 weeks to achieve steady-state concentrations. Similarly, with infliximab, which has a half-life of 9–10 days, it would take about 6 weeks to achieve steady state. This does not mean that it will take this long to elicit a therapeutic effect, but the time to onset of efficacy and the time to achieve optimal benefit will be much longer than if loading doses are used before switching to maintenance dosing. In this respect, use of mAbs is similar to that of small molecules such as corticosteroids, azathioprine and methotrexate, with which loading doses for induction have been used to induce remission of IBD for decades. However, for mAbs exhibiting TMDD and non-linear kinetics, the induction dose is also acting to saturate antigen targets. Following induction, the frequency of dosing and often the dose itself are reduced to maintain remission.

Infliximab was approved for induction and maintaining clinical remission of Crohn's disease in 2002, and it is particularly valuable to induce and maintain fistula healing. A Cochrane Review found that after induction, both infliximab and adalimumab maintain clinical remission, as well as having steroid-sparing effects (Behm and Bickston 2008). Certolizumab pegol was also found to maintain remission after induction. It is important to recognise that although all the anti-TNF mAbs share many properties, it is not appropriate to lump them together as equivalent. This is exemplified by etanercept, which has not been found to be effective in Crohn's disease. Many hypotheses have been made to explain its lack of efficacy, including that this might be related to its inability to induce apoptosis in the T cells of the lamina propria, the fact that etanercept blocks TNF-beta as well as alpha. TNF-beta is thought to regulate T-cell-dependent IgA production in the lamina propria and hence the microbiota composition.

6.3 Monoclonal Antibodies for Spondyloarthropathies

The spondylarthropathies (SpAs) are seronegative arthropathies (rheumatoid factor negative) frequently involving the axial skeleton and/or peripheral joints. They include ankylosing spondylitis, psoriatic arthritis and juvenile idiopathic arthritis but also have extra-articular manifestations such as those affecting the skin (psoriasis), eye (uveitis) and bowel (inflammatory bowel disease (IBD)). There is frequently a family history and many patients are positive for HLA-B27.

TNFα is one of the active inflammatory cytokines involved in these conditions, and anti-TNFs such as infliximab and adalimumab are licensed for treatment of SpA. Patients with both SpA and IBD generally respond to anti-TNFs, but new bone formation, which is a characteristic of ankylosing spondylitis, is not prevented by anti-TNFs. Interestingly, new bone formation is not seen in animal models involving soluble TNF overexpression but is a feature of an animal model overexpressing transmembrane TNF suggesting that the two forms of TNF play different roles in these SpAs.

Other inflammatory cytokines are also involved, in SpA. There is clear evidence of the role of the IL-23/IL-17 axis in the pathogenesis of SpA and IBD (Aggeletopoulou et al. 2018). Ustekinumab is a fully human IgG1kappa mAb that binds to the common p40 subunit of cytokines IL-12 and IL-23, thereby preventing p40 binding to the IL-12Rβ1 receptor protein, which is expressed on the surface of immune cells. IL-12 and IL-23 are secreted by activated APCs such as macrophages and dendritic cells. IL-12 stimulates NK cells and drives the differentiation of CD4+ T cells towards the Th1 phenotype, while IL-23 induces the Th17 pathway.

Ustekinumab has proven effective for Crohn's disease (Feagan et al. 2016) and for psoriasis and psoriatic arthritis but does not appear to be beneficial for ankylosing spondylitis. It seems that the pathogenesis of ankylosing spondylitis is different from that of psoriatic arthritis. Secukinumab is another IL-17 mAb, but unlike ustekinumab it has shown convincing efficacy in ankylosing spondylitis and conversely has not shown any benefit in Crohn's disease.

It should be noted that cases of new or exacerbations of existing IBD have been reported with use of certain mAbs such as secukinumab and ixekizumab.

6.3.1 Psoriasis and Psoriatic Arthropathy

Though psoriasis is classified as a spondyloarthropathy, it merits separate consideration. It is a very common autoimmune skin disease affecting 2–4% of the population, and although there are various types, about 90% of affected patients have plaque psoriasis with patches of red, dry, scaly and itchy plaques. For many patients, the condition can be kept under control by use of topical therapies including emollients, corticosteroid ointments, vitamin D analogues, coal tar, dithranol, topical immunosuppressants such as tacrolimus and phototherapy. However, about 20% of patients have more severe disease with or without arthritis. The arthritis does not resemble RA and is seronegative. It mainly affects the digital joints of hands and feet accompanied by dactylitis (painful swelling of the digits). The nails are also commonly affected. Immunosuppression with ciclosporin or methotrexate and the

retinoid acitretin have traditionally been used for these patients, but anti-TNFs such as adalimumab have proven effective for both plaque psoriasis and psoriatic arthritis.

In the last few years, additional cytokine targets have been identified. Several activate dendritic cells which then produce IlL-12 and IL-23. IL-23 binds to naïve T cells which then differentiate into Th17 cells. The differentiated cells secrete IL-17A, which is present in elevated concentration in blood of patients with psoriasis, the psoriatic plaques in the skin and the inflamed joints of patients with psoriatic arthritis. The cytokine binds to the IL-17 receptor (IL-17RA) on the cell surface and promotes keratinocyte proliferation and activation.

While the downstream effect is to reduce the activity on this cell surface receptor, the target binding of these newer interleukin inhibitors varies (Jeon et al. 2017). Thus, ustekinumab targets the p40 subunit of IL-12 and IL-23 preventing its binding to the receptor protein IL-12R. Secukinumab binds to the receptor IL-17RA and blocks the pro-inflammatory cytokines IL-17A, IL-17F, IL17A/F heterodimer and IL-25 whose genes are overexpressed in psoriatic plaques. The production of downstream mediators such as IL-6, GRO-alpha and G-CSF is also reduced. Brodalumab targets the IL-17A receptor (IL-17RA) on the cell surface. Guselkumab binds selectively to the regulatory cytokine IL-23 which affects the differentiation, expansion and survival of T-cell subsets and innate immune cell subsets which produce cytokines. These targets and activities are summarised in Table 5.

The skin response of skin lesions in plaque psoriasis is measured as the psoriasis area severity index (PASI). With the advent of these new cytokine modulators, the response rates have increased so that a substantial proportion of patients are responding with a 75% (PASI 75), 90% or even 100% reduction of the area affected within a few weeks or months of starting treatment. Response rates are dose related, increase over the first 6 months of treatment and are well maintained in the majority of patients at 1 year though relapses do occur. In some trials a PASI 90 is achieved in

Table 5 Monoclonal antibodies for the treatment of Psoriasis – targets and mechanism of action

mAb		Target	Mechanism of action
Ustekinumab	Human IgG1	p40 common subunit of IL-12 and IL-23	Prevents p40 binding to IL-12Rβ1 receptor protein IL-12R
Brodalumab	Human IgG2	IL-17RA on cell surface	Blocks activity of IL-17A, IL-17F, IL-17A/F heterodimer, IL-25 and associated induction of IL-6, GROα[a], G-CSF[b]
Secukinumab	Humanised IgG1	IL-17A	Blocks activity of IL-17A
Ixekizumab	Humanised IgG4	IL-17A and IL-17A/F	Blocks activity of IL-17A and IL-17A/F heterodimer
Guselkumab	Human IgG1	IL-23 p19	Reduces production of IL-17A, IL-17F, IL-22

[a]Growth-regulated oncogene-α now known as CXC ligand 1-a cytokine secreted by human melanoma cells expressed by macrophages, neutrophils and epithelial cells
[b]Granulocyte colony-stimulating factor

about 75% patients. Though not as impressive as the skin response, results with psoriatic arthritis are also positive with an ACR20 response rate typically over 50% by 12 weeks and >60% in the first year with clear evidence of radiological improvement.

7 Monoclonal Antibodies for the Treatment of Solid Tumours

In the same period of about 25 years that mAbs have become the mainstay of treatment for patients with severe inflammatory conditions, mAbs have also come to play an ever-increasing role in the treatment of malignancies, reflecting the ongoing molecular elucidation of oncological disease. Biologics now have a role in the treatment of solid tumours, leukaemias, lymphomas and myeloma. In this section, the discussion will concentrate on the use of some mAbs for treatment of solid tumours.

Until the recent advent of immune checkpoint inhibitors, the use of mAbs in oncology was mainly either to interfere with pathways essential for tumour progression and growth by blocking cell surface receptors (e.g. trastuzumab and cetuximab) or to act as an immune effector to direct Fc-mediated immunological activity against cells expressing a specific cell surface target (e.g. rituximab).

Examples of inhibition of blockade of cell surface receptors to inhibit tumorigenesis are trastuzumab and cetuximab which block epidermal growth factors that cause dimerisation of malignant cell surface receptors and the subsequent downstream events that promote tumour cell growth. Trastuzumab binds to human epidermal receptor 2 (HER2) overexpressed in 20–25% of breast cancers and 15–20% of gastric and gastroesophageal cancers. Cetuximab binds to epidermal growth factor receptor 1 (EGFR), which is overexpressed by many colorectal cancers and some squamous cell carcinomas of the head and neck.

7.1 Binding of Epidermal Growth Factors

Human epidermal growth factors (EGFs) drive tumorigenesis by regulating cell growth, survival and differentiation (Roskoski 2014). When EGFs bind to the extracellular domain of human epidermal growth factor receptors, dimerisation occurs. This results in autophosphorylation of specific tyrosine residues within the cytoplasmic kinase domain of the activated receptor, which initiates downstream signaling pathways such as mitogen-activated protein (MAP) kinase and phosphoinositide-3kinase (PI3K) that stimulate cell growth.

7.1.1 Trastuzumab

Trastuzumab is a humanised IgG1 mAb, which binds to the epidermal growth factor receptor, HER2, resulting in downregulation of the PI3K/Akt pathway (Nami et al. 2018) and G1 phase cell cycle arrest (Hudis 2007). It is indicated for breast,

gastric and gastroesophageal adenocarcinomas that overexpress HER2. Homo- and heterodimerisation can be prevented by binding of trastuzumab to an extracellular juxta-membrane region (domain IV). The resultant inhibition of downstream signaling causes growth arrest and apoptosis. Other pharmacodynamic properties of trastuzumab include ADCC and inhibition of angiogenesis; however, the main mechanism of action responsible for clinical efficacy is thought to be inhibition of receptor signal transduction (Capelan et al. 2013).

Following initial breakthrough approval by the FDA in 1998 (Blackwell et al. 2018), trastuzumab is currently approved as monotherapy and, in combination with chemotherapy with good response rates, prolongation of progression free survival and overall survival. It is administered by s.c. or i.v. routes, usually on a 3-week cycle. Apart from infusion-related reactions common to many mAbs (see Sect. 8), trastuzumab can produce cardiac dysfunction with congestive heart failure by mechanisms as yet unclear (Nemeth et al. 2017). Patients who have received taxanes and anthracyclines such as doxorubicin or epirubicin are particularly susceptible to this adverse effect which can be severe. All patients treated with trastuzumab should have cardiac function monitored during and after treatment.

Before treatment with trastuzumab can be initiated, evidence of HER2 positivity must be provided by testing the patient's tumour tissue using immunohistochemistry and/or HER2 gene amplification by fluorescent or chromogenic in situ hybridisation (FISH/CISH) (Krishnamurti and Silverman 2014). Before the advent of trastuzumab, the prognosis in terms of progression-free survival (PFS) and overall survival (OS) was worse than for most patients with breast cancer (Ross et al. 2009).

7.1.2 Pertuzumab

Pertuzumab is a humanised IgG1 mAb that also targets the extracellular dimerisation domain of HER2, preventing heterodimerisation with HER3, a requisite step for intracellular signal transduction pathways including PI3K/Akt (Nami et al. 2018). It is pharmacologically complementary to trastuzumab as together they achieve a more complete blockade of HER2-mediated signaling than either agent used alone (Gerratana et al. 2017). Pertuzumab is approved for treatment of breast cancer with HER2 overexpression in combination with trastuzumab and chemotherapy (Baselga et al. 2012) as neoadjuvant and adjuvant treatment and for metastatic breast cancer. In a Phase 3 trial, the response rate and overall and progression-free survival were all greater in patients receiving the combination of pertuzumab and trastuzumab with docetaxel than in those receiving placebo and trastuzumab with docetaxel (Swain et al. 2013).

Importantly, in a Phase 3 trial in HER2-positive metastatic gastric or gastroesophageal junction cancer, the combination of pertuzumab and trastuzumab with chemotherapy did not achieve a significant increase in overall survival compared with placebo and trastuzumab with chemotherapy (Tabernero et al. 2018). The apparent discrepancy between results with the combination in breast and gastric cancers indicates that improved patient outcomes with combinations of mAbs cannot be assumed despite common molecular pathologies.

7.1.3 Trastuzumab Emtansine

Trastuzumab emtansine is an antibody-drug conjugate (ADC) developed to enhance the efficacy of trastuzumab (Peddi and Hurvitz 2013). Trastuzumab is combined with DM1, a cytotoxic maytansinoid, via a hetero-bifunctional cross-linker. Trastuzumab emtansine is cleaved in lysosomes following drug-receptor internalisation, and resultant DM1 catabolites bind to tubulin, suppressing microtubule dynamic instability and leading to mitotic arrest (Oroudjev et al. 2010). Trials in breast cancer associated with HER2 overexpression have demonstrated benefit over and above standard trastuzumab therapy alone (von Minckwitz et al. 2019).

7.1.4 Cetuximab

Cetuximab is a recombinant chimeric IgG1 mAb which targets the epidermal growth factor receptor (EGFR or HER1) (Messersmith and Ahnen 2008). It is approved for treatment of squamous cell cancer of the head and neck (HNSCC) of which more than 90% of tumours express EGFR and for RAS wild-type metastatic colorectal cancer. It blocks binding of endogenous EGFR ligands, causes internalisation of the receptor and mediates ADCC (Lenz 2007). It also reduces tumour neovascularisation by inhibiting expression of angiogenic factors by tumour cells.

7.1.5 Panitumumab

Panitumumab is a fully human IgG2 mAb specific for EGFR and is approved for RAS wild-type metastatic colorectal cancer (Keating 2010). The outcome of late phase trials demonstrating non-inferiority as compared with cetuximab is noteworthy as panitumumab is an IgG2 mAb and thus not expected to trigger significant ADCC (Price et al. 2014).

7.2 Inhibition of Vascular Endothelial Growth Factor

Another approach to preventing tumorigenesis is to inhibit the growth of its blood supply (vasculogenesis and angiogenesis), and this can be achieved by inhibition of vascular endothelial growth factor (VEGF). Bevacizumab is a VEGF inhibitor first approved for colorectal cancer in 2004 with subsequent approval for treatment of several other tumour types in combination with chemotherapy. It is a humanised IgG1 which binds to VEGF and thereby prevents it binding to its receptors VEGFR-1 and VEGFR-2 on the surface of endothelial cells. Regression of tumour vascularisation as well as inhibition of new tumour vasculature results in tumour regression.

7.3 CD20 Antigen and Fc-Mediated B-Cell Lysis

Rituximab was first approved in 1997 for treatment of low-grade B-cell lymphoma (Maloney et al. 1997). It is a chimeric mouse/human mAb with human IgG1 constant regions and murine light and heavy chain variable region sequences. Its Fab domain

binds to the transmembrane antigen CD20 located on normal and malignant pre-B and mature B lymphocytes (Maloney 2012). The Fc domain can mediate B-cell lysis by recruitment of immune effector functions such as complement-dependent and antibody-dependent cellular cytotoxicity (CDC and ADCC, respectively). Its currently approved indications for malignancies are chronic lymphocytic leukaemia and non-Hodgkin's follicular lymphoma (Salles et al. 2017).

7.4 Immune Checkpoint Inhibition

More recently mAbs have been developed as immune checkpoint inhibitors, exploiting their ability to block protein-protein interactions. The approval of ipilimumab for advanced melanoma in 2011 (Cameron et al. 2011) and the subsequent and phenomenal success of immune checkpoint inhibitors targeting the programmed cell death-1 receptor/ligand system (PD-1/PD-L1) have firmly established cancer immunology as a therapeutic target for mAbs (Constantinidou et al. 2019). In doing so, they have led to the acceptance that malignant disease does not solely result from cell cycle control breakdown but also from failure of the immune system to recognise and remove abnormal cells.

Alongside broad advances in deciphering the workings of the immune system, the extent of tumour infiltration with lymphocytes was noted to correlate with improved prognosis among certain groups of cancer patients (Wirth and Kühnel 2017), suggesting that the immune system in general and T lymphocytes in particular can mitigate carcinogenesis. More direct confirmation that T cells can mediate anti-tumour responses was objective responses observed in melanoma patients following administration of autologous tumour infiltrating lymphocytes (TILs) (in combination with T effector cell promoting interleukin-2 therapy) (Schwartzentruber et al. 1994).

7.4.1 T Lymphocytes and Immune Checkpoints

T lymphocytes, the main effector cell of the adaptive immune system, are the main target for immunotherapy. They may be considered either T effector cells or T regulatory cells based upon the inverse nature of their function and differing affinities for self-antigens (Kumar et al. 2018). During development of the immune system, T-cell progenitor clones expressing T-cell receptors with high affinity for self-antigens are subject to negative selection. As negative selection is imperfect, clones with intermediate affinity for self-antigens develop into T regulatory cells, acquiring Fox3p expression, and serve to maintain peripheral self-tolerance (Josefowicz et al. 2012). Both CD4 (helper) and CD8 (cytotoxic) subtypes of T effector cells contribute to antigen-specific elimination of microbes and tumour cells (Groom 2019). Furthermore, certain T effector lineages are capable of transformation into T regulatory cells (Richards et al. 2015). Physiologically this is a mechanism to dampen down excessive immune responses and mitigate the risk of collateral tissue damage which may be exploited by tumour cells to prevent the establishment of an effective anticancer immune response (Kumar et al. 2018).

T-cell activation and proliferation are initiated following a complex series of signaling events at the immune synapse. In addition to the T-cell receptor specifically binding MHC-bound antigenic peptide fragments, there exist an array of co-signaling receptors on T cells, which, upon binding with ligands on target- or antigen-presenting cells (APCs), serve to impart a stimulatory or inhibitor influence on the T cell. Receptor/ligand systems with an inhibitory effect have been termed immune checkpoints (Chen and Flies 2013). Integration of the various stimulatory and inhibitory co-signaling events sets the context of TCR binding with a specific antigen and determines the onward fate of the T cell.

7.4.2 CTLA-4

The CD28/CTLA4-B7-1/B7-2 co-signaling system was one of the earliest systems to be described, and interest in CTLA4 as an immune checkpoint led to the development of ipilimumab, the first immune checkpoint inhibitor (Sondak et al. 2011). Ipilimumab is a fully human anti-CTLA-4 IgG1kappa mAb produced in CHO cells by recombinant DNA technology. It is indicated for advanced melanoma alone and in combination with another ICI nivolumab for advanced melanoma and renal cell carcinoma. When administered alone, it can produce durable responses in about 20% of patients with advanced melanoma (Hodi et al. 2010). Despite this comparatively modest effect by today's standards and a high incidence of mainly immune-related severe and serious adverse effects, the efficacy of ipilimumab represented a breakthrough in treatment of melanoma and established the principle of beneficial checkpoint inhibition.

It has been firmly established that tumours develop local escape mechanisms to subvert the immune system and prevent an effective immune response from being established (Chen and Han 2015). Such mechanisms have been collectively labelled 'immune evasion' (Vinay et al. 2015) and likely explain why TILs frequently exhibit an exhausted phenotype (Muenst et al. 2016). They also explain the dichotomy of progressive disease despite apparently normal and appropriate systemic anti-tumour responses (Rosenberg et al. 2005) resembling the findings in many anticancer vaccine trials (Melero et al. 2014). Immune evasion may occur at multiple levels of both the innate and adaptive immune systems, including:

- Defects in antigen presentation
- Defects in translocation of immune cells between periphery and the tumour microenvironment
- Defects in immune cell activation or proliferation
- Excessive T regulatory cell activity
- Immunosuppressed TILs due to suppression by molecular pathways

While the identification of mechanisms contributing to immune evasion is encouraging, immunotherapy aimed to counter immune evasion puts the patient at risk of immune-related adverse effects. Even when ipilimumab received approval, the objective tumour response (OR) observed in 15–20% treated patients occurred in the context of severe toxicity in approximately 30% (Hodi et al. 2010). Such an

unfavourable efficacy/toxicity profile may explain in part why it has never expanded its indications as a single agent. Intriguingly, despite the interest in CTLA4 as a therapeutic target, there exists no reliable evidence that it is upregulated in cancerous disease (Sanmamed and Chen 2018).

7.4.3 Programmed Cell Death-1 Receptor-Ligand System

In contrast to CTLA4, there is ample evidence that the programmed cell death-1 receptor (PD-1) and its ligand1 (PD-L1) immune checkpoint system are 'hijacked' to effect localised TME-specific immune evasion in patients with otherwise intact systemic immune systems (Ribas and Wolchok 2018). This phenomenon is termed 'adaptive immune resistance' (Taube et al. 2012). Elevated expression of both PD-1 and PD-L1 has been widely demonstrated in both haematological (Annibali et al. 2018) and solid organ malignancies (Gatalica et al. 2014). With regard to the latter, increased PD-L1 expression has been observed on both tumour and tumour-infiltrating myeloid cells including myeloid-derived suppressor cells (MDSCs) (Noman et al. 2014) and tumour-associated macrophages (TAMs), in addition to evidence of upregulation of the PD-1/PD-L1 system regulatory-type TILs (Belai 2014).

PD-1 was first described in 1992 (Ishida et al. 1992). It is a 288 amino acid protein belonging to the immunoglobulin superfamily of co-signaling receptors and exists as a monomer composed of four domains: (1) single extracellular N-terminal domain, (2) 20 amino acid linker, (3) transmembrane domain and (4) a cytoplasmic domain with two tyrosine-based signal motifs (Zhang et al. 2004). The ligand PD-L1 (initially termed B7-H1) was discovered in 2000 (Freeman et al. 2000) and is a type 1 glycoprotein with immunoglobulin domains. PD-L2, the second known ligand for PD-1, was discovered shortly after PD-L1 and shares a degree of homology, but its function and significance of PD-L2 remain unclear (Latchman et al. 2001).

Immune checkpoint inhibitors (ICIs) targeting the PD-1/PD-L1 system have achieved phenomenal success in recent years. They are indicated for treatment of an ever-expanding number of malignancies, including those not previously considered immunogenic. The success of PD-1/PD-L1 ICIs has spawned the 'masterswitch' concept whereby correction of a single molecular defect/aberration is sufficient to restore immune effector functionality in a previously immunosuppressive tumour microenvironment and has provided proof that it is not necessary to correct all immunological defects (Sanmamed and Chen 2018). In contrast to enhancement immunotherapy with ipilimumab, anti-PD-1/PD-L1 immunotherapy represents normalisation since the system is pathologically upregulated in malignancy (Muenst et al. 2016). Owing to the PD-1/PD-L1 system being largely inactive under normal conditions, tumour-specific PD-L1 expression means that PD-1/PD-L1 ICIs offer the promise of effecting a localised and specific immune response with reduced off-target toxicity (Sanmamed and Chen 2014). Furthermore, PD-1/PD-L1 ICIs are associated with durable responses due to the establishment of immunological memory (Sanmamed and Chen 2018).

Nivolumab is a human IgG4 mAb that blocks binding of the PD-1 receptor with PD-L1 (and PD-L2) (Weber et al. 2017). Approved indications include melanoma, non-small-cell lung cancer (NSCLC), squamous cell cancer of the head and neck (HNSCC), advanced renal cell carcinoma (RCC), advanced urothelial carcinoma, Hodgkin's lymphoma, hepatocellular carcinoma (HCC) and micro-satellite high (MSI-H) and deficient mismatch-repair (dMMR) colorectal carcinoma. Pembrolizumab is also an anti-PD1 human IgG4 mAb with similar indications to those of nivolumab though there are some differences (McDermott and Jimeno 2015). For example, certain approvals for pembrolizumab are contingent on PD-L1 expression levels. At the time of writing, it is approved for treatment of melanoma, NSCLC, HNSCC, Hodgkin's lymphoma, urothelial carcinoma, MSI-H/dMMR solid organ cancers, gastric and gastro-oesophageal carcinoma and cervical carcinoma. Atezolizumab an Fc-engineered, humanised IgG1; avelumab, a human IgG1; and durvalumab a human IgG1 kappa all bind to PD-L1 preventing its suppression of cytotoxic CD8 T cells by PD-L1. Most recently cemiplimab, an IgG4 that prevents binding of ligands PD-L1 and PD-L2 to PD-1, recently gained (FDA) approval for advanced cutaneous squamous cell carcinoma, a condition for which disease risk is strongly associated with immunosuppression and for which there is no other approved systemic treatment (Migden et al. 2018).

In keeping with the hypotheses that spurred their development, the toxicity profiles of PD-1/PD-L1 ICIs compare favourably with ipilimumab and standard chemotherapy (Nishijima et al. 2017). Indeed, in a direct comparative randomised clinical trial of nivolumab and ipilimumab in advanced melanoma, the latter resulted in a threefold increase in toxicity (Weber et al. 2017). However, PD-1/PD-L1 ICIs are associated with a number of mechanism-based and idiosyncratic toxicities with an unpredictable relationship to dosing (Haanen et al. 2017). These include rash, gastrointestinal inflammation (including colitis), thyroid dysfunction and pneumonitis. Immune-related adverse events (irAEs) appear as an inevitable trade-off with any ICI, and thus the observation that steroids (and similar immunosuppressive agents) seemingly do not compromise ICI efficacy is encouraging (Constantinidou et al. 2019).

It also becomes clear that PD-1/PD-L1 ICIs are not efficacious in all patients treated and that while striking and durable outcomes result, they only do so in a small cohort of patients suggestive of innate immunotherapy resistance. Furthermore, although responses are durable, acquired resistance to ICIs does develop in time (O'Donnell et al. 2018). Given the considerable cost plus unpredictable and potentially serious toxicity, it is imperative to understand better the optimal molecular context for deployment. To date, PD-L1 expression is the only approved predictive biomarker for PD-1/PD-L1 ICIs, but differences in the type of assay used, the type and handling of tissue, how 'positivity' is defined in the context of a continuously distributed biological phenomenon and the dynamic nature of PD-L1 expression all serve to muddy the waters and make confident reproduction challenging (Hirsch et al. 2017). As a result, only a few approvals are contingent on PD-L1 testing. It is also unclear whether these mAbs should be used alone or in combination with other agents and, if used in combination, what is the preferred sequence.

7.5 Antigenic Targets

Immune checkpoint inhibitors have served to demonstrate unequivocally that the tumour microenvironment does indeed contain antigenic targets against which an effective and specific immune response can be mounted. Cancer-associated antigens have long since attracted interest as vaccination candidates, although cancer vaccines have hitherto been underwhelming (Sanmamed and Chen 2018). The advent of ICIs may lead to a change in fortune as vaccination trials have repeatedly effected specific and appropriate systemic immune responses (Melero et al. 2014) but may require reversal of local immunosuppression by agents such as PD-1/PD-L1 ICIs in order to be effective.

Antigens capable of eliciting an effective immune response may be either tumour-associated (TAA) or tumour-specific (TSA) antigens. TAAs have frequently been targeted for immunotherapy and include those overexpressed in tumour tissue (e.g. HER2 in breast cancer, RAGE-1 in renal cell carcinoma, EGFR in epithelial cancers) and those with spatiotemporally restricted expression (e.g. cancer testis antigens such as MAGE-A1 (van der Bruggen et al. 1991) and NY-ESO-1 (Robbins et al. 2015)). As antigenic targets, TAAs are likely subject to central immunotolerance that may limit effective immune responses and coupled with the risk of collateral off-target damage (Wirth and Kühnel 2017), they have become less attractive as targets for immunotherapy. Indeed, while early phase trial with autologous T cells engineered to express NY-ESO-1 TCRs demonstrated some evidence of efficacy, significant 'off-target' toxicity was observed. TSAs are antigens not normally encoded for in the genome and include viral oncogenic proteins (e.g. SV40 from Epstein-Barr virus, E6/E7 from human papilloma virus) and neoantigens. Neoantigens result from mutations acquired during carcinogenesis and immunoediting, and their presence is associated with ICI success.

7.6 Immune Checkpoint Inhibitor Toxicity

Immune-mediated adverse events related to checkpoint inhibitors can be generalised such as fatigue, nausea and loss of appetite, but symptoms may be due to inflammatory effects on almost any organ in the body. Thus pneumonitis is likely to cause cough and dyspnoea; colitis will produce abdominal pain, diarrhoea and often gastrointestinal bleeding; effects on the skin are rashes and loss of pigmentation (vitiligo); and a variety of other symptoms and signs may result from hepatitis, myocarditis, vasculitis, arthritis, encephalitis, pituitary hypophysitis, thyroiditis and myasthenia gravis.

8 Adverse Effects and Limitations of Monoclonal Antibodies

While the specificity of mAbs largely negates off-target actions, mAbs are not the 'magic bullet' heralded by numerous early observers. A variety of adverse effects are associated with mAbs, often related to the intended pharmacodynamic action and frequently 'immune-related'. The latter include those resulting from immunosuppression, immunostimulation, autoimmunity and hypersensitivity. In addition, mAbs are also associated with a number of organ-specific toxicities of which relevant examples are provided in previous sections. The toxicity of mAbs presents clinical challenges as the half-life is generally much longer than that of small molecules, and hence the duration of adverse events may be prolonged. In addition, adverse reactions are often poorly predicted by preclinical studies.

Adverse reactions to mAbs are common and can be very serious. Attempts have been made to classify adverse events arising from mAb treatment, including complex classification schemes that consider the mechanism of action and structure of mAbs (Lee and Kavanaugh 2005). A simpler scheme (Pichler 2006) proposed to distinguish adverse events that resulted target engagement (i.e. those resulting from exaggerated on-target pharmacology) from agent-related adverse events (i.e. directly related to the innate properties of the mAb). Adverse reactions resulting from target engagement include infusion-related reactions (IRRs) and cytokine release syndrome (CRS) and, depending on any immunomodulatory outcomes, may also include infection, malignancy and autoimmune disease. Those related to the innate properties of mAbs include hypersensitivity reactions such as anaphylaxis and infusion-related reactions (See Sect. 8.1). Regardless of their classification, all adverse effects resulting from mAbs can be considered in the context of unique interaction between the host immune system, disease state, pharmacological target and the mAb (Sathish et al. 2013). For example, pro-malignant chronic inflammatory conditions may contribute to increase the risk of malignancy following immunosuppression resulting from the action of immunosuppressive mAbs.

8.1 Infusion-Related Reactions Including Cytokine Release Syndrome

Infusion-related reactions refer to acute reactions experienced during or following administration of a mAb, typically occurring with 2 h of administration, but the onset can sometimes be considerably later (Hansel et al. 2010). There are various causes of IRRs, but in broad terms they may be attributable either to the intended pharmacological activity of the mAb or to the inherent properties of the mAb including the potential for immunogenicity and stimulation of a specific host immune response.

The exact mechanisms of non-hypersensitive IRRs are not completely understood but are likely to involve Fc domain-mediated engagement of downstream immune pathways and the release of pro-inflammatory cytokines (e.g. IL-6, TNFα, IFNγ, etc.) (Descotes 2009). Additionally, activation of complement cascades may occur and are common, for example, with rituximab (van der Kolk et al. 2001). The

resultant clinical sequelae form a spectrum from local injection site reactions and flu-like symptoms through to cytokine release syndrome (CRS) of variable severity.

Mild CRS is frequently apparent as fever, nausea, headache, fatigue, rash, tachycardia and dyspnoea. More severe forms of the syndrome include muscle and joint pains, rigours, severe headache, vomiting, diarrhoea, dyspnoea, tachypnoea, hypotension, seizures, confusion and psychotic symptoms. Investigations may reveal hypoxia, abnormal cardiac function, disturbances of liver and renal function and coagulation. In its most severe form, it is known as 'cytokine storm'. The potential for multiple organ failure due to cytokine storm was demonstrated in the 2006 phase 1 trial of the CD28 'superagonist', TGN1412 (Hünig 2012). Monoclonal antibodies recognised to be associated with frequent and/or serious IRRs carry black box warnings.

8.1.1 Hypersensitivity Reactions

Immediate type 1 hypersensitivity reactions, including anaphylaxis, are mediated by mAb-specific IgE. Such reactions do not usually occur on first administration of a mAb as prior immune sensitisation is required, but there have been instances presumed to result from mAb epitope cross-reactivity to pre-existing host IgE. Anaphylactoid reactions, leading to symptoms clinically indistinguishable from type 1 hypersensitivity, can occur and are due to mast cell degranulation via non-IgE-mediated pathways (Doessegger and Banholzer 2015). When they occur, type 1 hypersensitivity reactions are likely to do so with rapid onset, within a few minutes of the start of administration of an infusion (Baldo 2013).

Other hypersensitivity reactions include angioedema, urticaria and serum sickness with arthralgias, myalgias, rash, pruritus, fever and lymphadenopathy (Descotes 2009). The onset of serum sickness is typically 1–5 days after administration of the first or subsequent injections. Omalizumab, a mAb indicated for prevention of exacerbations of severe allergic asthma and chronic spontaneous urticaria, is designed to bind to free and membrane-associated IgE except that bound to IgE Fc domain receptors on mast cells (Chang 2006). Consequently, it does not cause mast cell degranulation or anaphylaxis, but it can produce serum sickness with delayed type III reactions. These are thought to be due to immune complex formation and deposition resulting from development of antibodies against omalizumab. Serum sickness has also been recorded with rituximab and infliximab (Riegert-Johnson et al. 2002).

8.1.2 Nonimmune IRRs

Nonimmune IRRs do not involve IgE and, unlike type 1 hypersensitivity reactions, frequently occur on first administration and typically decline with repeat dosing. Rituximab administered i.v. or s.c. is associated with IRR in approximately 77% of patients receiving an initial infusion (10% severe), but the incidence declines with subsequent infusions (Roselló et al. 2017). The incidence with the HER2 mAb trastuzumab is approximately 40% on first administration, and that with the EGFR mAb cetuximab is about 15%, again with the incidence declining with subsequent administrations (Baldo 2013).

8.2 Antidrug Antibodies

All mAbs represent foreign antigenic and thus 'immunogenic' material capable of eliciting a specific T and/or B lymphocyte-mediated host immune response, and several mAbs are subject to black box warnings regarding the potential for adverse events related to such immune responses (Sathish et al. 2013). Although the transition from murine and chimeric mAbs towards humanised and fully human products has reduced immunogenicity, it is too simplistic to view the latter as simply being a function of the percentage homology (Clark 2000). Specific amino acid residues at certain positions can greatly and disproportionately impact upon immunogenicity. In addition, glycosylation residues can contribute to immunogenicity; acute anaphylactic reactions to cetuximab have been attributed to the formation of IgE antidrug antibodies (ADAs) against glycosylation residues on cetuximab (Chung et al. 2008). Regulatory agencies have issued formal guidelines for the assessment of immunogenicity during development (European Medicines Agency 2012), and various methods are employed to reduce immunogenicity of mAbs. The relevance of ADAs is highly variable and can include reduced efficacy, altered pharmacokinetics and infusion reactions including type 1 hypersensitivity and anaphylaxis. Depending on the steric consequences of ADA to mAb binding, ADAs may be classed as either neutralising or non-neutralising ADAs with respect to interaction of the mAb with its intended target though both may modulate pharmacokinetics, for example, by enhancing clearance. Persistent neutralising antibodies against natalizumab are associated with both reduced efficacy and incidence of IRRs (Cohen et al. 2008).

8.3 Tumour Lysis Syndrome

Tumour lysis syndrome is a combination of metabolic and electrolyte abnormalities that occur in patients with cancer often with a high tumour burden. It is characterised by excessive cell lysis resulting in hyperuricaemia, hyperphosphataemia, hyperkalaemia and hypocalcaemia. It is most common with lymphomas and leukaemias and other haematological malignancies but can also occur with solid tumours.

8.4 Neurotoxicity

Neurotoxicity is not common with mAbs but there have been some notable cases. In some cases, neurotoxicity may be related to immunosuppression and subsequent reactivation of dormant viruses. Natalizumab is a humanised mAb that is an antagonist of the cell adhesion molecule α4 integrin which prevents leukocyte trafficking in the CNS (Hutchinson 2007). It is a highly effective treatment for relapsing remitting multiple sclerosis (MS). It was licensed in 2004–2005 but was temporarily withdrawn by its manufacturer because of three cases of progressive multifocal leukoencephalopathy, a brain infection caused by a polyomavirus called John

Cunningham virus (JCV) (Saribas et al. 2010). JCV is a common virus which can persist without causing symptoms for many years, but natalizumab can cause reactivation of the virus resulting in a devastating, often multifocal fatal brain infection, particularly when combined with other immunosuppressant drugs such as interferon β1a. It was withdrawn in 2006 but was soon reintroduced partly due to pressure from the patient population, many of whom felt the benefit outweighed the risk (Singer 2017). The number of cases of PML has increased dramatically in recent years (Ho et al. 2017), but the mAb remains on the market in the USA and Europe for the treatment of relapsing-remitting MS and also Crohn's disease.

Neurotoxicity has also been observed frequently with the CD19-CD3 bispecific mAb blinatumomab indicated for treatment of Philadelphia chromosome-negative CD19-positive B-precursor acute lymphoblastic leukaemia (Jain and Litzow 2018). Encephalopathy, seizures, speech disorders, impaired consciousness and other neurological signs have been observed frequently and may be very serious. The toxicity may be related to the CD19 target.

Daclizumab was licensed for treatment of relapsing-remitting MS in 2016, but its use was soon restricted due to liver toxicity, and it was withdrawn from the market in March 2018 after 12 patients worldwide were reported to have developed serious inflammatory brain disorders including encephalitis and meningoencephalitis resulting in death of three patients (Lancet 2018).

Alemtuzumab is a mAb that binds to CD52, thereby causing destruction of lymphocytes (Jones and Coles 2014). When the humanised form was synthesised in the 1990s, it was called Campath-1H and was initially trialled in patients with non-Hodgkin's lymphomas. After several cycles of evaluation, it was eventually licensed for B-cell chronic lymphocytic leukaemia and relapsing-remitting MS. However, it is known to cause a number of adverse effects, including opportunistic infections often related to leucopenia and also stroke due to blood vessel damage in the brain (McCall 2019).

8.5 Other Limitations

Other limitations of mAbs relate to the fact that they must be administered parenterally, often intravenously. Subcutaneous injections frequently cause injection site reactions. They are also expensive, and their cost will remain high despite the advent of biosimilars. There will be no resemblance to the situation with small molecular entities, for which the price drops precipitously once the patent expires and the branded product can be replaced by bioequivalent generics. Antibodies are produced by biological manufacturing, and no two products are identical so that production remains expensive. A small change in glycosylation may be sufficient to produce a major change in the efficacy and/or safety profile so each biosimilar must undergo adequate testing to satisfy the regulatory authorities that it can be used as a 'biosimilar' in place of the original product. Prescribers and patients also need to be convinced.

9 The Future

9.1 Beyond mAbs: mAb-Based Biotherapeutics

There is a considerable effort in interrogating the clinical benefits of alternative mAb modalities, or mAb-based biologics, such as Fab domains, single-chain variable regions (scFv), bispecific antibodies and antibody drug conjugates (ADCs). While these will not be discussed in detail, a brief description of each and their potential clinical advantages are summarised.

Smaller units of mAbs, such as Fab fragments and scFvs, have been shown to have better tissue penetration, particularly in tumours and other relatively inaccessible tissues. With molecular weights of these smaller units often being less than one third those of conventional mAbs, access to antigen-binding sites may be enhanced (Grantab and Tannock 2012; Nelson 2010). Examples of fragments are ranibizumab, a humanised mAb fragment, and abciximab, a Fab fragment of the chimeric human-murine mAb 7E3. Ranibizumab is indicated for neovascular (wet) age-related macular degeneration and visual impairment due to choroidal neovascularisation, diabetic macular oedema and retinal vein occlusion. Abciximab inhibits platelet aggregation by binding to the integrin glycoprotein IIb/IIIA receptor, thereby preventing binding of fibrinogen and other large molecules to the IIb/IIIa receptor sites on activated platelets.

Bispecific (or bifunctional) mAbs are modalities in which two IgG chains with varying target binding specificity are fused into a single molecule, thereby facilitating two target antigens to be bound simultaneously by the same molecule. Bispecific molecules may afford pharmacodynamic synergy in their activity and/or increased convenience for patients by reducing the number of treatments required (Labrijn et al. 2019). For example, emicizumab binds two coagulation factors (factor IXa and factor X), acting in the place of activated factor VIII in the coagulation pathway and is used for prophylaxis against bleeding in some patients with haemophilia A (Labrijn et al. 2019). Another bispecific mAb, blinatumomab, interacts with CD3 on T cells and CD19, on precursor B-cell acute lymphoblastic leukaemia (ALL) cells, which is speculated to recruit cytotoxic T cells to kill the intended cell ALL population (Labrijn et al. 2019).

In the case of ADCs, mAbs are used to deliver a drug (typically small molecule called a payload) with increased specificity to a target tissue/cell type. The payload is conjugated to a mAb using approaches to prevent or slow the premature release of the payload into the systemic circulation and reduce the dose-limiting toxicities observed for the free payload while improving the efficacy of the unconjugated mAb. In this class of mAb-based biologics, brentuximab vedotin (Adcetris; CD30 targeted) and inotuzumab ozogamicin (Besponsa; CD22 targeted) have been approved for haematologic cancers, and one ADC, trastuzumab emtansine (Kadcyla; HER2 targeted), has been approved to treat breast cancer. In spite of the clear clinical benefits being demonstrated by these ADCs, the toxicity profiles are comparable with those of standard-of-care chemotherapeutics, with dose-limiting toxicities associated with the mechanism of activity of the cytotoxic warhead; thus additional opportunities exist for improving patient benefit with ADC modalities (Coats et al. 2019).

9.2 Altering the Drug-Ability

Protein engineering is being leveraged to optimise the drug-ability of properties of mAbs to broaden their therapeutic effects, as well as the patient experience. With regard to PK, evidence of improving mAb clearance to reduce the frequency of administration via enhancing FcRn interactions has been demonstrated clinically. For example, MEDI4893, a mAb which binds to alpha-toxin and contains a triple residue substitution (M252Y/S254T/T256E or YTE) within the Fc region that improves FcRn interaction, was estimated to have an elimination half-life of ~80 to ~112 days, which is ~fourfold longer than the systemic half-lives of other human IgGs (Yu et al. 2016). Along these lines, motavizumab-YTE showed an extended half-life of 70–100 days in healthy adults, which is also ~fourfold longer than that of wild-type motavizumab (Robbie 2013). The positive corroboration that a PK benefit can be manifested via enhanced FcRn interactions in humans for mAbs continues to leave a pillar for enhancing the drug-ability of next-generation mAb therapies.

Engineering the effector functions of mAbs is another trailblazing approach in which modifications in interactions with immune components allow vast opportunities to improve mAb efficacy and safety especially for mAbs directed at oncology and autoimmune disorders. Effector function modulation is predominantly interrogated through studies of complement-based or FcγR-based interactions of mAbs. There are also some examples of co-engagement of target antigens and FcγRs co-expressed on the same cells. For instance, a Toll-like receptor 4 (TLR4) mAb with Fc mutations that enhance binding to FcvRIIa showed increased potency compared to a mAb without the mutations in vitro (Loyau et al. 2014). The translation of the clinical significance of these types of engineering approaches remains an area that is mainly unexplored and of continued interrogation.

9.3 Autoimmune Disease

Due to FcRn protection of IgGs, the endogenous IgG half-life is typically 2–3 weeks. Reduction of the half-life can be achieved by blocking FcRn. This approach is being taken in an attempt to treat autoimmune diseases (Patel et al. 2011; Challa et al. 2013).

The checkpoint inhibitors used in oncology are antagonistic mAbs, and therefore the potential to reduce the activation of the immune system, via checkpoint agonism, is being actively pursued for the treatment of autoimmune diseases. A PD-1 agonist antibody is currently being tested in psoriasis, and agonist mAbs to B- and T-lymphocyte attenuator (BTLA), which mediates T-cell inhibition by interacting with TNF receptors, are also being evaluated in clinical studies.

9.4 Combination Therapies

As discussed in Sect. 7.1.2, the addition of pertuzumab to trastuzumab brings additional benefit to breast cancers overexpressing HER2, but results of the combination on HER2-expressing gastric tumours were less impressive. Combinations of immune checkpoint inhibitors (ICIs) targeting PD-1 or PD-L1 have been successful in treating various malignancies, but as is often the case, there are still tumours that do not respond to this approach. Therefore, the use of combined immune ICIs given either as two separate mAbs or a single bispecific mAb is being tested. The combination of lymphocyte activating gene 3 (LAG-3) and PD-1 mAbs synergistically improve anti-tumour responses in murine models and in early clinical trials (Yap et al. 2019).

Currently it is not known if inflammatory adverse events such as colitis, pneumonitis and thyroiditis, which are well recognised effects of ICIs, will be more frequent, severe or serious following inhibition of two checkpoint pathways.

9.5 Atopic Dermatitis

The use of mAbs to treat atopic dermatitis has lagged behind psoriasis treatment by many years. The first mAb to be approved for the treatment of atopic dermatitis is dupilumab which binds to the IL-4Rα subunit shared by the IL-4 and IL-13 receptor complexes. Dupilumab inhibits IL-4 signaling via the type I receptor and both IL-4 and IL-13 signaling through the type II receptor. It was approved in 2017 for the treatment of adult patients with moderate-to-severe atopic dermatitis whose disease is not adequately controlled with topical prescription therapies or when those therapies are not advisable.

IL-33 is involved in the pathogenesis of atopic dermatitis, and preliminary data suggest that ANB020, a mAb that binds to IL-33, is efficacious in this condition. It is envisaged that mAbs will be developed to treat atopic dermatitis and other allergic diseases such as asthma, eosinophilic esophagitis and eosinophilic chronic rhinosinusitis.

9.6 Bispecifics

Bispecific mAbs are designed to bind to two epitopes on a single antigen target with the aim of obtaining a greater benefit than achievable from binding at a single site. Catumaxomab was one of the earliest mAbs to be developed and is a bispecific antibody. It binds to the tumour antigen EpCAM and the CD3 receptor on T cells. Subsequently bispecific T cell-engaging (BiTE) antibodies have been developed, an example being blinatumomab for Philadelphia-negative acute lymphoblastic leukaemia. The bispecific targets CD19 overexpressed on malignant B-cell precursors and CD3 on the cytotoxic T cells. Further development of this concept has led to the development of ImmTACs (immune mobilising monoclonal T-cell receptors against

cancer). ImmTACs retain binding to CD3 and use peptide fragments derived from intracellular tumour antigens as their second binding site. The peptide fragment forms a 'peptide HLA' complex that is expressed on the surface of the malignant cell. Binding of the ImmTAC to both malignant cells and T cells allows T cell-dependent killing of the malignant cells.

Emicizumab is a bispecific IgG4 that bridges clotting factors IX and X in the clotting cascade to restore function of factor VIII which is deficient in haemophilia A. The treatment is effective as prophylaxis of bleeding episodes in these patients and has been licensed for this indication.

Bispecific mAbs to two cytokines have been developed, e.g. COVA322 (anti-TNF and anti-IL-17). One of the issues with this approach is that it is effectively a 'combination product' with a fixed ratio (1:1) for the two modes of action. If the dosage requirements are substantially e.g. fourfold greater than single agent therapy, this approach may not be viable. One potential approach to solving this problem is to build mAbs with three or more binding sites, thereby allowing ratios of 2:1 or 3:1 to be used.

Doses may also be higher for bispecific mAbs, as in this example each molecule of antibody can only bind one TNF and one IL-17 molecule. Taking the worst-case scenario, the dose of this dual anti-cytokine mAb would be double than that of the single agents to maintain the same number of binding sites (50 mg of anti-TNF and 50 mg anti-IL-17 = 100 mg of bispecific antibody).

A very large number of bispecifics have entered development for use in oncology where the engagement of cytotoxic cells will aid in the destruction of malignant cells. Redirection of T cells is the most common approach being taken, but to date success has been very limited. Bispecifics are also being developed as a possible approach for treatment of Alzheimer's disease using the transferrin receptor in the blood-brain barrier to enhance entry of the antibody into the brain.

9.7 Nanobodies

The structure of mAbs is not identical in all species. For example, camels, llamas and alpacas all have heavy chain antibodies (i.e. lacking light chains). These are referred to as camelid antibodies from which the variable regions can be isolated to form so-called nanobodies. These have a molecular weight of 12–15 kDa which results in improved solubility, tissue penetration and stability (both thermal and pH).

The improved tissue penetration allows them to be used as imaging agents or treatment of CNS conditions as they have the potential to cross the blood-brain barrier. Unfortunately, as a consequence of the low molecular weight, they have high renal clearance and hence a short half-life.

Caplacizumab is the first nanobody to be approved and is used to treat acquired thrombotic thrombocytopenic purpura (aTTP) which is a life-threatening autoimmune thrombotic microangiopathy manifested by systemic microvascular thrombosis, profound thrombocytopenia, haemolytic anaemia and organ ischemia. Caplacizumab is administered daily, but this is not an issue when treating this condition as patients are hospitalised.

9.8 Intrabodies

One of the limitations of conventional intact mAbs is that they can only bind to extracellular targets because of their size and polarity. The potential to express antibody genes in cells (and hence synthesise mAbs inside the cell) has now been achieved using retroviral delivery systems. The antibodies that are expressed intracellularly, known as intrabodies, remain within the cytosol or endoplasmic reticulum of the cell and allow highly specific targeting of intracellular proteins (Marschall and Dübel 2016). For example, it may be possible to inhibit a single function of a multifunctional protein by targeting one particular epitope, e.g. binding to only one splice variant of the protein. By targeting a specific splice variant, the potential for unwanted adverse effects may be reduced.

There are examples of successful intrabody mediated target knockdowns that include oncogenic proteins, and targets involving neurodegeneration and chronic viral infections. An intrabody targeting the host protein CCR5 which is involved in viral entry of HIV into host cells has been shown to protect cells from HIV infection.

Studies of intrabodies in animals (in vitro and in vivo) have shown that neurodegenerative diseases (Parkinson's disease, amyotrophic lateral sclerosis and Huntington's disease) can be successfully targeted.

In clinical terms, the use of intrabodies would essentially be a form of gene therapy rather than the administration of an antibody per se. This approach has yet to be tested in humans.

9.9 Stereospecific and Catalytic mAbs

The recognition of epitopes by mAb is generally divided into two types. One is linear epitope-specific recognition, while the other is conformational or discontinuous epitope specific (Tsumoto et al. 2018). Stereospecific mAbs that recognise 3D configurations of molecules offer advantages over linear epitope-specific mAbs, which only recognise 2D configuration.

CD73 (ecto-5′-nucleotidase) is considered a promising immuno-oncology target (Antonioli et al. 2016). MEDI9447 is a mAb that noncompetitively inhibits CD73 activity. This mAb antagonises CD73 through dual mechanisms of inter-CD73 dimer cross-linking and/or steric blocking that prevent CD73 from adopting a catalytically active conformation (Geoghegan et al. 2016); this is an example of stereospecific recognition by a mAb.

Bispecific mAbs that include stereospecific recognition may be particularly effective for detecting membranous antigens on cancer cells, which are not easily recognised with conventional linear-specific monoclonal antibodies. Catalytic antibodies can recognise and degrade target antigens. Hifumi et al. have developed a catalytic antibody capable of degrading the active site of the urease of *Helicobacter pylori* and eradicating the bacterial infection in a mouse stomach (Hifumi et al. 2008). Catalytic antibodies represent a potential new class of therapeutic mAbs.

9.10 Local Administration of Monoclonal Antibodies

Local administration of mAbs (Jones and Martino 2016) is already established in the treatment of ocular disease (ranibizumab, for age-related macular degeneration; bevacizumab, for corneal neovascularisation), intra-articular joint injections (e.g. anti-TNF for persistent inflammatory monoarthritis) and intra-tumoural injections. There is the potential for topical treatments to expand through the uses of Fabs and nanobodies allowing treatments of skin diseases (e.g. psoriasis, atopic dermatitis). In particular, the improved solubility, tissue penetration and stability of nanobodies would lend them to topical administration. Local application could reduce the potential for adverse events and reduce costs.

9.11 Monoclonal Antibodies for Less Common Diseases: Personalised Medicine

At one time 'Big Pharma' was mainly interested in 'blockbuster medicines' that were suitable for most patients with common diseases, e.g. angiotensin antagonists and calcium ion channel blockers for hypertension, but there is now much greater interest in developing medicines for comparatively uncommon diseases and subsets of patients with a particular disorder. For example, patients with eosinophilic asthma comprise less than 5% of all patients with asthma, amounting to about 100,000 individuals in the UK. This is a severe debilitating condition for which treatment of those failing to respond to high-dose inhaled corticosteroids was highly unsatisfactory. Three mAbs are now available for such patients. Mepolizumab and reslizumab bind to IL-5 and benralizumab to the IL-5 receptor IL-5R$_{\propto}$. These have been demonstrated to be highly effective treatments.

For patients with atopic asthma, which is mediated by IgE antibodies, omalizumab has proved an effective treatment. It is an anti-IgE mAb approved for treatment of moderate-to-severe IgE-mediated (allergic) asthma. It blocks free serum IgE, reducing its effector functions by inhibiting its binding to high-affinity receptors on inflammatory cells in the allergic cascade. Omalizumab is tolerated well and improves symptoms and exacerbations of atopic asthma and is steroid sparing. These mAbs for less common forms of asthma exemplify the use of mAbs as personalised medicines, which are bound to provide exciting advances in treatment of patients with less common diseases.

References

Aalberse RC, Stapel SO, Schuurman J, Rispens T (2009) Immunoglobulin G4: an odd antibody. Clin Exp Allergy 39:469–477

Aggeletopoulou I, Assimakopoulos SF, Konstantakis C, Triantos C (2018) Interleukin 12/Interleukin 23 pathway: biological basis and therapeutic effect in patients with Crohn's disease. World J Gastroenterol 24:4093–4103

Alegre ML, Peterson LJ, Xu D et al (1994) A non-activating "humanized" anti-CD3 monoclonal antibody retains immunosuppressive properties in vivo. Transplantation 57:1537–1543
Angal S, King DJ, Bodmer MW, Turner A, Lawson AD, Roberts G, Pedley B, Adair JR (1993) A single amino acid substitution abolishes the heterogeneity of chimeric mouse/human (IgG4) antibody. Mol Immunol 30:105–108
Annibali O et al (2018) PD-1 /PD-L1 checkpoint in hematological malignancies. Leuk Res 67:45–55
Antonioli L, Yegutkin GG, Pacher P, Blandizzi C, Hasko G (2016) Anti-CD-73 in cancer immunotherapy: awakening new opportunities. Trends Cancer 2:95–109
Bakri SJ, Snyder MR, Reid JM, Pulido JS, Ezzat MK, Singh R (2007) Pharmacokinetics of intravitreal ranibizumab (Lucentis). Ophthalmology 114:2179–2182
Baldo BA (2013) Adverse events to monoclonal antibodies used for cancer therapy: focus on hypersensitivity responses. Oncoimmunology e26333:2
Baselga J et al (2012) Pertuzumab plus trastuzumab plus docetaxel for metastatic breast cancer. N Engl Med 366:109–119
Bathon J, Martin RW, Fleischmann et al (2000) A comparison of etanercept and methotrexate in patients with early rheumatoid arthritis. New Engl J Med 343:1586–1593
Behm BW, Bickston SJ (2008) Tumor necrosis factor-alpha antibody for maintenance of remission in Crohn's disease. Cochrane Database Syst Rev:CD006893. https://doi.org/10.1002/14651858.CD006893
Belai EB (2014) PD-1 blockage delays murine squamous cell carcinoma development. Carcinogenesis 35:424–431
Blackwell K et al (2018) The global need for a trastuzumab biosimilar for patients with HER2-positive breast cancer. Clin Breast Cancer 18:95–113
Bolt S, Routledge E, Lloyd I et al (1993) The generation of a humanized, non-mitogenic CD3 monoclonal antibody which retains *in vitro* immunosuppressive properties. Eur J Immunol 23:403–411
Boswell CA, Tesar D, Mukhyala K, Thell FP, Fielder PJ, Khawli LI (2010) Effects of charge on antibody tissue distribution and pharmacokinetics. Bioconjug Chem 21:2153–2163
Braegger CP, Nicholls S, Murch SH, Stephens S, MacDonald TT (1992) Tumour necrosis factor alpha in stool as a marker of intestinal inflammation. Lancet 339:89–91
Brambell FW (1966) The transmission of immunity from mother to young and the catabolism of immunoglobulins. Lancet 2:1087–1093
Brambell FW, Hemmings WA, Morris IG (1964) A theoretical model of gamma-globulin catabolism. Nature 203:1352–1354
Brennan FM, Chantry D, Jackson A, Maini RN, Feldmann M (1989) Inhibitory effect of TNF alpha antibodies on synovial cell interleukin-1 production in rheumatoid arthritis. Lancet 2:244–247
Bruggemann M, Winter G, Waldmann H, Neuberger MS (1989) The immunogenicity of chimeric antibodies. J Exp Med 170:2153–2157
Buchan G, Barrett K, Turner M, Chantry D, Maini RN, Feldmann M (1988) Interleukin-1 and tumour necrosis factor mRNA expression in rheumatoid arthritis: prolonged production of IL-1 alpha. Clin Exp Immunol 73:449–455
Burmeister WP, Huber AH, Bjorkman PJ (1994) Crystal structure of the complex of rat neonatal Fc receptor with Fc. Nature 372:379–383
Cameron F, Whiteside G, Perry C (2011) Ipilimumab: first global approval. Drugs 71:1093–1104
Capelan M et al (2013) Pertuzumab: new hope for patients with HER2-positive breast cancer. Ann Oncol 24:273–282
Challa DK, Bussmeyer U, Khan T, Montoyo HP, Bansal P, Ober RJ, Ward ES (2013) Autoantibody depletion ameliorates disease in murine experimental autoimmune encephalomyelitis. MAbs 5:655–659
Chang TW (2006) Developing antibodies for targeting immunoglobulin and membrane-bound immunoglobulin E. Allergy Asthma Proc 27(Suppl 1):S7–S14

Chaparro-Riggers J, Liang H, DeVay RM, Bai L, Sutton JE, Chen W, Geng T, Lindquist K, Casas MG, Boustany LM et al (2012) Increasing serum half-life and extending cholesterol lowering in vivo by engineering antibody with pH-sensitive binding to PCSK9. J Biol Chem 287:11090–11097

Chapman K, Pullen NI, Coney L, Dempster M et al (2009) Preclinical development of monoclonal antibodies. Considerations for the use of non-human primates. MAbs 1:505–516

Chatenoud L, Baudrihaye MF, Chkoff N, Kreis H, Goldstein G, Bach J-F (1986) Restriction of the human *in vivo* immune response against the mouse monoclonal antibody OKT3. J Immunol 137:830–838

Chatenoud L, Primo J, Bach J-F (1997) CD3 antibody-induced dominant self-tolerance in overtly diabetic NOD mice. J Immunol 158:2947–2954

Chen L, Flies DB (2013) Molecular mechanisms of T cell co-stimulation and co-inhibition. Nat Rev Immunol 13:227–242

Chen L, Han X (2015) Anti-PD-1/PD-L1 therapy of human cancer: past, present, and future. J Clin Invest 125:3384–3391

Chung CH et al (2008) Cetuximab-induced anaphylaxis and IgE specific for galactose-alpha-1,3-galactose. New Engl J Med 358:1109–1117

Clark M (2000) Antibody humanization: a case of the 'Emperor's new clothes'? Immunol Today 21:397–402

Clarke J, Leach W, Pippig S, Joshi A, Wu B, House R, Beyer J (2004) Evaluation of a surrogate antibody for preclinical safety testing of an anti-CD11a monoclonal antibody. Regul Toxicol Pharmacol 40:219–226

Coats S, Williams M, Kebbie B et al (2019) Antibody-drug conjugates: future directions in clinical and translational strategies to improve the therapeutic index. Clin Cancer Res. https://doi.org/10.1158/1078-0432.CCR-19-0272

Cohen BA et al (2008) The implications of immunogenicity for protein-based multiple sclerosis therapies. J Neurol Sci 275:7–17

Constantinidou A, Alifieris C, Trafalis DT (2019) Targeting programmed cell death-1 (PD-1) and ligand (PD-L1): a new era in cancer active immunotherapy. Pharmacol Ther 194:84–106

Cope AP, Londei M, Chu R, Cohen SBA, Elliott MJ, Brennan FM, Maini RN, Feldmann M (1994) Chronic exposure to tumor necrosis factor (TNF) in vitro impairs the activation of T cells through the T cell receptor/CD3 complex; reversal in vivo by anti-TNF antibodies in patients with rheumatoid arthritis. J Clin Invest 94:749–760

Cortez-Jugo C, Qi A, Rajapaksa A, Friend JR, Yeo LY (2015) Pulmonary monoclonal antibody delivery via a portable micorfluidic nebulization platform. Biomicrofluidics 9:052603

Datta-Mannan A (2019) Mechanisms influencing the disposition of monoclonal antibodies and peptides. Drug Metab Dispos. https://doi.org/10.1124/dmd.119.086488

Datta-Mannan A, Croy JE, Schirtzinger L, Torgerson S, Breyer M, Wroblewski VJ (2016) Aberrant bispecific antibody pharmacokinetics linked to liver sinusoidal endothelium clearance mechanism in cynomolgus monkeys. MAbs 8:969–982

Datta-Mannan A, Choi H, Stokell D, Tang J, Murphy A, Wrobleski A, Feng Y (2018) The properties of cysteine-conjugated antibody-drug conjugates are impacted by the IgG subclass. AAPS J 20:103

Descotes J (2009) Immunotoxicity of monoclonal antibodies. MAbs 1:104–111

Doessegger L, Banholzer ML (2015) Clinical development methodology for infusion-related reactions with monoclonal antibodies. Clin Transl Immunol 4:e39

Elliot MJ, Maini RN, Feldmann M, Kalden JR, Antoni C, Smolen JS, Leeb B, Breedveld FC, Macfarlane JD, Bijl JA, Woody JN (1994) Randomised double-blind comparison of chimeric monoclonal antibody to tumour necrosis factor α (cA2) versus placebo in rheumatoid arthritis. Lancet 344:1105–1110

Elliott MJ, Maini RN, Feldmann M et al (1993) Treatment of rheumatoid arthritis with chimeric monoclonal antibodies to tumor necrosis factor alpha. Arthritis Rheum 36:1681–1690

Elliott MJ, Maini RN, Feldmann M et al (1994) Repeated therapy with monoclonal antibody to tumour necrosis factor alpha (cA2) in patients with rheumatoid arthritis. Lancet 344:1125

Ettlinger DE, Mitterhauser M, Wadsak W, Ostermann E, Farkouh A, Schueller J, Czejka M (2006) In vivo disposition of irinotecan (CPT-11) and its metabolites in combination with the monoclonal antibody cetuximab. Anticancer Res 26:1337–1341

Feagan BG, Sandborn WJ, Gasink C et al (2016) Ustekinumab as Induction and maintenance therapy for Crohn's disease. N Engl J Med 375:1946–1960

Feldmann M, Brennan FM, Maini RN (1996) Role of cytokines in rheumatoid arthritis. Annu Rev Immunol 14:397–440

Ferran C, Sheehan K, Dy M et al (1990) Cytokine-related syndrome following injection of anti-CD3 monoclonal antibody: further evidence for transient in vivo T cell activation. Eur J Immunol 20:509–515

Ferri N, Bellosta S, Baldessin L, Boccia D, Racagni G, Cosini A (2016) Pharmacokinetic interactions of monoclonal antibodies. Pharmacol Res 111:592–599

Freeman GJ et al (2000) Engagement of the PD-1 immunoinhibitory receptor by a novel B7 family member leads to negative regulation of lymphocyte activation. J Exp Med 192:1027–1034

Garg A, Balthasar JP (2007) Physiologically-based pharmacokinetic (PBPK) model to predict IgG tissue kinetics in wild-type and FcRn-knockout mice. J Pharmacokinet Pharmacodyn 34:687–709

Gatalica Z et al (2014) Programmed cell death 1 (PD-1) and its ligand (PD-L1) in common cancers and their correlation with molecular cancer type. Cancer Epidemiol Biomarkers Prev 23:2965–2970

Gaudreault J, Shiu V, Bricarello A et al (2005) Concomitant administration of bevacizumab, irinotecan, 5-fluorouracil, and leucovorin: nonclinical safety and pharmacokinetics. Int J Toxicol 24:357–363

Genovese MC, Becker J-C, Schiff M et al (2005) Abatacept for rheumatoid arthritis refractory to tumour necrosis factor α inhibition. New Engl J Med 353:1114–1123

Geoghegan JC, Diedrich G, Xiaojun L et al (2016) Inhibition of CD73 AMP hydrolysis by a therapeutic antibody with a dual, non-competitive mechanism of action. MAbs 8:454–467

Gerratana L et al (2017) Pertuzumab and breast cancer: another piece in the anti-HER2 puzzle. Exp Opin Biol Ther 17:365–374

Grantab RH, Tannock IF (2012) Penetration of anticancer drugs through tumour tissue as a function of cellular packing density and interstitial fluid pressure and its modification by bortezomib. BMC Cancer 12:214

Groom JR (2019) Regulators of T-cell fate: integration of cell migration, differentiation and function. Immunol Rev 289:101–114

Haanen JBAG et al (2017) Management of toxicities from immunotherapy: ESMO clinical practice guidelines for diagnosis, treatment and follow-up. Ann Dent 28(suppl 4):119–142

Hansel TT et al (2010) The safety and side effects of monoclonal antibodies. Nat Rev Drug Discov 9:325–338

Haworth C, Brennan FM, Chantry D, Turner M, Maini RN, Feldmann M (1991) Expression of granulocyte macrophage colony-stimulating factor (GM-CSF) in rheumatoid arthritis: regulation by tumor necrosis factor-alpha. Eur J Immunol 21:2575–2579

Henne KR, Ason B, Howard M et al (2015) Anti-PCSK9 antibody pharmacokinetics and low density lipoprotein-cholesterol pharmacodynamics in non-human primates are antigen affinity-dependent and exhibit limited sensitivity to neonatal Fc receptor-binding enhancement. J Pharmacol Exp Ther 353:119–131

Hifumi E, Morihara F, Hatiuchi K, Okuda T, Nishizono A, Uda T (2008) Catalytic features and eradication ability of antibody light chain UA15-L against *H. pylori*. J Biol Chem 283:899–907

Hirsch FR et al (2017) PD-L1 Immunohistochemistry assays for lung cancer: results from phase 1 of the blueprint PD-L1 IHC assay comparison project. J Thorac Oncol 12:208–222

Ho P-R et al (2017) Risk of natalizumab-associated progressive multifocal leukoencephalopathy in patients with multiple sclerosis: a retrospective analysis of data from four clinical studies. Lancet Neurol 16:925–933
Hodi FS et al (2010) Improved survival with ipilimumab in patients with metastatic melanoma. New Engl J Med 363:711–723
Hooks MA, Wade CS, Millikan WJ (1991) Muromonab CD-3: a review of its pharmacology, pharmacokinetics, and clinical use in transplantation. Pharmacotherapy 11:26–37
Hua F, Ribbing J, Reinisch W, Cataldi F, Martin S (2015) A pharmacokinetic comparison of anrukinzumab, an anti-IL-13 monoclonal antibody, among healthy volunteers, asthma and ulcerative colitis patients. Br J Clin Pharmacol 80:101–109
Hudis CA (2007) Trastuzumab--mechanism of action and use in clinical practice. New Engl J Med 357:39–51
Hünig T (2012) The storm has cleared: lessons from the CD28 superagonist TGN1412 trial. Nat Rev Immunol 12:317–318
Hutchinson M (2007) Natalizumab: A new treatment for relapsing remitting multiple sclerosis. Ther Clin Risk Manag 3:259–268
Ishida Y et al (1992) Induced expression of PD-1, a novel member of the immunoglobulin gene superfamily, upon programmed cell death. EMBO J 11:3887–3895
Jain T, Litzow MR (2018) No free rides: management of toxicities of novel immunotherapies in ALL, including financial. Blood Adv 2:3393–3403
Jefferis R (2009) Glycosylation of antibody therapeutics: optimisation for purpose. Methods Mol Biol 483:223–238
Jeon C, Sekhon S, Yan D, Ladan A, Nakamura M, Bhutani T (2017) Monoclonal antibodies inhibiting IL-12, -23, and -17 for the treatment of psoriasis. Human Vacc Immunother 13:2247–2259
Jones JL, Coles AJ (2014) Mode of action and clinical studies with alemtuzumab. Exp Neurol 262:37–43
Jones RG, Martino A (2016) Targeted localized use of therapeutic antibodies: a review of non-systemic, topical and oral applications. Crit Rev Biotechnol 36:506–520
Jones PT, Dear PH, Foote J, Neuberger MS, Winter G (1986) Replacing the complementarity-determining regions in a human antibody with those from a mouse. Nature 321:522–525
Josefowicz SZ, Lu L-F, Rudensky AY (2012) Regulatory T cells: mechanisms of differentiation and function. Annu Rev Immunol 30:531–564
Kaida-Yip F, Deshpande K, Saran T, Vyas D (2018) Biosimilars: review of current applications, obstacles, and their future in medicine. World J Clin Cases 16:161–166
Kauffman CL, Aria N, Toichi E, McCormick TS, Cooper KD, Gottlieb AB, Everitt DE, Frederick B, Zhu Y, Graham MA, Pendley CE, Mascelli MA (2004) A phase I study evaluating the safety, pharmacokinetics, and clinical response of a human IL-12 p40 antibody in subjects with plaque psoriasis. J Invest Dermatol 123:1037–1044
Keane J (2005) TNF-blocking agents and tuberculosis: new drugs illuminate an old topic. Rheumatology 44:714–720
Keating GM (2010) Panitumumab: a review of its use in metastatic colorectal cancer. Drugs 70:1059–1078
Kennett RH (1981) Hybridomas: a new dimension in biological analyses. In Vitro 17:1036–1050
Kim GW, Lee NR, Pi RH, Lim YS, Lee YM, Lee JM, Jeong HS, Chung SH (2015) IL-6 inhibitors for treatment of rheumatoid arthritis: past, present, and future. Arch Pharm Res 38:575–584
Knodler M, Korfer J, Kuzumann V et al (2018) Randomised phase II trial to investigate catumaxomab (anti-EpCAM × anti-CD3) for treatment of peritoneal carcinomatosis in patients with gastric cancer. Br J Cancer 119:296–302
Kohler G, Milstein C (1975) Continuous cultures of fused cells secreting antibody of predefined specificity. Nature 256:495–497
Krishnamurti U, Silverman JF (2014) HER2 in breast cancer: a review and update. Adv Anat Pathol 21:100–107

Kumar P, Bhattacharya P, Prabhakar BS (2018) A comprehensive review on the role of co-signaling receptors and Treg homeostasis in autoimmunity and tumor immunity. J Autoimmun 95:77–99
Kunert R, Reinhart D (2016) Advances in recombinant antibody manufacturing. Appl Microbiol Biotechnol 100:3451–3461
Kung P, Goldstein G, Reinherz EL, Schlossman SF (1979) Monoclonal antibodies defining distinctive human T cell surface antigens. Science 206:347–349
Labrijn AF, Janmaat ML, Reichert JM, Parren PWHI (2019) Bispecific antibodies: a mechanistic review of the pipeline. Nat Rev Drug Discov 18:585–608
The Lancet (2018) End of the road for daclizumab in multiple sclerosis. Lancet 391:1000
Latchman Y, Wood CR, Chernova T et al (2001) PD-L2 is a second ligand for PD-1 and inhibits T cell activation. Nat Immunol 2:261–268
Leach MW, Halpern WG, Johnson CW et al (2010) Use of tissue cross-reactivity studies in the development of antibody-based biopharmaceuticals: history, experience, methodology, and future directions. Toxicol Pathol 38:1138–1166
Lee SJ, Kavanaugh A (2005) Adverse reactions to biologic agents: focus on autoimmune disease therapies. J Allergy Clin Immunol 116:900–905
Lee DM, Weinblatt ME (2001) Rheumatoid arthritis. Lancet 358:903–911
Lenz H-J (2007) Cetuximab in the management of colorectal cancer. Biologics 1:77–91
Lipsky PE, van der Heijde DMFM, St Clair EW et al (2000) Infliximab and methotrexate in the treatment of RA. N Engl J Med 343:1594–1602
Liu L (2015) Antibody glycosylation and its impact on the pharmacokinetics and pharmacodynamics of monoclonal antibodies and Fc-fusion proteins. J Pharm Sci 104:1866–1884
Liu L (2018) Pharmacokinetics of monoclonal antibodies and Fc-fusion proteins. Protein Cell 9:15–32
Lobo PI, Patel HC (1997) Murine monoclonal IgG antibodies: differences in their IgG isotypes can affect the antibody effector activity when using human cells. Immunol Cell Biol 75:267–274
Loyau J, Malinge P, Daubeuf B et al (2014) Maximizing the potency of an anti-TLR4 monoclonal antibody by exploiting proximity to Fcγ receptors. MAbs 6:1621–1630
Lupo MG, Ferri N (2018) Angiopoietin-like 3 (ANGPTLS) and atherosclerosis: lipid and non-lipid related effects. J Caridiovasc Dev Dis 5:E39
Maloney DG (2012) Anti-CD20 antibody therapy for B-cell lymphomas. New Engl J Med 366:2008–2016
Maloney DG et al (1997) IDEC-C2B8 (rituximab) anti-CD20 monoclonal antibody therapy in patients with relapsed low-grade non-Hodgkin's lymphoma. Blood 90:2188–2195
Marschall ALJ, Dübel S (2016) Antibodies inside of a cell can change its outside: can intrabodies provide a new therapeutic paradigm? Comput Struct Biotechnol J 14:304–308
McCall B (2019) Alemtuzumab to be restricted pending review, says EMA. Lancet 393:1683
McDermott J, Jimeno A (2015) Pembrolizumab: PD-1 inhibition as a therapeutic strategy in cancer. Drugs Today (Barc) 51:7–20
Melero I et al (2014) Therapeutic vaccines for cancer: an overview of clinical trials. Nat Rev Clin Oncol 11:509–524
Messersmith WA, Ahnen DJ (2008) Targeting EGFR in colorectal cancer. New Engl J Med 359:1834–1836
Migden MR et al (2018) PD-1 blockade with cemiplimab in advanced cutaneous squamous-cell carcinoma. New Engl J Med 379:341–351
Monaco C, Nanchahal J, Taylor P, Feldmann M (2014) Anti-TNF therapy: past, present and future. Int Immunol 27:55–62
Morrison SL, Johnson MJ, Herzenberg LA, Oi VT (1984) Chimeric human antibody molecules: mouse antigen-binding domains with human constant region domains. Proc Natl Acad Sci U S A 81:6851–6855
Muenst S, Läubli H, Soysal SD, Zippelius A, Tzankov A, Hoeller S (2016) The immune system and cancer evasion strategies: therapeutic concepts. J Intern Med 279:541–562

Nami B, Maadi H, Wang Z (2018) Mechanisms underlying the action and synergism of trastuzumab and pertuzumab in targeting HER2-positive breast cancer. Cancer 10:342
Narazaki M, Tanaka T, Kishimoto T (2017) The role and therapeutic targeting of IL-6 in rheumatoid arthritis. Expert Rev Clin Immunol 13:535–551
Nelson AL (2010) Antibody fragments: hope and hype. MAbs 2:77–83
Nemeth BT, Varga ZV, Wu WJ, Pacher P (2017) Trastuzumab cardiotoxicity: from clinical trials to experimental studies. Br J Pharmacol 174:3727–3748
Nishijima TF, Shachar SS, Nyrop KA, Muss HB (2017) Safety and tolerability of PD-1/PD-L1 inhibitors compared with chemotherapy in patients with advanced cancer: a meta-analysis. Oncologist 22:470–479
Noman MZ et al (2014) PD-L1 is a novel direct target of HIF-1α, and its blockade under hypoxia enhanced MDSC-mediated T cell activation. J Exp Med 211:781–790
O'Donnell JS, Massi D, Teng MWL, Mandala M (2018) PI3K-AKT-mTOR inhibition in cancer immunotherapy, redux. Semin Cancer Biol 48:91–103
Oroudjev E, Lopus M, Wilson L et al (2010) Maytansinoid-antibody conjugates induce mitotic arrest by suppressing microtubule dynamic instability. Mol Cancer Ther 9:2700–2713
Patel DA, Puig-Canto A, Challa DK, Perez Montoyo H, Ober RJ, Ward ES (2011) Neonatal Fc receptor blockade by Fc engineering ameliorates in a murine model. J Immunol 187:1015–1022
Paulus HE, Egger MJ, Ward JR, Williams HJ (1990) Analysis of improvement in individual rheumatoid arthritis patients treated with disease-modifying antirheumatic drugs, based on the finding in patients treated with placebo. The Cooperative Systematic Studies of Rheumatic Diseases Group. Arthritis Rheum 33:477–484
Peddi PF, Hurvitz SA (2013) Trastuzumab emtansine: the first targeted chemotherapy for treatment of breast cancer. Future Oncol 9:319–326
Pichler WJ (2006) Adverse side-effects to biological agents. Allergy 61:912–920
Present DH, Rutgeerts P, Targan S, Hanauer SB, Mayer L, van Hogezand RA, Podolsky DK, Sands BE, Braakman T, DeWoody KL, Schaible TF, van Deventer SJ (1999) Infliximab for the treatment of fistulas in patients with Crohn's disease. N Engl J Med 340:1398–1405
Price TJ, Peeters M, Kim TW et al (2014) Panitumumab versus cetuximab in patients with chemotherapy-refractory wild-type KRAS exon 2 metastatic colorectal cancer (ASPECCT): a randomised, multicentre, open-label, non-inferiority phase 3 study. Lancet Oncol 15:569–579
Ribas A, Wolchok JD (2018) Cancer immunotherapy using checkpoint blockade. Science 359:1350–1355
Richards DM, Delacher M, Goldfarb Y et al (2015) Treg cell differentiation: from thymus to peripheral tissue. Prog Mol Biol Transl Sci 136:175–205
Riegert-Johnson DL, Godfrey JA, Myers JL et al (2002) Delayed hypersensitivity reaction and acute respiratory distress syndrome following infliximab infusion. Inflamm Bowel Dis 8:186–191
Robbie GJ, Criste R, Dall'acqua WF et al (2013) A novel investigational Fc-modified humanized monoclonal antibody, motavizumab-YTE, has an extended half-life in healthy adults. Antimicrob Agents Chemother 85:621–631
Robbins PF, Kassim SH, Tran TL et al (2015) A pilot trial using lymphocytes genetically engineered with an NY-ESO-1-reactive T-cell receptor: long-term follow-up and correlates with response. Clin Cancer Res 21:1019–1027
Roopenian DC, Akilesh S (2007) FcRn: the neonatal Fc receptor comes of age. Nat Rev Immunol 7:715–725
Roselló S, Biasco I, Garcia Fabregat L, Cervantes A, Jordan K, ESMO Clinical Practice Guidelines (2017) Management of infusion reactions to systemic anticancer therapy: ESMO clinical practice guidelines. Ann Oncol 28(suppl 4):iv100–iv118
Rosenberg SA, Sherry RM, Morton KE et al (2005) Tumor progression can occur despite the induction of very high levels of self/tumor antigen-specific CD8+ T cells in patients with melanoma. J Immunol 175:6169–6176

Roskoski R (2014) The ErbB/HER family of protein-tyrosine kinases and cancer. Pharmacol Res 79:34–74

Ross JS, Slodkowska EA, Symmans WF et al (2009) The HER-2 receptor and breast cancer: ten years of targeted anti-HER-2 therapy and personalized medicine. Oncologist 14:320–368

Ryman JT, Meibohm B (2017) Pharmacokinetics of monoclonal antibodies. CPT Pharmacometrics Syst Pharmacol 6:576–588

Salles G, Barrett M, Foà R et al (2017) Rituximab in B-cell hematologic malignancies: a review of 20 years of clinical experience. Adv Ther 34:2232–2273

Samineni D, Girish S, Li C (2016) Impact of Shed/Soluble targets on the PK/PD of approved therapeutic monoclonal antibodies. Expert Rev Clin Pharmacol 9:1557–1569

Sandborn WJ, Hanauer SB (2002) Infliximab in the treatment of Crohn's disease: a user's guide for clinicians. Am J Gastroenterol 97:2962–2972

Sanmamed MF, Chen L (2014) Inducible expression of B7-H1 (PD-L1) and its selective role in tumor site immune modulation. Cancer J 20:256–261

Sanmamed MF, Chen L (2018) A paradigm shift in cancer immunotherapy: from enhancement to normalization. Cell 175:313–326

Saribas AS, Ozdemir A, Lam C, Safak M (2010) JC virus-induced progressive multifocal leukoencephalopathy. Futur Virol 5:313–323

Sathish JG, Sethu S, Bielsky MC et al (2013) Challenges and approaches for the development of safer immunomodulatory biologics. Nat Rev Drug Discov 12:306–324

Schuurman J, van Ree R, Perdok GJ, van Doorn HR, Tan KY, Aalberse RC (1999) Normal human immunoglobulin G4 is bispecific: it has two different antigen-combining sites. Immunology 97:693–698

Schuurman J, Perdok GJ, Gorter AD, Aalberse RC (2001) The inter-heavy chain disulfide bonds of IgG4 are in equilibrium with intra-chain disulfide bonds. Mol Immunol 38:1–8

Schwartzentruber DJ, Hom SS, Dadmarz R et al (1994) In vitro predictors of therapeutic response in melanoma patients receiving tumor-infiltrating lymphocytes and interleukin-2. J Clin Oncol 12:1475–1483

Sedykh SE, Prinz VV, Buneva VN, Nevinsky GA (2018) Bispecific antibodies: design, therapy, perspectives. Drug Des Dev Ther 12:195–208

Sgro C (1995) Side-effects of a monoclonal antibody, muromonab CD3/orthoclone OKT3: bibliographic review. Toxicology 105:23–29

Singer BA (2017) The role of natalizumab in the treatment of multiple sclerosis: benefits and risks. Ther Adv Neurol Disord 10:327–336

Smolen JS, Aletaha D, McInnes IB (2016) Rheumatoid arthritis. Lancet 388:2023–2038

Sondak VK, Smalley KS, Kudchadkar R, Grippon S, Kirkpatrick P (2011) Ipilimumab. Nat Rev Drug Discov 10:411–412

Stubenrauch K, Wessels U, Regula JT, Kettenberger H, Schleypen J, Kohnert U (2010) Impact of molecular processing in the hinge region of therapeutic IgG4 antibodies on disposition profiles in cynomolgus monkeys. Drug Metab Dispos 38:84–91

Swain SM, Kim SB, Cortes J, Ro J, Siglazov V, Campone M, Ciruelos E, Ferrero JM, Schneeweiss A, Knott A, Clark E, Ross G, Benyunes MC, Baselga J (2013) Pertuzumab, trastuzumab, and docetaxel for HER2-positive breast cancer (CLEOPATRA study): overall survival results from a randomised, double-blind, placebo-controlled phase 3 study. Lancet Oncol 14:461–471

Tabernero J, Hoff PM, Shen L, Ohtsu A, Shah MA, Cheng K, Song C, Wu H, Eng-Wong J, Kim K, KangYK (2018) Pertuzumab plus trastuzumab and chemotherapy for HER2-positive metastatic gastric or gastro-oesophageal junction cancer (JACOB): final analysis of a double-blind, randomised, placebo-controlled phase 3 study. Lancet Oncol 19:1372–1384

Talbot JJ, Calamba D, Pai M, Ma M, Thway TM (2015) Measurement of free versus total therapeutic monoclonal antibody in pharmacokinetic assessment is modulated by affinity, incubation time, and bioanalytical platform. AAPSJ 17:1446–1454

Targan SR, Hanauer SB, van Deventer SJ, Mayer L, Present DH, Braakman T, DeWoody KL, Schaible TF, Rutgeerts PJ (1997) A short-term study of chimeric monoclonal antibody cA2 to tumor necrosis factor α for Crohn's disease. N Engl J Med 337:1029–1035

Taube JM, Anders RA, Young GD et al (2012) Colocalization of inflammatory response with B7-h1 expression in human melanocytic lesions supports an adaptive resistance mechanism of immune escape. Sci Trans Med 4:127ra37

Tsumoto K, Isozaki Y, Yagami H, Tomita M (2018) Future perspectives of therapeutic monoclonal antibodies. Immunotherapy 11:119–127

Vahle JL, Finch GL, Heidel SM et al (2010) Carcinogenicity assessments of biotechnology-derived pharmaceuticals: a review of approved molecules and best practice recommendations. Toxicol Pathol 38:522–553

van der Bruggen P, Traversari C, Chomez P et al (1991) A gene encoding an antigen recognized by cytolytic T lymphocytes on a human melanoma. Science 254:1643–1647

van der Kolk LE, Grillo-Lopez AJ, Baars JW, Hack CE, van Oers MH (2001) Complement activation plays a key role in the side-effects of rituximab treatment. Br J Haematol 115:807–811

Vidarsson G, Dekkers G, Rispens T (2014) IgG subclasses and allotypes: from structure to effector functions. Front Immunol 5:1–17

Vinay DS, Ryan EP, Pawelec G et al (2015) Immune evasion in cancer: mechanistic basis and therapeutic strategies. Semin Cancer Biol 35:S185–S198

von Minckwitz G, Huang CS, Mano MS et al (2019) Trastuzumab emtansine for residual invasive HER2-positive breast cancer. New Engl J Med 380:617–628

Wakefield I, Stephens S, Foulkes R, Nesbitt A, Bourne T (2011) The use of surrogate antibodies to evaluate the development and reproductive toxicity potential of an anti-TNFα PEGylated Fab' monoclonal antibody. Toxicol Sci 122:170–176

Wang DD, Zhang S, Zhao H, Men AY, Parivar K (2009) Fixed dosing versus body size-based dosing of monoclonal antibodies in adult clinical trials. J Clin Pharmacol 49:1012–1024

Wang J, Wang YM, Ahn HY (2014) Biological products for the treatment of psoriasis: therapeutic targets, pharmacodynamics and disease-drug-drug interaction implications. AAPS J 16:938–947

Ward ES, Zhou J, Ghetie V, Ober RJ (2003) Evidence to support the cellular mechanism involved in serum IgG homeostasis in humans. Int Immunol 15:187–195

Ward ES, Martinez C, Vaccaro C, Zhou J, Tang Q, Ober RJ (2005) From sorting endosomes to exocytosis: association of Rab4 and Rab11 GTPases with the Fc receptor, FcRn, during recycling. Mol Biol Cell 16:2028–2038

Weber J, Mandala M, Del Vecchio M et al (2017) Adjuvant nivolumab versus ipilimumab in resected stage III or IV melanoma. N Engl J Med 377:1824–1835

Weisman MH, Durez P, Hallegua D et al (2006) Reduction of inflammatory biomarker response by abatacept in treatment of rheumatoid arthritis. J Rheumatol 33:2162–2166

Wiley HS, Shvartsman SY, Lauffenburger DA (2003) Computational modeling of the EGF-receptor system: a paradigm for systems biology. Trends Cell Biol 13:43–50

Williams RO, Feldmann M, Maini RN (1992) Anti-tumor necrosis factor ameliorates joint disease in murine collagen induced arthritis. Proc Natl Acad Sci U S A 89:9784–9788

Wirth TC, Kühnel F (2017) Neoantigen targeting – dawn of a new era in cancer immunotherapy? Front Immunol 8:1031–1016

Xu L, Zuch CL, Lin YS, Modi NB, Lum BL (2008) Pharmacokinetics and safety of bevacizumab administered in combination with cisplatin and paclitaxel in cynomolgus monkeys. Cancer Chemother Pharmacol 61:607–614

Yamane-Ohnuki N, Satoh M (2009) Production of therapeutic antibodies with controlled fucosylation. MAbs 1:230–236

Yap TA, Papadopoulos KP, LoRusso P, Wong DJL, Hu-Lieskovan S, Holz J-B (2019) A first-in-human phase I study of FS118, an anti-LAG-3/PD-L1 bispecific antibody in patients with solid tumors that have progressed on prior PD-1/PD-L1 therapy. J Clin Oncol 37:TPS2652

Yu XQ, Robbie GJ, Wu Y et al (2016) Safety, tolerability, and pharmacokinetics of MEDI4893, an Investigational, extended-half-life, anti-staphylococcus aureus alpha-toxin human monoclonal antibody, in healthy adults. Antimicrob Agents Chemother 27:61. pii: e01020-16

Zhang X et al (2004) Structural and functional analysis of the costimulatory receptor programmed death-1. Immunity 20(3):337–347

Zhou H, Jang H, Fleischmann RM et al (2007) Pharmacokinetics and safety of golimumab, a fully human anti-TNF-alpha monoclonal antibody in subjects with rheumatoid arthritis. J Clin Pharmacol 47:383–396

Zinner RG, Glisson BS, Fossella FV et al (2004) Trastuzumab in combination with cisplatin and gemcitabine in patients with Her2 overexpressing, untreated, advanced non-small cell lung cancer: report of a phase II trial and findings regarding optimal identification of patients with Her2-overexpressing disease. Lung Cancer 44:99–110

100 Years of Drug Delivery to the Lungs

Federico Lavorini, Francesca Buttini, and Omar S. Usmani

Contents

1 The History of Therapeutic Aerosols: From the Ancient Time to Present 144
2 Current Inhalation Delivery Systems 146
2.1 The Development of Modern HFA pMDIs 146
2.2 The Emergence of Modern DPIs 149
2.3 Nebuliser Systems and Soft Mist Inhaler 150
3 Advances in Aerosol Science 152
3.1 Particle Sizing Techniques and In Vitro Measurements 152
3.2 Imaging Aerosol Deposition Techniques 153
3.3 Pharmacokinetics and Pharmacodynamics 154
4 Looking to the Future 155
5 Conclusions 156
References 157

Abstract

Inhalation therapy is one of the oldest approaches to the therapy of diseases of the respiratory tract. It is well recognised today that the most effective and safe means of treating the lungs is to deliver drugs directly to the airways. Surprisingly, the delivery of therapeutic aerosols has a rich history dating back more than 2,000 years to Ayurvedic medicine in India, but in many respects, the introduction of the first pressurised metered-dose inhaler (pMDI) in 1956 marked the beginning of the modern pharmaceutical aerosol industry. The pMDI was

F. Lavorini (✉)
Department of Experimental and Clinical Medicine, University of Florence, Florence, Italy
e-mail: federico.lavorini@unifi.it

F. Buttini
Food and Drug Department, University of Parma, Parma, Italy

O. S. Usmani
National Heart and Lung Institute, Imperial College London and Royal Brompton Hospital, London, UK

J. E. Barrett et al. (eds.), *Concepts and Principles of Pharmacology*,
Handbook of Experimental Pharmacology 260, https://doi.org/10.1007/164_2019_335

the first truly portable and convenient inhaler that effectively delivered drug to the lung and quickly gained widespread acceptance. Since 1956, the pharmaceutical aerosol industry has experienced dramatic growth. The signing of the Montreal Protocol in 1987 to reduce the use of CFCs as propellants for aerosols led to a surge in innovation that resulted in the diversification of inhaler technologies with significantly enhanced delivery efficiency, including modern pMDIs, dry powder inhalers and nebuliser systems. There is also great interest in tailoring particle size to deliver drugs to treat specific areas of the respiratory tract. One challenge that has been present since antiquity still exists, however, and that is ensuring that the patient has access to the medication and understands how to use it effectively. In this article, we will provide a summary of therapeutic aerosol delivery systems from ancient times to the present along with a look to the future.

Keywords
Aerosol · Dry powder inhalers · Inhalation medicines · Metered dose inhaler · Nebulisers

1 The History of Therapeutic Aerosols: From the Ancient Time to Present

The delivery of therapeutic vapours and aerosols through inhalation has been used for thousands of years in various cultures. Although the term "aerosol" was coined at the beginning of the twentieth century, the use of therapeutic aerosols dates back at least 4,000 years (Stein and Thiel 2017). The origins of inhalation therapy for asthma and other lung diseases may have arisen in the traditional therapies of Ayurvedic medicine in India around 2000 BC. The compounds smoked for medicinal purposes included herbal preparations, most notably *Datura* species, which contain potent alkaloids with anticholinergic properties (Stein and Thiel 2017). An Egyptian papyrus dating back to around 1500 BC describes patients breathing the vapour of the black henbane plant, a herb with anticholinergic bronchodilating properties, after being thrown onto a hot brick (Sanders 2007). One of the earliest inhaler devices is attributed to Hippocrates that consisted of a pot with a reed through which the vapour could be inhaled. By the first century AD, native cultures from Central and South America fashioned pipes to smoke tobacco and other plants. It is believed that these cultures had identified the smoking of plants with anticholinergic properties such as *Datura*, henbane and belladonna as therapeutic remedies to treat respiratory conditions (Sanders 2007). Variations on the Hippocrates's pot-and-reed design were used in the late of the eighteenth and early nineteenth century. The modern era of aerosol therapy began in 1778 with the English physician John Mudge who coined in his book *A Radical and Expeditious Cure for a Recent Catarrhous Cough* the term "inhaler" and described his device for inhaling opium vapour for the treatment of cough (Mudge 1778). The Mudge inhaler, the first known example of a marketed inhaler device, consisted of a pewter tankard with a mouthpiece covering the top and an air passage drilled through

the handle, so that, by inhaling through the mouthpiece, a patient can draw air through the liquid at the bottom of the vessel (Stein and Thiel 2017). Several models of ceramic inhalers followed the design of the Mudge inhaler and were popular from the nineteenth century onward. The last half of the nineteenth century saw unprecedented innovation in the technologies developed by the pharmaceutical for aerosol delivery. The first pressurised inhaler was the Sales-Giron's Pulverisateur in 1858. Many other nebulisers, asthma cigarettes containing stramonium and powders, were introduced in the late nineteenth and early twentieth century, and attempts were made to administer a number of medications by aerosol (Stein and Thiel 2017). The Aerohalor, developed by Abbott Laboratories and launched in 1948 for inhaled penicillin G powder, was the first truly commercially successful dry power inhaler (DPI). The device utilised a steel ball that moved when the patient inhaled and tapped the cartridge that contained the drug to aerosolise the powder. The device was a breakthrough in terms of commercial viability of a DPI device in spite of the fact that it was relatively inefficient in terms of dispersing the powder into a respirable aerosol.

Just over 60 years ago, Charlie Thiel and colleagues at Riker Laboratories (now 3M Pharmaceuticals, St Paul, Minnesota, USA) invented the pressurised metered-dose inhaler (pMDI) after Susie Maison, the daughter of a Riker Vice-President asked, "Why can't you make my asthma medicine like mother's hair spray?". The pMDI was a revolution and, with minor modifications, is still the most popular form of aerosol delivery (Roche and Dekhuijzen 2016). There have been remarkable advances in the technology of devices and formulations for inhaled drugs in the past 50 years since the development of the first pMDI. Jet, ultrasonic and vibrating mesh nebulisers have advanced with devices that are breath-actuated or breath-enhanced. Some milestones in the development of inhaler therapy are shown in Fig. 1.

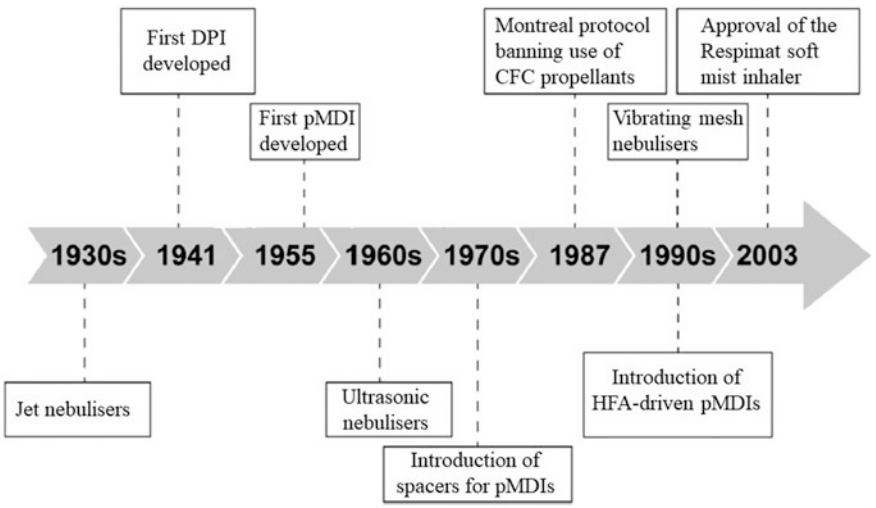


Fig. 1 Milestones in the development of inhaler therapy. Adapted from Iwanaga et al. (2019)

2 Current Inhalation Delivery Systems

Available inhalation devices include pMDIs (used with or without a spacer), DPIs, nebulisers and soft mist inhalers (SMIs). Each device type is associated with advantages and disadvantages, and these are summarised in Table 1.

2.1 The Development of Modern HFA pMDIs

First introduced in the 1950s, the pMDI is the inhaler most commonly used for drug delivery in the treatment of patients with asthma or chronic obstructive pulmonary disease (COPD). The pMDI consists of an aluminium canister, lodged in a plastic support, containing a pressurised suspension or solution of micronised drug particles dispersed in propellants (Roche and Dekhuijzen 2016). The key component of the pMDI is a metering valve, which delivers an accurately known volume of propellant, containing the micronised drug at each valve actuation. The operation principle of the present pMDIs remains similar to the original 1950s push-and-breath design:

Table 1 Strengths and weaknesses of each inhaler device type

	Strengths	Weaknesses
pMDI	Compact and portable Dose consistency Multidose Wide range of therapies	Contains propellant Requires good coordination of actuation and inhalation
pMDI + spacer	Easier to coordinate large drug doses delivered more conveniently Higher lung deposition than a pMDI alone less oropharyngeal deposition	Less portable than a pMDI plastic spacers may acquire static charge additional cost to a pMDI
DPI	Compact and portable Multidose Wide range of therapies No propellants Breath actuated (no coordination needed)	Moisture-sensitive Requires a minimum inspiratory flow Requires steps for preparation Dose inconsistency
SMI	Portable and compact multidose device High lung deposition does not contain propellants	Not breath actuated Requires some coordination of actuation and inhalation Requires priming before first use
Nebulisers	No specific inhalation technique required Vibrating mesh is portable and does not require an outside energy source High lung deposition (mesh nebulisers)	Jet and ultrasonic nebulisers require an outside energy source Treatment times can be long Performance varies between nebulisers Risk of bacterial contamination

pMDI pressurised metered-dose inhaler, *DPI* dry powder inhaler, *SMI* soft mist inhaler

pressing the bottom of the canister into the actuator seating causes decompression of the formulation within the metering valve, resulting in an explosive generation of a heterodisperse aerosol of droplets that consist of tiny drug particles contained within a shell of propellant (Fig. 2). The latter evaporates with time and distance, which reduces the size of the particles that use a propellant under pressure to generate a metered dose of an aerosol through an atomisation nozzle (Roche and Dekhuijzen 2016). Initially, pMDIs used chlorofluorocarbon (CFC) propellants, which were superseded by hydrofluoroalkane (HFA) propellants due to growing environmental concerns that CFC propellants were causing irreparable damage to the ozone layer in the atmosphere. Hydrofluoroalkanes were identified as a potential alternative, since they were considered inert with respect to environment. Although many of the physical properties of HFAs are similar to those of CFCs, direct translation of CFC formulations to HFA formulations was not possible. Historically, CFC formulations contained drug suspended in CFCs that were stabilised using surfactants. With the translation to HFA-based systems, it quickly became evident that the capacity for HFAs to solubilise these surfactants was not sufficient and thus a stable flocculated system could not be formed. Formulations of HFA-based pMDI systems are generally categorised as either suspension or solution technologies. Drug molecules conventionally used in pMDIs are not readily soluble in HFAs and thus require a co-solvent (Ganderton et al. 2002). Ethanol may be used as a co-solvent because it is miscible in HFAs and is also a good solvent for many hydrophobic pharmaceutical drugs. In general, solution-based pMDIs utilising volatile co-solvents result in higher fine-particle fractions, due to the small particle size of the dried aerosol. The particle size of solution-based pMDIs may be altered via the addition of non-volatile agents that are soluble in the HFA-co-solvent system; however, they will not evaporate during the aerosolisation process resulting in an increasing of final particle size (Buttini et al. 2014). Scanning electron microscopy images of particles generated from the glycerol-free and glycerol-containing formulations are shown in Fig. 3.

Some HFA-driven pMDIs have been extensively changed to obtain aerosol solution with small (mass median aerodynamic diameter ~1.3 μm) particle size (Lavorini et al. 2017). These pMDIs delivering small aerosol particles have shown

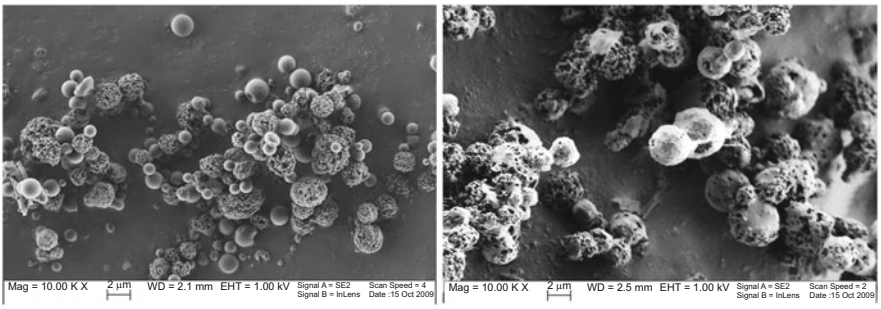


Fig. 2 The original pressurised metered-dose inhaler (pMDI) approved March 9, 1956 (left), and a modern pMDI with its main components (right)

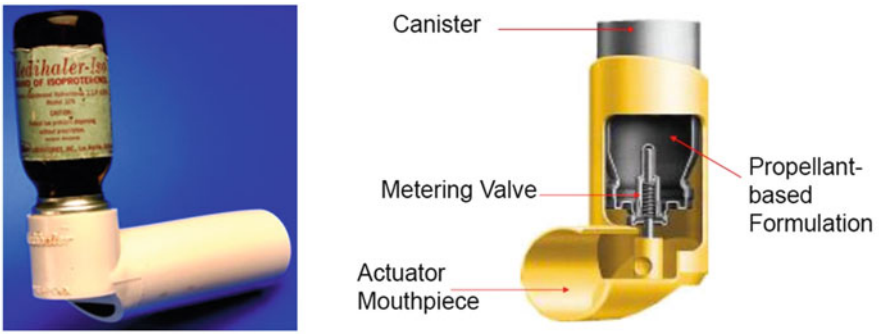


Fig. 3 Particles generated from the glycerol-free (left) and glycerol- (right) containing HFA pMDI formulations

to obtain a higher rate of pulmonary drug deposition than that achieved with the conventional pMDIs, i.e. those not emitting small aerosol particles (Lavorini et al. 2017). Generally, the velocity of the HFA spray is slower than that of the CFC, thus allowing a better distribution of the drugs along the respiratory airways and potentially making these devices more functional especially for elderly patients (Lavorini et al. 2016).

Correct use of the pMDI involves holding the inhaler in the correct position and performing a series of coordinated steps, and the complexity of this process can prove a challenge to some patients (Lavorini 2014). Suspension pMDIs also need to be shaken before use, a step commonly overlooked by both patients and healthcare professionals. In addition, for an efficient aerosol delivery to the lungs, pMDIs require slow, deep (i.e. an inspiratory flow rate of about 30 L/min roughly corresponding to a total inhalation time of 4–5 s) and steady inhalation starting just prior to device activation, with a subsequent short breath-hold of up to 10 s (Laube et al. 2011; Lavorini 2014). Unfortunately, most patients are not able to coordinate inhaler activation with inspiration and/or struggle to generate a deep enough inhalation, inhale too fast and/or fail to hold their breath for long enough, even after repeated tuition (Lavorini 2014). Importantly, misuse of pMDIs is associated with poorer asthma control, an increased number of exacerbations (Price et al. 2017) and worsening of COPD outcomes (Molimard et al. 2017).

To overcome the problems associated with poor pMDI use, spacers (Lavorini and Fontana 2009) and breath-actuated pMDIs are available (Lavorini et al. 2014). Spacers can be added to a pMDI to overcome problems with coordination and in doing so help to increase aerosol delivery to the peripheral airways. Spacers that feature a one-way inspiratory valve are termed valved holding chambers (VHCs). Spacers and VHCs can increase pulmonary deposition compared with pMDIs alone by reducing the velocity of the aerosol and filtering out larger, non-respirable particles (Lavorini and Fontana 2009). Breath-actuated pMDIs are useful for patients who struggle to time their inspiration properly, as they are triggered by airflow upon inspiration, although they still require an inspiratory flow rate of approximately 30 L/min and do not overcome the other disadvantages associated with pMDIs (Lavorini et al. 2014).

2.2 The Emergence of Modern DPIs

Much like the pMDI, DPIs are small, portable and widely available as either single-dose or multiple-dose devices (De Boer et al. 2017; Laube et al. 2011; Levy et al. 2019). At variance with pMDIs, all DPIs require a pre-inhalation dose-loading step to be completed successfully in order for them to function correctly. DPIs are actuated and driven by patient's inspiratory flow that drives the drug delivery; consequently, DPIs do not require coordination of inhaler actuation with inhalation thus resulting relatively simple to use for the majority of patients (De Boer et al. 2017; Laube et al. 2011; Levy et al. 2019). Most DPIs are formulated with their drug particles attached to excipient carrier molecules, such as lactose, or in the form of agglomerated pellets (De Boer et al. 2017). Consequently, DPIs are designed with an internal resistance that must be overcome by a forceful inhalation in order to generate a turbulent flow, de-aggregate the drug particles within, and produce fine particles for inhalation (De Boer et al. 2017; Laube et al. 2011; Levy et al. 2019). The currently available DPIs have varying internal resistance to airflow, which can be classified by the inhalation flow required to produce a 4 kPa pressure drop (Fig. 4). The force required to overcome the internal resistance, create a turbulent energy and generate an aerosol is the product of patient inhalation flow and the internal resistance of the device (Laube et al. 2011; Azouz and Chrystyn 2012). Subsequent lung deposition is a trade-off between generating sufficient power for particle de-aggregation and avoiding the increased oropharyngeal deposition that can occur at higher aerosol velocities (Azouz and Chrystyn 2012; De Boer et al. 2017). Therefore, a limitation of DPIs is their reliance on patients generating the necessary inspiratory force to de-aggregate the powder formulation into small respirable particles as efficiently as possible and, consequently, to ensure that the drug is delivered to the lungs (De Boer et al. 2017; Laube et al. 2011; Levy et al. 2019). The ability of certain patient populations to generate the required inspiratory force may impact an inhaler's efficacy thus substantially fine-particle dose delivered. Although most patients are capable of generating enough flow to

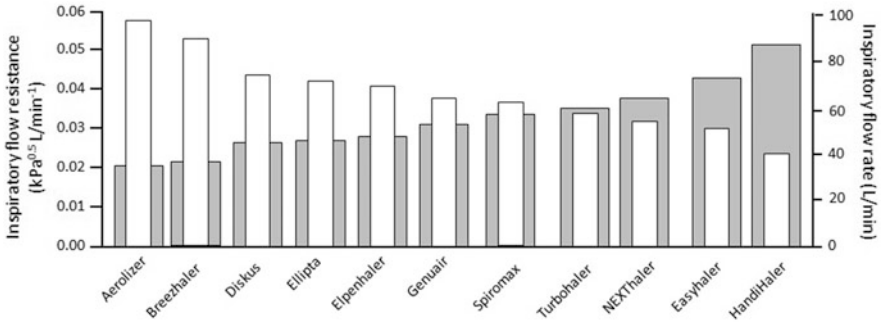


Fig. 4 Inspiratory resistance of dry powder inhalers (filled bars) and the corresponding flow (empty bars) required to achieve a 4 kPa pressure drop. See text for further details. Adapted from Lavorini et al. (2014)

operate a DPI efficiently, the need to inhale forcefully and, consequently, generate a sufficient inspiratory flow could be a problem for children aged <6 years or patients with severe airway obstruction (Laube et al. 2011).

Pulmonary administration of bronchodilators or corticosteroids as dry powders is primarily used to treat chronic obstructive airway diseases, such as asthma and COPD. However, in recent years an increasing interest has developed in the delivery of other types of drugs such as antibiotics to treat pulmonary infections. These medications have to be administered in higher doses (up to 200 mg) to achieve their therapeutic effect with limited amounts of excipients. As such, these types of formulations are regarded as "high powder dose drugs" (Sibum et al. 2018). Inhaled colistin (Colobreathe® Teva Pharmaceuticals Europe) and tobramycin (Tobi® Mylan Product Limited UK) are examples of high powder dose antibiotics for the treatment of cystic fibrosis patients. Notably, the efficacy of these inhaled antibiotics depends on the capability of the DPI device to load a consistent amount of powder and to modulate its release. For instance, the RS01 DPI (Plastiape Osnago, Italy), due to its delivery mechanism based on the spinning of the capsule, has shown to control the amount of a tobramycin powder emitted during the inhalation (Buttini et al. 2018a, b). Other DPIs releasing high dose of antibiotics are the Podhaler® (Mylan Product Limited UK), the Turbospin® (Forest Laboratories UK) the Orbital® (Pharmaxis, Australia), the Twincer® (Indes, the Netherlands) and the Cyclops® (PureIMS, The Netherlands) (Hoppentocht et al. 2015; Sibum et al. 2018).

2.3 Nebuliser Systems and Soft Mist Inhaler

Nebulisers are devices that convert a liquid in solution or suspension into small easily inhaled droplets. Solutions are comprised of drug dissolved in a carrier liquid, whereas suspensions are comprised of solid drug particles suspended in the carrier liquid. It is far more appropriate to refer to the "nebuliser system" in its entirety in which several components, other than the nebuliser itself, play a significant role in influencing aerosol delivery and its characteristics (Dolovich 2002). Among the components, the most influential are the compressor or line feed applied to the nebuliser, the volume fill, the residual volume and the driving gas flow and the use of a face mask or mouthpiece. Each of these significantly affects the total amount of drug received by the patients during therapy, the rate of nebulised aerosol output and the particle size distribution. If any component of the nebuliser system is replaced by another, then the nebuliser system has changed, and the aerosol output characteristics will have been significantly altered (Dolovich 2002).

Basically, there are three types of nebulisers: the jet nebuliser, the ultrasonic wave nebuliser and the vibrating mesh nebuliser.

2.3.1 Jet Nebulisers

Jet nebulisers are by far the most common type of nebulisers used worldwide. A jet nebuliser is powered by a compressed gas (usually air or oxygen) that draws medication through a capillary tube in the nebuliser' chamber, shearing the liquid

formulation and directing it to a baffle-generating aerosol. Coarse droplets impact on baffles, while smaller droplets may be inhaled or may land on internal walls returning to the reservoir for re-nebulisation (O'Callaghan and Barry 1997; Boe et al. 2001). There are four different designs of jet nebuliser: jet nebulisers with a reservoir tube, jet nebulisers with a collection bag, breath-enhanced nebulisers and breath-actuated jet nebulisers (O'Callaghan and Barry 1997; Boe et al. 2001). Jet nebulisers with a reservoir tube provide continuous aerosol generation during the entire breathing cycle, causing the release of aerosol to ambient air during exhalation and anytime when the patient is not breathing. Jet nebulisers with a collection bag generate aerosols by continuously filling a collection bag that acts as a reservoir. Both the breath-enhanced and breath-actuated jet nebulisers are modifications of the "conventional" jet nebulisers specifically designed to improve efficiency by increasing the amount of aerosol delivered to the patient with less wastage of aerosol during exhalation (O'Callaghan and Barry 1997; Boe et al. 2001).

2.3.2 Ultrasonic Wave Nebulisers

Ultrasonic wave nebulisers use a rapidly (>1 MHz) vibrating piezoelectric crystal to produce aerosol particles (O'Callaghan and Barry 1997; Boe et al. 2001). Ultrasonic vibrations from the crystal are transmitted to the surface of the drug solution where standing waves are formed. Droplets break free from the crest of these waves and are released as aerosol. The size of droplets produced by an ultrasonic nebuliser is related to the frequency of oscillation (O'Callaghan and Barry 1997; Boe et al. 2001). Although ultrasonic nebulisers can nebulise solutions more quickly than jet nebulisers, they are not suitable for suspensions, and the piezoelectric crystal can heat the drug to be aerosolised which can limit its use for heat-sensitive molecules.

2.3.3 Vibrating Mesh Nebulisers

The most recent innovation was made around 2005, with creation of the ultrasonic vibrating mesh technology. These modern nebuliser systems have been developed to address some of the limitations of conventional air jet nebulisers, in particular the long treatment time and inefficient utilisation of drug. The integral component of mesh nebulisers is a vibrating mesh plate, or aperture plate, that possesses precision-formed holes that control the size and flow of the aerosolised particles. An attached or separate power supply provides electricity to a vibrating piezoelectric element. As the plate begins to vibrate, the drug passes through the holes, thus producing a dense aerosol at low flow rates (Skaria and Smaldone 2010; Dhand 2002). Vibrating mesh devices such as the AeroNeb® Pro (Aerogen, Ireland) and the eFlow® (Pari, Germany) have a number of advantages over other nebuliser systems: they have greater efficiency, precision and consistency of drug delivery and are quieter and generally portable (Skaria and Smaldone 2010; Dhand 2002). On the downside, they are significantly more expensive than other types of nebuliser and require a significant amount of maintenance and cleaning after each use to prevent build-up of deposit and blockage of the apertures, especially when suspensions are aerosolised, and also to prevent colonisation by pathogens (Dhand 2002).

2.3.4 The Respimat Soft Mist Inhaler

The development of the "soft mist inhaler" (SMI) falls within the definition of a nebuliser, as SMIs transform aqueous liquid solution to liquid aerosol droplets suitable for inhalation. However, at variance with the traditional nebuliser designs, they are handheld multidose devices that have the potential to compete with both pMDIs and DPIs in the portable inhaler market. At present, the only SMI currently marketed is the Respimat® (Boehringer Ingelheim, Germany). This device does not require propellants since it is powered by the energy of a compressed spring inside the inhaler. Individual doses are delivered via a precisely engineered nozzle system as a slow-moving aerosol cloud, hence the term "soft mist" (Dalby et al. 2004; Iwanaga et al. 2019). Scintigraphic studies have shown that, compared to a CFC-based pMDI, lung deposition is higher (up to 50%) and oropharyngeal deposition is lower (Dalby et al. 2004; Iwanaga et al. 2019). Respimat is a "press-and-breathe" device, and the correct inhalation technique closely resembles that used with a pMDI. However, although coordination between firing and inhaling is required, the aerosol emitted from the Respimat is released very slowly, with a velocity of approximately four times less than that observed with a CFC-driven pMDI (Dalby et al. 2004). This greatly reduces the potential for drug impaction in the oropharynx. In addition, the relatively long duration over which the dose is expelled from the Respimat (about 1.2 s compared with 0.1 s from traditional pMDIs) would be expected to greatly reduce the need to coordinate actuation and inspiration, thus improving the potential for greater lung deposition (Iwanaga et al. 2019).

3 Advances in Aerosol Science

3.1 Particle Sizing Techniques and In Vitro Measurements

An aerosol can be defined as a system of solid particles or liquid droplets that can remain dispersed in a gas, usually air (Bisgaard et al. 2002). Naturally occurring aerosols, as well as those emitted by clinical aerosol generators, almost always contain a wide range of particle sizes. Because the aerodynamic behaviour of an aerosolised particle is critically influenced by its mass, it is important to precisely describe the size distribution of aerosolised particles. In clinical studies, the mass median aerodynamic diameter (MMAD) and the geometric standard deviation (GSD) are often used to characterise the dimension of an aerosol (Bisgaard et al. 2002). When the mass distribution of particles in an aerosol is fractioned and the cumulative particle distribution plotted as a lognormal distribution on probability paper, it often approximates a straight line. The MMAD represents the point in the distribution above which 50% of the mass resides, expressed as the diameter of a unit density sphere having the same terminal setting velocity as the aerosol particle in question, regardless of its shape and density (Bisgaard et al. 2002). The GSD is an indicator of the variability in particle diameters. If the particle size varies over a wide range (i.e. GSD $>$ 1.2), it is describe as having polydisperse particle distribution;

if the particles are of similar size (i.e. GSD < 1.2), the particle distribution is described as monodisperse. Monodisperse aerosols are usually encountered in research studies, whereas clinical aerosols are widely polydispersed (Bisgaard et al. 2002).

Particle size is an important factor in determining whether a particle will undergo nasopharyngeal, airway or alveolar deposition (Ziegler and Wacthel 2005). Methods for determination of particle size distribution are the light scattering or cascade impaction. The former is based on the principle that there is differential scattering of polarised light by particles of different size. In cascade impaction, particles at a set flow rate go through a series of apertures of decreasing diameter and impact on a series of plates if they fail to follow the air stream. The cascade impactor has been adopted as the method of choice for monitoring quality control in the manufacture of formulations for aerosol delivery, comparison of devices, and they can be used to estimate the amount of deposition in the respiratory tract. Characteristically, the cascade impactor is used to quantify the respirable fraction or fine-particle dose (usually the percentage of particles <5 μm diameter) as an estimate of lung delivery. Recommendations are available for assessment of particle size distributions and mass output of nebulisers, MDIs and DPIs. In vitro systems have been added to particle sizing devices in ways that more closely simulate the clinical scenario. Anatomic throats have been used with impactors instead of standard inlet manifolds. Radiolabelled aerosols have been delivered to anatomic lung models using simulated breathing patterns. Other measurements of "inhaled mass" from a nebuliser have used a patient or patient surrogate (piston pump) breathing from a nebuliser through filters.

Meaningful comparisons of the sizes of clinical aerosols should be compared only if obtained with identical techniques. Nevertheless, despite the technical difficulties encountered in measuring the size of polydisperse clinical aerosols, investigators have established that, when used with appropriate caution, data obtained by in vitro measurement of particle size do provide useful predictive data for subsequent clinical studies (Smaldone and Solomita 2009).

3.2 Imaging Aerosol Deposition Techniques

Several imaging modalities have been employed to quantify lung dose and the distribution of the dose of orally inhaled aerosols in vivo. Two-dimensional (2D or planar) imaging using gamma scintigraphy is the most widely used of these modalities. The gamma camera, invented by Anger in 1958, was used from the late 1970s to assess drug delivery to various organs, including the lungs. Two-dimensional gamma scintigraphy studies are accomplished using a single- or dual-headed gamma camera (Newman et al. 2003). The formulation to be tested is admixed with the gamma emitting radioisotope ^{99m}technetium, which serves as a surrogate for the drug. With this technique, total deposition should be assessed after identification of the right lung border and appropriate correction for tissue attenuation. Regional deposition should be quantified as a normalised outer/inner

deposition ratio and expressed as the penetration index (Newman et al. 2012). More recently, pulmonary drug delivery has been assessed with the three-dimensional imaging methods of single-photon emission computed tomography (SPECT) and positron emission tomography (Newman et al. 2003). SPECT is slightly superior to planar imaging for measuring total lung deposition. However, it is more complex to use, and for studies where total lung deposition is the endpoint, planar imaging is recommended. However, SPECT has been shown to be clearly superior to planar imaging for assessing regional distribution of aerosol and is the method of choice for this purpose. It therefore has applications in studying the influence of regional deposition on clinical effectiveness and also in validating computer models of deposition (Fleming et al. 2012).

3.3 Pharmacokinetics and Pharmacodynamics

Another approach to assessing respiratory drug delivery is to measure plasma levels of drug after absorption (pharmacokinetics) and to relate those levels to clinical efficacy and toxicity (pharmacodynamics) (Chrystyn 2001). The pharmacokinetic profile of a drug after inhalation may differ quite markedly from that seen after dosing by other routes of administration. Drugs may be administered to the lung to elicit a local action or as a portal for systemic delivery of the drug to its site of action elsewhere in the body (e.g. insulin). Some knowledge of pharmacokinetics is important for both locally and systemically acting drugs. For a systemically acting drug, the plasma concentration-time profile shares some similarities with drugs given by the oral or intravenous routes, since the plasma concentrations (after the distribution phase) will be in equilibrium with concentrations at the site of action. However, for a locally acting drug, such as an inhaled medication for the treatment of asthma or COPD, the plasma concentrations reflect its fate after it has been absorbed and removed from the airways, and not what is available to its site of action in the lung. Consequently, typical pharmacokinetic parameters which are determined from plasma concentration measurements (e.g. area under the curve, Cmax and tmax) may provide information on the deposition and absorption of drugs from the lung; however, the information from these parameters becomes more complicated to decipher for those drugs which are locally acting in the lung, and systemic levels are often used as a marker of toxicity (e.g. plasma levels of corticosteroids following inhalation). For instance, determination of pharmacokinetic profiles is difficult for inhaled drugs, because the low plasma levels require a sensitive assay and may be altered by drug absorbed from the gastrointestinal tract. In many cases, it is important to distinguish the relative contributions of lung and gastrointestinal tract absorption, as drug absorbed from the lung can be used as a surrogate for deposition. The influence of physiological and pathological factors needs also to be considered in the absorption of some inhaled drugs (Chrystyn 2001). The absorption of some hydrophilic drugs is influenced by the inspiratory manoeuver used during

initial inhalation of the drug and at later times after deposition. Similarly, the effects of smoking have been shown to increase lung permeability and increase the absorption of certain hydrophilic drugs (Chrystyn 2001).

4 Looking to the Future

Over the past decade, the efficiency of inhalers, as measured by total lung deposition, has increased from less than 10% to nearly 50% of the total dose, yet less than half the dose becomes available to the site of absorption (Hoppentocht et al. 2015). There is space for new technologies to improve these numbers, and pharmaceutical companies are continuously trying to innovate and improve on the existing inhalation technologies available. The incorporation of modern technology into inhaler devices is chiefly aimed at improving drug delivery, reducing device errors, improving patient adherence and monitoring and managing patients' disease state (Rogueda and Traini 2016).

New co-suspension technology uses low-density phospholipid particles to suspend micronised drug crystals in an HFA propellant, meaning multiple drugs can be administered via a single pMDI in a uniform manner (Ferguson et al. 2018). The low-density phospholipid particles increase the physiochemical stability of the drugs and can also reduce the effects of a shake-fire delay. In vitro and in vivo tests have shown highly reproducible, consistent drug delivery and effective lung deposition (Taylor et al. 2016). This was maintained across variations in flow rate, and drug delivery was constant under conditions of simulated patient handling errors, such as variable shake technique and delays between shaking and actuation (Doty et al. 2018).

One of the solutions envisaged to increase patient adherence to their therapies is the use of digital health solutions such as monitoring systems based on phone applications (apps) and electronic sensors. The first in-built inhaler monitoring technology was developed in the 1980s, mainly to assess adherence to medication, and this has evolved over the years to incorporate various other sensing functionalities (Kikidis et al. 2016). Development of the Smart Inhaler Tracker (Adherium) to store the dates and times of inhaler actuations led to the development of more sophisticated devices that incorporates a Global Positioning System (GPS) or functions capable of monitoring parameters such as inhalation flow and volume (Kikidis et al. 2016). The incorporation of dose-memory and dose-reminder functions in inhalers can have a positive effect on adherence and can increase confidence in self-management behaviour (Foster et al. 2017). In the 12-month STAAR study in children with asthma, for example, clinical review of electronic adherence monitoring data and dose reminders were shown to improve average adherence and reduce the number of courses of oral steroids and hospital admissions compared to non-review and no reminder function (Morton et al. 2017). In a randomised controlled trial in children with asthma, an electronic monitoring device with an audiovisual reminder function led to significant improvements in adherence to inhaled medications (Chan et al. 2015).

Digital health developments have also shown great utility in the management of device errors and are now able to provide detailed feedback on patients' device competence (Kikidis et al. 2016). The SmartMist™ (Aradigm) and MDILog™ (Westmed Technologies) have both included sensing capabilities to facilitate the assessment of inhalation technique. The MDILog™, which is widely used in clinical research, is designed to attach to the plastic casing of standard inhalers. The device includes an inhaler actuation sensor, as well as an accelerometer for the detection of inhaler shaking and a sensitive temperature sensor for the assessment of inhalation. Inhalation detection technologies can be used to coach patients on correct device technique. This kind of technology, along with other innovative e-health developments, such as mobile communication technology (mHealth), electronic reminders, telemedicine and inhaler tracker interventions, has the potential to reduce the resource burden on healthcare systems and provide optimal and personalised asthma management to patients (Bonini and Usmani 2018).

5 Conclusions

The pulmonary route of administration has proven to be effective in local and systemic delivery of miscellaneous drugs and biopharmaceuticals to treat pulmonary and non-pulmonary diseases. A successful pulmonary administration requires a harmonic interaction between the drug formulation, the inhaler device and the patient. Inhalation products are more complex compared to the conventional dosage forms: device and formulation interconnecting features should work together to aerosolise the drug and deliver it to the site of action. In this perspective, the development of inhalation products should be done using a Quality by Design approach in order to fully understand the interaction of the two product components and their contribution on the product performance enabling the minimisation of the patient errors during the device preparation and inhalation (Buttini et al. 2018a). In 2011 there were more than 230 different drug-device combinations available in Europe (Lavorini et al. 2011). Hoppentocht et al. (2015) list 32 different device technologies recently developed or in development currently devices technologies, and this count does not register recent generic devices. These numbers speak of inventiveness of technologists, but some of these inventions fail to translate into clinical improvements. Indeed, the biggest problem that accounts for the lack of desired effect or adverse outcomes is the incorrect use of the device. In addition, the myriad of devices and complex physics involved in the delivery often makes it difficult for clinicians to choose the right device for their patients. An ideal inhaler should deliver precise and consistent doses to a targeted region in the lungs and maintain the stability of the delivered drugs. It is also desirable that devices are small and simple enough to be easily used by patients. Dry powder inhalers are becoming more popular because of their ease of use; pMDIs are still facing challenges from the formulation and the design point of view. Nebulisers are being remodelled to broaden their applicability. Currently, there is no single device that fulfills the myriad of requirements to optimally deliver drugs having different physicochemical

properties, and therefore healthcare professionals need to fully understand the capabilities of each inhaler and relate that to the needs of the patient, according to their health status, to achieve the best therapeutic outcome.

References

Azouz W, Chrystyn H (2012) Clarifying the dilemmas about inhalation techniques for dry powder inhalers: integrating science with clinical practice. Prim Care Respir J 21:208–213

Boe J, Dennis JH, O'Driscoll BR et al (2001) European Respiratory Society guidelines on the use of nebulizers. Eur Respir J 18:228–242

Bonini M, Usmani OS (2018) Novel methods for device and adherence monitoring in asthma. Curr Opin Pulm Med 24:63–69

Buttini F, Miozzi M, Balducci AG, Royall PG, Brambilla G, Colombo P, Bettini R, Forbes B (2014) Differences in physical chemistry and dissolution rate of solid particle aerosols from solution pressurised inhalers. Int J Pharm 465(1–2):42–51

Buttini F, Rozou S, Rossi A, Zoumpliou V, Rekkas DM (2018a) The application of quality by design framework in the pharmaceutical development of dry powder inhalers. Eur J Pharm Sci 113:64–76

Buttini F, Balducci AG, Colombo G, Sonvico F, Montanari S, Pisi G, Rossi A, Colombo P, Bettini R (2018b) Dose administration maneuvers and patient care in tobramycin dry powder inhalation therapy. Int J Pharm 548:182–191

Chan AHY, Stewart AW, Harrison J et al (2015) The effect of an electronic monitoring device with audiovisual reminder function on adherence to inhaled corticosteroids and school attendance in children with asthma: a randomised controlled trial. Lancet Respir Med 3:210–219

Chrystyn H (2001) Methods to identify drug deposition in the lungs following inhalation. Br J Clin Pharmacol 51:289–299

Dalby R, Spallek M, Voshaar T (2004) A review of the development of Respimat Soft Mist Inhaler. Int J Pharm 283:1–9

De Boer AH, Hagedoorn P, Hoppentocht M, Buttini F, Grasmeijer F, Frijlink HW (2017) Dry powder inhalation: past, present and future. Expert Opin Drug Deliv 14(4):499–512

Dhand R (2002) Nebulisers that use a vibrating mesh or plate with multiple apertures to generate aerosol. Respir Care 47:1406–1416

Dolovich MB (2002) Assessing nebulizer performance. Respir Care 47:1290–1301

Doty A, Schroeder J, Vang K et al (2018) Drug delivery from an innovative LAMA/LABA co-suspension delivery technology fixed-dose combination MDI: evidence of consistency, robustness, and reliability. AAPS PharmSciTech 19:837–844

Ferguson GT, Hickey AJ, Dwivedi S (2018) Co-suspension delivery technology in pressurized metered-dose inhalers for multi-drug dosing in the treatment of respiratory diseases. Respir Med 134:16–23

Fleming J, Bailey DL, Chan H-K, Conway J, Kuehl PJ, Laube BL et al (2012) Standardization of techniques for using single-photon emission computed tomography (SPECT) for aerosol deposition assessment of orally inhaled products. J Aerosol Med Pulm Drug Deliv 25(Suppl 1): S29–S51

Foster JM, Reddel HK, Usherwood T et al (2017) Patient-perceived acceptability and behaviour change benefits of inhaler reminders and adherence feedback: a qualitative study. Respir Med 129:39–45

Ganderton D, Lewis D, Davies R, Meakin B, Brambilla G, Church T (2002) Modulite: a means of designing the aerosols generated by pressurized metered dose inhalers. Respir Med 96(Suppl D):S3–S8

Hoppentocht M, Akkerman OW, Hagedoorn P, Frijlink HW, De Boer AH (2015) The Cyclops for pulmonary delivery of aminoglycosides; a new member of the Twincer™ family. Eur J Pharm Biopharm 90:8–15

Iwanaga T, Tohda Y, Nakamura S, Suga Y (2019) The Respimat® soft mist inhaler: implications of drug delivery characteristics for patients. Clin Drug Investig. https://doi.org/10.1007/s40261-019-00835-z

Kikidis D, Konstantinos V, Tzovaras D et al (2016) The digital asthma patient: the history and future of inhaler based health monitoring devices. J Aerosol Med Pulm Drug Deliv 29:219–232

Laube BL, Janssens HM, de Jongh FH, Devadason SG, Dhand R, Diot P, Everard ML, Horvath I, Navalesi P, Voshaar T, Chrystyn H, European Respiratory Society; International Society for Aerosols in Medicine (2011) What the pulmonary specialist should know about the new inhalation therapies. Eur Respir J 37:1308–1331

Lavorini F (2014) Inhaled drug delivery in the hands of the patient. J Aerosol Med Pulm Drug Deliv 27:414–418

Lavorini F, Fontana GA (2009) Targeting drugs to the airways: the role of spacer devices. Expert Opin Drug Deliv 6:91–102

Lavorini F, Corrigan CJ, Barnes PJ, Dekhuijzen PR, Levy ML, Pedersen S, Roche N, Vincken W, Crompton GK (2011) Retail sales of inhalation devices in European countries: so much for a global policy. Respir Med 105:1099–1103

Lavorini F, Fontana GA, Usmani OS (2014) New inhaler devices – the good, the bad and the ugly. Respiration 88:3–15

Lavorini F, Mannini C, Chellini E, Fontana GA (2016) Optimising inhaled pharmacotherapy for elderly patients with chronic obstructive pulmonary disease: the importance of delivery devices. Drugs Aging 33:461–473

Lavorini F, Pedersen S, Usmani OS (2017) Dilemmas, confusion, and misconceptions related to small airways directed therapy. Chest 151:1345–1355

Levy ML, Carroll W, Izquierdo Alonso JL, Keller C, Lavorini F, Lehtimäki L (2019) Adv Ther. https://doi.org/10.1007/s12325-019-01066-6

Molimard M, Raherison C, Lignot S, Balestra A, Lamarque S, Chartier A, Droz-Perroteau C, Lassalle R, Moore N, Girodet PO (2017) Chronic obstructive pulmonary disease exacerbation and inhaler device handling: real-life assessment of 2935 patients. Eur Respir J 49. pii: 1601794

Morton RW, Elphick HE, Rigby AS et al (2017) STAAR: a randomised controlled trial of electronic adherence monitoring with reminder alarms and feedback to improve clinical outcomes for children with asthma. Thorax 72:347

Mudge J (1778) A radical and expeditious cure for a recent Catarrhous cough, 2nd edn. E. Allen, London, pp 131–147

Newman SP, Pitcairn GR, Hirst PH, Rankin L (2003) Radionuclide imaging technologies and their use in evaluating asthma drug deposition in the lungs. Adv Drug Deliv Rev 55:851–867

Newman S, Bennett WD, Biddiscombe M, Devadason SG, Dolovich MB, Fleming J et al (2012) Standardization of techniques for using planar (2D) imaging for aerosol deposition assessment of orally inhaled products. J Aerosol Med Pulm Drug Deliv 25(Suppl 1):S10–S28

O'Callaghan C, Barry PW (1997) The science of nebulised drug delivery. Thorax 52(Suppl 2): S31–S44

Price DB, Román-Rodríguez M, McQueen RB, Bosnic-Anticevich S, Carter V, Gruffydd-Jones K, Haughney J, Henrichsen S, Hutton C, Infantino A, Lavorini F, Law LM, Lisspers K, Papi A, Ryan D, Ställberg B, van der Molen T, Chrystyn H (2017) Inhaler errors in the CRITIKAL study: type, frequency, and association with asthma outcomes. J Allergy Clin Immunol Pract 5:1071–1081

Roche N, Dekhuijzen PNR (2016) The evolution of pressurized metered-dose inhalers from early to modern devices. J Aerosol Med Pulm Drug Deliv 29:1–17

Rogueda P, Traini D (2016) The future of inhalers: how can we improve drug delivery in asthma and COPD? Expert Rev Respir Med 10:1041–1044

Sanders M (2007) Inhalation therapy: an historical review. Prim Care Respir J 16:71–81

Sibum I, Hagedoorn P, de Boer AH, Frijlink HW, Grasmeijer F (2018) Challenges for pulmonary delivery of high powder doses. Int J Pharm 548:325–336

Skaria S, Smaldone GC (2010) Omron NE U22: comparison between vibrating mesh and jet nebulizer. J Aerosol Med Pulm Drug Deliv 23:173–180

Smaldone GC, Solomita M (2009) Predicting in vivo deposition in vitro. J Aerosol Med Pulm Drug Deliv 22:9–10

Stein SW, Thiel CG (2017) The history of therapeutic aerosols: a chronological review. J Aerosol Med Pulm Drug Deliv 30:20–41

Taylor G, Warren S, Orevillo C et al (2016) Pulmonary deposition of a novel LAMA/LABA co-suspension technology glycopyrronium/formoterol (GFF) MDI in healthy subjects. Eur Respir J 48:OA4834

Ziegler J, Wacthel H (2005) Comparison of cascade impactor and laser diffraction for particle size. J Aerosol Med 18:311–324

Ion Channel Pharmacology for Pain Modulation

Francesco De Logu and Pierangelo Geppetti

Contents

1 General Introduction 162
2 Calcium Channels 162
2.1 Introduction to Calcium Channels 162
2.2 LTCC and Pain 163
2.3 NTCC and Pain 164
2.4 $\alpha_2\delta$-1, Inhibitors and Pain 164
3 Transient Receptor Potential (TRP) Channels 165
3.1 Introduction to TRP Channels 165
3.2 TRPV1 and Pain 165
3.3 TRPA1 and Pain 167
3.4 TRPA1 and Chronic Neuropathic Pain 167
3.5 Other TRP Channels and Pain 169
4 Acid-Sensing Family of Ion Channels (ASICs) 170
4.1 Introduction to ASICs 170
4.2 Peripheral ASICs and Pain 171
5 PIEZO Channels 173
5.1 Introduction to PIEZO Channels 173
5.2 PIEZO and Pain Sensation 174
6 Purinergic $P2X_3$ Ion Channels 175
6.1 Introduction to $P2X_3$ Ion Channels 175
6.2 $P2X_3$ Ion Channels in Pain and Cough 176
References 177

Abstract

A large series of different ion channels have been identified and investigated as potential targets for new medicines for the treatment of a variety of human diseases, including pain. Among these channels, the voltage gated calcium

F. De Logu · P. Geppetti (✉)
Department of Health Sciences, Section of Clinical Pharmacology and Oncology, University of Florence, Florence, Italy
e-mail: geppetti@unfi.it

J. E. Barrett et al. (eds.), *Concepts and Principles of Pharmacology*,
Handbook of Experimental Pharmacology 260, https://doi.org/10.1007/164_2019_336

channels (VGCC) are inhibited by drugs for the treatment of migraine, neuropathic pain or intractable pain. Transient receptor potential (TRP) channels are emerging as important pain transducers as they sense low pH media or oxidative stress and other mediators and are abundantly found at sites of inflammation or tissue injury. Low pH may also activate acid sensing ion channels (ASIC) and mechanical forces stimulate the PIEZO channels. While potent agonists of TRP channels due to their desensitizing action on pain transmission are used as topical applications, the potential of TRP antagonists as pain therapeutics remains an exciting field of investigation. The study of ASIC or PIEZO channels in pain signaling is in an early stage, whereas antagonism of the purinergic $P2X_3$ channels has been reported to provide beneficial effects in chronic intractable cough. The present chapter covers these intriguing channels in great detail, highlighting their diverse mechanisms and broad potential for therapeutic utility.

Keywords
Acid-sensing ion channels · Calcium · Ion channels · Pain · PIEZO channels · TRP

1 General Introduction

Ion channels, due to their prominent localization in primary sensory neurons and other key structures in pain processing, are regarded as a major class of drug targets for modulating pain sensation and controlling chronic pain. Here, we review the implications of some ion channels in pain transduction and the preclinical and clinical investigations on drugs that target such channels in pain studies. The discussion will be limited to voltage-gated calcium channels, transient receptor potential channels, acid-sensing ion channels, and PIEZO channels. Descriptions of other important channels in pain signaling, including voltage-sensitive sodium channels, are reported elsewhere in this volume.

2 Calcium Channels

2.1 Introduction to Calcium Channels

Increases in calcium plasma concentrations are regulated by a variety of mechanisms and channels, of which one prominent class of proteins encompasses voltage-gated calcium channels (VGCCs). In excitable cells, the functions controlled by VGCCs include excitation–contraction coupling (Bannister and Beam 2013), excitation–transcription coupling (Wheeler et al. 2012), and transmitter release and hormone excretion (Catterall et al. 2013). VGCCs have been classified in the high voltage activated L-type CC (LTCC, long lasting, $Ca_v1.1$-$Ca_v1.4$), P/Q-type CC (PTCC, Purkinje, $Ca_v2.1$), N-type CC (NTCC, neuronal, $Ca_v2.2$), and R-type CC (RTCC, residual, $Ca_v2.3$), and low voltage activated T-type CC (TTCC, transient, $Ca_v3.1-Ca_v3.3$) (Zamponi et al. 2015). VGCCs have a complex structure formed

by the association of a variety of subunits ($\alpha_2\delta$, β, and δ) with the α_1 subunit differently encoded by a series of genes (*CACNA1A*, *1B*, *1C*, *1D*, *1E*, *1F*, *1G*, and *1S*), which are expressed with remarkable variability in different tissues and cells.

The different distribution of these channels accounts for the specificity expressed in skeletal muscle, smooth muscle, Purkinje cells, neurons, and many other cells. In addition, such diversity in structure and localization has offered the unique opportunity to identify drugs that, by targeting VGCC preferentially expressed in specific tissues/cells, exhibit remarkable selectivity. The role of VGCCs in different types of pain derives from both their heterogeneity in structure and tissue and cell localization. Although mutations of the various VGCCs have been linked to a series of diseases unrelated to pain, a mutation of the $Ca_v2.1$ (P/Q VGCC) was associated with familial hemiplegic migraine type 1 (FHM1), which is a very rare condition characterized by migraine attacks with weakness in half of the body and sometimes ataxia, coma, and cerebellar degeneration (Kors et al. 2001). However, more than one mutation has been successively identified in FHM1 patients exhibiting different phenotypes. As these mutations are absent in the vast majority of the exceedingly frequent migraine without aura (Nyholt et al. 2008), or have not been identified in genome-wide association studies (Gormley et al. 2016), doubts must be cast on the role of these mutations in migraine pain. It is possible that such mutations are linked to motor and other specific symptoms of FHM1.

2.2 LTCC and Pain

LTCC channels are widely diffused in peripheral tissues, with most localized in skeletal, smooth, and cardiac muscles. The two major classes of LTCC blockers, dihydropyridines (nifedipine, nimodipine, and many others) and phenylalkylamines (verapamil), are mainstays for the treatment of hypertension and arrythmia. However, the use of some of these drugs has also been considered for pain.

The observation that verapamil is effective in the treatment of cerebral vasospasm (Jun et al. 2010) may have some impact on the benefit that this drug offers in cluster headache patients (Petersen et al. 2019). From the first randomized clinical trial (Bussone et al. 1990), additional support for the current use of high doses of verapamil for the prophylaxis of cluster headache has derived from additional placebo controlled or open label trials (Blau and Engel 2004; Leone et al. 2000). Although LTCC are widely expressed in the central nervous system (CNS), it is possible that due to the key role of the hypothalamic clock in regulating circadian rhythms, and the strict chronobiology of the cluster headache attacks, the beneficial action of verapamil in this condition may be attributed to the LTCC localized at the level of this neural structure (McCarthy et al. 2016). Nimodipine and nifedipine have been proposed, although with conflicting results, for migraine prophylaxis, whereas flunarizine and cinnarizine are recommended in Europe for the treatment of this disease (Rossi et al. 2003; Stubberud et al. 2019). The latter drugs have multifaceted actions, being weak inhibitors of LTCCs and at the same time showing moderate antagonism for serotonin, histamine, and dopamine receptors. Inhibition of cortical

spreading depression, a wave of depolarization in the brain cortex considered a trigger of migraine attacks and other CNS actions of flunarizine, may explain the antimigraine action of the drug, along with some of the most relevant side effects (Parkinson-like symptoms, somnolence, and weight gain) (Reuter et al. 2002).

2.3 NTCC and Pain

Three calcium current components were described in dorsal root ganglion (DRG) neurons (Gross and Macdonald 1987; Nowycky et al. 1985) which included a dihydropyridine-sensitive (L-type) component, a low voltage activated (T-type) component, and a high voltage (N-type, comprising the $Ca_v2.2$ α_1 subunit) component. The latter component, which exerts a major contribution to neurotransmitter release (Hirning et al. 1988), was found to be selectively blocked by a toxin, ω-conotoxin GVIA, from the marine snail *Conus geographicus* (Boland et al. 1994; Plummer et al. 1989). Drug selectivity in blocking VGCCs was underlined by the failure of both dihydropyridines and ω-conotoxin GVIA to inhibit P-type channels (Llinas et al. 1989). Ziconotide was developed as a synthetic version of the $Ca_v2.2$ blocker, ω-conotoxin MVIIA, derived from a different marine snail, *Conus magus*, and is licensed for chronic pain. The pronounced analgesic effect of ziconotide for severe, intractable, and chronic cancer and non-malignant pain is, however, limited by the need to inject the drug via the intrathecal route of administration, and by its narrow therapeutic window (Sanford 2013). The inhibitory action on the release of proalgesic neurotransmitters, including calcitonin gene-related peptide (CGRP) (Santicioli et al. 1992), is one of the explanations for the robust analgesic activity of ziconotide. The absence of tolerance or addition by ziconotide is counterbalanced, however, by a series of serious adverse effects that confine its use to severe cases under strict medical control.

2.4 $\alpha_2\delta$-1, Inhibitors and Pain

Although initially developed as analogues of GABA, gabapentin and pregabalin were subsequently identified as ligands of the auxiliary VGCC $\alpha_2\delta$-1 subunit (Gee et al. 1996). The observation that $\alpha_2\delta$-1 null mice were resistant to the analgesic action of gabapentonoids robustly supported the proposal of the selective ability of these drugs to act on this calcium channel subunit (Field et al. 2006). Rather than a direct inhibitory action on neurotransmitter release, it is likely that gabapentonoids disrupt the axonal trafficking of $\alpha_2\delta$-1, which, being increased in injured nociceptors, is enhanced under circumstances underlying neuropathic pain. Interference with the $\alpha_2\delta$-1 subunit also affects calcium channel recycling between intracellular compartments and the synaptic membrane, thus inhibiting their function (Bauer et al. 2009). The anti-nociceptive activity of systemically administered gabapentin in behavioral assays displays a state-dependent effect (Field et al. 1997) that may explain its selectivity and tolerability. Nevertheless, ataxia, nausea, somnolence, and

dizziness may be experienced by patients on gabapentonoids. Thus, the search for more active and safer analgesics based on VGCC targeting focuses on domains different from the $\alpha_2\delta$-1 subunits, including $Ca_v2.2$. Indeed, elevated expression of $Ca_v2.2$ in presynaptic terminals of primary afferents and attenuated pain responses in $Ca_v2.2$ subunit deleted mice (Nieto-Rostro et al. 2018) still support future research for drugs that, by targeting this Ca channel subunit, may provide safer pain relief.

3 Transient Receptor Potential (TRP) Channels

3.1 Introduction to TRP Channels

The original observation that a Drosophila mutant, which was defective in sensing continuous light, showed only a transient receptor potential (TRP) instead of the normal sustained response (Cosens and Manning 1969) preceded the cloning of the gene responsible for this abnormal light response (Montell and Rubin 1989). After the initial discovery of some mammalian homologues (Wes et al. 1995; Zhu et al. 1995), the entire superfamily of mammalian TRPs consisting of 28 channels (27 in humans) has been identified to be composed by six subfamilies, distinguished on the basis of sequence homology, and not function (Nilius et al. 2012). These include the ankyrin (TRPA1), canonical (TRPC1–TRPC7), melastatin (TRPM1–TRPM8), mucolipin (TRPML1–TRPML3), polycystin (TRPP1–TRPP3), and vanilloid (TRPV1–TRPV6) subfamilies. Mutations of some TRP channels have been associated with hereditary diseases, such as mucolipodosis (TRPML) and polycystic kidney diseases (TRPP), from which their respective names are derived (Nilius et al. 2012).

The six families of TRP channels share a common sequence and structure, consisting of six transmembrane spanning regions (S1–S6), a pore-forming loop between (S5 and S6), and the possibility of functioning as homo- or hetero-tetramers (Nilius et al. 2012). TRP activation increases the influx of cations, which results in profound changes in intracellular calcium concentrations. TRP channels have been found in a large variety of cells in practically all tissues and organs, where they exert pleiotropic functions. This discussion will focus on one of the main roles of TRPs in primary sensory neurons, where they sense noxious stimuli and sustain pain signals.

3.2 TRPV1 and Pain

The cloning of the "capsaicin receptor," TRP vanilloid 1 (TRPV1) (Caterina et al. 1999), which is abundantly expressed in a subset of primary sensory neurons and causes burning pain, has promoted a new area of research for the discovery of more efficacious and safer analgesics. Additional TRPs expressed in primary afferents include TRPV3, TRPV4, TRPA1, TRPM2, TRPM3, and TRPM8 (Talavera et al. 2008; Voets et al. 2005). While TRPV1 and TRPA1 coexist in the same trigeminal (TG) and DRG neurons, which contain and release the proalgesic neuropeptides,

CGRP and substance P (SP), the menthol receptor, TRPM8, is found in a different subset of non-peptidergic sensory neurons (Bhattacharya et al. 2008). Due to the prominent role of CGRP in migraine and cluster headache, the ability of exogenous and endogenous agonists of TRPV1 and TRPA1 to cause, and of channel antagonists to prevent, headaches have been proposed (Marone et al. 2018; Nassini et al. 2012). Expression and pathophysiological functions of TRP channels are not limited to neurons, as TRPV4 has been identified in satellite glial cells that surround the neuronal cell body in TGs and DRGs (Rajasekhar et al. 2015), and TRPA1 in cells of the oligodendrocyte/Schwann cell lineage (De Logu et al. 2017, 2019; Hamilton et al. 2016). Initially, TRPs expressed by primary sensory neurons were found to encode specific temperature signals from noxious (TRPA1) to mild (TRPM8) cold and from mild (TRPV3) to noxious heat (TRPV1 and TRPV2) (Nilius et al. 2012). However, a more complex scenario emerged from the recent finding showing that perception of hot temperatures requires the simultaneous contribution of TRPV1, TRPA1, and TRPM3 (Vandewauw et al. 2018). If the understanding of the mechanisms responsible for acute pain is of theoretical importance, the identification of the pathways implicated in maintaining chronic pain is, however, crucial for the identification of targets for novel analgesics.

The implication of TRPV1 in pain transmission is linked to the ability of capsaicin (Szolcsanyi et al. 1975), and the ultrapotent agonist, resiniferatoxin (Szallasi and Blumberg 1989), to elicit acute burning pain followed by a prolonged thermal and mechanical hyperalgesia, thus indicating TRPV1 as a valuable target to attenuate chronic pain. A peculiar property of capsaicin-like agonists is that after the application of sufficiently high doses, these agents produce a time-dependent neuronal insensitivity to a large variety of painful stimuli. This inactivation, possibly due to a massive Ca^{2+}-influx and the ensuing breakdown of the nerve fiber cytoskeleton by Ca^{2+}-dependent proteases (Chard et al. 1995), transiently affects the function of the TRPV1-expressing nociceptors. Therefore, while TRPV1 deleted mice exhibit a phenotype with reduced responses solely to channel agonists (Caterina et al. 2000; Davis et al. 2000), desensitization to capsaicin and the ensuing complete nociceptor defunctionalization implies a broader antihyperalgesic action.

From this and additional evidence, a remarkable effort has been undertaken to develop desensitizing ointments with capsaicin or other agonists, and TRPV1 antagonists. Dermal patches containing 8% capsaicin (NGX-4010) (Noto et al. 2009) are currently approved in the European Union for various forms of peripheral neuropathic pain. While systemic resiniferatoxin is associated with severe side effects (hair loss and skin ulcers) in rats, its topical use seems to be safe and has been used via intravesical administration for catheter-related bladder discomfort (Zhang et al. 2012) and via the intrathecal route for intractable cancer pain (Heiss and Iadarola 2015). More than 10 years ago, the first TRPV1 antagonists entered clinical trials, showing two major side effects that were not predicted from animal studies: increase in body temperature and reduced sensitivity to noxious heat, an effect which may expose subjects to burn injuries. These side effects were more pronounced with certain antagonists, such as AMG517, which raised temperatures up to 40.2°C (Gavva 2008), than others, while two antagonists, PHE377 and

NEO6860, apparently did not elevate body temperature (Arsenault et al. 2018). However, the reason for the different ability to evoke temperature-related side effects by diverse antagonists is not completely understood.

3.3 TRPA1 and Pain

TRPA1, the only member of the ankyrin subfamily, is characterized by an abundant series of ankyrin repeats in the intracellular domain. As for other TRPs, including TRPV1 (Liao et al. 2013), TRPV2 (Huynh et al. 2016; Zubcevic et al. 2016), TRPV6 (Saotome et al. 2018), and TRPP2 (Shen et al. 2016), the structures of TRPA1 have been revealed by electron cryo-microscopy (Paulsen et al. 2015). These findings and mutagenesis studies have clarified that TRPA1 is exquisitely sensitive to the redox state of the milieu (Takahashi et al. 2011), and for this reason is the preferred target of a series of reactive oxygen, nitrogen, and carbonyl species (ROS, RNS, and RCS, respectively). Indeed, robust evidence has been accumulated showing that hydrogen peroxide, peroxynitrite, acrolein, and 4-hydroxynonenal (Andersson et al. 2008; Sawada et al. 2008; Taylor-Clark et al. 2009; Trevisani et al. 2007), and many other exogenous and endogenous reactive molecules via non-covalent or covalent binding to specific cysteine residues cause channel activation (Hinman et al. 2006; Macpherson et al. 2007).

ROS, RNS, and RCS are markedly increased at sites of inflammation or tissue injury (Kallenborn-Gerhardt et al. 2012; Nassini et al. 2014). Their ability to stimulate TRPA1-expressing cells and, in particular, peripheral terminals of nociceptors may explain that TRPA1 deleted mice or rodents treated with TRPA1 antagonists exhibit attenuated pain-like behaviors in a large variety of rodent models of inflammatory (Bonet et al. 2013; da Costa et al. 2010; Eid et al. 2008; McGaraughty et al. 2010; McNamara et al. 2007; Moilanen et al. 2012; Petrus et al. 2007), neuropathic (De Logu et al. 2017; Eid et al. 2008; Katsura et al. 2006; Trevisan et al. 2016), cancer (Antoniazzi et al. 2019), and migraine pain (Kunkler et al. 2018; Marone et al. 2018; Nassini et al. 2012). A gain of function mutation of TRPA1 has been associated with a familial episodic pain syndrome (Kremeyer et al. 2010), further supporting a role of TRPA1 in pain signaling.

3.4 TRPA1 and Chronic Neuropathic Pain

Neuropathic pain typically does not respond to cyclooxygenase inhibitors, whereas it is associated with excessive production of ROS that potentially may target TRPA1 in nociceptors. The Wallerian degeneration is characterized in injured nerve trunks by macrophage accumulation and increased oxidative burden, which markedly contribute to pain production (Gaudet et al. 2011; Ramer et al. 1997; von Hehn et al. 2012). However, the mechanism responsible for the macrophage-dependent and oxidative stress-mediated pain has been clarified only recently by investigating in a mouse model of trigeminal neuropathic pain the role of nociceptor TRPA1,

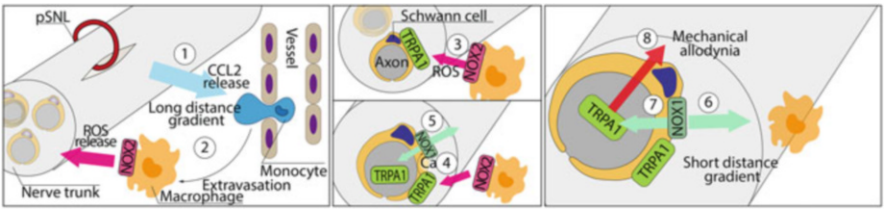


Fig. 1 Schematic representations of the cellular and molecular pathway that orchestrates chronic mechanical allodynia in a mouse model of neuropathic pain (partial sciatic nerve ligation, pSNL). Chemokine ligand 2 (CCL2) release from the injured nerve trunk (1) recruits macrophages to the site of nerve damage (2), which by a NADPH oxidase-2 (NOX2) mechanism generate an oxidative burst that targets TRPA1 in Schwann cells (3). The stimulated channel via a calcium-dependent mechanism (4) activates NOX1 to promote a bidirectional release of reactive oxygen species (ROS) (5): the outwardly directed release maintains the macrophage influx into the intra-neural space (6), while the inwardly directed release targets nociceptor TRPA1 (7) to signal pain (8) (modified from De Logu et al. 2017)

which senses oxidative stress generated by macrophages, thus signaling pain (Trevisan et al. 2016). However, a more recent study in a mouse model of neuropathic pain (partial sciatic nerve ligation) proposes a more complex pathway, pointing to the unpredicted TRPA1-dependent mechanism which causes chronic pain by sustaining neuroinflammation (De Logu et al. 2017). TRPA1 has been found to be expressed by Schwann cells surrounding the sciatic nerve fibers (De Logu et al. 2017), where it is targeted by the oxidative burst generated by macrophages infiltrating the injured nerve trunk. Stimulation of Schwann cell TRPA1 receptors results in a Ca^{2+}-dependent activation of NADPH oxidase-1 (NOX1) that produces a bidirectional ROS/RCS release. An outwardly directed release sustains macrophage influx inside the injured nerve trunk, while an inwardly directed release targets the neuronal TRPA1 to signal pain (De Logu et al. 2017) (Fig. 1). The presence of TRPA1 in human Schwann cells, where it generates a NOX1-dependent hydrogen peroxide release (De Logu et al. 2017), suggests that the pathway described in mice carries implications in human neuropathic pain.

However, it is possible to elicit neuropathic pain without a contribution of an inflammatory cell component, but with a critical role of Schwann cell TRPA1. Neuropathic pain is a frequent symptom associated with the peripheral neuropathy occurring in alcoholics (Chopra and Tiwari 2012). Although the short-lived ethanol metabolite acetaldehyde has been implicated, the mechanism responsible for the increased pain sensitivity in alcoholics is unknown. A recent study (De Logu et al. 2019) confirmed that acute pain caused by ethanol in mice is exclusively mediated by TRPV1 (Trevisani et al. 2002), but also revealed that the prolonged and robust mechanical allodynia that follows the acute pain sensation is entirely dependent on TRPA1. In this case, however, macrophages do not infiltrate peripheral nerve fibers, rather Schwann cells are found surrounding peripheral nerve endings and are thought to contribute to this situation. Acetaldehyde is produced by alcohol dehydrogenase in the liver and in Schwann cells themselves (De Logu et al. 2017, 2019)

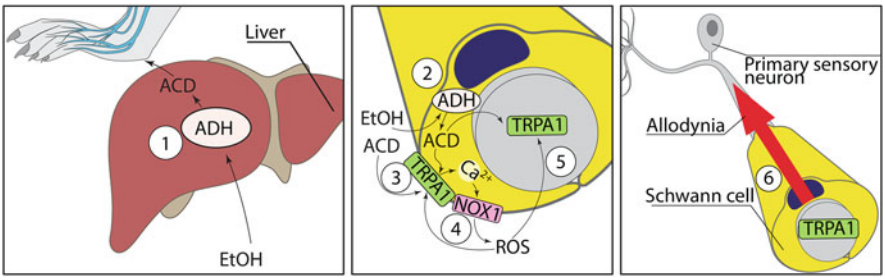


Fig. 2 Schematic representations of the cellular pathway that orchestrates prolonged ethanol-evoked neuropathic pain. Alcohol dehydrogenase (ADH) in the liver (1) and in Schwann cells (2) converts ethanol in the TRPA1 agonist, acetaldehyde (ACD). ACD targets TRPA1 in Schwann cells (3) to elicit, via a Ca^{2+}-dependent NADPH oxidase-1 (NOX1) pathway, a sustained release of reactive oxygen species (ROS) (4). ROS target nociceptor TRPA1 (5) to signal prolonged allodynia (6) (modified from De Logu et al. 2019)

and is a TRPA1 agonist (Bang et al. 2007). Thus, acetaldehyde, in part by an autocrine mechanism, targets Schwann cell TRPA1 receptors, which via NOX1-dependent ROS/RCS generation activates neuronal TRPA1 receptors to sustain prolonged mechanical allodynia (De Logu et al. 2019). Therefore, it would appear that different pathways driven by inflammatory cell accumulation, or by specific enzymatic pathways, share a common final mechanism mediated by Schwann cell-TRPA1 receptors to sustain chronic pain (Fig. 2). Further studies are required to confirm such pathways in additional animal models and in human chronic pain.

So far, only one TRPA1 antagonist, GRC 17536, has undergone investigation in a human phase-2 clinical trial in patients with painful diabetic neuropathy. No relevant side effects were reported, whereas reduced pain was observed (Khairatkar-Joshi et al. 2016). However, problems in the formulation hampered further development of the drug. There is, therefore, recent renewed interest in the pharmaceutical industry to develop improved TRPA1 antagonists for the treatment of pain.

3.5 Other TRP Channels and Pain

TRPM8, the menthol receptor (McKemy et al. 2002), when deleted in mice, failed to avoid cold temperatures (Knowlton et al. 2013) and to fully express cold allodynia (Bautista et al. 2007), suggested that this channel could contribute to cold hypersensitivity often observed in neuropathic pain patients. Mutations of TRPM8 have been associated with a migraine phenotype (Fu et al. 2019). The role of TRPM8 in cold sensation and allodynia may be supported by the use of menthol as topical analgesic in traditional medicine and by some findings in animal models (Proudfoot et al. 2006). However, these data should be considered with caution as the drug may be not selective for TRPM8. Only one TRPM8 antagonist, PF-05105679, has been tested in humans. The drug, which attenuated responses in the cold pressure test, did

not produce, as did other TRPM8 antagonists, small decreases in body temperature, but elicited an unpleasant hot sensation localized to the head and upper extremities (Andrews et al. 2015).

Expression of TRPV4 has been documented in primary afferents and TRPV4 antagonism or gene deletion attenuated inflammatory and neuropathic pain (Alessandri-Haber et al. 2008; Chen et al. 2007; Materazzi et al. 2012). In particular, this channel seems to be implicated in pancreatitis pain (Ceppa et al. 2010; Kanju et al. 2016) and cyclophosphamide-induced bladder cystitis (Everaerts et al. 2010). However, it should be noted that several mutations of TRPV4 have been associated with a variety of diseases, including skeletal dysplasia, arthropathies, congenital-distal spinal motor atrophy, and Charcot-Marie-Tooth disease type 2C (Nilius and Voets 2013), with little implication for proalgesic phenotypes. Thus, it is unclear whether TRV4 antagonists investigated in pulmonary edema may be tested in pain diseases.

Although its presence in nociceptors has been documented, much less has been reported regarding TRPV3 and pain. The selective antagonist GRC15300 failed to reduce neuropathic pain in a phase-2 clinical trial (Broad et al. 2016). It is unclear if this failure is due to the fact that it was not a full antagonist (Grubisha et al. 2014). The ability of TRPM2 to increase mechanical allodynia induced by nerve injury of joint or gut inflammation seems to be linked to increased tissue inflammation, rather than to a direct action on the nervous system (Haraguchi et al. 2012). TRPM3, selectively activated by pregnenolone sulfate, is expressed in a large subset of TRPV1-expressing sensory neurons (Vriens et al. 2011), and is required, along with TRPV1 and TRPA1, to elicit the acute heat-evoked pain response (Vandewauw et al. 2018). However, its contribution to chronic pain remains unknown.

4 Acid-Sensing Family of Ion Channels (ASICs)

4.1 Introduction to ASICs

The acid-sensing family of ion channels (ASICs) belongs to the superfamily of the voltage-insensitive, amiloride-sensitive epithelial sodium channel/degenerin (ENaC/DEG) cation channels (Kellenberger and Schild 2015; Krishtal 2003). To date, six ASICs have been identified that arise from four genes: ASIC1a and ASIC1b are splicing variants of the ASIC1 gene; ASIC2a and ASIC2b arise from the ASIC2 gene; ASIC3 and ASIC4. The ASIC4 protein does not appear to function either as a proton-gated or a modulatory channel (Akopian et al. 2000; Wemmie et al. 2006). ASIC2b is inactive when expressed alone, but seems to modify the properties of ASIC2a and ASIC3 when they are co-expressed (Lingueglia et al. 1997).

Although ASICs share only about 20–25% identity with the ENaC channels, they show all the structural features of the superfamily, which include two hydrophobic transmembrane domains, a large cysteine-rich extracellular loop, and short intracellular N- and C-termini (Holzer 2009; Wemmie et al. 2006). The recently described crystal structure of the chicken ASIC1 showed a channel consisting of three subunits which are required to form a functional channel (Baconguis et al. 2014; Baconguis

and Gouaux 2012; Dawson et al. 2012; Gonzales et al. 2009; Jasti et al. 2007). Each subunit consists of different extracellular domains, with the proton sensor distributed over multiple sites in the extracellular loop (Diochot et al. 2007; Holzer 2009). The amino acid sequences of ASIC subunits are well conserved between species: the mouse ASIC1A and the human ASIC1A share over 99% of their amino acid sequence identity (Wemmie et al. 2013).

ASICs are nonselective cation channels activated by a variety of exogenous chemicals or endogenous mediators, such as divalent and polyvalent cations, neuropeptides, arachidonic acid, protein kinases, and proteases. Following activation, they result preferentially permeable to Na^+, even if the homomeric ASIC1a channels also result to be permeable to Ca^{2+} (Waldmann et al. 1997; Wemmie et al. 2006). Functional ASICs are formed by homomultimers or heteromultimers (Benson et al. 2002; Waldmann and Lazdunski 1998) and, based on their structure, they possess distinct kinetics, pH sensitivity, ion selectivity, tissue distribution, and pharmacological properties (Hesselager et al. 2004; Waldmann et al. 1999). For instance, both ASIC1a and ASIC1b homomeric channels generate a rapidly activating and inactivating current. ASIC2a activates and inactivates more slowly, and ASIC3 generates a more rapidly activating and inactivating current (Holzer 2009). The recombinant and native channels are particularly sensitive to moderate extracellular low pH, with pH of half maximal activation (pH 0.5) ranging from 6.2 to 6.8 for ASIC1a, 5.1 to 6.2 for ASIC1b, 4.1 to 5 for ASIC2a, and 6.2 to 6.7 for ASIC3 (Benson et al. 2002; Chen et al. 1998; Price et al. 1996; Sutherland et al. 2001; Waldmann et al. 1997).

Although ASICs may be found in the intestine, liver, and other tissues, anatomical sites of major expression are the brain (ASIC1a, ASIC2a, and ASIC2b) (Wemmie et al. 2002) and the peripheral nervous system, where ASIC1b, ASIC2b, and ASIC3 are extensively expressed in small and medium nociceptive neurons (Benson et al. 2002; Chen et al. 1998), and ASIC2a and ASIC3 are mainly expressed in medium and large sensory neurons (Price et al. 2001). ASIC-like currents have also been measured from human DRG neurons (Baumann et al. 2004), and the presence of ASICs other than ASIC1a in the dorsal horn of spinal cord is less clear (Duan et al. 2007; Wu et al. 2004). Moreover, ASIC1a, ASIC2a, ASIC2b, and ASIC4 are widely expressed in the brain (Alvarez de la Rosa et al. 2003; Chen et al. 1998; Wemmie et al. 2002), and ASIC4, which is not activated by protons, has also been detected in the pituitary gland and retina (Brockway et al. 2002; Grunder et al. 2000).

4.2 Peripheral ASICs and Pain

Different physiological and pathological events, including muscle exercise, ischemia, and inflammation, may induce tissue acidification and changes in pH in the environment surrounding nociceptive nerves. The presence of ASIC channels both in the cell body and nerve terminals of pain-sensing neurons suggests their importance in acid-induced nociception (Deval and Lingueglia 2015; Sluka and Gregory

2015). The first observation of the putative role of ASIC channels in pain sensation emerged when amiloride, a nonselective ASIC blocker, and the NSAIDs, diclofenac and ibuprofen, selective inhibitors of ASIC1a and ASIC3, respectively (Voilley et al. 2001), attenuated pain evoked by intradermal acid infusion in humans (Jones et al. 2004; Ugawa et al. 2002).

Currently, different observations, obtained by using pharmacological inhibitors of the various ASIC channels and animals with genetic deletions of ASIC channels, support this hypothesis (Akopian et al. 2000; Holzer 2009). For instance, the pharmacological blockade and genetic knockdown of ASIC3 decreased primary and secondary hyperalgesia in a model of joint inflammation in rodents (Walder et al. 2010). In addition, in the inflammatory model induced by complete Freund's adjuvant (CFA), a 15-fold increase in ASIC3 mRNA in primary sensory neurons was observed, and a mixture of the proinflammatory mediators, including nerve growth factor, serotonin, interleukin-1, and bradykinin, increased ASIC3-like current density in isolated DRG neurons. Other ASICs showed controversial results in peripheral sensitization and pain, such as ASIC1a-knockout mice that showed changes in some pain behaviors but not in others (Ikeuchi et al. 2009; Radhakrishnan et al. 2003), and ASIC1a inhibition in the peripheral nervous system did not reduce thermal, mechanical, chemical/inflammatory, and muscle pain (Drew et al. 2004; Mogil et al. 2005). Strong evidence showing that peripheral ASIC3 activation is an essential sensor of cutaneous acidic pain in both normal and inflammatory conditions is provided by the use of small molecules, 2-guani-dine-4-methylquinazoline (GMQ) and the related endogenous polyamine, agmatine. The injection of GMQ into the mouse paw induced pain behaviors in wild-type but not ASIC3 knockout mice (Yu et al. 2010). It has also been shown that GMQ or agmatine binds ASIC3 at a site different from the proton acid-sensing domain (Yu et al. 2010).

These data indicate the existence of ASIC3 activators other than protons and suggest that endogenous molecules may activate ASICs to cause pain. The role of ASICs in inflammatory pain has been widely investigated, and the use of ASIC knockout mice seems to suggest that of the different ASICs, ASIC3s play a major role in primary inflammatory pain.

4.2.1 Central ASICs and Pain

ASIC channels are not solely present in primary sensory neurons, but have also been reported in the CNS where they contribute to central sensitization (Holzer 2009). Spinal dorsal horn neurons express a high density of homomeric ASIC1a channels, and the expression of these channels is upregulated by peripheral inflammation (Duan et al. 2007; Wu et al. 2004). Recent evidence reveals a role for ASICs in pain processing in the CNS. Intrathecal injection of the tarantula venom peptide, psalmotoxin 1 (PcTx1), selectively blocked homomultimeric ASIC1a and reduced thermal, mechanical, chemical, inflammatory, and neuropathic pain in rodents (Drew et al. 2004; Mogil et al. 2005).

In the spinal cord, ASIC1a and ASIC2a expression levels were increased by peripheral inflammation, suggesting a role for ASICs in central pain sensitization (Mogil et al. 2005; Yagi et al. 2006). Taken together, these studies suggest that ASICs

in the CNS contribute to pain processing. Although the mechanisms of ASIC action in central pain circuits are not yet clear, it is possible that ASICs alter neuron excitability or synaptic plasticity. One potential mechanism of how ASIC channels are activated suggest that the presence of the protons at synapses making them acidic (~pH 5.5) aids the release of neurotransmitter from vesicles (Voilley et al. 2001). For example, there is evidence that cholinergic, glutamatergic, and GABAergic synaptic vesicles have pH values of pH ~5.5 (Dietrich and Morad 2010; Michaelson and Angel 1980; Monshausen et al. 2016). However, other sources of protons, such as those generated by energy metabolism, might contribute to ASIC activation in the brain.

Altogether, these studies highlight the hypothesis that PNS and CNS use different combinations of ASIC subunits to mediate pain. Optimum signaling through ASICs at different anatomical sites may require different channel properties, and channel activity might be optimized through different combinations of ASIC subunits and ASIC modulators. Importantly, these studies have identified ASICs as potential targets for new pain medications. The ASIC inhibitor, amiloride, has been approved for use in humans, and a few small translational experiments have demonstrated its potential for reducing cutaneous pain and migraine (Hattori et al. 2009; Immke and McCleskey 2001).

5 PIEZO Channels

5.1 Introduction to PIEZO Channels

Mechanotransduction is an important physiologic process associated with a variety of biological functions, such as sensing of shear stress, pain and hearing, regulation of vasculature tone, urine flow, and bladder distention. Recently, a family of cation selective channels, called PIEZO, has been identified as essential proteins in mediating mechanosensory transduction in mammalian cells (Coste et al. 2010). Within this family, two members, PIEZO1 and PIEZO2, encoded by the *PIEZO1/FAM38A* and *PIEZO2/FAM38B* genes, respectively, have been identified in the murine neuroblastoma cell line N2A (Coste et al. 2010, 2012). PIEZO proteins are conserved among different species and possess unique features in their primary sequences, without an apparent sequence homology with any other known ion channels (Coste et al. 2010). They are large proteins composed of about 2,500 amino acids (2,521 and 2,752 amino acids for human PIEZO1 and human PIEZO2, respectively) with many (>30) transmembrane segments for each subunit (Coste et al. 2012). A recently resolved cryo-electron microscopy structure of the PIEZO1 channel showed a trimeric structure of the channel (Ge et al. 2015). PIEZO proteins were first designed only as components of mechanically activated ion channels and were shown to possess a pore region which allows the flow of ions, including calcium, inside the cell, following mechanical stimuli.

The PIEZO family of proteins has been shown to be directly gated by lipid layer tension (Cox et al. 2016; Syeda et al. 2016), displaying a rapid voltage-dependent desensitization and inactivation during the static phase of the stimulus in patches and

whole-cell mechanical assays (Coste et al. 2010). A functional difference has been reported between the two members, showing a more rapid inactivation for PIEZO2 compared to PIEZO1 (Coste et al. 2010). PIEZO proteins become non-inactivating with excessive mechanical stimulation, mainly caused by disruption of cytoskeletal support and/or membrane domain structure (Suchyna et al. 2004).

PIEZO1 is broadly expressed in non-sensory tissues exposed to fluid pressure, such as the skin, bladder, kidney, lung, endothelial cells, erythrocytes, and periodontal ligament cells (Ranade et al. 2014a, b; Jin et al. 2015), whereas PIEZO2 is predominantly found in sensory tissue, such as TG and DRG and Merkel cells (Coste et al. 2010; Huynh et al. 2016; Maksimovic et al. 2014). PIEZO1 also senses the local cellular environment (e.g., stochastic nanoroughness, confinement, or substrate stiffness) in neurons and other cells, thereby promoting downstream changes in specific cell–cell interactions and motility. The different distribution in non-sensory and sensory tissues of both PIEZO proteins seems to be conserved among the various species. Some cells, such as chondrocytes in cartilage, express both channels forming heteromeric channels and conferring a high-strain mechanosensitivity to articular cartilage (Lee et al. 2014). The presence of the PIEZO channels appears to be necessary for vertebrate survival, as a PIEZO1 knockout mouse does not survive after mid-gestation, mainly due to interrupted development of the vasculature system (Ranade et al. 2014a). Like PIEZO1, total knockout of PIEZO2 in mouse induces a lethal phenotype, with pups dying at birth (Ranade et al. 2014b), and different tissue-specific conditional knockout lines have shown that PIEZO2 mediates many responses to light, for example, skin-specific knockout of PIEZO2 leads to reduced light touch responses (Maksimovic et al. 2014; Woo et al. 2014).

5.2 PIEZO and Pain Sensation

One of the most sophisticated functions of the somatosensory system is the mechanosensation of touch and mechanical pain. About 70–80% of the neurons cultured from DRGs are characterized by having numerous and distinct cationic currents which probably mediate different sensations of touch proprioception and mechanical pain. PIEZO2 is expressed in more than 95% of low-threshold mechanoreceptors and mediates rapidly adapting currents in DRG neurons (Coste et al. 2010). Investigation and characterization of sensory neuron-specific PIEZO2-conditional deleted mice have revealed the critical role of PIEZO2 in sensing gentle touch and proprioception (Ranade et al. 2014b; Woo et al. 2015). Recent studies have demonstrated that PIEZO2 is also essential for mediating tactile allodynia in mice (Murthy et al. 2018; Szczot et al. 2018), and it has been verified that PIEZO2 is required for sensing gentle touch, proprioception, and tactile allodynia in humans as well (Szczot et al. 2018). Whether PIEZO2, whose cryo-electron microscopy structure has been recently revealed (Wang et al. 2019), functions as the mechanotransduction channel for sensing mechanical pain is ambiguous. It has been reported that tamoxifen-inducible *Advilin-CreERT2* mice with a deletion of PIEZO2 in ~82% of PIEZO2 positive DRG neurons had a striking behavioral defect

in gentle-touch sensation, but no changes in mechanical pain responses (Ranade et al. 2014b). In contrast, a more robust deletion of PIEZO2 in sensory neurons using the *HoxB8-Cre* mice resulted not only in defective touch and proprioception but also a partial impairment of the mechanical pain response (Murthy et al. 2018). However, ex vivo skin-nerve recordings found that only the firing frequency, but not the number of the nociceptive Aδ and C-fibers was reduced in the *HoxB8-Cre*-dependent PIEZO2 knockout mice (Murthy et al. 2018).

Another study with the use of in vivo calcium imaging of primary sensory neurons (Szczot et al. 2018) found that PIEZO2 deleted neurons were completely insensitive to gentle touch, but retained normal response to noxious mechanical stimuli. Importantly, human patients with loss-of-function mutations in PIEZO2 responded normally to noxious mechanical stimuli, but failed to develop tactile allodynia after skin inflammation (Szczot et al. 2018). This study suggests that PIEZO2, although dispensable for mechanical nociception, is critical for touch and mechanical allodynia (Szczot et al. 2018). In another recent study (Zhang et al. 2019), it was observed that the deletion of PIEZO2 in a portion of the low-threshold mechanoreceptors and a majority of mechano-nociceptors positive to isolectin B4 impairs touch, but unexpectedly sensitizes acute mechanical pain. Moreover, the ectopic expression of PIEZO1 sensitizes touch in normal mice and also saves the defective touch and proprioception of PIEZO2 knockout mice. Acute mechanical pain is suppressed rather than evoked, even when PIEZO1 is ectopically expressed in all DRG neurons (Zhang et al. 2019). Taken together, these data suggest that PIEZO channels play an important role in mediating touch and indirectly suppress acute pain.

6 Purinergic $P2X_3$ Ion Channels

6.1 Introduction to $P2X_3$ Ion Channels

In addition to a variety of important homeostatic functions, adenosine 5′-triphosphate (ATP) acts as a signaling molecule via ionotropic (ligand-gated ion channels, P2X) and metabotropic (G-protein coupled receptors, P2Y) receptors (Abbracchio and Burnstock 1994; Fredholm et al. 1994). P2X receptors allow the intracellular influx of cations, thus changing membrane potential and initiating subsequent cellular events (Abbracchio and Burnstock 1994; Fredholm et al. 1994). Among a variety of responses, ionic changes driven by activation of presynaptic P2X receptors modulate neurotransmitter release (Khakh et al. 2003; Nakatsuka and Gu 2001), whereas activation of postsynaptic P2X receptors results in fast excitatory signaling (Khakh 2001; Nörenberg 2000). The P2X family of receptors, widely expressed in all mammalian tissues, consists of the homomeric $P2X_1$, $P2X_2$, $P2X_3$, $P2X_4$, $P2X_5$, and $P2X_7$ channels and the heteromeric $P2X_{2/3}$ and $P2X_{1/5}$ receptors (Kawate et al. 2009). Notably, both homomeric $P2X_3$ and heteromeric P2X2/3 receptors have been found in small to medium diameter C-

and Aδ-fiber primary afferent neurons, suggesting their implication in the pain-sensing system (Bradbury et al. 1998; Dunn et al. 2001).

6.2 $P2X_3$ Ion Channels in Pain and Cough

A series of concurrent findings have underlined the role of the $P2X_3$ receptor in pain transmission. The injection into rodent skin of ATP or αβ-methylene ATP, a selective agonist for the $P2X_3$ receptor, produced nociceptive behaviors (Chen and Gu 2005; Cockayne et al. 2000; Kennedy 2005; Tsuda et al. 2000). By using selective antisense or short interfering RNA (siRNA) of the $P2X_3$ receptor, several studies have revealed its implication in models of neuropathic or inflammatory pain (Barclay et al. 2002; Dorn et al. 2004; Honore et al. 2002a). The role of the $P2X_3$ receptor in pain models was confirmed by using $P2X_3$ receptor antagonists, such as TNP-ATP (2′,3′-O-(2,4,6-trinitrophenyl)adenosine-5′-triphosphate) and pyridoxal phosphate-6-azo(benzene-2,4-disulfonic acid) (PPADS) (Honore et al. 2002b). However, it should be underlined that $P2X_3$ null mice showed normal transduction of sensory inputs, a part for the painful response to formalin, indicating a prominent role in inflammatory pain (Cockayne et al. 2000; Souslova et al. 2000).

$P2X_3$ receptors are associated with several biomarkers, which characterize three distinct neuronal subtypes. These are the TRPA1/Mas-related G-protein coupled receptor member D (MRGPRD) receptors, the nerve growth factor (NGF) receptor, TrkA/calcitonin gene-related peptide (CGRP), or the TRPA1/TRPV1 channels (Usoskin et al. 2015). Colocalizations may affect $P2X_3$ functions in different subpopulations of sensory neurons with diverse responses. For example, negative cooperativity between P2X3 receptor and ASIC channels has been reported. In fact, if $P2X_3$ is co-expressed with either ASIC1a, ASIC2a, or ASIC3, both proton-evoked ASIC currents and $P2X_3$-mediated responses are similarly attenuated as compared to conditions in which these channels are expressed alone (Stephan et al. 2018).

It should be underscored that $P2X_3$-mediated responses are not limited to pain but encompass the activation of additional protective reflex responses, including cough. $P2X_3$ receptors, which are activated by ATP released within airway tissue, are expressed in guinea pig vagal C-fibers (Kwong et al. 2008; Weigand et al. 2012). Furthermore, P2X receptor-dependent mechanisms exaggerate cough responses to tussive stimuli when guinea pigs are exposed to ATP and histamine aerosols (Kamei and Takahashi 2006; Kamei et al. 2005). These observations led to the hypothesis that a broad range of stimuli may cause responses, including cough, which are enhanced by $P2X_3$-receptor activation in terminals of primary afferents in airways or at the level of their central synapses (Khakh and North 2006; Prado et al. 2013; Vulchanova et al. 1997). This proposal has been studied in patients with chronic intractable cough who showed attenuated coughing after treatment with AF-219, a selective $P2X_3$ receptor antagonist (Abdulqawi et al. 2015). Further clinical phase III studies will show whether this interesting hypothesis can provide benefit to patients affected by this debilitating and undertreated condition.

References

Abbracchio MP, Burnstock G (1994) Purinoceptors: are there families of P2X and P2Y purinoceptors? Pharmacol Ther 64(3):445–475

Abdulqawi R, Dockry R, Holt K, Layton G, Mccarthy BG, Ford AP, Smith JA (2015) P2X3 receptor antagonist (AF-219) in refractory chronic cough: a randomised, double-blind, placebo-controlled phase 2 study. Lancet 385(9974):1198–1205

Akopian AN, Chen CC, Ding Y, Cesare P, Wood JN (2000) A new member of the acid-sensing ion channel family. Neuroreport 11(10):2217–2222

Alessandri-Haber N, Dina OA, Joseph EK, Reichling DB, Levine JD (2008) Interaction of transient receptor potential vanilloid 4, integrin, and SRC tyrosine kinase in mechanical hyperalgesia. J Neurosci 28(5):1046–1057

Alvarez de la Rosa D, Krueger SR, Kolar A, Shao D, Fitzsimonds RM, Canessa CM (2003) Distribution, subcellular localization and ontogeny of ASIC1 in the mammalian central nervous system. J Physiol 546(Pt 1):77–87

Andersson DA, Gentry C, Moss S, Bevan S (2008) Transient receptor potential A1 is a sensory receptor for multiple products of oxidative stress. J Neurosci 28(10):2485–2494

Andrews MD, Af Forselles K, Beaumont K, Galan SR, Glossop PA, Grenie M et al (2015) Discovery of a selective TRPM8 antagonist with clinical efficacy in cold-related pain. ACS Med Chem Lett 6(4):419–424

Antoniazzi CTD, Nassini R, Rigo FK, Milioli AM, Bellinaso F, Camponogara C et al (2019) Transient receptor potential ankyrin 1 (TRPA1) plays a critical role in a mouse model of cancer pain. Int J Cancer 144(2):355–365

Arsenault P, Chiche D, Brown W, Miller J, Treister R, Leff R et al (2018) NEO6860, modality-selective TRPV1 antagonist: a randomized, controlled, proof-of-concept trial in patients with osteoarthritis knee pain. Pain Rep 3(6):e696

Baconguis I, Gouaux E (2012) Structural plasticity and dynamic selectivity of acid-sensing ion channelspider toxin complexes. Nature 489(7416):400–405

Baconguis I, Bohlen CJ, Goehring A, Julius D, Gouaux E (2014) X-ray structure of acid-sensing ion channel 1-snake toxin complex reveals open state of a Na(+)-selective channel. Cell 156 (4):717–729

Bang S, Kim KY, Yoo S, Kim YG, Hwang SW (2007) Transient receptor potential A1 mediates acetaldehyde-evoked pain sensation. Eur J Neurosci 26(9):2516–2523

Bannister RA, Beam KG (2013) Ca(V)1.1: the atypical prototypical voltage-gated Ca(2)(+) channel. Biochim Biophys Acta 1828(7):1587–1597

Barclay J, Patel S, Dorn G, Wotherspoon G, Moffatt S, Eunson L et al (2002) Functional downregulation of P2X3 receptor subunit in rat sensory neurons reveals a significant role in chronic neuropathic and inflammatory pain. J Neurosci 22(18):8139–8147

Bauer CS, Nieto-Rostro M, Rahman W, Tran-Van-Minh A, Ferron L, Douglas L, Dolphin AC (2009) The increased trafficking of the calcium channel subunit alpha2delta-1 to presynaptic terminals in neuropathic pain is inhibited by the alpha2delta ligand pregabalin. J Neurosci 29 (13):4076–4088

Baumann TK, Chaudhary P, Martenson ME (2004) Background potassium channel block and TRPV1 activation contribute to proton depolarization of sensory neurons from humans with neuropathic pain. Eur J Neurosci 19:1343–1351

Bautista DM, Siemens J, Glazer JM, Tsuruda PR, Basbaum AI, Stucky CL et al (2007) The menthol receptor TRPM8 is the principal detector of environmental cold. Nature 448(7150):204–208

Benson CJ, Xie J, Wemmie JA, Price MP, Henss JM, Welsh MJ, Snyder PM (2002) Heteromultimers of DEG/ENaC subunits form H+-gated channels in mouse sensory neurons. Proc Natl Acad Sci U S A 99(4):2338–2343

Bhattacharya MR, Bautista DM, Wu K, Haeberle H, Lumpkin EA, Julius D (2008) Radial stretch reveals distinct populations of mechanosensitive mammalian somatosensory neurons. Proc Natl Acad Sci U S A 105(50):20015–20020

Blau JN, Engel HO (2004) Individualizing treatment with verapamil for cluster headache patients. Headache 44(10):1013–1018

Boland LM, Morrill JA, Bean BP (1994) Omega-conotoxin block of N-type calcium channels in frog and rat sympathetic neurons. J Neurosci 14(8):5011–5027

Bonet IJM, Fischer L, Parada CA, Tambeli CH (2013) The role of transient receptor potential A 1 (TRPA1) in the development and maintenance of carrageenan-induced hyperalgesia. Neuropharmacology 65:206–212

Bradbury EJ, Burnstock G, McMahon SB (1998) The expression of P2X3 purinoreceptors in sensory neurons: effects of axotomy and glial-derived neurotrophic factor. Mol Cell Neurosci 12(4–5):256–268

Broad LM, Mogg AJ, Eberle E, Tolley M, Li DL, Knopp KL (2016) TRPV3 in drug development. Pharmaceuticals (Basel) 9(3):55

Brockway LM, Zhou Z-H, Bubien JK, Jovov B, Benos DJ, Keyser KT (2002) Rabbit retinal neurons and glia express a variety of ENaC/DEG subunits. Am J Physiol Cell Physiol 283(1): C126–C134

Bussone G, Leone M, Peccarisi C, Micieli G, Granella F, Magri M et al (1990) Double blind comparison of lithium and verapamil in cluster headache prophylaxis. Headache 30(7):411–417

Caterina MJ, Rosen TA, Tominaga M, Brake AJ, Julius D (1999) A capsaicin-receptor homologue with a high threshold for noxious heat. Nature 398(6726):436–441

Caterina MJ, Leffler A, Malmberg AB, Martin WJ, Trafton J, Petersen-Zeitz KR et al (2000) Impaired nociception and pain sensation in mice lacking the capsaicin receptor. Science 288 (5464):306–313

Catterall WA, Leal K, Nanou E (2013) Calcium channels and short-term synaptic plasticity. J Biol Chem 288(15):10742–10749

Ceppa E, Cattaruzza F, Lyo V, Amadesi S, Pelayo JC, Poole DP et al (2010) Transient receptor potential ion channels V4 and A1 contribute to pancreatitis pain in mice. Am J Physiol Gastrointest Liver Physiol 299(3):G556–G571

Chard PS, Bleakman D, Savidge JR, Miller RJ (1995) Capsaicin-induced neurotoxicity in cultured dorsal root ganglion neurons: involvement of calcium-activated proteases. Neuroscience 65 (4):1099–1108

Chen M, Gu JG (2005) A P2X receptor-mediated nociceptive afferent pathway to lamina I of the spinal cord. Mol Pain 1:4

Chen CC, England S, Akopian AN, Wood JN (1998) A sensory neuron-specific, proton-gated ion channel. Proc Natl Acad Sci U S A 95(17):10240–10245

Chen X, Alessandri-Haber N, Levine JD (2007) Marked attenuation of inflammatory mediator-induced C-fiber sensitization for mechanical and hypotonic stimuli in TRPV4−/− mice. Mol Pain 3(31):31

Chopra K, Tiwari V (2012) Alcoholic neuropathy: possible mechanisms and future treatment possibilities. Br J Clin Pharmacol 73(3):348–362

Cockayne DA, Hamilton SG, Zhu QM, Dunn PM, Zhong Y, Novakovic S et al (2000) Urinary bladder hyporeflexia and reduced pain-related behaviour in P2X3-deficient mice. Nature 407 (6807):1011–1015

Cosens DJ, Manning A (1969) Abnormal electroretinogram from a Drosophila mutant. Nature 224 (5216):285–287

Coste B, Mathur J, Schmidt M, Earley TJ, Ranade S, Petrus MJ et al (2010) Piezo1 and Piezo2 are essential components of distinct mechanically activated cation channels. Science (New York, NY) 330(6000):55–60

Coste B, Xiao B, Santos JS, Syeda R, Grandl J, Spencer KS et al (2012) Piezo proteins are pore-forming subunits of mechanically activated channels. Nature 483(7388):176–181

Cox CD, Bae C, Ziegler L, Hartley S, Nikolova-Krstevski V, Rohde PR et al (2016) Removal of the mechanoprotective influence of the cytoskeleton reveals PIEZO1 is gated by bilayer tension. Nat Commun 7:10366

da Costa DSM, Meotti FC, Andrade EL, Leal PC, Motta EM, Calixto JB (2010) The involvement of the transient receptor potential A1 (TRPA1) in the maintenance of mechanical and cold hyperalgesia in persistent inflammation. Pain 148(3):431–437

Davis JB, Gray J, Gunthorpe MJ, Hatcher JP, Davey PT, Overend P et al (2000) Vanilloid receptor-1 is essential for inflammatory thermal hyperalgesia. Nature 405(6783):183–187

Dawson RJP, Benz J, Stohler P, Tetaz T, Joseph C, Huber S et al (2012) Structure of the acid-sensing ion channel 1 in complex with the gating modifier Psalmotoxin 1. Nat Commun 3:936

De Logu F, Nassini R, Materazzi S, Carvalho Gonçalves M, Nosi D, Rossi Degl'Innocenti D et al (2017) Schwann cell TRPA1 mediates neuroinflammation that sustains macrophage-dependent neuropathic pain in mice. Nat Commun 8(1):1887

De Logu F, Li Puma S, Landini L, Portelli F, Innocenti A, de Araujo DSM et al (2019) Schwann cells expressing nociceptive channel TRPA1 orchestrate ethanol-evoked neuropathic pain in mice. J Clin Invest 5:128022

Deval E, Lingueglia E (2015) Acid-sensing ion channels and nociception in the peripheral and central nervous systems. Neuropharmacology 94:49–57

Dietrich CJ, Morad M (2010) Synaptic acidification enhances GABAA signaling. J Neurosci 30 (47):16044–16052

Diochot S, Salinas M, Baron A, Escoubas P, Lazdunski M (2007) Peptides inhibitors of acid-sensing ion channels. Toxicon 49(2):271–284

Dorn G, Patel S, Wotherspoon G, Hemmings-Mieszczak M, Barclay J, Natt FJC et al (2004) siRNA relieves chronic neuropathic pain. Nucleic Acids Res 32(5):e49

Drew LJ, Rohrer DK, Price MP, Blaver KE, Cockayne DA, Cesare P, Wood JN (2004) Acid-sensing ion channels ASIC2 and ASIC3 do not contribute to mechanically activated currents in mammalian sensory neurones. J Physiol 556(3):691–710

Duan B, Wu L-J, Yu Y-Q, Ding Y, Jing L, Xu L et al (2007) Upregulation of acid-sensing ion channel ASIC1a in spinal dorsal horn neurons contributes to inflammatory pain hypersensitivity. J Neurosci 27(41):11139–11148

Dunn PM, Zhong Y, Burnstock G (2001) P2X receptors in peripheral neurons. Prog Neurobiol 65 (2):107–134

Eid SR, Crown ED, Moore EL, Liang HA, Choong KC, Dima S et al (2008) HC-030031, a TRPA1 selective antagonist, attenuates inflammatory- and neuropathy-induced mechanical hypersensitivity. Mol Pain 4:48

Everaerts W, Zhen X, Ghosh D, Vriens J, Gevaert T, Gilbert JP et al (2010) Inhibition of the cation channel TRPV4 improves bladder function in mice and rats with cyclophosphamide-induced cystitis. Proc Natl Acad Sci U S A 107(44):19084–19089

Field MJ, Oles RJ, Lewis AS, McCleary S, Hughes J, Singh L (1997) Gabapentin (neurontin) and S-(+)-3-isobutylgaba represent a novel class of selective antihyperalgesic agents. Br J Pharmacol 121(8):1513–1522

Field MJ, Cox PJ, Stott E, Melrose H, Offord J, Su TZ et al (2006) Identification of the alpha2-delta-1 subunit of voltage-dependent calcium channels as a molecular target for pain mediating the analgesic actions of pregabalin. Proc Natl Acad Sci U S A 103(46):17537–17542

Fredholm BB, Abbracchio MP, Burnstock G, Daly JW, Harden TK, Jacobson KA et al (1994) Actions of caffeine in the brain with special reference to factors that contribute to its widespread use. Pharmacol Rev 46(2):143–156

Fu X, Yang J, Wu X, Lin Q, Zeng Y, Xia Q et al (2019) Association between PRDM16, MEF2D, TRPM8, LRP1 gene polymorphisms and migraine susceptibility in the she ethnic population in China. Clin Invest Med 42(1):E21–E30

Gaudet AD, Popovich PG, Ramer MS (2011) Wallerian degeneration: gaining perspective on inflammatory events after peripheral nerve injury. J Neuroinflammation 8(110):110

Gavva NR (2008) Body-temperature maintenance as the predominant function of the vanilloid receptor TRPV1. Trends Pharmacol Sci 29(11):550–557

Ge J, Li W, Zhao Q, Li N, Chen M, Zhi P et al (2015) Architecture of the mammalian mechanosensitive Piezo1 channel. Nature 527(7576):64–69

Gee NS, Brown JP, Dissanayake VU, Offord J, Thurlow R, Woodruff GN (1996) The novel anticonvulsant drug, gabapentin (Neurontin), binds to the alpha2delta subunit of a calcium channel. J Biol Chem 271(10):5768–5776

Gonzales EB, Kawate T, Gouaux E (2009) Pore architecture and ion sites in acid-sensing ion channels and P2X receptors. Nature 460(7255):599–604

Gormley P, Anttila V, Winsvold BS, Palta P, Esko T, Pers TH et al (2016) Meta-analysis of 375,000 individuals identifies 38 susceptibility loci for migraine. Nat Genet 48(8):856–866

Gross RA, Macdonald RL (1987) Dynorphin A selectively reduces a large transient (N-type) calcium current of mouse dorsal root ganglion neurons in cell culture. Proc Natl Acad Sci U S A 84(15):5469–5473

Grubisha O, Mogg AJ, Sorge JL, Ball LJ, Sanger H, Ruble CL et al (2014) Pharmacological profiling of the TRPV3 channel in recombinant and native assays. Br J Pharmacol 171 (10):2631–2644

Grunder S, Geissler HS, Bassler EL, Ruppersberg JP (2000) A new member of acid-sensing ion channels from pituitary gland. Neuroreport 11(8):1607–1611

Hamilton NB, Kolodziejczyk K, Kougioumtzidou E, Attwell D (2016) Proton-gated Ca(2+)-permeable TRP channels damage myelin in conditions mimicking ischaemia. Nature 529 (7587):523–527

Haraguchi K, Kawamoto A, Isami K, Maeda S, Kusano A, Asakura K et al (2012) TRPM2 contributes to inflammatory and neuropathic pain through the aggravation of pronociceptive inflammatory responses in mice. J Neurosci 32(11):3931–3941

Hattori T, Chen J, Harding AMS, Price MP, Lu Y, Abboud FM, Benson CJ (2009) ASIC2a and ASIC3 heteromultimerize to form pH-sensitive channels in mouse cardiac dorsal root ganglia neurons. Circ Res 105(3):279–286

Heiss JD, Iadarola M (2015) Intrathecal resiniferatoxin for intractable cancer pain. J Neurosurg 123: A515-A

Hesselager M, Timmermann DB, Ahring PK (2004) pH dependency and desensitization kinetics of heterologously expressed combinations of acid-sensing ion channel subunits. J Biol Chem 279 (12):11006–11015

Hinman A, Chuang HH, Bautista DM, Julius D (2006) TRP channel activation by reversible covalent modification. Proc Natl Acad Sci U S A 103(51):19564–19568

Hirning LD, Fox AP, McCleskey EW, Olivera BM, Thayer SA, Miller RJ, Tsien RW (1988) Dominant role of N-type Ca2+ channels in evoked release of norepinephrine from sympathetic neurons. Science 239(4835):57–61

Holzer P (2009) Acid-sensitive ion channels and receptors. Handb Exp Pharmacol 194:283–332

Honore P, Kage K, Mikusa J, Watt AT, Johnston JF, Wyatt JR et al (2002a) Analgesic profile of intrathecal P2X(3) antisense oligonucleotide treatment in chronic inflammatory and neuropathic pain states in rats. Pain 99(1–2):11–19

Honore P, Mikusa J, Bianchi B, McDonald H, Cartmell J, Faltynek C, Jarvis MF (2002b) TNP-ATP, a potent P2X3 receptor antagonist, blocks acetic acid-induced abdominal constriction in mice: comparison with reference analgesics. Pain 96(1–2):99–105

Huynh KW, Cohen MR, Jiang J, Samanta A, Lodowski DT, Zhou ZH, Moiseenkova-Bell VY (2016) Structure of the full-length TRPV2 channel by cryo-EM. Nat Commun 7:11130

Ikeuchi M, Kolker SJ, Sluka KA (2009) Acid-sensing ion channel 3 expression in mouse knee joint afferents and effects of carrageenan-induced arthritis. J Pain 10(3):336–342

Immke DC, McCleskey EW (2001) Lactate enhances the acid-sensing NA+ channel on ischemia-sensing neurons. Nat Neurosci 4(9):869–870

Jasti J, Furukawa H, Gonzales EB, Gouaux E (2007) Structure of acid-sensing ion channel 1 at 1.9 Å resolution and low pH. Nature 449(7160):316–323

Jin Y, Li J, Wang Y, Ye R, Feng X, Jing Z, Zhao Z (2015) Functional role of mechanosensitive ion channel Piezo1 in human periodontal ligament cells. Angle Orthod 85(1):87–94

Jones NG, Slater R, Cadiou H, McNaughton P, McMahon SB (2004) Acid-induced pain and its modulation in humans. J Neurosci 24(48):10974–10979

Jun P, Ko NU, English JD, Dowd CF, Halbach VV, Higashida RT et al (2010) Endovascular treatment of medically refractory cerebral vasospasm following aneurysmal subarachnoid hemorrhage. AJNR Am J Neuroradiol 31(10):1911–1916

Kallenborn-Gerhardt W, Schroder K, Del Turco D, Lu R, Kynast K, Kosowski J, Schmidtko A (2012) NADPH oxidase-4 maintains neuropathic pain after peripheral nerve injury. J Neurosci 32(30):10136–10145

Kamei J, Takahashi Y (2006) Involvement of ionotropic purinergic receptors in the histamine-induced enhancement of the cough reflex sensitivity in Guinea pigs. Eur J Pharmacol 547 (1–3):160–164

Kamei J, Takahashi Y, Yoshikawa Y, Saitoh A (2005) Involvement of P2X receptor subtypes in ATP-induced enhancement of the cough reflex sensitivity. Eur J Pharmacol 528(1–3):158–161

Kanju P, Chen Y, Lee W, Yeo M, Lee SH, Romac J et al (2016) Small molecule dual-inhibitors of TRPV4 and TRPA1 for attenuation of inflammation and pain. Sci Rep 6:26894

Katsura H, Obata K, Mizushima T, Yamanaka H, Kobayashi K, Dai Y et al (2006) Antisense knock down of TRPA1, but not TRPM8, alleviates cold hyperalgesia after spinal nerve ligation in rats. Exp Neurol 200(1):112–123

Kawate T, Michel JC, Birdsong WT, Gouaux E (2009) Crystal structure of the ATP-gated P2X (4) ion channel in the closed state. Nature 460(7255):592–598

Kellenberger S, Schild L (2015) Structure, function, and pharmacology of acid-sensing ion channels and the epithelial Na+ channel. Pharmacol Rev 67(1):1–35

Kennedy C (2005) P2X receptors: targets for novel analgesics? Neuroscientist 11(4):345–356

Khairatkar-Joshi N, Tandon M, Gudi G, Keohane P, Godsafe Z, Inventors G. P. S. A. (2016) TRPA1 antagonist for the treatment of pain associated to diabetic neuropathic pain

Khakh BS (2001) Molecular physiology of P2X receptors and ATP signalling at synapses. Nat Rev Neurosci 2(3):165–174

Khakh BS, North RA (2006) P2X receptors as cell-surface ATP sensors in health and disease. Nature 442(7102):527–532

Khakh BS, Gittermann D, Cockayne DA, Jones A (2003) ATP modulation of excitatory synapses onto interneurons. J Neurosci 23(19):7426–7437

Knowlton WM, Palkar R, Lippoldt EK, McCoy DD, Baluch F, Chen J, McKemy DD (2013) A sensory-labeled line for cold: TRPM8-expressing sensory neurons define the cellular basis for cold, cold pain, and cooling-mediated analgesia. J Neurosci 33(7):2837–2848

Kors EE, Terwindt GM, Vermeulen FL, Fitzsimons RB, Jardine PE, Heywood P et al (2001) Delayed cerebral edema and fatal coma after minor head trauma: role of the CACNA1A calcium channel subunit gene and relationship with familial hemiplegic migraine. Ann Neurol 49 (6):753–760

Kremeyer B, Lopera F, Cox JJ, Momin A, Rugiero F, Marsh S et al (2010) A gain-of-function mutation in TRPA1 causes familial episodic pain syndrome. Neuron 66(5):671–680

Krishtal O (2003) The ASICs: signaling molecules? Modulators? Trends Neurosci 26(9):477–483

Kunkler PE, Zhang L, Johnson PL, Oxford GS, Hurley JH (2018) Induction of chronic migraine phenotypes in a rat model after environmental irritant exposure. Pain 159(3):540–549

Kwong K, Kollarik M, Nassenstein C, Ru F, Undem BJ (2008) P2X2 receptors differentiate placodal vs. neural crest C-fiber phenotypes innervating Guinea pig lungs and esophagus. Am J Physiol Lung Cell Mol Physiol 295(5):L858–L865

Lee W, Leddy HA, Chen Y, Lee SH, Zelenski NA, McNulty AL et al (2014) Synergy between Piezo1 and Piezo2 channels confers high-strain mechanosensitivity to articular cartilage. Proc Natl Acad Sci U S A 111(47):E5114–E5122

Leone M, D'Amico D, Frediani F, Moschiano F, Grazzi L, Attanasio A, Bussone G (2000) Verapamil in the prophylaxis of episodic cluster headache: a double-blind study versus placebo. Neurology 54(6):1382–1385

Liao M, Cao E, Julius D, Cheng Y (2013) Structure of the TRPV1 ion channel determined by electron cryo-microscopy. Nature 504(7478):107–112

Lingueglia E, De Weille JR, Bassilana F, Heurteaux C, Sakaif H, Waldmann R, Lazdunskig M (1997) A modulatory subunit of acid sensing ion channels in brain and dorsal root ganglion cells. J Biol Chem 272(47):29778–29783

Llinas RR, Sugimori M, Cherksey B (1989) Voltage-dependent calcium conductances in mammalian neurons. The P channel. Ann N Y Acad Sci 560:103–111

Macpherson LJ, Xiao B, Kwan KY, Petrus MJ, Dubin AE, Hwang S, Patapoutian A (2007) An ion channel essential for sensing chemical damage. J Neurosci 27(42):11412–11415

Maksimovic S, Nakatani M, Baba Y, Nelson AM, Marshall KL, Wellnitz SA, Lumpkin EA (2014) Epidermal Merkel cells are mechanosensory cells that tune mammalian touch receptors. Nature 509(7502):617–621

Marone IM, De Logu F, Nassini R, De Carvalho Goncalves M, Benemei S, Ferreira J et al (2018) TRPA1/NOX in the soma of trigeminal ganglion neurons mediates migraine-related pain of glyceryl trinitrate in mice. Brain 141(8):2312–2328

Materazzi S, Fusi C, Benemei S, Pedretti P, Patacchini R, Nilius B et al (2012) TRPA1 and TRPV4 mediate paclitaxel-induced peripheral neuropathy in mice via a glutathione-sensitive mechanism. Pflugers Arch 463(4):561–569

McCarthy MJ, Le Roux MJ, Wei H, Beesley S, Kelsoe JR, Welsh DK (2016) Calcium channel genes associated with bipolar disorder modulate lithium's amplification of circadian rhythms. Neuropharmacology 101:439–448

McGaraughty S, Chu KL, Perner RJ, Didomenico S, Kort ME, Kym PR (2010) TRPA1 modulation of spontaneous and mechanically evoked firing of spinal neurons in uninjured, osteoarthritic, and inflamed rats. Mol Pain 6(14):14

McKemy DD, Neuhausser WM, Julius D (2002) Identification of a cold receptor reveals a general role for TRP channels in thermosensation. Nature 416(6876):52–58

McNamara CR, Mandel-Brehm J, Bautista DM, Siemens J, Deranian KL, Zhao M et al (2007) TRPA1 mediates formalin-induced pain. Proc Natl Acad Sci U S A 104(33):13525–13530

Michaelson DM, Angel I (1980) Determination of delta pH in cholinergic synaptic vesicles: its effect on storage and release of acetylcholine. Life Sci 27(1):39–44

Mogil JS, Breese NM, Witty MF, Ritchie J, Rainville ML, Ase A et al (2005) Transgenic expression of a dominant-negative ASIC3 subunit leads to increased sensitivity to mechanical and inflammatory stimuli. J Neurosci 25(43):9893–9901

Moilanen LJ, Laavola M, Kukkonen M, Korhonen R, Leppanen T, Hogestatt ED et al (2012) TRPA1 contributes to the acute inflammatory response and mediates carrageenan-induced paw edema in the mouse. Sci Rep 2(380):380

Monshausen GB, Miller ND, Murphy AS, Gilroy S, Miesenböck G, de Angelis DA et al (2016) Thèse Identification et caractérisation d ' un canal chlorure, AtCLCg, impliqué dans la réponse au stress salin chez Arabidopsis thaliana. Biophys J 4(9):192–195

Montell C, Rubin GM (1989) Molecular characterization of the Drosophila trp locus: a putative integral membrane protein required for phototransduction. Neuron 2(4):1313–1323

Murthy S, Loud M et al (2018) The mechanosensitive ion channel Piezo2 mediates sensitivity to mechanical pain in mice. Sci Transl Med 10:eaat9897

Nakatsuka T, Gu JG (2001) ATP P2X receptor-mediated enhancement of glutamate release and evoked EPSCs in dorsal horn neurons of the rat spinal cord. J Neurosci 21(17):6522–6531

Nassini R, Materazzi S, Vriens J, Prenen J, Benemei S, De Siena G et al (2012) The "headache tree" via umbellulone and TRPA1 activates the trigeminovascular system. Brain 135(Pt 2):376–390

Nassini R, Materazzi S, Benemei S, Geppetti P (2014) The TRPA1 channel in inflammatory and neuropathic pain and migraine. Rev Physiol Biochem Pharmacol 167:1–43

Nieto-Rostro M, Ramgoolam K, Pratt WS, Kulik A, Dolphin AC (2018) Ablation of alpha2delta-1 inhibits cell-surface trafficking of endogenous N-type calcium channels in the pain pathway in vivo. Proc Natl Acad Sci U S A 115(51):E12043–E12052

Nilius B, Voets T (2013) The puzzle of TRPV4 channelopathies. EMBO Rep 14(2):152–163

Nilius B, Appendino G, Owsianik G (2012) The transient receptor potential channel TRPA1: from gene to pathophysiology. Pflugers Arch 464(5):425–458

Nörenberg W, Illes P (2000) Neuronal P2X receptors: localisation and functional properties. Naunyn Schmiedeberg's Arch Pharmacol 362(4):324–339
Noto C, Pappagallo M, Szallasi A (2009) NGX-4010, a high-concentration capsaicin dermal patch for lasting relief of peripheral neuropathic pain. Curr Opin Investig Drugs 10(7):702–710
Nowycky MC, Fox AP, Tsien RW (1985) Three types of neuronal calcium channel with different calcium agonist sensitivity. Nature 316(6027):440–443
Nyholt DR, LaForge KS, Kallela M, Alakurtti K, Anttila V, Farkkila M et al (2008) A high-density association screen of 155 ion transport genes for involvement with common migraine. Hum Mol Genet 17(21):3318–3331
Paulsen CE, Armache JP, Gao Y, Cheng Y, Julius D (2015) Structure of the TRPA1 ion channel suggests regulatory mechanisms. Nature 520(7548):511–517
Petersen AS, Barloese MCJ, Snoer A, Soerensen AMS, Jensen RH (2019) Verapamil and cluster headache: still a mystery. A narrative review of efficacy, mechanisms and perspectives. Headache 24(10):13603
Petrus M, Peier AM, Bandell M, Hwang SW, Huynh T, Olney N et al (2007) A role of TRPA1 in mechanical hyperalgesia is revealed by pharmacological inhibition. Mol Pain 3:40
Plummer MR, Logothetis DE, Hess P (1989) Elementary properties and pharmacological sensitivities of calcium channels in mammalian peripheral neurons. Neuron 2(5):1453–1463
Prado FC, Araldi D, Vieira AS, Oliveira-Fusaro MCG, Tambeli CH, Parada CA (2013) Neuronal P2X3 receptor activation is essential to the hyperalgesia induced by prostaglandins and sympathomimetic amines released during inflammation. Neuropharmacology 67:252–258
Price MP, Snydert PM, Welsh MJ (1996) Cloning and expression of a novel human brain Na+ channel. J Biol Chem 271(14):7879–7882
Price MP, McIlwrath SL, Xie J, Cheng C, Qiao J, Tarr DE et al (2001) The DRASIC cation channel contributes to the detection of cutaneous touch and acid stimuli in mice. Neuron 32 (6):1071–1083
Proudfoot CJ, Garry EM, Cottrell DF, Rosie R, Anderson H, Robertson DC et al (2006) Analgesia mediated by the TRPM8 cold receptor in chronic neuropathic pain. Curr Biol 16(16):1591–1605
Radhakrishnan R, Moore SA, Sluka KA (2003) Unilateral carrageenan injection into muscle or joint induces chronic bilateral hyperalgesia in rats. Pain 104(3):567–577
Rajasekhar P, Poole DP, Liedtke W, Bunnett NW, Veldhuis NA (2015) P2Y1 receptor activation of the TRPV4 ion channel enhances purinergic signaling in satellite glial cells. J Biol Chem 290 (48):29051–29062
Ramer MS, French GD, Bisby MA (1997) Wallerian degeneration is required for both neuropathic pain and sympathetic sprouting into the DRG. Pain 72(1–2):71–78
Ranade SS, Qiu Z, Woo S-H, Hur SS, Murthy SE, Cahalan SM et al (2014a) Piezo1, a mechanically activated ion channel, is required for vascular development in mice. Proc Natl Acad Sci U S A 111(28):10347–10352
Ranade SS, Woo S-H, Dubin AE, Moshourab RA, Wetzel C, Petrus M et al (2014b) Piezo2 is the major transducer of mechanical forces for touch sensation in mice. Nature 516(7529):121–125
Reuter U, Chiarugi A, Bolay H, Moskowitz MA (2002) Nuclear factor-kappaB as a molecular target for migraine therapy. Ann Neurol 51(4):507–516
Rossi P, Fiermonte G, Pierelli F (2003) Cinnarizine in migraine prophylaxis: efficacy, tolerability and predictive factors for therapeutic responsiveness. An open-label pilot trial. Funct Neurol 18 (3):155–159
Sanford M (2013) Intrathecal ziconotide: a review of its use in patients with chronic pain refractory to other systemic or intrathecal analgesics. CNS Drugs 27(11):989–1002
Santicioli P, Del Bianco E, Tramontana M, Geppetti P, Maggi CA (1992) Release of calcitonin gene-related peptide like-immunoreactivity induced by electrical field stimulation from rat spinal afferents is mediated by conotoxin-sensitive calcium channels. Neurosci Lett 136 (2):161–164
Saotome K, Singh AK, Sobolevsky AI (2018) Determining the crystal structure of TRPV6. CRC Press/Taylor & Francis, Boca Raton, pp 275–292

Sawada Y, Hosokawa H, Matsumura K, Kobayashi S (2008) Activation of transient receptor potential ankyrin 1 by hydrogen peroxide. Eur J Neurosci 27(5):1131–1142
Shen PS, Yang X, DeCaen PG, Liu X, Bulkley D, Clapham DE, Cao E (2016) The structure of the polycystic kidney disease channel PKD2 in lipid nanodiscs. Cell 167(3):763–773.e11
Sluka KA, Gregory NS (2015) The dichotomized role for acid sensing ion channels in musculoskeletal pain and inflammation. Neuropharmacology 94:58–63
Souslova V, Cesare P, Ding Y, Akopian AN, Stanfa L, Suzuki R et al (2000) Warm-coding deficits and aberrant inflammatory pain in mice lacking P2X3 receptors. Nature 407(6807):1015–1017
Stephan G, Huang L, Tang Y, Vilotti S, Fabbretti E, Yu Y et al (2018) The ASIC3/P2X3 cognate receptor is a pain-relevant and ligand-gated cationic channel. Nat Commun 9(1):1354
Stubberud A, Flaaen NM, McCrory DC, Pedersen SA, Linde M (2019) Flunarizine as prophylaxis for episodic migraine: a systematic review with meta-analysis. Pain 160(4):762–772
Suchyna TM, Tape SE, Koeppe RE, Andersen OS, Sachs F, Gottlieb PA (2004) Bilayer-dependent inhibition of mechanosensitive channels by neuroactive peptide enantiomers. Nature 430 (6996):235–240
Sutherland SP, Benson CJ, Adelman JP, McCleskey EW (2001) Acid-sensing ion channel 3 matches the acid-gated current in cardiac ischemia-sensing neurons. Proc Natl Acad Sci U S A 98(2):711–716
Syeda R, Florendo MN, Cox CD, Kefauver JM, Santos JS, Martinac B, Patapoutian A (2016) Piezo1 channels are inherently mechanosensitive. Cell Rep 17(7):1739–1746
Szallasi A, Blumberg PM (1989) Resiniferatoxin, a phorbol-related diterpene, acts as an ultrapotent analog of capsaicin, the irritant constituent in red pepper. Neuroscience 30(2):515–520
Szczot M, Liljencrantz J, Ghitani N, Barik A, Lam R, Thompson JH et al (2018) PIEZO2 mediates injury-induced tactile pain in mice and humans. Sci Transl Med 10:eaat9892
Szolcsanyi J, Jancso-Gabor A, Joo F (1975) Functional and fine structural characteristics of the sensory neuron blocking effect of capsaicin. Naunyn Schmiedeberg's Arch Pharmacol 287 (2):157–169
Takahashi N, Kuwaki T, Kiyonaka S, Numata T, Kozai D, Mizuno Y et al (2011) TRPA1 underlies a sensing mechanism for O2. Nat Chem Biol 7(10):701–711
Talavera K, Nilius B, Voets T (2008) Neuronal TRP channels: thermometers, pathfinders and lifesavers. Trends Neurosci 31(6):287–295
Taylor-Clark TE, Ghatta S, Bettner W, Undem BJ (2009) Nitrooleic acid, an endogenous product of nitrative stress, activates nociceptive sensory nerves via the direct activation of TRPA1. Mol Pharmacol 75(4):820–829
Trevisan G, Benemei S, Materazzi S, De Logu F, De Siena G, Fusi C, Nassini R (2016) TRPA1 mediates trigeminal neuropathic pain in mice downstream of monocytes/macrophages and oxidative stress. Brain 139:1361–1377
Trevisani M, Smart D, Gunthorpe MJ, Tognetto M, Barbieri M, Campi B et al (2002) Ethanol elicits and potentiates nociceptor responses via the vanilloid receptor-1. Nat Neurosci 5(6):546–551
Trevisani M, Siemens J, Materazzi S, Bautista DM, Nassini R, Campi B (2007) 4-Hydroxynonenal, an endogenous aldehyde, causes pain and neurogenic inflammation through activation of the irritant receptor TRPA1. Proc Natl Acad Sci U S A 104:13519–13524
Tsuda M, Koizumi S, Kita A, Shigemoto Y, Ueno S, Inoue K (2000) Mechanical allodynia caused by intraplantar injection of P2X receptor agonist in rats: involvement of heteromeric P2X2/3 receptor signaling in capsaicin-insensitive primary afferent neurons. J Neurosci 20(15):RC90
Ugawa S, Ueda T, Ishida Y, Nishigaki M, Shibata Y, Shimada S (2002) Amiloride-blockable acid-sensing ion channels are leading acid sensors expressed in human nociceptors. J Clin Investig 110(8):1185–1190
Usoskin D, Furlan A, Islam S, Abdo H, Lönnerberg P, Lou D, et al (2015) Unbiased classification of sensory neuron types by large-scale single-cell RNA sequencing. Nat Neurosci. 18:145–153
Vandewauw I, De Clercq K, Mulier M, Held K, Pinto S, Van Ranst N et al (2018) A TRP channel trio mediates acute noxious heat sensing. Nature 555(7698):662–666

Voets T, Talavera K, Owsianik G, Nilius B (2005) Sensing with TRP channels. Nat Chem Biol 1 (2):85–92
Voilley N, de Weille J, Mamet J, Lazdunski M (2001) Nonsteroid anti-inflammatory drugs inhibit both the activity and the inflammation-induced expression of acid-sensing ion channels in nociceptors. J Neurosci 21(20):8026–8033
von Hehn CA, Baron R, Woolf CJ (2012) Deconstructing the neuropathic pain phenotype to reveal neural mechanisms. Neuron 73(4):638–652
Vriens J, Owsianik G, Hofmann T, Philipp SE, Stab J, Chen X et al (2011) TRPM3 is a nociceptor channel involved in the detection of noxious heat. Neuron 70(3):482–494
Vulchanova L, Riedl MS, Shuster SJ, Buell G, Surprenant A, North RA, Elde R (1997) Immunohistochemical study of the P2X2 and P2X3 receptor subunits in rat and monkey sensory neurons and their central terminals. Neuropharmacology 36(9):1229–1242
Walder RY, Rasmussen LA, Rainier JD, Light AR, Wemmie JA, Sluka KA (2010) ASIC1 and ASIC3 play different roles in the development of hyperalgesia after inflammatory muscle injury. J Pain 11(3):210–218
Waldmann R, Lazdunski M (1998) H(+)-gated cation channels: neuronal acid sensors in the NaC/DEG family of ion channels. Curr Opin Neurobiol 8(3):418–424
Waldmann R, Champigny G, Bassilana F, Heurteaux C, Lazdunski M (1997) A proton-gated cation channel involved in acid-sensing. Nature 386(6621):173–177
Waldmann R, Champigny G, Lingueglia E, De Weille JR, Heurteaux C, Lazdunski M (1999) H(+)-gated cation channels. Ann N Y Acad Sci 868:67–76
Wang L, Zhou H, Zhang M, Liu W, Deng T, Zhao Q et al (2019) Structure and mechanogating of the mammalian tactile channel PIEZO2. Nature 573(7773):225–229
Weigand LA, Ford AP, Undem BJ (2012) A role for ATP in bronchoconstriction-induced activation of Guinea pig vagal intrapulmonary C-fibres. J Physiol 590(16):4109–4120
Wemmie JA, Chen J, Askwith CC, Hruska-Hageman AM, Price MP, Nolan BC et al (2002) The acid-activated ion channel ASIC contributes to synaptic plasticity, learning, and memory. Neuron 34(3):463–477
Wemmie JA, Price MP, Welsh MJ (2006) Acid-sensing ion channels: advances, questions and therapeutic opportunities. Trends Neurosci 29(10):578–586
Wemmie JA, Taugher RJ, Kreple CJ (2013) Acid-sensing ion channels in pain and disease. Nat Rev Neurosci 14(7):461–471
Wes PD, Chevesich J, Jeromin A, Rosenberg C, Stetten G, Montell C (1995) TRPC1, a human homolog of a Drosophila store-operated channel. Proc Natl Acad Sci U S A 92(21):9652–9656
Wheeler DG, Groth RD, Ma H, Barrett CF, Owen SF, Safa P, Tsien RW (2012) Ca(V)1 and Ca(V)2 channels engage distinct modes of Ca(2+) signaling to control CREB-dependent gene expression. Cell 149(5):1112–1124
Woo S-H, Ranade S, Weyer AD, Dubin AE, Baba Y, Qiu Z et al (2014) Piezo2 is required for Merkel-cell mechanotransduction. Nature 509(7502):622–626
Woo S-H, Lukacs V, de Nooij JC, Zaytseva D, Criddle CR, Francisco A et al (2015) Piezo2 is the principal mechanotransduction channel for proprioception. Nat Neurosci 18(12):1756–1762
Wu L-J, Duan B, Mei Y-D, Gao J, Chen J-G, Zhuo M et al (2004) Characterization of acid-sensing ion channels in dorsal horn neurons of rat spinal cord. J Biol Chem 279(42):43716–43724
Yagi J, Wenk HN, Naves LA, McCleskey EW (2006) Sustained currents through ASIC3 ion channels at the modest pH changes that occur during myocardial ischemia. Circ Res 99 (5):501–509
Yu Y, Chen Z, Li W-G, Cao H, Feng E-G, Yu F et al (2010) A nonproton ligand sensor in the acid-sensing ion channel. Neuron 68(1):61–72
Zamponi GW, Striessnig J, Koschak A, Dolphin AC (2015) The physiology, pathology, and pharmacology of voltage-gated calcium channels and their future therapeutic potential. Pharmacol Rev 67(4):821–870

Zhang N, Zhang P, Zhang X, Yang Y (2012) The efficacy of resiniferatoxin in prevention of catheter related bladder discomfort in patients after TURP - a pilot, randomized, open study. Transl Androl Urol 1(1):14–18

Zhang M, Wang Y, Geng J, Zhou S, Xiao B (2019) Mechanically activated piezo channels mediate touch and suppress acute mechanical pain response in mice. Cell Rep 26(6):1419–1431.e4

Zhu X, Chu PB, Peyton M, Birnbaumer L (1995) Molecular cloning of a widely expressed human homologue for the Drosophila trp gene. FEBS Lett 373(3):193–198

Zubcevic L, Herzik MA Jr, Chung BC, Liu Z, Lander GC, Lee SY (2016) Cryo-electron microscopy structure of the TRPV2 ion channel. Nat Struct Mol Biol 23(2):180–186

Exploiting the Diversity of Ion Channels: Modulation of Ion Channels for Therapeutic Indications

Yani Liu and KeWei Wang

Contents

1 Introduction 188
2 K^+ Channels 188
2.1 Introduction of K^+ Channels 188
2.2 Pharmacological Modulation and Implications of K^+ Channels 189
3 Sodium Channels 194
3.1 Introduction of Voltage-Gated Sodium (Nav) Channels 194
3.2 Pharmacological Modulations and Implications of Nav Channels 195
4 Chloride Channels 196
4.1 Introduction of Cl^- Channels 196
4.2 Target Validation and Drugs Targeting Cl^- for Medical Uses 197
References 202

Abstract

Ion channels are macromolecular proteins that form water-filled pores in cell membranes and they are critical for a variety of physiological and pharmacological functions. Dysfunctional ion channels can cause diseases known as channelopathies. Ion channels are encoded by approximately 400 genes, representing the second largest class of proven drug targets for therapeutic areas including neuropsychiatric disorders, cardiovascular and metabolic diseases, immunological diseases, nephrological diseases, gastrointestinal diseases, pulmonary/respiratory diseases, and many cancers. With more ion channel structures are being solved and functional robust assays are being developed, there are tremendous opportunities for identifying specific modulators targeting ion channels for new therapy.

Y. Liu · K. Wang (✉)
Department of Pharmacology, Qingdao University School of Pharmacy, Qingdao, China
e-mail: wangkw@qdu.edu.cn

J. E. Barrett et al. (eds.), *Concepts and Principles of Pharmacology*,
Handbook of Experimental Pharmacology 260, https://doi.org/10.1007/164_2019_333

Keywords
ANO1 · Arrhythmia · BK · Cancer · CFTR · Channelopathy · Drug target · Epilepsy · GABA · hERG · KCNQ · Kir · Kv7 · LRRC8A · Nav · Pain · TMEM16A

1 Introduction

Ion channels are macromolecular proteins that form water-filled pores in the plasma membranes of cells. In response to stimuli such as voltage change, mechanical stress, and neurotransmitters, ion channels open the pore through which ions diffuse in both a highly efficient and selective manner with $>10^7$ ions per second down the electrochemical gradient (Hille 2001). Ion channels can be classified based on their ion selectivity, gating (opening and closing) mechanism, and sequence homology. For the classification based on the gating mechanism, there are three main groups of ion channels: voltage-gated channels, ligand-gated channels, and mechanosensitive channels.

The human genome encodes approximately 400 ion channel genes (1.3%), representing the second largest class of membrane proteins among proven effect-mediating drug targets after G protein-coupled receptors (GPCRs). Ion channels are recognized as important and challenging drug targets for therapeutic areas including neuropsychiatric disorders, cardiovascular and metabolic diseases, immunological diseases, nephrology diseases, irritable bowel syndrome, pulmonary/respiratory diseases, and many cancers. Dysfunctional ion channels can cause diseases known as channelopathies.

2 K^+ Channels

2.1 Introduction of K^+ Channels

Potassium channels set the resting membrane potential, keep action potentials short, and shape the electrical activity of cells (Hille 2001; Miller 1986). Because of the gradient between intracellular K^+ ions (150 mM) and extracellular K^+ concentration (5 mM), opening K^+ channels favors an outpouring of positively charged ions and shifts the cell membrane voltage toward an equilibrium reversal potential (E_k) for hyperpolarization. Under this condition, the tendency for potassium ions to move down their concentration gradient is balanced by their tendency to move against their electrical gradient. Therefore, pharmacological activation of K^+ channels in excitable cells reduces excitability, whereas the channel inhibition has the opposite effect (Wulff et al. 2009).

The human genome encodes a superfamily of 80 K^+ channel genes that can be structurally classified as four groups, voltage-gated six transmembranes (6TM) and one pore (6TM/1P), calcium-activated seven transmembranes and one pore (7TM/1P), inwardly rectifying two transmembranes and one pore (2TM/1P), and

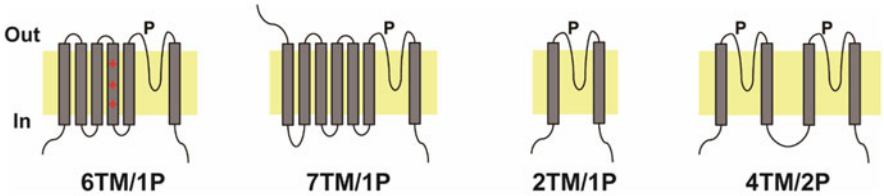


Fig. 1 Schematic membrane topology of potassium channels. All K^+ channels have a reentrant pore (P)-forming loop containing a unique three-amino acid GYG signature sequence that functions as selectivity filter. Each of Kv channels has a pore (P), six transmembrane α-helices (6TM) with positive charges in α helix 4 as voltage sensor and intracellular amino- and carboxy-termini. Calcium-activated K^+ channel subunit consisted of seven transmembranes (7TM) and one pore. An inward rectifier K^+ channel subunit has two transmembrane α-helices (2TM), a pore and intracellular amino- and carboxy-termini. Two-pore K^+ channel has two subunits with each subunit featuring four transmembrane α-helices (4TM) and a tandem-pore (2P) and intracellular amino- and carboxy-termini

tandem-pore domain four transmembranes and two pores (4TM/2P) channels, based on their membrane topology (Fig. 1). All potassium channels bear an identical and conserved three amino acid motif Gly/Tyr/Gly (GYG), known as signature sequence, in the pore.

The voltage-gated K^+ (Kv) channel α subunits encoded by 40 genes belong to the 6TM/1P and can be further grouped into 12 subfamilies, Kv1–12. K^+ channels are implicated in a range of diseases such as neuropsychiatric disorders, metabolic and cardiovascular conditions, autoimmune diseases, and cancer. Therefore, targeting K^+ channels with specific small molecules or biologics presents therapeutic strategies for a variety of indications. Here we highlight the therapeutic potential of some selected potassium channels.

2.2 Pharmacological Modulation and Implications of K^+ Channels

2.2.1 Kv7/KCNQ Channels

The voltage-gated Kv7 (or KCNQ) potassium channel subfamily comprises five members: Kv7.1–Kv7.5 encoded by *KCNQ1–KCNQ5* genes. Kv7.1 (KvLQT1) is expressed in the cardiac tissue and inner ear neurons. In the cardiac cells, Kv7.1 mediates I_{Ks} current responsible for repolarization of cardiac action potential. Kv7.2–Kv7.5 expressed in the nervous system are low-threshold, slowly activating K^+ currents originally termed the "M-current" (Brown and Adams 1980). In most neurons native M-channels are composed of heteromeric Kv7.2 and Kv7.3 subunits or homomeric Kv7.2 subunits (Wang et al. 1998). Kv7.4 subunits are predominantly expressed in the auditory and vestibular systems and also contribute to M-current in central dopaminergic neurons. Kv7.5 is primarily expressed in the brain and skeletal muscles.

Mutations of *KCNQ1–4* channel genes have been shown to cause genetic diseases. *KCNQ1* gene mutation is responsible for the autosomal dominant long QT syndrome of cardiac arrhythmia and the recessive Jervell and Lange-Nielsen cardio-auditory syndrome (Neyroud et al. 1997). Mutations in either the neuronal *KCNQ2* or *KCNQ3* gene cause benign familial neonatal convulsions (BFNC), a form of neonatal epilepsy (Biervert et al. 1998; Charlier et al. 1998; Singh et al. 1998). Kv7.4/KCNQ4 channels have been identified and linked to a form of inherited deafness (Kubisch et al. 1999).

The heteromeric expression of both Kv7.2 and Kv7.3 channels underlies the molecular identity of native neuronal M-current ($I_{k(M)}$) that was first described in peripheral sympathetic neurons and was inhibited by the muscarine through activating the M1, M3, and/or M5 acetylcholine receptor (mAChR) (Brown and Adams 1980). Kv7.2/Kv7.3 channels are dose-dependently blocked by two specific inhibitors, XE991 that was developed in an effort to ameliorate Alzheimer dementia and linopirdine as a prototypical compound that exhibited a cognitive-enhancing effect but did not pass phase 3 clinical trials. XE991 blocks Kv7.2/Kv7.3 with an IC_{50} around 0.6–0.98 μM or Kv7.1 homomeric channels with IC_{50} at 0.75 μM, but it is less potent against KvLQT1/minK channels with IC_{50} at 11.1 μM.

Because of their selective expressions in tissues, Kv7 channels are emerging targets that exhibit a unique pharmacology for therapeutic interventions. Retigabine or ezogabine is an anticonvulsant drug that was approved by the FDA in 2011 and is used as a treatment for partial-onset seizures. Retigabine primarily enhances Kv7 currents through a hyperpolarization shift of the channel voltage-dependent activation. Upregulating Kv7/KCNQ activity by retigabine repolarizes the membrane potential and inhibits repetitive firing that underlies epileptic activity. Retigabine activates four subtypes of Kv7/KCNQ channels, including Kv7.2, Kv7.3, Kv7.2/Kv7.3, Kv7.4, and Kv7.5 channels, with an effective concentration for half maximum response (EC_{50}) of 1.9 μM at −30 mV for Kv7.2. However, retigabine was withdrawn from the market in 2017 for side effects such as a blue-colored appearance of the skin or eyes. Flupirtine, a structural derivative of retigabine, functions as a centrally acting non-opioid analgesic that was also withdrawn in 2018 for liver toxicity. Recently, a 100-fold more potent and selective neuronal Kv7 opener, SCR2682, with an EC_{50} of 9.8 nM was identified and reported for antiepileptic activities in rodent models (Zhang et al. 2019), which may warrant further evaluation for clinical development of antiepileptic therapy.

2.2.2 hERG/Kv11.1 Channel

hERG channel or Kv11.1 is expressed in the heart and nervous tissues. In the heart, the hERG channel conducts the rapid component of the delayed rectifier current (I_{Kr}) (Sanguinetti et al. 1995; Trudeau et al. 1995). I_{Kr} is critical in determining the timing of the electrical repolarization of cardiac action potential (AP) in ventricular myocytes (Keating and Sanguinetti 2001). Genetic mutations in hERG channel gene can result in long QT syndrome (LQTS), a disorder that predisposes individuals to life-threatening risk of sudden death due to an arrhythmia known as *Torsades de pointes* (TdP). TdP is a polymorphic ventricular tachycardia that exhibits distinct

characteristics on the electrocardiogram (ECG). Arrhythmia can also be induced by a surprisingly diverse group of compounds or drugs that block hERG channels. This side effect is a common reason for drug failure in preclinical safety trials or for drugs withdrawn from the market.

hERG channel, encoded by the human *Ether-a-go-go*-Related Gene of *KCNH2*, is also a target of class III antiarrhythmic drugs, including amiodarone, sotalol, and dofetilide, which reduce the risk of reentrant arrhythmias by prolonging cardiac AP duration and refractory period without slowing conduction velocity in the myocardium of the heart. The hERG channel also promiscuously interacts with many compounds due to its atypical geometry of the central cavity surrounded by four deep hydrophobic pockets which are unusually sensitive to binding of innumerable non-cardiac drugs (Wang and MacKinnon 2017), including cisapride, loperamide, and iloperidone. As a result, since the mid-1990s, a wide variety of drugs found to be contaminated by this cardiac safety issue have been withdrawn from the market. As a routine in drug discovery and development, intensive screening efforts are carried out in order to eliminate hits that might cause cardiac liability concerns.

Dofetilide, approved by FDA in 1999, is a potent and selective class III antiarrhythmic agent for the maintenance of normal sinus rhythm in patients with atrial fibrillation (AF) and atrial flutter (AFL). Dofetilide selectively blocks the rapid component of I_{Kr}, thus increasing the effective refractory period and action potential duration without affecting the fast inward sodium current. Dofetilide has also been shown to block potassium currents such as K2P2.1 encoded by *KCNK2* gene and Kir2.2 encoded by *KCNJ12* gene.

Dronedarone is an anti-arrhythmia drug developed by Sanofi-Aventis in 2009. Dronedarone is a benzofuran derivative of amiodarone, an effective yet toxic antiarrhythmic drug. Unlike amiodarone, dronedarone does not contain iodine atoms and hence retains the efficacy of amiodarone without its unique toxicity profile. Dronedarone is a multichannel blocker inhibiting several inward potassium currents, such as rapid delayed rectifier, slow delayed rectifier, and ACh-activated inward rectifier. It is also believed to reduce inward rapid Na^+ current and current from L-type Ca^{2+} channels.

2.2.3 Inward Rectifier Potassium Channels (Kir1–7)

Inward rectifier potassium channels (Kir1–7), encoded by *KCNJ1–18* genes, play central roles in control of cellular excitability and K^+ ion homeostasis. Kir channels function in many tissues, such as the brain, heart, sensory, kidney, and endocrine. Kir channels are structurally distinct from the family of voltage-gated K^+ channels, possessing only two membrane-spanning helices without voltage sensor seen in Kv channels and a pore loop. As such, Kir channels have evolved distinct voltage-independent mechanisms for gating (opening and closing), including their gating by G proteins, protons, and ATP. Most Kir channels appear to form as α subunit homotetramers without β subunits. Some Kir6 members must co-assemble with sulfonylurea receptors (SUR) to form an octameric channel, with four Kir6 subunits forming the pore and four SUR subunits surrounding the central ion conductance pathway (Li et al. 2017).

Mutations in Kir channels or defects in their regulation have been shown to result in or associate with diseases, including neuronal degeneration, cardiovascular diseases, diabetes, defective insulin, thyrotoxic periodic paralysis (TPP), and autism spectrum disorders (ASDs) (Cheng et al. 2015; Dogan et al. 2019). Glibenclamide is an antidiabetic drug in a class of medications known as sulfonylureas that are closely related to sulfa drugs. It is a blood glucose-lowering agent used in the treatment of patients with non-insulin-dependent type 2 diabetes mellitus (T2DM). Glibenclamide inhibits the sulfonylurea receptor 1 (SUR1) in pancreatic β cells (Chen et al. 2003; Davies et al. 2005; Zunkler 2006), reducing K^+ current and causing cell membrane depolarization and activation of voltage-dependent calcium channel. Nateglinide, an amino acid d-phenylalanine derivative, belongs to the meglitinide class of blood glucose-lowering medications and acts by inhibiting ATP-dependent potassium channels in the β cells. Nateglinide is an antidiabetic drug for treatment of T2DM. Inhibition of K_{ATP} channel activity depolarizes β cells and causes voltage-gated calcium channels to open. The resulting calcium influx induces fusion of insulin-containing vesicles with the cell membrane, and insulin secretion occurs from the pancreas.

2.2.4 Two-Pore Domain K^+ Channels

The two-pore or tandem-pore K^+ channels (or K2P, KCNK channels), encoded by *KCNK* genes, are a subfamily of 15 members that form so-called the leak or background channels, which is characteristic of Goldman-Hodgkin-Katz rectification. KCNK channels are K^+ selective, and they tend to be constitutively open, thus exerting control over neuronal excitability by shaping the duration, frequency, and amplitude of action potentials. Activation of KCNK channels stabilizes the cell by setting hyperpolarization potential below the firing threshold, whereas the channel suppression facilitates membrane depolarization and cell excitability.

K2P channels play various roles in metabolic regulation, apoptosis, and thermo- and chemo-perception. The mechanically gated TREK (for TWIK-related K^+ channels) subfamily of KCNK channels is composed of three members: TREK-1 (KCNK2), TRAAK (KCNK4), and TREK-2 (KCNK10). These channels are modulated by several physical and chemical stimuli. The compound 2-aminoethoxydiphenyl borate (2-APB) was originally described as an inhibitor of IP_3-induced Ca^{2+} release, but 2-APB has been shown to act as either a blocker or an activator for several TRP ion channels.

K2P channels represent important clinical targets in the treatment of cardiovascular disease, pulmonary arterial hypertension (PAH), and neurological disorders, including pain and depression. Therefore, specific modulators targeting K2P channels can be therapeutically useful for the target validation and design of therapeutics for treatment of relevant diseases. Medications activating K2P channels include the halogenated volatile anesthetics, and drugs such as bupivacaine, quinidine, and fluoxetine can block K2P channels. Inhibitors targeting the extracellular allosteric ligand-binding sites of K2P channels may also be promising for potential therapy (Luo et al. 2017).

2.2.5 Large-Conductance Calcium- and Voltage-Activated Potassium (BK) Channels

The large-conductance calcium-activated potassium (BK) channels, also known as Maxi-K, Slo1, or KCa1.1, encoded by *KCNMA1* gene, have a large unitary conductance of ~100–300 pS. BK channel is dually activated by both membrane depolarization and increase of cytosolic Ca^{2+}, thus coupling electrical signaling to Ca^{2+}-mediated events such as muscle contraction, hormone secretion, and neuronal excitability. BK channel is tetrameric with each subunit containing seven transmembrane domains S0–S6 (Tao et al. 2017). Each subunit is featured of three structural components: a voltage sensor domain (VSD) comprising S1–S4 helices, a pore-forming unit (S5–S6), and a large cytosolic tail domain (CTD) comprising two tandem regulator of K^+ conductance (RCK) domains termed RCK1 and RCK2 being responsible for Ca^{2+} binding and sensing (Jiang et al. 2001). BK channels are also modulated by accessory β subunits, achieving functional and pharmacological diversities. There are four types of β subunits (β1–4) with each type displaying a distinct tissue-specific expression pattern and uniquely modifying gating properties of the channel.

BK channels, as an attractive drug target, play a pivotal and specific role in many pathophysiological conditions, including the regulation of smooth muscle tone and neuronal excitability. Dysfunctional BK channels have been linked to epilepsy (Du et al. 2005; Lee and Cui 2009), hypertension (Yang et al. 2013), urinary incontinence (Herrera et al. 2000; Meredith et al. 2004), and erectile dysfunction (Gonzalez-Corrochano et al. 2013). Attenuation of BK function has been shown to be related to transcriptional silencing of FMR1 gene that encodes fragile X mental retardation protein (FMRP). FMRP is required for normal brain development and neurotransmitter release (Deng et al. 2013). The genome-wide association studies reveal that single nucleotide polymorphisms in the *KCNMA1* gene encoding BK α subunit and the *KCNMAB2* gene encoding the regulatory subunit BKβ2 are strongly linked to the pathophysiology of Alzheimer's disease (AD) (Beecham et al. 2014). These observations suggest that BK channel activators may also have therapeutic potential for cognitive defects in AD and autism spectrum disorder patients.

A BKα gain-of-function mutation (D434G) showing larger macroscopic currents, increased intrinsic gating and Ca^{2+} sensitivity, was associated with generalized epilepsy and paroxysmal dyskinesia (Du et al. 2005). Enhancing neuronal BK activity as a result of D434G mutation increases the fast after-hyperpolarization (fAHP), thus resulting in faster recovery of Nav channels from inactivation and intrinsic neuronal hyperexcitability, ultimately leading to seizures (Matthews 2008; Wang et al. 2009). The residue D434 site is located in the vicinity of the RCK1 Ca^{2+} binding site of BK channel; its mutation likely affects the allosteric coupling between the Ca^{2+} binding and the opening of the channel (Yang et al. 2010). A de novo gain-of-function mutation N995S in BKα gene (*KCNMA1*) associated with epilepsy was also recently identified, and this variant shows increased sensitivity to the channel voltage-dependent activation, without affecting the Ca^{2+}-dependent activation (Li et al. 2017).

Currently there are no specific BK channel modulators used clinically. The design and validation of BK modulators that recognize tissue-specific subunits for therapeutics are challenging due to their diverse expression and pharmacological roles. The recent advances in structural cryo-EM revealing the atomic BK structures in the full length and the channel complex will facilitate the target-based discovery and development of novel BK therapeutics (Hite et al. 2017). Nevertheless, chlorothiazide, for instance, that blocks BK channel α subunit is a thiazide diuretic and antihypertensive agent. As a diuretic chlorothiazide helps prevent body from absorbing too much salt, which can cause fluid retention. Chlorzoxazone, an activator of BK, causes a leftward shift in the activation curve of BK channels, which promotes channel opening under physiological conditions. Chlorzoxazone is a centrally acting muscle relaxant used for treatment of muscle spasm and the resulting pain or discomfort.

3 Sodium Channels

3.1 Introduction of Voltage-Gated Sodium (Nav) Channels

The voltage-gated sodium (Nav) currents are essential for the initiation and propagation of action potentials in excitable cells of nerves, muscles, and heart tissues. Native sodium currents are encoded by genes of α subunits and also β subunits that affect the channel biochemistry and pharmacology. There are nine well-known mammalian isoforms of principal Nav α subunits (Nav1.1–1.9) encoded by nine genes (*SCN1A–11A*) in different chromosomal locations (Fig. 2) that are expressed in different excitable tissues. Nav1.1, 1.2, 1.3, and 1.6 are the primary sodium channels in the central nervous system (CNS). Nav1.7, 1.8, and 1.9 are expressed predominantly in unmyelinated and small diameter myelinated afferents that transmit nociceptive signals in the peripheral nervous system (PNS). Nav1.4 is the primary sodium channel for skeletal muscle contraction, whereas Nav1.5 is primary for the action potential in the heart.

Nav α subunits are composed of approximately 2,000 amino acid residues in a single peptide organized in 4 homologous repeats I–IV, with each repeat containing 6 transmembrane segments (S1–S6). The S4 segment of each repeat contains positively charged arginine and lysine residues acting as voltage sensors of cell membrane depolarization and repolarization. In combination, the S5 and S6 transmembrane helices from each repeat form the sodium channel pore (p loop) containing the Asp/Glu/Lys/Ala (EEKA) signature motif that serves as the Na^+ ion selectivity filter. Nav channels function in three states: closed (resting, nonconducting), open, and inactivated (nonconducting with pore open). The short intracellular loop connecting homologous repeats III and IV of α subunit is responsible for channel fast inactivation. Upon membrane depolarization, the S4 voltage sensors move outward, allowing the pore to open briefly (<1 ms), before fast and slow inactivation processes can occur that move the channel into a nonconducting inactivated state.

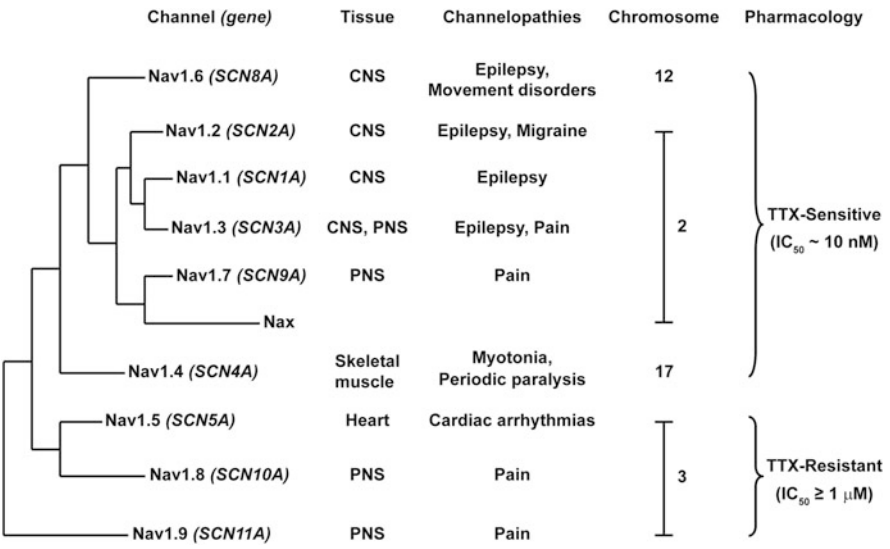


Fig. 2 Phylogenetic relationships, tissue distribution, chromosomal localization, and TTX sensitivity of human voltage-gated Na^+ channel α subunits. *CNS* central nervous system, *PNS* peripheral nervous system, *TTX* tetrodotoxin (Modified from Israel et al. 2017)

3.2 Pharmacological Modulations and Implications of Nav Channels

Natural toxins are powerful tools known to exert their effects through inhibiting Nav channels. Tetrodotoxin (TTX) is a potent marine neurotoxin that as a true blocker physically occludes the extracellular portion of channel pore. Nav1.1–1.9 channels have been broadly classified based on their pharmacology and gating kinetics with members of Nav1.1–Nav1.4 and Nav1.6–Nav1.7 being sensitive to block by tetrodotoxin (TTX-sensitive). Nav1.5, Nav1.8, and Nav1.9 are TTX-resistant, and they have much slower inactivation kinetics that produce persistent currents for up to several hundred milliseconds.

Genetic mutations of Nav1.1 and Nav1.2 channels are linked to epilepsy and CNS-related disorders. Periodic paralyses are caused by mutations in Nav1.4. Nav1.5 mutations have been linked to a variety of cardiac diseases, including cardiac arrhythmia such as long QT syndrome (LQTs), Brugada syndrome, cardiac conduction defect, atrial fibrillation, and dilated cardiomyopathy. Nav1.6 channel has been associated with cerebellar atrophy, behavioral deficits, and ataxia. Because of their differential expression and presence in sensory neurons for their role in pain transmission, Nav1.3, 1.7, 1.8, and 1.9 differentially expressed in peripheral sensory neurons have garnered much attention as promising targets for development of novel analgesics. Among the nine isoforms, Nav1.7 channel, in particular, has generated great interest as a promising target for development of pain therapeutics based on the identifications of genetic mutations of the channel and understanding of the

pharmacology of marketed drugs. The gain-of-function mutations in Nav1.7 have been shown to be associated with primary erythromelalgia characterized with burning pain in the extremities accompanied with hyperemia and inflammation (Yang et al. 2004). Conversely, loss-of-function mutations of Nav1.7 can cause congenital insensitivity to pain (Cox et al. 2006). These compelling findings have promoted considerable efforts to develop selective Nav1.7 inhibitors as analgesics.

Common small-molecule drugs blocking Nav channels are clinically prescribed as local anesthetics, anticonvulsants, and antiarrhythmics. Most of these drugs lack Nav subtype selectivity although exhibiting some distinct pharmacology based on differences in their binding affinities to the channel. Oxcarbazepine is an anticonvulsant or antiepileptic drug (AED), used primarily for the treatment of partial seizures in adults with epilepsy and for the adjunctive treatment of partial seizures in children. The precise mechanism by which oxcarbazepine exerts antiseizure effects remains elusive; however in vitro electrophysiological studies indicate that oxcarbazepine blocks voltage-sensitive sodium channels, resulting in the stabilization of hyperexcited neural membranes and inhibition of repetitive neuronal firing. Ranolazine has been shown to exert its antianginal and anti-ischemic effects without reducing heart rate or blood pressure. Ranolazine reduces Na^+ influx and ameliorates disturbed Na^+ and Ca^{2+} homeostasis. The principal mechanism underlying ranolazine's antiarrhythmic actions is thought to be primarily via inhibition of the persistent or late sodium currents in the ventricles. The effects of ranolazine on Nav1.7 and Nav1.8 channels also render it potentially useful in the treatment of neuropathic pain.

Nav channels have been implicated in a variety of diseases, but identifying subtype or functionally selective agents still remains a huge challenge in drug discovery. To date, no Nav1.7-selective drugs have reached the clinic. The demonstration that loss of Nav1.7 function is associated with upregulation of endogenous opioids and potentiation of mu- and delta-opioid receptor activities suggests that targeting only Nav1.7 may be insufficient for analgesia. Nevertheless, advances in screening technology and recent cryo-EM structures solved, such as human Nav1.7 in complex with auxiliary subunits and animal toxins or human Nav1.2, may establish the foundation of structure-based development of subtype specific analgesics or therapeutic agents (Shen et al. 2019).

4 Chloride Channels

4.1 Introduction of Cl^- Channels

Chloride ions are by far the most abundant anion in all organisms. Chloride channels present in every cell are probably the most important pathway to allow chloride to go through the cell membrane and are involved in many physiological functions, including control of transepithelial fluid secretion, muscle contraction, neuroexcitation, and regulation of cell volume and intracellular pH. Chloride channels are diversified in molecular structures, for example, Ca^{2+}-activated Cl^-

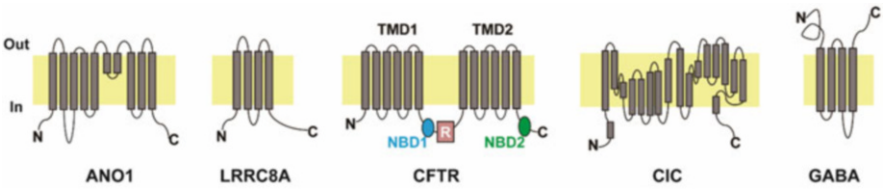


Fig. 3 Schematic membrane topology of chloride channels. Ca^{2+}-activated Cl^- channel ANO1 protein contains ten transmembrane segments with N- and C-terminal faced to the cytoplasm. Volume-regulated Cl^- channel LRRC8A protein consists of four transmembrane segments, intracellular N-terminal domain, and a C-terminal domain containing 15–17 predicted leucine-rich repeats. CFTR protein consists of two transmembrane domains (TMD) and two nucleotide-binding domains (NBD). Each TMD contains six transmembrane helixes, and two homologous TMD-NBD are linked by a specific cytosolic regulatory domain (R domain). ClC consists of two similar subunits with the double-pore architecture, and each subunit contains 18 α-helices to form the channel pore. The $GABA_A$ receptor contains five subunits with each subunit consisting of four transmembrane segments, with both the N- and C-terminal located extracellularly

channel ANO1 is identified as a homodimer with each subunit consisting of ten membrane-spanning α-helices, and volume-regulated Cl^- channel LRRC8A functions as a heterohexamer with each subunit consisting of four transmembrane segments (Fig. 3). Mammalian chloride channels can be roughly divided into five classes based on their regulation: calcium-activated chloride channels (CaCCs), volume-regulated chloride channel (VRCC), cystic fibrosis transmembrane conductance regulator (CFTR), voltage-gated chloride channels (ClCs), and ligand-gated chloride channels (GABA and glycine activated).

Under physiological conditions, the extracellular Cl^- concentration is higher than cytoplasmic concentration, and the equilibrium potential E_{Cl} is near the resting membrane potential due to the absence of the Cl^- pump. Thus, Cl^- channels can stabilize the membrane potential. Chloride channels are also permeable to many anions, including I^-, Br^-, NO_3^-, HCO_3^-, SCN^-, and some small organic acids, which are different from cation channels with high selectivity to a specific cation.

Chloride channels have received less attention for a very long time until the first chloride channel ClC-0 and CFTR were cloned (Jentsch et al. 1990; Riordan et al. 1989). Mutations in chloride channel genes have been identified to induce a variety of human diseases, including cystic fibrosis (CF) caused by mutations in CFTR, cancers and inflammatory airway diseases related to CaCCs dysfunctions, and osteopetrosis induced by mutations in ClCs. Chloride channels are considered to be potential drug targets for different diseases.

4.2 Target Validation and Drugs Targeting Cl^- for Medical Uses

4.2.1 Calcium-Activated Chloride Channels (CaCCs)

The calcium-activated chloride channels (CaCCs) were first described from *Xenopus* oocytes in 1982 (Miledi 1982). The intracellular Ca^{2+} rise during the fertilization

of oocytes activates CaCCs, and opening of CaCCs causes a depolarization of the membrane to prevent the fusion of additional sperm. CaCCs are also found in a variety of organs and tissues ranging from invertebrates to mammals, including epithelium, glands, smooth muscle, and neurons.

The molecular basis of CaCCs was disputed until 2008 when three independent groups reported that an anoctamin1 (ANO1), also named transmembrane protein 16A (TMEM16A), underlies the molecular identity of CaCCs (Caputo et al. 2008; Schroeder et al. 2008; Yang et al. 2008). ANO1 belongs to anoctamin family that has ten members (ANO1–ANO10) in vertebrates. ANO2 is also found to mediate Ca^{2+}-activated chloride currents in olfactory sensory neurons, photoreceptor terminals, and hippocampus (Huang et al. 2012b; Stephan et al. 2009; Stohr et al. 2009). It is not clear whether other ANOs also function as CaCCs or Ca^{2+}-dependent membrane phospholipid scramblases or other physiological proteins.

ANO1 channel functions as a homodimer with each subunit containing cytosolic N- and C-terminal domains and a transmembrane unit consisting of ten membrane-spanning α-helices. The α3–α7 helices form ion pore and are also critical for Ca^{2+} binding (Fig. 3) (Brunner et al. 2014; Dang et al. 2017; Paulino et al. 2017). The ubiquitous expression of CaCCs indicates a variety of important physiological functions, including regulation of epithelial Cl^- secretion, excitability of neuronal and cardiac cells, regulation of smooth muscle contraction, and nociception. Dysfunctional CaCCs can cause pathological conditions and diseases such as asthma, inflammatory airway and bowel diseases, diarrhea, hypertension, and cancer.

Because of its important physiological and pathological roles, ANO1 has become an emerging therapeutic target for different potential indications. Many small-molecule modulators from synthesis and natural products were identified, and, so far, there are no reported ANO1 modulators that are in clinical use or trials. The broad-spectrum blockers of CaCCs, such as NFA, NPPB, DIDS, and 9AC are commonly used as research tools. Recent CaCC blockers, including $CaCC_{inh}$-A01, tannic acid, $T16A_{inh}$-A01, MONNA, and Ani9, show higher efficacy and more selectivity over inhibition of ANO1 (De La Fuente et al. 2008; Namkung et al. 2011; Oh et al. 2013; Seo et al. 2016). These inhibitors have been used in many studies as pharmacological probes to investigate the physiological and pathological involvement of ANO1 in different tissues and diseases. Interestingly, a uricosuric agent benzbromarone used clinically shows an inhibitory effect on ANO1 with an IC_{50} of 10 μM (Huang et al. 2012a). Benzbromarone significantly impairs mucus secretion in primary human airway surface epithelial cells and reduces human airway smooth muscle contraction through inhibition of ANO1.

Until now, only a few activators of ANO1, including Eact, Fact, and some natural products (resveratrol, Ginsenoside Rb1, chitosan oligosaccharides) (Ji et al. 2019), have been reported in the last few decades, which can be useful for further understanding of ANO1 function and validation of the channel as potential target for ANO1 downregulated syndrome and disorders.

4.2.2 Volume-Regulated Chloride Channel (VRCC)

The volume-regulated chloride channel (VRCC) is widely expressed in most vertebrate cells and can be activated by cell swelling. Besides the important role in maintaining cell volume, VRCC is also implicated in many other physiological processes including cell proliferation, migration, apoptosis, and release of excitatory amino acids (Strange et al. 2019).

The family of leucine-rich repeat-containing 8 (LRRC8) proteins was recently identified as essential components of VRCC (Qiu et al. 2014; Voss et al. 2014). LRRC8 family contains five members (LRRC8A-E), whereas LRRC8A is the only obligatory subunit of VRCC, and at least one other isoform is still needed to form a functional channel (Syeda et al. 2016). The high-resolution cryo-EM structures of LRRC8 channel was recently described as a hexameric symmetry with each subunit consisting of four transmembrane segments, intracellular N-terminal domain, and a C-terminal domain containing 15–17 predicted leucine-rich repeats (Deneka et al. 2018; Kefauver et al. 2018).

The LRRC8A was first described in a female patient with congenital agammaglobulinemia characterized by defective B-cell development (Sawada et al. 2003). In the patient, the last 91 C-terminal amino acids were replaced by 35 noncoding amino acids. Mice lacking LRRC8A exhibit growth retardation, hind limb weakness, and infertility (Chen et al. 2019). Recent studies also suggest that LRRC8A is involved in several human diseases including stroke, diabetes, and cancer. Although the importance of physiological and pathological function of VRCC is identified, there is lack of a specific modulator and no mention to clinical use by now. There are only some anion channel inhibitors such as DIDS, DCPIB, and NPPB that are currently used in research (Friard et al. 2017).

4.2.3 Cystic Fibrosis Transmembrane Conductance Regulator (CFTR)

Cystic fibrosis transmembrane conductance regulator (CFTR) is a phosphor-regulated chloride channel that belongs to the ATP-binding cassette (ABC) family (Riordan et al. 1989). The CFTR is mainly expressed in the apical membrane of epithelia and involved in the epithelial fluid secretion (Moskwa et al. 2007). The CFTR consists of two membrane-spanning domains (MSD) and two nucleotide-binding domains (NBD). Each MSD contains six transmembrane helixes, and two homologous TMD-NBD are linked by a specific cytosolic regulatory domain (R domain). CFTR channels can open upon the phosphorylation of R domain by cAMP-dependent protein kinase (PKA) as well as NBDs binding with ATP (Csanady et al. 2019).

Mutations in the CFTR gene cause cystic fibrosis (CF), and over 2,000 variants have been identified (Strug et al. 2018). Among the 327 CF-related mutations, one of the most common CF-causing mutation is F508del, lacking a phenylalanine at position 508. F508del has a major effect on the stability of CFTR protein and shows poor channel activity.

The CFTR serves to be a fundamental therapeutic target for CF. Several CFTR potentiators or correctors have been approved for treatment of CF, and some are now in clinical trials (Burgener and Moss 2018; Gentzsch and Mall 2018; Strug et al.

2018). Ivacaftor is the first approved CFTR potentiator by FDA in 2012 for improvement in lung function. However, ivacaftor is only effective for nearly 10% of the population of CF patients because of the various mutations of CFTR. In 2015, a CFTR corrector lumacaftor combined with ivacaftor was approved by FDA for correction of F508del dysfunction. These two drugs given together can help 50% of CF patients but with significant side effects, including worsening shortness of breath and chest tightness. A new CFTR corrector, tezacaftor, in combination with ivacaftor was approved by FDA in 2018. The combination shows fewer side effects in F508del homozygous patients and also shows efficacious in heterozygous patients (Burgener and Moss 2018).

Several triple combination therapies for CF are now in clinical trials, including tezacaftor and ivacaftor combined with VX-445 or VX-659 in a phase III trial and tezacaftor and ivacaftor with VX-152 or VX-440 in a phase II trial. Several other compounds, including QBW251, FDL169, VX-561, GLPG1837, and GLPG2222, are now in phase II trials. There are still several CFTR correctors and potentiators in clinical phase I stage (Gentzsch and Mall 2018).

4.2.4 Voltage-Gated Chloride Channels (ClCs)

The first ClC family member ClC-0, most closely related to voltage-gated ClC-1 chloride channel encoded by *CLCN1* gene, was discovered in *Torpedo* in 1990 (Jentsch et al. 1990). Among the nine members within the family, four of them, ClC-1, ClC-2, ClC-Ka, and ClC-Kb, were found to function as chloride channels at the cell membrane, while the other five members (ClC-3 to ClC-7) function as Cl^-/H^+ exchangers and are generally localized in the cytoplasmic membrane, such as endosome and lysosome.

The high-resolution structures of ClCs reveal that ClCs consist of two similar subunits with the double-pore architecture and each subunit contains 18 α-helices to form the channel pore. Most helices are partially embedded in the membrane, forming the ion selectivity filter within the pore (Jentsch and Pusch 2018; Poroca et al. 2017).

Apart from the skeletal muscle-specific ClC-1 and neuronal-specific ClC-6, other ClC members are widely expressed. ClC proteins are involved in a variety of physiological processes. ClC-1 is critical for stabilization of membrane potential in skeletal muscle, and impairment of ClC-1 leads to myotonia. ClC-2 is widely expressed and is important to transepithelial transport of fluid. Mutations in *CLCN2* gene encoding ClC-2 cause leukodystrophy symptoms. ClC-K proteins are mainly expressed in the kidney and inner ear, mediating the fluid transepithelial transport. ClC-K functions in combination with Barttin which is a 40 kD protein with two transmembrane domains. The mutations of ClC-K or Barttin cause Bartter syndrome type 3 or 4. No human diseases have been found to associate with ClC-3 or ClC-6 by far. ClC-4 is related to mental retardation epilepsy, while ClC-5 contributes to Dent's disease characterized by tubular proteinuria, and ClC-7 is related to osteopetrosis/retinal degeneration/lysosomal storage disease.

Unfortunately, no high affinity and selective modulators are available for ClC proteins. Some classical nonspecific Cl^- channel inhibitors including 9-AC, DIDS,

and CPP have been shown to exert inhibitory effect on ClC-1 and ClC-K, but weakly sensitive to ClC-2, ClC-5, and ClC-7. A clofibric acid derivative (R-isofomer of CPP) and acetazolamide (a carbonic anhydrase inhibitor) were reported to increase ClC-1 current. Benzofuran derivatives were found to block ClC-K channel. SRA-36 was recently discovered to inhibit wild-type ClC-K channels and some hypertension-related mutations of ClC-K.

4.2.5 Ligand-Gated Chloride Channels (GABA)

Ligand-gated chloride channels predominantly refer to GABA and glycine receptors. GABA and glycine receptors are mainly located in the central nervous system and mediate synaptic inhibitory signaling. When the endogenous ligand binds to the receptor, the channel opens and allows Cl^- and HCO_3^- flux into cells, usually followed by inhibition of neuronal excitability (Martinez and Mohiuddin 2019).

GABA receptors contain two subfamily, ionotropic $GABA_A$ receptors and metabotropic $GABA_B$ receptors. $GABA_A$ receptors are cys-loop ligand-gated chloride channels which are found in the thalamus, hypothalamus, basal ganglia, and hippocampus synapse. The predominant synaptic $GABA_A$ receptors comprise two α1 subunits, two β2 subunits, and one γ2 subunit. Five subunits assemble in a pseudosymmetrical pentamer with each subunit containing four transmembrane segments (Zhu et al. 2018).

Mutations of GABA receptors have been identified in various human neurologic diseases, including epilepsy, autism spectrum disorder, schizophrenia, and addiction (Yuan et al. 2015). GABA is a therapeutic target for these neurologic diseases. A variety of clinical drugs are designed targeting $GABA_A$ receptors. Benzodiazepines function as GABA agonist and exhibit anticonvulsant effect. Barbiturates also target on GABA receptor and have sedation effect. Zolpidem is a GABA agonist which is indicated for the short-term treatment of insomnia characterized by difficulties with sleep initiation. Flumazenil is a GABA antagonist, exerting the reverse effects on GABA activation. There are many other clinical drugs that also target GABA receptor including acamprosate calcium for alcohol addition, propofol for anesthetic, zopiclone for insomnia, and etomidate for short-acting anesthetic and sedative. Several tool compounds are commonly used in research for blockage of GABA current, including bicuculline, picrotoxin, gabazine, and RU5135 (Clar and Maani 2019; George and Sadiq 2019).

In summary, ion channels are membrane proteins critical for a variety of physiological and pharmacological functions. Dysfunctional ion channels can cause diseases known as channelopathies. Ion channels represent the second largest class of proven drug targets for therapeutic areas of neuropsychiatric disorders, cardiovascular and metabolic diseases, immunological diseases, nephrological diseases, gastrointestinal diseases, pulmonary/respiratory diseases, and many cancers. Identifying specific modulators targeting ion channels with either small molecules or biologics is both challenging due to selectivity issues and hurdles to overcome and also providing exciting opportunities as more ion channel structures are being solved with emerging new computational techniques and functional robust assays with higher throughput are being developed.

References

Beecham GW, Hamilton K, Naj AC, Martin ER, Huentelman M, Myers AJ, Corneveaux JJ, Hardy J, Vonsattel JP, Younkin SG et al (2014) Genome-wide association meta-analysis of neuropathologic features of Alzheimer's disease and related dementias. PLoS Genet 10: e1004606

Biervert C, Schroeder BC, Kubisch C, Berkovic SF, Propping P, Jentsch TJ, Steinlein OK (1998) A potassium channel mutation in neonatal human epilepsy. Science 279:403–406

Brown DA, Adams PR (1980) Muscarinic suppression of a novel voltage-sensitive K+ current in a vertebrate neurone. Nature 283:673–676

Brunner JD, Lim NK, Schenck S, Duerst A, Dutzler R (2014) X-ray structure of a calcium-activated TMEM16 lipid scramblase. Nature 516:207–212

Burgener EB, Moss RB (2018) Cystic fibrosis transmembrane conductance regulator modulators: precision medicine in cystic fibrosis. Curr Opin Pediatr 30:372–377

Caputo A, Caci E, Ferrera L, Pedemonte N, Barsanti C, Sondo E, Pfeffer U, Ravazzolo R, Zegarra-Moran O, Galietta LJ (2008) TMEM16A, a membrane protein associated with calcium-dependent chloride channel activity. Science 322:590–594

Charlier C, Singh NA, Ryan SG, Lewis TB, Reus BE, Leach RJ, Leppert M (1998) A pore mutation in a novel KQT-like potassium channel gene in an idiopathic epilepsy family. Nat Genet 18:53–55

Chen M, Dong Y, Simard JM (2003) Functional coupling between sulfonylurea receptor type 1 and a nonselective cation channel in reactive astrocytes from adult rat brain. J Neurosci 23:8568–8577

Chen L, Konig B, Liu T, Pervaiz S, Razzaque YS, Stauber T (2019) More than just a pressure relief valve: physiological roles of volume-regulated LRRC8 anion channels. Biol Chem 400:1481–1496

Cheng CJ, Sung CC, Huang CL, Lin SH (2015) Inward-rectifying potassium channelopathies: new insights into disorders of sodium and potassium homeostasis. Pediatr Nephrol 30:373–383

Clar DT, Maani CV (2019) Physiology, ligand gated chloride channel. StatPearls, Treasure Island

Cox JJ, Reimann F, Nicholas AK, Thornton G, Roberts E, Springell K, Karbani G, Jafri H, Mannan J, Raashid Y et al (2006) An SCN9A channelopathy causes congenital inability to experience pain. Nature 444:894–898

Csanady L, Vergani P, Gadsby DC (2019) Structure, gating, and regulation of the Cftr anion channel. Physiol Rev 99:707–738

Dang S, Feng S, Tien J, Peters CJ, Bulkley D, Lolicato M, Zhao J, Zuberbuhler K, Ye W, Qi L et al (2017) Cryo-EM structures of the TMEM16A calcium-activated chloride channel. Nature 552:426–429

Davies GC, Thornton MJ, Jenner TJ, Chen YJ, Hansen JB, Carr RD, Randall VA (2005) Novel and established potassium channel openers stimulate hair growth in vitro: implications for their modes of action in hair follicles. J Invest Dermatol 124:686–694

De La Fuente R, Namkung W, Mills A, Verkman AS (2008) Small-molecule screen identifies inhibitors of a human intestinal calcium-activated chloride channel. Mol Pharmacol 73:758–768

Deneka D, Sawicka M, Lam AKM, Paulino C, Dutzler R (2018) Structure of a volume-regulated anion channel of the LRRC8 family. Nature 558:254–259

Deng PY, Rotman Z, Blundon JA, Cho Y, Cui J, Cavalli V, Zakharenko SS, Klyachko VA (2013) FMRP regulates neurotransmitter release and synaptic information transmission by modulating action potential duration via BK channels. Neuron 77:696–711

Dogan MF, Yildiz O, Arslan SO, Ulusoy KG (2019) Potassium channels in vascular smooth muscle: a pathophysiological and pharmacological perspective. Fundam Clin Pharmacol 33:504–523

Du W, Bautista JF, Yang HH, Diez-Sampedro A, You SA, Wang LJ, Kotagal P, Luders HO, Shi JY, Cui JM et al (2005) Calcium-sensitive potassium channelopathy in human epilepsy and paroxysmal movement disorder. Nat Genet 37:733–738

Friard J, Tauc M, Cougnon M, Compan V, Duranton C, Rubera I (2017) Comparative effects of chloride channel inhibitors on LRRC8/VRAC-mediated chloride conductance. Front Pharmacol 8:328

Gentzsch M, Mall MA (2018) Ion channel modulators in cystic fibrosis. Chest 154:383–393

George K, Sadiq NM (2019) GABA inhibitors. StatPearls, Treasure Island

Gonzalez-Corrochano R, La Fuente JM, Cuevas P, Fernandez A, Chen MX, de Tejada IS, Angulo J (2013) Ca2+-activated K+ channel (K-Ca) stimulation improves relaxant capacity of PDE5 inhibitors in human penile arteries and recovers the reduced efficacy of PDE5 inhibition in diabetic erectile dysfunction. Br J Pharmacol 169:449–461

Herrera GM, Heppner TJ, Nelson MT (2000) Regulation of urinary bladder smooth muscle contractions by ryanodine receptors and BK and SK channels. Am J Phys Regul Integr Comp Phys 279:R60–R68

Hille B (2001) Ion channels of excitable membranes, 3rd edn. Sinauer Associates Inc., Sunderland

Hite RK, Tao X, MacKinnon R (2017) Structural basis for gating the high-conductance Ca(2+)-activated K(+) channel. Nature 541:52–57

Huang F, Zhang H, Wu M, Yang H, Kudo M, Peters CJ, Woodruff PG, Solberg OD, Donne ML, Huang X et al (2012a) Calcium-activated chloride channel TMEM16A modulates mucin secretion and airway smooth muscle contraction. Proc Natl Acad Sci U S A 109:16354–16359

Huang WC, Xiao S, Huang F, Harfe BD, Jan YN, Jan LY (2012b) Calcium-activated chloride channels (CaCCs) regulate action potential and synaptic response in hippocampal neurons. Neuron 74:179–192

Israel MR, Tay B, Deuis JR, Vetter I (2017) Sodium channels and venom peptide pharmacology. Adv Pharmacol 79:67–116

Jentsch TJ, Pusch M (2018) CLC chloride channels and transporters: structure, function, physiology, and disease. Physiol Rev 98:1493–1590

Jentsch TJ, Steinmeyer K, Schwarz G (1990) Primary structure of Torpedo marmorata chloride channel isolated by expression cloning in Xenopus oocytes. Nature 348:510–514

Ji Q, Guo S, Wang X, Pang C, Zhan Y, Chen Y, An H (2019) Recent advances in TMEM16A: structure, function, and disease. J Cell Physiol 234:7856–7873

Jiang Y, Pico A, Cadene M, Chait BT, MacKinnon R (2001) Structure of the RCK domain from the E. coli K+ channel and demonstration of its presence in the human BK channel. Neuron 29:593–601

Keating MT, Sanguinetti MC (2001) Molecular and cellular mechanisms of cardiac arrhythmias. Cell 104:569–580

Kefauver JM, Saotome K, Dubin AE, Pallesen J, Cottrell CA, Cahalan SM, Qiu Z, Hong G, Crowley CS, Whitwam T et al (2018) Structure of the human volume regulated anion channel. elife 7

Kubisch C, Schroeder BC, Friedrich T, Lutjohann B, El-Amraoui A, Marlin S, Petit C, Jentsch TJ (1999) KCNQ4, a novel potassium channel expressed in sensory outer hair cells, is mutated in dominant deafness. Cell 96:437–446

Lee US, Cui JM (2009) Beta subunit-specific modulations of BK channel function by a mutation associated with epilepsy and dyskinesia. J Physiol 587:1481–1498

Li N, Wu JX, Ding D, Cheng J, Gao N, Chen L (2017) Structure of a pancreatic ATP-sensitive potassium channel. Cell 168:101–110 e110

Luo Q, Chen L, Cheng X, Ma Y, Li X, Zhang B, Li L, Zhang S, Guo F, Li Y et al (2017) An allosteric ligand-binding site in the extracellular cap of K2P channels. Nat Commun 8:378

Martinez AH, Mohiuddin SS (2019) Biochemistry, chloride channels. StatPearls, Treasure Island

Matthews EA, Weible AP, Shah S, Disterhoft JF (2008) The BK-mediated fAHP is modulated by learning a hippocampus-dependent task. Proc Natl Acad Sci U S A 105(39):15154–15159

Meredith AL, Thorneloe KS, Werner ME, Nelson MT, Aldrich RW (2004) Overactive bladder and incontinence in the absence of the BK large conductance Ca2+-activated K+ channel. J Biol Chem 279:36746–36752

Miledi R (1982) A calcium-dependent transient outward current in Xenopus laevis oocytes. Proc R Soc London Ser B 215:491–497

Miller C (ed) (1986) Ion channel reconstitution. Plenum, New York. https://doi.org/10.1007/978-1-4757-1361-9

Moskwa P, Lorentzen D, Excoffon KJ, Zabner J, McCray PB Jr, Nauseef WM, Dupuy C, Banfi B (2007) A novel host defense system of airways is defective in cystic fibrosis. Am J Respir Crit Care Med 175:174–183

Namkung W, Phuan PW, Verkman AS (2011) TMEM16A inhibitors reveal TMEM16A as a minor component of calcium-activated chloride channel conductance in airway and intestinal epithelial cells. J Biol Chem 286:2365–2374

Neyroud N, Tesson F, Denjoy I, Leibovici M, Donger C, Barhanin J, Faure S, Gary F, Coumel P, Petit C et al (1997) A novel mutation in the potassium channel gene KVLQT1 causes the Jervell and Lange-Nielsen cardioauditory syndrome. Nat Genet 15:186–189

Oh SJ, Hwang SJ, Jung J, Yu K, Kim J, Choi JY, Hartzell HC, Roh EJ, Lee CJ (2013) MONNA, a potent and selective blocker for transmembrane protein with unknown function 16/anoctamin-1. Mol Pharmacol 84:726–735

Paulino C, Kalienkova V, Lam AKM, Neldner Y, Dutzler R (2017) Activation mechanism of the calcium-activated chloride channel TMEM16A revealed by cryo-EM. Nature 552:421–425

Poroca DR, Pelis RM, Chappe VM (2017) ClC channels and transporters: structure, physiological functions, and implications in human chloride channelopathies. Front Pharmacol 8:151

Qiu Z, Dubin AE, Mathur J, Tu B, Reddy K, Miraglia LJ, Reinhardt J, Orth AP, Patapoutian A (2014) SWELL1, a plasma membrane protein, is an essential component of volume-regulated anion channel. Cell 157:447–458

Riordan JR, Rommens JM, Kerem B, Alon N, Rozmahel R, Grzelczak Z, Zielenski J, Lok S, Plavsic N, Chou JL et al (1989) Identification of the cystic fibrosis gene: cloning and characterization of complementary DNA. Science 245:1066–1073

Sanguinetti MC, Jiang C, Curran ME, Keating MT (1995) A mechanistic link between an inherited and an acquired cardiac arrhythmia: HERG encodes the IKr potassium channel. Cell 81(2):299–307

Sawada A, Takihara Y, Kim JY, Matsuda-Hashii Y, Tokimasa S, Fujisaki H, Kubota K, Endo H, Onodera T, Ohta H et al (2003) A congenital mutation of the novel gene LRRC8 causes agammaglobulinemia in humans. J Clin Invest 112:1707–1713

Schroeder BC, Cheng T, Jan YN, Jan LY (2008) Expression cloning of TMEM16A as a calcium-activated chloride channel subunit. Cell 134:1019–1029

Seo Y, Lee HK, Park J, Jeon DK, Jo S, Jo M, Namkung W (2016) Ani9, a novel potent small-molecule ANO1 inhibitor with negligible effect on ANO2. PLoS One 11:e0155771

Shen H, Liu D, Wu K, Lei J, Yan N (2019) Structures of human Nav1.7 channel in complex with auxiliary subunits and animal toxins. Science 363:1303–1308

Singh NA, Charlier C, Stauffer D, DuPont BR, Leach RJ, Melis R, Ronen GM, Bjerre I, Quattlebaum T, Murphy JV et al (1998) A novel potassium channel gene, KCNQ2, is mutated in an inherited epilepsy of newborns. Nat Genet 18:25–29

Stephan AB, Shum EY, Hirsh S, Cygnar KD, Reisert J, Zhao H (2009) ANO2 is the cilial calcium-activated chloride channel that may mediate olfactory amplification. Proc Natl Acad Sci U S A 106:11776–11781

Stohr H, Heisig JB, Benz PM, Schoberl S, Milenkovic VM, Strauss O, Aartsen WM, Wijnholds J, Weber BH, Schulz HL (2009) TMEM16B, a novel protein with calcium-dependent chloride channel activity, associates with a presynaptic protein complex in photoreceptor terminals. J Neurosci 29:6809–6818

Strange K, Yamada T, Denton JS (2019) A 30-year journey from volume-regulated anion currents to molecular structure of the LRRC8 channel. J Gen Physiol 151:100–117

Strug LJ, Stephenson AL, Panjwani N, Harris A (2018) Recent advances in developing therapeutics for cystic fibrosis. Hum Mol Genet 27:R173–R186

Syeda R, Qiu Z, Dubin AE, Murthy SE, Florendo MN, Mason DE, Mathur J, Cahalan SM, Peters EC, Montal M et al (2016) LRRC8 proteins form volume-regulated anion channels that sense ionic strength. Cell 164:499–511

Tao X, Hite RK, MacKinnon R (2017) Cryo-EM structure of the open high-conductance Ca(2+)-activated K(+) channel. Nature 541:46–51

Trudeau MC, Warmke JW, Ganetzky B, Robertson GA (1995) HERG, a human inward rectifier in the voltage-gated potassium channel family. Science 269(5220):92–95

Voss FK, Ullrich F, Munch J, Lazarow K, Lutter D, Mah N, Andrade-Navarro MA, von Kries JP, Stauber T, Jentsch TJ (2014) Identification of LRRC8 heteromers as an essential component of the volume-regulated anion channel VRAC. Science 344:634–638

Wang W, MacKinnon R (2017) Cryo-EM structure of the open human ether-a-go-go-related K(+) channel hERG. Cell 169:422–430 e410

Wang B, Rothberg BS, Brenner R (2009) Mechanism of increased BK channel activation from a channel mutation that causes epilepsy. J Gen Physiol 133(3):283–294

Wang HS, Pan Z, Shi W, Brown BS, Wymore RS, Cohen IS, Dixon JE, McKinnon D (1998) KCNQ2 and KCNQ3 potassium channel subunits: molecular correlates of the M-channel. Science 282:1890–1893

Wulff H, Castle NA, Pardo LA (2009) Voltage-gated potassium channels as therapeutic targets. Nat Rev Drug Discov 8:982–1001

Yang Y, Wang Y, Li S, Xu Z, Li H, Ma L, Fan J, Bu D, Liu B, Fan Z et al (2004) Mutations in SCN9A, encoding a sodium channel alpha subunit, in patients with primary erythermalgia. J Med Genet 41:171–174

Yang YD, Cho H, Koo JY, Tak MH, Cho Y, Shim WS, Park SP, Lee J, Lee B, Kim BM et al (2008) TMEM16A confers receptor-activated calcium-dependent chloride conductance. Nature 455:1210–1215

Yang Y, Li PY, Cheng J, Mao L, Wen J, Tan XQ, Liu ZF, Zeng XR (2013) Function of BKCa channels is reduced in human vascular smooth muscle cells from Han Chinese patients with hypertension. Hypertension 61:519–525

Yang J, Krishnamoorthy G, Saxena A, Zhang G, Shi J, Yang H et al (2010) An epilepsy/dyskinesia-associated mutation enhances BK channel activation by potentiating Ca2+ sensing. Neuron 66 (6):871–883

Yuan H, Low CM, Moody OA, Jenkins A, Traynelis SF (2015) Ionotropic GABA and glutamate receptor mutations and human neurologic diseases. Mol Pharmacol 88:203–217

Zhang F, Liu Y, Tang F, Liang B, Chen H, Zhang H, Wang K (2019) Electrophysiological and pharmacological characterization of a novel and potent neuronal Kv7 channel opener SCR2682 for antiepilepsy. FASEB J 33:9154–9166

Zhu S, Noviello CM, Teng J, Walsh RM Jr, Kim JJ, Hibbs RE (2018) Structure of a human synaptic GABAA receptor. Nature 559:67–72

Zunkler BJ (2006) Human ether-a-go-go-related (HERG) gene and ATP-sensitive potassium channels as targets for adverse drug effects. Pharmacol Ther 112:12–37

Genetically Encoded Fluorescent Calcium and Voltage Indicators

Irene Mollinedo-Gajate, Chenchen Song, and Thomas Knöpfel

Contents

1 Introduction 210
2 Calcium Versus Voltage Imaging 210
3 Advantages of Genetically Encoding Indicators 211
4 Key Features of Genetically Encoded Calcium and Voltage Indicators 212
4.1 GECIs 212
4.2 GEVIs 214
4.3 Features Common to GEVIs and GECIs 216
5 Currently Available Indicators 218
5.1 Genetically Encoded Calcium Indicators 218
5.2 Genetically Encoded Voltage Indicators 219
5.3 Bioluminescent Indicators 223
6 Applications 223
6.1 Monitoring Action Potentials in Awake Mouse Cortex 224
6.2 Mesoscopic Imaging of Neuronal Circuit Dynamics 224
6.3 Monitoring Astrocytic Activity 225
6.4 Combination of GECIs and GEVIs with Optogenetic Interference 225
6.5 Calcium and Voltage Imaging to Monitor Cardiac Cells 226
7 Outlook 226
References 226

Abstract

Fluorescent probes that indicate biologically important quantities are widely used for many different types of biological experiments across life sciences. During recent years, limitations of small molecule-based indicators have been overcome by the development of genetically encoded indicators. Here we focus on fluorescent calcium and voltage indicators and point to their applications mainly in neurosciences.

I. Mollinedo-Gajate · C. Song · T. Knöpfel (✉)
Laboratory for Neuronal Circuit Dynamics, Imperial College London, London, UK
e-mail: tknopfel@knopfel-lab.net

J. E. Barrett et al. (eds.), *Concepts and Principles of Pharmacology*,
Handbook of Experimental Pharmacology 260, https://doi.org/10.1007/164_2019_299

Keywords
GECI application · Genetically encoded calcium indicators · Genetically encoded voltage indicators · GEVI application

1 Introduction

Calcium concentration, transmembrane voltage, and in particular their spatiotemporal fluctuations are ubiquitous carriers of biological information in many cell types of animals and plants. Calcium and voltage signalling give rise to fundamental questions and associated measurement across life sciences. Intracellular calcium (Ca^{2+}_i) is a common secondary messenger known to regulate a large number of physiological processes including synaptic transmission and plasticity, muscle contraction, cell death, and fertilization. Voltage transients in the form of action potential (AP) and synaptic potentials carry, integrate, and propagate information in neurons. In non-excitable cells (cells that do not generate APs), membrane voltage relates to cellular differentiation and transmembrane transport of ions and larger molecules. Calcium indicators are typically loaded into the cytoplasm to report the intracellular Ca^{2+} concentration ($[Ca^{2+}_i]$), whereas voltage indicators sense the voltage (potential difference) across the plasma membrane.

In earlier times, recordings through microelectrodes had been the method of choice for calcium (ion-selective electrodes) and voltage measurements (e.g. intra- and extracellular electrodes or patch-clamp pipettes for membrane voltage measurements), but in the twenty-first century, optical methods took over more and more terrain. This shift in approaches was driven by the fact that optical methods facilitate upscaling of cellular-level measurements to simultaneously recording from a large number of cells and provide spatial information (imaging). This methodological shift has been paralleled and fuelled by the development of better performing fluorescent calcium and voltage indicators and optical instrumentations.

Here we focus on fluorescent calcium and voltage indicators and their application in neuroscience and cardiac science. We briefly mention bioluminescent indicators but do not cover other emerging imaging modalities such as photoacoustic imaging. General principles of calcium and voltage indicators hold broadly for applications throughout life sciences.

2 Calcium Versus Voltage Imaging

Depending on the research question (calcium- or voltage-signalling related), one would naturally measure calcium or voltage signals. A special case of broader importance is the measurement of action potentials in neuronal networks or cardiac tissue. Although action potentials are in the domain of electrical signalling, optical imaging of AP activities using calcium imaging has become a fruitful and widely used alternative to electrophysiology, in particular in neurosciences. Calcium

imaging is less demanding in terms of instrumentations, and even with the most advanced equipment, the signal-to-noise ratio (SNR) of calcium imaging usually outperforms that of voltage imaging. Nevertheless, voltage imaging cannot be replaced by calcium imaging if the signals of interest include membrane hyperpolarization and high-frequency oscillations.

3 Advantages of Genetically Encoding Indicators

The first broadly used synthetic chemical fluorescent calcium indicators (e.g. fura-2) did not permeate lipid membranes, and therefore needed to be injected into cells for measurement of cytosolic Ca^{2+} concentration. These organic compounds were then made membrane permeant by adding an ester group (e.g. fura-2 AM) that can be cleaved by intracellular esterase, enabling bulk loading approaches for multicellular calcium imaging experiments. Voltage-sensitive dyes were first used in mammalian cells by staining the outer leaflet of the plasma membrane following incubation in extracellular dye-containing solutions, and intracellular loading techniques were subsequently developed for single-cell-level experiments (Wu et al. 1998; Baker et al. 2005). These indicators therefore allow staining all cells or single cells within a preparation. This approach is problematic when working with preparations where different cell classes exhibit different calcium or voltage signalling properties. A solution to this issue of blindness to cellular diversity is offered by genetically encoded indicators, expressed by cell class-specific expression cassettes (Knöpfel et al. 2006). Genetic encoding offers several additional advantages, including reproducible preparations and avoidance of potentially harmful staining procedures. During recent years, genetically encoded calcium indicators have largely replaced organic dyes in many fields of application. However, organic dyes still have advantages against protein-based indicators, including lower molecular weight (resulting in higher diffusibility), greater photostability, inertness, and ease of use. Low molecular weight (LMW) voltage indicators have a long history of development and use in neurophysiology.

The first generation of genetically encoded voltage indicators (GEVIs, named "voltage-sensitive fluorescent proteins") that produced robust signals in mammalian cells, enabling physiological studies beyond methodological proof of principle studies, was published only in 2007 (Dimitrov et al. 2007). Since then, better performing GEVIs were generated at a fast pace, and it is probably fair to say that they are now outperforming organic voltage-sensitive dyes in many applications from subcellular to the systems level. Genetic encoding provides the same advantages as described above for genetically encoded calcium indicators (GECIs).

4 Key Features of Genetically Encoded Calcium and Voltage Indicators

4.1 GECIs

4.1.1 Structure and Functional Principles of GECIs

The design of GECIs is based on the molecular fusion of a fast calcium-binding protein moiety (typically calmodulin; CaM), a calmodulin-binding peptide from smooth muscle myosin light-chain kinase (RS20, also known as M13), and either a single fluorescent protein (FP) or a pair of FPs suitable for Förster resonance energy transfer (FRET) (Fig. 1). Upon Ca^{2+} binding to the Ca^{2+}-binding moiety, GECIs undergo conformational changes that lead to a modulation of fluorescence emission of the FP(s). In the case of single-FP GECIs, the conformational change modulates the absorbance (and to some extent the K_d; see next section) of the FP. Since fluorescence at a single wavelength is modulated as a function of [Ca^{2+}], these indicators are known as intensiometric. In the case of FRET-based GECIs, the Ca^{2+}-dependent conformation change affects the distance and orientation of the pair of FPs, and this alters their FRET efficacy. Consequently, the fluorescence emission of one of the FPs decreases while that of the other increases. Accordingly, this indicator class is termed "ratiometric" because the ratio between the two fluorescence intensities can be used to quantify [Ca^{2+}]. Recently, the concept of

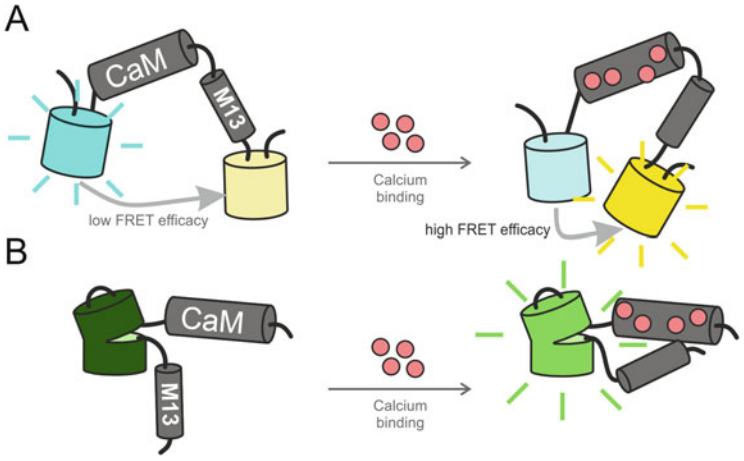


Fig. 1 Schematic depiction of genetically encoded calcium indicators. (**a**) Förster resonance energy transfer (FRET)-based calcium indicators use two fluorescent proteins (e.g. a cyan fluorescent protein, CFP, depicted in cyan colour, and a yellow fluorescent protein, YFP, depicted in yellow colour). Excitation of CFP results in cyan and (via FRET) yellow fluorescence. Binding of Ca^{2+} (red circles) to a calcium-binding domain (such as the calmodulin (CaM)– M13 complex shown here) increases the efficacy of FRET between CFP and YFP and therefore decreases cyan fluorescence and increases yellow fluorescence. (**b**) Single fluorescent protein indicators of the GCaMP type incorporate a circularly permuted fluorescent protein (cpFP, depicted in green hue). Binding of Ca^{2+} to CaM–M13 increases the cpFP fluorescence

FRET-based GECI has been expanded to bioluminescence-based indicators where a bioluminescent protein serves as the FRET donor.

Ca^{2+}-Binding Affinity of GECIs

The optical signal produced by GECIs is proportional to the fraction of indicator molecules occupied by Ca^{2+}; therefore the equilibrium constant of the calcium binding reaction is a crucial feature of each GECI. The ability of Ca^{2+} to bind to the calcium-binding moiety of the GECI is typically reported as the (apparent) equilibrium dissociation constant (K_d), which is defined by the balance between the rates of formation of the Ca^{2+}-bound fluorescent GECI conformation and the formation of the nonfluorescent Ca^{2+}-free conformation (Perez Koldenkova and Nagai 2013). At a $[Ca^{2+}]$ equal to K_d, half of the GECI molecules are Ca^{2+}-bound, and the sensitivity of the indicator is maximal. GECIs covering K_ds ranging from nM to mM are available. The choice of a particular K_d variant depends on the resting calcium concentration and the expected physiological fluctuations and hence on the subcellular locations of interest, cell types, and organisms. Indicators with large K_d (low affinity) are optimized for imaging large amplitude $[Ca^{2+}{}_i]$ transients or $[Ca^{2+}]$ in organelles with high resting $[Ca^{2+}]$. Conversely, small values of K_d represent a high affinity such that a large fraction of the indicator molecules bind Ca^{2+} during small transient changes in $[Ca^{2+}]$.

$$\text{GECI} + \text{Ca}^{2+} \underset{k_{\text{off}}}{\overset{k_{\text{on}}}{\rightleftharpoons}} \text{GECI-Ca}^{2+} \qquad k_d = k_{\text{off}}/k_{\text{on}}$$

k_{on} and k_{off}: rate constants of formation of bright and dim indicator state.

The Hill constant (h) describes the steepness of the indicator fluorescence – $[Ca^{2+}]$ relationship at a $[Ca^{2+}]$ equal to the indicator K_d. In the case of GECI, the Hill constant typically indicates the degree of cooperativity of multiple Ca^{2+} binding at sites of the calcium-binding protein. While high h increases the sensitivity to changes in $[Ca^{2+}]$ ranging around K_d, this advantage comes with lower sensitivity outside this range.

$$\Theta = \left[\text{Ca}^{2+}\right]^h / \left(\left[\text{Ca}^{2+}\right] + K_d\right)$$

h, Hill constant; Θ, fraction of the receptor protein concentration that is bound by the ligand.

Kinetic Properties

Ca^{2+} binding to organic calcium indicators is typically diffusion limited, and the unbinding (k_{off} (s^{-1}), see above) determines K_d. In the case of GECIs, kinetics are often more complex because the rate-limiting factor for the fluorescent conformations is not necessarily calcium binding but instead, for instance, the interaction of CaM and M13. Early-generation GECIs exhibited kinetics significantly slower than expected from their K_d, probably due to these more complex conformational changes. CaM-based GECIs display much faster Ca^{2+}-binding and

Ca^{2+}-unbinding properties than troponin-based indicators. Specific mutations of CaM can facilitate Ca^{2+}-binding while reducing h.

Selectivity and Interference

Most naturally occurring Ca^{2+}-binding proteins also bind magnesium ions (Mg^{2+}). Therefore, Ca^{2+}/Mg^{2+} selectivity of the Ca^{2+}-binding motifs of GECIs, derived from CaM or troponin C, may be of concern. In practice, however, Mg^{2+} sensitivity is not problematic since physiological [Mg^{2+}] varies very little in contrast to the very large dynamic range of physiological [Ca^{2+}] fluctuations (10^{-9}–10^{-3} M). GECIs may interact with other proteins as indicated by the discrepancy of K_d values obtained in cell-free conditions as compared to those measured in cellular cytosolic environment.

Calcium indicators act as Ca^{2+} buffers and hence interfere with Ca^{2+} dynamics. Thus, calcium indicators provide a distorted view of normal calcium dynamics (Neher 1995).

4.2 GEVIs

4.2.1 Structure and Functional Principles of GEVIs

GEVIs report membrane voltage by sensing the electrical field across the plasma membrane. To do so, they need to be localized within or in very close proximity to the plasma membrane. The first series of GEVIs that exhibited robust voltage reports in mammalian cells were constructed by the fusion of GFP derivatives and an isolated voltage-sensing domain (Dimitrov et al. 2007). These early GEVIs (dubbed Voltage-Sensitive Fluorescent Proteins; VSFPs) parented a large number of voltage-sensing domain-based GEVIs. In these GEVIs, coupling of voltage sensing and optical output is achieved either via FRET between a pair of FPs, sensitizing a single FP by circular permutation (cpFPs), or via mechanisms that await full explanation (Fig. 2). A second pedigree of GEVIs is rooted in the discovery of voltage-dependent fluorescence of some microbial opsins. The optical output of these opsin-based GEVIs is either the fluorescence of the opsin itself or the fluorescence of an attached FP that is quenched by voltage-dependent absorbance of the opsin. Most of the currently available GEVIs are known by their acronyms, but their structural design principles are rooted on these two distinct design scaffolds. Chemigenetic or hybrid GEVIs employ genetically targetable proteins which bind compounds that are not proteins or naturally occurring in the brain tissue.

Kinetic Properties

First-generation voltage-sensing domain (VSD)-based GEVIs responded to a quasi-instantaneous change of membrane voltage too slow to faithfully report fast action potentials. They were, however, successfully used to report synaptic population potentials in vitro and in vivo (Akemann et al. 2010). Recent versions of VSD-based GEVIs respond to voltage changes with millisecond kinetics and are

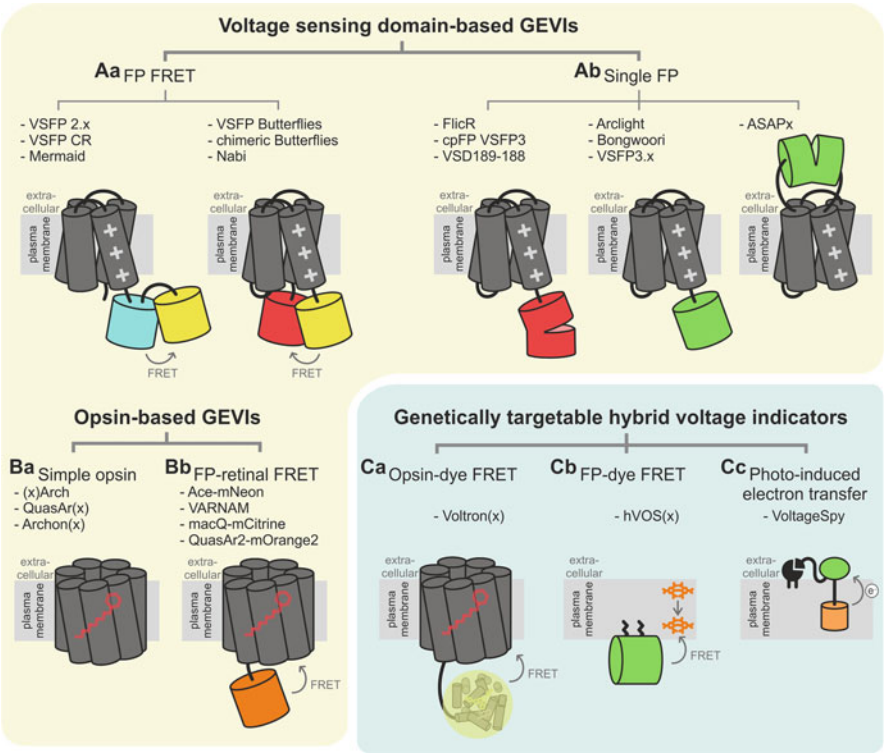


Fig. 2 Schematic depiction of genetically encoded voltage indicators. GEVIs can be classified based on their structural and functional design into three main groups: (**A**) Voltage-sensing domain (VSD)-based GEVIs, (**B**) opsin-based GEVIs, and (**C**) genetically targetable hybrid voltage indicators. Implementations of each of these design principles often come with different names of particular variants. (**Aa**) Fluorescent protein (FP) FRET-based GEVIs are engineered around a four transmembrane segment (S1–S4) voltage-sensing domain (dark grey structure) that spans in the plasma membrane S4 carries positive charges that sense changes of the transmembrane electric field. The efficacy of FRET increases upon plasma membrane depolarization as the voltage sensor adopts its activated state. The FPs can be either in tandem configuration (e.g. VSFP2.x, Mermaid) or flanking the VSD (e.g. VSFP Butterflies). (**Ab**) Single-FP GEVIs exhibit fluorescence voltage-dependent fluorescence quenching. The FP can be in its native configuration (VSFP3x, ArcLight, Bongwoori) or in the form of a circularly permuted FP (cpFP indicated by opening of FP structure), e.g. in VSFP3 and FlicR. Both fusion to the C-terminus of the VSD (former examples) and insertion into the extracellular loop between transmembrane segment S3 and S4 (ASAPx) have been successfully employed to develop GEVIs. (**Ba**) Simple opsin-based GEVIs exploit a voltage-dependent fluorescent state of their retinal chromophore. Prominent examples for this class are xArchs, QuasArs, and Archons. (**Bb**) To address the dimness of simple opsin-based GEVIs, they have been combined with a bright FP that is quenched by the opsin in voltage-dependent fashion (e.g. Ace-mNeon, VARNAM, macQ-mCitrine, QuasarAr2-mOrange2). (**Ca**) Opsin-dye FRET GEVIs use an organic dye captures by a high affinity binding engineered into the C-terminus of a voltage-dependent opsin (e.g. Voltrons). (**Cb**) FP-dye FRET GEVIs such as hVOS use a FP in combination with an organic quencher that distributes in the membrane in a voltage-dependent fashion, leading to voltage-dependent quenching of the FP. (**Cc**) Photoinduced electron transfer GEVIs exploit membrane voltage-dependent intramolecular quenching (e.g. VoltageSpy)

efficiently activated and deactivated during fast APs (St-Pierre et al. 2014; Gong et al. 2015). Notably, fast activation is required for APs to induce a large fluorescence signal, but fast deactivation kinetics imply that fluorescence needs to be measured at a fast sampling rate (kHz or above if the AP shape is of interest). Some opsin-based GEVIs use a mechanism with fast (<1 ms) kinetics (Kralj et al. 2011b).

Selectivity and Interference

GEVIs are selective for voltage signals but may also report pH changes in the case of those GEVIs that involve pH-sensitive FPs. Voltage sensing by movable charges in the membrane electric field corresponds to a capacitive current that interferes with voltage dynamics. To control for this effect, basic characterization of a new GEVI typically includes comparative measurements of AP width in GEVI-expressing and non-expressing neurons. Expression levels used in physiological experimentation are normally too slow to significantly affect AP widths, while the effect on dendritic integration of subthreshold potentials remains to be investigated in more detail (Akemann et al. 2009).

Because of the small changes in fluorescence, GEVI imaging in vivo has to carefully consider confounding optical signals triggered by changes in blood volume and oxygenation (Akemann et al. 2012).

4.3 Features Common to GEVIs and GECIs

4.3.1 Dynamic Range and Signal-to-Noise Ratio

The dynamic range is often considered as one of the important criteria when judging the performance of genetically encoded indicators. This dimensionless measurement states how many times brighter the Ca^{2+}-bound GECI state (or depolarized GEVI state) is when compared to the Ca^{2+}-free (hyperpolarized) state or how many times the maximal FRET ratio is larger than the minimal observed FRET ratio.

$$D_{\text{non-ratiometric}} = I_{\max}/I_{\min}$$

D, dynamic range; $I_{\max}$, maximal fluorescence intensity; $I_{\min}$, minimal fluorescence intensity.

However, dynamic range is usually measured under non-physiological conditions (including in the absence of Ca^{2+} and in the presence of very high Ca^{2+} concentrations for GECIs or exceedingly large membrane voltage changes for GEVIs), and often not in the cell type the indicator's use is envisaged. Therefore, more important than the biophysically defined dynamic range is the size of the optical signal in response to a physiological event of interest (e.g. synaptic activation of a receptor or an AP) in relation to the measured random photon fluctuations (noise) under concrete conditions of experimental applications. The SNR of an indicator is determined as the ratio between the fluorescence signal ($\Delta F = F_{\text{signal}} - F_{\text{baseline}}$) and the noise of the fluorescence (e.g. variance of total

fluorescence, F_{total}). Using high-end equipment, measured noise is typically shot (Poisson) noise-limited, which means that SNR increases with the square root of the measured fluorescence intensity.

$$\text{SNR} \sim \Delta F \times \sqrt{F_{\text{total}}}$$

SNR = signal to noise ratio; ΔF = fluorescence signal change; F_{total} = total fluorescence.

4.3.2 Brightness

Fluorescent indicators need to be bright. Brightness is determined by a combination of high fluorescence quantum yield and high photostability. Indicators of insufficient brightness require high illumination intensities where indicator photobleaching becomes a limiting factor. Indicator brightness is a critical feature in applications that demand for imaging at high frame rates, small ΔF values, high spatial resolution, and long-duration recordings. GECIs such as GCaMPs of generation 6 and above are sufficiently bright so that indicator bleaching is of no concern in most applications. For GEVIs that involve FPs, variants with the brightest and most photostable FPs (VSFP CR, Ace-mNeonGreen) are available. Opsin-based GEVIs are very photostable such that they may show little bleaching even with extended periods of very high illumination intensities.

4.3.3 Spectral Properties ("Colour")

The first ratiometric FRET-based and intensiometric GECIs were limited to the blue to green spectral range. More recently, a palette of GCaMP-type GECIs covering blue to near infrared becomes available. Similar expansion of the colour range also exists for GEVIs. In addition to the wide spectral range of available FPs, synthetic dyes of different colours have begun to be utilized for activity indicators. Such diverse spectral properties are especially attractive for multicolour imaging, where the activity from multiple distinct cell types can be simultaneously optically accessed via expression of activity indicators of different colours.

4.3.4 pH and Temperature

Modern FP variants used in genetically encoded indicators have been engineered towards a low p*K*a, making them relatively inert to physiological changes in pH. However, several intracellular organelles are normally acidic, and the cytosol of many cell types acidifies under cell stress conditions. In these situations, changes in pH may produce confounding optical GECI and GEVI signals. FPs are also temperature-dependent, and this needs to be considered when temperature changes are expected during an experimental paradigm. Specifically designed GECIs are available for monitoring Ca^{2+} dynamics within intracellular organelles, which may have a lower pH and/or a higher resting [Ca^{2+}] than the cytoplasm.

5 Currently Available Indicators

Description of the first protein-based indicators dates back to the middle of last century, where the luminescent protein aequorin from *Aequorea victoria* (a species of jellyfish) was purified (Shimomura et al. 1962) and then injected into barnacle muscle fibres (Ashley and Ridgway 1968). Three decades later, cloning of the genes encoding aequorin and the green fluorescent protein (GFP), and their molecular fusion to coding sequences of calcium-binding or voltage-sensing protein domains, led to the generation of the first genetically encoded calcium (Miyawaki et al. 1997; Persechini et al. 1997) and voltage-dependent indicators (Baker et al. 2008). GECIs evolved to a stage where their performance are probably close to theoretical limits with intensiometric GCaMP-type and X-CaMP variants covering the spectral range from blue to near infrared. GEVIs are still rapidly evolving with largely improved variants emerging at high pace. In the following Sects. 5.1 and 5.2, we give an overview of currently most used GECIs and GEVIs and briefly summarize the status of the emerging field of bioluminescent indicators.

5.1 Genetically Encoded Calcium Indicators

5.1.1 FRET-Based GECIs (Ratiometric)

The first FP-based calcium indicators used FRET between two FPs as the optical reporting mechanism. FRET relies on the transfer of excitation energy from a donor to a very closely localized (<10 nm distance) acceptor fluorophore through long-range dipole-dipole interactions (Miyawaki et al. 1997; Persechini et al. 1997). For FRET to occur, there has to be an overlap between FRET donor fluorescence emission spectrum and FRET acceptor absorbance spectrum. FRET can be used to report conformational changes of fusion proteins involving a pair of FPs, as FRET efficacy is very sensitive to their orientation and distance. In GECIs this mechanism is used to report the conformational changes triggered by Ca^{2+}-binding to a calcium-binding protein. These indicators are excited at a wavelength within the donor excitation spectrum (slightly blue shifted relative to the peak to minimize direct excitation of the acceptor), and fluorescence is split and measured within both the donor and acceptor emission wavelength bands. Ca^{2+}-dependent changes in FRET efficacy are quantified as a change in the ratio between acceptor and donor fluorescence.

Although imaging FRET-based GECIs have advantages over the use of intensiometric GECIs, such as a reduced interference by movement of the preparation, in practice they have been almost completely superseded by intensiometric indicators which have a higher dynamic range, require less demanding instrumentation (single channel versus dual channel), and are easier to be combined for multicolour imaging (simultaneous recording from different targets using different colour indicators).

Recently, a new generation of red-shifted FRET-based GECIs has been created (Waldeck-Weiermair et al. 2015), which possess green and orange or red FPs (GFP, OFP, RFP) as a FRET pair.

5.1.2 Single-FP (Intensiometric) GECIs

After several years of intense efforts by a number of groups, intensiometric GECIs of the GCaMP scaffold became the widely preferred fluorescent indicators for Ca^{2+} imaging. These probes are based on cpFP flanked by CaM at its C-terminus and the M13 domain of a myosin light chain kinase at its N-terminus. Upon Ca^{2+} binding, an intramolecular conformational change is triggered, and thereby the chromophore environment alters. This, in turn, results in a large modulation of the chromophore fluorescence output (Baird et al. 1999; Nagai et al. 2001).

The scaffold of GCaMP2 served as template to incorporate "superfolder GFP" mutations (Pedelacq et al. 2006) to generate highly sensitive GCaMPs with improved temporal and spatial resolution in vivo (Muto et al. 2011; Badura et al. 2014). Targeted mutagenesis of the linker sequences connecting M13/CaM to the cpGFP was optimized to yield the highly sensitive series of GCaMP5 (5A, 5D, 5G, 5K, and 5L) (Akerboom et al. 2013) with improved temporal and spatial resolution. Neuronal expression screening in in vivo imaging in zebrafish, flies, and mice heralded the improved brightness, dynamic range, and slow to fast kinetic properties of the ultrasensitive members of the GCaMP6 and jGCaMP7 series (Chen et al. 2013; Dana et al. 2019). With introduction of GCaMP6 and jGCaMP7, optimization of GCaMP-type GECIs for monitoring neuronal action potentials seemed, at least for the green fluorescent variety, to reach a stage where further improvements are difficult to achieve without trade-offs. A more recent suite of GECIs dubbed X-CaMPs demonstrated that tuning to specific requirements is still possible (Inoue et al. 2019). Green X-CaMPs (X-CaMP-G) are faster and more linear in response than its predecessors, features that are particularly important for estimating spike rates of fast spiking interneurons (Inoue et al. 2019). In addition, the increased baseline fluorescence of X-CaMPs facilitates the use of calcium imaging to monitor subthreshold events in spines and synaptic terminals (Inoue et al. 2019). This feature may, however, be of disadvantage in densely GECI-targeted tissue. New colour variants (blue, yellow, and red) in the X-CaMP series facilitate multicolour imaging with either 1-photon or 2-photon fluorescence excitation. Specifically tuned low-affinity intensiometric GECIs for targeting the endoplasmic reticulum R-CEPIA1er (with a K_d of 565 μM) and GCaMPer (10.19) (with a K_d of 400 μM) are tuned for monitoring very high [Ca^{2+}] such as in ER and complete the current toolbox of best performing intensiometric GECIs (Suzuki et al. 2016).

5.2 Genetically Encoded Voltage Indicators

5.2.1 Voltage-Sensing Domain-Based GEVIs

Over the last two decades, a palette of VSD-based GEVIs have been developed using fluorescent proteins with emission wavelengths covering a large portion of the visible spectrum. Irrespective of the origin organism of the VSD (either from *C. intestinalis* or *G. gallus*), VSD-based GEVIs can be categorized into two families – ratiometric FP-FP FRET GEVIs with emission at two spectral bands or single-FP GEVIs with emission in a single spectral band. The generation of FP-based activity

indicators is also largely influenced by the ongoing development of FPs; more recently, VSD-based GEVIs have also been further developed to expand the emission wavelength from the visible spectrum into the near-infrared range.

FP-FP FRET GEVIs

Ratiometric GEVIs function via the voltage-dependent VSD conformation change that alters the FRET energy transfer between a FP pair. Changes of membrane voltage hence result in the negatively correlated changes of fluorescence intensity of the FRET donor and acceptor FPs. The FP pair can either be attached to the VSD S4 domain in tandem (e.g. as in VSFP 2.3, VSFP CR, Mermaid, see Fig. 2 (Tsutsui et al. 2008; Akemann et al. 2010; Lam et al. 2012)), or the two FPs can flank the VSD S1-S4 transmembrane segments (e.g. as in VSFP Butterflies (Akemann et al. 2012) or Nabi (Sung et al. 2015)). Generation of VSD chimaeras via transplantation of segments from fast operating ion channels (e.g. Kv3.1 potassium channel) has been used to improve the kinetic properties of VSD-based GEVIs (e.g. chimeric VSFP Butterflies (Mishina et al. 2012)).

FRET-based GEVIs can be used in combination with dual-emission imaging setups that captures the donor and acceptor emission wavelength ranges simultaneously. This ratiometric imaging approach is particularly useful for in vivo voltage imaging where the negatively correlated changes of donor and acceptor emission intensity provide information on voltage activity, while movement and hemodynamic signals are identified by correlated changes in donor and acceptor fluorescence. For experiments in brain slices, it is sufficient (and possibly more efficient) to image either the acceptor or donor of FRET-based GEVIs.

Monochromatic FP GEVIs

Single FPs can be fused with a VSD to produce monochromatic GEVIs. Monochromatic GEVIs are of advantage where simplified optical configurations and a simpler optical imaging setup are needed. In contrast to ratiometric GEVIs, where sampling of photons is restricted to wavelength bands that minimize spectral overlap between the two FPs, monochromatic GEVIs allows one to capture across the (almost) full FP emission spectrum to maximize the sampling of useful photons. However, when using monochromatic GEVIs in vivo, especially in the mammalian brain, correction for haemodynamic and movement-related signals is more cumbersome. Strategies for such corrections often involve multiplexing GEVI imaging with a reference signal.

FPs with different colours can be used in monochromatic GEVIs of the VSFP3x type where they are attached to the fourth transmembrane domain of a VSD. The FP can be a cpFP (e.g. cpFP-VSFP3x, FlicR (Gautam et al. 2009; Abdelfattah et al. 2016)) or a FP in its native configuration (e.g. Arclight, Bongwoori, VSFP3x (Lundby et al. 2008; Perron et al. 2009; Jin et al. 2012; Lee et al. 2017)). cpFP has also been inserted extracellularly between the S3 and S4 transmembrane segment of *G. gallus* VSD to generate the ASAPx series with high sensitivity and fast kinetics (St-Pierre et al. 2014).

5.2.2 Opsin-Based GEVIs

More recently, voltage-dependent fluorescence of certain microbial opsins has been exploited to generate opsin-based indicators (Kralj et al. 2011a). Opsin-based GEVIs have been tuned to display high-voltage sensitivity and rapid kinetic properties, yet initial versions of these indicators suffered from low brightness and required high excitation intensity and highly specific optical imaging setups.

Opsin GEVIs Using Rhodopsin Fluorescence

Opsin-based GEVIs report voltage fluctuations via voltage-dependent protonation of the Schiff base of the opsin retinal component that leads to chromophore absorption. This led to the first opsin-based GEVI Arch with submillisecond kinetic properties, yet Arch is very dim and requires high excitation intensity and mediates an undesired photocurrent (Kralj et al. 2011a). To overcome these issues, Arch-based GEVIs have been further optimized to generate the first generation QuasAr variants (QuasAr1 and QuasAr2), with increased brightness and without the undesirable accompanying photocurrents (Hochbaum et al. 2014). However, first-generation QuasAr still suffered from suboptimal membrane localization and dimness. To further optimize these parameters, second-generation QuasArs (QuasAr3 and paQuasAr) were developed (Adam et al. 2019). QuasAr3 displayed improved membrane targeting but still required high illumination intensity. This need for high intensity is partially alleviated by photoactivated QuasAr (paQuasAr), where co-illumination of light at two different wavelengths helps to spatially confine excitation and thereby improve higher SNR (Adam et al. 2019).

Molecular evolution based on Arch also produced the Archon family of indicators, which display several-fold higher brightness than the parent protein (Piatkevich et al. 2018). Like other opsin-based indicators, Archon1 displays fast kinetic properties and large sensitivity.

Opsin absorbs at wavelengths >600 nm and displays fluorescence emission in the near-infrared wavelength range. These spectral properties facilitate all-optical electrophysiology, where in combination with blue-shifted optogenetic actuators, light is used to both control and monitor membrane voltage (Hochbaum et al. 2014).

Opsin GEVIs Using FP-Retinal FRET

Opsin-based FP-retinal FRET GEVIs have been designed with the idea to retain both the rapid kinetic properties of opsin-based GEVIs and the brightness of FPs. Their design exploits the spectral overlap between FP emission and opsin absorption spectrum, where the opsin can act as a FRET acceptor. Combination of the yellow FP mCitrine, the bright FP mNeonGreen, or the orange FP mRuby3 with rhodopsin cloned from *L. maculans* or *A. acetabulum* led to Ace-mNeonGreen and VARNAM, (Gong et al. 2014, 2015; Kannan et al. 2018). Since the rhodopsin variants used are only very weakly fluorescent, optical readout is the fluorescence quenching of the donor FP. These fast FRET opsin-based GEVIs operate with considerably lower illumination intensities as compared to opsin only-based GEVIs.

5.2.3 Chemigenetic Hybrid Voltage Indicators

Aside from the fully genetically encoded voltage indicators, voltage indicators with a genetically encoded component have also been developed. The genetically encoded component of these hybrid indicators (also known as chemigenetic indicators) allows for targeting indicator localization to specific cell types. The second component may be a fluorescent dye that needs to be delivered to the tissue (this may be achieved via a simple i.v. injection). The advantage of fluorescent dyes is that they can be much brighter than FPs. These indicators offer increasing potential for applications in vivo.

Opsin-Dye FRET Chemigenetic Indicators

Like FP-retinal FRET opsin-based indicators in section "Opsin GEVIs Using FP-Retinal FRET", opsin-dye FRET chemigenetic indicators utilize the voltage sensitivity and submillisecond kinetic properties of microbial opsin and combine this with a bright fluorescent counterpart. Cell type-specific expression of hybrid indicators of this class is hence achieved via genetic encoding of the voltage sensing component (i.e. the opsin). One of the most promising new families of GEVIs of this class are Voltron and Positron, hybrid indicators where the opsin domain from *A. acetabulum* is fused to a self-labelling protein tag (e.g. HaloTag or SNAP-tag) that allows the covalent binding of a synthetic fluorophore dye ligand (e.g. Janelia Fluor dyes). Voltage-dependent changes of the opsin absorption spectrum then modulate fluorescence quenching of the JF dye component via FRET. JF dye covers a large portion of the visible and near-infrared wavelength spectra. Along with distinct self-labelling proteins, Voltrons and Positrons are prepared for multicolour imaging from different cell classes.

FP-Dye FRET Hybrid Indicators

Hybrid voltage indicators can also be targeted to specific cell populations by genetic targeting of its fluorescent component. In hybrid indicators of the hVOS type, genetic targeting is achieved via expression of a membrane-anchored FP (Wang et al. 2010). The component to be applied externally is dipicrylamine (DPA) which quenches the FP with an efficacy that depends on its voltage-dependent distribution in the plasma membrane. Like with GEVI FRET opsin-based GEVIs, a voltage-dependent reversible quenching of a FP informs on membrane voltage transients.

Indicators Using Photoinduced Electron Transfer

Departing further from protein-based voltage indicators, voltage-sensitive fluorescent dyes can also be modified to be genetically targeted to specific neuronal populations. One of the recent promising examples is VoltageSpy, a series of fluorescein-based voltage-sensitive fluorescent dyes conjugated to a peptide (SpyTag) that can covalently attach to peptide-based cell surface receptors (SpyCatcher) (Grenier et al. 2019). Voltage sensing in these indicators is achieved by photoinduced electron transfer via a synthetic molecular wire in the membrane to reversibly quench a synthetic fluorescent reporter. Genetic encoding of such indicators would be conferred by the cell type-specific expression of SpyCatcher.

5.3 Bioluminescent Indicators

The first protein-based optical indicator used for live science applications was the bioluminescent protein aequorin. Aequorin is a calcium-activated photoprotein which, when in its calcium-bound state, converts its prosthetic group, coelenterazine, into an excited state that in turn emits blue light when returning to its ground state. Notably, aequorin occurs in jellyfish together with the green fluorescent protein to produce green light by resonant energy transfer.

The development of well-performing bioluminescent indicators has not progressed as fast as fluorescent indicators over the past decades. More recent hybrids between bioluminescent proteins and FPs led to improved bioluminescent indicators both for calcium (Suzuki et al. 2018) and voltage (Inagaki et al. 2019). These sensors exploit the catalysis of a chemical substrate by a luciferase enzyme and produce light via FRET to FP.

Recent versions, like the Orange CaMBI (orange calcium-modulated bioluminescent indicator), have been reported to track calcium dynamics in single cells (HeLa cells and cardiomyocytes derived from human-induced pluripotent stem cells) and in whole organs of a transgenic mouse, in a noninvasive manner. Different variants from Orange CaMBI have been engineered with a broader range of affinities (Oh et al. 2013, 2019).

Likewise, bioluminescent proteins have been combined with voltage indicators. Bioluminescent voltage indicators consist of a VSD, a luciferase, and a fluorescent protein in a configuration analogous to FRET-based GEVIs. The increase of membrane voltage causes a structural change of the VSD, enhancing FRET between the light-emitting luciferase and the fluorescent protein. The ratiometric LOTUS-V indicator is a promising example of bioluminescent voltage indicators (Inagaki et al. 2019) that enables voltage imaging from neurons and in freely moving mice using a fibre-free system (Inagaki et al. 2019). Advantage of bioluminescent genetically encoded indicators is low invasiveness, no need for excitation light to fuel fluorescence, and simple instrumentation for detection. As mammalian tissues lack endogenous bioluminescence, even relatively small number of photons can be detected over background with bioluminescence. However, limited photon flux translates into a limited SNR. Their major disadvantage is the requirement of substrate application.

6 Applications

Although optimization to obtain new better performing genetically encoded indicators is still ongoing, GECIs are being broadly used in neurophysiological and cardiac studies since several years. The use of GEVIs for physiological experimentation beyond technical feasibility reports is only emerging. The following example applications have been selected with a bias on studies in the mouse cerebral cortex and heart.

6.1 Monitoring Action Potentials in Awake Mouse Cortex

One major challenge in neuroscience is to unravel how the information is encoded, stored, and processed within the brain circuits. The traditional physiological approach to study representation of neuronal information is by correlating APs occurrence with sensory input or motor output. To understand processing of information in neuronal circuits, it is necessary to observe AP occurrence in a large number of cells simultaneously. Ca^{2+} imaging is a powerful optical approach for multineuronal AP recordings due to that fact that APs in mammalian cortical neurons are generally accompanied by substantial influx of Ca^{2+}. GECIs combined with 2-photon microscopy have revolutionized the study of neural activity in model organisms, since they allow non- or minimally invasive recordings and can be used with better spatial resolution than electrophysiological techniques in awake, behaving animals (Chen et al. 2013). Since emergence of GCaMP2, GECIs based on this general scaffold have been routinely used for in vivo imaging, for tracking presynaptic calcium transients, individual spikes, and large populations of neurons (Diez-Garcia et al. 2005; Chen et al. 2013). In vivo AP recordings using GCaMPs have not been limited to 2-photon approaches. For instance, the fast GCaMP6f variant has been used to study the neuronal networks of behaving mice with a high temporal resolution by single-photon wide-field imaging using miniature microscopes (Aharoni et al. 2019). GCaMP6s have been highlighted for being a slow-kinetics indicator, able to yield high detection rates for single spikes and resolving subthreshold Ca^{2+} transients (Chen et al. 2013).

The promising multicolour suite of XCaMPs – especially the green and yellow variants – possesses faster kinetics. These GECIs enable detection of single AP within 3–10 ms of spike onset, thereby having the potential to resolve spike trains from fast-spiking parvalbumin interneurons in the cortex of freely moving mice (Inoue et al. 2019).

Although calcium imaging is currently the most preferred optical method of systems level assessment of patterns of APs, due to the superior SNR values compared to those of GEVIs, well-performing voltage indicators that directly report membrane potential transients are still on the list of "the most wanted". Considerable progress in the field of GEVI development led to the first voltage indicators that allow monitoring APs of a few cells simultaneously in the awake mouse hippocampus (Adam et al. 2019).

6.2 Mesoscopic Imaging of Neuronal Circuit Dynamics

Optical imaging of cortical activity patterns across milimetre-size cortical areas has traditionally been the realm of voltage imaging. The first in vivo experiments using GEVIs build on these traditional mesoscopic voltage imaging approaches (Akemann et al. 2010, 2012; Carandini et al. 2015). While not providing cellular resolution data, mesoscopic GEVI imaging is a powerful approach to map cortical

representations (such as retinotopy), establish long-range interactions as well as higher level functional organization principles.

6.3 Monitoring Astrocytic Activity

Neuronal networks are constantly interacting and being modulated by other cell populations, particularly astrocytes, which represent the most abundant glial cells in the brain. Astrocytes sense neuronal activity and respond with intracellular Ca^{2+} elevations that in turn evoke the release of gliotransmitters (Losi et al. 2019). Expression of GECIs can be targeted exclusively to astrocytes by using the glial fibrillary acidic protein (GFAP) promoter. Expression of GCaMP6s in astrocytes of the somatosensory cortex has enabled the demonstration of robust astrocytic Ca^{2+} transients associated with local neuronal activity evoked by sensory stimulation (Sonoda et al. 2018).

6.4 Combination of GECIs and GEVIs with Optogenetic Interference

Optogenetic interference exploits the use of light-gated ion channels to activate (using inward current-generating opsins such as channelrhodopsin 2, ChR2) or silence (using outward current-generating opsins such as ACR) the activity of neurons. In contrast to electrode-based methods, optogenetic interference can be targeted to specific genetically defined cell populations. Optogenetic interference in combination with GECI- and GEVI-based monitoring of cellular activity enables all-optical electrophysiology (AOE). Initial efforts towards AOE have been hampered by difficulties to spectrally separate indicators and actuators. For instance, green GCaMPs cannot be used in cells that express ChR2, halorhodopsin (HR), or archaerhodopsin (Arch), since the GECIs excitation light spectrum overlaps with the activation spectra of light-gated channels. In this combination, illumination light for optical recordings would continuously activate the light-gated channel unless expression of actuators and indicators and light delivery are spatially well separated. The development of red fluorescent Ca^{2+} indicators based on mRuby and mApple only partially resolved this issue (Dana et al. 2016). However, far red-shifted genetically encoded indicators have been successfully combined with blue-shifted optogenetic activators to image and modulate specific cells and circuits simultaneously, as the wavelength of the light used for monitoring is sufficiently longer than the wavelengths in the activation band of blue-shifted ChR2 variants.

An example of a red GECI that has succeeded in being used in combination with optogenetics is the jRGECO1a variant (Dana et al. 2019). GEVIs suitable for AOE are QuasAr, Archon, and NIR butterflies.

6.5 Calcium and Voltage Imaging to Monitor Cardiac Cells

Development of GECIs and GEVIs has been driven by interests to use them as tools in neuroscience research. Cardiac scientists soon became interested in these tools as well. Indeed, transgenic mice and fish have been engineered to specifically express first-generation GECIs and GEVIs in cardiomyocytes to sense cardiac electrical activity and Ca^{2+} signalling (Chi et al. 2008; Chang Liao et al. 2015; van Opbergen et al. 2018). GCaMPs were used to demonstrate that human embryonic stem cell-derived cardiomyocytes can electrically couple with host myocytes upon transplantation to infarcted hearts (Shiba et al. 2012). Experiments using a GEVI expressed either in cardiomyocytes or myofibroblast provided direct electrophysiological evidence of hetero-cellular electrotonic coupling in native myocardium (Quinn et al. 2016).

7 Outlook

Development of genetically encoded indicators progressed in parallel to considerable advances in instrumentations for functional optical recordings. These include two-photon microscopy, miniature microscopes, and fibre photometry, which allow monitoring of calcium and voltage dynamics in cells, tissue sections, and behaving subjects. At the time of writing this chapter (June 2019), cutting-edge AP detection using calcium imaging in living mice benchmarks at about 10,000 cells with about 20 Hz temporal resolution but only a handful of neurons but at 1 kHz using GEVI imaging. Considerable further progress is expected for GEVI imaging along with specialized instrumentation.

References

Abdelfattah AS, Farhi SL, Zhao Y et al (2016) A bright and fast red fluorescent protein voltage indicator that reports neuronal activity in organotypic brain slices. J Neurosci 36:2458–2472. https://doi.org/10.1523/JNEUROSCI.3484-15.2016

Adam Y, Kim JJ, Lou S et al (2019) Voltage imaging and optogenetics reveal behaviour-dependent changes in hippocampal dynamics. Nature 569:413–417. https://doi.org/10.1038/s41586-019-1166-7

Aharoni D, Khakh BS, Silva AJ, Golshani P (2019) All the light that we can see: a new era in miniaturized microscopy. Nat Methods 16:11–13. https://doi.org/10.1038/s41592-018-0266-x

Akemann W, Lundby A, Mutoh H, Knöpfel T (2009) Effect of voltage sensitive fluorescent proteins on neuronal excitability. Biophys J 96:3959–3976

Akemann W, Mutoh H, Perron A, Rossier J, Knöpfel T (2010) Imaging brain electric signals with genetically targeted voltage-sensitive fluorescent proteins. Nat Methods 7:643–649

Akemann W, Mutoh H, Perron A, Park YK, Iwamoto Y, Knöpfel T (2012) Imaging neural circuit dynamics with a voltage-sensitive fluorescent protein. J Neurophysiol 108:2323–2337

Akerboom J, Carreras Calderon N, Tian L et al (2013) Genetically encoded calcium indicators for multi-color neural activity imaging and combination with optogenetics. Front Mol Neurosci 6:2. https://doi.org/10.3389/fnmol.2013.00002

Ashley CC, Ridgway EB (1968) Simultaneous recording of membrane potential, calcium transient and tension in single muscle fibers. Nature 219:1168–1169. https://doi.org/10.1038/2191168a0

Badura A, Sun XR, Giovannucci A, Lynch LA, Wang SSH (2014) Fast calcium sensor proteins for monitoring neural activity. Neurophotonics 1:025008

Baird GS, Zacharias DA, Tsien RY (1999) Circular permutation and receptor insertion within green fluorescent proteins. Proc Natl Acad Sci U S A 96:11241–11246. https://doi.org/10.1073/pnas.96.20.11241

Baker BJ, Kosmidis EK, Vucinic D, Falk CX, Cohen LB, Djurisic M, Zecevic D (2005) Imaging brain activity with voltage- and calcium-sensitive dyes. Cell Mol Neurobiol 25:245–282

Baker BJ, Mutoh H, Dimitrov D et al (2008) Genetically encoded fluorescent sensors of membrane potential. Brain Cell Biol 36:53–67

Carandini M, Shimaoka D, Rossi LF, Sato TK, Benucci A, Knöpfel T (2015) Imaging the awake visual cortex with a genetically encoded voltage indicator. J Neurosci 35:53–63. https://doi.org/10.1523/JNEUROSCI.0594-14.2015

Chang Liao ML, de Boer TP, Mutoh H et al (2015) Sensing cardiac electrical activity with a cardiac myocyte--targeted optogenetic voltage indicator. Circ Res 117:401–412. https://doi.org/10.1161/CIRCRESAHA.117.306143

Chen T-W, Wardill TJ, Sun Y et al (2013) Ultrasensitive fluorescent proteins for imaging neuronal activity. Nature 499:295–300. https://doi.org/10.1038/nature12354

Chi NC, Shaw RM, Jungblut B et al (2008) Genetic and physiologic dissection of the vertebrate cardiac conduction system. PLoS Biol 6:e109. https://doi.org/10.1371/journal.pbio.0060109

Dana H, Mohar B, Sun Y et al (2016) Sensitive red protein calcium indicators for imaging neural activity. Elife 5:e12727. https://doi.org/10.7554/eLife.12727

Dana H, Sun Y, Mohar B et al (2019) High-performance calcium sensors for imaging activity in neuronal populations and microcompartments. Nat Methods 16:649–657. https://doi.org/10.1038/s41592-019-0435-6

Diez-Garcia J, Matsushita S, Mutoh H et al (2005) Activation of cerebellar parallel fibers monitored in transgenic mice expressing a fluorescent Ca2+ indicator protein. Eur J Neurosci 22:627–635

Dimitrov D, He Y, Mutoh H, Baker BJ, Cohen L, Akemann W, Knöpfel T (2007) Engineering and characterization of an enhanced fluorescent protein voltage sensor. PLoS One 2:e440

Gautam SG, Perron A, Mutoh H, Knöpfel T (2009) Exploration of fluorescent protein voltage probes based on circularly permuted fluorescent proteins. Front Neuroeng 2:14

Gong Y, Wagner MJ, Zhong Li J, Schnitzer MJ (2014) Imaging neural spiking in brain tissue using FRET-opsin protein voltage sensors. Nat Commun 5:3674. https://doi.org/10.1038/ncomms4674

Gong Y, Huang C, Li JZ, Grewe BF, Zhang Y, Eismann S, Schnitzer MJ (2015) High-speed recording of neural spikes in awake mice and flies with a fluorescent voltage sensor. Science 350:1361–1366

Grenier V, Daws BR, Liu P, Miller EW (2019) Spying on neuronal membrane potential with genetically targetable voltage indicators. J Am Chem Soc 141:1349–1358. https://doi.org/10.1021/jacs.8b11997

Hochbaum DR, Zhao Y, Farhi SL et al (2014) All-optical electrophysiology in mammalian neurons using engineered microbial rhodopsins. Nat Methods 11:825–833. https://doi.org/10.1038/nmeth.3000

Inagaki S, Agetsuma M, Ohara S et al (2019) Imaging local brain activity of multiple freely moving mice sharing the same environment. Sci Rep 9:7460. https://doi.org/10.1038/s41598-019-43897-x

Inoue M, Takeuchi A, Manita S et al (2019) Rational engineering of XCaMPs, a multicolor GECI suite for in vivo imaging of complex brain circuit dynamics. Cell 177:1346–1360.e1324. https://doi.org/10.1016/j.cell.2019.04.007

Jin L, Han Z, Platisa J, Wooltorton JRA, Cohen LB, Pieribone VA (2012) Single action potentials and subthreshold electrical events imaged in neurons with a fluorescent protein voltage probe. Neuron 75:779–785

Kannan M, Vasan G, Huang C et al (2018) Fast, in vivo voltage imaging using a red fluorescent indicator. Nat Methods 15:1108–1116. https://doi.org/10.1038/s41592-018-0188-7

Knöpfel T, Diez-Garcia J, Akemann W (2006) Optical probing of neuronal circuit dynamics: genetically encoded versus classical fluorescent sensors. Trends Neurosci 29:160–166

Kralj JM, Douglass AD, Hochbaum DR, Maclaurin D, Cohen AE (2011a) Optical recording of action potentials in mammalian neurons using a microbial rhodopsin. Nat Methods 9:90–95. https://doi.org/10.1038/nmeth.1782

Kralj JM, Hochbaum DR, Douglass AD, Cohen AE (2011b) Electrical spiking in *Escherichia coli* probed with a fluorescent voltage-indicating protein. Science 333:345–348. https://doi.org/10.1126/science.1204763

Lam AJ, St-Pierre F, Gong Y et al (2012) Improving FRET dynamic range with bright green and red fluorescent proteins. Nat Methods 9:1005–1012. https://doi.org/10.1038/nmeth.2171

Lee S, Geiller T, Jung A, Nakajima R, Song YK, Baker BJ (2017) Improving a genetically encoded voltage indicator by modifying the cytoplasmic charge composition. Sci Rep 7:8286. https://doi.org/10.1038/s41598-017-08731-2

Losi G, Lia AM, Gomez-Gonzalo M, Zonta M, Carmignoto G (2019) Optogenetic interneuron stimulation and calcium imaging in astrocytes. Methods Mol Biol 1925:173–182. https://doi.org/10.1007/978-1-4939-9018-4_16

Lundby A, Mutoh H, Dimitrov D, Akemann W, Knöpfel T (2008) Engineering of a genetically encodable fluorescent voltage sensor exploiting fast Ci-VSP voltage-sensing movements. PLoS One 3:e2514

Mishina Y, Mutoh H, Knöpfel T (2012) Transfer of Kv3.1 voltage sensor features to the isolated Ci-VSP voltage-sensing domain. Biophys J 103:669–676

Miyawaki A, Llopis J, Heim R, McCaffery JM, Adams JA, Ikura M, Tsien RY (1997) Fluorescent indicators for Ca2+ based on green fluorescent proteins and calmodulin. Nature 388:882–887. https://doi.org/10.1038/42264

Muto A, Ohkura M, Kotani T, Higashijima S-I, Nakai J, Kawakami K (2011) Genetic visualization with an improved GCaMP calcium indicator reveals spatiotemporal activation of the spinal motor neurons in zebrafish. Proc Natl Acad Sci U S A 108:5425–5430

Nagai T, Sawano A, Park ES, Miyawaki A (2001) Circularly permuted green fluorescent proteins engineered to sense Ca2+. Proc Natl Acad Sci U S A 98:3197–3202. https://doi.org/10.1073/pnas.051636098

Neher E (1995) The use of fura-2 for estimating Ca buffers and Ca fluxes. Neuropharmacology 34:1423–1442

Oh HJ, Hwang DW, Youn H, Lee DS (2013) In vivo bioluminescence reporter gene imaging for the activation of neuronal differentiation induced by the neuronal activator neurogenin 1 (Ngn1) in neuronal precursor cells. Eur J Nucl Med Mol Imaging 40:1607–1617. https://doi.org/10.1007/s00259-013-2457-0

Oh Y, Park Y, Cho JH et al (2019) An orange calcium-modulated bioluminescent indicator for non-invasive activity imaging. Nat Chem Biol 15:433–436. https://doi.org/10.1038/s41589-019-0256-z

Pedelacq JD, Cabantous S, Tran T, Terwilliger TC, Waldo GS (2006) Engineering and characterization of a superfolder green fluorescent protein. Nat Biotechnol 24:79–88. https://doi.org/10.1038/nbt1172

Perez Koldenkova V, Nagai T (2013) Genetically encoded Ca(2+) indicators: properties and evaluation. Biochim Biophys Acta 1833:1787–1797. https://doi.org/10.1016/j.bbamcr.2013.01.011

Perron A, Mutoh H, Launey T, Knöpfel T (2009) Red-shifted voltage-sensitive fluorescent proteins. Chem Biol 16:1268–1277

Persechini A, Lynch JA, Romoser VA (1997) Novel fluorescent indicator proteins for monitoring free intracellular Ca2+. Cell Calcium 22:209–216

Piatkevich KD, Jung EE, Straub C et al (2018) A robotic multidimensional directed evolution approach applied to fluorescent voltage reporters. Nat Chem Biol 14:352–360

Quinn TA, Camelliti P, Rog-Zielinska EA et al (2016) Electrotonic coupling of excitable and nonexcitable cells in the heart revealed by optogenetics. Proc Natl Acad Sci U S A 113:14852–14857. https://doi.org/10.1073/pnas.1611184114

Shiba Y, Fernandes S, Zhu WZ et al (2012) Human ES-cell-derived cardiomyocytes electrically couple and suppress arrhythmias in injured hearts. Nature 489:322–325. https://doi.org/10.1038/nature11317

Shimomura O, Johnson FH, Saiga Y (1962) Extraction, purification and properties of aequorin, a bioluminescent protein from the luminous hydromedusan, Aequorea. J Cell Comp Physiol 59:223–239

Sonoda K, Matsui T, Bito H, Ohki K (2018) Astrocytes in the mouse visual cortex reliably respond to visual stimulation. Biochem Biophys Res Commun 505:1216–1222. https://doi.org/10.1016/j.bbrc.2018.10.027

St-Pierre F, Marshall JD, Yang Y, Gong Y, Schnitzer MJ, Lin MZ (2014) High-fidelity optical reporting of neuronal electrical activity with an ultrafast fluorescent voltage sensor. Nat Neurosci 17:884–889. https://doi.org/10.1038/nn.3709

Sung U, Sepehri-Rad M, Piao HH, Jin L, Hughes T, Cohen LB, Baker BJ (2015) Developing fast fluorescent protein voltage sensors by optimizing FRET interactions. PLoS One 10:e0141585

Suzuki J, Kanemaru K, Iino M (2016) Genetically encoded fluorescent indicators for organellar calcium imaging. Biophys J 111:1119–1131. https://doi.org/10.1016/j.bpj.2016.04.054

Suzuki K, Onishi T, Nakada C et al (2018) Uninterrupted monitoring of drug effects in human-induced pluripotent stem cell-derived cardiomyocytes with bioluminescence Ca(2+) microscopy. BMC Res Notes 11:313. https://doi.org/10.1186/s13104-018-3421-7

Tsutsui H, Karasawa S, Okamura Y, Miyawaki A (2008) Improving membrane voltage measurements using FRET with new fluorescent proteins. Nat Methods 5:683–685. https://doi.org/10.1038/nmeth.1235

van Opbergen CJM, Koopman CD, Kok BJM et al (2018) Optogenetic sensors in the zebrafish heart: a novel in vivo electrophysiological tool to study cardiac arrhythmogenesis. Theranostics 8:4750–4764. https://doi.org/10.7150/thno.26108

Waldeck-Weiermair M, Bischof H, Blass S et al (2015) Generation of red-shifted chameleons for imaging Ca(2)(+) dynamics of the endoplasmic reticulum. Sensors 15:13052–13068. https://doi.org/10.3390/s150613052

Wang D, Zhang Z, Chanda B, Jackson MB (2010) Improved probes for hybrid voltage sensor imaging. Biophys J 99:2355–2365. https://doi.org/10.1016/j.bpj.2010.07.037

Wu JY, Lam YW, Falk CX et al (1998) Voltage-sensitive dyes for monitoring multineuronal activity in the intact central nervous system. Histochem J 30:169–187

Mechanistic Image-Based Modelling: Concepts and Applications

Denis Menshykau and Simon Tanaka

Contents

1 Introduction 232
2 Mathematical Description of Spatially Resolved Systems 233
2.1 Chemical Gradients: Reaction-Diffusion System 233
2.2 Stress, Strain and Deformation: Solid Mechanics 236
2.3 Blood Flow: Fluid Mechanics 237
3 Mechanistic Image-Based Modelling Approach: Concepts and Examples from Developmental Biology 238
3.1 Branching Morphogenesis 238
3.2 Limb Bud Elongation 241
3.3 Genetics of Geometry 244
4 Image-Based Modelling for Biomedical Applications 244
4.1 Solid Mechanics: Dental Implantology 244
4.2 Fluid Dynamics: Vasculatures, Stents, Valves and Diagnostics 245
4.3 Image-Based Multi-scale Heart Modelling 249
5 Image-Based Modelling in Drug Discovery and Development 250
5.1 ADME Applications 250
5.2 Towards Pharmacodynamic Applications 253
6 Outlook 254
References 255

D.M. is employed by Bayer AG, Wuppertal, Germany.
S.T. is employed by Cyfex AG, Zürich, Switzerland.

The opinions expressed in this manuscript are those of the authors and do not necessarily reflect the views of their employers.

D. Menshykau (✉)
Düsseldorf, Germany

S. Tanaka
Zurich, Switzerland

J. E. Barrett et al. (eds.), *Concepts and Principles of Pharmacology*,
Handbook of Experimental Pharmacology 260, https://doi.org/10.1007/164_2019_328

Abstract

Advancements in imaging techniques have led to a rapid growth of available imaging data. Interpretation of the imaging data and extraction of biologically, physiologically and/or medically relevant information, however, remains challenging. In contrast, mechanistic computational modelling provides a means to formalise and dissect mechanisms governing the behaviour of complex systems. However, its application often is limited due to the lack of relevant data for model building and validation. Exploitation of the imaging data to build, parameterise and validate computational models gives rise to an image-based modelling approach. In this chapter, we introduce the basics of the mechanistic image-based modelling approach and review its application in developmental biology and biomedical research as well as for medical device development and drug discovery and development. Implementation of image-based modelling in pharmaceutical industry holds promise to further advance model-informed drug discovery and development and aids substantially in our understanding of drug pharmacokinetic, pharmacodynamic and ultimately de-risk drug development.

Keywords

Biomedical engineering · Developmental biology · Drug discovery and development · Image-based modelling · Mechanistic modelling · Modelling and simulation · Model informed drug development

1 Introduction

A mathematical model is a simplified and formalised description of a real-world system. As real-world systems are inherently complex, in particular biological systems, mathematical models typically focus only on certain aspects of system composition and/or behaviour. Theoretical analysis of an adequately established mathematical model provides insight into properties and behaviour of a real-world system it represents. Due to the complexity of real-world systems, their mathematical models are also complex, and their analysis is conducted numerically on a computer giving rise to computational modelling. Development and application of computational models can be roughly classified as "scientifically" motivated with a goal of gaining novel insight into the mechanisms governing system behaviour, and "engineeringly" motivated with a goal to forecast a system's behaviour under certain conditions. "Scientifically" motivated modelling typically precedes engineering applications; the latter are employed to optimise problems which are qualitatively understood. In any case, a mechanistic computational model requires substantial knowledge about the inner mechanics of the system, as well as exact values (or their distributions) of parameters describing system properties, which are often not known

and require substantial effort to measure them. For biological systems, measurements can be extremely challenging and time consuming. On the other hand, advances in imaging techniques have led to a rapid growth of available imaging data; for instance, a routine imaging experiment could generate more than 1 terabyte of data (Kherlopian et al. 2008; Chen et al. 2014; Troy and Kris 2017). Extracting biologically relevant, human interpretable knowledge from these large spatio-temporal datasets is challenging due to the lack of adequate tools and sheer size (Myers 2012). Development of spatially resolved computational models based on imaging data gives rise to image-based modelling. The strength of this approach is in that it directly uses imaging data to build, parameterise and validate mechanistic computational models, and it formalises hypothesis testing based on imaging data. Ultimately, it is an effective approach to transform large experimentally acquired datasets into human interpretable knowledge.

In this chapter, we introduce basics of the image-based modelling approach and review its applications in developmental biology and biomedical research as well as for medical device development and drug discovery and development. Examples from developmental biology and basic biomedical research illustrate applications of image-based modelling to gain understanding of mechanisms governing systems behaviour. Other examples demonstrate applications of already established image-based models for medical device and drug development. To facilitate understanding of various application areas of image-based modelling by a diverse audience, we added Sect. 2 with a very brief description of typical mathematical models describing fluid flow, solid mechanics and reaction-diffusion systems, which provide mathematical formalisms for mechanistic image-based modelling. Computational methods for image processing, computational mesh generation, solution of spatially distributed models, etc. are out of scope of this review.

2 Mathematical Description of Spatially Resolved Systems

In this section, we provide a very brief introduction into the mathematical description of chemical and physical laws governing behaviour of biological systems described in Sects. 3–5. A detailed discussion of mathematical formalism, its analysis and implications for physiological and biochemical systems is available elsewhere (Murray 2003; Keener and Sneyd 2009).

2.1 Chemical Gradients: Reaction-Diffusion System

Description of a complex system like a human lung or heart requires explicit consideration of spatial dimensions. Partial differential equations (PDEs) provide a mathematical formalism to describe system evolution in space over time. The overall structure of partial differential equations is defined by nature of the phenomena under consideration. A morphogen is a chemical signalling molecule, whose non-uniform distribution establishes a pattern of tissue growth and differentiation during foetal development. Typically, morphogens are synthesised by source cells

and spread through tissue by passive diffusion as well as being involved into biochemical reactions of degradation, binding, etc. Morphogen concentration $L = L(\boldsymbol{x}, t)$ in time and space could be described by the reaction-diffusion partial differential equation:

$$\underbrace{\frac{\partial L}{\partial t}}_{\text{time derivative}} = \underbrace{D_L \Delta L}_{\text{diffusion}} + \underbrace{\rho_L}_{\text{production}} - \underbrace{\delta_L L}_{\text{degradation}} + \underbrace{F(L, \ldots)}_{\text{other reactions}} \tag{1}$$

On the left-hand side of the equation, the time derivative $\frac{\partial L}{\partial t}$ describes the local change in molecular concentration. On the right-hand side, the diffusion term $D_L \Delta L$ describes ligand diffusion in the tissue along the concentration gradient $\frac{\partial L}{\partial x}$. D_L refers to the diffusion coefficient and Δ to the Laplace operator. The production term ρ_L describes the constitutive production rate, and the degradation term $-\delta_L L$ describes ligand degradation under the assumption that depletion of the local concentration in time is proportional to the immediate local concentration. $F(L, \ldots)$ describes other potential reactions the ligand L is involved into. Overall, the local change of concentration in time is equal to the change induced by all the reactions, plus diffusional flux. Detailed derivations and discussions of reaction-diffusion equations from first principles are available elsewhere (Murray 2003).

The structural Eq. (1) has to be accompanied by initial and boundary conditions. Initial conditions define the values of all variables at the temporal moment chosen to be the starting point of the model. In case of a reaction-diffusion system described by Eq. (1), the initial value describes the concentration of the ligand L along the spatial coordinate x. A simple example of initial values would be a uniform distribution of concentration C along the entire domain at time zero: $L(x, t = 0) = C$. In addition to the structural equation and initial conditions, PDEs have to be accompanied by boundary conditions which define system behaviour at the boundary of the spatial domain under consideration. The most frequently encountered boundary conditions for reaction-diffusion systems are fixed concentration boundary conditions (also known as Dirichlet boundary conditions) $L|_\Omega = f$ and fixed flux boundary conditions (also known as Neumann boundary conditions) $\frac{\partial L}{\partial x}\big|_\Omega = f$.

Let us now consider a simple, yet realistic example of a reaction-diffusion system. Figure 1a depicts a line of cells comprising two cell types: ligand secreting cells (in blue) and cells which do not secret any ligand (in green). This situation resembles the one observed in the developing fly wings (Fig. 1b) (Kicheva et al. 2007). Let us assume that the concentration of the ligand in the proximity of ligand secreting blue cells is at steady state and equal to the ratio of ligand production and elimination rate constants: ρ_L/δ_L. On the domain comprised of green cells, only ligand degradation is observed. The structural equation describing this system is depicted in Fig. 1a, accompanied by the boundary condition $L|_{x\,=\,0} = \rho_L/\delta_L$ on the left-hand side of the domain and no flux boundary conditions on the right-hand side of the domain $\frac{\partial L}{\partial x}\big|_{x=x0} = 0$. The time dependent solution of this partial differential equation shows that L propagates from its source on the left side of the domain into the depth of the tissue until steady state is reached. At the steady state, local degradation rate is equal

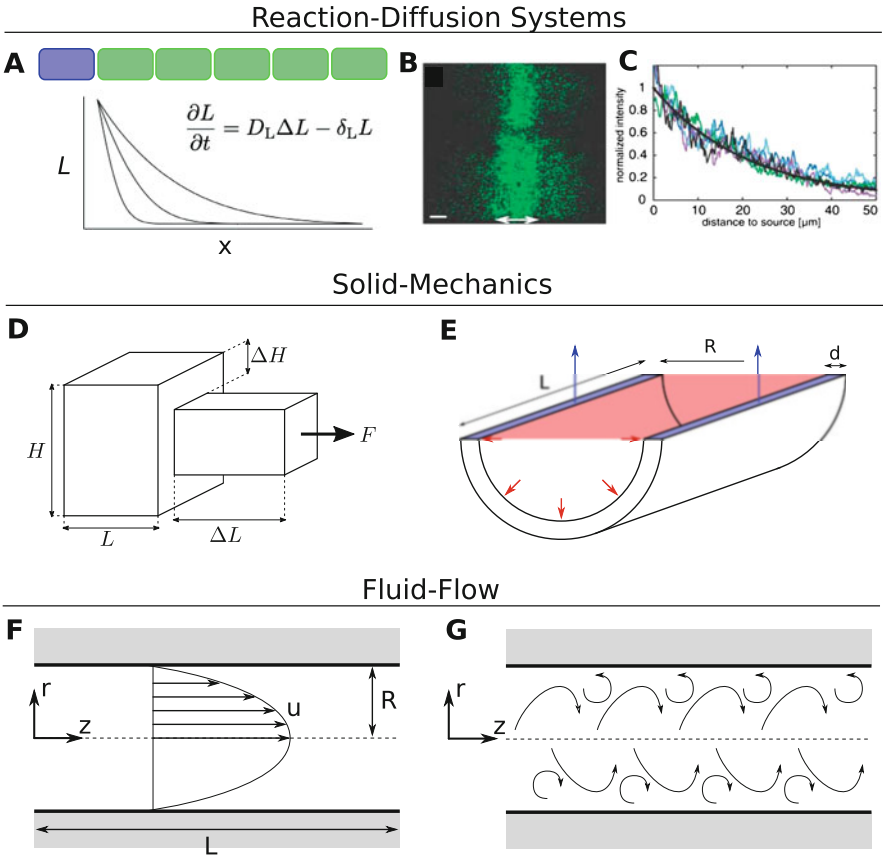


Fig. 1 Mathematical description of spatially resolved biological systems. (**a**–**c**) Reaction-diffusion systems. (**a**) Solution of diffusion – linear degradation equation on a one-dimensional domain. (**b**) Wing disc showing GFP-Dpp (green) expressed in the endogenous source (double arrow). (**c**) Normalised average fluorescence in the receiving territory of five GFP-Dpp expressing discs as a function of the distance to the source. Black curve, exponential fit to the black trace. (**b**, **c**) Reproduced with permission from (Kicheva et al. 2007). (**d**–**e**) Solid mechanics. (**d**) Upon applying a force F, a hypothetical solid material is stretched. The Young modulus $E = \frac{F/A}{\Delta L}$ is defined as the proportionality factor between force F per area A, and the relative length change $\Delta L/L$. (**e**) Assuming a tube of length L, radius R, wall thickness d and a static pressure difference of Δp between the interior and exterior, the interior pressure exerts forces perpendicular to the inner tube surface (red arrows). By virtually cutting the tube horizontally and integrating all those forces on one half of the tube, a net force can be computed. In order to fulfill the balance of forces, this net force has to be counteracted by opposing forces: those forces are brought up by the tube walls (blue arrows). (**f**–**g**) Fluid flow. (**f**) A linear Newtonian fluid is flowing at fully developed, steady-state and laminar conditions in a tube of radius R and length L. The resulting velocity profile only has a non-zero velocity component in z direction, which is distributed parabolically across the tube diameter (Eq. 7). (**g**) A schematic representation of turbulent flow

to the diffusional flux of the ligand from the source $\delta_L L = D_L \Delta L$. This differential equation has the following analytical solution:

$$L = \rho_L / \delta_L \times e^{-x(\delta_L / D_L)^{1/2}} \tag{2}$$

which describes an exponential decay of a ligand concentration with a distance from the source.

Experimentally measured morphogen *Decapentaplegic* (Dpp) and *Wingless* (Wg) concentrations in fruit fly wing (Fig. 1b) disc were shown to be well described by this reaction-diffusion equation (Fig. 1c) (Kicheva et al. 2007).

2.2 Stress, Strain and Deformation: Solid Mechanics

Solid mechanics studies the motion and deformation of solid materials under application of forces, e.g. the deformation of a solid cube in response to mechanical stress (Fig. 1d). In solid mechanics, the following two measures are of paramount importance: stress $\sigma = F/A$, which is defined as acting force F per area A; and strain $\epsilon = \Delta L/L$, which is the relative length change of a material. Solid materials can respond differently to applied force; here we consider the elastic response only. A material is defined to be elastic if it deforms under stress but relaxes to its original shape upon stress removal. Elasticity is characterised by the Young modulus $E = \sigma/\epsilon$, which is defined as the proportionality factor between stress σ and strain ϵ (Fig. 1d).

Let us consider stress and strain in a blood vessel wall. A segment of a blood vessel could be approximated by a tube of length L, radius R, wall thickness d and a static pressure difference of Δp across the wall (Fig. 1e). By virtually cutting the tube in half, the net force acting on each of the halves (red area in Fig. 1e) is $F_p = 2RL\Delta p$, which has to be counteracted by forces of equal amplitude but opposite direction. These counteracting forces are brought up by the vessel wall. To compute the mechanical tangential stress in the tube wall, this counteracting net force is simply divided by the total wall surface (blue area in Fig. 1e):

$$\sigma_\theta = \frac{F_p}{2dL} = \frac{R\Delta p}{d} \tag{3}$$

The change in the vessel radius in response to pressure change could be calculated as:

$$\frac{\Delta R}{R} = \frac{1}{E}\sigma_\theta = \frac{R\Delta p}{Ed} \tag{4}$$

where E is the Young modulus of the blood vessel wall.

This relationship defines the blood vessels ability to stretch in response to a pressure pulse. It has been found that blood vessels show a non-linear stress-strain relationship, i.e. they have a lower Young modulus at low stress, which increases at

higher stresses (Wagenseil and Mecham 2012). The result is that the blood vessels are stretched relatively easily at low blood pressure but stiffen more and more with increasing blood pressure (Silva Vieira et al. 2015).

2.3 Blood Flow: Fluid Mechanics

Unlike solids, a fluid continually deforms (flows) under an applied external force. If for solid bodies stress and strain is of paramount importance (Sect. 2.2), for fluids these are stress and strain rate (the latter being defined as deformation per time). The proportionality factor between stress and velocity is viscosity μ. If viscosity μ is constant over a wider range of applied stress, the fluid is called Newtonian. In reality, there is no perfectly Newtonian fluid, but the Newtonian fluid model turns out to be a good approximation of many practically relevant fluids under normal condition, including air and water flow.

The dynamics of incompressible Newtonian fluids is described by the Navier-Stokes equation (Eq. 5):

$$\rho\left(\frac{\partial \boldsymbol{u}}{\partial t} + (\nabla \cdot \boldsymbol{u})\boldsymbol{u}\right) = -\nabla p + \mu\left(\Delta \boldsymbol{u} + \frac{1}{3}\nabla(\nabla \cdot \boldsymbol{u})\right) + \boldsymbol{f}$$
$$\rho \nabla \cdot \boldsymbol{u} = \mathcal{S} \tag{5}$$

where ρ is the mass density, $\boldsymbol{u} = \boldsymbol{u}(\boldsymbol{x}, t)$ the fluid velocity at spatial location $\boldsymbol{x}$ and time t, $\partial u/\partial t$ the temporal partial derivative vector of $\boldsymbol{u}$, $\nabla \cdot \boldsymbol{u}$ the divergence operator of the velocity field $\boldsymbol{u}$, $p = p(\boldsymbol{x}, t)$ the pressure, ∇p the gradient vector of p, μ the dynamic viscosity and $\boldsymbol{f}$ an external force. $\mathcal{S} = \mathcal{S}(\boldsymbol{x}, t)$ is a source term typically equal to zero for physical liquids. However, it could be modelled to be non-zero to capture local tissue growth (Sect. 3.2).

When describing flow conditions, one dimensionless number is of outmost importance: the Reynolds number $Re = uL\rho/\mu$. This number describes the ratio between inertial and viscous forces. Lets consider a fully developed steady-state laminar flow ($Re < 2{,}300$) of Newtonian fluid with dynamic viscosity μ in a tube with radius R and length L, so-called Hagen-Poiseuille flow (White 1981). Since the geometry is axisymmetric, the Navier-Stokes Eq. (5) can be rewritten in cylindrical coordinates, i.e. $\boldsymbol{u} = \boldsymbol{u}(r, \phi, z)$ with r being the distance from the centre line, ϕ the azimuth angle and z the distance along the centre line (Vitturi 2016):

$$\frac{1}{r}\frac{\partial}{\partial r}\left(r\frac{\partial u_z}{\partial r}\right) = \frac{1}{\mu}\frac{\partial p}{\partial z} \tag{6}$$

This partial differential equation can be solved analytically, and the result is, alongside $u_r = 0$ and $u_\phi = 0$:

$$u_z(r,\phi,z) = -\frac{\Delta p}{4L\mu}\left(R^2 - r^2\right) \tag{7}$$

where Δp is the pressure drop along the tube length L.

Equation (7) states that the velocity profile is parabolic across the tube diameter, i.e. zero at the tube wall and maximal on the centre line (Fig. 1f).

At high fluid speed, large flow geometry and smaller viscosity, the Reynolds number is high. At $Re > 2{,}300$, a flow is typically becoming unstable and transiting into the turbulent regime (Fig. 1g). Real-world examples are aircraft moving in air, air being in- and exhaled in lungs, blood flow in large arteries after branch points, stenotic arteries and stenotic heart valves.

3 Mechanistic Image-Based Modelling Approach: Concepts and Examples from Developmental Biology

Section 2 introduced mathematical descriptions for a number of simple biochemical and physiological systems. These examples demonstrated that mathematical description of spatially resolved biological systems requires knowledge of: structural equations governing behaviours of the systems; boundary and initial conditions as well as the geometry of the domain on which the system exists. Furthermore, exact values for system properties (parameters), such as rate constants, diffusion coefficients, etc. are also required. It is often the case that all of these are known only to a limited extent, and extensive data is required to build and validate a computational model. Advances in imaging techniques (Kherlopian et al. 2008; Chen et al. 2014; Troy and Kris 2017) across scales and modalities lead to a rapid growth of available imaging data. Methods to extract biologically and physiologically meaningful information from this data are being developed (Myers 2012). The image-based modelling approach naturally bundles mechanistic computational modelling with abundant imaging data. The strength of the approach is that, firstly, based on imaging data it provides realistic computational domains to define mechanistic models on, and secondly, imaging data can be directly used to parameterise, test and select models.

3.1 Branching Morphogenesis

In this section, we introduce the image-based modelling approach by discussing the example of branching morphogenesis in embryonic kidneys and lungs, and demonstrate how it can be applied to establish mechanisms governing behaviour of complex systems. Mechanisms for branching morphogenesis have been studied for decades, and a number of alternative morphogen-based models have been proposed to govern growth area specification (Iber and Menshykau 2013). Image-based modelling provides a means to quantitatively evaluate the capacity of alternative models to correctly predict areas of growth during branching morphogenesis. To this

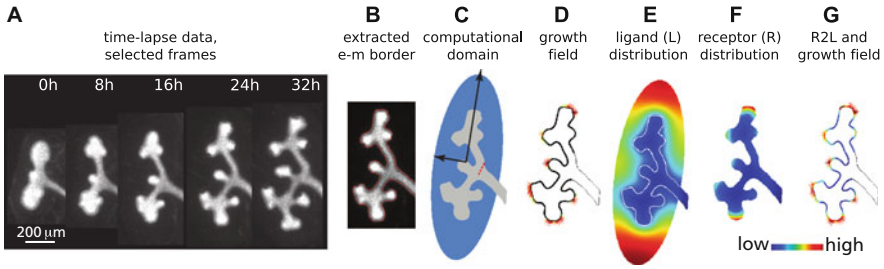


Fig. 2 The image-based modelling approach. (**a**) Snapshots from the time-lapse movie of kidney branching morphogenesis at the indicated time points. (**b**) The red curve marks the extracted border of the epithelium of the ureteric bud. (**c**) The computational domain comprises the epithelium (grey) and the mesenchyme (blue). (**d**) The growth field (red arrows) and the epithelial border (black line). (**e**, **f**) The computed distribution of the (**e**) ligand L, (**f**) receptor R. (**f**) The predicted ligand-receptor signalling (R^2L, solid line) at the epithelium-mesenchyme border and the growth field (vectors). (**d**–**g**) The relative strength of the signalling and of the growth field is encoded according to the colour bar. Adapted from (Menshykau et al. 2019)

end, time-lapse movies of kidney branching morphogenesis were recorded ex vivo (Fig. 2a). The images were segmented, borders delineating the shape of the renal epithelium at a given time point were extracted (Fig. 2b), and computational domains were defined on the extracted geometries (Fig. 2c). A reaction-diffusion type model (Sect. 2.1) describing receptor-ligand interactions (Eq. 8) were defined and solved on the computational domain in the shape of embryonic kidney explant (Fig. 2c) (Menshykau et al. 2019).

$$\begin{aligned}
\text{Epithelium}: \quad & \underbrace{\frac{\partial R}{\partial t}}_{\text{time derivative}} = \underbrace{D_R \Delta R}_{\text{diffusion}} + \underbrace{\gamma\left(a - R + R^2 L\right)}_{\text{biochemical reactions}} \\
\text{Epithelium}: \quad & \underbrace{\frac{\partial L}{\partial t}}_{\text{time derivative}} = \underbrace{D_L \Delta L}_{\text{diffusion}} - \underbrace{\gamma\left(-R^2 L\right)}_{\text{biochemical reactions}} \\
\text{Mesenchyme}: \quad & \underbrace{\frac{\partial L}{\partial t}}_{\text{time derivative}} = \underbrace{D_L \Delta L}_{\text{diffusion}} + \underbrace{\gamma(b - L)}_{\text{biochemical reactions}}
\end{aligned} \tag{8}$$

As described in Sect. 2, on the left-hand side of the equations are the time derivatives describing a local change in molecular concentration. On the right-hand side are the diffusion terms, $D_R \Delta R$ and $D_L \Delta L$, and the reaction terms. Here, D_i refers to the diffusion coefficient and Δ to the Laplace operator, where $i = \{R, L\}$. Receptors are produced at rate a in the epithelium, and ligands are produced at rate b in the mesenchyme. The receptors and ligands are turned over independently by linear decay at a unit rate $-R$ and $-L$, accordingly. R^2L is a lump term for receptor-ligand binding, sequestration and induced increase in receptor concentration. γ describes the

relative strength of reaction terms as compared to diffusional terms. Details of the derivation and mathematical analysis are described in (Menshykau et al. 2012, 2019; Menshykau and Iber 2013; Kurics et al. 2014).

Exemplary solution of Eq. (8) on the computational domain in the shape of embryonic kidney explant (Fig. 2c) is depicted in Fig. 2e, f. A solution of the alternative models could also be obtained on the same computational domain, and therefore quantitative criteria for model evaluation have to be formulated. As the model's aim is to predict areas of growth observed during branching morphogenesis, this information was extracted from the experimentally registered time-lapse data in Fig. 2d. Figure 2g depicts an overlay of the growth field extracted from the time-lapse data and model predicted growth areas. The quantitative difference between predicted and computed growth areas can be calculated according to

$$\Delta = \sqrt{\int_{\partial\Omega} (C - E)^2 d\Omega}, \tag{9}$$

where E is the normalised magnitude of the growth field extract from time-lapse data, C is the normalised computed intensity of receptor-ligand signalling and Ω is an epithelium-mesenchyme border.

Quantitative evaluation of competing models for kidney branching morphogenesis demonstrated that the Turing-type ligand-receptor model (Fig. 3a, Eq. (8)) yielded the smallest deviation, Δ (Eq. 9), between the spatial distribution of signalling strength C and the measured experimental growth fields E for the vast majority of the analysed time frames (Fig. 3b, black) as well as the smallest global deviation for the entire time-lapse movie (Fig. 3c, black) compared to the alternative non-Turing models (Fig. 3a). Visual inspection of the simulations with a single globally optimal parameter set (Fig. 3d) and of simulations with different optimal parameter sets for each stage confirms that the Turing-type ligand-receptor model (Fig. 3a: T1) recapitulates the experimentally observed growth fields well, both on each separate time frame in a series of static computational domains.

In the example above, we have illustrated the application of an image-based modelling approach to establish a core mechanism for kidney branching morphogenesis based on 2D time-lapse data. The image-based modelling approach can well be applied to the analysis of 3D imaging data as depicted in Fig. 4. A number of models have been proposed to explain the control of the branching events during lung branching morphogenesis (Bellusci et al. 1997a; Clément et al. 2012a, b; Nelson et al. 2006; Gleghorn et al. 2012). To quantitatively test receptor-ligand-based Turing-type mechanisms and alternative models, we used a sequence of 3D geometries of mouse embryonic lungs (Blanc et al. 2012). 3D imaging data was converted into computational meshes and displacement fields showing areas of growth (Fig. 4a–c) (Menshykau et al. 2014). Next, alternative models were formulated and solved on the obtained computational domains. In particular, we tested various receptor-ligand-based Turing-type models (Eq. 8), models based on the epithelium to mesenchyme distance (Bellusci et al. 1997b), gradient-based

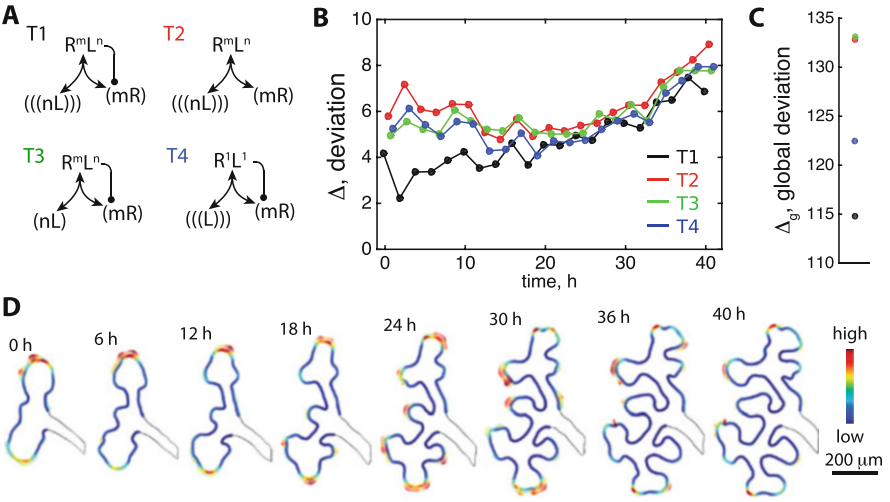


Fig. 3 Image-based data from wild-type kidneys supports a ligand-receptor-based Turing mechanism. (**a**) Schematic representation of the tested models: (T1) ligand-receptor-based Turing model; (T2) like T1 but without receptor upregulation; (T3) like T1 but with equal diffusion coefficients for receptors and ligands; (T4) like T1 but with 1:1 stoichiometry of the ligand-receptor complex. (**b**, **c**) Minimal deviation between the spatial distribution of signalling strengths *C* and the experimentally measured growth field *E* for (**b**) each time frame (Δ, Eq 9) and (**c**) globally. The colours represent the different models, T1 – black, T2 – red, T3 – green and T4 – blue. (**d**) The growth areas predicted by the ligand-receptor-based model with the globally optimal parameter set (solid colour) match the growth fields extracted from the experimental data (vectors). The relative strength of the signalling and of the growth field is encoded according to the colour bar. Adapted from (Menshykau et al. 2019)

models (Clément et al. 2012b) and models based on the tissue-specific expression (Nelson et al. 2006). We further calculated deviation Δ (Eq. 9) between the growth field predicted by models and that extracted from the experimental data (Fig. 4c). We found that the receptor-ligand-based Turing-type model yields the minimal deviation Δ as compared to all alternative models (Fig. 4d, e). Visual examination further demonstrates that the receptor-ligand-based Turing-type models correctly predict all areas of growth (Fig. 4d′), where alternative models fail to do so (Fig. 4e′). Analysis of the later stages shows that, despite more complex geometries of the lung, receptor-ligand-based Turing-type mechanisms successfully predict areas of growth on these complex geometries.

3.2 Limb Bud Elongation

Application of mechanistic image-based modelling in developmental biology is not limited to reaction-diffusion systems. An image-based modelling approach was applied to reveal mechanisms governing morphogenesis in a number of developmental systems: limb growth, leaf and corolla development, drosophila gene patterning,

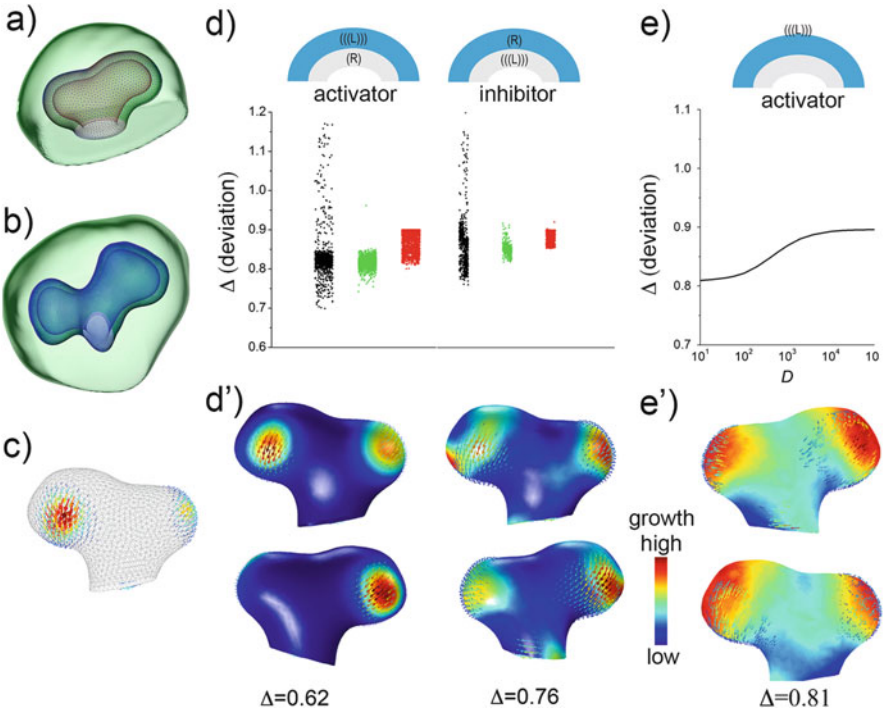


Fig. 4 Receptor-ligand-based Turing mechanism recapitulates areas of growth observed during lung branching morphogenesis. (**a**, **b**) Computational domains in the shape of the R.Cr lobe of mouse embryonic lungs; (Blanc et al. 2012) (**c**) growth areas from stage (**a**) to (**b**); (**d**) deviation, Δ, of the growth areas predicted by the receptor-ligand-based Turing mechanism from that observed experimentally. Black, red and green dots represent parameter sets from the Turing space, out of the Turing space and those which cannot be classified reliably; (**d′**) receptor-ligand-based Turing type models reproduce well the growth field extracted from the experimental data; (**e**) deviation, Δ, computed for the best performing alternative model – the ligand is expressed in the mesothelium and activates epithelial bud outgrowth; (**e′**) growth areas predicted by an alternative model cannot reproduce the experimental growth field (Menshykau et al. 2014). Reproduced with permission from (Menshykau et al. 2014)

etc. (Hiscock and Megason 2015; Sbalzarini 2013; Sharpe 2017; Coen and Rebocho 2016; Coen et al. 2017). Below, we demonstrate the application of image-based modelling to various developmental systems with different governing models describing the systems behaviour.

The vertebrate limb is a classical model system in developmental biology (Scott 2013). Growth of the bud is strongly differentiated – extension in the distal direction (away from the body) is dramatic, while, by comparison, the increase in width and height is much slower. Boehm et al. (Boehm et al. 2010) tested contributions of isotropic cell division and directional cell behaviour during limb morphogenesis by solving computational models for tissue growth on experimentally acquired 3D

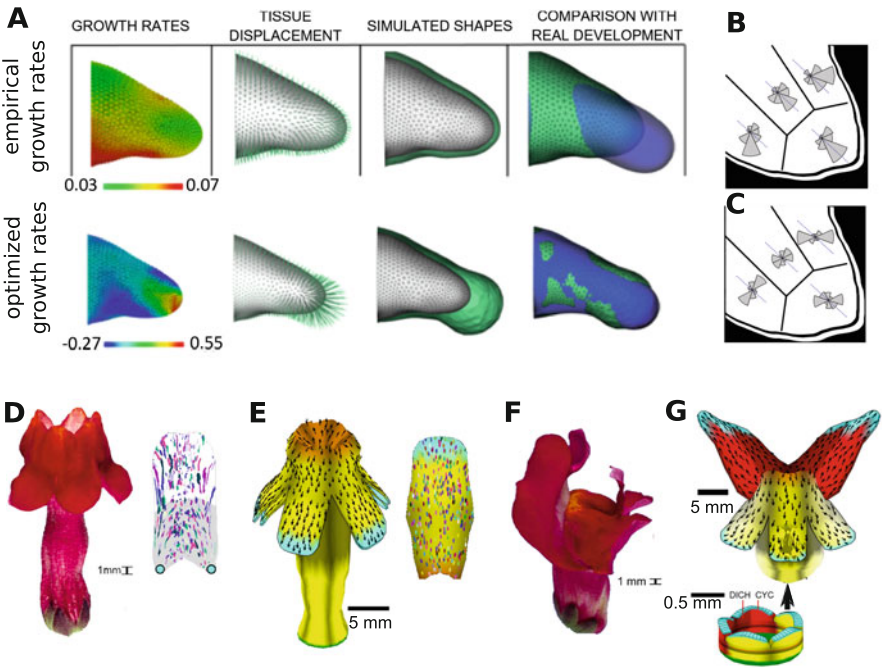


Fig. 5 Interplay of directional growth and mechanical constrains in limb and corolla morphogenesis. Spatially controlled cell proliferation in limb bud morphogenesis (**a**–**b**). (**a**) Simulated limb bud shapes with empirical (upper row) and optimised (lower row) proliferation rates. (**b**, **c**) Observed golgi (**b**) and cell division (**c**) orientations of limb bud mesenchymal cells. Adapted from (Boehm et al. 2010). Emergence of complex shape during corolla morphogenesis (**c**–**f**). (**c**, **d**) Observed and simulated mature *cyc dich* mutant *Antirrhinum* flower and observed clone pattern. (**e**, **f**) Observed and simulated mature *Antirrhinum* flower. Adapted from (Green et al. 2010)

shapes of mouse embryonic limb bud. Over a long timescale, mesenchymal tissue displays a liquid-like characteristic; therefore the Navier-Stokes-type equation (Eq. 5) with a source term S, corresponding to tissue proliferation, was employed to mathematically describe limb bud shape evolution over time.

Computer simulations according to the Navier-Stokes equation with a source term, corresponding to tissue proliferation (Eq. 5), demonstrate that experimentally acquired isotropic tissue proliferation rates imposed on the initial shape of the limb bud extracted from the optical projection tomography data does not account for observed elongation of the limb bud along distal direction (Fig. 5a, upper row). In a computer simulation, isotropic tissue growth rates could be optimised to recapitulate limb bud shape at the later stage (Fig. 5a, lower row); however, simulation requires biologically unrealistically high cell division and death rates. Therefore, authors postulated and experimentally confirmed (Fig. 5b, c) that directional cell activities play a role during limb bud morphogenesis.

3.3 Genetics of Geometry

Plants generate a spectacular diversity of complex shapes and, therefore, serve well as a model system for elucidating mechanisms of morphogenesis. Planar nature of tissues comprising leaves and flowers makes these systems well suited for imaging and clonal analysis. In a series of publications, Coen's lab applied image-based modelling methodology to establish how the combination of genetics, controlling tissue growth and polarity with physical constrains introduced by tissue connectivity, leads to the generation of shape diversity in leaves and flowers (Coen and Rebocho 2016; Coen et al. 2017; Green et al. 2010). Local rates and orientations of growth within *Antirrhinum* flower petal are evident from clonal analysis (Fig. 5d) and are modulated by a complex regulatory network (Green et al. 2010). Correct shape development of *Antirrhinum* flower could be recapitulated in silico (Fig. 5d–g) if modulation of local tissue proliferation rates and orientations of tissue growth by a genetic network are incorporated into the model. Development of floral dorsoventral asymmetry, as observed in wild-type *Antirrhinum* flower (Fig. 5f), requires dorsoventral polariser. Genetic knockout experiments demonstrated that dorsally expressed *CYCLOIDEA (CYC)* and *DICHOTOMA (DICH)* genes, among other functions, modulate relative growth of two surfaces of the petal canvas, preventing bending back of the dorsal lobes. In their absence, flower asymmetry fails to be established with all petals folded back (Fig. 5d–e). A computational model with incorporated CYC and DICH genes correctly recapitulates *Antirrhinum* flower asymmetry (Fig. 5g). Further examples of image-based modelling applications to reveal mechanisms governing biological and developmental systems can be found in the following reviews (Hiscock and Megason 2015; Sbalzarini 2013; Sharpe 2017).

4 Image-Based Modelling for Biomedical Applications

In the section above, we have considered several examples of application of image-based mechanistic modelling to establish mechanisms responsible for emergence of complex shapes, in particular those of organs. Organs evolved to have complex shapes to enable their physiological functions. Below we review selected applications of image-based modelling to establish relationships between organ shape, physiology and the function, as well as functional decline in disease state and its restoration by medical intervention.

4.1 Solid Mechanics: Dental Implantology

Computational image-based modelling has been widely used to study dental and bone prosthetics and is gaining importance in clinical practice (Geng et al. 2001; Alper et al. 2012). Below, we review and discuss a few representative examples.

A mandibulectomy is a surgical removal of all or part of a jaw (mandible) performed to eradicate certain infectious diseases and cancers. Mandibulectomy is typically followed up with bone grafting to reconstruct mandibular anatomy for subsequent abutment implantation (Khan 2012). Grafting material is preferably osteotomised from the individual itself (Sakkas et al. 2017). Two common sources of grafting material for mandibular reconstruction are the patient's ribs (Longacre and Destefano 1957) and fibula (Hidalgo 1989; Louis 2011). In order to study the difference of the mechanical situation in a non-grafted and in a fibula grafted mandible, finite element analysis of mandible solid mechanics model (see Sect. 2.2) extracted from an individual patient's computed tomography (CT) data has been performed (Park and Kwon 2013). It was found that the mechanical stresses do not differ significantly between the original and the fibula grafted mandible and confirmed that fibula grafting is indeed a mechanically appropriate procedure (Fig. 6a–d) (Park and Kwon 2013).

To find a suitable abutment implant position, angle and depth, dentists rely on a range of different X-ray and CT technologies. In recent years, tools have been developed to process various sources of patient data into a virtual patient model to plan dental and surgical procedures digitally (Beuer et al. 2008; Neugebauer et al. 2010). Once an implant position has been found, it has to be executed as precisely as possible. Most commonly, a personalised surgical guide will be produced, i.e. a template which precisely fits on the patient's teeth and features a guide for the drill. Using this approach, even most complex surgical treatments, such as zygoma and pterygoid implants, can be executed safely and precisely (Fig. 6e–g) (Vrielinck et al. 2003). Finite element solution of solid mechanics models (see Sect. 2.2) based on CT data is increasingly employed to predict and assess the mechanical integrity of implant positioning (Chang et al. 2018).

Not only the abutment implants, but also dental crowns and bridges are subject to high forces. Since these are designed for each patient individually, it is of great importance that dentists have tools at hand to ensure mechanical integrity and longevity. Commercial solutions are already used clinically on a daily basis and are a steadily growing segment in the portfolio of all major suppliers of dental technology. Based on intraoral optical scans, dental restorations are designed virtually using specialised computer-aided design (CAD) tools. These tools based on computer models enable dentists to compute mechanical stresses, anticipate weaknesses and iteratively adapt the implant design accordingly (Dentsply Sirona 2019).

Image-based modelling finds applications in bone implant design also beyond dentistry, in particular in design of orthopaedic devices (Prendergast 1997; Leondes 2003). Typically, image-based modelling could be applied to optimise implant shape to minimise implant-induced stress.

4.2 Fluid Dynamics: Vasculatures, Stents, Valves and Diagnostics

Computational fluid dynamics (CFD) based on solving the three-dimensional Navier-Stokes equations (Eq. 5, Sect. 2.3) for blood flow has been used since the

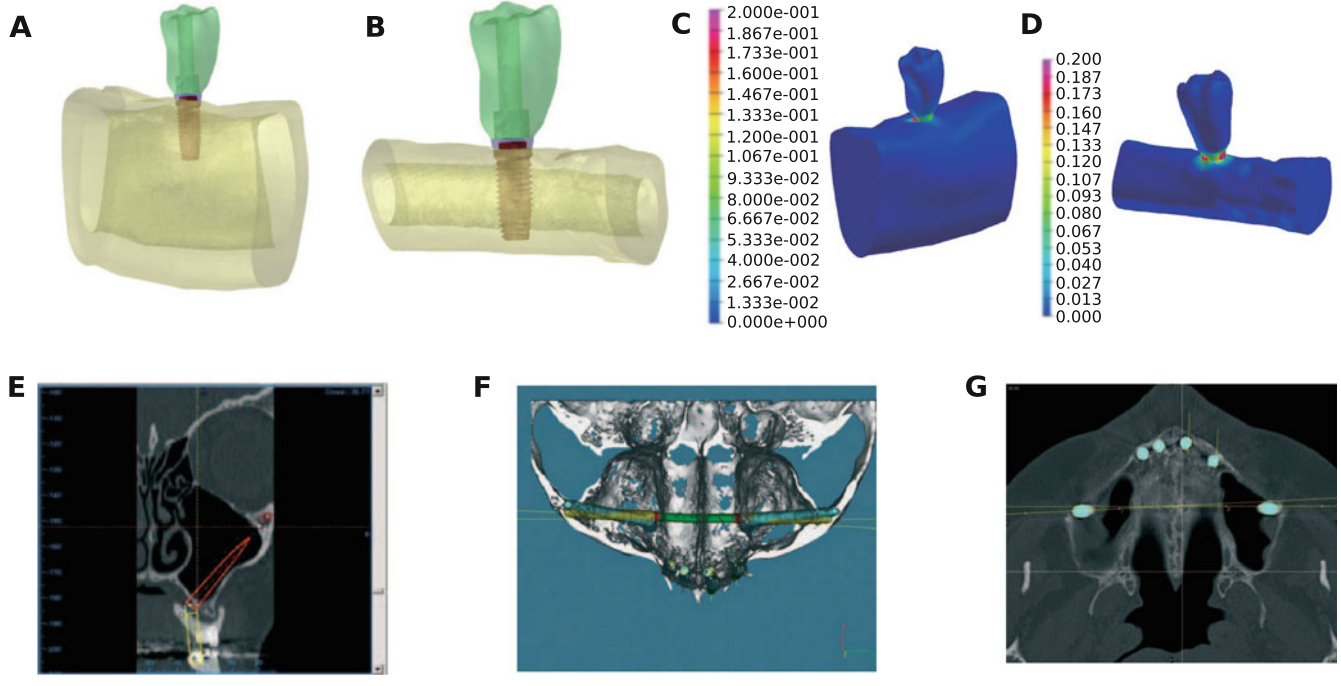


Fig. 6 Image-based modelling for dental prosthetics. (**a**–**d**) Finite element analysis of mandibular abutment implants. Abutment implant in mandibular (**a**) and fibular (**b**) bone. The bone structure has been segmented of an individual patient's computed tomography scan. Finite element analysis has been performed to compute the stress distribution in mandibular (**c**) and fibular bone (**d**). Adapted and reprinted with permission (Park and Kwon 2013). (**e**–**g**) Digital planning of zygoma and pterygoid implant positions. (**e**) In a 2D cross section, a drill channel for an implant is positioned by the user. (**f**) The implant positions can also be examined in a 3D view. To this end, isosurfaces have been segmented from the CT data to get a representation of the bone surfaces. (**g**) Once the positions have been found, the data is used to produce a personalised drill guide. Pre- and post-surgical comparison reveals the accuracy of the implants (planned position, yellow; actual position, blue). Adapted and reprinted with permission (Vrielinck et al. 2003)

1980s to examine the effects of anatomical features, blood rheology, vessel wall compliance, etc. on blood flow patterns and their relationship with the onset of vascular disease (Steinman and Taylor 2005; Xu and Collins 1990). CFD methods were applied to examine haemodynamic effects of surgical interventions initially in idealised geometries (Steinman and Taylor 2005). Advancement in the field came from application of image-based methodology: solution of fluid-flow equations on the geometries extracted from magnetic resonance imaging (MRI) data (Fig. 7a, b) has demonstrated substantial interindividual variability in computed wall shear stress patterns (Fig. 7c, d) as well as their difference from those computed on idealised geometries (Fig. 7e). In particular spiralling contours of the wall shear stresses and strong anterior-posterior asymmetry of the haemodynamic patterns are observed in models solved on geometries extracted from the MRI data, however, not on the idealised geometry (Fig. 7c–e) (Milner et al. 1998). This (Milner et al. 1998) and other early studies (Steinman and Taylor 2005) highlighted the importance of using "real" individual-specific geometries. The image-based modelling has further advanced from modelling flow and pressure pattern into modelling of surgical interventions (Steinman and Taylor 2005; Charles and Taylor 2010; Morris et al. 2016). CFD modelling is currently adopted for stent and valve and in the design of other cardiovascular devices (Morris et al. 2016; Morlacchi and Migliavacca 2013; Sotiropoulos and Borazjani 2009). Advancement in image-based modelling approach led to its application to modelling of haemodynamic changes induced by stent or valve implantation or other surgical interventions in real patient-specific geometries (de Zélicourt et al. 2006). CFD image-based modelling is predominantly used by medical device developers and academics, however, not in routine clinical practice (Morris et al. 2016). Adaptation of this approach in clinics requires robust, easy to use implementation as well as regulatory approval. HeartFlow FFR_{CT} demonstrates that the approach has sufficiently matured to reach these criteria. Fractional flow reserve (FFR) is a fraction of the maximal blood flow in the supplying coronary artery achieved in the presence of a stenosis. FFR has been used to estimate the likelihood that the stenosis impedes oxygen delivery to the heart muscle and causes myocardial ischemia. FFR serves as a criterion for revascularisation procedure. In clinical practice, calculation of FFR had been requiring invasive coronary angiography (ICA) and had been calculated as a fraction of distal coronary pressure to the proximal aortic pressure measured during sustained hyperaemia (Pijls et al. 1993; De Bruyne et al. 2014; Min et al. 2015). A particular software implementation (HeartFlow FFR_{CT}) of mechanistic image-based modelling approach was acknowledged by the FDA as Type 2 device to estimate FFR non-invasively (FDA 2014). In this approach, FFR is computed by solving Navier-Stokes (Eq. 5) equations in 3D geometry extracted from a CT scan of a patient's coronary arteries (Min et al. 2015).

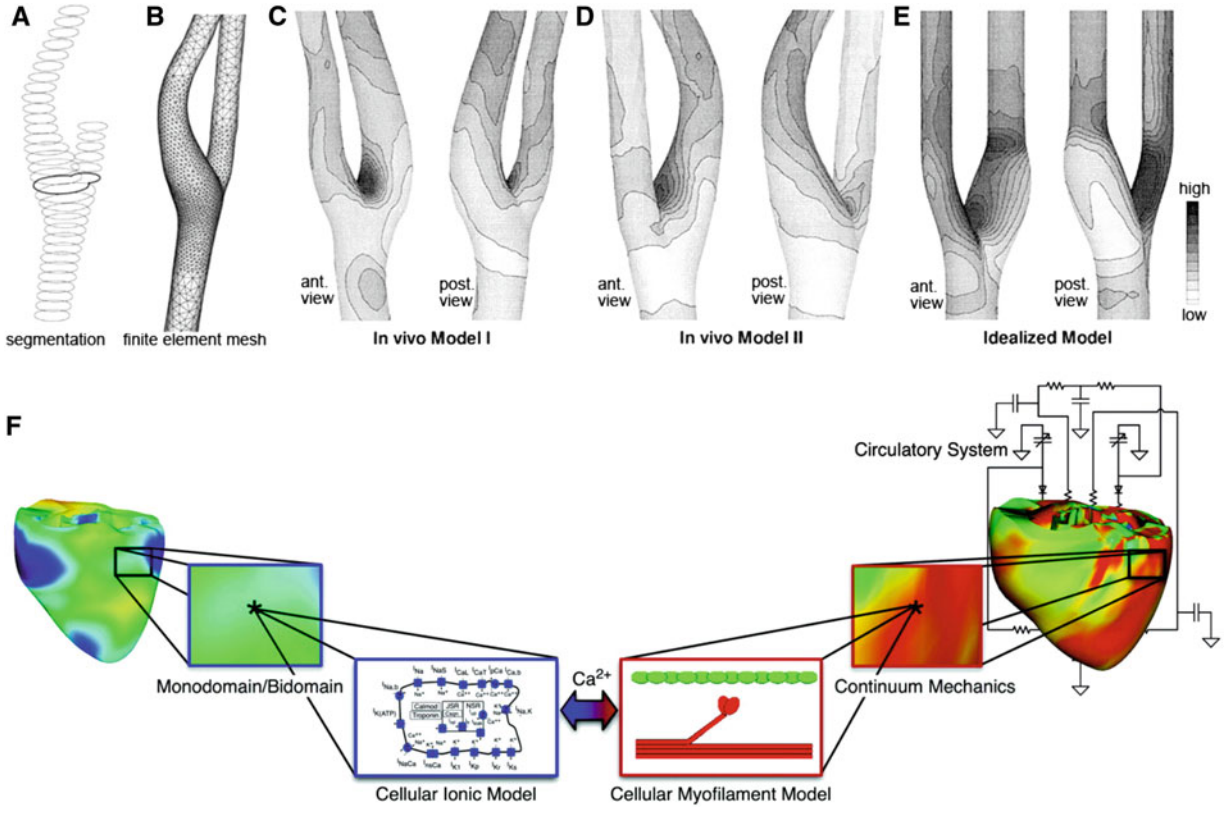


Fig. 7 Biomedical applications of image-based modelling. (**a**–**f**) Haemodynamics of human carotid artery bifurcations. (**a**) The lumen contours were derived from the MR images. (**b**) The surface of the volume finite element mesh. (**c**–**e**) Time-averaged wall shear stress magnitude (normalised to the value at the common carotid artery) for the in vivo I (**c**), in vivo II (**d**) and idealised models (**e**). (**a**–**e**) Reproduced with permission from (Milner et al. 1998). (**f**) Schematic of the approach to modelling cardiac electromechanical function. Reproduced with permission from (Trayanova 2011)

4.3 Image-Based Multi-scale Heart Modelling

In the sections above, we have reviewed applications of mechanistic image-based modelling to describe systems governed mostly by a single law of physics at a time: reaction-diffusion equation for pattern formation, Navier-Stokes for blood flow and solid mechanics equations for bone mechanics. We have already demonstrated above that interaction of several phenomena could take place and have to be accounted for in the model to gain an adequate description of the system, e.g. interaction of genetics and tissue mechanics during plant development, interaction of fluid flow and elastic properties of the vessel wall during pulsatile blood flow. In this section we review whole-heart modelling as an example of truly multiphysics modelling (Noble 2002; Trayanova 2011; Niederer et al. 2019), where a reaction-diffusion equation (Sect. 2.1) describes propagation of electrical and chemical waves of the transmembrane potential. Intracellular calcium released during this wave propagation induces tension generation by the myocyte contraction, what in turn leads to macroscopic contraction of cardiac tissue constrained by the equations describing tissue continuum mechanics (Sect. 2.2). Macroscopic muscle heart tissue contraction ejects blood from a heart chamber with a blood flow described by continuous flow mechanics (Sect. 2.3). As in the previous examples, imaging data contributes here in multiple ways: providing information regarding the shape of the cardiac tissue on which to formulate a computational model, as well as providing information regarding fibre and sheet structure of the tissue, which defines electrical conductivity and passive mechanical properties of the tissue. Additionally, time-lapse imaging data can be used to define boundary conditions in the model (Fig. 7f). With decades of development and refinement at every scale, electrophysiological whole-heart models contributed to understanding of the mechanisms underlying ventricular and atrial arrhythmias, as well as for in silico optimisation of clinical intervention protocols (Niederer et al. 2019). In particular, image-based electrophysiological whole-heart models predicted the optimal target for tissue ablation in patients with scar-related ventricular tachycardia (Ashikaga et al. 2013) and atrial fibrillation (Ruchat et al. 2007), optimal location for implantable defibrillator leads in a paediatric and congenital heart disease patients (Rantner et al. 2013), as well as optimal left ventricle lead location in heart failure patients undergoing cardiac resynchronisation therapy (Lee et al. 2017).

Applications of image-based modelling in biomedical research is not limited to the organs and systems reviewed above, as mechanistic image-based modelling methodology was applied to modelling of blood clotting (Kadri et al. 2019; Voronov et al. 2013), skin biophysics (Limbert 2017) and other systems (Neal and Kerckhoffs 2010; Tavares and Jorge 2012; Viceconti and Hunter 2016).

5 Image-Based Modelling in Drug Discovery and Development

The pharmaceutical industry is tackling increasingly complex diseases, which has significantly contributed to the increase of R&D spending in large pharmaceutical companies during the past decade (Arrowsmith 2012). Despite increasing costs, the success rate for Phase II clinical trials is ≈30%, and the estimate for overall success rate for progression through clinical development is as low as ≈10% (Smietana et al. 2016). However, the years 2017 and 2018 showed a spike in a number of new drugs approved by the FDA (Mullard 2018). The main reasons for drug candidate failure during Phase II and III trials are clinical toxicity and lack of efficacy, with complex diseases like cancer, neurodegenerative and cardiovascular disorders exhibiting the highest attrition numbers (Arrowsmith 2011). Model-informed drug discovery and development (MID3) is becoming a routine practice in the pharmaceutical industry and is believed to be associated with increased R&D productivity (Morgan et al. 2012, 2018; Visser et al. 2014; Lippert et al. 2016; Topp et al. 2019; Marshall et al. 2019). In this section we review selected applications of mechanistic image-based modelling approaches in drug discovery and development.

5.1 ADME Applications

To exert the desired pharmacological action, a drug needs to be delivered into a relevant tissue at sufficient concentration for a required time interval. Drug absorption, distribution, metabolism and excretion (ADME) studies are conducted to address these challenges. Physiologically based pharmacokinetic modelling (PBPK) is a mathematical and computational modelling approach to interpret ADME studies, perform cross-species and cross-population translation, predict drug pharmacokinetics, etc. (Eissing 2011). PBPK models are multi-compartmental models, where compartments are defined a priori and represent known anatomical and physiological structures of the body. This formalism assumes an absence of gradients within a computational compartment; however, this is not valid in all cases. Below we will review several examples of image-based modelling applied to gain insights into ADME processes as well as its integration with established PBPK framework.

The liver is the central organ for detoxification of xenobiotics in the body. In pharmacokinetic modelling, hepatic metabolisation capacity is typically quantified as hepatic clearance computed assuming a first-order degradation process in a well-stirred compartment. However, assumptions of the well-stirred model might be violated during the first instance after drug injection. To simulate the first pass perfusion and clearance, an image-based model for liver vasculature was developed (Fig. 8a) and coupled with a PBPK model (Schwen et al. 2014). Molecular concentration in plasma vs. time (c-t) profiles simulated for several test compounds with an image-based liver model exhibited a temporal delay and more smeared-out peak as compared to profiles simulated with well-stirred model. These effects in c-t profiles

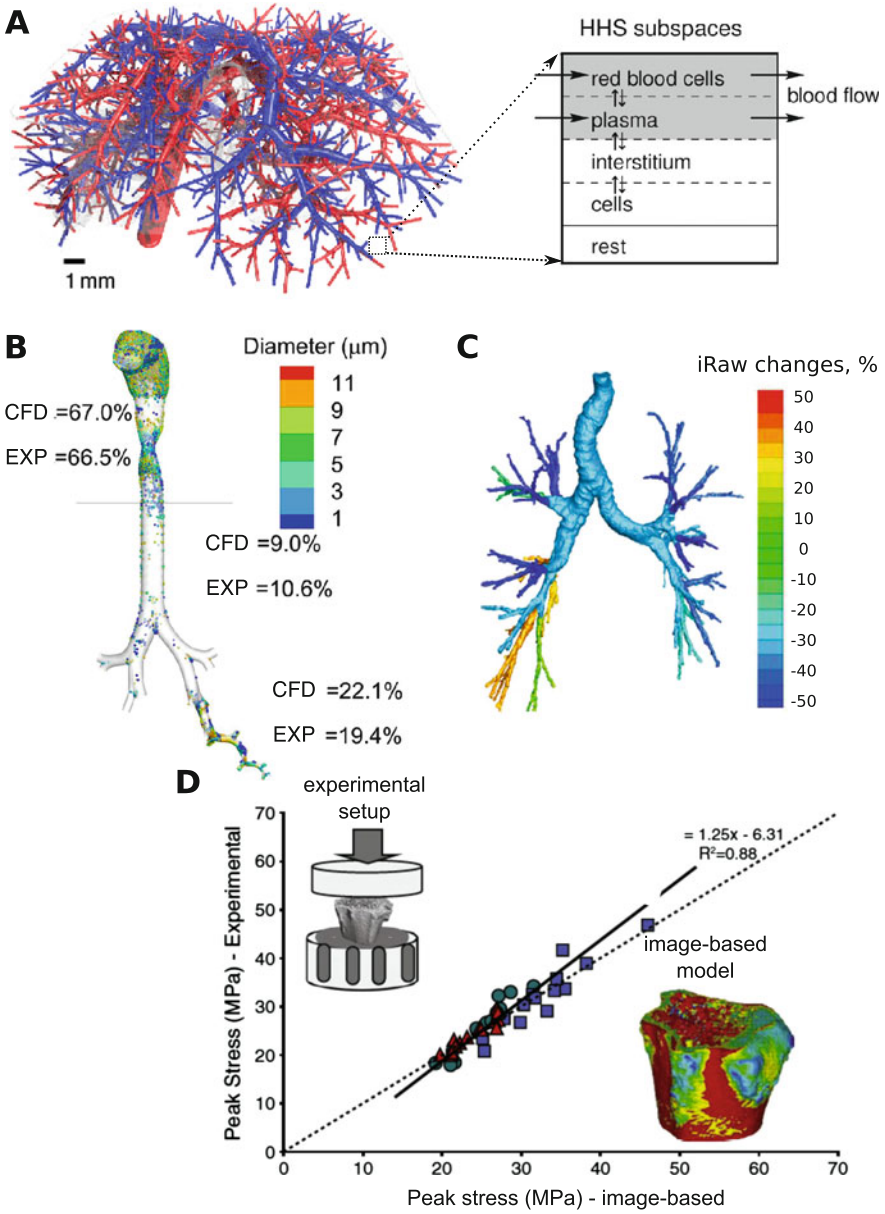


Fig. 8 Image-based models on drug discovery and development. (**a**) Schematic of the liver model. 3D rendering of liver vasculature: the supplying (red) and draining (blue) vascular systems. A homogenised hepatic space (HHS) consists of several subspaces compatible with PBPK modelling, adapted from (Schwen et al. 2014); (**b**) Comparison of in vivo and calculated predictions of deposition fraction (DF) in different regions of the airways for the Respimat, adapted from (Tian et al. 2015); (**c**) Change in image-based calculated airway resistance (iRaw) in a COPD patient responding to treatment with roflumilast, reproduced with permission of the © ERS 2019 (de Backer et al. 2014). (**d**) Comparison between the peak stress experimentally obtained and the estimated

are due to the geometry driven variety of transition times required for the substance to flow from the supplying to the draining vasculature. The model was further applied to simulate the impact of pathophysiological states and chemical injury of the liver on c-t profiles of the test compounds (Schwen et al. 2014). Methodologically similar image-based models of the liver were also developed to simulate biliary fluid dynamics (Meyer et al. 2017), liver regeneration after chemical damage (Hoehme et al. 2010) and to develop a strategy for pharmacological interventions in liver diseases (Ghallab et al. 2016). Adequate drug uptake by the target tissue is crucial for the treatment efficacy, in particular in oncology, as cancers exhibit substantial spatial heterogeneity and drug uptake varies significantly by the tumour type and size. Overall factors defining tumour drug uptake are understood (Baxter and Jain 1989, 1990, 1991); however, uptake of a particular drug with given physico-chemical and biochemical properties by a given tumour is challenging to forecast. A mechanistic image-based modelling approach similar to that applied to model interplay of perfusion and distribution in the liver (Schwen et al. 2014) (Fig. 8a) was applied to model whole tumour perfusion and drug uptake (D'Esposito et al. 2018; Sweeney et al. 2019). An image-based model was established based on ex vivo optical projection tomography data for colorectal tumour xenografts and validated with in vivo imaging data obtained in the same tumour prior to dissection. Image-based models were shown to adequately predict vascular perfusion and drug delivery into the tumours of different heterogeneity (D'Esposito et al. 2018; Sweeney et al. 2019). Image-based models also find application in modelling growth and response to chemotherapy of brain and other tumours (Baldock et al. 2013; Bhandari et al. 2018; Karolak et al. 2018). Further pharmacodynamic applications of image-based modelling are reviewed below (Sect. 5.2).

Inhalation is an important route of administration for pulmonary drugs enabling selective drug action in the lung. Particle deposition in the lung is also crucial for characterising hazards of exposure to toxic substances. Due to the complex geometry of the lung image-based modelling is well suited to asses particle deposition patterns in the lung and their dependence on aerosol properties, inhalation regimens, etc. (Lambert et al. 2011; Longest and Holbrook 2012). Direct comparison of experimentally determined deposition fractions with those calculated in realistic geometries extracted from imaging data indicate a good agreement (relative error $<10\%$) between measured and simulated deposition fractions across particle sizes and inhalation regiments (Fig. 8b) (Tian et al. 2015). Similar to the image-based liver model (Schwen et al. 2014), an image-based lung model can be integrated with a whole-body PBPK model to simulate exposure in systemic regions (Haghnegahdar et al. 2019). Other examples of image-based modelling applications to address ADME-related questions include drug distribution within the cerebrospinal fluid (Kuttler et al. 2010) and all the way up to the whole-body

Fig. 8 (continued) stress with image-based modelling. The black line represents the linear regression of the data, and the dotted line corresponds to the line of identity. Reproduced with permission from (Mathers et al. 2013)

image-based modelling (Viceconti and Hunter 2016; Hunter and Borg 2003; Neufeld et al. 2013; Brolin et al. 2015).

5.2 Towards Pharmacodynamic Applications

Image-based modelling in drug development and discovery can be applied not only to address ADME-related questions but also to asses exerted pharmacodynamic effects. Figure 8c depicts reduction in computed airway resistance (iRaw) based on CT data and fluid dynamics model (Sect. 2.3) in a chronic obstructive pulmonary disease (COPD) patient responding to treatment with roflumilast (de Backer et al. 2014). Image-based modelling-derived iRaw exhibits spatial variation of pharmacodynamic effect within a lung, an information not accessible with a typical functional lung test. Therefore image-based modelling end points, if sufficiently validated, could serve as surrogate biomarkers in clinical trials. Additional clinical studies confirmed that image-based modelling end points correlate with lung functional tests yet are more sensitive (de Backer et al. 2012, 2015). A placebo-controlled study of glycopyrrolate/formoterol fumarate in moderate-to-severe COPD patients demonstrated that treatment induced 71% reduction in image-based derived airway resistance iRaw, as compared to a placebo group (de Backer et al. 2018). Changes in iRaw were accompanied by improvement in lung functional tests. Another example of the use of image-based modelling end points as a surrogate clinical biomarker comes from odanacatib development for prevention of postmenopausal osteoporosis (Visser et al. 2014). An image-based modelling approach for bone stress calculation (see Sect. 2.2 for details) was validated against ex vivo measured values in non-human primates (Fig. 8d) (Mathers et al. 2013). An image-based modelling-derived biomarker was demonstrated to be more sensitive than a direct imaging read-out, and it was included into odanacatib clinical trials as an exploratory end point. Odanacatib placebo-controlled clinical trials indicated an increase in estimated bone strength in postmenopausal women after a year of treatment (Visser et al. 2014; Brixen et al. 2013). Methodologically similar image-based modelling approaches were included into other clinical studies for osteoporosis treatment, e.g. idoxifene's effect on calcaneal bone mechanical parameters (Rietbergen 2002) and teriparatide and alendronate effects on vertebral strength (Keaveny et al. 2007).

Even though to date most of the applications of the whole-heart image-based modelling concern questions arising in basic research and surgical interventions (Sect. 4.3), applications related to drug discovery and development are emerging. Image-based heart modelling was applied to assess drug-induced arrhythmia (Okada et al. 2015, 2018). In this approach exposure-response relationships were obtained in in vitro experiments to assess ion channel blocking properties of the test compounds and were transferred into the whole-heart model, coupled to the human torso model to simulate electrocardiogram at increasing concentration of the test compounds. Simulations correctly recapitulated concentration-dependent characteristic types of ventricular arrhythmia. The advantage of the whole-heart models over simpler cell and/or tissue electrophysiological models is in that they explicitly consider interplay between cell electrophysiology and feedback originating from the travelling

electrochemical depolarisation wave (Fig. 7f) (Okada et al. 2015). Application of whole-heart image-based modelling for drug discovery and development requires quantification and consideration of uncertainty and variability in model prediction (Ni et al. 2018). The methods for accounting for naturally occurring patient to patient variability in models without explicit spatial dimension are well established (Ette and Williams 2007; Owen and Fiedler-Kelly 2014) and could be applied to the electrophysiological part of the whole-heart model; however, methodology for incorporating variability concerning geometries and other spatially distributed model properties (e.g. tissue conductivity) are being developed (Ni et al. 2018).

6 Outlook

Above we have reviewed applications of the mechanistic image-based modelling in areas ranging from developmental biology and biomedical research to diagnostics, medical device and drug development. With varying timing across these application areas, image-based modelling is emerging from research methodology available only to a few computational experts to becoming a tool which could be integrated into the daily practice of pharmacologists, medical doctors, biomarker experts, etc. This evolution is facilitated by the advancement of methods at every step of the image-based modelling pipeline (Fig. 2) as well as by increased availability of imaging data of all modalities. "Industrialisation" of image-based modelling requires increased robustness and automation of all steps, in particular, perhaps the automation of the image segmentation step, which in early applications was often performed manually or semi-manually. Recent progress in deep learning, including the development of convolutional neural networks (Ronneberger et al. 2015; Cicek et al. 2016) could enable robust image segmentation solutions to be included into an image-based modelling pipeline.

Applications of image-based modelling in drug discovery and development as reviewed above are based on imaging data from animal models and patients. In the past decade microphysiological systems, also known as organ-on-a-chip, underwent rapid development with examples of their applications to address ADME-, pharmacology- and toxicity-related questions (Bhatia and Ingber 2014; Taylor et al. 2019). Among other benefits, microphysiological systems offer the following advantages over conventional 2D cultures: 3D tissue organisation comprising several cell types and the presence of controllable mechanical and molecular cues. Data generated in these microphysiological systems could be combined with computational modelling to translate from micro to macro scale and, if necessary, to compensate for potentially missing biochemical and physiological feedback loops. Microphysiological systems are typically set in transparent housing that enables monitoring of system response with optical microscopy (Wenzel et al. 2014, 2015; Menshykau 2017; Peel

et al. 2019). Modelling based on imaging data acquired in microphysiological systems holds promise for facilitating translation from lab to clinic.

References

Alper B, Gultekin P, Yalci S (2012) Application of finite element analysis in implant dentistry. Finite Elem Anal – New Trends Dev 85:585–598. http://www.intechopen.com/books/finite-element-analysis-new-trends-and-developments/application-of-finite-element-analysis-in-implant-dentistry

Arrowsmith J (2011) Trial watch: phase II failures: 2008–2010. Nat Rev Drug Discov 10:328–329. https://doi.org/10.1038/nrd3439

Arrowsmith J (2012) A decade of change. Nat Rev Drug Discov 11:17–18. https://doi.org/10.1038/nrd3630

Ashikaga H et al (2013) Feasibility of image-based simulation to estimate ablation target in human ventricular arrhythmia. Heart Rhythm 10:1109–1116. https://doi.org/10.1371/journal.pone.0178059

Baldock AL et al (2013) From patient-specific mathematical neuro-oncology to precision medicine. Front Oncol 3:1–11. https://www.ncbi.nlm.nih.gov/pmc/articles/PMC3613895/

Baxter LT, Jain RK (1989) Transport of fluid and macromolecules in tumors. I. Role of interstitial pressure and convection. Microvasc Res 37:77–104. https://linkinghub.elsevier.com/retrieve/pii/0026286289900745

Baxter LT, Jain RK (1990) Transport of fluid and macromolecules in tumors. II. Role of heterogeneous perfusion and lymphatics. Microvasc Res 40:246–263. https://linkinghub.elsevier.com/retrieve/pii/002628629090023K

Baxter LT, Jain RK (1991) Transport of fluid and macromolecules in tumors. Microvasc Res 41:5–23. https://linkinghub.elsevier.com/retrieve/pii/002628629190003T

Bellusci S et al (1997a) Involvement of sonic hedgehog (shh) in mouse embryonic lung growth and morphogenesis. Development 124:53–63. https://www.ncbi.nlm.nih.gov/pubmed/9006067

Bellusci S, Grindley J, Emoto H, Itoh N, Hogan B (1997b) Fibroblast growth factor 10(fgf10) and branching morphogenesis in the embryonic mouse lung. Development 124:4867–4878. https://www.ncbi.nlm.nih.gov/pubmed/9428423

Beuer F, Schweiger J, Edelhoff D (2008) Digital dentistry: an overview of recent developments for CAD/CAM generated restorations. Br Dent J 204:505–511. https://www.ncbi.nlm.nih.gov/pubmed/18469768

Bhandari A, Bansal A, Singh A, Sinha N (2018) DCE-MRI based voxelized computational model for chemotherapeutic drug transport in human brain tumor. Int J Adv Eng Sci Appl Math 10:252–262. https://www.ncbi.nlm.nih.gov/pubmed/28623038

Bhatia SN, Ingber DE (2014) Microfluidic organs-on-chips. Nat Biotechnol 32:760–772. https://www.ncbi.nlm.nih.gov/pubmed/25093883

Blanc P et al (2012) A role for mesenchyme dynamics in mouse lung branching morphogenesis. PLoS One 7:e41643. https://www.ncbi.nlm.nih.gov/pubmed/22844507

Boehm B et al (2010) The role of spatially controlled cell proliferation in limb bud morphogenesis. PLoS Biol 8:1–21. https://doi.org/10.1371/journal.pbio.1000420.

Brixen K et al (2013) Bone density, turnover, and estimated strength in postmenopausal women treated with odanacatib: a randomized trial. J Clin Endocrinol Metab 98:571–580. https://www.ncbi.nlm.nih.gov/pubmed/23337728

Brolin G, Gustafsson J, Ljungberg M, Gleisner KS (2015) Pharmacokinetic digital phantoms for accuracy assessment of image-based dosimetry in 177Lu-DOTATATE peptide receptor radionuclide therapy. Phys Med Biol 60:6131–6149. http://stacks.iop.org/0031-9155/60/i=15/a=6131?key=crossref.9fb9097c0c4d2a17e58252a9272be52d

Chang Y, Tambe AA, Maeda Y, Wada M, Gonda T (2018) Finite element analysis of dental implants with validation: to what extent can we expect the model to predict biological phenomena? A literature review and proposal for classification of a validation process. Int J Implant Dent 4:7. https://www.ncbi.nlm.nih.gov/pubmed/29516219

Charles A, Taylor DAS (2010) Image-based modeling of blood flow and vessel wall dynamics: applications, methods and future directions. Ann Biomed Eng 38:1188–1203. https://www.ncbi.nlm.nih.gov/pubmed/20087775

Chen B-C et al (2014) Lattice light-sheet microscopy: imaging molecules to embryos at high spatiotemporal resolution. Science 346:1257998. http://www.sciencemag.org/lookup/doi/10.1126/science.1257998

Cicek O, Abdulkadir A, Lienkamp SS, Brox T, Ronneberger O (2016) 3D U-net: learning dense volumetric. Med Image Comput Comput Interv – MICCAI 2016:424–432. http://arxiv.org/abs/1606.06650

Clément R, Blanc P, Mauroy B, Sapin V, Douady S (2012a) Shape self-regulation in early lung morphogenesis. PLoS One 7:e36925. https://www.ncbi.nlm.nih.gov/pmc/articles/PMC3353953/

Clément R, Douady S, Mauroy B (2012b) Branching geometry induced by lung self-regulated growth. Phys Biol 9:066006. https://www.ncbi.nlm.nih.gov/pubmed/23160420

Coen E, Rebocho AB (2016) Commentary resolving conflicts: modeling genetic control of plant morphogenesis. Dev Cell 38:579–583. https://doi.org/10.1016/j.devcel.2016.09.006.

Coen E, Kennaway R, Whitewoods C (2017) On genes and form. Development:4203–4213. https://www.ncbi.nlm.nih.gov/pubmed/29183934

Dentsply Sirona (2019) inLab check user manual. https://manuals.sirona.com/home.HomeDmsDocument.download.html?id=57329

D'Esposito A et al (2018) Computational fluid dynamics with imaging of cleared tissue and of in vivo perfusion predicts drug uptake and treatment responses in tumours. Nat Biomed Eng 2:773–787. https://www.ncbi.nlm.nih.gov/pubmed/3101564

de Backer LA et al (2012) The acute effect of budesonide/formoterol in COPD: a multi-slice computed tomography and lung function study. Eur Respir J 40:298–305. https://www.ncbi.nlm.nih.gov/pubmed/22183484

de Backer W et al (2014) The effect of roflumilast in addition to LABA/LAMA/ICS treatment in COPD patients. Eur Respir J 44:527–529. http://erj.ersjournals.com/cgi/doi/10.1183/09031936.00011714

de Backer J et al (2015) The effects of extrafine beclometasone/formoterol (BDP/F) on lung function, dyspnea, hyperinflation, and airway geometry in COPD patients: novel insight using functional respiratory imaging. J Aerosol Med Pulm Drug Deliv 28:88–99. http://www.liebertpub.com/doi/10.1089/jamp.2013.1064

de Backer W et al (2018) A randomized study using functional respiratory imaging to characterize bronchodilator effects of glycopyrrolate/formoterol fumarate delivered by a metered dose inhaler using co-suspension delivery technology in patients with COPD. Int J COPD 13:2673–2684. https://www.ncbi.nlm.nih.gov/pubmed/30214185

De Bruyne B et al (2014) Fractional flow reserve–guided pci for stable coronary artery disease. N Engl J Med 371:1208–1217. https://doi.org/10.1056/NEJMoa1408758.

de Zélicourt DA et al (2006) Flow study of an extracardiac connection with persistent left superior vena cava. J Thorac Cardiovasc Surg 131:785–791. https://linkinghub.elsevier.com/retrieve/pii/S0022522305020015

Eissing TA (2011) Computational systems biology software platform for multiscale modeling and simulation: integrating whole-body physiology, disease biology, and molecular reaction networks. Front Physiol 2:4. http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=3070480&tool=pmcentrez&rendertype=abstract

Ette EI, Williams PJ (eds) (2007) Pharmacometrics. Wiley, Hoboken. https://doi.org/10.1002/0470087978

FDA (2014) https://www.accessdata.fda.gov/cdrh_docs/reviews/DEN130045.pdf

Geng J-P, Tan KB, Liu G-R (2001) Application of finite element analysis in implant dentistry: a review of the literature. J Prosthet Dent 85:585–598. https://linkinghub.elsevier.com/retrieve/pii/S0022391301574314

Ghallab A et al (2016) Model-guided identification of a therapeutic strategy to reduce hyperammonemia in liver diseases. J Hepatol 64:860–871. https://www.ncbi.nlm.nih.gov/pubmed/26639393

Gleghorn JP, Kwak J, Pavlovich AL, Nelson CM (2012) Inhibitory morphogens and monopodial branching of the embryonic chicken lung. Dev Dyn 241:852–862. https://www.ncbi.nlm.nih.gov/pubmed/22410853

Green AA, Kennaway JR, Hanna AI, Bangham JA, Coen E (2010) Genetic control of organ shape and tissue polarity. PLoS Biol 8:e1000537. http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=2976718&tool=pmcentrez&rendertype=abstract

Haghnegahdar A, Zhao J, Kozak M, Williamson P, Feng Y (2019) Development of a hybrid CFD-PBPK model to predict the transport of xenon gas around a human respiratory system to systemic regions. Heliyon 5:e01461. https://doi.org/10.1016/j.heliyon.2019.e01461

Hidalgo DA (1989) Fibula free flap: a new method of mandible reconstruction. Plast Reconstr Surg 84:71–79. http://www.ncbi.nlm.nih.gov/pubmed/2734406

Hiscock TW, Megason SG (2015) Mathematically guided approaches to distinguish models of periodic patterning. Development:409–419. https://www.ncbi.nlm.nih.gov/pubmed/25605777

Hoehme S et al (2010) Prediction and validation of cell alignment along microvessels as order principle to restore tissue architecture in liver regeneration. Proc Natl Acad Sci 107:10371–10376. http://www.pnas.org/content/107/23/10371.full.pdf+html?with-ds=yes%5Cn

Hunter PJ, Borg TK (2003) Integration from proteins to organs: the Physiome project. Nat Rev Mol Cell Biol 4:237–243. http://www.ncbi.nlm.nih.gov.ezproxy.lib.monash.edu.au/pubmed/12612642

Iber D, Menshykau D (2013) The control of branching morphogenesis. Open Biol 3:130088. https://www.ncbi.nlm.nih.gov/pubmed/24004663

Okada JI et al (2018) Arrhythmic hazard map for a 3D whole-ventricle model under multiple ion channel block. Br J Pharmacol 175:3435–3452. https://www.ncbi.nlm.nih.gov/pubmed/29745425

Kadri OE, Chandran VD, Surblyte M, Voronov RS (2019) In vivo measurement of blood clot mechanics from computational fluid dynamics based on intravital microscopy images. Comput Biol Med 106:1–11. https://doi.org/10.1016/j.compbiomed.2019.01.001

Karolak A, Markov DA, Mccawley LJ, Rejniak KA (2018) Towards personalized computational oncology: from spatial models of tumour spheroids, to organoids, to tissues. J R Soc Interface 15:20170703. https://www.ncbi.nlm.nih.gov/pubmed/29367239

Keaveny TM et al (2007) Effects of teriparatide and alendronate on vertebral strength as assessed by finite element modeling of QCT scans in women with osteoporosis. J Bone Miner Res 22:149–157. https://doi.org/10.1359/jbmr.061011

Keener J, Sneyd J (2009) Mathematical physiology. Volume II: systems physiology. Springer, Berlin. https://doi.org/10.1007/978-0-387-79388-7

Khan SBRBBH (2012) Current therapy in oral and maxillofacial surgery. Elsevier, Amsterdam. https://www.us.elsevierhealth.com/current-therapy-in-oral-and-maxillofacial-surgery-9781416025276.html

Kherlopian AR et al (2008) A review of imaging techniques for systems biology. BMC Syst Biol 2:1–18. https://www.ncbi.nlm.nih.gov/pubmed/18700030

Kicheva A et al (2007) Kinetics of morphogen gradient formation. Science 315:521–525. https://www.ncbi.nlm.nih.gov/pubmed/17255514

Kurics T, Menshykau D, Iber D (2014) Feedback, receptor clustering, and receptor restriction to single cells yield large Turing spaces for ligand-receptor-based Turing models. Phys Rev E 90:022716. https://www.ncbi.nlm.nih.gov/pubmed/25215767

Kuttler A et al (2010) Understanding pharmacokinetics using realistic computational models of fluid dynamics: biosimulation of drug distribution within the CSF space for intrathecal drugs. J Pharmacokinet Pharmacodyn 37:629–644. https://www.ncbi.nlm.nih.gov/pubmed/21132572

Lambert AR, O'Shaughnessy PT, Tawhai MH, Hoffman EA, Lin CL (2011) Regional deposition of particles in an image-based airway model: large-eddy simulation and left-right lung ventilation asymmetry. Aerosol Sci Tech 45:11–25. http://www.tandfonline.com/doi/abs/10.1080/02786826.2010.517578

Lee AW et al (2017) Biophysical modeling to determine the optimization of left ventricular pacing site and AV/VV delays in the acute and chronic phase of cardiac resynchronization therapy. J Cardiovasc Electrophysiol 28:208–215. https://www.ncbi.nlm.nih.gov/pubmed/27885749

Leondes CT (2003) Automated three dimensional finite element modeling techniques of bone and bone implant systems. In: Computational methods in biophysics, biomaterials, biotechnology and medical systems. Springer, Boston, pp 378–415. https://doi.org/10.1007/0-306-48329-7_10

Limbert G (2017) Mathematical and computational modelling of skin biophysics: a review. Proc R Soc A Math Phys Eng Sci 473:20170257. https://doi.org/10.1098/rspa.2017.0257

Lippert J et al (2016) Modeling and simulation of in vivo drug effects. Springer, Cham, pp 313–329. https://doi.org/10.1007/164_2015_21

Longacre JJ, Destefano GA (1957) Further observations of the behavior of autogenous split-rib grafts in reconstruction of extensive defects of the cranium and face. Plast Reconstr Surg (1946) 20:281–296. http://www.ncbi.nlm.nih.gov/pubmed/13494206

Longest PW, Holbrook LT (2012) In silico models of aerosol delivery to the respiratory tract – development and applications. Adv Drug Deliv Rev 64:296–311. https://linkinghub.elsevier.com/retrieve/pii/S0169409X11001244

Louis PJ (2011) Bone grafting the mandible. Dent Clin N Am 55:673–695. https://linkinghub.elsevier.com/retrieve/pii/S0011853211001121

Marshall S et al (2019) Model-informed drug discovery and development: current industry good practice and regulatory expectations and future perspectives. CPT Pharmacometrics Syst Pharmacol 8:87–96. https://www.ncbi.nlm.nih.gov/pubmed/30411538

Mathers PD et al (2013) High-resolution peripheral quantitative computed tomography and finite element analysis of bone strength at the distal radius in ovariectomized adult rhesus monkey demonstrate efficacy of odanacatib and differentiation from alendronate. Bone 56:497–505. http://www.ncbi.nlm.nih.gov/pubmed/23791777%5Cn

Menshykau D (2017) Emerging technologies for prediction of drug candidate efficacy in the preclinical pipeline. Drug Discov Today 22:1598–1603. http://www.sciencedirect.com/science/article/pii/S1359644617302386

Menshykau D, Iber D (2013) Kidney branching morphogenesis under the control of a ligand-receptor based Turing mechanism. Phys Biol 10:046003. https://www.ncbi.nlm.nih.gov/pubmed/23770927

Menshykau D, Kraemer C, Iber D (2012) Branch mode selection during early lung development. PLoS Comp Biol 8:e1002377. https://www.ncbi.nlm.nih.gov/pmc/articles/PMC3280966/

Menshykau D, Blanc P, Unal E, Vincent S, Iber D (2014) An interplay of geometry and signaling enables robust branching morphogenesis. Development 141:4526–4536. https://www.ncbi.nlm.nih.gov/pubmed/25359721

Menshykau D et al (2019) Image-based modeling of kidney branching morphogenesis reveals GDNF-RET based Turing-type mechanism and pattern-modulating WNT11 feedback. Nat Commun 10:239. https://doi.org/10.1038/s41467-018-08212-8

Meyer K et al (2017) A predictive 3D multi-scale model of biliary fluid dynamics in the liver lobule. Cell Syst 4:277–290.e9. https://doi.org/10.1016/j.cels.2017.02.008

Milner JS, Moore JA, Rutt BK, Steinman DA (1998) Hemodynamics of human carotid artery bifurcations: computational studies with models reconstructed from magnetic resonance imaging of normal subjects. J Vasc Surg 28:143–156. http://www.sciencedirect.com/science/article/pii/S0741521498702101

Min JK et al (2015) Noninvasive fractional flow reserve derived from coronary ct angiography: clinical data and scientific principles. JACC Cardiovasc Imaging 8:1209–1222. http://www.sciencedirect.com/science/article/pii/S1936878X15006038

Morgan P et al (2012) Can the flow of medicines be improved? Fundamental pharmacokinetic and pharmacological principles toward improving phase II survival. Drug Discov Today 17:419–424. https://doi.org/10.1016/j.drudis.2011.12.020.

Morgan P et al (2018) Impact of a five-dimensional framework on R&D productivity at AstraZeneca. Nat Rev Drug Discov 17:167–181. https://doi.org/10.1038/nrd.2017.244

Morlacchi S, Migliavacca F (2013) Modeling stented coronary arteries: where we are, where to go. Ann Biomed Eng 41:1428–1444. https://www.ncbi.nlm.nih.gov/pubmed/23090621

Morris PD et al (2016) Computational fluid dynamics modelling in cardiovascular medicine. Heart 102:18–28. https://www.ncbi.nlm.nih.gov/pubmed/26512019

Mullard A (2019) FDA drug approvals. Nat Rev Drug Discov 18:85–89. https://doi.org/10.1038/nrd.2018.4

Murray J (2003) Mathematical biology. In: Volume II: spatial models and biomedical applications, vol 2, 3rd edn. Springer, Berlin. https://www.springer.com/gp/book/9780387952284

Myers G (2012) Why bioimage informatics matters. Nat Methods 9:659–660. https://www.nature.com/articles/nmeth.2024

Neal ML, Kerckhoffs R (2010) Current progress in patient-specific modeling. Brief Bioinform 11:111–126. https://doi.org/10.1093/bib/bbp049

Nelson C, Vanduijn M, Inman J, Fletcher D (2006) Tissue geometry determines sites of mammary branching morphogenesis in organotypic cultures. Science 314:298–300. https://www.ncbi.nlm.nih.gov/pmc/articles/PMC2933179/

Neufeld E, Szczerba D, Chavannes N, Kuster N (2013) A novel medical image data-based multi-physics simulation platform for computational life sciences. Interface Focus 3:1–6. https://www.ncbi.nlm.nih.gov/pmc/articles/PMC3638470/

Neugebauer J et al (2010) Computer-aided manufacturing technologies for guided implant placement. Expert Rev Med Devices 7:113–129. https://www.ncbi.nlm.nih.gov/pubmed/20021243

Ni H, Morotti S, Grandi E (2018) A heart for diversity: simulating variability in cardiac arrhythmia research. Front Physiol 9:1–19. https://www.ncbi.nlm.nih.gov/pmc/articles/PMC6062641/

Niederer SA, Lumens J, Trayanova NA (2019) Comput Models Cardiol. https://doi.org/10.1038/s41569-018-0104-y

Noble D (2002) Modeling the heart – from genes to cells to the whole organ. Science 295:1678–1682. https://doi.org/10.1126/science.1069881

Okada JI et al (2015) Screening system for drug-induced arrhythmogenic risk combining a patch clamp and heart simulator. Sci Adv 1:1–8. https://www.ncbi.nlm.nih.gov/pubmed/26601174

Owen JS, Fiedler-Kelly J (2014) Introduction to population pharmacokinetic/pharmacodynamic analysis with nonlinear mixed effects models. Wiley, Hoboken. https://doi.org/10.1002/9781118784860

Park YS, Kwon HB (2013) Three-dimensional finite element analysis of implant-supported crown in fibula bone model. J Adv Prosthodont 5:326–332. https://www.ncbi.nlm.nih.gov/pmc/articles/PMC3774948/

Peel S et al (2019) Introducing an automated high content confocal imaging approach for organs-on-chips. Lab Chip 19:410–421. https://doi.org/10.1039/C8LC00829A

Pijls NHJ, Son JAMV, Kirkeeide RL, Bruyne BD, Gould KL (1993) Experimental basis of determining maximum coronary, myocardial, and collateral blood flow by pressure measurements for assessing functional stenosis Severit before and after percutaneous transluminal coronary angioplasty. Circulation:1354–1367. https://www.ncbi.nlm.nih.gov/pubmed/8462157

Prendergast P (1997) Finite element models in tissue mechanics and orthopaedic implant design. Clin Biomech 12:343–366. https://linkinghub.elsevier.com/retrieve/pii/S0268003397000181

Rantner LJ, Vadakkumpadan F, Spevak PJ, Crosson JE, Trayanova NA (2013) Placement of implantable cardioverter-defibrillators in paediatric and congenital heart defect patients: a pipeline for model generation and simulation prediction of optimal configurations. J Physiol 591:4321–4334. https://www.ncbi.nlm.nih.gov/pubmed/23798492

Rietbergen BV (2002) High-resolution MRI and micro-FE for the evaluation of changes in bone mechanical properties during longitudinal clinical trials: application to calcaneal bone in postmenopausal women after one year of idoxifene treatment. Clin Biomech 17:81–88

Ronneberger O, Fischer P, Brox T (2015) U-net: convolutional networks for biomedical image segmentation. Lect Notes Comput Sci 9351:234–241. https://www.ncbi.nlm.nih.gov/pubmed/11832257

Ruchat P et al (2007) A biophysical model of atrial fibrillation to define the appropriate ablation pattern in modified maze. Eur J Cardio-Thoracic Surg 31:65–69. https://doi.org/10.1016/j.ejcts.2006.10.015

Sakkas A, Wilde F, Heufelder M, Winter K, Schramm A (2017) Autogenous bone grafts in oral implantology-is it still a "gold standard"? A consecutive review of 279 patients with 456 clinical procedures. Int J Implant Dentistry 3:23. http://journalimplantdent.springeropen.com/articles/10.1186/s40729-017-0084-4

Sbalzarini IF (2013) Modeling and simulation of biological systems from image data. Bioessays 35:482–490. https://www.ncbi.nlm.nih.gov/pubmed/23533152

Schwen LO et al (2014) Spatio-temporal simulation of first pass drug perfusion in the liver. PLoS Comput Biol 10:e1003499. https://www.ncbi.nlm.nih.gov/pubmed/24625393

Scott G (2013) Developmental biology, 10th edn. Sinauer Associates, Sunderland

Sharpe J (2017) Computer modeling in developmental biology: growing today, essential tomorrow. Development 144:4214–4225. https://www.ncbi.nlm.nih.gov/pubmed/29183935

Silva Vieira M, Hussain T, Alberto Figueroa C (2015) Patient-specific image-based computational modeling in congenital heart disease: a clinician perspective. J Cardiol Ther 2:436–448. http://www.ghrnet.org/index.php/jct/article/view/1512

Smietana K, Siatkowski M, Møller M (2016) Trends in clinical success rates. Nat Rev Drug Discov 15:379–380. https://doi.org/10.1038/nrd.2016.85

Sotiropoulos F, Borazjani I (2009) A review of state-of-the-art numerical methods for simulating flow through mechanical heart valves. Med Biol Eng Comput 47:245–256. https://www.ncbi.nlm.nih.gov/pmc/articles/PMC2717171/

Steinman DA, Taylor CA (2005) Flow imaging and computing: large artery hemodynamics. Ann Biomed Eng 33:1704–1709. https://doi.org/10.1007/s10439-005-8772-2.

Sweeney PW, D'Esposito A, Walker-Samuel S, Shipley RJ (2019) Modelling the transport of fluid through heterogeneous, whole tumours in silico. PLoS Comput Biol 15:e1006751. https://www.ncbi.nlm.nih.gov/pubmed/31226169

Tavares J, Jorge N (2012) Lecture notes in computational vision and biomechanics. Springer, Berlin. https://link.springer.com/book/10.1007/978-94-007-4270-3

Taylor DL et al (2019) Harnessing human microphysiology systems as key experimental models for quantitative systems pharmacology. In: Handbook of experimental pharmacology. Springer, Berlin. http://link.springer.com/10.1007/164_2019_239

Tian G, Hindle M, Lee S, Longest PW (2015) Validating CFD predictions of pharmaceutical aerosol deposition with in vivo data. Pharm Res 32:3170–3187. http://link.springer.com/10.1007/s11095-015-1695-1

Topp B, Trujillo ME, Sinha V (2019) Industrialization of quantitative systems pharmacology. CPT Pharmacometrics Syst Pharmacol 8(6):356–358. https://www.ncbi.nlm.nih.gov/pubmed/31087476

Trayanova NA (2011) Whole-heart modeling: applications to cardiac electrophysiology and electromechanics. Circ Res 108:113–128. https://www.ncbi.nlm.nih.gov/pubmed/21212393

Troy F, Kris I (2017) Medical imaging: technology and applications. CRC Press, Boca Raton. https://www.crcpress.com/Medical-Imaging-Technology-and-Applications/Farncombe-Iniewski/p/book/9781138072282

Viceconti M, Hunter P (2016) The virtual physiological human: ten years after. Annu Rev Biomed Eng 18:103–123. https://www.ncbi.nlm.nih.gov/pubmed/27420570

Visser SA, de Alwis DP, Kerbusch T, Stone JA, Allerheiligen SR (2014) Implementation of quantitative and systems pharmacology in large pharma. CPT Pharmacometrics Syst Pharmacol 3:1–10. https://www.ncbi.nlm.nih.gov/pubmed/25338195

Vitturi MdM (2016) Navier-stokes equations in cylindrical coordinates. https://vhub.org/resources/4126

Voronov RS, Stalker TJ, Brass LF, Diamond SL (2013) Simulation of intrathrombus fluid and solute transport using in vivo clot structures with single platelet resolution. Ann Biomed Eng 41:1297–1307. https://doi.org/10.1007/s10439-013-0764-z. https://linkinghub.elsevier.com/retrieve/pii/S0022202X15370834

Vrielinck L, Politis C, Schepers S, Pauwels M, Naert I (2003) Image-based planning and clinical validation of zygoma and pterygoid implant placement in patients with severe bone atrophy using customized drill guides. Preliminary results from a prospective clinical follow-up study. Int J Oral Maxillofac Surg 32:7–14. https://www.ncbi.nlm.nih.gov/pubmed/12653226

Wagenseil JE, Mecham RP (2012) Elastin in large artery stiffness and hypertension. J Cardiovasc Transl Res 5:264–273. https://www.ncbi.nlm.nih.gov/pubmed/22290157

Wenzel C et al (2014) 3D high-content screening for the identification of compounds that target cells in dormant tumor spheroid regions. Exp Cell Res 323:131–143. https://www.ncbi.nlm.nih.gov/pubmed/24480576

Wenzel C, Otto S, Prechtl S, Parczyk K, Steigemann P (2015) A novel 3D high-content assay identifies compounds that prevent fibroblast invasion into tissue surrogates. Exp Cell Res 339:35–43. https://doi.org/10.1016/j.yexcr.2015.10.003

White F (1981) Fluid mechanics. In: Rank F, White M (eds) International student edition, McGraw-Hill Kogakusha, Tokyo, 1979. 701 p. £5.95 advances in applied mechanics, vol 20. Academic Press, 1980. 235 p. $31.50. Laminar-turbulent transition. PPLER RE (ed) J Fluid Mech 111:498. http://www.journals.cambridge.org/abstract_S0022112081222482

Xu XY, Collins MW (1990) A review of the numerical analysis of blood flow in arterial bifurcations. Proc Inst Mech Eng Part H J Eng Med 204:205–216. https://doi.org/10.1243/PIME_PROC_1990_204_258_02

Pharmacometabonomics: The Prediction of Drug Effects Using Metabolic Profiling

Jeremy R. Everett

Contents

1 Introduction 264
2 Discovery of Pharmacometabonomics 275
3 Recent Developments in Pharmacometabonomics and the Delivery of Personalised Medicine 280
3.1 Prediction of Drug Pharmacokinetics (PK) 281
3.2 Prediction of Drug Metabolism 283
3.3 Prediction of Drug Efficacy 284
3.4 Prediction of Drug Safety 285
3.5 Not Pharmacometabonomics! 286
3.6 Prediction of Interventions Other Than Drug Treatment: Predictive Metabonomics 287
4 Conclusions 288
Glossary 289
References 291

Abstract

Metabonomics, also known as metabolomics, is concerned with the study of metabolite profiles in humans, animals, plants and other systems in order to assess their health or other status and their responses to experimental interventions. Metabonomics is thus widely used in disease diagnosis and in understanding responses to therapies such as drug administration. Pharmacometabonomics, also known as pharmacometabolomics, is a related methodology but with a prognostic as opposed to diagnostic thrust. Pharmacometabonomics aims to predict drug effects including efficacy, safety, metabolism and pharmacokinetics, prior to drug administration, via an analysis of pre-dose metabolite profiles. This article will review the development of pharmacometabonomics as a new field of science that

J. R. Everett (✉)
Medway Metabonomics Research Group, University of Greenwich, Kent, UK
e-mail: j.r.everett@greenwich.ac.uk

J. E. Barrett et al. (eds.), *Concepts and Principles of Pharmacology*,
Handbook of Experimental Pharmacology 260, https://doi.org/10.1007/164_2019_316

has much promise in helping to deliver more effective personalised medicine, a major goal of twenty-first century healthcare.

Keywords
Metabolic phenotyping · Metabolomics · Metabonomics · Metabotypes · NMR spectroscopy · Personalised medicine · Pharmacometabolomics · Pharmacometabonomics · Precision medicine · Systems medicine

1 Introduction

Metabolic profiling of biological fluids has a long history going back hundreds, if not thousands, of years, to simple methods for detecting sweet-tasting urine as a biomarker for diabetes (Burt and Nandal 2016; Lindon and Wilson 2016). The science of metabolic profiling developed rapidly in the 1980s as huge advances were made in the power and sensitivity of the nuclear magnetic resonance (NMR) spectroscopy and mass spectrometry (MS) detection technologies used in most metabolic profiling studies. Then in the late 1990s the sciences of metabonomics and metabolomics were named and defined. Metabonomics was defined in an interventional, i.e. experimental paradigm by the groups of Jeremy Nicholson and Jeremy Everett at Birkbeck College/Imperial College and Pfizer respectively as "the quantitative measurement of the multiparametric metabolic response of living systems to pathophysiological stimuli or genetic modification" (Lindon et al. 2000). The alternative term metabolomics was defined in an observational fashion a few years later by Fiehn as "a comprehensive analysis in which all the metabolites of a biological system are identified and quantified" (Fiehn 2002). The two terms are now used inter-operatively in spite of the stark differences between the definitions. The blanket term metabolic profiling is also used interchangeably with both terms (Lindon et al. 2007, 2019).

Metabonomics has many uses in the clinical arena including studies of disease mechanisms and biomarkers, disease diagnosis, detection of inborn errors of metabolism, the effects of therapeutic interventions on patients, drug metabolism, drug efficacy and drug safety (Lindon et al. 2007, 2019). The experiments are typically performed using NMR and MS technologies to detect and identify large numbers of metabolites in biological fluids such as urine, blood plasma, sweat, cerebrospinal fluid, tears, etc., but occasionally in body tissues as well. The metabolites detected in these metabonomics experiments are derived from a variety of sources including human endogenous, non-human endogenous (mainly the microbiome) and exogenous (external) sources including food, drink, drugs and the exposome. The phenotype of an organism is dictated by both the metabolites and the proteins that it contains and these may derive from many sources (Fig. 1).

Metabonomics experiments are typically conducted in an interventional or a diagnostic paradigm. Differences in metabolite profiles following an experimental intervention such as drug treatment are used to interpret the biological and biochemical effects of that treatment. In some cases, the intervention will produce a simple

metabolites			proteins: functional and structural		
human endogenous	non-human endogenous	environmental exogenous	human endogenous	non-human endogenous	environmental exogenous
metabonome					
proteome	bacteriome		post-translational modification		
transcriptome	mycome	food	transcriptome		food
epigenome	virome	drugs	epigenome	microbiome	drugs
genome	parasitome	exposome	genome	parasitome	

Fig. 1 The metabolites and proteins found in the human body may originate from inside the body (endogenous) or from various sources outside (exogenous). The pathway from gene to product is shown for the human endogenous metabolites and proteins, and the origins of non-human endogenous and exogenous metabolites and proteins are given

change, such as the reduction or increase in the concentration of one or a small number of key metabolites. In other cases, the intervention may produce widespread changes in the concentrations of a large number of metabolites and multivariate statistical analysis methods such as principal components analysis (PCA) can be used to simplify the data analysis and visualise the changes in metabolite space (Fig. 2a). In this "event interpretation" mode of metabonomics, the changes from pre-intervention (open circles) to post-intervention metabolic state (black squares) are interpreted in relation to the nature of the intervention applied. Another typical use of metabonomics is to distinguish between different groups of subjects, such as patients with a disease, such as liver failure (orange squares), compared to age- and gender-matched healthy human controls (green circles, Fig. 2b). In fact, the diagnostic paradigm of metabonomics is equivalent to the interventional paradigm if one considers that the intervention could be, for example, the presence or absence of a disease.

The technologies with which metabonomics experiments are conducted are important. There are two main technologies in use for metabolite detection and identification in biological fluids and tissues/tissue extracts today: mass spectrometry (MS), usually hyphenated together with a separation technology such as HPLC, UPLC, GC or CE, and NMR spectroscopy (Lindon et al. 2007, 2019; Markley et al. 2017; Nicholson et al. 2016; Wehrens and Salek 2019; Wilson 2015). Good protocols and guides for conducting the experiments by MS (Chen et al. 2016; Scalbert et al. 2009) or NMR (Beckonert et al. 2007; Gowda and Raftery 2017) and good methodologies for identifying the metabolites by MS (Kind and Fiehn 2010; Watson 2013) or NMR (Dona et al. 2016; Markley et al. 2017) are available.

Metabonomics experiments typically analyse the concentrations of metabolites before and after an intervention (Fig. 2a). Modern NMR spectrometers are capable

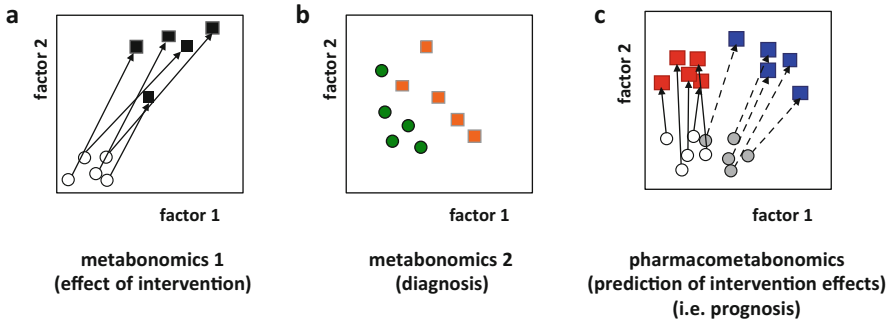


Fig. 2 Schematic representations of the outcomes from the two key experimental approaches to metabolic phenotyping, based on multivariate analysis, e.g. principal components scores, of the metabolic profiles of a number of individuals and showing the first two components (factor 1 and factor 2). Each square or circle represents an individual subject in the study. (**a**) Metabonomics approach 1 (effect of intervention), where open circles represent pre-intervention biofluid metabolic spectral profiles, and black squares represent post-intervention metabolic profiles in the same individuals, where some metabolic perturbation has occurred. The arrows indicate the metabolic trajectory that each individual underwent across metabolic hyperspace as a consequence of the intervention; (**b**) metabonomics approach 2: diagnosis. The metabolite profiles of patients with a disease (orange squares) are distinct from those of healthy controls (green circles) and thus a diagnosis can be made; (**c**) the predictive or prognostic approach. The difference in the *pre-intervention* metabolic profiles of two sub-groups of subjects (white circles v grey circles) allows *prediction* of different post-intervention states for these sub-groups (red and blue squares, respectively). For pharmacometabonomics, the intervention will be drug treatment and the prediction will be of drug PK, metabolism, efficacy or toxicity

of accurately quantifying the biofluid concentrations of dozens to hundreds of metabolites in a few minutes (Fig. 3).

A comparison of the attributes of MS and NMR for conducting metabonomics experiments is given in Table 1. Although far more studies are reported using MS-based detection (see Table 2 below), there is currently a trend to the increasing use of NMR due to its greater stability, ease of automation and reliability, which are important when dealing with large sample number studies, often the case in a clinical setting.

Metabonomics experiments can however be conducted not just by measuring metabolite levels. Metabolite concentration trajectories through time, metabolite entropies and metabolite correlations or networks can also be measured (Fig. 4) and these can often give additional information relative to that obtained from simple concentration measurements before and after an intervention (Everett et al. 2019).

Metabonomics experiments are sometimes categorised as to whether they are targeted or untargeted (Wishart 2016). In the targeted experiments, a selected group of metabolites is analysed, often quantifying the metabolite concentrations relative to an authentic reference standard. In the untargeted experiments, an unbiased approach is used and all the metabolites detected above the sensitivity threshold of the technology employed are analysed. Given the current lack of knowledge of mammalian biology and the complexities of genome – microbiome interactions, adopting

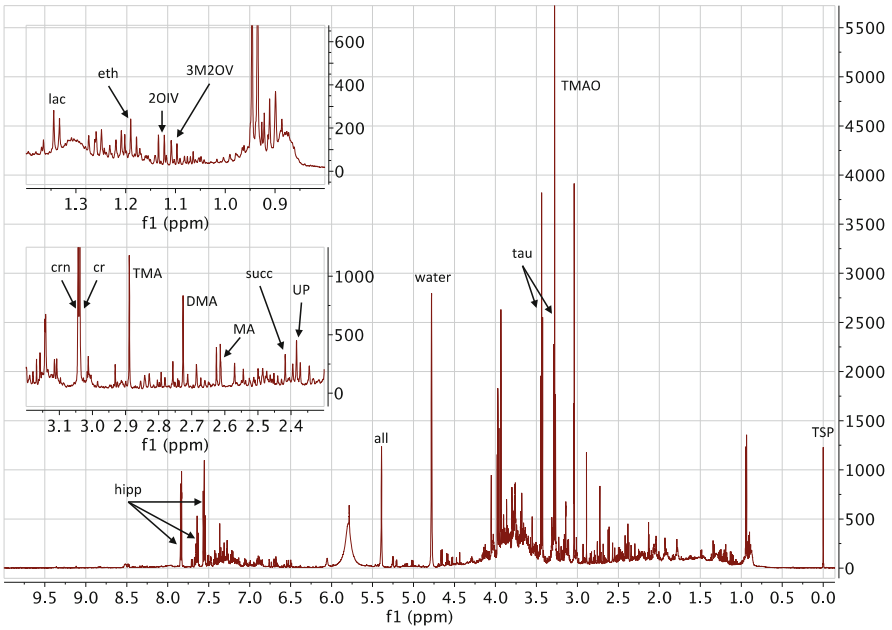


Fig. 3 The 600 MHz ^{1}H NMR spectrum of the urine of a control, male C57BL/6 mouse together with expansions of two low frequency regions, demonstrating the large number of metabolites that can be detected. The identities of some key metabolites are given: *2OIV* 2-oxoisovalerate, *3M2OV* 3-methyl-2-oxovalerate, *all* allantoin, *cr* creatine, *crn* creatinine, *eth* ethanol, *lac* lactate, *DMA* dimethylamine, *hipp* hippurate, *MA* methylamine, *succ* succinate, *TMA* trimethylamine, *TSP* trimethylsilylpropionate-d4 (the chemical shift and quantification reference), *UP* ureidopropionate, *water* the residual signal after water suppression

a targeted approach to metabonomics is only recommended when there is a well-understood biological hypothesis regarding the subjects and the intervention of the experiment. Many surprising and important discoveries are to be made by untargeted methods particularly because our understanding of mammalian biology is so primitive.

Metabonomics experiments can be conducted on a wide variety of sample types including biological fluids such as urine, blood plasma, cerebrospinal fluid, breath condensate, joint fluids, etc. (Lindon et al. 2007). The choice of sample will influence the sort of information that the experiments can provide. Analysis of breath condensates will provide information on the large number of volatile, low molecular weight compounds in exchange with lung tissue, whereas the analysis of blood plasma will provide information on low molecular weight metabolites including sugars, organic acids and amino acids, together with macromolecular compounds such as proteins, glycoproteins and lipoproteins. The analysis of urine (Emwas et al. 2015) can be advantageous: it contains a wide variety of metabolites including amines, organic acids, amino acids and sugars and in mammals, reports on both endogenous mammalian and endogenous microbial metabolites, as well as

Table 1 The attributes and capabilities of mass spectrometry and NMR spectroscopy in metabonomics experiments

NMR spectroscopy	Mass spectrometry
Powerful structure elucidation capability for small molecules in solution giving information on molecular structure, isomerism, conformations and dynamics	Powerful structure analysis capability to generate metabolite mass and molecular fragment information together with molecular formulae at high resolution
Relatively insensitive, but sensitivity improved recently with digital spectrometers, cryoprobes and low volume probes	Highly sensitive
Instrumentation expensive but per sample cost relatively low	Instrumentation relatively inexpensive but running costs high and isotopically-labelled reference standards for quantitation can be expensive
Absolute quantitative measurements and no reference standard required when used with ERETIC technology(Bharti and Roy 2012)	Not absolutely quantitative in absence of specific reference standards, but has relative quantification capability
Highly stable as no contact between sample and spectrometer Little effect of history on data Suitable for large-scale experiments on hundreds to thousands of samples in full automation	Relatively unstable, and may have detector gain changes with large sample numbers Column and spectrometer performance can be affected by history Large sample number runs are difficult due to challenges of maintaining instrument stability
Minimal sample preparation and direct analysis of biological samples	Generally requires a chromatographic separation step prior to MS analysis Gas chromatographic (GC) analysis requires metabolite derivatisation in order to obtain metabolite volatilisation
One set of unique signals for each isomer of each metabolite	Soft ionisation mass spectra may be complicated by multiple adduct formation with multiple spectra for different metal ion and solvent adducts observed for each metabolite GC-MS analyses may be complicated by formation of multiple derivatives
Completely non-destructive technique: Samples can be stored and re-analysed	Sample destroyed in analysis

mammalian – microbial co-metabolites. When the metabolism of a mammal such as a human is perturbed by disease or perhaps the effects of another intervention, such as drug treatment, the metabolic control systems will try to re-establish homeostasis. This will frequently occur by the elimination of unwanted metabolites via the urine, leaving the plasma less affected, thus giving the opportunity to identify the nature of the metabolic perturbation. Frequently, changes to the status of the gut microbiome can be detected by the observation of metabolic perturbations in urine samples.

Although most metabonomics experiments are diagnostic or interventional in mode, some experiments can be prognostic; that is, the metabolite patterns observed can be used to predict future events. The rest of this chapter will be devoted to prognostic metabonomics.

Table 2 A list of pharmacometabonomics studies from 2006 to 2019, sorted by study type and date order

#	Study and reference	Species	Metabolite profiling technology
Prediction of pharmacokinetics (PK)			
1	Prediction of tacrolimus PK in healthy volunteers (Phapale et al. 2010)	Human	LC-MS
2	Prediction of pharmacokinetics of triptolide (Liu et al. 2012)	Rat	GC-MS
3	Prediction of atorvastatin pharmacokinetics in healthy volunteers (Huang et al. 2015)	Human	GC-MS
4	Prediction of methotrexate clearance in patients with lymphoid malignancies (Kienana et al. 2016)	Human	GC-MS
5	Prediction of midazolam clearance in female volunteers (Shin et al. 2016)	Human	GC-MS
6	Pharmacometabonomic prediction of busulphan clearance in haematopoietic stem cell transplant recipients (Navarro et al. 2016)	Human	LC-MS
7	Prediction of intravenous busulphan clearance by endogenous plasma biomarkers using global pharmacometabolomics (Lin et al. 2016)	Human	LC-MS
8	Prediction of busulphan AUC in haematopoietic stem cell transplantation patients (Kim et al. 2017)	Human	LC-MS
9	Prediction of d4-cholic acid pharmacokinetics (Zhang et al. 2017b)	Rat	LC-MS
10	Integrated use of pharmacometabonomics and pharmacogenomics to predict the pharmacokinetics of a novel transient receptor potential vanilloid type 1 (TRPV1) antagonist (Oh et al. 2018)	Human	LC-MS
11	Prediction of zonisamide pharmacokinetics parameters in volunteers (Martinez-Avila et al. 2018a, b)	Human	LC-MS
12	Prediction of methylphenidate PK in healthy volunteers (Kaddurah-Daouk et al. 2018)	Human	LC-MS
13	Prediction of midazolam clearance in *male* volunteers (Lee et al. 2019)	Human	GC-MS
Prediction of drug metabolism			
1	Prediction of paracetamol/acetaminophen metabolism (Clayton et al. 2006) ** *First demonstration of pharmacometabonomics*	Rat	NMR
2	Prediction of metabolism of paracetamol/acetaminophen in human volunteers (Clayton et al. 2009) ** *First demonstration of pharmacometabonomics in humans*	Human	NMR

(continued)

Table 2 (continued)

#	Study and reference	Species	Metabolite profiling technology
3	Prediction of CYP3A4 induction in volunteer twins (Rahmioglu et al. 2011)	Human	NMR
4	Prediction of CYP3A activity in healthy volunteers (Shin et al. 2013)	Human	GC-MS
5	Prediction of losartan metabolism in healthy volunteers (He et al. 2018)	Human	NMR and LC-MS
6	Prediction of methylphenidate (Ritalin [for ADHD]) metabolism in healthy genotyped volunteers (Kaddurah-Daouk et al. 2018)	Human	LC-MS
Prediction of drug efficacy			
1	Prediction of simvastatin efficacy in patients on the cholesterol and pharmacogenomics study (Kaddurah-Daouk et al. 2010; Trupp et al. 2012)	Human	TLC plus GC and GC-MS
2	Prediction of chemotherapy efficacy in breast cancer patients (Stebbing et al. 2012)	Human	NMR
3	Prediction of citalopram/escitalopram response in patients with major depressive disorder (MDD) (Ji et al. 2011) ** *First demonstration of pharmacometabonomics-informed pharmacogenomics approach to personalised medicine* See also Abo et al. (2012) and Gupta et al. (2016)	Human	GC-MS and LC-ECA (LC-electrochemical coulometric array detection)
4	Prediction of sertraline and placebo responses in patients with MDD (Kaddurah-Daouk et al. 2011, 2013; Zhu et al. 2013)	Human	LC-ECA and GC-MS
5	Prediction of efficacy of anti-psychotics in schizophrenia patients (Condray et al. 2011)	Human	LC-ECA
6	Prediction of response to aspirin in healthy volunteers (Ellero-Simatos et al. 2014; Lewis et al. 2013; Yerges-Armstrong et al. 2013)	Human	LC-MS and GC-MS
7	Prediction of efficacy with anti-TNF therapies in rheumatoid arthritis (Kapoor et al. 2013)	Human	NMR
8	Prediction of thiopurine-*S*-methyltransferase phenotype in Estonian volunteers (Karas-Kuzelicki et al. 2014)	Human	HPLC
9	Prediction of efficacy of L-carnitine therapy for patients with septic shock (Evans et al. 2019; Puskarich et al. 2015, 2018)	Human	NMR and LC-MS
10	Prediction of acamprosate treatment outcomes in alcohol-dependent patients (Nam et al. 2015)	Human	LC-MS

(continued)

Table 2 (continued)

#	Study and reference	Species	Metabolite profiling technology
11	Prediction of blood pressure lowering in hypertensive patients treated with atenolol and hydrochlorothiazide (Rotroff et al. 2015)	Human	GC-MS
12	Prediction of response in lung cancer patients (Hao et al. 2016a)	Human	NMR and GC-MS
13	Prediction of patient response to trastuzumab-paclitaxel neoadjuvant therapy in HER-2 positive breast cancer (Miolo et al. 2016)	Human	LC-MS
14	Prediction of patient response in SSRI treatment of major depressive disorder (Gupta et al. 2016)	Human	LC-ECA
15	Prediction of clopidogrel high on treatment platelet reactivity (HTPR) in CAD patients [NMR] (Amin et al. 2017)	Human	NMR
16	Prediction of chemosensitivity of treatment of AML patients with cytarabine and anthracycline (Tan et al. 2017)	Human	LC-MS
17	Prediction of efficacy in pancreatic ductal adenocarcinoma patients receiving gemcitabine (Phua et al. 2017)	Human	GC-TOFMS
18	Prediction of blood pressure lowering by hydrochlorothiazide [lipidomics and pharmacogenomics] (Shahin et al. 2017)	Human	
19	Prediction of efficacy of gemcitabine and carboplatin treatment of metastatic breast cancer patients (Jiang et al. 2018)	Human	NMR
20	Prediction of gemcitabine efficacy in pancreatic ductal adenocarcinoma patients (Phua et al. 2018)	Human	GC-MS
21	Prediction of response to metformin treatment in early T2DM patients (Park et al. 2018)	Human	GC-MS
22	Prediction of efficacy of propranolol in reducing hepatic venous pressure gradient (HPVG) in patients with liver cirrhosis (Reverter et al. 2019)	Human	LC-MS
23	Prediction of efficacy of meglumine antimonite efficacy if patients with cutaneous leishmaniasis (Alejandro Vargas et al. 2019)	Human	LC-MS
24	QUASI-prediction of dexamethasone steroid treatment efficacy in pre-term infants with respiratory syndrome (Cao et al. 2019)	Human	GC-TOF-MS

(continued)

Table 2 (continued)

#	Study and reference	Species	Metabolite profiling technology
25	Prediction of warfarin efficacy in atrial fibrillation patients (Bawadikji et al. 2019)	Human	NMR
Prediction of adverse events			
1	Prediction of toxicity from paracetamol/acetaminophen dosing (Clayton et al. 2006) ** *First demonstration of pharmacometabonomics*	Rat	NMR
2	Prediction of weight gain in breast cancer patients undergoing chemotherapy (Keun et al. 2009) ** *First demonstration of pharmacometabonomics in patients*	Human	NMR
3	Prediction of onset of diabetes in rats administered with streptozotocin (Li et al. 2007)	Rat	GC-MS
4	Prediction of liver injury markers in patients treated with ximelagatran (Andersson et al. 2009)	Human	NMR, GC-MS and LC-MS
5	Prediction of toxicity of paracetamol/acetaminophen ("early-onset pharmacometabonomics") (Winnike et al. 2010)	Human	NMR
6	Prediction of nephrotoxicity of cisplatin (Kwon et al. 2011)	Rat	NMR
7	Prediction of toxicity in patients with inoperable colorectal cancer treated with capecitabine (Backshall et al. 2011)	Human	NMR
8	Prediction of toxicity of isoniazid in rats (Cunningham et al. 2012)	Rat	NMR
9	Prediction of hyperglycaemia in Caucasian hypertensive patients on the PEAR study with atenolol (Weng et al. 2016)	Human	LC-MS
10	Prediction of variability in response to galactosamine treatment (Coen et al. 2012)	Rat	NMR
11	Prediction of hyperglycaemia in Caucasian hypertensive patients on the PEAR study with atenolol (de Oliveira et al. 2016)	Human	GC-TOF-MS and genomics
12	Prediction of toxicity from lipopolysaccharide treatment in rats (Dai et al. 2016)	Rat	LC-MS and GC-MS
13	Prediction of 'high on treatment platelet reactivity (HTPR)' in patients on clopidogrel anti-platelet therapy to prevent stent thrombosis in urine (Amin et al. 2017)	Human	NMR
14	Prediction of nephrotoxicity of cisplatin in rats (Zhang et al. 2017a)	Rat	GC-MS and LC-MS

(continued)

Table 2 (continued)

#	Study and reference	Species	Metabolite profiling technology
15	prediction of "high on treatment platelet reactivity (HTPR)" in patients on clopidogrel anti-platelet therapy to prevent stent thrombosis in plasma (Amin et al. 2018)	Human	NMR
16	Prediction of peripheral neuropathy in breast cancer patients treated with Paclitaxel (Sun et al. 2018)	Human	NMR
17	Prediction of irinotecan gastrointestinal toxicity (Gao et al. 2019)	Rat	GC-MS and LC-MS
Predictive metabonomics			
1	Prediction of developing diabetes (Wang et al. 2011) ** *First predictive metabonomics study*	Human	LC-MS
2	Prediction of pre-diabetes (Wang-Sattler et al. 2012)	Human	LC-MS and flow-injection analysis-MS
3	Prediction of renal function recovery after relief of obstructive uropathy (Dong et al. 2013)	Human	NMR
4	Prediction of all-cause death (Fischer et al. 2014)	Human	NMR
5	Prediction of stroke recurrence after transient ischemic attack (Jove et al. 2015)	Human	LC-MS
6	Prediction of breast cancer risk (Bro et al. 2015)	Human	NMR
7	Prediction of preeclampsia and gestational hypertension (Austdal et al. 2015)	Human	NMR
8	Prediction of development of obesity (Ni et al. 2015)	Human	LC-MS
9	Prediction of 1-year outcome in subarachnoid haemorrhage (Sjoberg et al. 2015)	Human	GC-MS
10	Prediction of survival of lung cancer patients undergoing treatment (Hao et al. 2016a, b)	Human	GC-MS and NMR
11	A predictive metabolic signature for the transition from gestational diabetes to type 2 diabetes (Allalou et al. 2016)	Human	GC-MS and LC-MS
12	Prediction of survival of patients with decompensated cirrhosis s 2016 (McPhail et al. 2016)	Human	NMR and LC-MS
13	Prediction of postoperative hypoxaemia (Maltesen et al. 2016)	Human	NMR
14	Prediction of ALS clinical progression (Blasco et al. 2018)	Human	LC-MS
15	Prediction of all-cause death (Deelan et al. 2019)	Human	NMR

Significant studies are highlighted with *double asterisks in italic* **
Some studies have several publications associated with them
The table is unlikely to be exhaustive due to the different keywords used for some studies

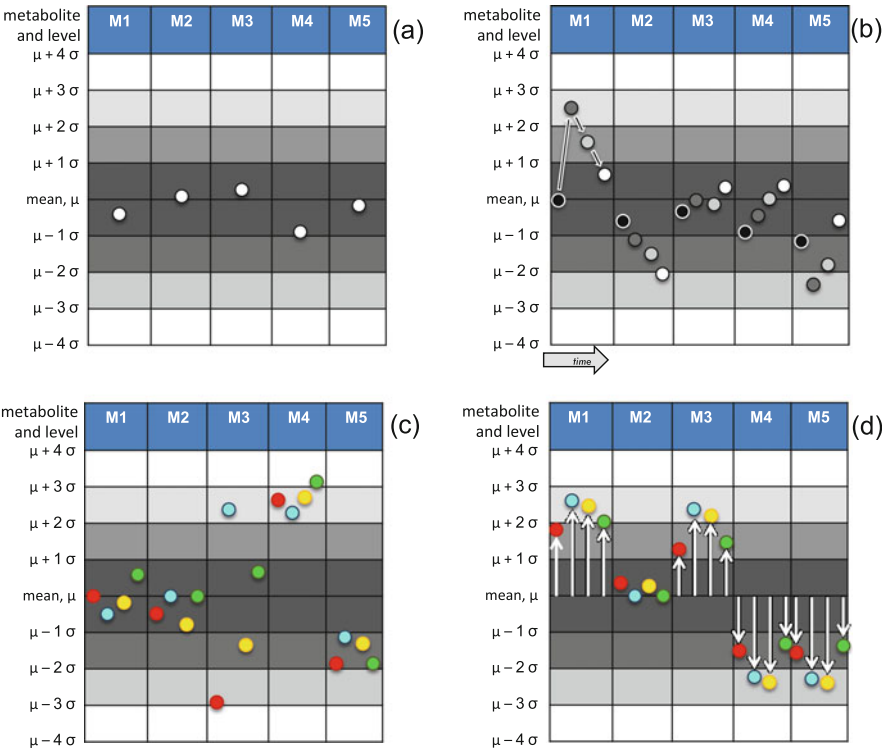


Fig. 4 A *schematic* representation of the four principal approaches to the measurement of metabonomic data: via (**a**) metabolite levels, (**b**) trajectories, (**c**) entropies or (**d**) correlations/dependencies. In Box (**a**), the *levels* of five different metabolites M1–M5 in one normal individual (white circles) are superimposed on a chart that represents the normal population distribution of metabolite levels. Symbols μ and σ are the mean and standard deviations of the levels of the metabolites for that population and with normal distribution. In Box (**b**), we see the *trajectories* over time for the same five metabolites M1–M5 in one individual subjected to an intervention of some kind. The time course of the metabolic trajectory moves from left to right in each metabolite column and is represented by circles, whose shading gets lighter over time. Arrows connect the time points for metabolite M1 but others are omitted for clarity. It can be seen that as a result of the challenge, the levels of some metabolites (M1, M2 and M5) undergo positive and negative excursions from the normal population values, whereas other metabolites are less affected. In Box (**c**), the metabolic *entropies* of a cohort of four individuals that have been subjected to a challenge are represented. The metabolite level for each individual is coloured differentially (red, blue, yellow and green circles represent individuals 1, 2, 3 and 4, respectively). It can be seen that in this cohort there is high metabolic entropy for metabolite M3 (metabolite levels are distributed across a very wide range of values/configurational states following the intervention) and significant disturbances in the metabolite *levels* for M4 and M5, but much lower metabolic *entropy* for metabolites, M1, M2, M4 and M5. In Box (**d**), the metabolite correlations seen for five metabolites (M1–M5) in four human subjects (red, blue, yellow and green circles represent individuals 1, 2, 3 and 4, respectively) are shown following an intervention. The intervention causes a significant increase in the concentrations of metabolite M1 for all four subjects (white vertical arrows), although to differing degrees. The same pattern of disturbance is seen for metabolite M3 in all four subjects. It is clear that the concentrations of metabolites M1 and M3 are correlated, with the excursions from the mean

2 Discovery of Pharmacometabonomics

A high degree of "biological variation", i.e. widely varying results, was often observed in early drug metabolism and drug safety studies in Beecham Pharmaceuticals and Pfizer R & D in the 1980s and 1990s. The causes of this variance were unknown but could lead to widely disparate results, sometimes to the extent that doubts were raised as to whether the drug in question had been dosed properly. Pfizer and Imperial College had established a panomics study of early drug safety signals in the 1990s. At a collaboration meeting in Amboise, France on 18th October 2000, the topic of widely varying safety data on galactosamine and isoniazid was discussed. The notion emerged from the meeting that the metabolic phenotype of the animals *prior to dosing* was influencing differential responses to the drug *post-dose*. A series of experiments was designed to test this notion, and the concept of pharmacometabonomics was born.

The first key experiment was to test the hypothesis that pre-dose rat metabolite profiles could predict post-dose drug metabolism and safety for the common analgesic paracetamol, also known as acetaminophen (Clayton et al. 2006). A dose of 600 mg/kg was administered to 65 Sprague-Dawley rats, and urine samples were collected both pre- and post-dosing and then analysed by 600 MHz ^{1}H NMR spectroscopy. A validated projection to latent structure (PLS) model showed a statistically significant correlation between pre-dose urine metabolite concentrations and the post-dose ratio of the metabolite paracetamol glucuronide (G) to the parent drug paracetamol (P, Fig. 5).

In addition, unbiased principal components analysis (PCA) of the pre-dose urine ^{1}H NMR spectra showed a partial correlation between the mean liver histopathology score (MHS) and principal component 2 (PC2) of the data (Fig. 6). A Mann–Whitney *U* test showed the statistical significance of the separation of the *pre-dose* NMR data for rats in class 1 (minimal/no liver pathology) and class 3 (significant liver pathology) with $p = 0.002$. The pre-dose levels of taurine were negatively correlated with the post-dose degree of liver pathology, consistent with taurine's known role in protecting against paracetamol toxicity (Waters et al. 2001). The taurine levels may have reflected the availability of inorganic sulphate to individual rats. Inorganic sulphate is needed for the biosynthesis of both taurine and for the paracetamol-sulphating agent phosphoadenosine phosphosulphate (PAPS). Consistent with this, rats with a high degree of liver necrosis showed a low degree of paracetamol sulphation (Clayton et al. 2006).

Thus it was clearly demonstrated that pre-dose metabolite profiles could enable the prediction of post-dose effects including drug metabolism and toxicity. This is pharmacometabonomics, which was defined as "the prediction of the outcome (for

Fig. 4 (continued) greatest for the yellow and blue subjects. By contrast the concentrations of metabolites M4 and M5 are anti-correlated with those of M1 and M3. It could be inferred from the correlations of the concentrations of these metabolites that they may be in the same or a related biochemical pathway. The levels of metabolite M2 are relatively undisturbed

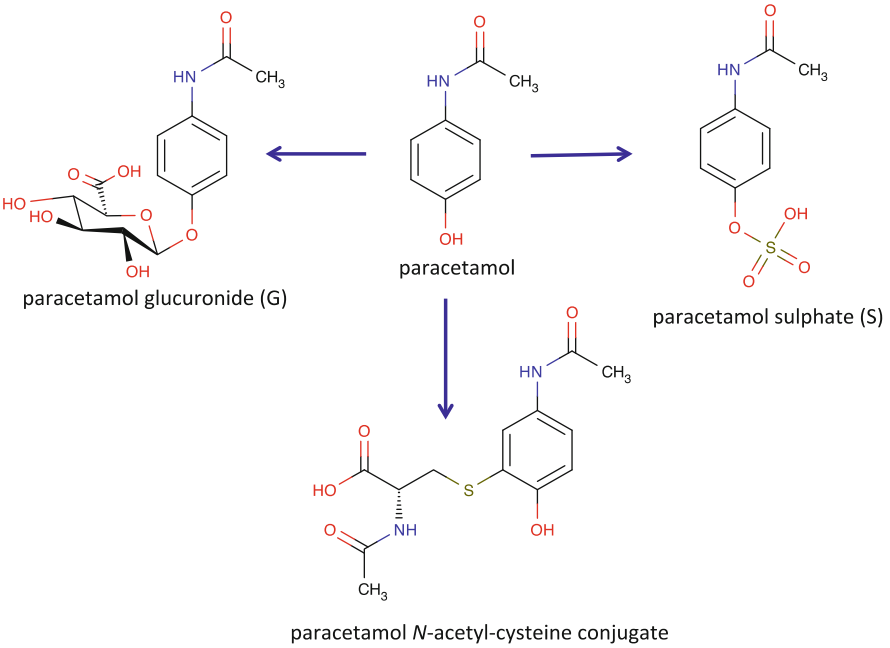


Fig. 5 The molecular structures of paracetamol (P) and its major metabolites

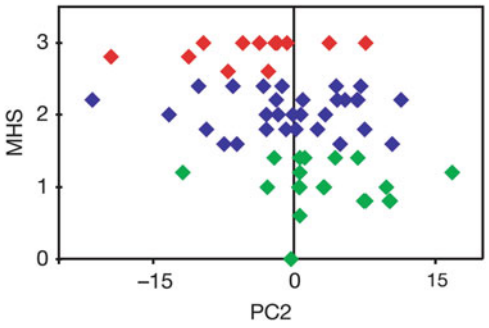


Fig. 6 A plot of paracetamol liver toxicity as measured by the mean liver histopathology score (MHS) against principal component 2 (PC2) of the *pre-dose* urine NMR spectral data. A partial class separation is observed. Each point represents a single rat and is colour-coded by its histology class with increasing degree of liver pathology: class 1 is green (minimal/no pathology), class 2 is blue (intermediate pathology), class 3 is red (significant pathology). Figure reproduced from Nature Publishing Group (Clayton et al. 2006)

example, efficacy or toxicity) of a drug or xenobiotic intervention in an individual based on a mathematical model of pre-intervention metabolite signatures" (Clayton et al. 2006). Pharmacometabonomics is a prognostic or predictive methodology, in contrast to the diagnostic mode of metabonomics and it is the metabolic equivalent

of pharmacogenomics, which is the use of genetic information to predict drug effects in advance of dosing (Salari et al. 2012).

The initial success of pharmacometabonomics experiments in animals prompted the question of whether the method would work in humans. A Pfizer/Imperial College research team therefore set up an experiment to test the hypothesis that pre-dose urine metabolite profiles could predict post-dose drug metabolism, again in the analgesic paracetamol. A normal clinical dose of paracetamol (two 500 mg tablets with water) was administered to 100, normal, male volunteers in March and April 2003. Urine samples were collected both pre-dose and 0–3 and 3–6 h post-dose and these were analysed by both 600 MHz ^{1}H NMR spectroscopy (Fig. 7) and UPLC-MS (Clayton et al. 2009).

The pre-dose ^{1}H NMR spectrum of volunteer 1 (Fig. 7a) showed signals from microbial metabolites such as hippurate (2) and human metabolites such as citrate (5) in addition to an unknown metabolite (4) with a singlet methyl signal at ca 2.35 ppm and second-order aromatic doublet signals between ca 7.2 and 7.3 ppm. This volunteer excreted more paracetamol glucuronide metabolite (8) than paracetamol sulphate (7) as is clear in the ^{1}H NMR spectrum of the 0–3 h post-dose urine (Fig. 7b) where both methyl group singlet and second-order aromatic doublet signals for these metabolites are clearly visible. By contrast, volunteer 2 excreted no visible quantity of unknown metabolite 4 pre-dose but excreted a much higher ratio of paracetamol sulphate (7) to glucuronide (8) post-dose (Fig. 7c, d).

Analysis of the remaining urinary ^{1}H NMR data showed that this pattern was present across all of the volunteers (Fig. 8).

It is clear from Fig. 8 that when the pre-dose ratio of metabolite 4 normalised to creatinine is greater than 0.06, then the post-dose paracetamol sulphate (S) to paracetamol glucuronide (G) ratio is always less than 0.8. The same pattern was found when the 3–6 h post-dose urines were analysed. Mann–Whitney U tests in conjunction with a Bonferroni correction to counter the effects of multiple hypothesis testing showed that the association of high metabolite 4 to creatinine ratios with low S/G ratios was statistically significant for both the 0–3 h ($p = 0.0001$) and 3–6 h ($p = 0.00012$) post-dose urines. With a Bonferroni correction of 100, the p value for statistical significance is 0.0005 instead of 0.05 (Broadhurst and Kell 2006).

Thus, it was clear that there was a statistically significant correlation, between the presence of metabolite 4 at high levels pre-dose and diminished paracetamol sulphate (S) to paracetamol glucuronide (G) ratios post-dose. It therefore became important to identify unknown metabolite 4.

Metabolite 4 possesses a singlet, three proton signal at ca 2.35 ppm indicating the presence of a methyl group attached to an sp^2 carbon on the basis of its chemical shift. The metabolite also possessed two, second-order aromatic doublet signals of two hydrogens each, indicating that metabolite 4 had a methyl group attached to a benzene ring with a substituent para to the methyl group. Metabolite 4 was identified as 4-cresolsulphate (HMDB11635) (Wishart et al. 2018) by both unambiguous chemical synthesis and spiking and by enzymatic desulphation to 4-cresol in situ (Fig. 9) (Clayton et al. 2009).

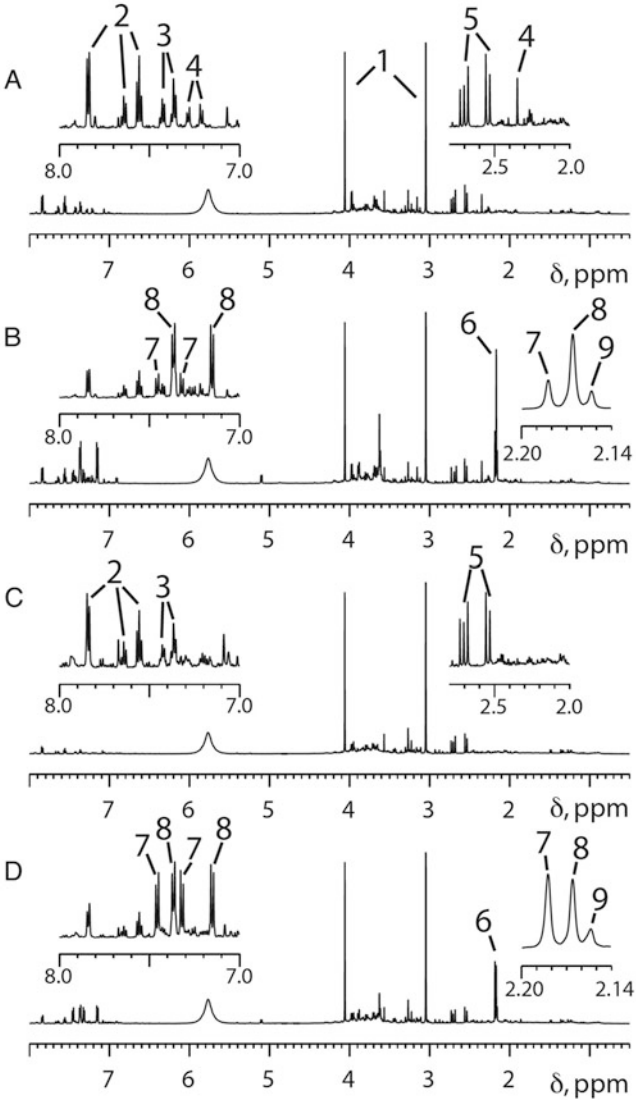


Fig. 7 600 MHz ^{1}H NMR spectra of the urines of volunteers taking a 1 g oral dose of paracetamol. (**a**) Spectrum of pre-dose urine of volunteer 1 together with expansions of the aromatic and lower frequency regions. (**b**) 0–3 h post-dose urine spectrum of volunteer 1. (**c** and **d**) The corresponding pre-dose and post-dose urine spectra of volunteer 2, respectively. Key to NMR signal numbers: 1, creatinine; 2, hippurate; 3, phenylacetylglutamine; 4, unknown metabolite; 5, citrate; 6, cluster of signals from N-acetyl groups from paracetamol-related compounds that resolves into 7, 8 and 9 on expansion; 7, paracetamol sulphate; 8, paracetamol glucuronide; 9, other paracetamol-related compounds. Reproduced with permission from PNAS (Clayton et al. 2009)

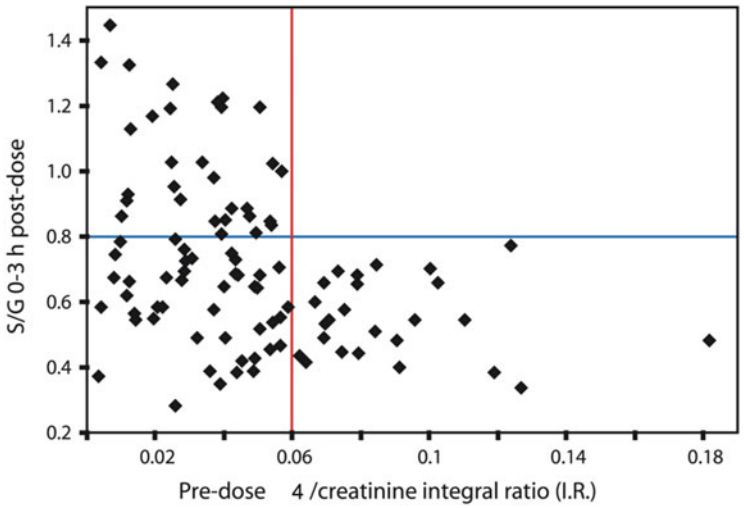


Fig. 8 The urinary ratio of paracetamol sulphate (S) to paracetamol glucuronide (G) excreted 0–3 h post-dose plotted against the pre-dose ratio of metabolite 4 normalised to creatinine. Reproduced with permission from PNAS (Clayton et al. 2009)

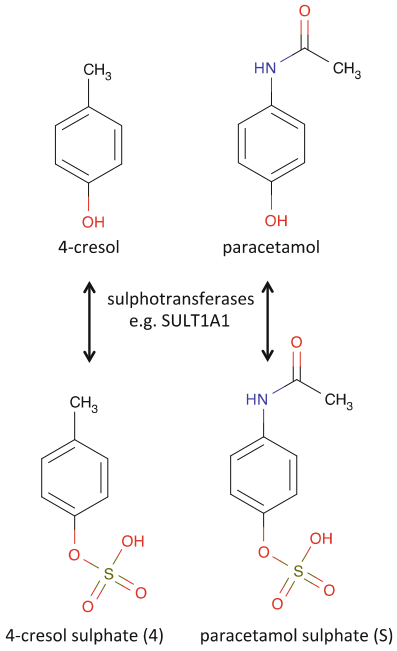


Fig. 9 The molecular structures of 4-cresol and paracetamol and their corresponding sulphate metabolites

The revelation that 4-cresolsulphate was a biomarker that, at least in part, could enable the prediction of the metabolic fate of paracetamol in humans was a surprise. 4-Cresolsulphate is made in humans by the sulphation of 4-cresol, itself a product of gut bacteria, particularly *Clostridia* species. Thus the human metabolism of the widely used analgesic paracetamol (acetaminophen) is at least in part under the control of gut bacterial metabolites. The influence of the gut microbiome on drug properties was not widely recognised at this time and this paper helped to highlight these important effects (Wilson 2009).

The reason for the relationship between 4-cresol and paracetamol metabolism is evident from an inspection of Fig. 9. The molecular structures of 4-cresol and paracetamol are quite similar and both are sulphated by the same sulphotransferases, particularly SULT1A1. In humans, as opposed to rodents, 4-cresol is metabolised almost exclusively by sulphation with no significant glucuronidation. However, this sulphation requires the sulphate donor cofactor 3-phosphoadenosine 5-phosphosulfate (PAPS) and its supply is limited in humans (Gamage et al. 2006). Therefore in a human with a high 4-cresol burden due to their gut microbiome, a significant amount of PAPS is used in 4-cresol sulphation and a challenge to the body of that person of a large dose of a drug requiring sulphation, results in the body turning to the alternative elimination pathway of glucuronidation and the consequent decreased S/G metabolite ratios. Note that these findings have implications for all drugs metabolised by sulphation and implications also for endogenous metabolism involving sulphation (Clayton et al. 2009). Finally, it is worth noting that a number of diseases including childhood autism, childhood hyperactivity and Parkinson's disease are associated with increased 4-cresolsulphate levels or altered S/G ratios after paracetamol administration and it is therefore likely that there is a microbiome influence on these disease states (Clayton et al. 2009).

3 Recent Developments in Pharmacometabonomics and the Delivery of Personalised Medicine

The prediction of paracetamol metabolism and safety described above represented the first definitive demonstration of pharmacometabonomics. Since that study was published (Clayton et al. 2006) numerous other studies have emerged demonstrating the ability of pharmacometabonomics methodologies to predict drug pharmacokinetics, metabolism, efficacy and safety in animals and humans (Burt and Nandal 2016; Everett 2016; Everett et al. 2013, 2016). These studies are important because they promise a new way to help deliver personalised medicine, which is a key objective of twenty-first century healthcare (Nicholson et al. 2011, 2016). The aim of personalised medicine is to select treatments that provide optimal efficacy with minimal toxicity or side effects for a given patient group, rather than giving the same standard treatment to all patients regardless of outcomes. It is a shocking fact that many drugs are ineffective or even unsafe in a high percentage of patients. It has been estimated that in the USA in 1994, over two million patients had serious

adverse drug reactions (ADRs), resulting in hospitalisation, disability or, in 106,000 cases, death (Lazarou et al. 1998). A more recent study put the cost of ADRs to the US economy in the range $30 billion to $100 billion per year. Thus the need to be able to prescribe medicines that are both effective and also safe for patients is clear.

Pharmacogenomics, i.e. the use of patient genetic information to predict drug effects has been important in enabling the development of personalised medicine in some areas especially in the prediction of the effects of "drug metabolising" enzymes such as cytochrome P450s on drug efficacy and safety (Lee et al. 2014). However, in many complex, multi-factorial diseases, the use of pharmacogenomics information has had more limited success (Pirmohamed 2014). Given the impact of environmental factors on drug effects, such as the status of the gut microbiome (Clayton et al. 2009), and the impact of drug-drug interactions, especially in phenoconversion (Shah and Smith 2015), it is not surprising that human pharmacogenomics studies have encountered challenges in progressing from success in the laboratory to success in clinical practice (Pirmohamed 2014). It is therefore encouraging that metabolic studies in the form of pharmacometabonomics can assist in the prediction of drug effects and with the implementation of personalised medicine. We will review progress in this area in the remainder of this chapter.

Table 2 provides an overview of the key pharmacometabonomics and predictive metabonomics studies that we are aware of, using the keywords pharmacometabonomics and pharmacometabolomics in PubMed. However, the list is unlikely to be exhaustive as some authors do not put these terms in either the title or keyword list. In addition, there are a minority of authors who are using the term pharmacometabonomics or pharmacometabonomics to describe diagnostic experiments with no prognostic elements.

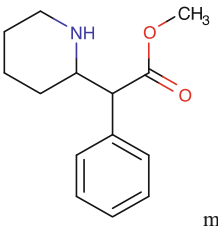
It can be seen from Table 2 that there are 13 studies dealing with the prediction of drug pharmacokinetics, 6 on prediction of drug metabolism, 25 on prediction of drug efficacy, 17 on prediction of adverse events and a further 15 predictive metabonomics studies where the prediction is based on an intervention other than drug administration. Thus we have at least 61 pharmacometabonomics studies in the literature to date. Of these 61 pharmacometabonomics and 15 predictive metabonomics studies, 65 were conducted in humans and 11 in the rat.

The development of pharmacometabonomics has been significant over the past 10 years especially. Several reviews of the field have already appeared (Burt and Nandal 2016; Everett 2016), so in the remainder of this chapter, we will focus on recent developments in the four key areas of prediction of drug pharmacokinetics, metabolism efficacy and safety.

3.1 Prediction of Drug Pharmacokinetics (PK)

The prediction of drug PK is especially important in situations where the therapeutic index (TI) of a drug is relatively low and also variable. Inappropriately high drug doses may lead to adverse effects in individual patients. The group of Rima Kaddurah-Daouk et al. used LC-MS methodologies to measure correlations between

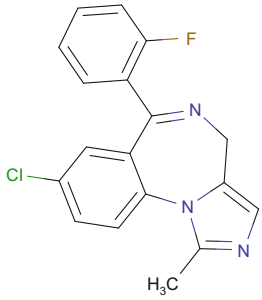
baseline plasma lipids of healthy volunteers and the PK of methylphenidate, trade name Ritalin (Kaddurah-Daouk et al. 2018).



methylphenidate, Ritalin

The phosphatidylcholine PC(38:5) was negatively correlated with the drug AUC and the blood plasma C_{max} values and the ceramide Cer(d18:1/24:1) was positively correlated with the plasma half-life of the drug metabolite ritalinic acid. Carboxylesterase 1(CES1) metabolises methylphenidate and other drugs such as cocaine and heroin via amide and ester bond hydrolysis. It was suggested that CES1 has a role in lipid metabolism and that the findings could be used for the prediction of the PK not only of methylphenidate, but other drugs metabolised by CES1 (Kaddurah-Daouk et al. 2018).

Differences in cytochrome P450 3A activities are a major source of variability in patient drug responses. Lee et al. (2019) developed a model for the prediction of CYP3A activity in the presence of inhibitors and inducers that was able to predict the clearance of midazolam with $r^2 = 0.75$. GC and GC-MS methodology was used in a targeted fashion to measure the concentrations of a small number of endogenous steroids in human volunteer urine and plasma samples.

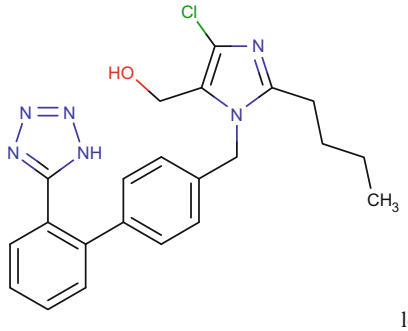


midazolam

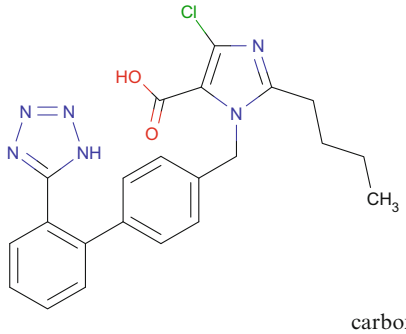
These data were amalgamated together with CYP3A5 genotype information to develop a model for the prediction of midazolam clearance. It was concluded that use of the model could be valuable for predicting CYP3A activities generally in drug development but that further validation was required.

3.2 Prediction of Drug Metabolism

He et al. have shown that pre-dose profiling by NMR spectroscopy of volunteer blood plasma could allow prediction of some metabolic and PK characteristics of losartan and its metabolite EXP3174 (He et al. 2018).



losartan



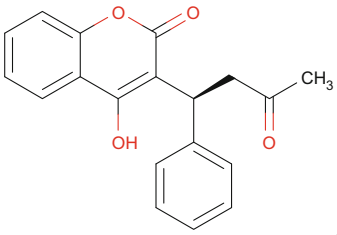
carboxylosartan, E3174

Losartan and its bioactive metabolite EXP3174 show a large degree of inter-individual differences in blood plasma concentrations that impact upon efficacy and safety. He et al. showed that pre-dose LDL/VLDL, lactate, citrate, creatine and glucose concentrations were positively correlated with, and HDL, creatinine, choline, glycine and phosphorylcholine concentrations were negatively correlated with the ratio of AUCs of EXP3174 and losartan. Pre-dose LDL/VLDL, lactate and glucose concentrations were positively correlated with, and choline, citrate concentrations were negatively correlated with the ratio of C_{max} values of EXP3174 and losartan. The switch of citrate from positively correlating with the ratio of AUCs to negatively correlating with the ratio of C_{max} values of EXP3174 and losartan was not commented upon. However, as Table 2 in the paper shows that the FDR value for citrate in the pathway analysis was 0.64, i.e. a >60% chance of a false discovery, then perhaps that switch is not surprising. Simple formulae

involving creatinine and lactate and also choline and glucose were derived for calculating the ratios of the AUCs and the C_{max} values of EXP3174 and losartan, respectively (He et al. 2018).

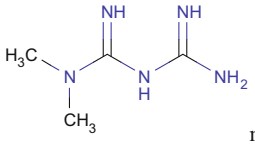
3.3 Prediction of Drug Efficacy

NMR spectroscopy of blood plasma was used to show a discrimination between atrial fibrillation patients on warfarin treatment that had stable versus unstable blood thickness. However, the study was not able to demonstrate any such discrimination for patients who were newly treated with warfarin and it was concluded that further studies were required (Bawadikji et al. 2019).



warfarin

Park and co-workers used GC-MS analysis of urine metabolites in early-phase type 2 diabetes mellitus (T2DM) patients to show that baseline levels of citrate and hippurate were significantly different for responders and non-responders to metformin treatment (Park et al. 2018).

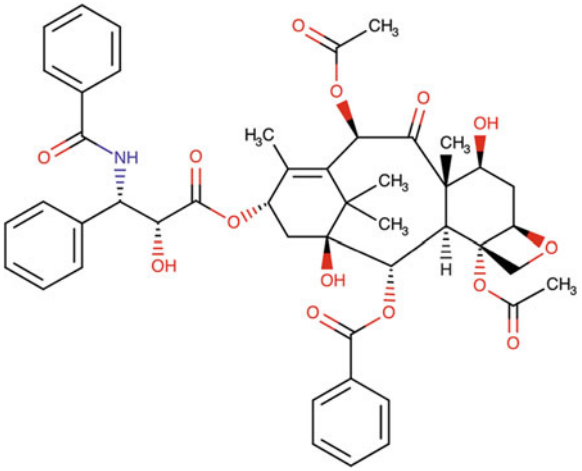


metformin

The response to treatment was assessed on the basis of changes in glycated haemoglobin A1c (HbA1c) levels from baseline. Pre-dose levels of myo-inositol were also marginally significantly different between these groups. This study was seen to be important in the context of developing personalised medicine, given the significant global burden of T2DM and the variability of patient response to treatment with metformin, a key medicine for treatment of the disease.

3.4 Prediction of Drug Safety

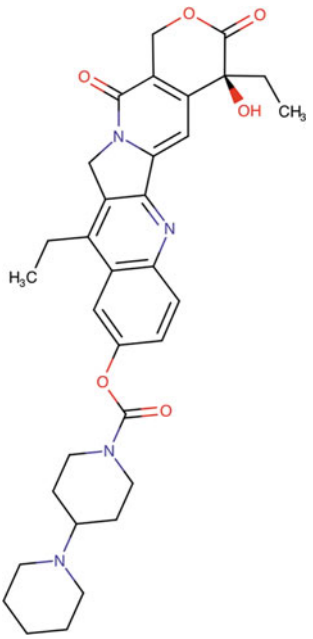
Paclitaxel (brand name Taxol) is a natural product widely used in the treatment of breast cancer. However, its usage is limited by many side effects including the development of peripheral neuropathy, which causes treatment delays or discontinuation in about one quarter of the patients (Sun et al. 2018).



paclitaxel, HMDB0015360, (Wishart et al. 2018)

The group of Sun et al. used an NMR spectroscopic approach to show that pre-treatment levels of blood histidine, phenylalanine and threonine were inversely associated with maximal change in the peripheral neuropathy index CIPN8 (Sun et al. 2018). This work promises to inform personalised medicine approaches to the selection of patients for treatment who will not suffer peripheral pain side effects.

Colorectal cancer is commonly treated with the topoisomerase I inhibitor, irinotecan.



irinotecan (HMDB14900, (Wishart et al. 2018)

However, several adverse effects are associated with its use, including gastrointestinal toxicity (delayed onset diarrhoea) and myelosuppression. Gao et al. used untargeted GC-MS and LC-MS as well as other targeted metabonomics methods to analyse biofluids from rats treated with the drug (Gao et al. 2019). OPLS-DA analysis of pre-dose serum metabolites showed a significant discrimination between sensitive rats displaying adverse drug side effects and non-sensitive rats. The bile acids cholic acid, deoxycholic acid and glycocholic acid together with phenylalanine were predictors for late-onset diarrhoea. The ketogenic amino acids phenylalanine, lysine and tryptophan were predictive of myelosuppression (Gao et al. 2019).

3.5 Not Pharmacometabonomics!

One issue that readers should be aware of is that many studies purporting to be pharmacometabonomics studies are merely metabonomics studies of the effects of drugs and nothing to do with predicting the effects of drug treatment. This growing confusion in the literature is to be regretted and resisted (Balashova et al. 2018; Kaddurah-Daouk et al. 2015).

3.6 Prediction of Interventions Other Than Drug Treatment: Predictive Metabonomics

In the original discovery of pharmacometabonomics, it was envisaged that the methodology would work for interventions other than drug treatment, such as diet changes, physical exercise or even just the passage of time (Clayton et al. 2006). This type of experiment is termed predictive metabonomics rather than pharmacometabonomics. Indeed, pharmacometabonomics is one member of the broader class of predictive metabonomics experiments, where the intervention is drug treatment. Predictive metabonomics has been defined as "the prediction of the outcome of an intervention in an individual based on a mathematical model of pre-intervention metabolite signatures" (Everett 2015). We will now illustrate the application of this prognostic methodology with some recent examples (see Table 2 for a fuller listing).

Pulmonary dysfunction resulting in hypoxaemia is a common complication following cardiac surgery. No predictive biomarkers are available to help identify patients that might suffer from this disease, which is characterised by low partial pressure of oxygen in arterial blood (PaO_2). Maltesen et al. used ^{1}H NMR spectroscopy to study blood serum taken from the pulmonary artery and left atrium of 47 coronary artery bypass graft patients, 16 h after weaning off their cardiopulmonary bypass (Maltesen et al. 2016). At day 3 post-operation, 32 patients had developed hypoxaemia. It was found that levels of carnitine, arachidonic and eicosapentaenoic acid, glycoprotein, citrate, phenylalanine, glycine, plasmalogen, and lysophosphocholine (Lyso-PC) were the most significant in the prediction of day 3 hypoxaemia from day 1 serum analysis. The concentrations of several of these metabolites were found to be individually correlated to day 3, PaO_2 levels. Thus predictive metabonomics methods are capable of prognosing later adverse effects of surgery well before any clinical sign. The results are promising in terms of targeting further treatments to affected patients and also directing research for new drugs to treat this disease on the basis of the perturbed metabolic pathways discovered.

Estimating mortality risk in ageing patients is important for decisions on treatment options. Current methods of mortality prediction are limited and some of the parameters, including systolic blood pressure and total cholesterol show opposite trends in the elderly compared with middle-aged people (Deelan et al. 2019). A predictive method based on metabolite profiles would find great clinical utility. Fischer and co-workers used ^{1}H NMR spectroscopy of the plasma of 9,842 individuals (randomly sampled from the Estonian Biobank) to elucidate that albumin, glycoprotein acetyls, citrate and the mean diameter of VLDL particles are associated with all-cause and cause-specific (cardiovascular and cancer) mortality (Fischer et al. 2014). The group of Deelan et al. recently published the results of a much larger study of 44,168 individuals, 5,512 of who died during follow-up, using ^{1}H NMR spectroscopy of EDTA plasma and serum (Deelan et al. 2019). A set of 14 metabolic biomarkers was found to independently associate with all-cause mortality. A mortality score based on gender and these 14 biomarkers led to an improved risk prediction compared to the conventional risk score. The biomarkers

included albumin, glycoprotein acetyls and mean diameter of VLDL particles, as found by Fischer, but also included acetoacetate, glucose and a number of amino acids. It was concluded that predictive metabonomics methods could be used in the future to guide patient care, if further validated in other clinical settings (Deelan et al. 2019).

4 Conclusions

Metabonomics and predictive metabonomics, including pharmacometabonomics, are starting to have an impact on biomedical and medical research, and in the future it is expected that these technologies will be widely used for both diagnostic and prognostic applications in real clinical settings. Metabolic profiling will be used synergistically with genomic analyses to assist in the delivery of personalised medicine. The approach of metabolite profiling, as opposed to genetic analysis, benefits hugely from the fact that it is a systems biology approach that integrates both genetic and environmental information and gives insights into the real-time status of a subject, as opposed to information on genetic risk factors that may not develop into a disease phenotype. In this context it is encouraging to see the work done by the groups of Kaddurah-Daouk and Weinshilboum on the development of pharmacometabonomics-led pharmacogenomics (Ji et al. 2011; Neavin et al. 2016).

The advent of large biobanks and the development of phenome centres such as those in London, Singapore and Birmingham, amongst others, gives the opportunity for very large-scale clinical studies that will undoubtedly lead to new insights into patient treatments in many disease areas. The great stability and automation capabilities of NMR spectroscopy as a metabolite detection technology are well matched to the task of analysing these huge numbers of samples, as was seen above in the work of Deelan and co-workers on all-cause mortality prediction (Deelan et al. 2019).

One major area that still needs much attention is metabolite identification. This is a significant challenge for both NMR- and MS-based technologies and in spite of the many advances in both areas in recent years, it is still the fact that most metabolites detected in most untargeted metabonomics experiments are unidentified. Progress on this issue promises to enable much more comprehensive biochemical insights into complex organisms including humans and their diseases.

The future for the use of metabolic profiling in clinical pharmacology and medicine in general is very bright.

Acknowledgements I would like to acknowledge productive and enjoyable collaborations with Professor Jeremy Nicholson, Professor John Lindon, Professor Ian Wilson, Professor Elaine Holmes and Professor Elizabeth Shephard over the past 35 years or more. I have gained very much from these stimulating interactions.

Glossary

Area under the curve (AUC) The integral over time of the concentration of a drug in blood plasma: a measure of the exposure of a patient to the drug.

Capillary electrophoresis (CE) An electrophoretic separation methodology based on molecular charge and mobility that can be hyphenated to mass spectrometry.

C_{max} The maximal blood plasma concentration achieved by a drug.

Diagnosis The characterisation of an organism, disease state, phenotype or response to an intervention.

GC Gas chromatography: a powerful method for the separation of volatile compounds. For use in metabonomics, pre-derivatisation of metabolites is required in order to achieve volatility.

HDL High density lipoprotein.

HPLC High performance liquid chromatography: a powerful analytical separation technology often hyphenated with mass spectrometry.

LDL Low density lipoprotein.

Metabolic entropy The degree of disorder of metabolite concentrations in an individual or in a group of subjects.

Metabolic phenotype Multicomponent metabolic characteristics that result from the cumulative interactions of genetic variation, gene products and environmental exposures and that can be related directly to disease risks and therapeutic responses: also known as the metabotype.

Metabolic trajectory The changes in metabolite concentrations over time in response to an intervention.

Metabolite A compound in a biological matrix of an organism that is produced in that organism by an enzymatic pathway.

Metabolome The full set of metabolites within, or that can be secreted from, a biological system such as a cell type or tissue.

Metabolomics Metabolic profiling defined in an observational fashion as "a comprehensive analysis in which all the metabolites of a biological system are identified and quantified".

Metabonome The full set of metabolites contained within an organism, i.e. the sum of all the metabolomes.

Metabonomics Metabolic profiling defined in an experimental fashion as "the quantitative measurement of the multiparametric metabolic response of living systems to pathophysiological stimuli or genetic modification".

Metabotype A probabilistic, multiparametric description of an organism in a given physiological state based on analysis of its cell types, biofluids and tissues: see metabolic phenotype.

Microbiome The collection of microorganisms present both in and on an organism, in a variety of environmental niches.

MS Mass spectrometry: a sensitive analytical methodology for the detection and characterisation of metabolites in biological matrices.

Multivariate analysis: MVA Multivariate (statistical) analysis: a method for the analysis of multiple variables in an experiment or observation at a time and the

simplification of the analysis problem by reduction of the large number of initial variables to a small number of key factors.

NMR spectroscopy Nuclear magnetic resonance spectroscopy: the most powerful method for molecular structure identification in solution, including metabolites in biological fluids.

OPLS-DA Orthogonal projection to latent structures with discriminant analysis: a supervised (and therefore potentially biased) approach to multivariate data analysis with the aim of finding metabolites that are statistically significantly discriminating between two groups, e.g. responders and non-responders, and which also discards metabolite variations that are orthogonal to the group discrimination.

Personalised medicine The use of genomic, molecular and clinical information to select treatments or medicines that are more likely to be both effective and safe for that patient: also known as precision medicine or stratified medicine.

Pharmacogenomics The prediction of the effects of a drug on the basis of individual genetic profiles.

Pharmacokinetics (PK) The measurement of the time course of the absorption, distribution, metabolism and excretion of a drug.

Pharmacometabolomics This term is used synonymously with pharmacometabonomics (see below), but is sometimes erroneously used to describe the investigation of the effects of a drug on an organism: this is just diagnostic metabonomics.

Pharmacometabonomics The prediction of the effects of a drug on the basis of a mathematical model of pre-dose metabolite profiles.

Phenotype The quantitative or qualitative measurement of specific parameters or traits that characterise individual functional biological classes or groups.

Predictive metabolic phenotyping or predictive metabonomics The prediction of the outcome of an intervention in an individual based on a mathematical model of pre-intervention metabolite profiles. The intervention could be a change in diet, exercise, the passage of time, surgical treatment, etc. Pharmacometabonomics is one case of predictive metabonomics, which covers the prognosis of any intervention.

Principal components analysis (PCA) An unsupervised (and therefore unbiased) multivariate statistical method for analysing high dimensional data, such as spectral data from metabonomics experiments. The PCA effects a drastic dimensionality reduction and transformation so that new principal components readily display the variance present in the dataset and therefore patterns in the data like clusters or groupings can be readily discerned and outliers identified.

Prognosis The prediction of disease onset, disease outcome or the outcome of an intervention such as drug treatment.

T2DM Type 2 diabetes mellitus.

Therapeutic index (TI) The TI measures the ratio of the effective dose of a drug for 50% of patients (expressed as ED50) to the toxic dose expressed as the TD50.

Usually a minimal TI of 10 is required in drug development: some companies will aim for a more conservative TI of 30.

UPLC Ultra-performance liquid chromatography: a more efficient and effective form of HPLC using smaller column packings and higher pressures.

VLDL Very low density lipoprotein.

References

Abo R, Hebbring S, Ji Y, Zhu H, Zeng ZB, Batzler A, Jenkins GD, Biernacka J, Snyder K, Drews M, Fiehn O, Fridley B, Schaid D, Kamatani N, Nakamura Y, Kubo M, Mushiroda T, Kaddurah-Daouk R, Mrazek DA, Weinshilboum RM (2012) Merging pharmacometabolomics with pharmacogenomics using '1000 Genomes' single-nucleotide polymorphism imputation: selective serotonin reuptake inhibitor response pharmacogenomics. Pharmacogenet Genomics 22:247–253. https://doi.org/10.1097/FPC.0b013e32835001c9

Alejandro Vargas D, Dario Prieto M, Jose Martinez-Valencia A, Cossio A, Burgess KEV, Burchmore RJS, Adelaida Gomez M (2019) Pharmacometabolomics of meglumine antimoniate in patients with cutaneous leishmaniasis. Front Pharmacol 10:657. https://doi.org/10.3389/fphar.2019.00657

Allalou A, Nalla A, Prentice KJ, Liu Y, Zhang M, Dai FF, Ning X, Osborne LR, Cox BJ, Gunderson EP, Wheeler MB (2016) A predictive metabolic signature for the transition from gestational diabetes mellitus to type 2 diabetes. Diabetes 65:2529–2539. https://doi.org/10.2337/db15-1720

Amin AM, Chin LS, Teh CH, Mostafa H, Noor DAM, Kader M, Hay YK, Ibrahim B (2017) H-1 NMR based pharmacometabolomics analysis of urine identifies metabolic phenotype of clopidogrel high on treatment platelets reactivity in coronary artery disease patients. J Pharm Biomed Anal 146:135–146. https://doi.org/10.1016/j.jpba.2017.08.018

Amin AM, Chin LS, Teh C-H, Mostafa H, Noor DAM, Kader MASKA, Hay YK, Ibrahim B (2018) Pharmacometabolomics analysis of plasma to phenotype clopidogrel high on treatment platelets reactivity in coronary artery disease patients. Eur J Pharm Sci 117:351–361. https://doi.org/10.1016/j.ejps.2018.03.011

Andersson U, Lindberg J, Wang S, Balasubramanian R, Marcusson-Stahl M, Hannula M, Zeng C, Juhasz PJ, Kolmert J, Backstrom J, Nord L, Nilsson K, Martin S, Glinghammar B, Cederbrant K, Schuppe-Koistinen I (2009) A systems biology approach to understanding elevated serum alanine transaminase levels in a clinical trial with ximelagatran. Biomarkers 14:572–586. https://doi.org/10.3109/13547500903261354

Austdal M, Tangeras LH, Skrastad RB, Salvesen KA, Austgulen R, Iversen A-C, Bathen TF (2015) First trimester urine and serum metabolomics for prediction of preeclampsia and gestational hypertension: a prospective screening study. Int J Mol Sci 16:21520–21538. https://doi.org/10.3390/ijms160921520

Backshall A, Sharma R, Clarke SJ, Keun HC (2011) Pharmacometabonomic profiling as a predictor of toxicity in patients with inoperable colorectal cancer treated with capecitabine. Clin Cancer Res 17:3019–3028. https://doi.org/10.1158/1078-0432.ccr-10-2474

Balashova EE, Maslov DL, Lokhov PG (2018) A metabolomics approach to pharmacotherapy personalization. J Pers Med 8:28. https://doi.org/10.3390/jpm8030028

Bawadikji AA, Teh C-H, Sheikh Abdul Kader MAB, Abdul Wahab MJB, Syed Sulaiman SA, Ibrahim B (2019) Plasma metabolites as predictors of warfarin outcome in atrial fibrillation. Am J Cardiovasc Drugs. https://doi.org/10.1007/s40256-019-00364-2

Beckonert O, Keun HC, Ebbels TMD, Bundy JG, Holmes E, Lindon JC, Nicholson JK (2007) Metabolic profiling, metabolomic and metabonomic procedures for NMR spectroscopy of urine, plasma, serum and tissue extracts. Nat Protoc 2:2692–2703. https://doi.org/10.1038/nprot.2007.376

Bharti SK, Roy R (2012) Quantitative H-1 NMR spectroscopy. Trac-Trend Anal Chem 35:5–26. https://doi.org/10.1016/j.trac.2012.02.007

Blasco H, Patin F, Descat A, Garcon G, Corcia P, Gele P, Lenglet T, Bede P, Meininger V, Devos D, Gossens JF, Pradat P-F (2018) A pharmaco-metabolomics approach in a clinical trial of ALS: identification of predictive markers of progression. PLoS One 13:e0198116. https://doi.org/10.1371/journal.pone.0198116

Bro R, Kamstrup-Nielsen MH, Engelsen SB, Savorani F, Rasmussen MA, Hansen L, Olsen A, Tjonneland A, Dragsted LO (2015) Forecasting individual breast cancer risk using plasma metabolomics and biocontours. Metabolomics 11:1376–1380. https://doi.org/10.1007/s11306-015-0793-8

Broadhurst DI, Kell DB (2006) Statistical strategies for avoiding false discoveries in metabolomics and related experiments. Metabolomics 2:171–196. https://doi.org/10.1007/s11306-006-0037-z

Burt T, Nandal S (2016) Pharmacometabolomics in early-phase clinical development. CTS Clin Trans Sci 9:128–138. https://doi.org/10.1111/cts.12396

Cao Z, Miller MS, Lubet RA, Grubbs CJ, Beger RD (2019) Pharmacometabolomic pathway response of effective anticancer agents on different diets in rats with induced mammary tumors. Meta 9. https://doi.org/10.3390/metabo9070149

Chen YH, Xu J, Zhang RP, Abliz Z (2016) Methods used to increase the comprehensive coverage of urinary and plasma metabolomes by MS. Bioanalysis 8:981–997. https://doi.org/10.4155/bio-2015-0010

Clayton T, Lindon J, Cloarec O, Antti H, Charuel C, Hanton G, Provost J, Le Net J, Baker D, Walley R, Everett J, Nicholson J (2006) Pharmaco-metabonomic phenotyping and personalized drug treatment. Nature 440:1073–1077. https://doi.org/10.1038/nature04648

Clayton TA, Baker D, Lindon JC, Everett JR, Nicholson JK (2009) Pharmacometabonomic identification of a significant host-microbiome metabolic interaction affecting human drug metabolism. Proc Natl Acad Sci U S A 106:14728–14733. https://doi.org/10.1073/pnas.0904489106

Coen M, Goldfain-Blanc F, Rolland-Valognes G, Walther B, Robertson DG, Holmes E, Lindon JC, Nicholson JK (2012) Pharmacometabonomic investigation of dynamic metabolic phenotypes associated with variability in response to galactosamine hepatotoxicity. J Proteome Res 11:2427–2440. https://doi.org/10.1021/pr201161f

Condray R, Dougherty GG, Keshavan MS, Reddy RD, Haas GL, Montrose DM, Matson WR, McEvoy J, Kaddurah-Daouk R, Yao JK (2011) 3-Hydroxykynurenine and clinical symptoms in first-episode neuroleptic-naive patients with schizophrenia. Int J Neuropsychopharmacol 14:756–767. https://doi.org/10.1017/s1461145710001689

Cunningham K, Claus SP, Lindon JC, Holmes E, Everett JR, Nicholson JK, Coen M (2012) Pharmacometabonomic characterization of xenobiotic and endogenous metabolic phenotypes that account for inter-individual variation in isoniazid-induced toxicological response. J Proteome Res 11:4630–4642. https://doi.org/10.1021/pr300430u

Dai D, Tian Y, Guo H, Zhang P, Huang Y, Zhang W, Xu F, Zhang Z (2016) A pharmacometabonomic approach using predose serum metabolite profiles reveals differences in lipid metabolism in survival and non-survival rats treated with lipopolysaccharide. Metabolomics 12. https://doi.org/10.1007/s11306-015-0892-6

de Oliveira FA, Shahin MH, Gong Y, McDonough CW, Beitelshees AL, Gums JG, Chapman AB, Boerwinkle E, Turner ST, Frye RF, Fiehn O, Kaddurah-Daouk R, Johnson JA, Cooper-DeHoff RM (2016) Novel plasma biomarker of atenolol-induced hyperglycemia identified through a metabolomics-genomics integrative approach. Metabolomics 12. https://doi.org/10.1007/s11306-016-1076-8

Deelan J et al (2019) A metabolic profile of all-cause mortality risk identified in an observational study of 44,168 individuals. Nat Commun 10:1–8. https://doi.org/10.1038/s41467-019-11311-9

Dona AC, Kyriakides M, Scott F, Shephard EA, Varshavi D, Veselkov K, Everett JR (2016) A guide to the identification of metabolites in NMR-based metabonomics/metabolomics

experiments. Comput Struct Biotechnol J 14:135–153. https://doi.org/10.1016/j.csbj.2016.02.005

Dong B, Jia J, Hu W, Chen Q, Jiang C, Pan J, Huang Y, Xue W, Gao H (2013) Application of H-1 NMR metabonomics in predicting renal function recoverability after the relief of obstructive uropathy in adult patients. Clin Biochem 46:346–353. https://doi.org/10.1016/j.clinbiochem.2012.11.012

Ellero-Simatos S, Lewis JP, Georgiades A, Yerges-Armstrong LM, Beitelshees AL, Horenstein RB, Dane A, Harms AC, Ramaker R, Vreeken RJ, Perry CG, Zhu H, Sanchez CL, Kuhn C, Ortel TL, Shuldiner AR, Hankemeier T, Kaddurah-Daouk R (2014) Pharmacometabolomics reveals that serotonin is implicated in aspirin response variability. CPT Pharmacometrics Syst Pharmacol 3:e125. https://doi.org/10.1038/psp.2014.22

Emwas A-H, Luchinat C, Turano P, Tenori L, Roy R, Salek RM, Ryan D, Merzaban JS, Kaddurah-Daouk R, Zeri AC, Gowda GAN, Raftery D, Wang Y, Brennan L, Wishart DS (2015) Standardizing the experimental conditions for using urine in NMR-based metabolomic studies with a particular focus on diagnostic studies: a review. Metabolomics 11:872–894. https://doi.org/10.1007/s11306-014-0746-7

Evans CR, Karnovsky A, Puskarich MA, Michailidis G, Jones AE, Stringer KA (2019) Untargeted metabolomics differentiates L-carnitine treated Septic shock 1-year survivors and nonsurvivors. J Proteome Res 18:2004–2011. https://doi.org/10.1021/acs.jproteome.8b00774

Everett JR (2015) Pharmacometabonomics in humans: a new tool for personalized medicine. Pharmacogenomics 16:737–754. https://doi.org/10.2217/pgs.15.20

Everett JR (2016) From metabonomics to pharmacometabonomics: the role of metabolic profiling in personalized medicine. Front Pharmacol 7:15. https://doi.org/10.3389/fphar.2016.00297

Everett JR, Loo RL, Pullen FS (2013) Pharmacometabonomics and personalized medicine. Ann Clin Biochem 50:523–545. https://doi.org/10.1177/0004563213497929

Everett JR, Lindon JC, Nicholson JK (2016) Pharmacometabonomics and predictive metabonomics: new tools for personalized medicine. In: Holmes E, Jeremy K, Darzi AW, Lindon JC (eds) Metabolic phenotyping in personalized and public healthcare. Academic Press, London, pp 138–165

Everett JR, Holmes E, Veselkov KA, Lindon JC, Nicholson JK (2019) A unified conceptual framework for metabolic phenotyping in diagnosis and prognosis. Trend Pharmacol Sci. https://doi.org/10.1016/j.tips.2019.08.004

Fiehn O (2002) Metabolomics--the link between genotypes and phenotypes. Plant Mol Biol 48:155–171

Fischer K, Kettunen J, Wurtz P, Haller T, Havulinna AS, Kangas AJ, Soininen P, Esko T, Tammesoo M-L, Maegi R, Smit S, Palotie A, Ripatti S, Salomaa V, Ala-Korpela M, Perola M, Metspalu A (2014) Biomarker profiling by nuclear magnetic resonance spectroscopy for the prediction of all-cause mortality: an observational study of 17,345 persons. PLoS Med 11:e1001606. https://doi.org/10.1371/journal.pmed.1001606

Gamage N, Barnett A, Hempel N, Duggleby RG, Windmill KF, Martin JL, McManus ME (2006) Human sulfotransferases and their role in chemical metabolism. Toxicol Sci 90:5–22. https://doi.org/10.1093/toxsci/kfj061

Gao Y, Li W, Chen J, Wang X, Lv Y, Huang Y, Zhang Z, Xu F (2019) Pharmacometabolomic prediction of individual differences of gastrointestinal toxicity complicating myelosuppression in rats induced by irinotecan. Acta Pharm Sin B 9:157–166. https://doi.org/10.1016/j.apsb.2018.09.006

Gowda GAN, Raftery D (2017) Recent advances in NMR-based metabolomics. Anal Chem 89:490–510. https://doi.org/10.1021/acs.analchem.6b04420

Gupta M, Neavin D, Liu D, Biernacka J, Hall-Flavin D, Bobo WV, Frye MA, Skime M, Jenkins GD, Batzler A, Kalari K, Matson W, Bhasin SS, Zhu H, Mushiroda T, Nakamura Y, Kubo M, Wang L, Kaddurah-Daouk R, Weinshilboum RM (2016) TSPAN5, ERICH3 and selective serotonin reuptake inhibitors in major depressive disorder: pharmacometabolomics-informed pharmacogenomics. Mol Psychiatry 21:1717–1725. https://doi.org/10.1038/mp.2016.6

Hao D, Sarfaraz MO, Farshidfar F, Bebb DG, Lee CY, Card CM, David M, Weljie AM (2016a) Temporal characterization of serum metabolite signatures in lung cancer patients undergoing treatment. Metabolomics 12:58. https://doi.org/10.1007/s11306-016-0961-5

Hao D, Sarfaraz MO, Farshidfar F, Bebb DG, Lee CY, Card CM, David M, Weljie AM (2016b) Temporal characterization of serum metabolite signatures in lung cancer patients undergoing treatment (vol 12, 58, 2016). Metabolomics 12:122. https://doi.org/10.1007/s11306-016-1068-8

He C, Liu Y, Wang Y, Tang J, Tan Z, Li X, Chen Y, Huang Y, Chen X, Ouyang D, Zhou H, Peng J (2018) H-1 NMR based pharmacometabolomics analysis of metabolic phenotype on predicting metabolism characteristics of losartan in healthy volunteers. J Chromatogr B Anal Technol Biomed Life Sci 1095:15–23. https://doi.org/10.1016/j.jchromb.2018.07.016

Huang Q, Aa J, Jia H, Xin X, Tao C, Liu L, Zou B, Song Q, Shi J, Cao B, Yong Y, Wang G, Zhou G (2015) A pharmacometabonomic approach to predicting metabolic phenotypes and pharmacokinetic parameters of atorvastatin in healthy volunteers. J Proteome Res 14:3970–3981. https://doi.org/10.1021/acs.jproteome.5b00440

Ji Y, Hebbring S, Zhu H, Jenkins GD, Biernacka J, Snyder K, Drews M, Fiehn O, Zeng Z, Schaid D, Mrazek DA, Kaddurah-Daouk R, Weinshilboum RM (2011) Glycine and a glycine dehydrogenase (GLDC) SNP as citalopram/escitalopram response biomarkers in depression: pharmacometabolomics-informed pharmacogenomics. Clin Pharmacol Ther 89:97–104. https://doi.org/10.1038/clpt.2010.250

Jiang L, Lee SC, Ng TC (2018) Pharmacometabonomics analysis reveals serum formate and acetate potentially associated with varying response to gemcitabine-carboplatin chemotherapy in metastatic breast cancer patients. J Proteome Res 17:1248. https://doi.org/10.1021/acs.jproteome.7b00859

Jove M, Mauri-Capdevila G, Suarez I, Cambray S, Sanahuja J, Quilez A, Farre J, Benabdelhak I, Pamplona R, Portero-Otin M, Purroy F (2015) Metabolomics predicts stroke recurrence after transient ischemic attack. Neurology 84:36–45

Kaddurah-Daouk R, Baillie RA, Zhu HJ, Zeng ZB, Wiest MM, Nguyen UT, Watkins SM, Krauss RM (2010) Lipidomic analysis of variation in response to simvastatin in the cholesterol and pharmacogenetics study. Metabolomics 6:191–201. https://doi.org/10.1007/s11306-010-0207-x

Kaddurah-Daouk R, Boyle SH, Matson W, Sharma S, Matson S, Zhu H, Bogdanov MB, Churchill E, Krishnan RR, Rush AJ, Pickering E, Delnomdedieu M (2011) Pretreatment metabotype as a predictor of response to sertraline or placebo in depressed outpatients: a proof of concept. Transl Psychiatry 1:1–7. https://doi.org/10.1038/tp.2011.22

Kaddurah-Daouk R, Bogdanov MB, Wikoff WR, Zhu H, Boyle SH, Churchill E, Wang Z, Rush AJ, Krishnan RR, Pickering E, Delnomdedieu M, Fiehn O (2013) Pharmacometabolomic mapping of early biochemical changes induced by sertraline and placebo. Transl Psychiatry 3:e223. https://doi.org/10.1038/tp.2012.142

Kaddurah-Daouk R, Weinshilboum R, Pharmacometabolomics Res N (2015) Metabolomic signatures for drug response phenotypes: pharmacometabolomics enables precision medicine. Clin Pharmacol Ther 98:71–75. https://doi.org/10.1002/cpt.134

Kaddurah-Daouk R, Hankemeier T, Scholl EH, Baillie R, Harms A, Stage C, Dalhoff KP, Jurgens G, Taboureau O, Nzabonimpa GS, Motsinger-Reif AA, Thomsen R, Linnet K, Rasmussen HB, INDICES Consortium, Pharmacometabolomics Research Network (2018) Pharmacometabolomics informs about pharmacokinetic profile of methylphenidate. CPT Pharmacometrics Syst Pharmacol 7:525–533. https://doi.org/10.1002/psp4.12309

Kapoor SR, Filer A, Fitzpatrick MA, Fisher BA, Taylor PC, Buckley CD, McInnes IB, Raza K, Young SP (2013) Metabolic profiling predicts response to anti-tumor necrosis factor alpha therapy in patients with rheumatoid arthritis. Arthritis Rheum 65:1448–1456. https://doi.org/10.1002/art.37921

Karas-Kuzelicki N, Smid A, Tamm R, Metspalu A, Mlinaric-Rascan I (2014) From pharmacogenetics to pharmacometabolomics: SAM modulates TPMT activity. Pharmacogenomics 15:1437–1449. https://doi.org/10.2217/pgs.14.84

Keun HC, Sidhu J, Pchejetski D, Lewis JS, Marconell H, Patterson M, Bloom SR, Amber V, Coombes RC, Stebbing J (2009) Serum molecular signatures of weight change during early

breast cancer chemotherapy. Clin Cancer Res 15:6716–6723. https://doi.org/10.1158/1078-0432.ccr-09-1452

Kienana M, Benz-de Bretagne I, Nadal-Desbarats L, Blasco H, Gyan E, Choquet S, Montigny F, Emond P, Barin-Le Guellec C (2016) Endogenous metabolites that are substrates of Organic Anion Transporter's (OATs) predict methotrexate clearance. Pharmacol Res 118:121. https://doi.org/10.1016/j.phrs.2016.05.021

Kim B, Lee JW, Hong KT, Yu KS, Jang IJ, Park KD, Shin HY, Ahn HS, Cho JY, Kang HJ (2017) Pharmacometabolomics for predicting variable busulfan exposure in paediatric haematopoietic stem cell transplantation patients. Sci Rep 7:1711. https://doi.org/10.1038/s41598-017-01861-7

Kind T, Fiehn O (2010) Advances in structure elucidation of small molecules using mass spectrometry. Bioanal Rev 2:23–60. https://doi.org/10.1007/s12566-010-0015-9

Kwon HN, Kim M, Wen H, Kang S, Yang H-J, Choi M-J, Lee HS, Choi D, Park IS, Suh YJ, Hong S-S, Park S (2011) Predicting idiopathic toxicity of cisplatin by a pharmacometabonomic approach. Kidney Int 79:529–537. https://doi.org/10.1038/ki.2010.440

Lazarou J, Pomeranz BH, Corey PN (1998) Incidence of adverse drug reactions in hospitalized patients - a meta-analysis of prospective studies. JAMA 279:1200–1205. https://doi.org/10.1001/jama.279.15.1200

Lee JW, Aminkeng F, Bhavsar AP, Shaw K, Carleton BC, Hayden MR, Ross CJD (2014) The emerging era of pharmacogenomics: current successes, future potential, and challenges. Clin Genet 86:21–28. https://doi.org/10.1111/cge.12392

Lee J, Yoon SH, Yi S, Kim AH, Kim B, Lee S, Yu K-S, Jang I-J, Cho J-Y (2019) Quantitative prediction of hepatic CYP3A activity using endogenous markers in healthy subjects after administration of CYP3A inhibitors or inducers. Drug Metab Pharmacokinet 34:247–252. https://doi.org/10.1016/j.dmpk.2019.04.002

Lewis JP, Yerges-Armstrong LM, Ellero-Simatos S, Georgiades A, Kaddurah-Daouk R, Hankemeier T (2013) Integration of pharmacometabolomic and pharmacogenomic approaches reveals novel insights into antiplatelet therapy. Clin Pharmacol Ther 94:570–573. https://doi.org/10.1038/clpt.2013.153

Li H, Ni Y, Su M, Qiu Y, Zhou M, Qiu M, Zhao A, Zhao L, Jia W (2007) Pharmacometabonomic phenotyping reveals different responses to xenobiotic intervention in rats. J Proteome Res 6:1364–1370. https://doi.org/10.1021/pr060513q

Lin YS, Kerr SJ, Randolph T, Shireman LM, Senn T, McCune JS (2016) Prediction of intravenous busulfan clearance by endogenous plasma biomarkers using global pharmacometabolomics. Metabolomics 12:161. https://doi.org/10.1007/s11306-016-1106-6

Lindon JC, Wilson ID (2016) The development of metabolic phenotyping - a historical perspective. In: Holmes E, Nicholson JK, Darzi A, Lindon JC (eds) Metabolic phenotyping in personalized and public healthcare. Elsevier, Oxford pp 17–48

Lindon J, Nicholson J, Holmes E, Everett J (2000) Metabonomics: metabolic processes studied by NMR spectroscopy of biofluids. Concepts Magn Reson 12:289–320. https://doi.org/10.1002/1099-0534(2000)12:5<289::AID-CMR3>3.0.CO;2-W

Lindon JC, Nicholson JK, Holmes E (2007) The handbook of metabonomics and metabolomics. Elsevier, Amsterdam

Lindon J, Nicholson JK, Holmes E (2019) The handbook of metabolic phenotyping. Elsevier, Oxford

Liu L, Cao B, Aa J, Zheng T, Shi J, Li M, Wang X, Zhao C, Xiao W, Yu X, Sun R, Gu R, Zhou J, Wu L, Hao G, Zhu X, Wang G (2012) Prediction of the pharmacokinetic parameters of Triptolide in rats based on endogenous molecules in pre-dose baseline serum. PLoS One 7:e43389. https://doi.org/10.1371/journal.pone.0043389

Maltesen RG, Hanifa MA, Kucheryavskiy S, Pedersen S, Kristensen SR, Rasmussen BS, Wimmer R (2016) Predictive biomarkers and metabolic hallmark of postoperative hypoxaemia. Metabolomics 12:87. https://doi.org/10.1007/s11306-016-1018-5

Markley JL, Brüschweiler R, Edison AS, Eghbalnia HR, Powers R, Raftery D, Wishart DS (2017) The future of NMR-based metabolomics. Curr Opin Biotechnol 43:34–40. https://doi.org/10.1016/j.copbio.2016.08.001

Martinez-Avila JC, Garcia Bartolome A, Garcia I, Dapia I, Tong HY, Diaz L, Guerra P, Frias J, Carcas Sansuan AJ, Borobia AM (2018a) Pharmacometabolomics applied to zonisamide pharmacokinetic parameter prediction (vol 14, 70, 2018). Metabolomics 14:87. https://doi.org/10.1007/s11306-018-1378-0

Martinez-Avila JC, Garcia-Bartolome A, Garcia I, Dapia I, Tong HY, Diaz L, Guerra P, Frias J, Carcas Sansuan AJ, Borobia AM (2018b) Pharmacometabolomics applied to zonisamide pharmacokinetic parameter prediction. Metabolomics 14:70. https://doi.org/10.1007/s11306-018-1365-5

McPhail MJW, Shawcross DL, Lewis MR, Coltart I, Want EJ, Antoniades CG, Veselkov K, Triantafyllou E, Patel V, Pop O, Gomez-Romero M, Kyriakides M, Zia R, Abeles RD, Crossey MME, Jassem W, O'Grady J, Heaton N, Auzinger G, Bernal W, Quaglia A, Coen M, Nicholson JK, Wendon JA, Holmes E, Taylor-Robinson SD (2016) Multivariate metabotyping of plasma predicts survival in patients with decompensated cirrhosis. J Hepatol 64:1058–1067. https://doi.org/10.1016/j.jhep.2016.01.003

Miolo G, Muraro E, Caruso D, Crivellari D, Ash A, Scalone S, Lombardi D, Rizzolio F, Giordano A, Corona G (2016) Phamacometabolomics study identifies circulating spermidine and tryptophan as potential biomarkers associated with the complete pathological response to trastuzumab-paclitaxel neoadjuvant therapy in HER-2 positive breast cancer. Oncotarget 7:39809. https://doi.org/10.18632/oncotarget.9489

Nam HW, Karpyak VM, Hinton DJ, Geske JR, Ho AMC, Prieto ML, Biernacka JM, Frye MA, Weinshilboum RM, Choi DS (2015) Elevated baseline serum glutamate as a pharmacometabolomic biomarker for acamprosate treatment outcome in alcohol-dependent subjects. Transl Psychiatry 5:e621. https://doi.org/10.1038/tp.2015.120

Navarro SL, Randolph TW, Shireman LM, Raftery D, McCune JS (2016) Pharmacometabonomic prediction of busulfan clearance in hematopoetic cell transplant recipients. J Proteome Res 15:2802–2811. https://doi.org/10.1021/acs.jproteome.6b00370

Neavin D, Kaddurah-Daouk R, Weinshilboum R (2016) Pharmacometabolomics informs pharmacogenomics. Metabolomics 12:121. https://doi.org/10.1007/s11306-016-1066-x

Ni Y, Zhao L, Yu H, Ma X, Bao Y, Rajani C, Loo LM, Shvetsov YB, Yu H, Chen T, Zhang Y, Wang C, Hu C, Su M, Xie G, Zhao A, Jia W, Jia W (2015) Circulating unsaturated fatty acids delineate the metabolic status of obese individuals. EBioMedicine 2:1513–1522. https://doi.org/10.1016/j.ebiom.2015.09.004

Nicholson JK, Wilson ID, Lindon JC (2011) Pharmacometabonomics as an effector for personalized medicine. Pharmacogenomics 12:103–111. https://doi.org/10.2217/pgs.10.157

Nicholson JK, Darzi A, Holmes E, Lindon JC (eds) (2016) Metabolic phenotyping in personalized and public healthcare. Academic Press, London

Oh J, Yi S, Gu N, Shin D, Yu K-S, Yoon SH, Cho J-Y, Jang I-J (2018) Utility of integrated analysis of pharmacogenomics and pharmacometabolomics in early phase clinical trial: a case study of a new molecular entity. Genom Inform 16:52–58. https://doi.org/10.5808/gi.2018.16.3.52

Park J-E, Jeong G-H, Lee I-K, Yoon Y-R, Liu K-H, Gu N, Shin K-H (2018) A pharmacometabolomic approach to predict response to metformin in early-phase type 2 diabetes mellitus patients. Molecules 23:E1579. https://doi.org/10.3390/molecules23071579

Phapale PB, Kim SD, Lee HW, Lim M, Kale DD, Kim YL, Cho JH, Hwang D, Yoon YR (2010) An integrative approach for identifying a metabolic phenotype predictive of individualized pharmacokinetics of tacrolimus. Clin Pharmacol Ther 87:426–436. https://doi.org/10.1038/clpt.2009.296

Phua LC, Goh S, Tai DWM, Leow WQ, Alkaff SMF, Chan CY, Kam JH, Lim TKH, Chan ECY (2017) Metabolomic prediction of treatment outcome in pancreatic ductal adenocarcinoma patients receiving gemcitabine. Cancer Chemother Pharmacol 81:277. https://doi.org/10.1007/s00280-017-3475-6

Phua LC, Goh S, Tai DWM, Leow WQ, Alkaff SMF, Chan CY, Kam JH, Lim TKH, Chan ECY (2018) Metabolomic prediction of treatment outcome in pancreatic ductal adenocarcinoma patients receiving gemcitabine. Cancer Chemother Pharmacol 81:277–289. https://doi.org/10.1007/s00280-017-3475-6

Pirmohamed M (2014) Personalized pharmacogenomics: predicting efficacy and adverse drug reactions. Annu Rev Genomics Hum Genet 15:349–370. https://doi.org/10.1146/annurev-genom-090413-025419

Puskarich MA, Finkel MA, Karnovsky A, Jones AE, Trexel J, Harris BN, Stringer KA (2015) Pharmacometabolomics of l-carnitine treatment response phenotypes in patients with septic shock. Ann Am Thorac Soc 12:46–56. https://doi.org/10.1513/AnnalsATS.201409-415OC

Puskarich MA, Evans CR, Karnovsky A, Das AK, Jones AE, Stringer KA (2018) Septic shock nonsurvivors have persistently elevated acylcarnitines following carnitine supplementation. Shock 49:412–419. https://doi.org/10.1097/shk.0000000000000997

Rahmioglu N, Le Gall G, Heaton J, Kay KL, Smith NW, Colquhoun IJ, Ahmadi KR, Kemsley EK (2011) Prediction of variability in CYP3A4 induction using a combined H-1 NMR metabonomics and targeted UPLC-MS approach. J Proteome Res 10:2807–2816. https://doi.org/10.1021/pr200077n

Reverter E, Lozano JJ, Alonso C, Berzigotti A, Seijo S, Turon F, Baiges A, Martinez-Chantar ML, Mato JM, Martinez-Arranz I, La Mura V, Hernandez-Gea V, Bosch J, Garcia-Pagan JC (2019) Metabolomics discloses potential biomarkers to predict the acute HVPG response to propranolol in patients with cirrhosis. Liver Int 39:705. https://doi.org/10.1111/liv.14042

Rotroff DM, Shahin MH, Gurley SB, Zhu H, Motsinger-Reif A, Meisner M, Beitelshees AL, Fiehn O, Johnson JA, Elbadawi-Sidhu M, Frye RF, Gong Y, Weng L, Cooper-DeHoff RM, Kaddurah-Daouk R (2015) Pharmacometabolomic assessments of atenolol and hydrochlorothiazide treatment reveal novel drug response phenotypes. CPT Pharmacometrics Syst Pharmacol 4:669–679. https://doi.org/10.1002/psp4.12017

Salari K, Watkins H, Ashley EA (2012) Personalized medicine: hope or hype? Eur Heart J 33:1564–1570. https://doi.org/10.1093/eurheartj/ehs112

Scalbert A, Brennan L, Fiehn O, Hankemeier T, Kristal BS, van Ommen B, Pujos-Guillot E, Verheij E, Wishart D, Wopereis S (2009) Mass-spectrometry-based metabolomics: limitations and recommendations for future progress with particular focus on nutrition research. Metabolomics 5:435–458. https://doi.org/10.1007/s11306-009-0168-0

Shah RR, Smith RL (2015) Addressing phenoconversion: the Achilles' heel of personalized medicine. Br J Clin Pharmacol 79:222–240. https://doi.org/10.1111/bcp.12441

Shahin MH, Gong Y, Frye RF, Rotroff DM, Beitelshees AL, Baillie RA, Chapman AB, Gums JG, Turner ST, Boerwinkle E, Motsinger-Reif A, Fiehn O, Cooper-DeHoff RM, Han X, Kaddurah-Daouk R, Johnson JA (2017) Sphingolipid metabolic pathway impacts thiazide diuretics blood pressure response: insights from genomics, metabolomics, and lipidomics. J Am Heart Assoc 7: e006656. https://doi.org/10.1161/JAHA.117.006656

Shin KH, Choi MH, Lim KS, Yu KS, Jang IJ, Cho JY (2013) Evaluation of endogenous metabolic markers of hepatic CYP3A activity using metabolic profiling and midazolam clearance. Clin Pharmacol Ther 94:601–609. https://doi.org/10.1038/clpt.2013.128

Shin KH, Ahn LY, Choi MH, Moon JY, Lee J, Jang IJ, Yu KS, Cho JY (2016) Urinary 6β-hydroxycortisol/cortisol ratio most highly correlates with midazolam clearance under hepatic CYP3A inhibition and induction in females: a pharmacometabolomics approach. AAPS J 18:1254–1261. https://doi.org/10.1208/s12248-016-9941-y

Sjoberg RL, Bergenheim T, Moren L, Antti H, Lindgren C, Naredi S, Lindvall P (2015) Blood metabolomic predictors of 1-year outcome in subarachnoid hemorrhage. Neurocrit Care 23:225–232. https://doi.org/10.1007/s12028-014-0089-2

Stebbing J, Sharma A, North B, Athersuch TJ, Zebrowski A, Pchejetski D, Coombes RC, Nicholson JK, Keun HC (2012) A metabolic phenotyping approach to understanding relationships between metabolic syndrome and breast tumour responses to chemotherapy. Ann Oncol 23:860–866. https://doi.org/10.1093/annonc/mdr347

Sun Y, Kim JH, Vangipuram K, Hayes DF, Smith EML, Yeomans L, Henry NL, Stringer KA, Hertz DL (2018) Pharmacometabolomics reveals a role for histidine, phenylalanine, and threonine in the development of paclitaxel-induced peripheral neuropathy. Breast Cancer Res Treat 171:657–666. https://doi.org/10.1007/s10549-018-4862-3

Tan GG, Zhao BB, Li YQ, Liu X, Zou ZL, Wan J, Yao Y, Xiong H, Wang YY (2017) Pharmacometabolomics identifies dodecanamide and leukotriene B4 dimethylamide as a predictor of chemosensitivity for patients with acute myeloid leukemia treated with cytarabine and anthracycline. Oncotarget 8:88697–88707. https://doi.org/10.18632/oncotarget.20733

Trupp M, Zhu H, Wikoff WR, Baillie RA, Zeng ZB, Karp PD, Fiehn O, Krauss RM, Kaddurah-Daouk R (2012) Metabolomics reveals amino acids contribute to variation in response to simvastatin treatment. PLoS One 7:e38386. https://doi.org/10.1371/journal.pone.0038386

Wang TJ, Larson MG, Vasan RS, Cheng S, Rhee EP, McCabe E, Lewis GD, Fox CS, Jacques PF, Fernandez C, O'Donnell CJ, Carr SA, Mootha VK, Florez JC, Souza A, Melander O, Clish CB, Gerszten RE (2011) Metabolite profiles and the risk of developing diabetes. Nat Med 17:448–453. https://doi.org/10.1038/nm.2307

Wang-Sattler R, Yu Z, Herder C, Messias AC, Floegel A, He Y, Heim K, Campillos M, Holzapfel C, Thorand B, Grallert H, Xu T, Bader E, Huth C, Mittelstrass K, Doering A, Meisinger C, Gieger C, Prehn C, Roemisch-Margl W, Carstensen M, Xie L, Yamanaka-Okumura H, Xing G, Ceglarek U, Thiery J, Giani G, Lickert H, Lin X, Li Y, Boeing H, Joost H-G, de Angelis MH, Rathmann W, Suhre K, Prokisch H, Peters A, Meitinger T, Roden M, Wichmann HE, Pischon T, Adamski J, Illig T (2012) Novel biomarkers for pre-diabetes identified by metabolomics. Mol Syst Biol 8:615. https://doi.org/10.1038/msb.2012.43

Waters E, Wang JH, Redmond HP, Wu QD, Kay E, Bouchier-Hayes D (2001) Role of taurine in preventing acetaminophen-induced hepatic injury in the rat. Am J Physiol Gastrointest Liv Physiol 280:G1274–G1279

Watson DG (2013) A rough guide to metabolite identification using high resolution liquid chromatography mass spectrometry in metabolomic profiling in metazoans. Comput Struct Biotechnol J 4:1–10. https://doi.org/10.5936/csbj.201301005

Wehrens R, Salek R (2019) Metabolomics: practical guide to design and analysis, 1st edn. Chapman and Hall/CRC Press, Boca Raton

Weng L, Gong Y, Culver J, Gardell SJ, Petucci C, Morse AM, Frye RF, Turner ST, Chapman A, Boerwinkle E, Gums J, Beitelshees AL, Borum PR, Johnson JA, Garrett TJ, McIntyre LM, Cooper-DeHoff RM (2016) Presence of arachidonoyl-carnitine is associated with adverse cardiometabolic responses in hypertensive patients treated with atenolol. Metabolomics 12. https://doi.org/10.1007/s11306-016-1098-2

Wilson ID (2009) Drugs, bugs, and personalized medicine: pharmacometabonomics enters the ring. Proc Natl Acad Sci U S A 106:14187–14188

Wilson ID (2015) In: Everett JR, Lindon JC, Harris RK (eds) NMR in pharmaceutical science, 1st edn. Wiley, Hoboken

Winnike JH, Li Z, Wright FA, Macdonald JM, O'Connell TM, Watkins PB (2010) Use of pharmaco-metabonomics for early prediction of acetaminophen-induced hepatotoxicity in humans. Clin Pharmacol Ther 88:45–51. https://doi.org/10.1038/clpt.2009.240

Wishart DS (2016) Emerging applications of metabolomics in drug discovery and precision medicine. Nat Rev Drug Discov 15:473–484. https://doi.org/10.1038/nrd.2016.32

Wishart DS, Feunang YD, Marcu A, Guo AC, Liang K, Vazquez-Fresno R, Sajed T, Johnson D, Li CR, Karu N, Sayeeda Z, Lo E, Assempour N, Berjanskii M, Singhal S, Arndt D, Liang YJ, Badran H, Grant J, Serra-Cayuela A, Liu YF, Mandal R, Neveu V, Pon A, Knox C, Wilson M, Manach C, Scalbert A (2018). HMDB 4.0: the human metabolome database for 2018) Nucleic Acids Res 46:D608–D617. https://doi.org/10.1093/nar/gkx1089

Yerges-Armstrong LM, Ellero-Simatos S, Georgiades A, Zhu H, Lewis JP, Horenstein RB, Beitelshees AL, Dane A, Reijmers T, Hankemeier T, Fiehn O, Shuldiner AR, Kaddurah-Daouk R, Pharmacometabolomics Research Network (2013) Purine pathway

implicated in mechanism of resistance to aspirin therapy: pharmacometabolomics-informed pharmacogenomics. Clin Pharmacol Ther 94:525–532. https://doi.org/10.1038/clpt.2013.119

Zhang P, Li W, Chen J, Li R, Zhang Z, Huang Y, Xu F (2017a) Branched-chain amino acids as predictors for individual differences of cisplatin nephrotoxicity in rats: a pharmacometabonomics study. J Proteome Res 16:1753. https://doi.org/10.1021/acs.jproteome.7b00014

Zhang ZX, Gu H, Zhao HZ, Liu YH, Fu S, Wang ML, Zhou WJ, Xie ZY, Yu HH, Huang ZH, Gao XY (2017b) Pharmacometabolomics in endogenous drugs: a new approach for predicting the individualized pharmacokinetics of cholic acid. J Proteome Res 16:3529–3535. https://doi.org/10.1021/acs.jproteome.7b00218

Zhu H, Bogdanov MB, Boyle SH, Matson W, Sharma S, Matson S, Churchill E, Fiehn O, Rush JA, Krishnan RR, Pickering E, Delnomdedieu M, Kaddurah-Daouk R, Network PR (2013) Pharmacometabolomics of response to sertraline and to placebo in major depressive disorder - possible role for methoxyindole pathway. PLoS One 8:e68283. https://doi.org/10.1371/journal.pone.0068283

The Microbiome and Its Potential for Pharmacology

Aries Chavira, Pedro Belda-Ferre, Tomasz Kosciolek, Farhana Ali, Pieter C. Dorrestein, and Rob Knight

Contents

1 Introduction 302
2 What Is the Human Microbiome and What Impacts It? 304
3 What Diseases and Biological Processes Is the Human Microbiome Linked to? 306

A. Chavira
Division of Biological Sciences, University of California San Diego, La Jolla, CA, USA

P. Belda-Ferre
Department of Pediatrics, University of California San Diego, La Jolla, CA, USA

T. Kosciolek
Department of Pediatrics, University of California San Diego, La Jolla, CA, USA

Małopolska Centre of Biotechnology, Jagiellonian University, Krakow, Poland

F. Ali
Division of Gastroenterology, Department of Pediatrics, University of California San Diego, La Jolla, CA, USA

P. C. Dorrestein
Department of Pediatrics, University of California San Diego, La Jolla, CA, USA

Skaggs School of Pharmacy and Pharmaceutical Sciences, University of California San Diego, La Jolla, CA, USA

Collaborative Mass Spectrometry Innovation Center, University of California San Diego, La Jolla, CA, USA

Center for Microbiome Innovation, University of California San Diego, La Jolla, CA, USA

R. Knight (✉)
Department of Pediatrics, University of California San Diego, La Jolla, CA, USA

Center for Microbiome Innovation, University of California San Diego, La Jolla, CA, USA

Department of Computer Science and Engineering, University of California San Diego, La Jolla, CA, USA

Department of Bioengineering, University of California San Diego, La Jolla, CA, USA
e-mail: robknight@ucsd.edu

J. E. Barrett et al. (eds.), *Concepts and Principles of Pharmacology*,
Handbook of Experimental Pharmacology 260, https://doi.org/10.1007/164_2019_317

4 How Do We Find Out About the Microbiome and Metabolome? 310
5 How Do Drugs Affect the Microbiome? 312
6 How Does the Microbiome Affect Drug Metabolism or Activity? 313
7 How Can Drugs Be Isolated from the Microbiome? 314
8 Conclusions 316
References 316

Abstract

The human microbiota (the microscopic organisms that inhabit us) and microbiome (their genes) hold considerable potential for improving pharmacological practice. Recent advances in multi-"omics" techniques have dramatically improved our understanding of the constituents of the microbiome and their functions. The implications of this research for human health, including microbiome links to obesity, drug metabolism, neurological diseases, cancer, and many other health conditions, have sparked considerable interest in exploiting the microbiome for targeted therapeutics. Links between microbial pathways and disease states further highlight a rich potential for companion diagnostics and precision medicine approaches. For example, the success of fecal microbiota transplantation to treat *Clostridium difficile* infection has already started to redefine standard of care with a microbiome-directed therapy. In this review we briefly discuss the nature of human microbial ecosystems and with pathologies and biological processes linked to the microbiome. We then review emerging computational metagenomic, metabolomic, and wet lab techniques researchers are using today to learn about the roles host-microbial interactions have with respect to pharmacological purposes and vice versa. Finally, we describe how drugs affect the microbiome, how the microbiome can impact drug response in different people, and the potential of the microbiome itself as a source of new therapeutics.

Keywords

Drug discovery · Immunology · Immunotherapy · Live biotherapeutics · Metabolism · Microbiology · Microbiome · Microbiota · Patient stratification · Precision medicine

1 Introduction

From the skin on our bodies to the contents of our digestive tract, the human body teems with microbes (the microbiota). According to current estimates (Sender et al. 2016), this vast consortium consists of ~40 trillion microbial cells (outnumbering our ~30 trillion human cells) and harbors a microbiome of 2–20 million microbial genes, vastly outnumbering our ~20,000 "human" genes in our germline genomes (Qin et al. 2010). These ecosystems of commensal microbes perform countless metabolic pathways and produce trillions of metabolites (loosely termed the metabolome) (Fischbach 2018). The microbiome and metabolome form an enormously complex

interrelationship with the human body. Researchers have barely scratched the surface of the parts, let alone the interactions. Furthermore, exploitation of these new discoveries for pharmacological purposes is just the beginning. In the last two decades, technological advances in metabolomics and metagenomics, together with animal experiments to clarify mechanisms, have highlighted many important emerging roles for the microbiome in human health. Various studies have linked the microbiome to obesity (Turnbaugh et al. 2006), inflammatory bowel disorders (Sokol et al. 2006), neurological processes (Gonzalez et al. 2011; Sharon et al. 2016), cancer (Kilkkinen et al. 2008), cardiovascular health and disease (Koren et al. 2011), immunology (Lee and Mazmanian 2010), and many other conditions, processes, and systems.

The success in the use of fecal microbiota transplantation (FMT), literally transplanting gut microbes from one person to another, to treat gastrointestinal diseases such as *C. difficile* infection is the key example to date of a clinical application for commensal microbes as live biotherapeutic agents. *C. difficile* is a toxin-producing Gram-positive bacterium that is common in our food and healthcare facilities. The healthy gut microbial ecosystem provides a line of defense that prevents infection, but this line of defense can be breached unintentionally. A major risk factor associated with *C. difficile* infection is antibiotic use, especially clindamycin (Nowak et al. 2019). Restoring the gut microbiota via FMT has been shown to have high success rates not seen with other treatment methods, notably repeated antibiotic use which tends to lead to relapse (Tvede et al. 2015). Other types of microbiome transplantations, such as skin microbes to treat atopic dermatitis, have also been shown to be effective (Nakatsuji et al. 2017).

Using the microbiome to stratify patients prior to therapy for likely response to treatment is also very promising. For instance, the human microbiome plays an important role in whether a patient will respond well to chemotherapeutic cancer treatments or experience adverse events. Namely, one therapeutic for colon cancer is CPT-11 (irinotecan), which has severe diarrhea as a common adverse side effect. Through a series of reactions, CPT-11 is converted into inactive SN-38G in the human liver (Wallace et al. 2010). Inactive SN-38G is excreted into the gastrointestinal tract via the biliary ducts, where β-glucuronidase enzymes produced by *Streptococcus agalactiae*, *Escherichia coli*, and *Bacteroides fragilis* hydrolyze it into active SN-38 (Wallace et al. 2010, 2015). Increased levels of active SN-38 in the gastrointestinal tract are linked to the patients' diarrhea and have been shown to decrease CPT-11 efficacy (Kurita et al. 2011).

In this review, we highlight recent advances in metabolomic and metagenomic techniques and how they have been used to deepen our understanding of the role the microbiome plays in various disease states. We also connect these research results to prospects for improved pharmacological understanding and practice.

2 What Is the Human Microbiome and What Impacts It?

As first observed by Antonie van Leeuwenhoek in the 1680s, the microbes in different parts of the human body harbor distinct communities. However, modern DNA-sequencing techniques (Kashyap et al. 2013), together with improved methods of comparing microbial communities (Lozupone and Knight 2005) and vast databases of human and environmental samples sequenced using the same methods (McDonald et al. 2018; Thompson et al. 2017), allow us to put this diversity in perspective. Amazingly, the difference between two sites on the human body can be as different in their microbiomes as soil is from seawater (McDonald et al. 2018), with different parts of the body dominated by completely different taxa (Fig. 1).

The human-microbial interrelationship begins no later than birth. If a neonate passes through the birth canal, the trillions of microbes that constitute the mother's vaginal flora coat the skin and become the newborn's first microbial ecosystem, homogeneous across the body (Dominguez-Bello et al. 2010). Unlike adults, the gut, oral, and skin microbial consortia of a newborn are almost indistinguishable from each other. In addition, these communities are most closely related to that of their mother's vaginal microbiota with an overabundance of *Lactobacillus* and *Prevotella* spp. (Dominguez-Bello et al. 2010). Children born via Caesarean section also have homogenous microbial ecosystems, but rather than resembling the mother's vaginal community, they instead most closely resemble their mother's skin and are rich in *Staphylococcus* spp. (Dominguez-Bello et al. 2010). The microbial ecosystems of different body sites experience the most change over the first 3 years of life. By the age of 2–3 years, children have gut microbial communities that closely resemble that of adults (Koenig et al. 2011). As an infant ages, the oral, gut, and skin microbial body habitats begin to differentiate and form distinct niches. Many factors such as the environment, long-term diets, antibiotic and other drug use (see below), geographic location, culture, and sleep patterns are linked to the spatial and temporal differentiation observed among body habitat communities, while other "obvious" factors such as biological sex have surprisingly little effect (Costello et al. 2009).

However, the most astounding observation about the human microbiome is the incredible variation of microbial communities observed from person to person. While we are only ~0.1% genetically different within our human genome, we can be up to ~90% different in terms of our microbial genomes (Parfrey and Knight 2012). Intriguingly, relatively little of this variation is linked to the host genome (Goodrich et al. 2014; Rothschild et al. 2018; Turnbaugh et al. 2009). Because there is so much variation in the gut microbiome and some of this variation is strongly linked to phenotype, searching for microbial genes linked to phenotypic outcomes can be far more efficient than looking for human ones via genome-wide association studies (Rothschild et al. 2018).

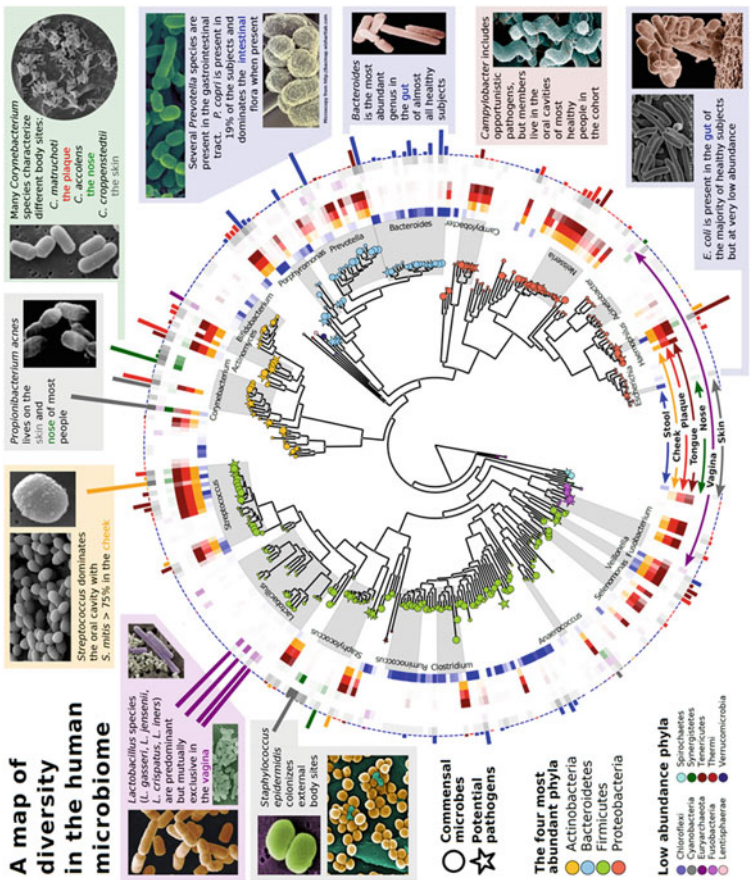


Fig. 1 "A map of microbial diversity in the human microbiome," reproduced from Morgan et al. (2013). A phylogenetic tree representing the four main phyla of microbes that dominate the human microbiome: *Actinobacteria*, *Bacteroidetes*, *Firmicutes*, and *Proteobacteria*. Extending from the tree are seven color-coded body sites and the corresponding taxa present. The heights of the outward bars indicate taxa abundance pertaining to one body site [e.g., *Lactobacillus* dominates the vaginal community, *Bacteroides* the gut]. The levels of abundance, and dominance, of bacterial taxa across the four phyla highlight the tremendous microbial diversity expressed across the human body

3 What Diseases and Biological Processes Is the Human Microbiome Linked to?

The gut microbiota, which is the best studied compartment of the human microbiome, has important roles in health and disease. One of the first links discovered was between the microbiota and inflammatory bowel disease (IBD). IBD is a complex pathological condition of chronic inflammation in the gastrointestinal tract that includes two clinical subtypes, Crohn's disease and ulcerative colitis, differentiated based on location of disease and inflammation. The etiology of IBD is unknown, but it is speculated that the pathogenesis is multifactorial, involving our immune system, genetic and environmental profiles, and modulation by our gut microbiota (Khan et al. 2019). Dysbiosis and decreased microbiota and microbiome complexity have been implicated in IBD. Factors that can perturb the complexity and stability of the microbiome, such as genetics, diet, drugs, and stress, have been linked to the diseased state, leading to overgrowth of Gram-negative bacteria, oxidative stress, altered metabolite production, and inflammation (Kostic et al. 2014) (Fig. 2).

The gut microbiota and/or microbiome has also been linked to metabolic disease states including obesity, diabetes mellitus, and nonalcoholic fatty liver disease (NAFLD). The four dominant phyla in the human gut are *Firmicutes*, *Bacteroidetes*, *Actinobacteria*, and *Proteobacteria*, with 90% of our microbes coming from the *Firmicutes* and *Bacteroidetes*. *Bifidobacterium* is a genus of commensal microbes found in the human intestine associated with protective health benefits (Mokhtari et al. 2017). In obesity, decreased diversity is observed both at a phylogenetic and metagenomic level. Overweight children have reduced levels of *Bifidobacterium*, and obese mice systematically overrepresent *Firmicutes* relative to *Bacteroidetes* (although this does not reproduce across human cohorts) (Turnbaugh et al. 2006;

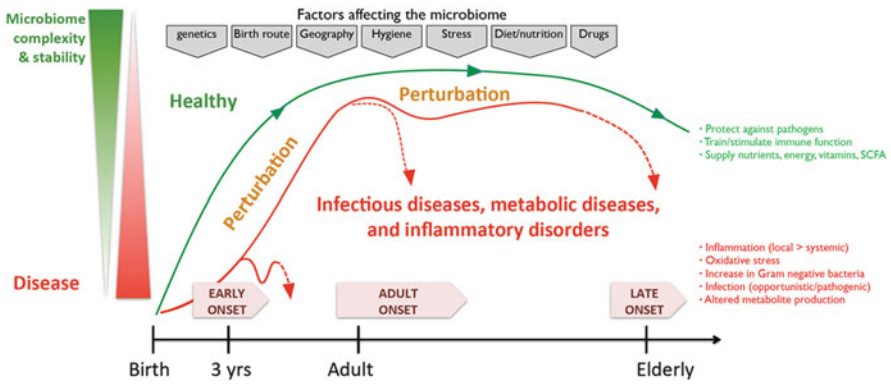


Fig. 2 "Factors affecting the stability and complexity of the gut microbiome in health and disease" from (Kostic et al. 2014). The microbiome and key characteristics are influenced during the progression of time from birth to old age. Many factors impact the microbiome (indicated in the top grey boxes). Some of these factors introduce perturbations that can affect the complexity and stability of the microbiome, potentially inducing imbalances (microbial dysbiosis)

Walters et al. 2014). *Bifidobacterium* is also reduced in mouse models of type 2 diabetes (Mokhtari et al. 2017). Improved hygiene and reduced infection rates in humans have been associated with lower *Bifidobacterium* and higher *Bacteroides* abundance in the first 3 years of life, together with an increased risk of autoimmune diseases like diabetes, asthma, and allergies (Vatanen et al. 2016). *Bacteroides* lipopolysaccharide (LPS) showed low immune acute responses in peripheral mononuclear blood cells (PMBC) and monocyte-derived dendritic cells, which failed to induce long-term endotoxin tolerance and increased the incidence of diabetes in an animal model (Vatanen et al. 2016), revealing the strong influence of the microbiome in immune response maturation. Patients with NAFLD have been shown to have bacterial overgrowth, increased intestinal permeability, and elevated serum LPS levels. Bacterial overgrowth increases the production of LPS, a hepatotoxic product, and intestinal permeability leads to translocation of bacteria and endotoxins. The passage of endotoxins into the portal vein of the liver activates hepatic inflammatory cells and increases risks for developing NAFLD (Mokhtari et al. 2017).

The gut microbiome has also been shown to play a role in cardiovascular disease through microbial metabolic pathways. Trimethylamine N-oxide (TMAO) is a biologically active amine oxide linked to atherosclerosis and a potential cause of cardiac and renal disease. TMAO's precursor, trimethylamine (TMA), is primarily generated by gut bacteria in their metabolism of choline and choline-containing compounds, carnitine, and betaine from diet and bile products. TMA and TMAO are also found in high concentrations in dietary seafood. Gut microbiota are required for the conversion of nutrients to TMA, where TMA is then absorbed from the intestines and oxidized to TMAO in the liver. TMAO is subsequently excreted in urine, sweat, and breath. The exact mechanism how TMAO potentiates cardiovascular disease is unknown, though it regulates aspects of cholesterol and sterol metabolism. Patients at risk for cardiovascular disease and stroke have high serum TMAO levels, choline, carnitine, and betaine (Zeisel and Warrier 2017) (Fig. 3).

Perhaps most remarkable is our recent expanded understanding of the "gut-brain axis," with identified links between the microbiota and neurologic/psychiatric disorders including anxiety and depression. In anxiety and depression, gut microbiota alterations are associated with changes in intestinal motility and increased translocation of bacterial products. The release of gut peptides and signaling molecules then act as sensory information to the central nervous system (Lach et al. 2018). In healthy individuals, bacterial taxa including *Bacteroides*, *Parabacteroides*, and *Escherichia* produce large quantities of gamma-aminobutyric acid (GABA), an inhibitory neurotransmitter involved in depressive disease and detectable in stool. Neuroimaging of patients with major depressive disorder showed that high relative abundance of *Bacteroides* was negatively correlated with brain imaging signatures seen in depression, suggesting that microbially derived GABA can influence the host in neurological disease states (Strandwitz et al. 2019) (Fig. 4).

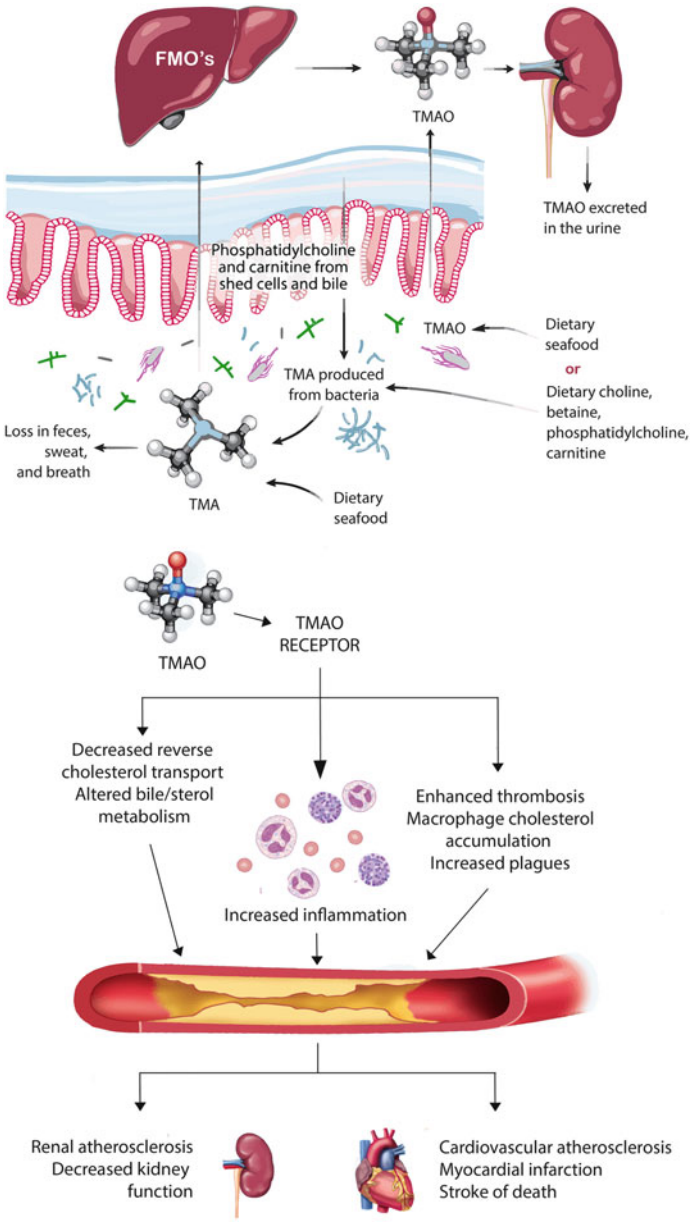


Fig. 3 Pathway for formation and excretion of TMAO. TMA is primarily produced in intestinal lumen by gut microbiota but also present in dietary seafood. Once TMA is absorbed through the intestines, it's oxidized to flavin-dependent monooxygenase (FMO) isoform in the liver, where then converted to TMAO and available for excretion primarily through urine. TMAO regulates metabolism of sterols and cholesterol and is implicated in renal and cardiac atherosclerosis. [From (Zeisel and Warrier 2017)]

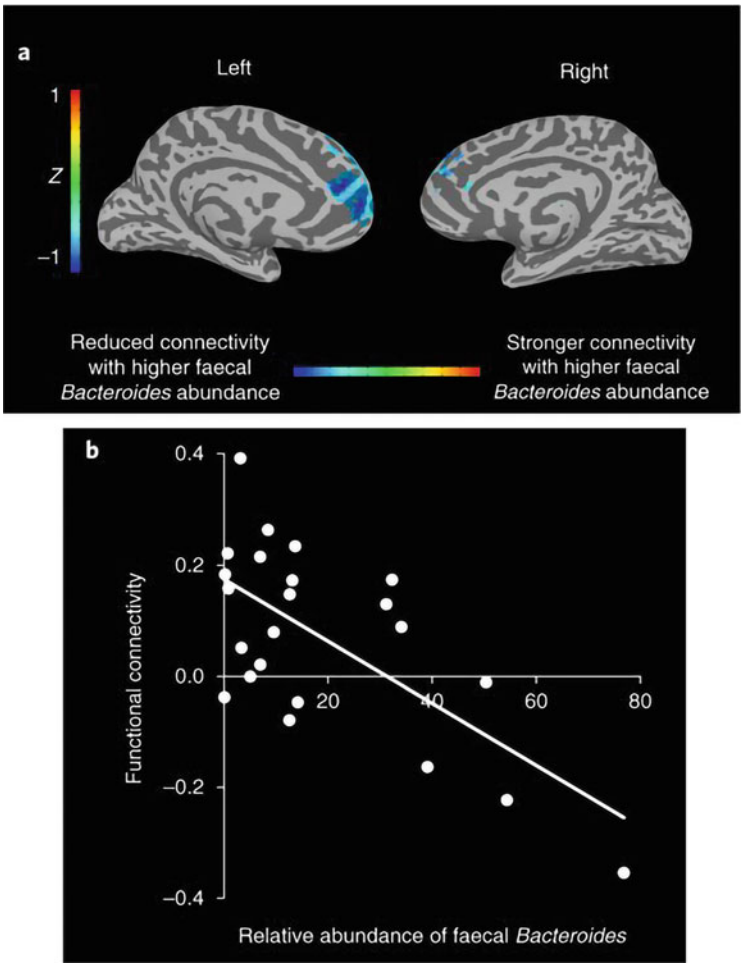


Fig. 4 "Fecal *Bacteroides* relative abundance inversely correlates with functional connectivity between left DLPFC and DMN structures in patients with major depressive disorder (MDD)" from Strandwitz et al. (2019). 3D plot of the medial surface of the cerebral hemispheres in patients with MDD. The left dorsolateral prefrontal cortex (DLPRC) area is known to be hypoactive in depression, and the default mode network (DMN) area is involved with negative rumination of self-referential processing in depression. (**a**) Demonstrates inverse correlation with fecal *Bacteroides* relative abundance and functional connectivity in the left DLPFC. (**b**) Scatter plot shows the average functional connectivity (*Z* score) over a sphere of radius 5 mm centered at the voxel of peak significance, related to abundance of fecal *Bacteroides*

4 How Do We Find Out About the Microbiome and Metabolome?

From the nineteenth century until recently, microbes were mainly studied by growing individual strains on solid or liquid media. However, this approach could only access a small fraction of microbes observed under the microscope – a phenomenon dubbed "the great plate count anomaly" (Staley and Konopka 1985). In the 1970s, the focus shifted to describing microbes by their DNA sequences. The most useful tool has been the 16S rRNA gene that is unique to and shared by all bacteria and archaea, which evolves quickly enough to provide near-species resolution yet has slowly evolving regions that can act as primer sites for the polymerase chain reaction (PCR) (Olsen et al. 1986; Woese and Fox 1977). Culturing and co-culturing techniques are currently undergoing a revival via new methods such as intestine-on-a-chip (Jalili-Firoozinezhad et al. 2019) and culturomics (Lagier et al. 2015, 2018). Still, many gut microbes (35–65%) remain uncultured (Lagkouvardos et al. 2017), and global efforts exist to culture, biobank, and preserve diverse important gut species (Bello et al. 2018; Rabesandratana 2018).

16S rRNA gene amplicon sequencing revolutionized microbiome research. This approach now has well-established and standardized protocols (Caporaso et al. 2012; Minich et al. 2018; Walters et al. 2016), including those established by the Earth Microbiome Project, which enabled massive-scale surveys of diverse environments (Thompson et al. 2017). For 16S rRNA gene amplicon studies, data analysis protocols are also well-established (Bolyen et al. 2019; Caporaso et al. 2010). For DNA-based analyses, shotgun metagenomics, which sequences fragments of all genes of all organisms present in a given sample, is rapidly becoming the standard (Lloyd-Price et al. 2017). Shotgun metagenomics not only gives access to higher taxonomic resolution by analysis of more genes but also helps describe metabolic pathways and identify microbial genes and pathways associated with host phenotype (Zeevi et al. 2019). Shallow shotgun sequencing approaches can be nearly as cost-effective as 16S rRNA studies, scaling to thousands of samples (Hillmann et al. 2018). Differentiating live from dead cells using DNA-based analyses is not standard and requires additional sample processing steps such as PMA treatment (Emerson et al. 2017).

However, DNA-based analyses do not provide access to actual expression levels of RNA or proteins. Meta-transcriptomics provides expression data for RNA and metaproteomics for proteins. Metaproteomics can now analyze hundreds of samples (Lloyd-Price et al. 2019; Rechenberger et al. 2019). The scalability of these methods remains a major bottleneck relative to shotgun metagenomics or for 16S rRNA sequencing, where analysis at scales of thousands and tens of thousands of samples, respectively, are now possible in individual studies (McDonald et al. 2018; Thompson et al. 2017). Interestingly, some evidence exists that good correlations between metagenomic and metaproteomic results at the taxonomic level can be achieved (Mills et al. 2019).

Metabolomics is the readout of the small molecules present in a sample, including microbial and host-derived metabolites. There are two main approaches to metabolomics: targeted and untargeted. Targeted metabolomics uses calibrated reference standards to read out a small number of molecules (typically, a few hundred) quantitatively, with known limits of detection. Untargeted metabolomics is more discovery-based, producing a huge number of non- or semiquantitative molecule identifications. Still, most mass peaks or spectra derived from untargeted approaches cannot be currently identified and require reference databases in order to understand them (much as was required in DNA sequencing, where most microbes in the human gut can now be identified, but 20 years ago, most were unknown). Currently reference databases are biased toward commercially available molecules. Most microbial molecules are not commercially available. The main technologies used for metabolomics (Wishart 2016) are nuclear magnetic resonance (NMR) and mass spectrometry (MS), with the latter becoming much more prevalent in recent years due to higher throughput, improved computational tools, and better limits of detection. MS is essentially a method of weighing ions very precisely according to their mass to charge ratio; in tandem MS (MS/MS), the initial ion is fragmented, and the fragments themselves are subjected to MS in order to yield a "fragmentation spectrum" that assists in molecule identification. Typically, molecules are ionized from a matrix-assisted laser desorption ionization, desorbed by a solvent under atmospheric conditions followed by electrospray ionization or separated by liquid chromatography or gas chromatography to reduce the complexity of the mixture. Computational processing for MS data is still in its infancy, with molecular networking proving an especially useful approach for untargeted metabolomics (Quinn et al. 2017; Wang et al. 2016), and community-curated reference databases of MS/MS from large numbers of samples are just beginning to be assembled and used (Wang et al. 2016). Furthermore, search tools that would provide the equivalent of BLAST for molecules and repository-scale analysis are still in development (Jarmusch et al. 2019; Wang et al. 2019).

An important consideration in techniques generating multiple data layers (e.g., DNA, RNA, proteins, metabolites, and live bacteria for culture) is that the same sample collection methods will not work for all biospecimens and analytes (Allaband et al. 2019). Benchmarking these approaches is still at relatively early stages, but this principle should be considered early in the experimental design phase, e.g., if live bacteria are desired, the samples cannot be preserved in ethanol or another fixative that kills cells, and if RNA-based analyses are desired, the samples must be immediately frozen in an ethanol-dry ice slurry or liquid nitrogen or preserved in high-salt reagents that will prevent metabolomics assays from being performed later (Metwally et al. 2015).

An influential experimental technique establishing the causality between microbiome composition and phenotype is the transplant of microbial communities into germ-free mice (or other species) (Rawls et al. 2006; Sommer and Backhed 2013). Essentially, such studies reveal if personalized characteristics of the donor (whether another mouse or a human) can be transmitted from one organism to another, demonstrating that the microbiome can be causal in the development of

such traits (at least in the germ-free mouse model). Such studies have clarified the role of the gut microbiome in many conditions, including obesity (Goodrich et al. 2014; Ridaura et al. 2013), Parkinson's disease (Sampson et al. 2016), multiple sclerosis (Cekanaviciute et al. 2017), neurodevelopmental disorders (Hsiao et al. 2013; Sharon et al. 2019), and major depressive disorder (Zheng et al. 2016).

5 How Do Drugs Affect the Microbiome?

Drugs commonly used in humans can exert an impact on the microbiome, influencing its functionality and composition. For example, antibiotics are widely used to treat bacterial infections. However, broad-spectrum antibiotics target not only the causative agent of the infection but also harmless and beneficial members of the microbiome (Dethlefsen and Relman 2011). When broad-spectrum antibiotics are used for prolonged periods of time, overall microbial diversity is reduced, enabling opportunistic pathogens such as *C. difficile* to overgrow and cause disease (as noted above). The microbiome disturbance posed by antibiotics in early life stages can also lead to higher risk of autoimmune diseases (Cox et al. 2014; Livanos et al. 2016; Russell et al. 2013; Schulfer et al. 2018; Zanvit et al. 2015). These long-lasting side effects of antibiotics have traditionally been ignored. Recent efforts aim to restore the antibiotic-induced dysbiosis using probiotics or FMT (Suez et al. 2018), which might if successful become standard of care after antibiotics. However, these efforts are still at the research stage as of this writing.

Antibiotics also impose an ecological pressure on the microbiome, selecting bacteria that resist their effects. This selection can occur in hours to days (Baym et al. 2016a). Antibiotic resistance has been increasing steadily, with increasing numbers of infections not responding to any of the available antibiotics: there is therefore a need for new molecules to treat multidrug-resistant bacteria and new treatment strategies that prevent or reverse the development of such resistance (Baym et al. 2016b).

How other drug classes affect the microbiome has been typically overlooked, although many drugs have been reported to produce gastrointestinal side effects (Leong and Chan 2006). Recent research has revealed that non-antibiotic drugs can also influence the composition of the gut microbiome (Falony et al. 2016; Maier et al. 2018). For example, metformin impacts the gut microbiome in diabetic patients, with a signature larger than that of diabetes itself (Forslund et al. 2015). These microbiome changes may have therapeutic effects in managing blood glucose levels (Wu et al. 2017). Gliptins, a group of dipeptidyl peptidase 4 (DPP-4) inhibitor drugs used for blood glucose control, have been proposed as a potential new therapeutic approach to treat colitis (Salaga et al. 2017). Vildagliptin alters the gut microbiome in mouse models, inhibiting both bacterial DPP-4 activity in fecal contents and growth of commensals. These microbiome changes improve mucosal barrier function, increase host-produced antimicrobial peptides, and improve crypt depth in the ileum (Olivares et al. 2018). Targeting bacterial DPP-4 homologs precisely might provide a new therapeutic avenue for colitis.

Proton pump inhibitors are commonly prescribed drugs that inhibit gastric acid production, increasing the stomach's pH. Although generally deemed as safe, they reduce alpha diversity in the gut microbiome and increase the proportion of pharyngeal bacteria in the gut (Jackson et al. 2016). This reduced alpha diversity could explain the increased risk of *C. difficile* associated with proton pump inhibitors usage (Linsky et al. 2010; Trifan et al. 2017).

A variety of non-antibiotic drugs inhibit the growth of commensal bacteria (Maier et al. 2018; Le Bastard et al. 2018). Repurposing those drugs as new antibiotic families could provide a promising strategy for altering the commensal microbiome and preventing or treating multidrug-resistant bacteria infections (Younis et al. 2015). However, tailoring these antimicrobial effects toward specific bacteria only under pathogenic conditions should improve the outcome of the treatment. In 2018, Zhu et al. proposed the inhibition of molybdenum-cofactor-dependent microbial respiratory pathways using tungstate (Zhu et al. 2018). These pathways are active in colitis-associated *Enterobacteriaceae* family exclusively during inflammatory episodes. By inhibiting this pathway, colitis is reversed without affecting the microbiome in noninflammatory states.

Taken together, these results suggest that many microbial pathways could be inhibited by either existing or yet-to-be-discovered compounds. Devising rules for druggable targets in the microbiome, and companion diagnostics that allow precision approaches based on an individual's microbiome, therefore hold considerable potential.

6 How Does the Microbiome Affect Drug Metabolism or Activity?

Considerable efforts have expended the identification of human genes linked to differential drug metabolism, with a particular focus on degradation pathways. These efforts have identified many genetic variants among human individuals that are involved in drug metabolism (Thorn et al. 2013). Pharmacogenetics and pharmacogenomics initiatives have enabled the development of diagnostic tests that assist in drug selection (companion diagnostics). However, the microbiome accounts for over 99% of the genetic repertoire in the human body, with human genes contributing less than 1%. This vast genetic diversity encodes many enzymes that can potentially degrade drugs. This field of research, pharmacomicrobiomics, is starting to decipher new degradation pathways by which bacteria can metabolize drugs (Spanogiannopoulos et al. 2016).

The microbiome can contribute to activation of prodrugs, leading to active metabolites. For example, sulfasalazine, used to treat rheumatoid arthritis, requires bacterial azoreductase activity to cleave the azo bond that links the two active components, mesalazine and sulfapyridine (Peppercorn and Goldman 1972). However, the microbiome can also contribute to inactivation of drugs, reducing their bioavailability. Microbial degradation pathways can be critical for narrow therapeutic range drugs like digoxin. Digoxin is used to treat cardiac arrhythmia and heart failure

and is subject to continuous bloodstream monitoring to ensure that therapeutic levels are achieved and that toxic levels are not reached. In a subset of patients that carry specific strains of *Eggerthella lenta* harboring the *cgr* operon, digoxin is metabolized into the inactive compound dihydrodigoxin (Haiser et al. 2013). This can lead to subtherapeutic systemic digoxin concentrations in those patients. Interestingly, the activity of the *cgr* can be reduced by increasing the intake of dietary protein, which could enable implementation of concomitant dietary changes for patients hosting such *E. lenta* strains, without increasing digoxin dosing.

The microbiome can also limit the detoxifying capacity of drugs, by producing metabolites that are metabolized through similar pathways. Patients with high bacterially mediated production of *p*-cresol have limited capacity to sulfonate acetaminophen, increasing the hepatotoxicity of the drug (Clayton et al. 2009). Other nonsteroidal anti-inflammatory drugs, such as diclofenac or indomethacin, are reabsorbed into intestinal epithelial cells due to bacterial-mediated β-glucuronidase activity, leading to increased risk of mucosal ulceration. By inhibiting the β-glucuronidase bacterial activity using the small molecule Inh-1, mucosal ulceration can be greatly reduced in vivo in animal models (Saitta et al. 2014).

The influence of the microbiome on the outcome of different cancer treatments received intense recent attention. Cyclophosphamide, an alkylating chemotherapy, exerts its immune-mediated effects by inducing impairment of the barrier function in the small intestine, enabling Gram-positive bacteria to translocate into lymph nodes and spleen (Viaud et al. 2013) This promotes pTh17 and Th17 anticancer immune responses, which can be mitigated with antibiotics. Similarly, the lack of microbiota impaired treatment response to CpG and oxaliplatin in a mouse model of melanoma, reducing the innate immune antitumor response (Iida et al. 2013). Similar observations were made in mice treated with immune checkpoint inhibitor drugs, whose therapeutic effects are abolished with the simultaneous administration of antibiotics (Routy et al. 2018). It has also been shown that the human gut microbiome composition influences the outcome of checkpoint inhibitors for treatment of melanomas and sarcomas (Routy et al. 2018; Gopalakrishnan et al. 2018; Matson et al. 2018). Microbiome-based patient stratification and modifying the immunogenicity of the microbiome could therefore dramatically improve success rates for cancer immunotherapy treatments.

7 How Can Drugs Be Isolated from the Microbiome?

In addition to affecting drug response, the microbiome can itself be a source for the discovery of new drugs. Therapeutic molecules with varied applications have been obtained from microbes isolated from soil (Grzelak et al. 2019; Ling et al. 2015) and marine environments (Molinski et al. 2009). Recent advances in metagenomics enabled mining the human microbiome for new biosynthetic gene clusters that produce natural products analogous to known drugs (Donia et al. 2014), already allowing isolation of several small molecules of potential therapeutic use (Donia and Fischbach 2015).

As in other ecosystems, microbes compete for space and resources. Bacteria use quorum-sensing systems to detect environmental conditions that might require a community-wide response in order to survive (Solano et al. 2014). In high-density ecosystems such as biofilms, quorum sensing is essential for detecting nutrient-deficient and toxic microenvironments. Quorum quenching molecules could provide a promising avenue for treating biofilm infections (Muras et al. 2018). Bacteriophages are strong modulators of bacterial communities that maintain their genetic diversity (Rodriguez-Valera et al. 2009). Given the high specificity of bacteriophages for particular bacterial targets, they have been proposed for high-precision treatment of multidrug-resistant bacteria (Schooley et al. 2017) and could potentially be engineered to increase efficacy (Dedrick et al. 2019).

Given the influence of microbiome composition has on host health, beneficial bacteria known as probiotics have been proposed to treat diseases such as necrotizing enterocolitis (Patel and Underwood 2018), oral diseases (Lopez-Lopez et al. 2017), allergic diseases (West et al. 2016), and major depressive disorder (Strandwitz et al. 2019). However, the administration of a single strain is not enough to achieve a therapeutic effect in some diseases, including *C. difficile* infection. In such cases, FMT from healthy individuals can be highly effective (van Nood et al. 2013). Recent research showed that autologous FMTs with fecal samples collected prior to antibiotic use are more effective than probiotics in restoring the gut mucosal microbiome after antibiotics (Suez et al. 2018). However, FMT poses important risks, given the wide variety of diseases associated with the microbiome that have demonstrated to be transmissible by fecal or cecal transplant in mice noted above. Further studies are needed to assess which bacterial strains engraft in patients, so that patients and donors can be matched effectively (Smillie et al. 2018). Defined clonal bacterial consortiums might be a safer and reproducible alternative to FMT, which has been shown to be effective in cancer treatment (Tanoue et al. 2019), colitis, and allergic diarrhea (Atarashi et al. 2013).

Finally, engineering microbiome commensals for different purposes is becoming a promising field of research. The gut commensal *Escherichia coli* Nissle has been engineered to sense and disperse *Pseudomonas aeruginosa* biofilms (Hwang et al. 2017). *E. coli* Nissle has also been genetically modified to supplement the lack of phenylalanine-metabolizing enzymes in phenylketonuria animal models, effectively reducing serum phenylalanine levels (Isabella et al. 2018), and is being tested in humans in clinical trials (NCT03516487). Drug delivery to the diseased body site can be also engineered in bacteria, using thermal bioswitches tunable with MRI-guided ultrasounds, reducing unwanted side effects in other body sites (Piraner et al. 2017). Additionally, engineered bacteria can be integrated into minimally invasive sensor devices that assist in diagnosis at the molecular level (Mimee et al. 2018).

8 Conclusions

Microbial interactions are clearly important for many human disease states where their involvement was not even suspected 20 years ago. Because they affect many biological processes by a range of mechanisms, many different and complementary lines of research are converging to show how microbiomes can affect pharmacological action and even be sources of novel therapeutic agents. Old culturing techniques are being redefined to grow a far larger fraction of our microbial inhabitants, and new fields such as pharmacomicrobiomics are beginning to enable a new understanding at the molecular level of the functional roles that microbes can play. Such multifaceted approaches have, in research settings, already enabled microbiome-based patient stratification, precision, and companion diagnostics and targeted therapeutics that can improve pharmaceutical response. Translating these research studies into clinical applications remains an important challenge for the field.

Important future directions that will improve our mechanistic understanding of microbiome processes and/or deliver benefits for patients include better integration of metagenomic, metabolomic, and metaproteomic techniques; improved characterization of viruses, parasites, and fungi (which are often understudied but which have been shown to interact with the human-microbial landscape in various disease states); and better characterization of the therapeutic capabilities of genetically modified microbes (Takiishi et al. 2012). The advances described in this review highlight the importance of collaboration among clinicians and researchers in a range of different disciplines for understanding host-microbial interactions and their relationships to human health.

References

Allaband C, McDonald D, Vazquez-Baeza Y, Minich JJ, Tripathi A, Brenner DA, Loomba R, Smarr L, Sandborn WJ, Schnabl B, Dorrestein P, Zarrinpar A, Knight R (2019) Microbiome 101: studying, analyzing, and interpreting gut microbiome data for clinicians. Clin Gastroenterol Hepatol 17(2):218–230. https://doi.org/10.1016/j.cgh.2018.09.017

Atarashi K, Tanoue T, Oshima K, Suda W, Nagano Y, Nishikawa H, Fukuda S, Saito T, Narushima S, Hase K, Kim S, Fritz JV, Wilmes P, Ueha S, Matsushima K, Ohno H, Olle B, Sakaguchi S, Taniguchi T, Morita H, Hattori M, Honda K (2013) Treg induction by a rationally selected mixture of clostridia strains from the human microbiota. Nature 500(7461):232–236. https://doi.org/10.1038/nature12331

Baym M, Lieberman TD, Kelsic ED, Chait R, Gross R, Yelin I, Kishony R (2016a) Spatiotemporal microbial evolution on antibiotic landscapes. Science 353(6304):1147–1151. https://doi.org/10.1126/science.aag0822

Baym M, Stone LK, Kishony R (2016b) Multidrug evolutionary strategies to reverse antibiotic resistance. Science 351(6268):aad3292. https://doi.org/10.1126/science.aad3292

Bello MGD, Knight R, Gilbert JA, Blaser MJ (2018) Preserving microbial diversity. Science 362 (6410):33–34. https://doi.org/10.1126/science.aau8816

Bolyen E, Rideout JR, Dillon MR, Bokulich NA, Abnet CC, Al-Ghalith GA, Alexander H, Alm EJ, Arumugam M, Asnicar F, Bai Y, Bisanz JE, Bittinger K, Brejnrod A, Brislawn CJ, Brown CT, Callahan BJ, Caraballo-Rodriguez AM, Chase J, Cope EK, Da Silva R, Diener C, Dorrestein PC, Douglas GM, Durall DM, Duvallet C, Edwardson CF, Ernst M, Estaki M, Fouquier J,

Gauglitz JM, Gibbons SM, Gibson DL, Gonzalez A, Gorlick K, Guo J, Hillmann B, Holmes S, Holste H, Huttenhower C, Huttley GA, Janssen S, Jarmusch AK, Jiang L, Kaehler BD, Kang KB, Keefe CR, Keim P, Kelley ST, Knights D, Koester I, Kosciolek T, Kreps J, Langille MGI, Lee J, Ley R, Liu YX, Loftfield E, Lozupone C, Maher M, Marotz C, Martin BD, McDonald D, McIver LJ, Melnik AV, Metcalf JL, Morgan SC, Morton JT, Naimey AT, Navas-Molina JA, Nothias LF, Orchanian SB, Pearson T, Peoples SL, Petras D, Preuss ML, Pruesse E, Rasmussen LB, Rivers A, Robeson MS 2nd, Rosenthal P, Segata N, Shaffer M, Shiffer A, Sinha R, Song SJ, Spear JR, Swafford AD, Thompson LR, Torres PJ, Trinh P, Tripathi A, Turnbaugh PJ, Ul-Hasan S, van der Hooft JJJ, Vargas F, Vazquez-Baeza Y, Vogtmann E, von Hippel M, Walters W, Wan Y, Wang M, Warren J, Weber KC, Williamson CHD, Willis AD, Xu ZZ, Zaneveld JR, Zhang Y, Zhu Q, Knight R, Caporaso JG (2019) Reproducible, interactive, scalable and extensible microbiome data science using QIIME 2. Nat Biotechnol 37 (8):852–857. https://doi.org/10.1038/s41587-019-0209-9

Caporaso JG, Kuczynski J, Stombaugh J, Bittinger K, Bushman FD, Costello EK, Fierer N, Pena AG, Goodrich JK, Gordon JI, Huttley GA, Kelley ST, Knights D, Koenig JE, Ley RE, Lozupone CA, McDonald D, Muegge BD, Pirrung M, Reeder J, Sevinsky JR, Turnbaugh PJ, Walters WA, Widmann J, Yatsunenko T, Zaneveld J, Knight R (2010) QIIME allows analysis of high-throughput community sequencing data. Nat Methods 7(5):335–336. https://doi.org/10.1038/nmeth.f.303

Caporaso JG, Lauber CL, Walters WA, Berg-Lyons D, Huntley J, Fierer N, Owens SM, Betley J, Fraser L, Bauer M, Gormley N, Gilbert JA, Smith G, Knight R (2012) Ultra-high-throughput microbial community analysis on the Illumina HiSeq and MiSeq platforms. ISME J 6 (8):1621–1624. https://doi.org/10.1038/ismej.2012.8

Cekanaviciute E, Yoo BB, Runia TF, Debelius JW, Singh S, Nelson CA, Kanner R, Bencosme Y, Lee YK, Hauser SL, Crabtree-Hartman E, Sand IK, Gacias M, Zhu Y, Casaccia P, Cree BAC, Knight R, Mazmanian SK, Baranzini SE (2017) Gut bacteria from multiple sclerosis patients modulate human T cells and exacerbate symptoms in mouse models. Proc Natl Acad Sci U S A 114(40):10713–10718. https://doi.org/10.1073/pnas.1711235114

Clayton TA, Baker D, Lindon JC, Everett JR, Nicholson JK (2009) Pharmacometabonomic identification of a significant host-microbiome metabolic interaction affecting human drug metabolism. Proc Natl Acad Sci U S A 106(34):14728–14733. https://doi.org/10.1073/pnas.0904489106

Costello EK, Lauber CL, Hamady M, Fierer N, Gordon JI, Knight R (2009) Bacterial community variation in human body habitats across space and time. Science 326(5960):1694–1697. https://doi.org/10.1126/science.1177486

Cox LM, Yamanishi S, Sohn J, Alekseyenko AV, Leung JM, Cho I, Kim SG, Li H, Gao Z, Mahana D, Zarate Rodriguez JG, Rogers AB, Robine N, Loke P, Blaser MJ (2014) Altering the intestinal microbiota during a critical developmental window has lasting metabolic consequences. Cell 158(4):705–721. https://doi.org/10.1016/j.cell.2014.05.052

Dedrick RM, Guerrero-Bustamante CA, Garlena RA, Russell DA, Ford K, Harris K, Gilmour KC, Soothill J, Jacobs-Sera D, Schooley RT, Hatfull GF, Spencer H (2019) Engineered bacteriophages for treatment of a patient with a disseminated drug-resistant Mycobacterium abscessus. Nat Med 25(5):730–733. https://doi.org/10.1038/s41591-019-0437-z

Dethlefsen L, Relman DA (2011) Incomplete recovery and individualized responses of the human distal gut microbiota to repeated antibiotic perturbation. Proc Natl Acad Sci U S A 108(Suppl 1):4554–4561. https://doi.org/10.1073/pnas.1000087107

Dominguez-Bello MG, Costello EK, Contreras M, Magris M, Hidalgo G, Fierer N, Knight R (2010) Delivery mode shapes the acquisition and structure of the initial microbiota across multiple body habitats in newborns. Proc Natl Acad Sci U S A 107(26):11971–11975. https://doi.org/10.1073/pnas.1002601107

Donia MS, Fischbach MA (2015) Human microbiota. Small molecules from the human microbiota. Science 349(6246):1254766. https://doi.org/10.1126/science.1254766

Donia MS, Cimermancic P, Schulze CJ, Wieland Brown LC, Martin J, Mitreva M, Clardy J, Linington RG, Fischbach MA (2014) A systematic analysis of biosynthetic gene clusters in the human microbiome reveals a common family of antibiotics. Cell 158(6):1402–1414. https://doi.org/10.1016/j.cell.2014.08.032

Emerson JB, Adams RI, Roman CMB, Brooks B, Coil DA, Dahlhausen K, Ganz HH, Hartmann EM, Hsu T, Justice NB, Paulino-Lima IG, Luongo JC, Lymperopoulou DS, Gomez-Silvan C, Rothschild-Mancinelli B, Balk M, Huttenhower C, Nocker A, Vaishampayan P, Rothschild LJ (2017) Schrodinger's microbes: tools for distinguishing the living from the dead in microbial ecosystems. Microbiome 5(1):86. https://doi.org/10.1186/s40168-017-0285-3

Falony G, Joossens M, Vieira-Silva S, Wang J, Darzi Y, Faust K, Kurilshikov A, Bonder MJ, Valles-Colomer M, Vandeputte D, Tito RY, Chaffron S, Rymenans L, Verspecht C, De Sutter L, Lima-Mendez G, D'Hoe K, Jonckheere K, Homola D, Garcia R, Tigchelaar EF, Eeckhaudt L, Fu J, Henckaerts L, Zhernakova A, Wijmenga C, Raes J (2016) Population-level analysis of gut microbiome variation. Science 352(6285):560–564. https://doi.org/10.1126/science.aad3503

Fischbach MA (2018) Microbiome: focus on causation and mechanism. Cell 174(4):785–790. https://doi.org/10.1016/j.cell.2018.07.038

Forslund K, Hildebrand F, Nielsen T, Falony G, Le Chatelier E, Sunagawa S, Prifti E, Vieira-Silva S, Gudmundsdottir V, Pedersen HK, Arumugam M, Kristiansen K, Voigt AY, Vestergaard H, Hercog R, Costea PI, Kultima JR, Li J, Jorgensen T, Levenez F, Dore J, Meta HIT consortium, Nielsen HB, Brunak S, Raes J, Hansen T, Wang J, Ehrlich SD, Bork P, Pedersen O (2015) Disentangling type 2 diabetes and metformin treatment signatures in the human gut microbiota. Nature 528(7581):262–266. https://doi.org/10.1038/nature15766

Gonzalez A, Stombaugh J, Lozupone C, Turnbaugh PJ, Gordon JI, Knight R (2011) The mind-body-microbial continuum. Dialogues Clin Neurosci 13(1):55–62

Goodrich JK, Waters JL, Poole AC, Sutter JL, Koren O, Blekhman R, Beaumont M, Van Treuren W, Knight R, Bell JT, Spector TD, Clark AG, Ley RE (2014) Human genetics shape the gut microbiome. Cell 159(4):789–799. https://doi.org/10.1016/j.cell.2014.09.053

Gopalakrishnan V, Spencer CN, Nezi L, Reuben A, Andrews MC, Karpinets TV, Prieto PA, Vicente D, Hoffman K, Wei SC, Cogdill AP, Zhao L, Hudgens CW, Hutchinson DS, Manzo T, Petaccia de Macedo M, Cotechini T, Kumar T, Chen WS, Reddy SM, Szczepaniak Sloane R, Galloway-Pena J, Jiang H, Chen PL, Shpall EJ, Rezvani K, Alousi AM, Chemaly RF, Shelburne S, Vence LM, Okhuysen PC, Jensen VB, Swennes AG, McAllister F, Marcelo Riquelme Sanchez E, Zhang Y, Le Chatelier E, Zitvogel L, Pons N, Austin-Breneman JL, Haydu LE, Burton EM, Gardner JM, Sirmans E, Hu J, Lazar AJ, Tsujikawa T, Diab A, Tawbi H, Glitza IC, Hwu WJ, Patel SP, Woodman SE, Amaria RN, Davies MA, Gershenwald JE, Hwu P, Lee JE, Zhang J, Coussens LM, Cooper ZA, Futreal PA, Daniel CR, Ajami NJ, Petrosino JF, Tetzlaff MT, Sharma P, Allison JP, Jenq RR, Wargo JA (2018) Gut microbiome modulates response to anti-PD-1 immunotherapy in melanoma patients. Science 359(6371):97–103. https://doi.org/10.1126/science.aan4236

Grzelak EM, Choules MP, Gao W, Cai G, Wan B, Wang Y, McAlpine JB, Cheng J, Jin Y, Lee H, Suh JW, Pauli GF, Franzblau SG, Jaki BU, Cho S (2019) Strategies in anti-Mycobacterium tuberculosis drug discovery based on phenotypic screening. J Antibiot (Tokyo). https://doi.org/10.1038/s41429-019-0205-9

Haiser HJ, Gootenberg DB, Chatman K, Sirasani G, Balskus EP, Turnbaugh PJ (2013) Predicting and manipulating cardiac drug inactivation by the human gut bacterium Eggerthella lenta. Science 341(6143):295–298. https://doi.org/10.1126/science.1235872

Hillmann B, Al-Ghalith GA, Shields-Cutler RR, Zhu Q, Gohl DM, Beckman KB, Knight R, Knights D (2018) Evaluating the information content of shallow shotgun metagenomics. mSystems 3(6). https://doi.org/10.1128/mSystems.00069-18

Hsiao EY, McBride SW, Hsien S, Sharon G, Hyde ER, McCue T, Codelli JA, Chow J, Reisman SE, Petrosino JF, Patterson PH, Mazmanian SK (2013) Microbiota modulate behavioral and physiological abnormalities associated with neurodevelopmental disorders. Cell 155(7):1451–1463. https://doi.org/10.1016/j.cell.2013.11.024

Hwang IY, Koh E, Wong A, March JC, Bentley WE, Lee YS, Chang MW (2017) Engineered probiotic Escherichia coli can eliminate and prevent Pseudomonas aeruginosa gut infection in animal models. Nat Commun 8:15028. https://doi.org/10.1038/ncomms15028

Iida N, Dzutsev A, Stewart CA, Smith L, Bouladoux N, Weingarten RA, Molina DA, Salcedo R, Back T, Cramer S, Dai RM, Kiu H, Cardone M, Naik S, Patri AK, Wang E, Marincola FM, Frank KM, Belkaid Y, Trinchieri G, Goldszmid RS (2013) Commensal bacteria control cancer response to therapy by modulating the tumor microenvironment. Science 342(6161):967–970. https://doi.org/10.1126/science.1240527

Isabella VM, Ha BN, Castillo MJ, Lubkowicz DJ, Rowe SE, Millet YA, Anderson CL, Li N, Fisher AB, West KA, Reeder PJ, Momin MM, Bergeron CG, Guilmain SE, Miller PF, Kurtz CB, Falb D (2018) Development of a synthetic live bacterial therapeutic for the human metabolic disease phenylketonuria. Nat Biotechnol 36(9):857–864. https://doi.org/10.1038/nbt.4222

Jackson MA, Goodrich JK, Maxan ME, Freedberg DE, Abrams JA, Poole AC, Sutter JL, Welter D, Ley RE, Bell JT, Spector TD, Steves CJ (2016) Proton pump inhibitors alter the composition of the gut microbiota. Gut 65(5):749–756. https://doi.org/10.1136/gutjnl-2015-310861

Jalili-Firoozinezhad S, Gazzaniga FS, Calamari EL, Camacho DM, Fadel CW, Bein A, Swenor B, Nestor B, Cronce MJ, Tovaglieri A, Levy O, Gregory KE, Breault DT, Cabral JMS, Kasper DL, Novak R, Ingber DE (2019) A complex human gut microbiome cultured in an anaerobic intestine-on-a-chip. Nat Biomed Eng 3(7):520–531. https://doi.org/10.1038/s41551-019-0397-0

Jarmusch AK, Wang M, Aceves CM, Advani RS, Aguire S, Aksenov AA, Aleti G, Aron AT, Bauermeister A, Bolleddu S, Bouslimani A, Caraballo-Rodriguez AM, Chaar R, Coras R, Elijah EO, Ernst M, Gauglitz JM, Gentry EC, Husband M, Jarmusch SA, Jones KL, Kamenik Z, Le Gouellec A, Lu A, McCall L-I, McPhail KL, Meehan MJ, Melnik AV, Menezes RC, Montoya-Giraldo YA, Nguyen NH, Nothias LF, Nothias-Esposito M, Panitchpakdi M, Petras D, Quinn R, Sikora N, van der Hooft JJJ, Vargas F, Vrbanac A, Weldon K, Knight R, Bandeira N, Dorrestein PC (2019) Repository-scale Co- and Re-analysis of tandem mass spectrometry data. bioRxiv:750471. https://doi.org/10.1101/750471

Kashyap PC, Marcobal A, Ursell LK, Smits SA, Sonnenburg ED, Costello EK, Higginbottom SK, Domino SE, Holmes SP, Relman DA, Knight R, Gordon JI, Sonnenburg JL (2013) Genetically dictated change in host mucus carbohydrate landscape exerts a diet-dependent effect on the gut microbiota. Proc Natl Acad Sci U S A 110(42):17059–17064. https://doi.org/10.1073/pnas.1306070110

Khan I, Ullah N, Zha L, Bai Y, Khan A, Zhao T, Che T, Zhang C (2019) Alteration of gut microbiota in inflammatory bowel disease (IBD): cause or consequence? IBD treatment targeting the gut microbiome. Pathogens 8(3). https://doi.org/10.3390/pathogens8030126

Kilkkinen A, Rissanen H, Klaukka T, Pukkala E, Heliovaara M, Huovinen P, Mannisto S, Aromaa A, Knekt P (2008) Antibiotic use predicts an increased risk of cancer. Int J Cancer 123(9):2152–2155. https://doi.org/10.1002/ijc.23622

Koenig JE, Spor A, Scalfone N, Fricker AD, Stombaugh J, Knight R, Angenent LT, Ley RE (2011) Succession of microbial consortia in the developing infant gut microbiome. Proc Natl Acad Sci U S A 108(Suppl 1):4578–4585. https://doi.org/10.1073/pnas.1000081107

Koren O, Spor A, Felin J, Fak F, Stombaugh J, Tremaroli V, Behre CJ, Knight R, Fagerberg B, Ley RE, Backhed F (2011) Human oral, gut, and plaque microbiota in patients with atherosclerosis. Proc Natl Acad Sci U S A 108(Suppl 1):4592–4598. https://doi.org/10.1073/pnas.1011383107

Kostic AD, Xavier RJ, Gevers D (2014) The microbiome in inflammatory bowel disease: current status and the future ahead. Gastroenterology 146(6):1489–1499. https://doi.org/10.1053/j.gastro.2014.02.009

Kurita A, Kado S, Matsumoto T, Asakawa N, Kaneda N, Kato I, Uchida K, Onoue M, Yokokura T (2011) Streptomycin alleviates irinotecan-induced delayed-onset diarrhea in rats by a mechanism other than inhibition of beta-glucuronidase activity in intestinal lumen. Cancer Chemother Pharmacol 67(1):201–213. https://doi.org/10.1007/s00280-010-1310-4

Lach G, Schellekens H, Dinan TG, Cryan JF (2018) Anxiety, depression, and the microbiome: a role for gut peptides. Neurotherapeutics 15(1):36–59. https://doi.org/10.1007/s13311-017-0585-0

Lagier JC, Hugon P, Khelaifia S, Fournier PE, La Scola B, Raoult D (2015) The rebirth of culture in microbiology through the example of culturomics to study human gut microbiota. Clin Microbiol Rev 28(1):237–264. https://doi.org/10.1128/CMR.00014-14

Lagier JC, Dubourg G, Million M, Cadoret F, Bilen M, Fenollar F, Levasseur A, Rolain JM, Fournier PE, Raoult D (2018) Culturing the human microbiota and culturomics. Nat Rev Microbiol:540–550. https://doi.org/10.1038/s41579-018-0041-0

Lagkouvardos I, Overmann J, Clavel T (2017) Cultured microbes represent a substantial fraction of the human and mouse gut microbiota. Gut Microbes 8(5):493–503. https://doi.org/10.1080/19490976.2017.1320468

Le Bastard Q, Al-Ghalith GA, Gregoire M, Chapelet G, Javaudin F, Dailly E, Batard E, Knights D, Montassier E (2018) Systematic review: human gut dysbiosis induced by non-antibiotic prescription medications. Aliment Pharmacol Ther 47(3):332–345. https://doi.org/10.1111/apt.14451

Lee YK, Mazmanian SK (2010) Has the microbiota played a critical role in the evolution of the adaptive immune system? Science 330(6012):1768–1773. https://doi.org/10.1126/science.1195568

Leong RW, Chan FK (2006) Drug-induced side effects affecting the gastrointestinal tract. Expert Opin Drug Saf 5(4):585–592. https://doi.org/10.1517/14740338.5.4.585

Ling LL, Schneider T, Peoples AJ, Spoering AL, Engels I, Conlon BP, Mueller A, Schaberle TF, Hughes DE, Epstein S, Jones M, Lazarides L, Steadman VA, Cohen DR, Felix CR, Fetterman KA, Millett WP, Nitti AG, Zullo AM, Chen C, Lewis K (2015) A new antibiotic kills pathogens without detectable resistance. Nature 517(7535):455–459. https://doi.org/10.1038/nature14098

Linsky A, Gupta K, Lawler EV, Fonda JR, Hermos JA (2010) Proton pump inhibitors and risk for recurrent Clostridium difficile infection. Arch Intern Med 170(9):772–778. https://doi.org/10.1001/archinternmed.2010.73

Livanos AE, Greiner TU, Vangay P, Pathmasiri W, Stewart D, McRitchie S, Li H, Chung J, Sohn J, Kim S, Gao Z, Barber C, Kim J, Ng S, Rogers AB, Sumner S, Zhang XS, Cadwell K, Knights D, Alekseyenko A, Backhed F, Blaser MJ (2016) Antibiotic-mediated gut microbiome perturbation accelerates development of type 1 diabetes in mice. Nat Microbiol 1(11):16140. https://doi.org/10.1038/nmicrobiol.2016.140

Lloyd-Price J, Mahurkar A, Rahnavard G, Crabtree J, Orvis J, Hall AB, Brady A, Creasy HH, McCracken C, Giglio MG, McDonald D, Franzosa EA, Knight R, White O, Huttenhower C (2017) Strains, functions and dynamics in the expanded human microbiome project. Nature 550(7674):61–66. https://doi.org/10.1038/nature23889

Lloyd-Price J, Arze C, Ananthakrishnan AN, Schirmer M, Avila-Pacheco J, Poon TW, Andrews E, Ajami NJ, Bonham KS, Brislawn CJ, Casero D, Courtney H, Gonzalez A, Graeber TG, Hall AB, Lake K, Landers CJ, Mallick H, Plichta DR, Prasad M, Rahnavard G, Sauk J, Shungin D, Vazquez-Baeza Y, White RA 3rd, Investigators I, Braun J, Denson LA, Jansson JK, Knight R, Kugathasan S, McGovern DPB, Petrosino JF, Stappenbeck TS, Winter HS, Clish CB, Franzosa EA, Vlamakis H, Xavier RJ, Huttenhower C (2019) Multi-omics of the gut microbial ecosystem in inflammatory bowel diseases. Nature 569(7758):655–662. https://doi.org/10.1038/s41586-019-1237-9

Lopez-Lopez A, Camelo-Castillo A, Ferrer MD, Simon-Soro A, Mira A (2017) Health-associated niche inhabitants as oral probiotics: the case of Streptococcus dentisani. Front Microbiol 8:379. https://doi.org/10.3389/fmicb.2017.00379

Lozupone C, Knight R (2005) UniFrac: a new phylogenetic method for comparing microbial communities. Appl Environ Microbiol 71(12):8228–8235. https://doi.org/10.1128/AEM.71.12.8228-8235.2005

Maier L, Pruteanu M, Kuhn M, Zeller G, Telzerow A, Anderson EE, Brochado AR, Fernandez KC, Dose H, Mori H, Patil KR, Bork P, Typas A (2018) Extensive impact of non-antibiotic drugs on human gut bacteria. Nature 555(7698):623–628. https://doi.org/10.1038/nature25979

Matson V, Fessler J, Bao R, Chongsuwat T, Zha Y, Alegre ML, Luke JJ, Gajewski TF (2018) The commensal microbiome is associated with anti-PD-1 efficacy in metastatic melanoma patients. Science 359(6371):104–108. https://doi.org/10.1126/science.aao3290

McDonald D, Hyde E, Debelius JW, Morton JT, Gonzalez A, Ackermann G, Aksenov AA, Behsaz B, Brennan C, Chen Y, DeRight Goldasich L, Dorrestein PC, Dunn RR, Fahimipour AK, Gaffney J, Gilbert JA, Gogul G, Green JL, Hugenholtz P, Humphrey G, Huttenhower C, Jackson MA, Janssen S, Jeste DV, Jiang L, Kelley ST, Knights D, Kosciolek T, Ladau J, Leach J, Marotz C, Meleshko D, Melnik AV, Metcalf JL, Mohimani H, Montassier E, Navas-Molina J, Nguyen TT, Peddada S, Pevzner P, Pollard KS, Rahnavard G, Robbins-Pianka A, Sangwan N, Shorenstein J, Smarr L, Song SJ, Spector T, Swafford AD, Thackray VG, Thompson LR, Tripathi A, Vazquez-Baeza Y, Vrbanac A, Wischmeyer P, Wolfe E, Zhu Q, American Gut Consortium, Knight R (2018) American gut: an open platform for citizen science microbiome research. mSystems 3(3). https://doi.org/10.1128/mSystems.00031-18

Metwally H, McAllister RG, Konermann L (2015) Exploring the mechanism of salt-induced signal suppression in protein electrospray mass spectrometry using experiments and molecular dynamics simulations. Anal Chem 87(4):2434–2442. https://doi.org/10.1021/ac5044016

Mills RH, Vazquez-Baeza Y, Zhu Q, Jiang L, Gaffney J, Humphrey G, Smarr L, Knight R, Gonzalez DJ (2019) Evaluating metagenomic prediction of the metaproteome in a 4.5-year study of a patient with crohn's disease. mSystems 4(1). https://doi.org/10.1128/mSystems.00337-18

Mimee M, Nadeau P, Hayward A, Carim S, Flanagan S, Jerger L, Collins J, McDonnell S, Swartwout R, Citorik RJ, Bulovic V, Langer R, Traverso G, Chandrakasan AP, Lu TK (2018) An ingestible bacterial-electronic system to monitor gastrointestinal health. Science 360(6391):915–918. https://doi.org/10.1126/science.aas9315

Minich JJ, Humphrey G, Benitez RAS, Sanders J, Swafford A, Allen EE, Knight R (2018) High-throughput miniaturized 16S rRNA amplicon library preparation reduces costs while preserving microbiome integrity. mSystems 3(6). https://doi.org/10.1128/mSystems.00166-18

Mokhtari Z, Gibson DL, Hekmatdoost A (2017) Nonalcoholic fatty liver disease, the gut microbiome, and diet. Adv Nutr 8(2):240–252. https://doi.org/10.3945/an.116.013151

Molinski TF, Dalisay DS, Lievens SL, Saludes JP (2009) Drug development from marine natural products. Nat Rev Drug Discov 8(1):69–85. https://doi.org/10.1038/nrd2487

Morgan XC, Segata N, Huttenhower C (2013) Biodiversity and functional genomics in the human microbiome. Trends Genet 29(1):51–58. https://doi.org/10.1016/j.tig.2012.09.005

Muras A, Mayer C, Romero M, Camino T, Ferrer MD, Mira A, Otero A (2018) Inhibition of Streptococcus mutans biofilm formation by extracts of Tenacibaculum sp. 20J, a bacterium with wide-spectrum quorum quenching activity. J Oral Microbiol 10(1):1429788. https://doi.org/10.1080/20002297.2018.1429788

Nakatsuji T, Chen TH, Narala S, Chun KA, Two AM, Yun T, Shafiq F, Kotol PF, Bouslimani A, Melnik AV, Latif H, Kim JN, Lockhart A, Artis K, David G, Taylor P, Streib J, Dorrestein PC, Grier A, Gill SR, Zengler K, Hata TR, Leung DY, Gallo RL (2017) Antimicrobials from human skin commensal bacteria protect against Staphylococcus aureus and are deficient in atopic dermatitis. Sci Transl Med 9(378). https://doi.org/10.1126/scitranslmed.aah4680

Nowak A, Hedenstierna M, Ursing J, Lidman C, Nowak P (2019) Efficacy of routine fecal microbiota transplantation for treatment of recurrent Clostridium difficile infection: a retrospective cohort study. Int J Microbiol 2019:7395127. https://doi.org/10.1155/2019/7395127

Olivares M, Neyrinck AM, Potgens SA, Beaumont M, Salazar N, Cani PD, Bindels LB, Delzenne NM (2018) The DPP-4 inhibitor vildagliptin impacts the gut microbiota and prevents disruption of intestinal homeostasis induced by a Western diet in mice. Diabetologia 61(8):1838–1848. https://doi.org/10.1007/s00125-018-4647-6

Olsen GJ, Lane DJ, Giovannoni SJ, Pace NR, Stahl DA (1986) Microbial ecology and evolution: a ribosomal RNA approach. Annu Rev Microbiol 40:337–365. https://doi.org/10.1146/annurev.mi.40.100186.002005

Parfrey LW, Knight R (2012) Spatial and temporal variability of the human microbiota. Clin Microbiol Infect 18(Suppl 4):8–11. https://doi.org/10.1111/j.1469-0691.2012.03861.x

Patel RM, Underwood MA (2018) Probiotics and necrotizing enterocolitis. Semin Pediatr Surg 27 (1):39–46. https://doi.org/10.1053/j.sempedsurg.2017.11.008

Peppercorn MA, Goldman P (1972) The role of intestinal bacteria in the metabolism of salicylazosulfapyridine. J Pharmacol Exp Ther 181(3):555–562

Piraner DI, Abedi MH, Moser BA, Lee-Gosselin A, Shapiro MG (2017) Tunable thermal bioswitches for in vivo control of microbial therapeutics. Nat Chem Biol 13(1):75–80. https://doi.org/10.1038/nchembio.2233

Qin J, Li R, Raes J, Arumugam M, Burgdorf KS, Manichanh C, Nielsen T, Pons N, Levenez F, Yamada T, Mende DR, Li J, Xu J, Li S, Li D, Cao J, Wang B, Liang H, Zheng H, Xie Y, Tap J, Lepage P, Bertalan M, Batto JM, Hansen T, Le Paslier D, Linneberg A, Nielsen HB, Pelletier E, Renault P, Sicheritz-Ponten T, Turner K, Zhu H, Yu C, Li S, Jian M, Zhou Y, Li Y, Zhang X, Li S, Qin N, Yang H, Wang J, Brunak S, Dore J, Guarner F, Kristiansen K, Pedersen O, Parkhill J, Weissenbach J, Meta HITC, Bork P, Ehrlich SD, Wang J (2010) A human gut microbial gene catalogue established by metagenomic sequencing. Nature 464(7285):59–65. https://doi.org/10.1038/nature08821

Quinn RA, Nothias LF, Vining O, Meehan M, Esquenazi E, Dorrestein PC (2017) Molecular networking as a drug discovery, drug metabolism, and precision medicine strategy. Trends Pharmacol Sci 38(2):143–154. https://doi.org/10.1016/j.tips.2016.10.011

Rabesandratana T (2018) Microbiome conservancy stores global fecal samples. Science 362 (6414):510–511. https://doi.org/10.1126/science.362.6414.510

Rawls JF, Mahowald MA, Ley RE, Gordon JI (2006) Reciprocal gut microbiota transplants from zebrafish and mice to germ-free recipients reveal host habitat selection. Cell 127(2):423–433. https://doi.org/10.1016/j.cell.2006.08.043

Rechenberger J, Samaras P, Jarzab A, Behr J, Frejno M, Djukovic A, Sanz J, Gonzalez-Barbera EM, Salavert M, Lopez-Hontangas JL, Xavier KB, Debrauwer L, Rolain JM, Sanz M, Garcia-Garcera M, Wilhelm M, Ubeda C, Kuster B (2019) Challenges in clinical metaproteomics highlighted by the analysis of acute leukemia patients with gut colonization by multidrug-resistant enterobacteriaceae. Proteomes 7(1). https://doi.org/10.3390/proteomes7010002

Ridaura VK, Faith JJ, Rey FE, Cheng J, Duncan AE, Kau AL, Griffin NW, Lombard V, Henrissat B, Bain JR, Muehlbauer MJ, Ilkayeva O, Semenkovich CF, Funai K, Hayashi DK, Lyle BJ, Martini MC, Ursell LK, Clemente JC, Van Treuren W, Walters WA, Knight R, Newgard CB, Heath AC, Gordon JI (2013) Gut microbiota from twins discordant for obesity modulate metabolism in mice. Science 341(6150):1241214. https://doi.org/10.1126/science.1241214

Rodriguez-Valera F, Martin-Cuadrado AB, Rodriguez-Brito B, Pasic L, Thingstad TF, Rohwer F, Mira A (2009) Explaining microbial population genomics through phage predation. Nat Rev Microbiol 7(11):828–836. https://doi.org/10.1038/nrmicro2235

Rothschild D, Weissbrod O, Barkan E, Kurilshikov A, Korem T, Zeevi D, Costea PI, Godneva A, Kalka IN, Bar N, Shilo S, Lador D, Vila AV, Zmora N, Pevsner-Fischer M, Israeli D, Kosower N, Malka G, Wolf BC, Avnit-Sagi T, Lotan-Pompan M, Weinberger A, Halpern Z, Carmi S, Fu J, Wijmenga C, Zhernakova A, Elinav E, Segal E (2018) Environment dominates over host genetics in shaping human gut microbiota. Nature 555(7695):210–215. https://doi.org/10.1038/nature25973

Routy B, Le Chatelier E, Derosa L, Duong CPM, Alou MT, Daillere R, Fluckiger A, Messaoudene M, Rauber C, Roberti MP, Fidelle M, Flament C, Poirier-Colame V, Opolon P, Klein C, Iribarren K, Mondragon L, Jacquelot N, Qu B, Ferrere G, Clemenson C, Mezquita L, Masip JR, Naltet C, Brosseau S, Kaderbhai C, Richard C, Rizvi H, Levenez F, Galleron N, Quinquis B, Pons N, Ryffel B, Minard-Colin V, Gonin P, Soria JC, Deutsch E, Loriot Y,

Ghiringhelli F, Zalcman G, Goldwasser F, Escudier B, Hellmann MD, Eggermont A, Raoult D, Albiges L, Kroemer G, Zitvogel L (2018) Gut microbiome influences efficacy of PD-1-based immunotherapy against epithelial tumors. Science 359(6371):91–97. https://doi.org/10.1126/science.aan3706

Russell SL, Gold MJ, Willing BP, Thorson L, McNagny KM, Finlay BB (2013) Perinatal antibiotic treatment affects murine microbiota, immune responses and allergic asthma. Gut Microbes 4 (2):158–164. https://doi.org/10.4161/gmic.23567

Saitta KS, Zhang C, Lee KK, Fujimoto K, Redinbo MR, Boelsterli UA (2014) Bacterial beta-glucuronidase inhibition protects mice against enteropathy induced by indomethacin, ketoprofen or diclofenac: mode of action and pharmacokinetics. Xenobiotica 44(1):28–35. https://doi.org/10.3109/00498254.2013.811314

Salaga M, Mokrowiecka A, Zielinska M, Malecka-Panas E, Kordek R, Kamysz E, Fichna J (2017) New peptide inhibitor of dipeptidyl peptidase IV, EMDB-1 extends the half-life of GLP-2 and attenuates colitis in mice after topical administration. J Pharmacol Exp Ther 363(1):92–103. https://doi.org/10.1124/jpet.117.242586

Sampson TR, Debelius JW, Thron T, Janssen S, Shastri GG, Ilhan ZE, Challis C, Schretter CE, Rocha S, Gradinaru V, Chesselet MF, Keshavarzian A, Shannon KM, Krajmalnik-Brown R, Wittung-Stafshede P, Knight R, Mazmanian SK (2016) Gut microbiota regulate motor deficits and Neuroinflammation in a model of Parkinson's disease. Cell 167(6):1469–1480.e12. https://doi.org/10.1016/j.cell.2016.11.018

Schooley RT, Biswas B, Gill JJ, Hernandez-Morales A, Lancaster J, Lessor L, Barr JJ, Reed SL, Rohwer F, Benler S, Segall AM, Taplitz R, Smith DM, Kerr K, Kumaraswamy M, Nizet V, Lin L, McCauley MD, Strathdee SA, Benson CA, Pope RK, Leroux BM, Picel AC, Mateczun AJ, Cilwa KE, Regeimbal JM, Estrella LA, Wolfe DM, Henry MS, Quinones J, Salka S, Bishop-Lilly KA, Young R, Hamilton T (2017) Development and use of personalized bacteriophage-based therapeutic cocktails to treat a patient with a disseminated resistant acinetobacter baumannii infection. Antimicrob Agents Chemother 61(10). https://doi.org/10.1128/AAC.00954-17

Schulfer AF, Battaglia T, Alvarez Y, Bijnens L, Ruiz VE, Ho M, Robinson S, Ward T, Cox LM, Rogers AB, Knights D, Sartor RB, Blaser MJ (2018) Intergenerational transfer of antibiotic-perturbed microbiota enhances colitis in susceptible mice. Nat Microbiol 3(2):234–242. https://doi.org/10.1038/s41564-017-0075-5

Sender R, Fuchs S, Milo R (2016) Are we really vastly outnumbered? Revisiting the ratio of bacterial to host cells in humans. Cell 164(3):337–340. https://doi.org/10.1016/j.cell.2016.01.013

Sharon G, Sampson TR, Geschwind DH, Mazmanian SK (2016) The central nervous system and the gut microbiome. Cell 167(4):915–932. https://doi.org/10.1016/j.cell.2016.10.027

Sharon G, Cruz NJ, Kang DW, Gandal MJ, Wang B, Kim YM, Zink EM, Casey CP, Taylor BC, Lane CJ, Bramer LM, Isern NG, Hoyt DW, Noecker C, Sweredoski MJ, Moradian A, Borenstein E, Jansson JK, Knight R, Metz TO, Lois C, Geschwind DH, Krajmalnik-Brown R, Mazmanian SK (2019) Human gut microbiota from autism spectrum disorder promote behavioral symptoms in mice. Cell 177(6):1600–1618.e17. https://doi.org/10.1016/j.cell.2019.05.004

Smillie CS, Sauk J, Gevers D, Friedman J, Sung J, Youngster I, Hohmann EL, Staley C, Khoruts A, Sadowsky MJ, Allegretti JR, Smith MB, Xavier RJ, Alm EJ (2018) Strain tracking reveals the determinants of bacterial engraftment in the human Gut following fecal microbiota transplantation. Cell Host Microbe 23(2):229–240.e5. https://doi.org/10.1016/j.chom.2018.01.003

Sokol H, Seksik P, Rigottier-Gois L, Lay C, Lepage P, Podglajen I, Marteau P, Dore J (2006) Specificities of the fecal microbiota in inflammatory bowel disease. Inflamm Bowel Dis 12 (2):106–111. https://doi.org/10.1097/01.MIB.0000200323.38139.c6

Solano C, Echeverz M, Lasa I (2014) Biofilm dispersion and quorum sensing. Curr Opin Microbiol 18:96–104. https://doi.org/10.1016/j.mib.2014.02.008

Sommer F, Backhed F (2013) The gut microbiota – masters of host development and physiology. Nat Rev Microbiol 11(4):227–238. https://doi.org/10.1038/nrmicro2974

Spanogiannopoulos P, Bess EN, Carmody RN, Turnbaugh PJ (2016) The microbial pharmacists within us: a metagenomic view of xenobiotic metabolism. Nat Rev Microbiol 14(5):273–287. https://doi.org/10.1038/nrmicro.2016.17

Staley JT, Konopka A (1985) Measurement of in situ activities of nonphotosynthetic microorganisms in aquatic and terrestrial habitats. Annu Rev Microbiol 39:321–346. https://doi.org/10.1146/annurev.mi.39.100185.001541

Strandwitz P, Kim KH, Terekhova D, Liu JK, Sharma A, Levering J, McDonald D, Dietrich D, Ramadhar TR, Lekbua A, Mroue N, Liston C, Stewart EJ, Dubin MJ, Zengler K, Knight R, Gilbert JA, Clardy J, Lewis K (2019) GABA-modulating bacteria of the human gut microbiota. Nat Microbiol 4(3):396–403. https://doi.org/10.1038/s41564-018-0307-3

Suez J, Zmora N, Zilberman-Schapira G, Mor U, Dori-Bachash M, Bashiardes S, Zur M, Regev-Lehavi D, Ben-Zeev Brik R, Federici S, Horn M, Cohen Y, Moor AE, Zeevi D, Korem T, Kotler E, Harmelin A, Itzkovitz S, Maharshak N, Shibolet O, Pevsner-Fischer M, Shapiro H, Sharon I, Halpern Z, Segal E, Elinav E (2018) Post-antibiotic gut mucosal microbiome reconstitution is impaired by probiotics and improved by autologous FMT. Cell 174 (6):1406–1423.e16. https://doi.org/10.1016/j.cell.2018.08.047

Takiishi T, Korf H, Van Belle TL, Robert S, Grieco FA, Caluwaerts S, Galleri L, Spagnuolo I, Steidler L, Van Huynegem K, Demetter P, Wasserfall C, Atkinson MA, Dotta F, Rottiers P, Gysemans C, Mathieu C (2012) Reversal of autoimmune diabetes by restoration of antigen-specific tolerance using genetically modified Lactococcus lactis in mice. J Clin Invest 122 (5):1717–1725. https://doi.org/10.1172/JCI60530

Tanoue T, Morita S, Plichta DR, Skelly AN, Suda W, Sugiura Y, Narushima S, Vlamakis H, Motoo I, Sugita K, Shiota A, Takeshita K, Yasuma-Mitobe K, Riethmacher D, Kaisho T, Norman JM, Mucida D, Suematsu M, Yaguchi T, Bucci V, Inoue T, Kawakami Y, Olle B, Roberts B, Hattori M, Xavier RJ, Atarashi K, Honda K (2019) A defined commensal consortium elicits CD8 T cells and anti-cancer immunity. Nature 565(7741):600–605. https://doi.org/10.1038/s41586-019-0878-z

Thompson LR, Sanders JG, McDonald D, Amir A, Ladau J, Locey KJ, Prill RJ, Tripathi A, Gibbons SM, Ackermann G, Navas-Molina JA, Janssen S, Kopylova E, Vazquez-Baeza Y, Gonzalez A, Morton JT, Mirarab S, Zech Xu Z, Jiang L, Haroon MF, Kanbar J, Zhu Q, Jin Song S, Kosciolek T, Bokulich NA, Lefler J, Brislawn CJ, Humphrey G, Owens SM, Hampton-Marcell J, Berg-Lyons D, McKenzie V, Fierer N, Fuhrman JA, Clauset A, Stevens RL, Shade A, Pollard KS, Goodwin KD, Jansson JK, Gilbert JA, Knight R, Earth Microbiome Project Consortium (2017) A communal catalogue reveals earth's multiscale microbial diversity. Nature 551(7681):457–463. https://doi.org/10.1038/nature24621

Thorn CF, Klein TE, Altman RB (2013) PharmGKB: the pharmacogenomics knowledge base. Methods Mol Biol 1015:311–320. https://doi.org/10.1007/978-1-62703-435-7_20

Trifan A, Stanciu C, Girleanu I, Stoica OC, Singeap AM, Maxim R, Chiriac SA, Ciobica A, Boiculese L (2017) Proton pump inhibitors therapy and risk of Clostridium difficile infection: systematic review and meta-analysis. World J Gastroenterol 23(35):6500–6515. https://doi.org/10.3748/wjg.v23.i35.6500

Turnbaugh PJ, Ley RE, Mahowald MA, Magrini V, Mardis ER, Gordon JI (2006) An obesity-associated gut microbiome with increased capacity for energy harvest. Nature 444 (7122):1027–1031. https://doi.org/10.1038/nature05414

Turnbaugh PJ, Hamady M, Yatsunenko T, Cantarel BL, Duncan A, Ley RE, Sogin ML, Jones WJ, Roe BA, Affourtit JP, Egholm M, Henrissat B, Heath AC, Knight R, Gordon JI (2009) A core gut microbiome in obese and lean twins. Nature 457(7228):480–484. https://doi.org/10.1038/nature07540

Tvede M, Tinggaard M, Helms M (2015) Rectal bacteriotherapy for recurrent Clostridium difficile-associated diarrhoea: results from a case series of 55 patients in Denmark 2000-2012. Clin Microbiol Infect 21(1):48–53. https://doi.org/10.1016/j.cmi.2014.07.003

van Nood E, Vrieze A, Nieuwdorp M, Fuentes S, Zoetendal EG, de Vos WM, Visser CE, Kuijper EJ, Bartelsman JF, Tijssen JG, Speelman P, Dijkgraaf MG, Keller JJ (2013) Duodenal infusion of donor feces for recurrent Clostridium difficile. N Engl J Med 368(5):407–415. https://doi.org/10.1056/NEJMoa1205037

Vatanen T, Kostic AD, d'Hennezel E, Siljander H, Franzosa EA, Yassour M, Kolde R, Vlamakis H, Arthur TD, Hamalainen AM, Peet A, Tillmann V, Uibo R, Mokurov S, Dorshakova N, Ilonen J, Virtanen SM, Szabo SJ, Porter JA, Lahdesmaki H, Huttenhower C, Gevers D, Cullen TW, Knip M, DIABIMMUNE Study Group, Xavier RJ (2016) Variation in microbiome LPS immunogenicity contributes to autoimmunity in humans. Cell 165(4):842–853. https://doi.org/10.1016/j.cell.2016.04.007

Viaud S, Saccheri F, Mignot G, Yamazaki T, Daillere R, Hannani D, Enot DP, Pfirschke C, Engblom C, Pittet MJ, Schlitzer A, Ginhoux F, Apetoh L, Chachaty E, Woerther PL, Eberl G, Berard M, Ecobichon C, Clermont D, Bizet C, Gaboriau-Routhiau V, Cerf-Bensussan N, Opolon P, Yessaad N, Vivier E, Ryffel B, Elson CO, Dore J, Kroemer G, Lepage P, Boneca IG, Ghiringhelli F, Zitvogel L (2013) The intestinal microbiota modulates the anticancer immune effects of cyclophosphamide. Science 342(6161):971–976. https://doi.org/10.1126/science.1240537

Wallace BD, Wang H, Lane KT, Scott JE, Orans J, Koo JS, Venkatesh M, Jobin C, Yeh LA, Mani S, Redinbo MR (2010) Alleviating cancer drug toxicity by inhibiting a bacterial enzyme. Science 330(6005):831–835. https://doi.org/10.1126/science.1191175

Wallace BD, Roberts AB, Pollet RM, Ingle JD, Biernat KA, Pellock SJ, Venkatesh MK, Guthrie L, O'Neal SK, Robinson SJ, Dollinger M, Figueroa E, McShane SR, Cohen RD, Jin J, Frye SV, Zamboni WC, Pepe-Ranney C, Mani S, Kelly L, Redinbo MR (2015) Structure and inhibition of microbiome beta-glucuronidases essential to the alleviation of cancer drug toxicity. Chem Biol 22(9):1238–1249. https://doi.org/10.1016/j.chembiol.2015.08.005

Walters WA, Xu Z, Knight R (2014) Meta-analyses of human gut microbes associated with obesity and IBD. FEBS Lett 588(22):4223–4233. https://doi.org/10.1016/j.febslet.2014.09.039

Walters W, Hyde ER, Berg-Lyons D, Ackermann G, Humphrey G, Parada A, Gilbert JA, Jansson JK, Caporaso JG, Fuhrman JA, Apprill A, Knight R (2016) Improved bacterial 16S rRNA gene (V4 and V4-5) and fungal internal transcribed spacer marker gene primers for microbial community surveys. mSystems 1(1). https://doi.org/10.1128/mSystems.00009-15

Wang M, Carver JJ, Phelan VV, Sanchez LM, Garg N, Peng Y, Nguyen DD, Watrous J, Kapono CA, Luzzatto-Knaan T, Porto C, Bouslimani A, Melnik AV, Meehan MJ, Liu WT, Crusemann M, Boudreau PD, Esquenazi E, Sandoval-Calderon M, Kersten RD, Pace LA, Quinn RA, Duncan KR, Hsu CC, Floros DJ, Gavilan RG, Kleigrewe K, Northen T, Dutton RJ, Parrot D, Carlson EE, Aigle B, Michelsen CF, Jelsbak L, Sohlenkamp C, Pevzner P, Edlund A, McLean J, Piel J, Murphy BT, Gerwick L, Liaw CC, Yang YL, Humpf HU, Maansson M, Keyzers RA, Sims AC, Johnson AR, Sidebottom AM, Sedio BE, Klitgaard A, Larson CB, A Boya PC, Torres-Mendoza D, Gonzalez DJ, Silva DB, Marques LM, Demarque DP, Pociute E, O'Neill EC, Briand E, Helfrich EJN, Granatosky EA, Glukhov E, Ryffel F, Houson H, Mohimani H, Kharbush JJ, Zeng Y, Vorholt JA, Kurita KL, Charusanti P, McPhail KL, Nielsen KF, Vuong L, Elfeki M, Traxler MF, Engene N, Koyama N, Vining OB, Baric R, Silva RR, Mascuch SJ, Tomasi S, Jenkins S, Macherla V, Hoffman T, Agarwal V, Williams PG, Dai J, Neupane R, Gurr J, Rodriguez AMC, Lamsa A, Zhang C, Dorrestein K, Duggan BM, Almaliti J, Allard PM, Phapale P, Nothias LF, Alexandrov T, Litaudon M, Wolfender JL, Kyle JE, Metz TO, Peryea T, Nguyen DT, VanLeer D, Shinn P, Jadhav A, Muller R, Waters KM, Shi W, Liu X, Zhang L, Knight R, Jensen PR, Palsson BO, Pogliano K, Linington RG, Gutierrez M, Lopes NP, Gerwick WH, Moore BS, Dorrestein PC, Bandeira N (2016) Sharing and community curation of mass spectrometry data with Global Natural Products Social Molecular Networking. Nat Biotechnol 34(8):828–837. https://doi.org/10.1038/nbt.3597

Wang M, Jarmusch AK, Vargas F, Aksenov AA, Gauglitz JM, Weldon K, Petras D, Silva RD, Quinn R, Melnik AV, van der Hooft JJJ, Caraballo Rodríguez AM, Nothias LF, Aceves CM, Panitchpakdi M, Brown E, Di Ottavio F, Sikora N, Elijah EO, Labarta-Bajo L, Gentry EC, Shalapour S, Kyle KE, Puckett SP, Watrous JD, Carpenter CS, Bouslimani A, Ernst M, Swafford AD, Zúñiga EI, Balunas MJ, Klassen JL, Loomba R, Knight R, Bandeira N,

Dorrestein PC (2019) MASST: a web-based basic mass spectrometry search tool for molecules to search public data. bioRxiv:591016. https://doi.org/10.1101/591016

West CE, Jenmalm MC, Kozyrskyj AL, Prescott SL (2016) Probiotics for treatment and primary prevention of allergic diseases and asthma: looking back and moving forward. Expert Rev Clin Immunol 12(6):625–639. https://doi.org/10.1586/1744666X.2016.1147955

Wishart DS (2016) Emerging applications of metabolomics in drug discovery and precision medicine. Nat Rev Drug Discov 15(7):473–484. https://doi.org/10.1038/nrd.2016.32

Woese CR, Fox GE (1977) Phylogenetic structure of the prokaryotic domain: the primary kingdoms. Proc Natl Acad Sci U S A 74(11):5088–5090. https://doi.org/10.1073/pnas.74.11.5088

Wu H, Esteve E, Tremaroli V, Khan MT, Caesar R, Manneras-Holm L, Stahlman M, Olsson LM, Serino M, Planas-Felix M, Xifra G, Mercader JM, Torrents D, Burcelin R, Ricart W, Perkins R, Fernandez-Real JM, Backhed F (2017) Metformin alters the gut microbiome of individuals with treatment-naive type 2 diabetes, contributing to the therapeutic effects of the drug. Nat Med 23 (7):850–858. https://doi.org/10.1038/nm.4345

Younis W, Thangamani S, Seleem MN (2015) Repurposing non-antimicrobial drugs and clinical molecules to treat bacterial infections. Curr Pharm Des 21(28):4106–4111. https://doi.org/10.2174/1381612821666150506154434

Zanvit P, Konkel JE, Jiao X, Kasagi S, Zhang D, Wu R, Chia C, Ajami NJ, Smith DP, Petrosino JF, Abbatiello B, Nakatsukasa H, Chen Q, Belkaid Y, Chen ZJ, Chen W (2015) Antibiotics in neonatal life increase murine susceptibility to experimental psoriasis. Nat Commun 6:8424. https://doi.org/10.1038/ncomms9424

Zeevi D, Korem T, Godneva A, Bar N, Kurilshikov A, Lotan-Pompan M, Weinberger A, Fu J, Wijmenga C, Zhernakova A, Segal E (2019) Structural variation in the gut microbiome associates with host health. Nature 568(7750):43–48. https://doi.org/10.1038/s41586-019-1065-y

Zeisel SH, Warrier M (2017) Trimethylamine N-oxide, the microbiome, and heart and kidney disease. Annu Rev Nutr 37:157–181. https://doi.org/10.1146/annurev-nutr-071816-064732

Zheng P, Zeng B, Zhou C, Liu M, Fang Z, Xu X, Zeng L, Chen J, Fan S, Du X, Zhang X, Yang D, Yang Y, Meng H, Li W, Melgiri ND, Licinio J, Wei H, Xie P (2016) Gut microbiome remodeling induces depressive-like behaviors through a pathway mediated by the host's metabolism. Mol Psychiatry 21(6):786–796. https://doi.org/10.1038/mp.2016.44

Zhu W, Winter MG, Byndloss MX, Spiga L, Duerkop BA, Hughes ER, Buttner L, de Lima Romao E, Behrendt CL, Lopez CA, Sifuentes-Dominguez L, Huff-Hardy K, Wilson RP, Gillis CC, Tukel C, Koh AY, Burstein E, Hooper LV, Baumler AJ, Winter SE (2018) Precision editing of the gut microbiota ameliorates colitis. Nature 553(7687):208–211. https://doi.org/10.1038/nature25172

Harnessing Human Microphysiology Systems as Key Experimental Models for Quantitative Systems Pharmacology

D. Lansing Taylor, Albert Gough, Mark E. Schurdak, Lawrence Vernetti, Chakra S. Chennubhotla, Daniel Lefever, Fen Pei, James R. Faeder, Timothy R. Lezon, Andrew M. Stern, and Ivet Bahar

Contents

1 Introduction 328
1.1 Quantitative Systems Pharmacology 329
1.2 Human Microphysiology Systems (MPS) 337
2 QSP Involves Iterative Application of Experimental and Computational Models 341
2.1 Identifying Differential Omics from Patient Samples 341
2.2 Inferring Pathways of Disease from Omics Data 343
2.3 Identifying Drugs, Targets, and Pathways by Machine Learning for Drug Repurposing and as a Starting Point for Developing Novel Therapeutics 344
2.4 Computational Models of Disease 346
2.5 Computational Models of ADME-Tox 346
3 Human Organ Microphysiology Systems (MPS) Complement Animal Models of Disease and ADME-Tox 348
3.1 Designing Human Organ MPS 348
3.2 Example of a Liver MPS 351
3.3 Human Liver MPS Experimental Model of Nonalcoholic Fatty Liver Disease 351
3.4 Testing Drugs in Human MPS 354
3.5 Critical Role of the Microphysiology Systems Database (MPS-Db) 354
4 Summary and Conclusions 355
References 356

D. L. Taylor (✉) · A. Gough · M. E. Schurdak · L. Vernetti · C. S. Chennubhotla · F. Pei · J. R. Faeder · T. R. Lezon · A. M. Stern · I. Bahar
University of Pittsburgh Drug Discovery Institute, Pittsburgh, PA, USA

Department of Computational and Systems Biology, University of Pittsburgh, Pittsburgh, PA, USA
e-mail: dltaylor@pitt.edu

D. Lefever
University of Pittsburgh Drug Discovery Institute, Pittsburgh, PA, USA

J. E. Barrett et al. (eds.), *Concepts and Principles of Pharmacology*,
Handbook of Experimental Pharmacology 260, https://doi.org/10.1007/164_2019_239

Abstract

Two technologies that have emerged in the last decade offer a new paradigm for modern pharmacology, as well as drug discovery and development. Quantitative systems pharmacology (QSP) is a complementary approach to traditional, target-centric pharmacology and drug discovery and is based on an iterative application of computational and systems biology methods with multiscale experimental methods, both of which include models of ADME-Tox and disease. QSP has emerged as a new approach due to the low efficiency of success in developing therapeutics based on the existing target-centric paradigm. Likewise, human microphysiology systems (MPS) are experimental models complementary to existing animal models and are based on the use of human primary cells, adult stem cells, and/or induced pluripotent stem cells (iPSCs) to mimic human tissues and organ functions/structures involved in disease and ADME-Tox. Human MPS experimental models have been developed to address the relatively low concordance of human disease and ADME-Tox with engineered, experimental animal models of disease. The integration of the QSP paradigm with the use of human MPS has the potential to enhance the process of drug discovery and development.

Keywords

Computational models of ADME-Tox · Computational models of disease · DILI · Drug development · Drug discovery · Drug repurposing · Induced pluripotent stem cells · Microphysiology systems · Omics analyses · PBPK · Personalized medicine · Quantitative systems pharmacology · Toxicology

1 Introduction

Over the last 30 years, the primary drug discovery and development paradigm has been based on target-centric discovery methods and the use of simple 2D cellular models along with animal models of disease and ADME-Tox (Sorger et al. 2011; Stern et al. 2016). Although some very valuable therapeutics have been discovered and delivered to patients based on this paradigm, the efficiency has been very low. In fact, after the investment of significant time and money, the failure rate is still ca. 80% for those new drug candidates that enter phase 2 clinical trials (Arrowsmith and Miller 2013), although in recent years there has been some improvement concurrent with an increase in the percentage of biologics and a more critical triage of candidates (Smietana et al. 2016). The primary causes of failure have been identified as a lack of efficacy with some unpredicted toxicity (Alex et al. 2015; Arrowsmith and Miller 2013). This knowledge has led to a widely held view that there is need for a new paradigm, together with the use of more sophisticated human multicellular, 3D experimental tissue/organ models (Sorger et al. 2011; Stern et al. 2016). This chapter explores the application of QSP as an alternative approach to

drug discovery and development and the role of human MPS (e.g., organs-on-a-chip) to complement animal models of disease and ADME-Tox in the practice of QSP.

1.1 Quantitative Systems Pharmacology

The identification of small molecules that modulate disease-relevant animal phenotypic models, which can then serve both as probes of pathophysiology and starting points for drug discovery and development, has its roots in classical pharmacology (Sorger et al. 2011). Historically, the structure-activity profiles of these small molecules were used to classify receptors, infer their biological functions, and then guide their molecular characterization (i.e., "target identification") (Ahlquist 1948; Black et al. 1972; Lands et al. 1967; Sorger et al. 2011). Recent technological advancements including genomics and high-content screening (phenotypic screening) have resulted in the development of more sophisticated experimental models (human cells, human 3D, MPS models, and small organisms) exhibiting quantifiable, clinically relevant features (phenotypes) (Haasen et al. 2017; Horvath et al. 2016; Taylor 2012). Phenotypic discovery has been enabled with more extensive structurally and mechanistically diverse chemical libraries (https://drugdiscovery.msu.edu/facilities/addrc/compound-libraries/) and orthogonal methods that facilitate target identification of phenotypic screening "hits" (Mateus et al. 2016; Schenone et al. 2013). Juxtaposed to classical pharmacology, recent pharmacology has evolved an alternative reductionist approach, exploiting phenotypic screening, extensive datasets, chemical libraries, and antibody collections, to identify small molecules and biologics that directly target well-annotated receptors to effect specific biological responses (Sorger et al. 2011).

Phenotypic and target-centric pharmacological approaches are complementary, can be used in tandem, and have resulted in the approval of nearly 4,000 drugs having a profound benefit on human health worldwide. The reduction in mortality among a large segment of our population, through the pharmacological management of cardiovascular risk factors, and the modification of the lethal HIV infection into a clinically manageable chronic disease through combination therapy targeting the virus life cycle are two remarkable examples among many. Despite this success, diseases range widely in their complexity and prevalence, from cancers, opioid addiction, Alzheimer's disease, type 2 diabetes, and nonalcoholic fatty liver disease (NAFLD) to the more than 7,000 rare diseases for which there are no effective treatments. To address this extensive unmet need, the field of pharmacology continues to evolve, incorporating an explosion of knowledge and unprecedented advances in technology in this post-genomic era, into a modular, highly integrated platform termed quantitative systems pharmacology (QSP) (Gadkar et al. 2016a; Hansen and Iyengar 2013; Iyengar et al. 2012; Sorger et al. 2011; Stern et al. 2016; Zhao and Iyengar 2012).

QSP focuses on determining disease mechanisms and drug modes of action and their intrinsic relationships, to facilitate the repurposing of existing drugs, as well as

to develop novel therapeutics and therapeutic strategies. Discovering and developing therapeutics is a multiscale challenge. QSP addresses this challenge through the iterative use of experimental and computational models, starting with analysis of human clinical data and continuing with the analysis of molecular results from in vitro experimental and animal models relative to the human data.

Historically, advancements in the development of preclinical human models of barrier epithelia, coupled with the development of mathematical models of pharmacokinetics, improved our ability to predict human pharmacokinetic (PK) profiles, optimizing compound and dose selection for early stages of clinical development (Ferreira and Andricopulo 2019; Kola and Landis 2004; Sorger et al. 2011; Torras et al. 2018). The combined use of human, cell-based experimental models and computational models of physiologically based pharmacokinetics (PBPK) paved the way for overcoming a major hurdle in traditional drug development (Lave et al. 2016; Rowland et al. 2011). Addressing this particular challenge has nevertheless unmasked others. Today, attrition in the drug discovery pipeline results mainly from lack of efficacy in phase 2, as well as toxicity that can be observed at any stage of development including post-market surveillance (phase 4). A lack of efficacy can be observed despite evidence for drug-target engagement and that toxicity has often been determined to be on-mechanism, suggesting that medicinal chemistry per se is not limiting drug development, but that our knowledge of underlying human biological mechanisms and pathophysiology is insufficient.

Paradoxically, it appears that for complex diseases a phenotypic approach may be particularly useful in combination with elements of the target-based approach (Haasen et al. 2017). This observation suggests, at least for certain diseases and targets, that cellular context at the level of network regulation may be an important determinant for achieving efficacy and that a more comprehensive systems-based approach, in contrast to a focused yet restrictive target-based approach, may be indicated to identify emergent biology (Hopkins 2008). Likewise, mechanism-based toxicity can also be context-dependent, resulting from the expression of the target in different tissues/organs or in the presence of particular comorbidities (Ferdinandy et al. 2018; Marnett 2009). In addition to these intricacies, successful drug development requires the study of a drug candidate's mode of action in systems that span a wide range of biological complexity and diversity (i.e., from purified subcellular components to patient populations), involving timescales from milliseconds to life spans (Sorger et al. 2011). Together these considerations emphasize the need to make comprehensive systems-based measurements in experimental models that are also iteratively coupled to computational models, resulting in predictions that lead to new experiments. This iterative process leads to the refinement of the computational models with the goal to define mathematically the alterations leading to disease and toxicity (Woodhead et al. 2017). In QSP, hypotheses are tested across experimental models of increasing clinical relevance, including increasingly sophisticated human cell-based MPS, to help verify a mechanistic link between drug mode of action and the underlying pathophysiology in patients (see Sect. 1.2). The identification of pharmacodynamic (PD) markers that take into account context-dependent emergent properties to quantitate drug-target interactions and reliably predict efficacy is an

important deliverable for QSP. It is important to note that the pharmaceutical industry has been implementing QSP (Visser et al. 2014).

Figure 1 describes the implementation of QSP based on a platform for repurposing drugs and developing novel therapeutics. QSP begins with a focus on *patient sample analytics* (Fig. 1a) where patient biospecimens (adjacent "normal" and disease biopsies and longitudinal samples where possible) are obtained and analyzed by methods including DNA sequencing, transcriptomics, epigenetics, proteomics, metabolomics, and computational pathology (Spagnolo et al. 2016, 2017). Despite the challenge of obtaining longitudinal tissue biopsies, single time point measurements can often provide valuable insights into disease progression. For example, in the case of rapidly evolving diseases such as metastatic cancer, mutational analysis of the primary tumor and patient-matched metastases could indicate those clones from the primary tumor with the highest metastatic potential, routes of dissemination from one site to the other, and the selection of therapy-resistant clones (Macintyre et al. 2017). More generally, noninvasive blood sampling and single cell "omics" (Keating et al. 2018) can be used to compute real- and pseudo-time trajectories (Trapnell et al. 2014), and an in situ proteomics-based computational pathology platform (Keating et al. 2018) can exploit spatial heterogeneity to infer the evolution of disease-associated phenotypes.

These analyses are used to *infer pathways of disease progression* (Fig. 1b). In the example of RNASeq from "normal" and disease tissue samples, the outputs from comprehensive data analyses are differentially expressed gene sets that can be used to infer disease-associated pathways through the implementation of validated systems-based computational tools (Ge et al. 2018; Lee et al. 2008). Large numbers of inferred pathways can be reduced to a smaller most significant number by applying thresholding statistics and allow making inferences on causal molecular networks (Hill et al. 2016). The selected pathways allow identification of known molecular targets that serve as candidate targets for pharmacological and/or genomic perturbations to investigate disease mechanism. Although superficially this stage of QSP implementation may appear to take on the character of the target-centric approach, there are significant differences that may ultimately bear on the high rate of attrition due to lack of efficacy. For example, many candidate targets, in contrast to one or a limited few, are being considered in parallel, and each is inferred from a comprehensive unbiased dataset derived directly from patient samples.

Machine learning (ML) tools can then *predict drug-target interactions* (*DTIs*) or chemical-target interactions (Fig. 1c) from databases such as DrugBank (Wishart et al. 2018) and STITCH (Kuhn et al. 2010), in order to identify a focused library of disease mechanism "probes" that includes known and predicted drugs (Chen et al. 2016; Cobanoglu et al. 2013, 2015; Keiser et al. 2009; Liu et al. 2016) and is complemented by RNAi- and cDNA-based probes (Martz et al. 2014).

The predicted drug/chemical "probes" are then investigated as *test drugs/chemicals in human MPS* (Fig. 1d) that recapitulate critical functions of normal organs and clinically relevant disease states. The disease state MPS can be constructed using patient-derived cells and/or by exposure to established disease-potentiating environmental factors (see Sect. 1.2 below). MPS experimental models

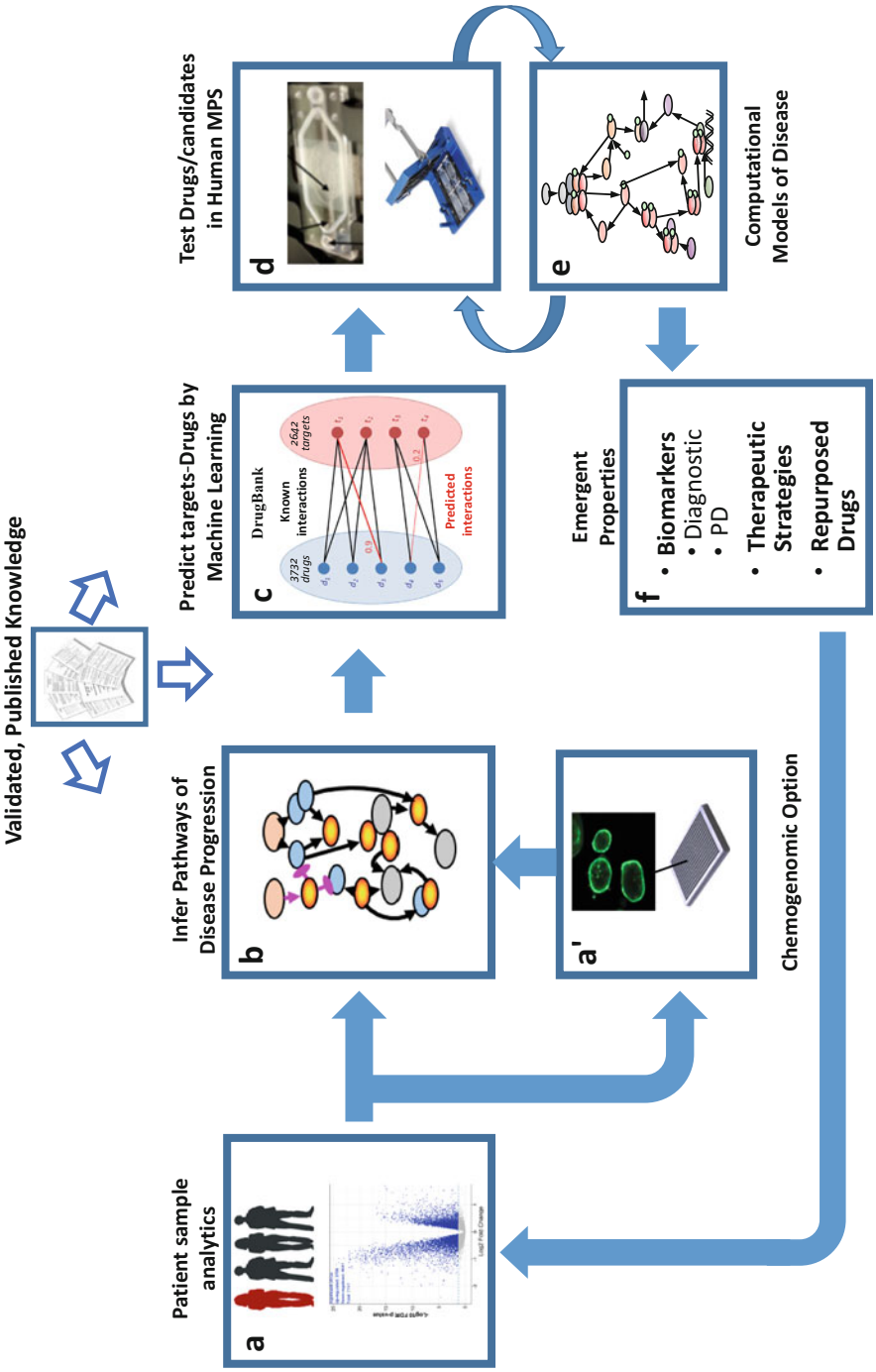


Fig. 1 Quantitative systems pharmacology platform for repurposing drugs and developing novel therapeutics

are amenable to phenotypic screening with the set of "probes" using high-content screening (HCS) platforms to quantify disease-specific phenotypes. The goal of these "probe" studies is to identify probes or probe combinations that reverse the phenotype or genotype of the disease experimental models back to "normal." The selection of "probes" can be expanded through the use of computational medicinal chemistry tools such as homology modeling, druggability assessment (Bakan et al. 2012; Volkamer et al. 2012), pharmacophore modeling (Sanders et al. 2012), and molecular simulations (De Vivo et al. 2016). The predicted "probes" can also be modified through medicinal chemistry to identify drug candidates that selectively modulate specific molecular targets rather than the canonical targets. The quantification of pharmacodynamic and disease-modifying effects of each probe enables drug mode of action to be studied in relation to disease mechanism. Successful probes from the in vitro studies can also be tested in animal models of disease. Probes or probe combinations that are approved drugs can be the starting point for drug repurposing.

The datasets resulting from use of the "probes," coupled with publication-validated knowledge, are used to construct *computational models of disease* (Fig. 1e), which are refined and optimized through iterative experimental and computational analyses (Sorger et al. 2011). The computational models can make predictions based on selected perturbations, and these can be tested in the experimental models (e.g., using well-annotated drug sets and gene/protein knockdown studies).

The computational models ultimately predict *emergent properties* (Fig. 1f), including diagnostic and pharmacodynamic biomarkers associated with the disease, and therapeutic strategies (including drug combinations) that utilize novel and/or repurposed drugs. These strategies can be tested in personalized MPS experimental disease models using patient-derived cells (primary, adult stem cell-derived, and induced pluripotent stem cells (iPSCs)) in a "preclinical trial" on a range of patient genetic and disease backgrounds (see Sect. 1.2 below). The results from the "preclinical trial" studies and clinical trial data are used to refine hypotheses of the mechanisms of disease progression. Since some of the measurements made in the experimental models are the same as those made in the patient samples, biomarkers identified in the model can be retrospectively analyzed in the patient samples to cross-validate the preclinical studies and establish a strong rationale for clinical trial design. In the case of rare diseases where biospecimens may be scarce, the implementation of QSP could be initiated with MPS models. Furthermore, as discussed below, MPS models could be used to predict both on-mechanism and off-target toxicities (Vernetti et al. 2016). It is important to note that validated, published knowledge about the disease, targets, pathways, biomarkers, and drugs can be used as input information at any point in the QSP platform (Stern et al. 2016). Selected key examples of applying QSP in developing therapeutic strategies are presented in Table 1.

Chemogenomic option (Fig. 1a'). A specific chemogenomic version of the platform can be applied at the beginning of the pipeline, preferably using higher-throughput human, 3D models of disease. In silico chemogenomic approach

Table 1 Selected key examples of applying QSP in developing therapeutic strategies

Authors	Title
Schoeberl et al. (2009)	Therapeutically targeting ErbB3: a key node in ligand-induced activation of the ErbB receptor-PI3K axis
Gadkar et al. (2016b)	Evaluation of HDL-modulating interventions for cardiovascular risk reduction using a systems pharmacology approach
Dziuba et al. (2014)	Modeling effects of SGLT-2 inhibitor dapagliflozin treatment versus standard diabetes therapy on cardiovascular and microvascular outcomes
Howell et al. (2014)	A mechanistic model of drug-induced liver injury aids the interpretation of elevated liver transaminase levels in a phase 1 clinical trial
Pei et al. (2017)	Connecting neuronal cell protective pathways and drug combinations in a Huntington's disease model through the application of quantitative systems pharmacology
Vaidya et al. (2019)	Combining multiscale experimental and computational systems pharmacological approaches to overcome resistance to HER2-targeted therapy in breast cancer
Yin et al. (2018)	Quantitative systems pharmacology analysis of drug combination and scaling to humans: the interaction between noradrenaline and vasopressin in vasoconstriction
Pei et al. (2019)	Quantitative systems pharmacological analysis of drugs of abuse reveals the pleiotropy of their targets and the effector role of mTORC1

(Fig. 1a'–f), inferring the molecular mechanisms of a phenotype of interest based on a collection of chemicals identified through phenotypic screening, offers an alternative framework to identify novel therapeutics (Bredel and Jacoby 2004; Brennan et al. 2009; Digles et al. 2016; Pei et al. 2017; Prathipati and Mizuguchi 2016). The collection of chemicals is used therein to mine the DTI or chemical-target interaction databases (Gaulton et al. 2017; Kooistra et al. 2016; Szklarczyk et al. 2016; Wishart et al. 2018) and extract ML-based information on associated targets (Cobanoglu et al. 2015; Gfeller et al. 2014; Nickel et al. 2014; Yamanishi et al. 2014), which may be further linked to enriched pathways and gene ontology (GO) annotations (Huntley et al. 2015; Kanehisa et al. 2017; Slenter et al. 2018). Thus, the cellular pathways and environment and the biological functions and processes affected by the chemicals are systematically explored. Such system-level analyses (Bian et al. 2019; Pei et al. 2017, 2019; Wei et al. 2018; Wu et al. 2019; Xu et al. 2016) assist in deciphering polypharmacological effects and disease mechanisms (Fig. 2).

1.1.1 Challenges and Opportunities in Applying QSP

The unprecedented molecular and cellular characterization of patient samples, in conjunction with well-documented electronic health records, enables the comprehensive and unbiased QSP platform to determine complex disease mechanisms and inform optimal therapeutic strategies, including the identification of emergent properties (Fig. 1). The paradigm of iterative experimental and computational modeling provides testable mechanistic hypotheses serving to connect the actual pathogenesis to the ensemble of modules comprising QSP, despite the large spatiotemporal scales they encompass. For example, the presence in the MPS models of the disease-specific pathways inferred from the patient data can be determined,

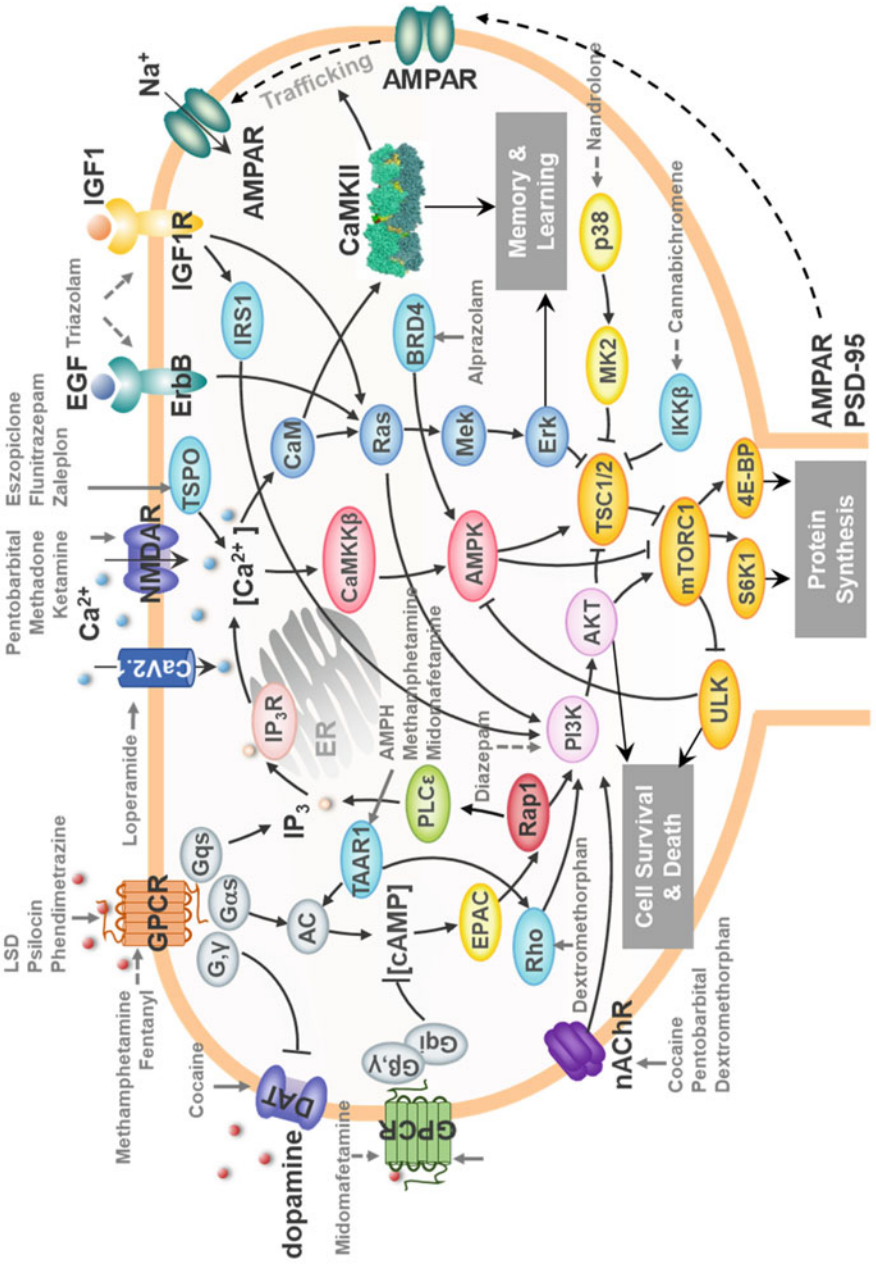


Fig. 2 A unified signaling network generated through the chemogenomic approach (see Fig. 1a' and Pei et al. 2017, 2019) in investigation of drugs of abuse. *Black arrows* represent the activation, inhibition, and translocation events during signal transduction. *Solid gray arrows* represent the known drug-target

Fig. 2 (continued) interactions. *Dashed gray arrows* represent predicted drug-target interactions. The diagram illustrates the targets of several drugs of abuse belonging to different categories: loperamide, fentanyl, heroin, morphine, and methadone from opioids; midomafetamine, ketamine, dextromethorphan, LSD, and psilocin from hallucinogens; triazolam, diazepam, alprazolam, pentobarbital, eszopiclone, flunitrazepam, and zaleplon from CNS depressants; cannabichromene, 2-AG, cannabinol, and dronabinol from cannabinoids; methamphetamine, cocaine, AMPH, and phendimetrazine from CNS stimulants; and nandrolone from anabolic steroids. mTORC1 emerges as a hub where the effects on several targets of addictive drugs appear to be consolidated to lead to cell death and/or protein synthesis in the CNS and in particular AMPAR/PSD95 synthesis that induces morphological changes in the dendrites. Figure originally published in Pei et al. (2019)

thereby providing one level of cross-validation. Chemical and genetic probes predicted to modulate these pathways can then be tested in the MPS models to determine their effect on disease phenotypes recapitulated in the model, providing a second level of clinical relevance. In parallel, systems modeling of these inferred pathways could be used to predict disease-specific biomarker profiles that can form the rationale for an observational study, establishing yet another critical connection between the preclinical model and the patient. Finally, epidemiological analysis of clinical outcome in those patients being treated for a comorbidity could provide complementary evidence for a particular disease mechanism and could establish a strong basis for drug repurposing. The QSP platform provides the critical nexus between pharmacodynamic markers and disease mechanism that promises to reduce attrition and facilitate the regulatory process. This focus on connecting disease mechanism with drug mode of action enables QSP to be effective for identifying drugs that can be repurposed and for optimizing combination therapies, particularly for the treatment of complex diseases. This approach also functions as a starting point for harnessing medicinal chemistry to evolve novel therapeutics that have higher specificity and efficacy than the DTI taking advantage of the polypharmacology of drugs. A key target other than the canonical target may be critical, and the "other" target engagements can be optimized.

The strength of QSP lies in its transdisciplinary approach, and this presents its greatest challenge, in the form of organizational barriers. For the full potential of QSP to be realized, multidisciplinary teams need to be assembled, likely across two or more institutions, under leaders with expertise not only in one particular field but also possessing the sophisticated set of skills to manage critical interfaces across several disciplines. The requirement for this paradigm shift in basic research and translational medicine is increasingly being recognized by industry, academia, and government. Consequently, we anticipate that the organizational barriers to full implementation of QSP will be significantly reduced.

1.2 Human Microphysiology Systems (MPS)

The development of human MPS has grown out of the recognition that animal models and simple 2D monocultures of cells do not reflect the complexity and specificity of human physiology, toxicology, and disease mechanisms (Hartung 2009; Seok et al. 2013; Sorger et al. 2011; Stern et al. 2016). The challenge has been to develop in vitro human experimental models using patient-derived cells, either primary, tissue-resident adult stem cells (AdSCs), embryonic stem cells (ESCs), or induced pluripotent stem cells (iPSCs), that recapitulate enough tissue/organ functions to serve as useful models in the drug discovery and development pipeline. There is the added potential to create a personalized platform for preclinical trials using these patient-derived cells. Another opportunity is to evolve these personalized MPS models into tissue replacement therapeutics (Xie and Tang 2016). The use of HCS methods to acquire temporal-spatial information and quantitative phenotypes from the 3D, multicellular MPS systems has been critical (Stern et al. 2016; Taylor 2012).

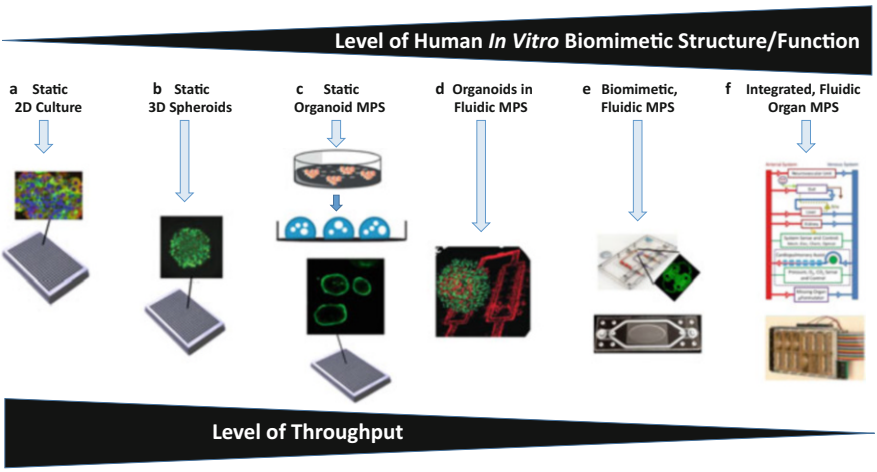


Fig. 3 Human in vitro experimental models span a broad range of experimental throughput and biomimetic structure and function

Figure 3 illustrates a range of in vitro human experimental models that increase in tissue/organ biomimetic structure and function from left to right. A common goal in the discovery and development pipeline is to select the optimal model for the stage ("fit-for-purpose"). The continuum in Fig. 3a–f involves sacrificing throughput vs biomimetic complexity.

Static 2D cultures (Fig. 3a), typically consisting of cell lines analyzed in microplates, have dominated biomedical research for over 30 years. These static 2D cultures have been used extensively in high-throughput screening (HTS) and high-content screening (HCS) applications for target ID, target validation, screening, hit to lead, and early toxicology testing (Fang and Eglen 2017). However, there has been a shift away from static 2D cultures since they do not adequately reflect human physiology/pathology. The more recent use of 3D experimental models dates back to the early 1900s, and the historical timeline of the evolution of 3D approaches has recently been summarized (Simian and Bissell 2017).

Static 3D spheroids (Fig. 3b) were originally developed by Sutherland and collaborators to better recapitulate the functional phenotypes of human cancer cells in response to radiation therapy and as a general model for tumors (Fang and Eglen 2017; Sutherland et al. 1970). Spheroids are produced by a variety of methods that form spheres of cells (ca. 100–400 um diameter) that have more physiological cell-cell and cell-matrix interactions and can generate gradients of nutrients, oxygen, signaling molecules, and metabolites from the outer layers of cells to the center, mimicking a solid tumor better than 2D models (Fang and Eglen 2017). Most spheroids use a single cell type such as cancer cells or hepatocytes, but it is possible to construct spheroids with more than one cell type. Static 3D spheroids have been used for both HTS and HCS in microplates, but confocal imaging is required for single cell resolution.

Static organoid MPS (Fig. 3c) have been defined in multiple ways (Simian and Bissell 2017); however, we prefer the following broad definition: an organoid is an in vitro 3D cellular cluster derived exclusively from primary tissue, ESCs, AdSCs, or iPSCs, optimally capable of self-renewal and self-organization, and exhibiting similar organ functionality as the tissue of origin (slightly modified from Fatehullah et al. 2016). Static organoids developed from either AdSCs, ESCs, iPSCs, or primary patient cells (sometimes mixed with some human cell lines for selected cell types) recreate a partial biomimetic for many types of organs, which can be used for drug discovery, development, and the exploration of disease mechanisms (Dutta et al. 2017; Low and Tagle 2017; McCauley and Wells 2017; Prestigiacomo et al. 2017; Schwartz et al. 2015; Shamir and Ewald 2014; Skardal et al. 2016; van den Berg et al. 2019). Like static spheroids, static organoids can be investigated by HTS and HCS in microplates using confocal imaging.

The same principles used in developing static organoids can be applied to *organoids in fluidic MPS* (Fig. 3d). In fact, it is possible to combine the organoid technology with *biomimetic, fluidic MPS* (Fig. 3e) engineering principles to better address the limitations of each approach (Edington et al. 2018; Takebe et al. 2017; Wevers et al. 2016). Organoids in fluidic MPS enable the physiologically relevant shear stress required by many tissues, coupled with at least partial spatial cell-cell and cell-matrix interactions. They can also be linked to other organ MPS for integrated functions using fluidic connections.

Biomimetic, fluidic MPS (Fig. 3e) are devices designed to maximize physiologically relevant structural relationships between cells, natural gradients of physiological parameters (e.g., oxygen tension, hormones), matrix materials, mechanical cues including shear stress of vascular flow, mechanical movements, innervation, and immune system communication (Bhatia and Ingber 2014; Low and Tagle 2017; Watson et al. 2017). The focus is on constructing an organ model that is as close to a functional unit (e.g., liver acinus, cardiac muscle fibers, lung) as possible (Huh et al. 2010; Li et al. 2018; Lind et al. 2017). Early advances were stimulated in particular by research of Don Ingber and his colleagues at the Harvard Wyss Institute, including a lung biomimetic, fluidic MPS (Bhatia and Ingber 2014; Huh et al. 2010).

Presently, static organoids, organoids in fluidic MPS, and biomimetic, fluidic MPS are being created for most normal and diseased organs (Esch et al. 2015; Fang and Eglen 2017; Low and Tagle 2017; van den Berg et al. 2019). The complexity of current biomimetic, fluidic MPS models is not amenable to high-throughput studies, yet the high content of the structure and functionality are optimal to validate findings from higher-throughput models, as well as to investigate the mechanisms of disease progression using HCS over extended time periods (ca. 1 month or longer).

The most ambitious platform is the *integrated, fluidic organ MPS* (Fig. 3f) where multiple organ MPS are linked together either functionally (Vernetti et al. 2017) or physically (Edington et al. 2018; Low and Tagle 2017; Oleaga et al. 2019; Satoh et al. 2017; Skardal et al. 2016). Michael Shuler and his colleagues have been pioneers, demonstrating in the 1990s that linking multiple organ systems allowed organ-organ communications that could be used to identify toxicity and to perform physiologically based pharmacokinetics (PBPK) (Sin et al. 2004; Sweeney et al.

Table 2 Selected examples of human experimental MPS disease models

Authors	Title
Jain et al. (2018)	Primary human lung alveolus-on-a-chip model of intravascular thrombosis for assessment of therapeutic clinical pharmacology and therapeutics
Blutt et al. (2017)	Gastrointestinal microphysiological systems
Workman et al. (2017)	Engineered human pluripotent stem cell-derived intestinal tissues with a functional enteric nervous system
Hachey and Hughes (2018)	Applications of tumor chip technology
Clark et al. (2018)	A model of dormant-emergent metastatic breast cancer progression enabling exploration of biomarker signatures
Vernetti et al. (2016)	A human liver microphysiology platform for investigating physiology, drug safety, and disease models
Atchison et al. (2017)	A tissue-engineered blood vessel model of Hutchinson-Gilford progeria syndrome using human iPSC-derived smooth muscle cells

1995). This concept has been extended and applied with an integrated, fluidic organ MPS to explore ADME and PK/PD using QSP approaches (Yu et al. 2015). There are many challenges and opportunities in developing and applying these "body-on-a-chip" systems, but the progress over the last 5 years has been impressive (Low and Tagle 2017; Shuler 2017; Skardal et al. 2016; Wikswo et al. 2013b).

A consortium of pharmaceutical company representatives (IQ Consortium), participating in the National Center for Advancing Translational Sciences (NCATS) microphysiology systems program, recently wrote an article discussing the translation of MPS models from the laboratory to commercial use by the pharmaceutical industry (Ewart et al. 2017). It is clear that the industry understands the great potential of these systems and is giving important guidance to the field. In addition, the FDA and the EPA have collaborated with NCATS to learn of the potential of these systems and to provide their insights into the needed functionalities and reproducibility. Furthermore, the dramatic advances in the development of the biology, materials science, and microfluidics have led to the formation of numerous companies offering platforms that will accelerate the biomedical sciences, drug industry, and clinical applications based on some emerging standards (Zhang and Radisic 2017). Recently, May et al. (2017) explored the advantages and disadvantages of organoids, biomimetic, fluidic MPS, and integrated, fluidic organ MPS. Table 2 lists selected examples of human experimental MPS disease models.

1.2.1 Challenges and Opportunities in Developing and Applying MPS

The MPS field has exploded during the last 5 years, and dramatic advances have been made in microfluidic devices, in-line as well as cellular biosensors, the development of renewable cells from iPSCs (where there is still the need to mature the iPSCs to the adult genotype and phenotype), optimizing matrix biochemical content and stiffness, the optimization of media for normal and disease states in different organs, the exploration of a "universal" medium for connected organs, the

involvement of the innate and adaptive immune systems, as well as the role of chemical, electrical, and mechanical cues on functions. NCATS has involved the pharmaceutical industry, the FDA, and the EPA in the MPS programs, and there has been great feedback to guide developments. Further technical developments, as well as the demonstration of reproducibility of the models from day to day and between distinct sites, will position MPS to have a major impact on the drug discovery and development process, as well as to help to define the progression of diseases in human, in vitro models. MPS models are projected by many to refine, reduce, and ultimately replace animal models of disease and ADME-Tox sometime in the future.

2 QSP Involves Iterative Application of Experimental and Computational Models

The iterative use of experimental and computational models of disease and ADME-Tox is the hallmark of the practice of QSP (Fig. 1). This section discusses in more detail the key role of computational methods in the QSP platform, while Sect. 3 discusses in more detail the application of MPS in the QSP platform.

2.1 Identifying Differential Omics from Patient Samples

2.1.1 Early Omics and Implications for Human Disease: The GWAS Era

Omics generally refers to technologies that profile the entirety of the biological domain of interest (Hasin et al. 2017), which allows the investigator to take an unbiased data-driven, instead of a focused hypothesis-driven, approach to research. The first omics field to emerge was genomics, driven by the "SNP chip" (reviewed by LaFramboise 2009), which allowed high-throughput genotyping of individuals across common variants, termed genome-wide association studies (GWAS) (Visscher et al. 2012a). Some early-disease GWAS results were translational successes. The best example is age-related macular degeneration, where over half the disease heritability was explained by the GWAS results that guided drug discovery (Black and Clark 2016). However, this was not true for other complex diseases as the results could only explain a tiny portion of heritability. For schizophrenia (Visscher et al. 2012b) and obesity (Weedon et al. 2006), only 1–2% of heritability could be attributed to the GWAS-identified SNPs (Visscher et al. 2012a). This limitation applies to complex traits as well. For example, a study examining height across 253,288 individuals found 697 SNPs, which together explained ~20% of heritability (Wood et al. 2014). Further, the effect sizes of identified SNPs from most GWAS are typically vanishingly small, which necessitates huge sample sizes (Visscher et al. 2012a). Taken together, the leading paradigm is that complex diseases are polygenic and are therefore caused by complex interactions of genes as opposed to single genes (Wray et al. 2018). While the knowledge obtained using

GWAS has provided priceless insights into the biology of complex disease and drug discovery (Floris et al. 2018), it is only one piece of the puzzle.

2.1.2 Post-GWAS Era Omics Technologies and Strategies for Their Use in Human Disease

GWAS results often implicate numerous variants which have some degree of association with the disease; however, a mechanistic understanding of how these variants contribute to the disease phenotype remains largely incomplete (Wray et al. 2018). Other omics technologies (summarized in Table 3) offer the chance to close the gap left by GWAS (Karczewski and Snyder 2018). For example, the independent role of the epigenome in type 1 diabetes was shown by identifying differentially methylated regions using monozygotic twins as case controls (Paul et al. 2016). In another example, transcriptome data from patients with inflammatory bowel disease was used to identify potentially repurposable drugs (Dudley et al. 2011). Metabolomics has emerged relatively recently and shows great potential in further characterizing human disease (Wishart 2016). Lipidomics is another discipline which gained importance in the last decade with advances in mass spectrometry, driven by the tight association of lipids with many diseases including cardiovascular diseases, diabetes, stroke, NAFLD, neurological disorders, and cancer (Yang and Han 2016).

Since the cost of omics technologies continues to fall, investigators are increasingly combining multiple types of omics to obtain a more complete picture of the underlying biology (Hasin et al. 2017; Karczewski and Snyder 2018). One approach is to combine gene expression profiling with GWAS to identify quantitative trait loci, that is, variants which are associated with gene expression (Karczewski and Snyder 2018). A number of studies, reviewed in Sharma et al. (2015), have tied several genes – most notably PNPLA3 – to NAFLD progression. Interestingly, metabolic profiles associated with risk variants do not directly correlate with risk of disease (Sliz et al. 2018), underscoring the complex nature of the disease and the value of complementary multi-omics approaches for studying NAFLD.

Table 3 List of key omics analyses and reviews

Domain	Applications	Reference
DNA	Genome: Variant calling, GWAS, SNPs	Laurie et al. (2016) and He et al. (2017)
	Epigenome: Chip-Seq, BS-seq	Bailey et al. (2013) and Kurdyukov and Bullock (2016)
RNA	Transcriptome: gene expression profiling	Conesa et al. (2016) and Koch et al. (2018)
Protein	Proteomics: protein abundance	Larance and Lamond (2015)
Metabolites	Metabolomics: metabolite abundance	Johnson et al. (2016) and Wishart (2016)
Lipids	Lipidomics: lipid classes and pathways	Yang and Han (2016)

Most of the early omics technologies have been based on tissue samples that do not preserve the spatial relationships between cells, matrix, and tissue structures (e.g., blood vessels, ducts) and "average" the analyses among many cells. For example, the omics sampling of tumor samples has until recently relied on cores of tissue that do not consider the spatial heterogeneity in the tumors. We now understand that heterogeneity within a tumor is critical to understanding the evolution of the tumor. Recently, a variety of single cell methods have emerged to address this challenge (Keating et al. 2018). One of these methods is hyperplexed fluorescence imaging (Gough et al. 2014; Spagnolo et al. 2016, 2017). This method is based on computational and systems pathology using iterative fluorescence labeling of specific targets within formalin-fixed paraffin-embedded (FFPE) tissue sections or tissue microarrays (TMAs), imaging, quenching of the fluorescence, and then repeating the cycle for dozens of biomarkers in the same sample (Gerdes et al. 2013). Spatial analytics are then applied to the samples (Spagnolo et al. 2016). This method preserves the spatial relationships within tissues while allowing omics analyses based on the spatial connections within microdomains. Recently, this platform has been applied to a colon cancer patient cohort and a risk recurrence prognostic analysis demonstrated (Uttam et al. 2019).

2.1.3 Remaining Challenges of Using Omics to Study Human Disease

While omics continue to further our understanding of human disease, there are remaining challenges. Omics datasets are high-dimensional, with many more observations (e.g., genes, metabolites, proteins, lipids, etc.) assayed than samples (e.g., patients) taken (Teschendorff 2018). However, there has been considerable effort in developing specialized statistical methods for omics including the development of mixed graphical models (Manatakis et al. 2018) for learning disease models. Batch effects or technical confounding factors introduced by experimental design have historically been (Lambert and Black 2012) and continue to be (Goh et al. 2017; Goh and Wong 2018) the bane of omics studies. In a study examining epigenome of obese men as compared to lean controls, the authors found ~5.5% to be differentially methylated; however, these differences were entirely attributable to batch effects (Buhule et al. 2014). In another dramatic example, failure to account for technical variability introduced by using a different platform for the cases and controls led to the retraction (Sebastiani et al. 2011) of a paper originally published in *Science* (Sebastiani et al. 2010). There is great potential power in the use of omics approaches, but significant controls and data optimization steps are required.

2.2 Inferring Pathways of Disease from Omics Data

There is increasing interest in using patient-derived omics data for drug discovery to help increase therapeutic efficiency (Floris et al. 2018; Hodos et al. 2016). GWAS data has been used to help guide drug discovery; however, these data alone do not usually provide sufficient information for rational drug design (Pushpakom et al. 2018). Gene expression data can be an excellent type of omics to use for drug

discovery, and transcriptomic data was found to predict drug sensitivity of breast cancer cells better than genomic, epigenomic, and proteomic data (Costello et al. 2014). Other omics data still have their value: proteomics data, for example, provide details on posttranslational modifications that are not visible at the transcript level yet may provide insights into the nature of signaling in disease (Erdem et al. 2016).

Inferring pathways of disease progression begins with defining the difference between "diseased" and "healthy" states in terms of specific omics measurements. For example, in transcriptomic analysis, one might identify differentially expressed genes (DEGs) as those genes with transcript levels that change significantly between disease samples and healthy controls. Exactly defining "diseased" and "healthy" states themselves however is often difficult due to the inherent noise of biological data and inter-sample variability. Once statistically significant differences between diseased and healthy states are identified, the biological mechanisms that give rise to these differences can be hypothesized. For example, pathways containing higher than expected numbers of DEGs are commonly implicated in disease progression and subject to further investigation. Similarly, pathways upstream of transcriptional regulators of DEGs may also be implicated in disease progression. Connectivity mapping can then be used to find drugs which "reverse" the gene expression pattern (Musa et al. 2017).

2.3 Identifying Drugs, Targets, and Pathways by Machine Learning for Drug Repurposing and as a Starting Point for Developing Novel Therapeutics

Drug repurposing (also known as drug repositioning) refers to the process of identifying new therapeutic indications for approved drugs or investigational drugs (Allarakhia 2013; Ashburn and Thor 2004; Keiser et al. 2009). Drug repurposing takes advantage of established pharmacology of existing drugs to drastically reduce risk and cost of development, making it an attractive track for drug discovery and development (Pushpakom et al. 2018). A well-known example of drug repurposing is the anticancer drug imatinib, which was originally developed in 2001 for the treatment of chronic myeloid leukemia and, later in 2008, approved by the US Food and Drug Administration (FDA) for treating gastrointestinal stromal tumors (Al-Hadiya et al. 2014). A key step of drug repurposing is to identify new DTIs. However, experimental identification of DTIs is time-consuming, costly, and limited. For example, the current version (v5.1.1) of DrugBank (Wishart et al. 2018) contains data on 16,959 interactions between 10,562 drugs and 4,493 targets, while the presence or absence of the remaining interactions (99.96% of the complete space of interactions) is yet to be determined. Therefore, developing machine learning (ML)-based computational methods (Fig. 1c) for efficient DTI prediction is of great need.

To date, both supervised and semi-supervised ML methods have been adopted in DTI predictions (Chen et al. 2016, 2018). Most supervised learning methods, including kernel regression (Yamanishi et al. 2008), random forest (Cao et al.

2014), bipartite local models (Bleakley and Yamanishi 2009), regularized least-square classifier (van Laarhoven et al. 2011), kernelized Bayesian matrix factorization, and similarity-based deep learning (Zong et al. 2017), use the known DTIs as positive samples and consider the rest as negative ones. The structural and physicochemical properties of drugs, such as 2D fingerprints, 3D conformations, topological descriptors, and the sequence, structure, and expression data of targets such as protein sequence and structural motifs and gene expression profiles, are utilized to generate feature vectors of drugs or targets or to calculate drug-drug similarities and target-target similarities. Other supervised learning methods such as probabilistic matrix factorization (Cobanoglu et al. 2013) and integrated neighborhood-based method (Chen et al. 2016) utilize the known DTI patterns to compute drug-drug similarities and target-target similarities and predict novel DTIs, independent of the structural or physicochemical properties of drugs and targets. Semi-supervised methods, on the other hand, use labeled data (known DTIs) to infer labels for unknown DTIs, and these inferred DTIs play a role in the training process. Examples include the manifold Laplacian regularized least-square method (Xia et al. 2010) based on integrated data from known DTIs, chemical structures and genomic sequences, and the deep learning-based framework (Wen et al. 2017).

Most current ML-based methods simply regard DTI as an on-off relationship. Development of selective and potent drugs may require further consideration of specific binding poses and affinities. ML-based DTI prediction serves as a first step for identifying new associations, while further computational biophysical and medicinal chemistry tools help characterize the mechanistic aspects and specificities of predicted DTIs. For example, if the drug-binding site on the target is unclear or new (e.g., allosteric) sites beyond those (orthosteric) traditionally targeted are of interest, a useful method of approach is to perform druggability simulations (Bakan et al. 2012; Ivetac and McCammon 2012; Lexa and Carlson 2011; Loving et al. 2014). These simulations are conducted in the presence of a series of probes representative of drug-like fragments, whose simulated binding properties disclose the high-affinity binding sites as well as favorable binding poses on the target. Statistical analysis of these binding events permits us to build pharmacophore models (see, e.g., Bakan et al. 2015; Mustata et al. 2009), which, in turn, are used for screening virtual libraries of small compounds and identifying best matching compounds, termed "hits." Top hits identified at this stage are experimentally tested (e.g., via binding affinity assays (Pollard 2010)), and the feedback from experiments is used to revise computational models. In addition, with a set of bioactive hits, a numerical description of molecular structure/properties to known biological activity can be generated via quantitative structure-activity relationship (QSAR) (Wang et al. 2015) analysis, which further guides the rational structural optimization of the hits into lead compounds. The combined computational and experimental methods are performed iteratively until the refinement of the compounds to achieve desirable biological activity in the MPS models.

2.4 Computational Models of Disease

A central element of QSP is the iterative computational/experimental feedback loop. In order to understand the biological mechanisms of disease onset and progression, it is helpful to formalize certain aspects of the experimental system into a mathematical model that can be manipulated in silico. When dealing with in vitro systems, computational disease models are usually limited to interactions within and between a small number of cells and most often take the form of either agent-based models (ABMs) or systems of ordinary differential equations (ODEs). In an ABM, each biological cell is represented as an autonomous entity that interacts with its environment and neighboring cells according to pre-defined rules. The behavior of the system as a whole is therefore an emergent property of this collection of agents. Although computationally more expensive than ODE models, ABMs are easily interpretable in terms of cellular features and are readily adaptable to novel geometries such as those found in MPS experiments. ABMs have been used to explore a range of diseases, including tumor growth (Szabo and Merks 2013) and liver fibrosis (Dutta-Moscato et al. 2014). ODE models are typically higher resolution than ABMs and represent the system at the level of molecules rather than cells. As they are computationally efficient and mathematically straightforward, these are a popular choice for modeling signaling pathways and regulatory networks. The standard ODE approach assumes that the molecular components of cellular chemistry are contained in a well-mixed system that obeys mass action kinetics, although more complex, spatially realistic models (represented by partial differential equations, PDEs) and/or stochastic models (described by stochastic differential equations) are gaining popularity, especially in the description of complex microphysiological processes (e.g., MCell for modeling synaptic transmission) (Bartol et al. 2015; Kaya et al. 2018).

In the context of the computational model, the difference between "diseased" and "healthy" states arises from changes in parameters, such as reaction rates or molecule numbers. For example, differences in computationally predicted transcript profiles between healthy and diseased cells might arise as the result of an altered binding affinity and/or posttranslational modification in the computational model (Fig. 4). Changes in the computational model that promote the disease phenotype indicate hypothetical mechanisms of disease progression. If rectifying these changes (e.g., via drugs) in the in vitro system reverses the disease state, then the computational model has successfully identified a disease mechanism; if not, then the computational model is refined, and another hypothesis is generated and experimentally tested. For example, by using separate compartments, an ODE was able to capture the effects of liver zonation on steatosis (Ashworth et al. 2016).

2.5 Computational Models of ADME-Tox

Since the days of Fortran programs such as MODFIT (Allen 1990), drug discovery researchers recognized the advantages in storing, managing, and analyzing large

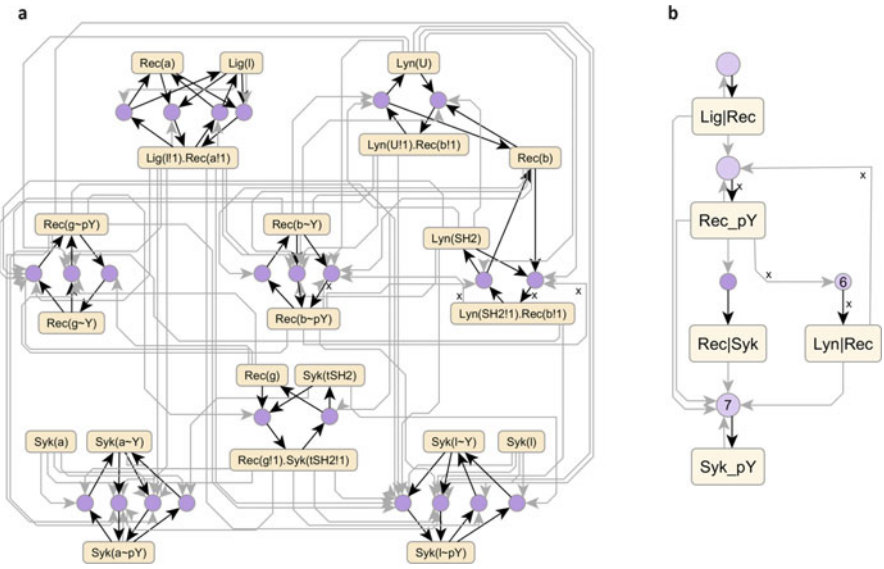


Fig. 4 Two views of a detailed computational model of immunoreceptor signaling mediated by the high-affinity receptor for IgE (Fc epsilon R1). Panel (**a**) shows the molecular components (yellow rectangles) and processes (purple circles) that govern the flow of activity in the network. Each process represents either a binding interaction between the components or posttranslational modification of a component (e.g., phosphorylation). Enormous complexity is generated just from the basic interactions that include binding and phosphorylation. Although this complexity does not limit our ability to simulate the dynamics of such systems, it does limit our ability to understand the dynamics. Through a process of static analysis, we can reduce the complexity and interpret the dynamics in terms of simple motifs and mechanisms, such as the positive feedback loop that is illustrated in panel (**b**) (edges marked with "x"). Modified from Sekar et al. (2017)

amounts of pharmacokinetic (PK) data. Drug developers require computational tools that have a good correlation between in silico, in vitro, and in vivo absorption, distribution, metabolism, and excretion (ADME) data to address the challenges of predicting PK behavior in drug development in order to determine dosing regimens, target organ exposure, and identify compounds or their reactive metabolites for off-target liabilities. The current computational modeling for drug development is evolving from simple classical PK compartmental models that describe the disposition of drugs in the body and the component ADME properties to physiologically based PK (PBPK) models that predict PK based on the physiochemical properties of the drugs and knowledge of the physiology of the organism. Although the concept of PBPK modeling has been around since 1937, it is the relatively recent advances in computing power and preclinical physiologic data that enable effective PBPK modeling. Computational approaches now include in silico predictors for drug metabolism, pharmacokinetics, and toxicology using ordinary differential equations, machine learning neural networks, Bayesian, recursive partitioning, and support vector machine algorithms (Byvatov et al. 2003; Hou et al. 2001; Li et al. 2007;

Muller et al. 2005; Sadowski and Kubinyi 1998; Wagener and van Geerestein 2000; Walters and Murcko 2002; Zernov et al. 2003). There are many commercial and academic computational tools available (R Project, GastroPlus, DILIsym, Simcyp, and MATLAB among others) for PBPK prediction (Lin et al. 2017; Tan et al. 2018; Zhuang and Lu 2016). Toxicology tools based on R-group structural alerts (DEREK), QSAR (MC4PC, MDL-QSAR, TopKat, and ADMET Predictor® among others) and molecular descriptors (PaDel) are also being developed for predicting human organ and systemic toxicity (Chen et al. 2014; Wu and Wang 2018).

All of these computational models depend on the availability of experimental data accurately representing the clinical physiology. The advanced physiological relevance of human MPS models is well suited to providing such data. In particular, liver MPS models are useful in predicting intrinsic hepatic clearance, which can then be applied to predict other PK parameters (Ewart et al. 2018; Tsamandouras et al. 2017). In addition to predicting PK, data from MPS models also allow for modeling of pharmacodynamic (PD) properties, enabling PK/PD modeling to guide drug development decisions. Finally, as MPS models can utilize patient-specific cells, PK/PD and toxicology modeling can be applied to individual genetic and physiologic backgrounds to guide the development of precision medicine models (Tsamandouras et al. 2017). The combination of MPS models and the advancing computational modeling will aid in reducing the time and cost of preclinical drug discovery.

3 Human Organ Microphysiology Systems (MPS) Complement Animal Models of Disease and ADME-Tox

As discussed above, the minimal concordance between animal models of disease and toxic liabilities, and human disease and toxicity, is one of the factors in the low success rate for drug candidates entering phase 2 clinical trials. However, animal models are still the gold standard in research and development; and regulatory agencies still require animal data before going into humans. Continued developments in the MPS field have the potential to initially complement animal models and then refine, reduce, and ultimately replace animal testing.

3.1 Designing Human Organ MPS

As illustrated in Fig. 2, MPS models include a continuum from simple 2D models to complex, integrated multi-organ systems. The design and implementation of an MPS is a “systems engineering” challenge that must take into account the complete platform consisting of microdevices, control systems, cells, extracellular matrices, media, readouts, and data analysis (Wikswo et al. 2013a). This becomes more important and challenging when integrating multiple organ MPS, requiring consideration of issues of organ scaling, sequencing, media composition, volume, and flow

(Wikswo et al. 2013b). The rapid growth in the development of MPS is partially driving, and partially driven by, the rapid development of component technologies, which provides a diversity of choices, but can also complicate the design and optimization of the model. The ultimate goal is to create a multi-organ human-on-a-chip that will recapitulate a wide range of human physiology for experimentally modeling complex systemic diseases and toxicities, but such a complex model is not needed for many studies. Because all experimental models have limitations, and the simplest model that provides the required information is usually the best choice, perhaps the most important considerations in designing an MPS are how the model will be used and what the key functional indications will be.

Models can be roughly divided into two types: (1) self-assembly models that range from cells spreading on a 2D substrate to multilayer organoids in fluidic chambers and (2) biomimetic models in which the design of the device and/or the assembly of the model promotes cellular organization that mimics the in vivo organization. Generally, self-assembly models are easier to apply in high-throughput applications, while biomimetic models provide deeper functional information. In either case, many choices go into the design of an MPS. Here, we will focus on the design of biomimetic models, though many of the same considerations apply to simpler models.

A major focus in the development of biomimetic models is the engineering of the device to recapitulate the organization of cells in vivo and also, in some cases, to engineer active elements that mimic functions such as breathing in the lung (Huh et al. 2010), contraction of muscle (Truskey et al. 2013), the beating of the heart (Benam et al. 2015; Lind et al. 2017), as well as others. To facilitate the prototyping of these systems, polydimethylsiloxane (PDMS) has been the material of choice due to low cost and ease of rapid casting in a laboratory setting. PDMS is also oxygen permeant, reducing the need to provide for additional oxygenation in the design of the model. However, PDMS is hydrophobic and readily absorbs hydrophobic molecules including some drugs and other test molecules, especially those with a higher logP and few or no hydrogen-bond donor groups (Auner et al. 2019). There are now many commercial devices that are glass and/or plastic, reducing the likelihood of compound binding (Lenguito et al. 2017; Ribas et al. 2018). Existing commercial devices have less flexibility for customizing model architecture and require more attention to oxygenation of the cells in the model, but many have already been used to implement specific organ models and therefore provide a good starting point for design or development. Driving flow in the MPS is also an important consideration and has been accomplished by using gravity, either through rocking or media transfers between outlet and inlet, pressurized systems, syringe pumps, and peristaltic pumps. In all cases, it is important that the pumping system can provide the required range of flow rates and that a physiological shear stress on the model tissues is attained.

Because a major goal in the development of MPS is to model human physiology, the focus has been on the use of human cells. While there is some interest in developing MPS models using animal cells, both for validation of the model with respect to the larger number of compounds that have been tested in animal models

and for the prediction of preclinical animal safety, relatively few MPS have been constructed with animal cells. For human cells, the choice is between primary cells, ESC, AdSC, iPSC, and cell lines. Primary cells are still the gold standard for adult-like human organ function, but specific functions may vary from donor to donor, and therefore a specific lot of cells may need to be selected and then used for the duration of a project to minimize variability from the cells. iPSCs hold promise to provide an unlimited supply of human cells, including isogenic cells for models with multiple cell types, but improved protocols to generate adult-like cells are still in development (Besser et al. 2018). The inclusion of one or more human cell lines in a multicell MPS is still an attractive option for higher-throughput applications or where the functional role of the cell type is adequately provided by a cell line. The media used in an MPS is typically selected to support the cells used, but this can become difficult in multicellular models where different cell types require different media compositions. This is further complicated in integrated organ systems. Mixing media has been one approach (Vernetti et al. 2017), but there is some evidence that creating vascularized organ systems with media that is optimized for the endothelial cells in the vascular channel, while parenchymal cells are perfused with a cell-specific media, may be a good solution, especially in coupled organ systems. In addition to the media consideration, selecting and optimizing an appropriate extracellular matrix (ECM) material is important in most models. Collagen 1 is widely used, but other hydrogels have also been used, and achieving physiologically relevant biochemistry and stiffness has been shown to be an important factor in some models (Barry et al. 2017; Kalli and Stylianopoulos 2018; Sun et al. 2018).

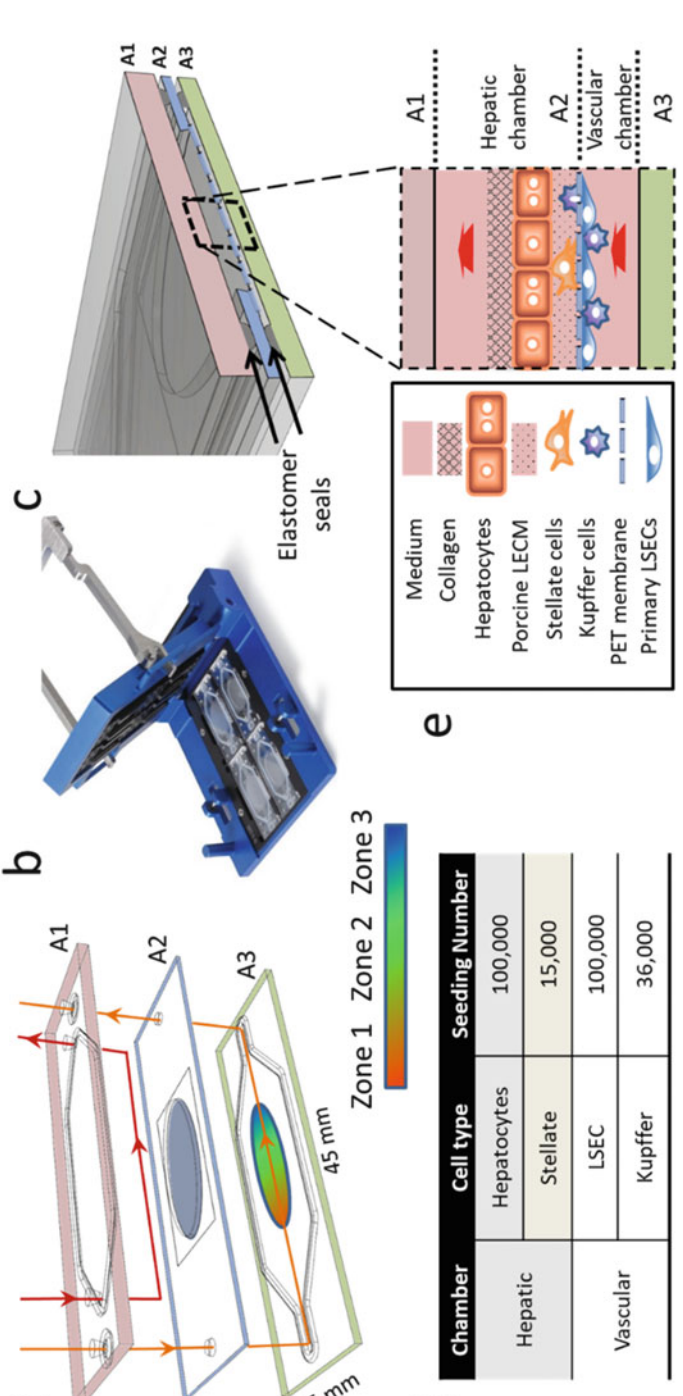
The most important aspect of an MPS is the functional performance in the particular application. A wide range of assay types have been developed and used in MPS to demonstrate basic organ functions as well as disease- and toxicity-associated responses. From a systems perspective, it is important to consider the planned readouts in the design of the model. Readouts in MPS often include secreted factors (proteins, cytokines, free fatty acids, etc.), imaging, biosensors, expression profiling, metabolism, and spatial characterization. Sampling the media efflux or from the media recirculation in microfluidic systems is typically sufficient to allow assays of secreted factors, metabolites, and cytokines, although sensitivity may be limiting in systems with high flow rates or large media volumes, key considerations in system design. Imaging, especially with the many commercial and custom biosensors (Newman and Zhang 2014; Senutovitch et al. 2015), can provide important real-time functional readouts including cell tracking, protein expression, ion concentration, enzyme activity, ROS, apoptosis, and other functions, provided the device design supports online imaging. Imaging of the 3D spatial relationships in the model can be important in establishing the organization of the cells, and interrogating subsets of the cells, such as the growth of cancer cells in an organ model of a metastatic niche (Miedel et al. 2019; Rao et al. 2019). For high-resolution confocal imaging, it is important that the device is constructed with an optical-quality, coverslip-thick "window" through which to image the cells and that the cells in the device are within the working distance of the objective, which may be >1 mm at $20\times$ and <0.2 mm at $40\times$ (Vernetti et al. 2016).

3.2 Example of a Liver MPS

The optimal MPS design will likely result from an evolution of models of increasing capability with respect to organ functions and its intended use (Beckwitt et al. 2018; Clark et al. 2016). As an example, the vascularized liver acinus MPS (vLAMPS) model (Fig. 5) currently in use at the University of Pittsburgh (Li et al. 2018) started as a micro-grooved prototype cast from PDMS and bonded to a glass coverslip for imaging (Bhushan et al. 2013). Although the prototype was functional by several metrics, the connections were unreliable, and the evaporation rate from the large surface area of PDMS was too high. To address these issues, we moved the model into the Nortis (Seattle, WA) chip, which is also cast from PDMS and attached to a coverslip but encased in plastic with metal ferrules for tubing connections. The robustness of this device provided a reproducible model for further optimization that included, along with the primary human hepatocytes and endothelial cells, the addition of human stellate and Kupffer-like cells. This model was shown to be stable out to 28 days and provides multiple functional readouts. It responded appropriately to toxic compounds (binding of test compounds to PDMS was tested); exhibited canalicular efflux, a fibrotic response (Vernetti et al. 2016); and supported the development and validation of multiple biosensors (Senutovitch et al. 2015). Further development of this model included the addition of a space of Disse using a porcine liver ECM, the incorporation of liver-specific endothelial cells, and alteration of flow rates, by which the oxygen tension in the device could be controlled to simulate oxygen zones in the liver, enabling the demonstration of zone-specific biology (Lee-Montiel et al. 2017; Soto-Gutierrez et al. 2017). However, the oxygen permeability of the PDMS made it difficult to create the continuous zonation of the in vivo liver and complicated the use of the device for screening compounds, due to the potential for absorption discussed above. Furthermore, although the model was successfully used to demonstrate organ-organ interactions (Vernetti et al. 2017), the lack of a vascular channel limited the prospects for direct coupling with other organ models, a key application for a metabolically competent liver model. To address these limitations, the model was transferred to the Micronit (Enschede, Netherlands) organ-on-a-chip platform which is glass, supports continuous oxygen zonation, and has a vascular channel for connection to other organs, as well as introduction of circulating immune cells (Li et al. 2018). Presently, multiple liver disease models, both stand-alone liver MPS and liver coupled to other organ MPS, containing isogenic primary cells or iPSC-derived cells from normal and diseased patients are in development. The liver biomimetic MPS will continue to evolve based on technological advances.

3.3 Human Liver MPS Experimental Model of Nonalcoholic Fatty Liver Disease

The liver performs ca. 500 critical functions making it vulnerable to many diseases including NAFLD, a disorder that is rapidly increasing in parallel with the



Chamber	Cell type	Seeding Number
Hepatic	Hepatocytes	100,000
	Stellate	15,000
Vascular	LSEC	100,000
	Kupffer	36,000

Fig. 5 The vascularized liver acinus microphysiology system (vLAMPS). (**a**) The vLAMPS model is assembled in a three-layer glass microfluidic device from Micronit. The center layer (A2) has an 8 × 16 mm elliptical hole with a porous PET membrane on which matrices and cells are layered. The media flow in the hepatic and vascular chambers, combined with the oxygen consumption by the hepatocytes, creates an oxygen gradient mimicking the in vivo liver acinus, creating Zones 1–3 microenvironments. (**b**) The three layers are held together in a clamp for robust connections and imaging. (**c**) The independent flow channels are sealed with elastomer. (**d**) The proportions of the four human cell types used to construct the model were chosen based on the proportions in the human liver. (**e**) The organization of the cells and matrices in the assembled model. Adapted from Li et al. (2018)

worldwide obesity and diabetes epidemics. NAFLD encompasses a spectrum of liver damage ranging from simple steatosis (or NAFL) to nonalcoholic steatohepatitis (NASH), cirrhosis, and hepatocellular carcinoma (HCC). Genetic and environmental factors, as well as disease drivers such as inflammatory cytokines, including adipokines, bacterial products, and metabolites originating from the intestine and adipose tissue, contribute to the development and progression of NAFLD (Satapathy and Sanyal 2015). The pathologic hallmarks of NAFLD include steatosis, inflammatory infiltrate, fibrosis, and hepatocyte ballooning, leading to decreased hepatocellular functions and eventually cirrhosis and HCC (Jain et al. 2015; Pacana and Sanyal 2015).

The application of QSP to NAFLD starts with studies of patient data (Fig. 1a) that have identified several NAFLD associated single-nucleotide polymorphisms (SNPs) (Speliotes et al. 2011) and gene signatures. The most relevant and reproducible SNP identified across GWAS studies is in the patatin-like phospholipase domain-containing 3 (PNPLA3) gene (rs738409 C>G p.I1e148Met) which is strongly associated with hepatic steatosis, fibrosis, cirrhosis, and HCC (Schulze et al. 2015). Despite its strong association with NAFLD, the functional significance of the PNPLA3 SNP is unknown. A major limitation in the elucidation of a mechanistic role of PNPLA3 in NAFLD has been the interspecies differences in its expression and tissue-specific distribution (Anstee and Day 2013). In particular, PNPLA3 is expressed predominantly in the human liver, whereas in mice it is mainly expressed in the adipose tissue (Smagris et al. 2015). Therefore, human patient-derived MPS are needed to study the pathogenesis of NAFLD and to test novel therapies. The use of molecular manipulation technologies is currently being used to engineer human IPSCs for specific gene knockouts and knock-ins to generate specific genetic disease models (Wu et al. 2018).

Based on network inference (Erdem et al. 2016; Grimes et al. 2019; Lezon et al. 2006; Subramanian et al. 2005), molecular interactions and signaling pathways are identified (Fig. 1b) that may be involved in the progression of early NAFLD. A consensus gene network is constructed using published interaction information (*see* Fig. 1), including regulation, protein-protein interactions, and functional relationships that are used to define on the network a disease neighborhood. Pathways containing members of the disease neighborhood are flagged as potentially disease-associated. Potential DTIs in these pathways are computationally predicted with a latent factor model such as BalestraWeb (Fig. 1c) (Cobanoglu et al. 2015), and then predicted drugs are screened in the vLAMPS models (Fig. 1d), along with compounds currently in development, to identify drugs that halt and/or reverse the disease phenotypes.

The experimental strategy is to recapitulate the early stages of human NAFLD progression (NAFL and NASH) in the vLAMPS experimental models using primary human hepatocytes and non-parenchymal cells (LSECs, stellate and Kupffer cells) from patients. The models are investigated over a 1-month period with and without the addition of known molecular and cellular drivers of NAFLD progression. Early NAFLD models are compared with normal liver models, along with clinical findings in the MPS-Db (see below) using a panel of phenotypic/functional measures.

vLAMPS models are also post-processed to produce H&E stained sections to compare the pathology with the original patient tissue. The resulting data are used to create computational models of disease progression (Fig. 1e) that are iteratively used to refine the selection of biomarkers and potential therapeutics (Fig. 1f).

3.4 Testing Drugs in Human MPS

Human MPS models are projected to have great potential to bridge the efficacy gap between animals and humans by offering drug testing in a complex, physiologically relevant human organ or multi-organ system. For many decades, animal models have served the pharmaceutical industry well for testing single target therapeutics for antibiotics, blood pressure control, or cholesterol reduction but were ineffective or even misleading when testing compounds for complex human diseases such as cancer, obesity, liver diseases, and neural degenerative diseases (van der Worp et al. 2010). Although the biomimetic fluidic MPS platforms are not high-throughput at this time, progress is being made in that direction (Satoh et al. 2017; Trietsch et al. 2013; Wevers et al. 2016). Importantly though, many biomimetic MPS models have been tested and shown to be sufficiently robust and repeatable for routine use in compound testing (Sakolish et al. 2018). Progress toward confirming correlation between the test systems and human safety and efficacy is expected to reduce the number of drugs that fail in clinical trials, despite promising findings in preclinical test species (Cirit and Stokes 2018). Preclinical animal models for toxicity assessment are still required, despite multiple examples of lead compounds that failed in clinical trials due to toxicity and despite demonstrated safety in animal models. Human MPS organ models and multi-organ models will increasingly be used along with animal models for toxicology assessment and disease efficacy models. Finally, biologic therapies such as peptides, proteins, antibodies, and cells are notoriously difficult to assess for safety liabilities in the standard preclinical toxicology models due to foreign antigen recognition and immune response. Here, again, human MPS models will offer a convenient and species-specific method to assess off-target liabilities.

3.5 Critical Role of the Microphysiology Systems Database (MPS-Db)

To accelerate the development and application of MPS in the biopharmaceutical and pharmaceutical industries, as well as in basic biomedical research, a centralized resource is required to manage the detailed design, application, and performance data that enables industry and research scientists to select, optimize, and/or develop new MPS solutions. We have built and implemented a microphysiology systems database (Gough et al. 2016) which is an open-source, simple icon-driven interface as a resource for MPS researchers (accessible at https://mps.csb.pitt.edu). The MPS-Db enables users to design and implement multifactor, multichip studies,

capture and standardize MPS experimental data and metadata (description of the experimental design and conditions), and provide tools to analyze, model, and interpret results in the context of human physiology and toxicology. The MPS-Db is designed to capture and aggregate data from multiple organ models using any type of platform from microplates to sophisticated, microfluidic devices and associate that data with reference data from chemical, biochemical, preclinical, clinical, and post-marketing sources, in order to support the design, development, validation, and interpretation of organ models. A key benefit of the MPS-Db is the standardization of metadata and data, which simplifies intra- and inter-study comparison of the results for testing and validating the performance of MPS models.

The vision for the MPS-Db is to support all MPS technologies, from organ model design to applications. Portals have been developed to aid in the design of organ and disease models by linking to databases to collate information on organs or disease biology, along with MPS data. This new information, together with the existing links to compound and clinical information, enables the user to more efficiently design and analyze proof-of-concept studies, in order to establish the model performance. Independent validation of the models is supported by tools to design studies, for example, by selecting compounds and concentrations to test and distributing those to the chips in the study, identify the best or most relevant clinical readouts, and apply statistical tools to assess the reproducibility of the model. Links to clinical data enable evaluation of clinical concordance and developing physiologically based pharmacokinetics (PBPK) models that will provide a basis for predicting exposure and clearance.

In summary, the MPS-Db supports data providers (e.g., academic and industry researchers) with tools to capture, manage, and disseminate data from experimental models, and data consumers (e.g., researchers and regulatory agencies) with a platform to analyze data and interpret results in the context of human physiology, and design computational and experimental models and studies. The variety and types of data collected and incorporated into the MPS-Db allow scientists to build predictive tools that will link the pathways or molecular events of drug toxicity and efficacy to higher-order pathways, cells, tissues, and organs. The MPS-Db is an innovative advancement for the MPS community and is the first and only publicly accessible, comprehensive resource for sharing and disseminating data and information on MPS.

4 Summary and Conclusions

The last decade has seen an explosion in the number of computational studies in the field of quantitative systems pharmacology, with the realization that current challenges in drug discovery and development require approaches well beyond traditional chemically driven efforts at the single-molecule level. In parallel, there has been progress in experimental models from classical animal models to human microphysiology systems (MPS) based on the use of human primary cells, adult stem cells, and/or iPSCs, as a powerful tool to mimic not only the structure and

morphology of human cells, tissue, and organs but also their biological or physiological functions. The combined use of these novel computational and experimental methods, complemented by classical PK and PD approaches, QSAR analyses, and ADME-Tox assessments, holds promise for overcoming the attrition effect that has long stalled progress in rational design of new therapies.

Acknowledgments Support from the National Institutes of Health awards UG3DK119973 (DLT), R01DK0017781 (DLT), 1UO1 TR002383 (DLT), U24TR002632 (AG/MS), SBIR HHSN271201800008C UO1CA204826 (CC), DA035778 (IB), and P41 GM103712 (IB) is gratefully acknowledged. We also thank the members of the University of Pittsburgh Drug Discovery Institute, the Department of Computational and Systems Biology, and other collaborators at the University of Pittsburgh and beyond for critical discussions.

References

Ahlquist RP (1948) A study of the adrenotropic receptors. Am J Phys 153:586–600. https://doi.org/10.1152/ajplegacy.1948.153.3.586

Alex A, Harris CJ, Keighley WW, Smith DA (2015) Compound Attrition in Phase II/III. In: Attrition in the pharmaceutical industry. https://doi.org/10.1002/9781118819586.ch4

Al-Hadiya BM, Bakheit AH, Abd-Elgalil AA (2014) Imatinib mesylate. Profiles Drug Subst Excip Relat Methodol 39:265–297. https://doi.org/10.1016/b978-0-12-800173-8.00006-4

Allarakhia M (2013) Open-source approaches for the repurposing of existing or failed candidate drugs: learning from and applying the lessons across diseases. Drug Des Devel Ther 7:753–766. https://doi.org/10.2147/dddt.S46289

Allen GD (1990) MODFIT: a pharmacokinetics computer program. Biopharm Drug Dispos 11:477–498

Anstee QM, Day CP (2013) The genetics of NAFLD. Nat Rev Gastroenterol Hepatol 10:645–655. https://doi.org/10.1038/nrgastro.2013.182

Arrowsmith J, Miller P (2013) Trial watch: phase II and phase III attrition rates 2011-2012. Nat Rev Drug Discov 12:569. https://doi.org/10.1038/nrd4090

Ashburn TT, Thor KB (2004) Drug repositioning: identifying and developing new uses for existing drugs. Nat Rev Drug Discov 3:673–683. https://doi.org/10.1038/nrd1468

Ashworth WB, Davies NA, Bogle ID (2016) A computational model of hepatic energy metabolism: understanding zonated damage and steatosis in NAFLD. PLoS Comput Biol 12:e1005105. https://doi.org/10.1371/journal.pcbi.1005105

Atchison L, Zhang H, Cao K, Truskey GA (2017) A tissue engineered blood vessel model of hutchinson-gilford progeria syndrome using human iPSC-derived smooth muscle cells. Sci Rep 7:8168. https://doi.org/10.1038/s41598-017-08632-4

Auner A, Tasneem KM, Markov DA, McCawley LJ, Hutson MS (2019) Chemical-PDMS binding kinetics and implications for bioavailability in microfluidic devices. Lab Chip. https://doi.org/10.1039/C8LC00796A

Bailey T et al (2013) Practical guidelines for the comprehensive analysis of ChIP-seq data. PLoS Comput Biol 9:e1003326. https://doi.org/10.1371/journal.pcbi.1003326

Bakan A, Nevins N, Lakdawala AS, Bahar I (2012) Druggability assessment of allosteric proteins by dynamics simulations in the presence of probe molecules. J Chem Theory Comput 8:2435–2447. https://doi.org/10.1021/ct300117j

Bakan A, Kapralov AA, Bayir H, Hu F, Kagan VE, Bahar I (2015) Inhibition of peroxidase activity of cytochrome c: de novo compound discovery and validation. Mol Pharmacol 88:421–427. https://doi.org/10.1124/mol.115.097816

Barry C et al (2017) Uniform neural tissue models produced on synthetic hydrogels using standard culture techniques. Exp Biol Med 242:1679–1689. https://doi.org/10.1177/1535370217715028
Bartol TM, Bromer C, Kinney J, Chirillo MA, Bourne JN, Harris KM, Sejnowski TJ (2015) Nanoconnectomic upper bound on the variability of synaptic plasticity. elife 4:e10778. https://doi.org/10.7554/eLife.10778
Beckwitt CH, Clark AM, Wheeler S, Taylor DL, Stolz DB, Griffith L, Wells A (2018) Liver 'organ on a chip'. Exp Cell Res 363:15–25. https://doi.org/10.1016/j.yexcr.2017.12.023
Benam KH et al (2015) Engineered in vitro disease models. Annu Rev Pathol 10:195–262. https://doi.org/10.1146/annurev-pathol-012414-040418
Besser RR, Ishahak M, Mayo V, Carbonero D, Claure I, Agarwal A (2018) Engineered microenvironments for maturation of stem cell derived cardiac myocytes. Theranostics 8:124–140. https://doi.org/10.7150/thno.19441
Bhatia SN, Ingber DE (2014) Microfluidic organs-on-chips. Nat Biotechnol 32:760–772. https://doi.org/10.1038/nbt.2989
Bhushan A et al (2013) Towards a three-dimensional microfluidic liver platform for predicting drug efficacy and toxicity in humans. Stem Cell Res Ther 4(Suppl 1):S16. https://doi.org/10.1186/scrt377
Bian YM, He XB, Jing YK, Wang LR, Wang JM, Xie XQ (2019) Computational systems pharmacology analysis of cannabidiol: a combination of chemogenomics-knowledgebase network analysis and integrated in silico modeling and simulation. Acta Pharmacol Sin 40:374–386. https://doi.org/10.1038/s41401-018-0071-1
Black JR, Clark SJ (2016) Age-related macular degeneration: genome-wide association studies to translation. Genet Med 18:283–289. https://doi.org/10.1038/gim.2015.70
Black JW, Duncan WA, Durant CJ, Ganellin CR, Parsons EM (1972) Definition and antagonism of histamine H2-receptors. Nature 236:385–390
Bleakley K, Yamanishi Y (2009) Supervised prediction of drug-target interactions using bipartite local models. Bioinformatics 25:2397–2403. https://doi.org/10.1093/bioinformatics/btp433
Blutt SE et al (2017) Gastrointestinal microphysiological systems. Exp Biol Med 242:1633–1642. https://doi.org/10.1177/1535370217710638
Bredel M, Jacoby E (2004) Chemogenomics: an emerging strategy for rapid target and drug discovery. Nat Rev Genet 5:262–275. https://doi.org/10.1038/nrg1317
Brennan RJ, Nikolskya T, Bureeva S (2009) Network and pathway analysis of compound-protein interactions. Methods Mol Biol (Clifton, NJ) 575:225–247. https://doi.org/10.1007/978-1-60761-274-2_10
Buhule OD et al (2014) Stratified randomization controls better for batch effects in 450K methylation analysis: a cautionary tale. Front Genet 5:354. https://doi.org/10.3389/fgene.2014.00354
Byvatov E, Fechner U, Sadowski J, Schneider G (2003) Comparison of support vector machine and artificial neural network systems for drug/nondrug classification. J Chem Inf Comput Sci 43:1882–1889. https://doi.org/10.1021/ci0341161
Cao DS, Zhang LX, Tan GS, Xiang Z, Zeng WB, Xu QS, Chen AF (2014) Computational prediction of drugtarget interactions using chemical, biological, and network features. Mol Inf 33:669–681. https://doi.org/10.1002/minf.201400009
Chen M, Bisgin H, Tong L, Hong H, Fang H, Borlak J, Tong W (2014) Toward predictive models for drug-induced liver injury in humans: are we there yet? Biomark Med 8:201–213. https://doi.org/10.2217/bmm.13.146
Chen X, Yan CC, Zhang X, Zhang X, Dai F, Yin J, Zhang Y (2016) Drug-target interaction prediction: databases, web servers and computational models. Brief Bioinform 17:696–712. https://doi.org/10.1093/bib/bbv066
Chen R, Liu X, Jin S, Lin J, Liu J (2018) machine learning for drug-target interaction prediction. Molecules 23. https://doi.org/10.3390/molecules23092208
Cirit M, Stokes CL (2018) Maximizing the impact of microphysiological systems with in vitro-in vivo translation. Lab Chip 18:1831–1837. https://doi.org/10.1039/c8lc00039e

Clark AM, Ma B, Taylor DL, Griffith L, Wells A (2016) Liver metastases: microenvironments and ex-vivo models. Exp Biol Med 241:1639–1652. https://doi.org/10.1177/1535370216658144

Clark AM et al (2018) A model of dormant-emergent metastatic breast cancer progression enabling exploration of biomarker signatures. Mol Cell Proteomics 17:619–630. https://doi.org/10.1074/mcp.RA117.000370

Cobanoglu MC, Liu C, Hu F, Oltvai ZN, Bahar I (2013) Predicting drug-target interactions using probabilistic matrix factorization. J Chem Inf Model 53:3399–3409. https://doi.org/10.1021/ci400219z

Cobanoglu MC, Oltvai ZN, Taylor DL, Bahar I (2015) BalestraWeb: efficient online evaluation of drug-target interactions. Bioinformatics (Oxford, England) 31:131–133. https://doi.org/10.1093/bioinformatics/btu599

Conesa A et al (2016) A survey of best practices for RNA-seq data analysis. Genome Biol 17:13. https://doi.org/10.1186/s13059-016-0881-8

Costello JC et al (2014) A community effort to assess and improve drug sensitivity prediction algorithms. Nat Biotechnol 32:1202 EP. https://doi.org/10.1038/nbt.2877. https://www.nature.com/articles/nbt.2877#supplementary-information

De Vivo M, Masetti M, Bottegoni G, Cavalli A (2016) Role of molecular dynamics and related methods in drug discovery. J Med Chem 59:4035–4061. https://doi.org/10.1021/acs.jmedchem.5b01684

Digles D et al (2016) Open PHACTS computational protocols for in silico target validation of cellular phenotypic screens: knowing the knowns. Med Chem Commun 7:1237–1244. https://doi.org/10.1039/c6md00065g

Dudley JT et al (2011) Computational repositioning of the anticonvulsant topiramate for inflammatory bowel disease. Sci Transl Med 3:96ra76

Dutta D, Heo I, Clevers H (2017) Disease modeling in stem cell-derived 3D organoid systems. Trends Mol Med 23:393–410. https://doi.org/10.1016/j.molmed.2017.02.007

Dutta-Moscato J, Solovyev A, Mi Q, Nishikawa T, Soto-Gutierrez A, Fox IJ, Vodovotz Y (2014) A multiscale agent-based in silico model of liver fibrosis progression. Front Bioeng Biotechnol 2:18. https://doi.org/10.3389/fbioe.2014.00018

Dziuba J et al (2014) Modeling effects of SGLT-2 inhibitor dapagliflozin treatment versus standard diabetes therapy on cardiovascular and microvascular outcomes. Diabetes Obes Metab 16:628–635. https://doi.org/10.1111/dom.12261

Edington CD et al (2018) Interconnected microphysiological systems for quantitative biology and pharmacology studies. Sci Rep 8:4530. https://doi.org/10.1038/s41598-018-22749-0

Erdem C et al (2016) Proteomic screening and lasso regression reveal differential signaling in insulin and insulin-like growth factor I (IGF1) pathways. Mol Cell Proteomics 15:3045–3057. https://doi.org/10.1074/mcp.M115.057729

Esch EW, Bahinski A, Huh D (2015) Organs-on-chips at the frontiers of drug discovery. Nat Rev Drug Discov 14:248–260. https://doi.org/10.1038/nrd4539

Ewart L et al (2017) Navigating tissue chips from development to dissemination: a pharmaceutical industry perspective. Exp Biol Med 242:1579–1585. https://doi.org/10.1177/1535370217715441

Ewart L et al (2018) Application of microphysiological systems to enhance safety assessment in drug discovery. Annu Rev Pharmacol Toxicol 58:65–82. https://doi.org/10.1146/annurev-pharmtox-010617-052722

Fang Y, Eglen RM (2017) Three-dimensional cell cultures in drug discovery and development. SLAS Discov Adv Life Sci R & D 22:456–472. https://doi.org/10.1177/1087057117696795

Fatehullah A, Tan SH, Barker N (2016) Organoids as an in vitro model of human development and disease. Nat Cell Biol 18:246–254. https://doi.org/10.1038/ncb3312

Ferdinandy P et al (2018) Definition of hidden drug cardiotoxicity: paradigm change in cardiac safety testing and its clinical implications. Eur Heart J. https://doi.org/10.1093/eurheartj/ehy365

Ferreira LLG, Andricopulo AD (2019) ADMET modeling approaches in drug discovery. Drug Discov Today. https://doi.org/10.1016/j.drudis.2019.03.015

Floris M, Olla S, Schlessinger D, Cucca F (2018) Genetic-driven druggable target identification and validation. Trends Genet 34:558–570. https://doi.org/10.1016/j.tig.2018.04.004

Gadkar K, Kirouac D, Parrott N, Ramanujan S (2016a) Quantitative systems pharmacology: a promising approach for translational pharmacology. Drug Discov Today Technol 21–22:57–65. https://doi.org/10.1016/j.ddtec.2016.11.001

Gadkar K, Lu J, Sahasranaman S, Davis J, Mazer NA, Ramanujan S (2016b) Evaluation of HDL-modulating interventions for cardiovascular risk reduction using a systems pharmacology approach. J Lipid Res 57:46–55. https://doi.org/10.1194/jlr.M057943

Gaulton A et al (2017) The ChEMBL database in 2017. Nucleic Acids Res 45:D945–D954. https://doi.org/10.1093/nar/gkw1074

Ge SX, Son EW, Yao R (2018) iDEP: an integrated web application for differential expression and pathway analysis of RNA-Seq data. BMC Bioinf 19:534. https://doi.org/10.1186/s12859-018-2486-6

Gerdes MJ et al (2013) Highly multiplexed single-cell analysis of formalin-fixed, paraffin-embedded cancer tissue. Proc Natl Acad Sci U S A 110:11982–11987. https://doi.org/10.1073/pnas.1300136110

Gfeller D, Grosdidier A, Wirth M, Daina A, Michielin O, Zoete V (2014) SwissTargetPrediction: a web server for target prediction of bioactive small molecules. Nucleic Acids Res 42:W32–W38. https://doi.org/10.1093/nar/gku293

Goh WWB, Wong L (2018) Dealing with confounders in omics analysis. Trends Biotechnol 36:488–498. https://doi.org/10.1016/j.tibtech.2018.01.013

Goh WWB, Wang W, Wong L (2017) Why batch effects matter in omics data, and how to avoid them. Trends Biotechnol 35:498–507. https://doi.org/10.1016/j.tibtech.2017.02.012

Gough A, Lezon T, Faeder J, Chennubhotla C, Murphy R, Critchley-Thorne R, Taylor DL (2014) High-content analysis with cellular and tissue systems biology: a bridge between cancer cell biology and tissue-based diagnostics. In: Mendelsohn J, Gray JW, Howley PM, Israel MA, Thompson CB (eds) The molecular basis of cancer, 4th edn. Elsevier, New York, pp 369–392. e367. https://doi.org/10.1016/B978-1-4557-4066-6.00025-1

Gough A, Vernetti L, Bergenthal L, Shun TY, Taylor DL (2016) The microphysiology systems database for analyzing and modeling compound interactions with human and animal organ models. Appl In Vitro Toxicol 2:103–117. https://doi.org/10.1089/aivt.2016.0011

Grimes T, Potter SS, Datta S (2019) Integrating gene regulatory pathways into differential network analysis of gene expression data. Sci Rep 9:5479. https://doi.org/10.1038/s41598-019-41918-3

Haasen D, Schopfer U, Antczak C, Guy C, Fuchs F, Selzer P (2017) How phenotypic screening influenced drug discovery: lessons from five years of practice. Assay Drug Dev Technol 15:239–246. https://doi.org/10.1089/adt.2017.796

Hachey SJ, Hughes CCW (2018) Applications of tumor chip technology. Lab Chip 18:2893–2912. https://doi.org/10.1039/c8lc00330k

Hansen J, Iyengar R (2013) Computation as the mechanistic bridge between precision medicine and systems therapeutics. Clin Pharmacol Ther 93:117–128. https://doi.org/10.1038/clpt.2012.199

Hartung T (2009) Toxicology for the twenty-first century. Nature 460:208–212. https://doi.org/10.1038/460208a

Hasin Y, Seldin M, Lusis A (2017) Multi-omics approaches to disease. Genome Biol 18:83. https://doi.org/10.1186/s13059-017-1215-1

He KY, Ge D, He MM (2017) Big data analytics for genomic medicine. Int J Mol Sci 18. https://doi.org/10.3390/ijms18020412

Hill SM et al (2016) Inferring causal molecular networks: empirical assessment through a community-based effort. Nat Methods 13:310–318. https://doi.org/10.1038/nmeth.3773

Hodos RA, Kidd BA, Shameer K, Readhead BP, Dudley JT (2016) In silico methods for drug repurposing and pharmacology. Wiley Interdiscip Rev Syst Biol Med 8:186–210

Hopkins AL (2008) Network pharmacology: the next paradigm in drug discovery. Nat Chem Biol 4:682–690. https://doi.org/10.1038/nchembio.118

Horvath P et al (2016) Screening out irrelevant cell-based models of disease. Nat Rev Drug Discov 15:751–769. https://doi.org/10.1038/nrd.2016.175

Hou T, Qiao X, Xu X (2001) Research and development of 3D molecular structure database of traditional Chinese drugs. Acta Chim Sin 59:1788–1792

Howell BA, Siler SQ, Shoda LK, Yang Y, Woodhead JL, Watkins PB (2014) A mechanistic model of drug-induced liver injury AIDS the interpretation of elevated liver transaminase levels in a phase I clinical trial. CPT Pharmacometrics Syst Pharmacol 3:e98. https://doi.org/10.1038/psp.2013.74

Huh D, Matthews BD, Mammoto A, Montoya-Zavala M, Hsin HY, Ingber DE (2010) Reconstituting organ-level lung functions on a chip. Science 328:1662–1668. https://doi.org/10.1126/science.1188302

Huntley RP, Sawford T, Mutowo-Meullenet P, Shypitsyna A, Bonilla C, Martin MJ, O'Donovan C (2015) The GOA database: gene ontology annotation updates for 2015. Nucleic Acids Res 43:D1057–D1063. https://doi.org/10.1093/nar/gku1113

Ivetac A, McCammon JA (2012) A molecular dynamics ensemble-based approach for the mapping of druggable binding sites. Methods Mol Biol (Clifton, NJ) 819:3–12. https://doi.org/10.1007/978-1-61779-465-0_1

Iyengar R, Zhao S, Chung SW, Mager DE, Gallo JM (2012) Merging systems biology with pharmacodynamics. Sci Transl Med 4:126ps127. https://doi.org/10.1126/scitranslmed.3003563

Jain MR et al (2015) Saroglitazar, a novel PPAR alpha/gamma agonist with predominant PPAR alpha activity, shows lipid-lowering and insulin-sensitizing effects in preclinical models. Pharmacol Res Perspect 3:e00136. https://doi.org/10.1002/prp2.136

Jain A et al (2018) Primary human lung alveolus-on-a-chip model of intravascular thrombosis for assessment of therapeutics. Clin Pharmacol Ther 103:332–340. https://doi.org/10.1002/cpt.742

Johnson CH, Ivanisevic J, Siuzdak G (2016) Metabolomics: beyond biomarkers and towards mechanisms. Nat Rev Mol Cell Biol 17:451–459. https://doi.org/10.1038/nrm.2016.25

Kalli M, Stylianopoulos T (2018) Defining the role of solid stress and matrix stiffness in cancer cell proliferation and metastasis. Front Oncol 8:55–55. https://doi.org/10.3389/fonc.2018.00055

Kanehisa M, Furumichi M, Tanabe M, Sato Y, Morishima K (2017) KEGG: new perspectives on genomes, pathways, diseases and drugs. Nucleic Acids Res 45:D353–D361. https://doi.org/10.1093/nar/gkw1092

Karczewski KJ, Snyder MP (2018) Integrative omics for health and disease. Nat Rev Genet 19:299–310. https://doi.org/10.1038/nrg.2018.4

Kaya C et al (2018) Heterogeneities in axonal structure and transporter distribution lower dopamine reuptake efficiency. eNeuro 5. https://doi.org/10.1523/ENEURO.0298-17.2017

Keating SM et al (2018) Opportunities and challenges in implementation of multiparameter single cell analysis platforms for clinical translation. Clin Transl Sci 11:267–276. https://doi.org/10.1111/cts.12536

Keiser MJ et al (2009) Predicting new molecular targets for known drugs. Nature 462:175–181. https://doi.org/10.1038/nature08506

Koch CM, Chiu SF, Akbarpour M, Bharat A, Ridge KM, Bartom ET, Winter DR (2018) A beginner's guide to analysis of RNA sequencing data. Am J Respir Cell Mol Biol 59:145–157. https://doi.org/10.1165/rcmb.2017-0430TR

Kola I, Landis J (2004) Can the pharmaceutical industry reduce attrition rates? Nat Rev Drug Discov 3:711–715. https://doi.org/10.1038/nrd1470

Kooistra AJ, Kanev GK, van Linden OP, Leurs R, de Esch IJ, de Graaf C (2016) KLIFS: a structural kinase-ligand interaction database. Nucleic Acids Res 44:D365–D371. https://doi.org/10.1093/nar/gkv1082

Kuhn M et al (2010) STITCH 2: an interaction network database for small molecules and proteins. Nucleic Acids Res 38:D552–D556. https://doi.org/10.1093/nar/gkp937

Kurdyukov S, Bullock M (2016) DNA methylation analysis: choosing the right method. Biology 5. https://doi.org/10.3390/biology5010003

LaFramboise T (2009) Single nucleotide polymorphism arrays: a decade of biological, computational and technological advances. Nucleic Acids Res 37:4181–4193. https://doi.org/10.1093/nar/gkp552

Lambert CG, Black LJ (2012) Learning from our GWAS mistakes: from experimental design to scientific method. Biostatistics (Oxford, England) 13:195–203. https://doi.org/10.1093/biostatistics/kxr055

Lands AM, Arnold A, McAuliff JP, Luduena FP, Brown TG Jr (1967) Differentiation of receptor systems activated by sympathomimetic amines. Nature 214:597–598

Larance M, Lamond AI (2015) Multidimensional proteomics for cell biology. Nat Rev Mol Cell Biol 16:269–280. https://doi.org/10.1038/nrm3970

Laurie S et al (2016) From wet-lab to variations: concordance and speed of bioinformatics pipelines for whole genome and whole exome sequencing. Hum Mutat 37:1263–1271. https://doi.org/10.1002/humu.23114

Lave T, Caruso A, Parrott N, Walz A (2016) Translational PK/PD modeling to increase probability of success in drug discovery and early development. Drug Discov Today Technol 21–22:27–34. https://doi.org/10.1016/j.ddtec.2016.11.005

Lee E, Chuang HY, Kim JW, Ideker T, Lee D (2008) Inferring pathway activity toward precise disease classification. PLoS Comput Biol 4:e1000217. https://doi.org/10.1371/journal.pcbi.1000217

Lee-Montiel FT et al (2017) Control of oxygen tension recapitulates zone-specific functions in human liver microphysiology systems. Exp Biol Med 242:1617–1632. https://doi.org/10.1177/1535370217703978

Lenguito G et al (2017) Resealable, optically accessible, PDMS-free fluidic platform for ex vivo interrogation of pancreatic islets. Lab Chip 17:772–781. https://doi.org/10.1039/c6lc01504b

Lexa KW, Carlson HA (2011) Full protein flexibility is essential for proper hot-spot mapping. J Am Chem Soc 133:200–202. https://doi.org/10.1021/ja1079332

Lezon TR, Banavar JR, Cieplak M, Maritan A, Fedoroff NV (2006) Using the principle of entropy maximization to infer genetic interaction networks from gene expression patterns. Proc Natl Acad Sci U S A 103:19033–19038. https://doi.org/10.1073/pnas.0609152103

Li Q, Bender A, Pei J, Lai L (2007) A large descriptor set and a probabilistic kernel-based classifier significantly improve druglikeness classification. J Chem Inf Model 47:1776–1786. https://doi.org/10.1021/ci700107y

Li X, George SM, Vernetti L, Gough AH, Taylor DL (2018) A glass-based, continuously zonated and vascularized human liver acinus microphysiological system (vLAMPS) designed for experimental modeling of diseases and ADME/TOX. Lab Chip 18:2614–2631. https://doi.org/10.1039/c8lc00418h

Lin Z, Jaberi-Douraki M, He C, Jin S, Yang RSH, Fisher JW, Riviere JE (2017) Performance assessment and translation of physiologically based pharmacokinetic models from acslX to Berkeley Madonna, MATLAB, and R Language: oxytetracycline and gold nanoparticles as case examples. Toxicol Sci 158:23–35. https://doi.org/10.1093/toxsci/kfx070

Lind JU et al (2017) Cardiac microphysiological devices with flexible thin-film sensors for higher-throughput drug screening. Lab Chip 17:3692–3703. https://doi.org/10.1039/c7lc00740j

Liu Y, Wu M, Miao C, Zhao P, Li XL (2016) Neighborhood regularized logistic matrix factorization for drug-target interaction prediction. PLoS Comput Biol 12:e1004760. https://doi.org/10.1371/journal.pcbi.1004760

Loving KA, Lin A, Cheng AC (2014) Structure-based druggability assessment of the mammalian structural proteome with inclusion of light protein flexibility. PLoS Comput Biol 10:e1003741. https://doi.org/10.1371/journal.pcbi.1003741

Low LA, Tagle DA (2017) Tissue chips – innovative tools for drug development and disease modeling. Lab Chip 17:3026–3036. https://doi.org/10.1039/c7lc00462a

Macintyre G, Van Loo P, Corcoran NM, Wedge DC, Markowetz F, Hovens CM (2017) How subclonal modeling is changing the metastatic paradigm. Clin Cancer Res 23:630–635. https://doi.org/10.1158/1078-0432.CCR-16-0234

Manatakis DV, Raghu VK, Benos PV (2018) piMGM: incorporating multi-source priors in mixed graphical models for learning disease networks. Bioinformatics 34:i848–i856. https://doi.org/10.1093/bioinformatics/bty591

Marnett LJ (2009) The COXIB experience: a look in the rearview mirror. Annu Rev Pharmacol Toxicol 49:265–290. https://doi.org/10.1146/annurev.pharmtox.011008.145638

Martz CA et al (2014) Systematic identification of signaling pathways with potential to confer anticancer drug resistance. Sci Signal 7:ra121. https://doi.org/10.1126/scisignal.aaa1877

Mateus A, Maatta TA, Savitski MM (2016) Thermal proteome profiling: unbiased assessment of protein state through heat-induced stability changes. Proteome Sci 15:13. https://doi.org/10.1186/s12953-017-0122-4

May S, Evans S, Parry L (2017) Organoids, organs-on-chips and other systems, and microbiota. Emerg Top Life Sci 1:385–400. https://doi.org/10.1042/ETLS20170047

McCauley HA, Wells JM (2017) Pluripotent stem cell-derived organoids: using principles of developmental biology to grow human tissues in a dish. Development 144:958–962. https://doi.org/10.1242/dev.140731

Miedel MT, Gavlock DC, Jia S, Gough A, Taylor DL, Stern AM (2019) Modeling the effect of the metastatic microenvironment on phenotypes conferred by estrogen receptor mutations using a human liver microphysiology system. Sci Rep (final review)

Muller KR, Ratsch G, Sonnenburg S, Mika S, Grimm M, Heinrich N (2005) Classifying 'drug-likeness' with kernel-based learning methods. J Chem Inf Model 45:249–253. https://doi.org/10.1021/ci049737o

Musa A et al (2017) A review of connectivity map and computational approaches in pharmacogenomics. Brief Bioinform 18:903. https://doi.org/10.1093/bib/bbx023

Mustata G et al (2009) Discovery of novel Myc-Max heterodimer disruptors with a three-dimensional pharmacophore model. J Med Chem 52:1247–1250. https://doi.org/10.1021/jm801278g

Newman RH, Zhang J (2014) The design and application of genetically encodable biosensors based on fluorescent proteins. Methods Mol Biol (Clifton, NJ) 1071:1–16. https://doi.org/10.1007/978-1-62703-622-1_1

Nickel J et al (2014) SuperPred: update on drug classification and target prediction. Nucleic Acids Res 42:W26–W31. https://doi.org/10.1093/nar/gku477

Oleaga C et al (2019) Long-term electrical and mechanical function monitoring of a human-on-a-chip system. Adv Funct Mater 29:1805792. https://doi.org/10.1002/adfm.201805792

Pacana T, Sanyal AJ (2015) Recent advances in understanding/management of non-alcoholic steatohepatitis. F1000Prime Rep 7:28. https://doi.org/10.12703/P7-28

Paul DS et al (2016) Increased DNA methylation variability in type 1 diabetes across three immune effector cell types. Nat Commun 7:13555. https://doi.org/10.1038/ncomms13555. https://www.nature.com/articles/ncomms13555#supplementary-information

Pei F et al (2017) Connecting neuronal cell protective pathways and drug combinations in a huntington's disease model through the application of quantitative systems pharmacology. Sci Rep 7:17803. https://doi.org/10.1038/s41598-017-17378-y

Pei F, Li H, Liu B, Bahar I (2019) Quantitative systems pharmacological analysis of drugs of abuse reveals the pleiotropy of their targets and the effector role of mTORC1. Front Pharmacol 10:191. https://doi.org/10.3389/fphar.2019.00191

Pollard TD (2010) A guide to simple and informative binding assays. Mol Biol Cell 21:4061–4067. https://doi.org/10.1091/mbc.E10-08-0683

Prathipati P, Mizuguchi K (2016) Systems biology approaches to a rational drug discovery paradigm. Curr Top Med Chem 16:1009–1025

Prestigiacomo V, Weston A, Messner S, Lampart F, Suter-Dick L (2017) Pro-fibrotic compounds induce stellate cell activation, ECM-remodelling and Nrf2 activation in a human 3D-multicellular model of liver fibrosis. PLoS One 12:e0179995. https://doi.org/10.1371/journal.pone.0179995

Pushpakom S et al (2018) Drug repurposing: progress, challenges and recommendations. Nat Rev Drug Discov. https://doi.org/10.1038/nrd.2018.168

Rao SS, Kondapaneni RV, Narkhede AA (2019) Bioengineered models to study tumor dormancy. J Biol Eng 13:3. https://doi.org/10.1186/s13036-018-0137-0

Ribas J, Pawlikowska J, Rouwkema J (2018) Microphysiological systems: analysis of the current status, challenges and commercial future. Microphysiol Syst 2

Rowland M, Peck C, Tucker G (2011) Physiologically-based pharmacokinetics in drug development and regulatory science. Annu Rev Pharmacol Toxicol 51:45–73. https://doi.org/10.1146/annurev-pharmtox-010510-100540

Sadowski J, Kubinyi H (1998) A scoring scheme for discriminating between drugs and nondrugs. J Med Chem 41:3325–3329. https://doi.org/10.1021/jm9706776

Sakolish C et al (2018) Technology transfer of the microphysiological systems: a case study of the human proximal tubule tissue chip. Sci Rep 8:14882. https://doi.org/10.1038/s41598-018-33099-2

Sanders MP, Barbosa AJ, Zarzycka B, Nicolaes GA, Klomp JP, de Vlieg J, Del Rio A (2012) Comparative analysis of pharmacophore screening tools. J Chem Inf Model 52:1607–1620. https://doi.org/10.1021/ci2005274

Satapathy SK, Sanyal AJ (2015) Epidemiology and natural history of nonalcoholic fatty liver disease. Semin Liver Dis 35:221–235. https://doi.org/10.1055/s-0035-1562943

Satoh T et al (2017) A multi-throughput multi-organ-on-a-chip system on a plate formatted pneumatic pressure-driven medium circulation platform. Lab Chip 18:115–125. https://doi.org/10.1039/c7lc00952f

Schenone M, Dancik V, Wagner BK, Clemons PA (2013) Target identification and mechanism of action in chemical biology and drug discovery. Nat Chem Biol 9:232–240. https://doi.org/10.1038/nchembio.1199

Schoeberl B et al (2009) Therapeutically targeting ErbB3: a key node in ligand-induced activation of the ErbB receptor-PI3K axis. Sci Signal 2:ra31. https://doi.org/10.1126/scisignal.2000352

Schulze K et al (2015) Exome sequencing of hepatocellular carcinomas identifies new mutational signatures and potential therapeutic targets. Nat Genet 47:505–511. https://doi.org/10.1038/ng.3252

Schwartz MP et al (2015) Human pluripotent stem cell-derived neural constructs for predicting neural toxicity. Proc Natl Acad Sci U S A 112:12516–12521. https://doi.org/10.1073/pnas.1516645112

Sebastiani P et al (2010) Genetic signatures of exceptional longevity in humans. Science 2010. https://doi.org/10.1126/science.1190532

Sebastiani P et al (2011) Retraction. Science 333:404. https://doi.org/10.1126/science.333.6041.404-a

Sekar JAP, Tapia J-J, Faeder JR (2017) Automated visualization of rule-based models. PLoS Comput Biol 13:e1005857. https://doi.org/10.1371/journal.pcbi.1005857

Senutovitch N, Vernetti L, Boltz R, DeBiasio R, Gough A, Taylor DL (2015) Fluorescent protein biosensors applied to microphysiological systems. Exp Biol Med 240:795–808. https://doi.org/10.1177/1535370215584934

Seok J et al (2013) Genomic responses in mouse models poorly mimic human inflammatory diseases. Proc Natl Acad Sci U S A 110:3507–3512. https://doi.org/10.1073/pnas.1222878110

Shamir ER, Ewald AJ (2014) Three-dimensional organotypic culture: experimental models of mammalian biology and disease. Nat Rev Mol Cell Biol 15:647–664. https://doi.org/10.1038/nrm3873

Sharma M, Mitnala S, Vishnubhotla RK, Mukherjee R, Reddy DN, Rao PN (2015) The riddle of nonalcoholic fatty liver disease: progression from nonalcoholic fatty liver to nonalcoholic steatohepatitis. J Clin Exp Hepatol 5:147–158. https://doi.org/10.1016/j.jceh.2015.02.002

Shuler ML (2017) Organ-, body- and disease-on-a-chip systems. Lab Chip 17:2345–2346. https://doi.org/10.1039/c7lc90068f

Simian M, Bissell MJ (2017) Organoids: a historical perspective of thinking in three dimensions. J Cell Biol 216:31–40. https://doi.org/10.1083/jcb.201610056

Sin A, Chin KC, Jamil MF, Kostov Y, Rao G, Shuler ML (2004) The design and fabrication of three-chamber microscale cell culture analog devices with integrated dissolved oxygen sensors. Biotechnol Prog 20:338–345. https://doi.org/10.1021/bp034077d

Skardal A, Shupe T, Atala A (2016) Organoid-on-a-chip and body-on-a-chip systems for drug screening and disease modeling. Drug Discov Today 21:1399–1411. https://doi.org/10.1016/j.drudis.2016.07.003

Slenter DN et al (2018) WikiPathways: a multifaceted pathway database bridging metabolomics to other omics research. Nucleic Acids Res 46:D661–D667. https://doi.org/10.1093/nar/gkx1064

Sliz E et al (2018) NAFLD risk alleles in PNPLA3, TM6SF2, GCKR and LYPLAL1 show divergent metabolic effects. Hum Mol Genet 27:2214–2223. https://doi.org/10.1093/hmg/ddy124

Smagris E et al (2015) Pnpla3I148M knockin mice accumulate PNPLA3 on lipid droplets and develop hepatic steatosis. Hepatology 61:108–118. https://doi.org/10.1002/hep.27242

Smietana K, Siatkowski M, Moller M (2016) Trends in clinical success rates. Nat Rev Drug Discov 15:379–380. https://doi.org/10.1038/nrd.2016.85

Sorger P et al (2011) Quantitative and systems pharmacology in the post-genomic era: new approaches to discovering drugs and understanding therapeutic mechanisms. An NIH White Paper by the QSP Workshop Group – October, 2011

Soto-Gutierrez A, Gough A, Vernetti LA, Taylor DL, Monga SP (2017) Pre-clinical and clinical investigations of metabolic zonation in liver diseases: the potential of microphysiology systems. Exp Biol Med 242:1605–1616. https://doi.org/10.1177/1535370217707731

Spagnolo DM et al (2016) Pointwise mutual information quantifies intratumor heterogeneity in tissue sections labeled with multiple fluorescent biomarkers. J Pathol Inf 7:47. https://doi.org/10.4103/2153-3539.194839

Spagnolo DM et al (2017) Platform for quantitative evaluation of spatial intratumoral heterogeneity in multiplexed fluorescence images. Cancer Res 77:e71–e74. https://doi.org/10.1158/0008-5472.CAN-17-0676

Speliotes EK et al (2011) Genome-wide association analysis identifies variants associated with nonalcoholic fatty liver disease that have distinct effects on metabolic traits. PLoS Genet 7:e1001324. https://doi.org/10.1371/journal.pgen.1001324

Stern AM, Schurdak ME, Bahar I, Berg JM, Taylor DL (2016) A perspective on implementing a quantitative systems pharmacology platform for drug discovery and the advancement of personalized medicine. J Biomol Screen 21:521–534. https://doi.org/10.1177/1087057116635818

Subramanian A et al (2005) Gene set enrichment analysis: a knowledge-based approach for interpreting genome-wide expression profiles. Proc Natl Acad Sci U S A 102:15545–15550. https://doi.org/10.1073/pnas.0506580102

Sun M et al (2018) Effects of matrix stiffness on the morphology, adhesion, proliferation and osteogenic differentiation of mesenchymal stem cells. Int J Med Sci 15:257–268. https://doi.org/10.7150/ijms.21620

Sutherland RM, Inch WR, McCredie JA, Kruuv J (1970) A multi-component radiation survival curve using an in vitro tumour model. Int J Radiat Biol Relat Stud Phys Chem Med 18:491–495

Sweeney LM, Shuler ML, Babish JG, Ghanem A (1995) A cell culture analogue of rodent physiology: application to naphthalene toxicology. Toxicol In Vitro 9:307–316

Szabo A, Merks RM (2013) Cellular potts modeling of tumor growth, tumor invasion, and tumor evolution. Front Oncol 3:87. https://doi.org/10.3389/fonc.2013.00087

Szklarczyk D, Santos A, von Mering C, Jensen LJ, Bork P, Kuhn M (2016) STITCH 5: augmenting protein-chemical interaction networks with tissue and affinity data. Nucleic Acids Res 44:D380–D384. https://doi.org/10.1093/nar/gkv1277

Takebe T, Zhang B, Radisic M (2017) Synergistic engineering: organoids meet organs-on-a-chip. Cell Stem Cell 21:297–300. https://doi.org/10.1016/j.stem.2017.08.016

Tan YM, Worley RR, Leonard JA, Fisher JW (2018) Challenges associated with applying physiologically based pharmacokinetic modeling for public health decision-making. Toxicol Sci 162:341–348. https://doi.org/10.1093/toxsci/kfy010

Taylor DL (2012) A new vision of drug discovery and development. Eur Pharm Rev 17:20–24

Teschendorff AE (2018) Avoiding common pitfalls in machine learning omic data science. Nat Mater. https://doi.org/10.1038/s41563-018-0241-z

Torras N, Garcia-Diaz M, Fernandez-Majada V, Martinez E (2018) Mimicking epithelial tissues in three-dimensional cell culture models. Front Bioeng Biotechnol 6:197. https://doi.org/10.3389/fbioe.2018.00197

Trapnell C et al (2014) The dynamics and regulators of cell fate decisions are revealed by pseudotemporal ordering of single cells. Nat Biotechnol 32:381–386. https://doi.org/10.1038/nbt.2859

Trietsch SJ, Israels GD, Joore J, Hankemeier T, Vulto P (2013) Microfluidic titer plate for stratified 3D cell culture. Lab Chip 13:3548–3554. https://doi.org/10.1039/c3lc50210d

Truskey GA et al (2013) Design considerations for an integrated microphysiological muscle tissue for drug and tissue toxicity testing. Stem Cell Res Ther 4(Suppl 1):S10. https://doi.org/10.1186/scrt371

Tsamandouras N, Kostrzewski T, Stokes CL, Griffith LG, Hughes DJ, Cirit M (2017) Quantitative assessment of population variability in hepatic drug metabolism using a perfused three-dimensional human liver microphysiological system. J Pharmacol Exp Ther 360:95–105. https://doi.org/10.1124/jpet.116.237495

Uttam S, Chennubhotla C, Stern AM, Taylor DL (2019) Computational and systems pathology analytics platform applied to hyperplexed fluorescence-labeled patient tissues predicts risk of colorectal cancer recurrence and infers relevant signaling networks. Nat Biotech. in review

Vaidya TR, Ande A, Ait-Oudhia S (2019) Combining multiscale experimental and computational systems pharmacological approaches to overcome resistance to HER2-targeted therapy in breast cancer. J Pharmacol Exp Ther. https://doi.org/10.1124/jpet.118.255752

van den Berg A, Mummery CL, Passier R, van der Meer AD (2019) Personalised organs-on-chips: functional testing for precision medicine. Lab Chip 19:198–205. https://doi.org/10.1039/c8lc00827b

van der Worp HB, Howells DW, Sena ES, Porritt MJ, Rewell S, O'Collins V, Macleod MR (2010) Can animal models of disease reliably inform human studies? PLoS Med 7:e1000245. https://doi.org/10.1371/journal.pmed.1000245

van Laarhoven T, Nabuurs SB, Marchiori E (2011) Gaussian interaction profile kernels for predicting drug-target interaction. Bioinformatics 27:3036–3043. https://doi.org/10.1093/bioinformatics/btr500

Vernetti LA, Senutovitch N, Boltz R, DeBiasio R, Shun TY, Gough A, Taylor DL (2016) A human liver microphysiology platform for investigating physiology, drug safety, and disease models. Exp Biol Med 241:101–114. https://doi.org/10.1177/1535370215592121

Vernetti L et al (2017) Functional coupling of human microphysiology systems: intestine, liver, kidney proximal tubule, blood-brain barrier and skeletal muscle. Sci Rep 7:42296. https://doi.org/10.1038/srep42296

Visscher PM, Brown MA, McCarthy MI, Yang J (2012a) Five years of GWAS discovery. Am J Hum Genet 90:7–24. https://doi.org/10.1016/j.ajhg.2011.11.029

Visscher PM, Goddard ME, Derks EM, Wray NR (2012b) Evidence-based psychiatric genetics, AKA the false dichotomy between common and rare variant hypotheses. Mol Psychiatry 17:474–485. https://doi.org/10.1038/mp.2011.65

Visser SA, de Alwis DP, Kerbusch T, Stone JA, Allerheiligen SR (2014) Implementation of quantitative and systems pharmacology in large pharma. CPT Pharmacometrics Syst Pharmacol 3:e142. https://doi.org/10.1038/psp.2014.40

Volkamer A, Kuhn D, Rippmann F, Rarey M (2012) DoGSiteScorer: a web server for automatic binding site prediction, analysis and druggability assessment. Bioinformatics 28:2074–2075. https://doi.org/10.1093/bioinformatics/bts310

Wagener M, van Geerestein VJ (2000) Potential drugs and nondrugs: prediction and identification of important structural features. J Chem Inf Comput Sci 40:280–292

Walters WP, Murcko MA (2002) Prediction of 'drug-likeness'. Adv Drug Deliv Rev 54:255–271

Wang T, Wu MB, Lin JP, Yang LR (2015) Quantitative structure-activity relationship: promising advances in drug discovery platforms. Expert Opin Drug Discovery 10:1283–1300. https://doi.org/10.1517/17460441.2015.1083006

Watson DE, Hunziker R, Wikswo JP (2017) Fitting tissue chips and microphysiological systems into the grand scheme of medicine, biology, pharmacology, and toxicology. Exp Biol Med 242:1559–1572. https://doi.org/10.1177/1535370217732765

Weedon MN et al (2006) A common haplotype of the glucokinase gene alters fasting glucose and birth weight: association in six studies and population-genetics analyses. Am J Hum Genet 79:991–1001. https://doi.org/10.1086/509517

Wei Y et al (2018) GCDB: a glaucomatous chemogenomics database for in silico drug discovery. Database (Oxford) 2018. epublish. https://doi.org/10.1093/database/bay117

Wen M, Zhang Z, Niu S, Sha H, Yang R, Yun Y, Lu H (2017) Deep-learning-based drug-target interaction prediction. J Proteome Res 16:1401–1409. https://doi.org/10.1021/acs.jproteome.6b00618

Wevers NR et al (2016) High-throughput compound evaluation on 3D networks of neurons and glia in a microfluidic platform. Sci Rep 6:38856. https://doi.org/10.1038/srep38856

Wikswo JP et al (2013a) Engineering challenges for instrumenting and controlling integrated organ-on-chip systems. IEEE Trans Biomed Eng 60:682–690. https://doi.org/10.1109/TBME.2013.2244891

Wikswo JP, Curtis EL, Eagleton ZE, Evans BC, Kole A, Hofmeister LH, Matloff WJ (2013b) Scaling and systems biology for integrating multiple organs-on-a-chip. Lab Chip 13:3496–3511. https://doi.org/10.1039/c3lc50243k

Wishart DS (2016) Emerging applications of metabolomics in drug discovery and precision medicine. Nat Rev Drug Discov 15:473–484. https://doi.org/10.1038/nrd.2016.32

Wishart DS et al (2018) DrugBank 5.0: a major update to the DrugBank database for 2018. Nucleic Acids Res 46:D1074–D1082. https://doi.org/10.1093/nar/gkx1037

Wood AR et al (2014) Defining the role of common variation in the genomic and biological architecture of adult human height. Nat Genet 46:1173–1186. https://doi.org/10.1038/ng.3097

Woodhead JL, Watkins PB, Howell BA, Siler SQ, Shoda LKM (2017) The role of quantitative systems pharmacology modeling in the prediction and explanation of idiosyncratic drug-induced liver injury. Drug Metab Pharmacokinet 32:40–45. https://doi.org/10.1016/j.dmpk.2016.11.008

Workman MJ et al (2017) Engineered human pluripotent-stem-cell-derived intestinal tissues with a functional enteric nervous system. Nat Med 23:49–59. https://doi.org/10.1038/nm.4233

Wray NR, Wijmenga C, Sullivan PF, Yang J, Visscher PM (2018) Common disease is more complex than implied by the core gene omnigenic model. Cell 173:1573–1580. https://doi.org/10.1016/j.cell.2018.05.051

Wu Y, Wang G (2018) Machine learning based toxicity prediction: from chemical structural description to transcriptome analysis. Int J Mol Sci 19. https://doi.org/10.3390/ijms19082358

Wu M et al (2018) Conditional gene knockout and reconstitution in human iPSCs with an inducible Cas9 system. Stem Cell Res 29:6–14. https://doi.org/10.1016/j.scr.2018.03.003

Wu N, Feng Z, He X, Kwon W, Wang J, Xie XQ (2019) Insight of captagon abuse by chemogenomics knowledgebase-guided systems pharmacology target mapping analyses. Sci Rep 9:2268. https://doi.org/10.1038/s41598-018-35449-6

Xia Z, Wu LY, Zhou X, Wong ST (2010) Semi-supervised drug-protein interaction prediction from heterogeneous biological spaces. BMC Syst Biol 4(Suppl 2):S6. https://doi.org/10.1186/1752-0509-4-s2-s6

Xie N, Tang B (2016) The application of human iPSCs in neurological diseases: from bench to bedside. Stem Cells Int 2016:6484713. https://doi.org/10.1155/2016/6484713

Xu X, Ma S, Feng Z, Hu G, Wang L, Xie XQ (2016) Chemogenomics knowledgebase and systems pharmacology for hallucinogen target identification-Salvinorin A as a case study. J Mol Graph Model 70:284–295. https://doi.org/10.1016/j.jmgm.2016.08.001

Yamanishi Y, Araki M, Gutteridge A, Honda W, Kanehisa M (2008) Prediction of drug-target interaction networks from the integration of chemical and genomic spaces. Bioinformatics 24: i232–i240. https://doi.org/10.1093/bioinformatics/btn162

Yamanishi Y, Kotera M, Moriya Y, Sawada R, Kanehisa M, Goto S (2014) DINIES: drug-target interaction network inference engine based on supervised analysis. Nucleic Acids Res 42:W39–W45. https://doi.org/10.1093/nar/gku337

Yang K, Han X (2016) Lipidomics: techniques, applications, and outcomes related to biomedical sciences. Trends Biochem Sci 41:954–969. https://doi.org/10.1016/j.tibs.2016.08.010

Yin A, Yamada A, Stam WB, van Hasselt JGC, van der Graaf PH (2018) Quantitative systems pharmacology analysis of drug combination and scaling to humans: the interaction between noradrenaline and vasopressin in vasoconstriction. Br J Pharmacol 175:3394–3406. https://doi.org/10.1111/bph.14385

Yu J et al (2015) Quantitative systems pharmacology approaches applied to microphysiological systems (MPS): data interpretation and multi-MPS integration. CPT Pharmacometrics Syst Pharmacol 4:585–594. https://doi.org/10.1002/psp4.12010

Zernov VV, Balakin KV, Ivaschenko AA, Savchuk NP, Pletnev IV (2003) Drug discovery using support vector machines. The case studies of drug-likeness, agrochemical-likeness, and enzyme inhibition predictions. J Chem Inf Comput Sci 43:2048–2056. https://doi.org/10.1021/ci0340916

Zhang B, Radisic M (2017) Organ-on-a-chip devices advance to market. Lab Chip 17:2395–2420. https://doi.org/10.1039/c6lc01554a

Zhao S, Iyengar R (2012) Systems pharmacology: network analysis to identify multiscale mechanisms of drug action. Annu Rev Pharmacol Toxicol 52:505–521. https://doi.org/10.1146/annurev-pharmtox-010611-134520

Zhuang X, Lu C (2016) PBPK modeling and simulation in drug research and development. Acta Pharm Sin B 6:430–440. https://doi.org/10.1016/j.apsb.2016.04.004

Zong N, Kim H, Ngo V, Harismendy O (2017) Deep mining heterogeneous networks of biomedical linked data to predict novel drug-target associations. Bioinformatics 33:2337–2344. https://doi.org/10.1093/bioinformatics/btx160

The Future of Clinical Trial Design: The Transition from Hard Endpoints to Value-Based Endpoints

Matthijs D. Kruizinga, Frederik E. Stuurman, Geert J. Groeneveld, and Adam F. Cohen

Contents

1 Introduction 372
2 Historical Overview of Clinical Trial Design 373
3 Increasing Complexity and Obstructions to Clinical Trials 375
4 Event-Based Versus Value-Based Endpoint Trials 381
5 Areas of Opportunity for Value-Based Trials 383
6 The Road Towards New Clinical Endpoints 384
7 Movement Towards Monocenter Studies 384
8 Technology Allows Home-Based Trials 385
9 Electronic Patient-Reported Outcomes (ePROs) Represent Value for Study Participants ... 387
10 Frequent Measurements May Allow for Precision and Personalized Medicine 388
11 Validation of New Endpoints 389
12 Hype, Hope, and the Drawbacks of Continuous Innovation 389
13 Home-Based Sampling of Alternative Matrices 391
14 Integration of Value-Based Endpoints in the Clinical Trial of the Future 391
References 393

Abstract

Clinical trials have been conducted since 500 BC. Currently, the methodological gold standard is the randomized controlled clinical trial, introduced by Austin Bradford Hill. This standard has produced enormous amounts of high-quality

M. D. Kruizinga
Centre for Human Drug Research, Leiden, The Netherlands

Juliana Children's Hospital, HAGA Teaching Hospital, The Hague, The Netherlands

F. E. Stuurman
Centre for Human Drug Research, Leiden, The Netherlands

G. J. Groeneveld · A. F. Cohen (✉)
Centre for Human Drug Research, Leiden, The Netherlands

Leiden University Medical Center, Leiden, The Netherlands
e-mail: ac@chdr.nl

J. E. Barrett et al. (eds.), *Concepts and Principles of Pharmacology*,
Handbook of Experimental Pharmacology 260, https://doi.org/10.1007/164_2019_302

evidence, resulting in evidence-based clinical guidelines for physicians. However, the current trial paradigm needs to evolve because of the ongoing decrease of the incidence of hard endpoints and spiraling trial costs. While new trial designs, such as adaptive clinical trials, may lead to an increase in efficiency and decrease in costs, we propose a shift towards value-based trial design: a paradigm that mirrors value-based thinking in business and health care. Value-based clinical trials will use technology to focus more on symptoms and endpoints that patients care about, will incorporate fewer research centers, and will measure a state or consequence of disease at home or at work. Furthermore, they will measure the subjective experience of subjects in relation to other objective measurements. Ideally, the endpoints are suitable for individual assessment of the effect of an intervention. The value-based clinical trial of the future will have a low burden for participants, allowing for the inclusion of neglected populations such as children and the elderly, will be data-rich due to a high frequency of measurements, and can be conducted with technology that is already available.

Keywords
Clinical trial · Endpoint · Future · Technology · Value-based · Wearable

1 Introduction

Clinical trials are prospective research studies on human participants designed to answer specific questions about biomedical or behavioral interventions. This includes new pharmacological and non-pharmacological interventions (such as drugs, dietary supplements, vaccines, diets, medical devices, behavior change, and surgical procedures) but also known interventions that warrant further study and comparison. The methodological gold standard is the randomized controlled clinical trial, which was introduced in the 1940s. The standard has served the medical profession well and has led to enormous amounts of high-quality evidence guiding our health-care decisions. However, there are more and more reasons to change the current trial paradigm. Spiraling costs for well-designed clinical trials preclude their application for all but the most expensive interventions. Trials focus on hard endpoints, like mortality and major clinical events, which are, thanks to the achievements made in the last 70 years, increasingly rare and therefore require increasing patient numbers to reach sufficient statistical power. Additionally, these events have decreasing consequences for the majority of the patients. The attrition rate, which is the percentage of new compounds that fail in clinical trials, has been 87–90% since the turn of the century, so a large amount of this effort is not benefitting patients directly (Hay et al. 2014; Wong et al. 2019). Moreover, the therapeutic applications are evolving from single compound interventions that are widely usable, to precisely directed treatments and combinations of molecules and other interventions like devices or surgery. Modern technology and trial designs are therefore essential to sustain the evidence base for the increasing number of possible interventions.

2 Historical Overview of Clinical Trial Design

The first clinical trial occurs in the book of Daniel in the Old Testament (Box 1) and was conducted by king Nebuchadnezzar of Babylon (500 BC). The king ordered his servants to only consume meat and wine, a diet he believed to be superior. However, Daniel and several of his followers opted to only eat vegetables and drink water. They eventually gained authorization to do so for 10 days, after which they looked healthier than the servants of the king and were given their choice of food in perpetuity (The Bible 2017). While this investigation does not concern a treatment and the quality of this trial and the resulting evidence are questionable (although probably correct), it is the first documented health-care decision based on evidence gathered via a controlled experiment.

Box 1 The First Documented Clinical Trial (The Bible 2017)
Daniel said to the guard whom the chief official had appointed:

> Please test your servants for ten days: Give us nothing but vegetables to eat and water to drink.
> Then compare our appearance with that of the young men who eat the royal food, and treat your servants in accordance with what you see.

So he agreed to this and tested them for 10 days. At the end of the 10 days, they looked healthier and better nourished than any of the young men who ate the royal food.

The first novel therapy was not investigated in a trial until 1557, although completely by accident. When Ambroise Paré, a French surgeon working on the battlefield, ran out of the standard oil used to cauterize and treat wounds, he resorted to a surprising alternative. He documented the following: "at length my oil lacked and I was constrained to apply in its place a digestive made of yolks of eggs, oil of roses and turpentine." While he feared the worst for his patients, the alternative treatment appeared to be a big improvement compared to cauterization. The patients were "feeling but little pain, their wounds neither swollen nor inflamed. The others to whom I had applied the boiling oil were feverish with much pain and swelling about their wounds." This revelation led him to "never again burn thus so cruelly, the poor wounded by arquebuses" (Donaldson 2015).

However, this trial was uncontrolled and still extremely anecdotal. In 1747, the first controlled clinical trial took place, and it was on the open sea. James Lind was a surgeon on a ship and was greatly dismayed by the toll scurvy had on his fellow seafarers. He decided to conduct a trial with no less than six treatment arms with two patients each. They consisted of the most promising treatments adopted by physicians until then. Patients were administered a quart of cyder, elixir vitriol, vinegar, seawater, an electuary recommended by another surgeon, or, finally, two oranges and one lemon a day. The patients who received fruit were found to be

Fig. 1 Sir Austin Bradford Hill, the statistician and researcher credited with the invention of modern randomization methods

significantly better off compared to their crewmates (Lind 1753), although Lind was hesitant to recommend the preventative treatment for all sailors in service because of the costs of lemons.

During this trial and the various investigations conducted in the centuries afterwards, treatment was allocated at the discretion of the investigator or physician, which meant that both patient and doctor were aware of the novel treatment. This changed in 1943, when the Medical Research Council (MRC) in the United Kingdom carried out the first double-blind controlled trial investigating the efficacy of the antimycotic patulin as treatment for the common cold (MRC Patulin Clinical Trials Committee 1944). During this trial, patulin or placebo was allocated by a nurse using the method of alternation (or rotation) in an isolated room. While alternation was common during that time, this usually meant that the physician and patient were able to discern the used treatment arm. Besides the specifically designated room, two control groups and two treatment groups were used in this trial to reduce the possibility of advance knowledge of the allocations among the physicians who were also responsible for recruitment.

The strict double-blinded treatment allocation was a big step forward in the prevention of observer and selection bias. However, the alternation method used was not truly random and therefore vulnerable to contamination of the study by mentioned bias. Austin Bradford Hill (Fig. 1), a statistician and researcher based in London, introduced a revolutionary random process of treatment allocation. He incorporated randomization in 1946 in the study design of a trial investigating the efficacy of streptomycin in the treatment of tuberculosis. In the resulting paper, the authors state the following: “the details of the (allocation) series were unknown to any of the investigators or to the coordinator and were contained in a set of sealed

envelopes, each bearing on the outside only the name of the hospital and a number" (Raistrick et al. 1948). Interestingly, it was deemed unnecessary to employ blinding in this landmark trial, as there was no possibility of bias when determining the presence of the primary endpoint: death.

During the decades following the patulin and streptomycin trials, randomization and double blinding became international standards, except for a small rebellion of researchers opposed to the burdensome processes which were the result of blinding and randomization in the 1970s (Gehan and Freireich 1974; Doll 2009). Both were enthusiastically propagated by Bradford Hill, who can truly be considered the father of the modern clinical trial, and his colleagues (Doll 2009). Since 1948, more than 200,000 interventional clinical trials have been registered in international trial registries, many of them using randomization (ClinicalTrials.gov 2019). Little has changed in general trial design since then.

3 Increasing Complexity and Obstructions to Clinical Trials

While the general design has been relatively stable, treatments investigated in clinical trials have become increasingly complex. Complexity of clinical trials is at least partly a natural result of the increasing complexity of health care. The first well-designed clinical trials investigated a single antibiotic, but over the subsequent decades, this evolved to drug combinations for tuberculosis or HIV, lifestyle interventions, and, finally, combinations with medical devices such as drug-eluting stents for coronary artery disease and electrodes for deep brain stimulation in Parkinson's disease. The combination of multiple drugs, lifestyle programs, and medical devices leads to the conclusion that we should no longer simply speak of investigational product in clinical trials but rather of a new health-care intervention consisting of several components. Investigations of multifaceted health-care interventions will need complex trials to elucidate the individual value of each component.

A second factor is an increase in the size of clinical trials and the accompanying regulation. During the landmark streptomycin trial, participants were not aware they were included in a medical research study which is unthinkable in our current research practice. The circumstances and underlying rationale regarding the ratification of the declaration of Helsinki and the introduction of Good Clinical Practice guidelines have been well-documented and are beyond the scope of this article. The regulations have undoubtedly increased trial quality, subject safety, and data integrity. However, overinterpretations of the guidelines have irreversibly led to the consequence that it is now very difficult to conduct trials the exact same way in multiple research locations. As a result, costly and burdensome monitoring procedures and bureaucracy have made clinical trials much more difficult and expensive to conduct.

The increasing complexity of the health-care interventions and especially the increase in bureaucracy led to increased costs and loss of efficiency (Fig. 2). Where the scurvy trial and the landmark streptomycin trial would probably cost no more

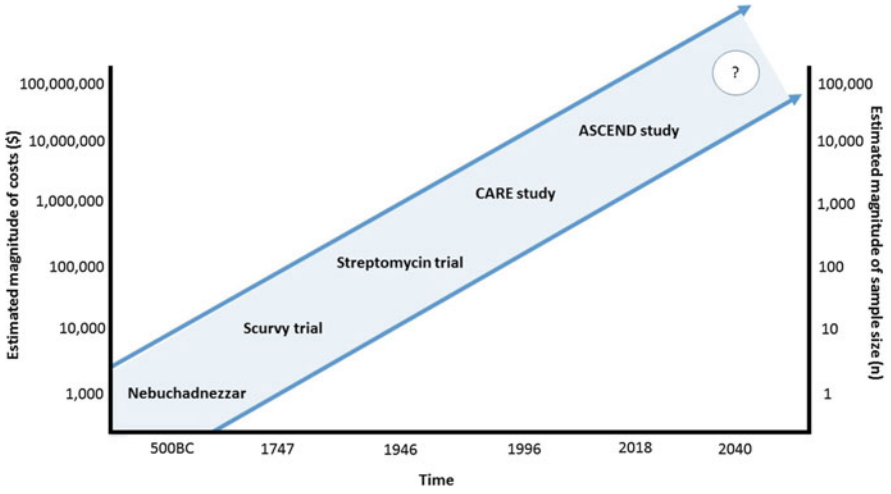


Fig. 2 The exponential rise of funds and sample sizes needed to conduct a clinical trial

than 100,000 dollars to conduct in the present day (not considering the current costs of a seaworthy wooden ship), current average costs of a phase II or phase III trial are approximately 8.6 and 21.4 million dollars, respectively, but can be much higher, and levels of 75 million dollars have been quoted (Martin et al. 2017). The clinical trial of the future will have to counteract the spiraling costs because these contribute to the uncontrolled rise in cost of health care. Several advancements have already been made in this area.

For example, in an attempt to improve efficiency and flexibility, some more recent adaptive trial designs utilize the results gathered in the trial to modify the trial's course according to pre-specified rules (Pallmann et al. 2018; Thorlund et al. 2018). This is a requirement sometimes forgotten by proponents of adaptive clinical trials, which carries the risk of undermining the trial validity and integrity (Chow 2014). In addition, some experiment with umbrella and basket designs in cancer trials (Simon 2017), but most of the designs that use treatment allocation or modification based on earlier findings increase the potential for bias. Although there are some good examples, ultimately these designs may not be sufficient for most situations and are only an evolution of the current paradigms in trial design (Box 2) (Montgomery et al. 2003; Li et al. 2015; Baer and Ivanova 2013).

Box 2 Overview of Innovative Trial Designs Since 1946 (Pallmann et al. 2018; Thorlund et al. 2018; Simon 2017; Montgomery et al. 2003; Li et al. 2015; Baer and Ivanova 2013)

Since the introduction of randomization and blinding in the 1940s, clinical trial design has seen only few new innovations. Designs of consequence were,

(continued)

Box 2 (continued)
among others, the sequential design, crossover design, factorial design, and adaptive design. The designs are meant to improve efficiency, reduce the number of participants, or improve the chances of finding clinically relevant outcomes. Some more specific designs utilized in oncology may improve clinical trial efficiency in the field as well. However, the several innovations come with flaws, such as the introduction of bias and preclusion of use in common situations. Still, some of the newer clinical trial designs, such as the adaptive trial, may significantly improve trial efficiency. Here, we will discuss the advantages and disadvantages of a select number of designs.

Crossover design: Crossover designs allocate each participant to a sequence of interventions. A simple randomized example is an "AB/BA" design in which participants are randomized initially to intervention A or intervention B and then "crossover" to intervention B or intervention A, respectively. The major advantage is that subjects are used as their own, perfectly matched, "controls." This improves the statistical power of the trial and therefore the efficiency. However, there are important conditions to be met regarding the treatment before a crossover design can be utilized. First, the disease should be chronic and stable and the first treatment should not cure the disease. Second, a washout period must be implemented to allow for complete reversibility of drug effects. Besides the washout period, the investigator must be absolutely certain there is no carry-over effect. Also, treatment effects should be quickly observable in order to prevent natural progression of the disease to influence trial results.

Factorial design: As time progressed, more and more treatments became available. Subsequently, investigators also needed a method to research the effects of combinations of treatments. Factorial designs enable efficient simultaneous investigation of two or more interventions by randomizing participants in a treatment group receiving one, multiple, or no intervention. The simplest form is the 2×2 factorial design investigating two interventions (A and B). In this design, participants receive either A alone, B alone, both A and B or neither A nor B (control). A major advantage is the option to investigate both the individual treatment benefits and effects and interactions of receiving multiple interventions together. When there are no treatment interactions, this design greatly increases the efficiency of a trial. However, interactions usually cannot be reliably excluded during trial design. Sample size then becomes a major factor in the trial, for if the trial is to have adequate power to detect an interaction, the sample size increases dramatically. This makes the factorial design still inefficient for many health-care intervention studies and therefore relatively rare.

Sequential design: In the late 1950s, clinical trials started to adopt sequential designs. Here instead of a predefined sample size, a pair of statistical

(continued)

Box 2 (continued)
borders is drawn: one to decide the rejection of the null hypothesis and the other to accept. Trial results are analyzed continuously or during planned interim analyses, and after each analysis an accept, reject, or continue decision is made. This allows for early discontinuation of futile trials but, more importantly, also for a quicker and more efficient road towards acceptance of a new health-care intervention. However, the design was the subject of several critical opinion pieces. It was argued that clinical trials were not about making "accept" or "reject" decisions but rather about estimating the range effects between treatments and therefore it should remain necessary to conduct and complete adequately powered trials. While the sequential design has since then become a rarity, concepts have been integrated in some adaptive trial designs.

Adaptive trial design: Adaptive designs add a review-adapt loop to the regular, linear paradigm. This way, scheduled interim analyses are conducted in order to apply and allow pre-specified changes to the trial's protocol or sample size. All changes have to be applied while retaining the validity and integrity of the trial. Common adaptive designs are the interim sample size reassessment, adaptation of the allocation ratio towards the superior treatment and the dropping of inferior treatments, addition of new treatment arms to save time and resources, and population "enrichment" to narrow scope of the clinical trial. While the potential of adaptive clinical trials looks good on paper, the planning and implementation of the design requires a lot of effort and time, which may trump the potential gains in efficiency. Furthermore, specialized statistical knowledge is necessary to conduct simulations and gain insight in all possible consequences of the possible adaptations. Therefore, adaptive designs are still relatively rare, despite being available for more than 25 years.

Umbrella and basket design: In oncology, new trial designs have been developed to combine the principles of individualized medicine with histology and specific genomic changes of the tumor. Umbrella trials and platform trials maintain the single histology focus of traditional clinical trials but stratify treatment evaluation based on pre-specified genomic biomarkers. In contrast to umbrella designs, basket designs do not focus on disease histology but on patients with a specific genomic change. Patients are assigned a regimen that is expected to be active for tumors containing that alteration. Often this expectation is based on knowledge of the target of the drug and its role in the progression of the disease as well as previous approval of the drug, or a similar drug, for patients with the same genomic alteration in some specified histology.

Parallel groups remain the standard: In the end, the classic parallel group design, introduced around the time of Austin Bradford Hill, is used most often, and this design will most likely remain the cornerstone of health-care

(continued)

Box 2 (continued)
intervention research. New, innovative trial designs may make trials more efficient and smaller and thereby reduce costs. However, all new trial designs introduce limitations or some sort of bias and therefore may never be enough to solve the current problems facing the process of introducing new health-care interventions. In the end, the concepts of blinding and randomization are strong concepts that well-designed clinical trials may never go without.

Besides changes in the general trial design, utilization of the principles of question-based clinical trial design can aid investigators in designing trials that answer the most pressing questions involved with a particular health-care intervention (Cohen et al. 2014). This ideology (Fig. 3) encourages investigators to answer questions covered in six distinct domains during the process of developing a health-care intervention. The domains cover the fundamentals of mechanistic health-care intervention research, important pharmacokinetic aspects of new compounds such as absorption and excretion, but also the pharmacological, physiological, and clinical effects a health-care intervention could induce (Cohen et al. 2014). The domains may state the obvious, but research indicates that in almost half of drug development projects, mechanistic aspects are not investigated thoroughly enough and therefore

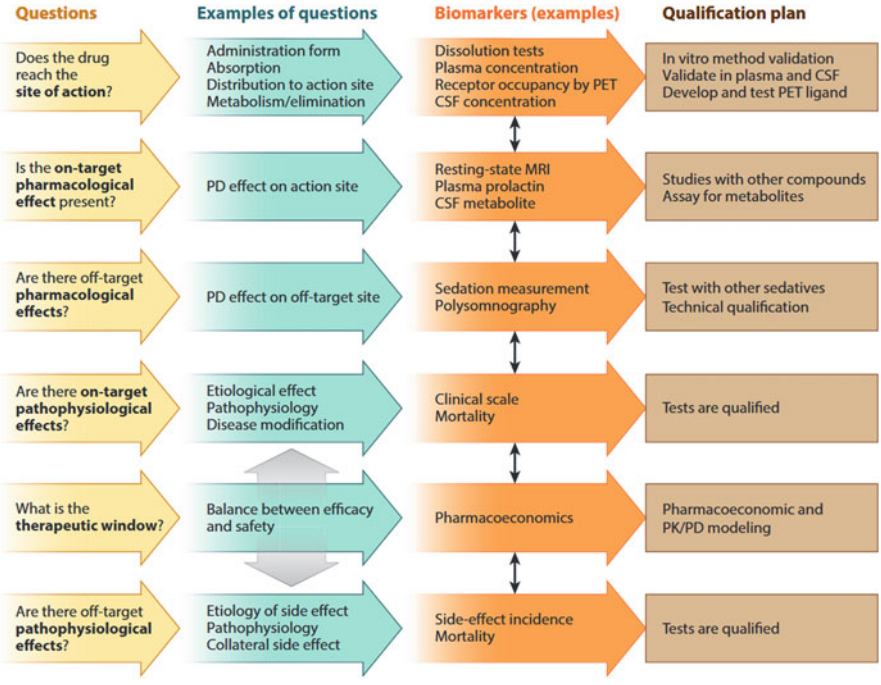


Fig. 3 Domains of question-based drug development (Cohen et al. 2014)

have a considerable chance of failure (Feltner et al. 2011). Following the question-based principles could therefore increase efficiency in developmental processes, while ensuring drug attrition occurs as early in the developmental process as possible and sharply reducing costs.

Furthermore, the health-care intervention development process could be made more efficient by incorporating the principles of health technology assessments (HTA) in early phase clinical trials. HTA is the systematic evaluation of properties, effects, and/or impacts of health technology. The main purpose of conducting an assessment is to inform a policy decision-making regarding reimbursement and decide on incorporation in treatment guidelines (Perry and Thamer 1999). Usually, HTA is conducted at the end of the clinical drug development process, partly while using data that was gathered at an early phase. Utilization of the data as it becomes available at an early stage may identify compounds that are doomed to fail while also allowing for the allocation of more resources towards health-care intervention that shows promise in early assessments (Jönsson 2015).

While incorporation of innovative and question-based designs and early HTA in clinical trials will undoubtedly lead to efficiency gains and cost savings, outcome parameters obviously are an important factor. The role and importance of correct endpoint measurements may have remained a relatively underemphasized area, perhaps because of the importance that has been given to the so-called hard endpoints. Trial outcomes that evaluate the incidence of major health events, such as mortality or vascular or neurological events, are doubtlessly important. However, as their incidence has become lower with better health care, there is a requirement for even more patients in a trial, among which an increasing number that would never have experienced the event, whatever the treatment.

A good example is the recent ASCEND study investigating whether aspirin is of additional value for the primary prevention of cardiovascular disease in patients with type 2 diabetes. For this trial, a sample size of 15,000 subjects followed for 7.5 years was necessary on the basis of an event rate of 1.2–1.3% per year. At the end of the study, a barely significant effect (odds ratio 0.88 [0.80–0.97]) was reported on the composite endpoint "any serious vascular event including TIA" (ASCEND 2018). While statistically significant, the slightly lower chance of a group of complications does not represent great value for the individual patient who is part of a majority that will never experience any of these events whatever the treatment. On the contrary, the composite endpoint "any adverse event" is rarely included in clinical trials (Warren 2019). This imbalance invariably skews results when comparing advantages and disadvantages of new health-care interventions. Furthermore, the sample and effect size of the ASCEND study are in stark contrast with sample sizes in the early clinical trials. In the streptomycin trial of Austin Bradford Hill, only 109 subjects were included in the study, and the death rate was halved from a control mortality of 45% (Raistrick et al. 1948).

To cope with the increased complexity, size, and costs of the current clinical trials, a more radical change in general clinical trial design is needed, particularly in the way we choose trial endpoints: the value-based clinical trial in opposition to the

event-based trial. In this approach we follow the principles of value-based health care (Porter 2010).

4 Event-Based Versus Value-Based Endpoint Trials

The idea of a value-based clinical trial is born out of the introduction of value-based thinking in business and health care. This concept of shared value in business was first introduced in 2006 by Michael E. Porter, professor of economics at Harvard, as a way of developing profitable business strategies that deliver tangible social benefits (Porter and Kramer 2006). The paradigm was further expanded in a 2012 report regarding measurement strategies of shared value (Porter et al. 2012). Porter and his colleagues concluded that the measurement of shared value strategies is as important as the implementation, since this allows quantification of value, provides insight in areas for improvement of the strategy, and allows scaling towards larger implementation of the strategy in the organization.

After proving the benefits of shared value concepts in business, Porter turned his attention to introducing value-based thinking in health care (Porter 2010). Here, value is captured in the formula "health outcomes that matter to patients/costs of delivering the outcomes." The simple formula indicates that one can create value in health care either by improving health outcomes or by lowering costs during the care for a patient. This approach encourages to focus more on collaboration between health-care providers and on sharing data to measure outcomes easily. Furthermore, the approach encourages health-care providers to stop asking themselves how a patient fits in a specific treatment strategy but rather how the provider can help the individual patient sitting before them in the best way. Finally, it encourages health-care providers to use big data and modern technology to assist in decision-making and to evaluate the health-care results. During the last 9 years, integration of value-based health care has accelerated, partly due to the accompanying incorporation of financial incentives in some countries to embrace the concept (Scott et al. 2018).

The value-based concepts in business and health-care focus on the measurement of outcomes that matter to consumers or patients, e.g., social improvement and health benefits and on the analysis of costs. What is odd is that value-based thinking has not reached clinical trial design yet, as trials should generally focus on measuring the effect and hopefully improvement of health-care interventions in patients' lives (Table 1). Concurrently, the incorporation of modern technology and big data in clinical trials is slow (Izmailova et al. 2018). Instead, clinical trials stay focused on the measurement of events and (composite) hard endpoints. While this approach was feasible and preferable during the streptomycin trial of Bradford Hill and the introduction of other radical new therapies, such as aspirin and metformin, medical research has now made our treatments quite advanced. So advanced that, in the developed world, patients generally do not die of pulmonary tuberculosis anymore, cardiovascular mortality following a myocardial infarction is as low as 3%, and glycemic control is quite good with the current arsenal of anti-diabetic medication (Smilowitz et al. 2017; Lipska et al. 2017).

Table 1 Characteristics of value-based concepts in business and health care and our proposal for value-based clinical trial design

	Business	Health care	Clinical research
Priority and focus	Generate economic and social value	Improve outcomes patients care about	Conduct trials with endpoints patients care about
Determination of value	Calculation of the relationship between social improvement and economic value creation of a process	$\frac{\text{Health outcomes that matter to patients}}{\text{costs of delivering the outcomes}}$	$\frac{\text{Value-based + hard endpoint}}{\text{costs of conducting the trial}}$
Main distinct goals	Improve social outcomes and increase economic value of a business	Improve health care by improving important patient outcomes while staying cost-effective	Improve clinical trials by using value-based outcomes combined with hard endpoints, while improving efficiency to save costs
Personalized approach	No	Yes	Yes
Utilize technology	Yes	Yes	Yes
Collect more data	Yes	Yes	Yes

5 Areas of Opportunity for Value-Based Trials

The current paradigm has also led to gaps in knowledge regarding the real-life impact and actual value of our health-care interventions. This could underestimate the negative effects of treatments. For example, we know statins have a generalized negative effect on cellular mitochondria and consequently can lead to muscle complaints (van Diemen et al. 2017). A larger than expected proportion of patients that suffer from muscle aches was reported already in 1991 (Scott and Lintott 1991). Nevertheless, the adverse event was not even mentioned during the landmark CARE trial, which demonstrated significant survival benefit for patients with coronary artery disease (Sacks et al. 1996). The effects of statins on general physical activity, an endpoint which directly measures the value of the therapy and the general well-being of all patients on the treatment, were not investigated in a randomized controlled trial until 2012 (Parker et al. 2013; Noyes and Thompson 2017). They demonstrated a negative treatment effect on mobility, particularly in older patients. In a value-based system, this endpoint would have been included in the very first clinical trials.

Several other research fields have gaps in knowledge regarding the real-life impact of disease and could benefit from a more value-based approach, such as psychiatry. The cornerstones of current trials investigating depression are (validated) questionnaires and depression scales, such as the Hamilton Depression Scale. While used frequently, one could debate the value of scales and questionnaires that, at best, are a subjective measurement of only 40% of depression symptoms (Fried 2017). While it is easy to criticize the objective value of the response to statements such as "my life is pretty full," included in the Zung Self-Rating Depression Scale (Zung 1965), a value-based trial would focus less on asking questions like these every other visit. Instead, it would focus more on objectively measurable parameters that directly impact a patients' well-being. Examples include measuring the amount of social interaction a subject engages in by using their phone, monitoring a patient's radius of action around their home, or monitoring the properties of the sound of their voice. Current technology allows an easily and cheaply implementation (Hashim et al. 2017; Mohr et al. 2017).

The introduction of value-based thinking could also improve the execution of trials, including those in vulnerable populations like children. Pediatric trials are notoriously difficult to conduct because of a difficult ethical approval process, slow recruitment, outcome measures that cannot be directly derived from adult trials, and subjective data obtained from parents which introduces recall and respondent bias that clouds trial results (Joseph et al. 2015). Here, value could be generated by the introduction of technology in the home situation. This could reduce the burden and visits for the patient, generate more useful data, and – in the process – ease the process of obtaining informed consent. Also, value-based trials would focus more on the individual child and objectively measurable behavior, such as the effect a certain asthma intervention has on general physical activity and sleep quality, which we believe to be important indicators of general health (Chaput et al. 2016; Janssen and Leblanc 2010). This would allow to answer questions which have been present in the

Table 2 Guidelines for the value-based endpoints

Evaluate symptoms and consequences of disease patients care about
Focus on home measurement to increase real-life relevance and decrease burden for participants
Employ continuous monitoring to elucidate day-to-day variability in symptoms
Increase the role of ePROs in order to adequately value the subjective experience of patients
Allow for individual assessment of the effect of an intervention

field for decades and enables reevaluation of interventions that have provided conflicting results in conventional clinical trials, such as the efficacy of montelukast in pediatric asthma and recurrent wheezing (Bush 2015; Broughton et al. 2017).

6 The Road Towards New Clinical Endpoints

The value-based endpoints are opposed to the current hard endpoints that evaluate aspects of disease which are reliable to measure but lost their immediate relevance with the improvement of modern medicine. We propose the following guidelines for new value-based endpoints, on which we will further elaborate in the following pages (Table 2).

First, the basis of value-based trial design should be that clinical trials evaluate symptoms and endpoints that patients care about. They should be assessments that directly or indirectly measure an aspect of the disease that, if relieved, improved, or prevented, would be meaningful to patients. Second, value-based trials should trend towards incorporating less research centers in their trial and incorporate value-based endpoints which ideally measure a state or consequence of disease in the natural environment of the participant: home and work. While clinical research units allow standardized measurements by trained personnel, the environment usually does not induce the usual behavior of patients. Third, assuming that none of these endpoints is stable over time, endpoints should allow much more frequent measurements than weekly or monthly assessments. Ideally, the endpoints are suitable for individual assessment of the effect of an intervention. Finally, while the patients are in their natural environment, measurement of the subjective experience with the use of electronic patient-reported outcomes (ePROs) is essential, especially in relation to other objective measurements.

7 Movement Towards Monocenter Studies

The large sample sizes needed for the hard endpoints in event-based trials, as well as regulatory directives requiring local sites in pivotal studies, invariably lead to a multicenter study in multiple countries. As a consequence, the accompanying administrative burden, monitoring procedures, and costs have risen enormously. We expect incorporation of value in trial design will lead to smaller sample sizes and therefore will reduce the need for large multicenter studies, when accompanied with

a relaxation of regulations. Eventually, this could lead to the return of the monocenter study as the standard in clinical trial design. When a trial holds value for the patient, a short travel time will not deter participation. This will benefit trials by standardizing the conduct of complex measurements and procedures. Reductions in costs due to more streamlined logistics and efficient data management procedures are other advantages of the monocenter study. Evidently, when a sample size necessary for the trial exceeds the capacity of the service area of a clinical research unit, patients will have to travel longer for participation in the trial, which will inevitably lead to a multicenter approach. However, this effect could be countered by the embrace of technology to allow for home measurements, either in a hybrid form with few visits at the research unit or in the form of a completely home-based trial.

8 Technology Allows Home-Based Trials

Home measurements include use of technological advances to employ wearable devices and other @*home* devices to measure disease activity outside of the clinical research unit. This has several advantages. First, directly measuring symptoms in daily life allows for more realistic data capture and will aid in the determination of real-world effects of health-care interventions. Furthermore, the noninvasive nature of most devices for home use will reduce the burden for study participants, while automated data streams reduce the administrative burden and data entry errors at the research sites. The uncontrolled environment where study assessments will take place and the reduced threshold to drop out of a study are pitfalls of home-based trials, although there is little data available to substantiate this.

Wearable devices in health care and research are generally hyped and a popular topic for publication (Izmailova et al. 2018; Mohr et al. 2017; Kamiŝalić et al. 2018; Lu et al. 2016; Colbert et al. 2017; Dunn et al. 2018). While a PubMed search in 2010 for the word "wearable" would yield a mere 1,092 results, this number has increased to 8,827 in March 2019 (Pubmed 2019). Nevertheless, incorporation of wearables in clinical trials has lagged behind the hype. A review regarding mobile device-related endpoints in clinical trials identified only 22 interventional studies using mobile data as an endpoint (Perry et al. 2018). Of these, ten trials used device data as their primary endpoint. Still, there is no doubt home-based measurement and wearables hold promise in improving clinical trials. Exciting anecdotal reports exist of wearable technologies assisting in the diagnosis of Lyme disease and the tracking of inflammatory disease (Colbert et al. 2017). A simpler, more obvious, and already widely used example is the measurement of blood pressure at home, which negates the well-known issue of white-coat hypertension (Casiglia et al. 2016).

An interesting development is the integration of smartwatches and other wrist-worn sensors in clinical trials (Lu et al. 2016). They allow for the continuous measurement of parameters such as physical activity, sleep, and heart rate, and some are also capable of measuring blood pressure and environmental factors such as temperature and altitude. The array of possibilities could be expanded in the near future with sensors capable of reliably measuring blood glucose, oxygen saturation,

and a variety of environmental factors such as air pollutants, background sounds, and ambient light levels (Kamišalić et al. 2018). However, the validity of measurements is a factor that should be investigated in individual watch models, particularly in the case of heart rate analysis. There appears to be some discordance (Wang et al. 2017), and devices are generally not medical grade or meant for use by patients. Furthermore, measurement devices may also be used incorrectly by patients as they are no longer assisted by extensively trained trial staff. However, when the frequency of measurements is high enough, occasional results that do not correspond to the gold standard are suboptimal but not disqualifying. With further progression of our technological capabilities, accuracy of wrist-worn sensors will improve as well.

In addition, several devices have been developed for home use that can easily measure vital signs and other outcomes like ECG. Smartphone-based ECG recording systems are already validated for the evaluation of rhythm disorders and outperform conventional Holter monitoring in specific populations in both accuracy and patient satisfaction (Macinnes et al. 2019). Another example is the use of spirometry. Where research participants used to come to a clinical research unit to perform a spirometry test to evaluate treatment, patients with respiratory disease can now easily perform a complete spirometry maneuver by connecting a mobile spirometer to their phone (Ramos Hernández et al. 2018). This could be combined with big data regarding environmental exposures such as pollution, pollen counts, and general weather in the vicinity of a patient in order to create a personalized profile of what exposure leads to reduced pulmonary function in a specific patient, an approach that could join the concepts of value-based thinking with the hard endpoint FEV1. Other examples that could add valuable home monitoring to clinical trials are the Abbott Freestyle Libre for glucose monitoring in diabetes and the biometric shirt system Hexoskin for continuous monitoring of biometric parameters (Pion-Massicotte et al. 2019; Fokkert et al. 2017).

Finally, the device that may show the most promise for incorporation in clinical trials is a device virtually all participants already own: their smartphone. Every smartphone has a range of sensors that could collect data continuously, such as an accelerometer, light sensor, GPS, microphone, and a variety of apps which frequency of use may indicate how a patient is feeling. Some of this data is already being collected by tech companies and could also be used for clinical research, with the caveat that any privacy concerns should be adequately covered. Customized apps could provide participants with simple but valuable test assessments. Furthermore, video-observed administration of medication by study subjects could be superior compared to directly observed administration, which is standard practice in research units (Story et al. 2019). One of the first studies that solely used the smartphone in clinical research is the mPower study (Pratap et al. 2016). In this study, patients with Parkinson's disease downloaded the study app on their iPhone. They were asked to perform a memory, tapping and voice activity on their phone, as well as a performing a walking activity and questionnaire. Patients could complete each activity three times a day but were allowed to skip assessments as they saw fit. Study designs like these could elucidate the day-to-day variability of symptoms and effects of healthcare interventions in several, if not all, diseases while being extremely noninvasive.

A second novelty of the mPower study was the implementation of electronic consent (eConsent). The concept of eConsent ideally allows patients to obtain all relevant information regarding the trial, ask questions to the responsible investigators, and provide adequate written consent. The concept could make the mobile interventional clinical trial a reality, but there are several limitations that should be overcome before widespread implementation. For example, investigators should be wary of a "Facebook effect," where subjects are barely aware of the consequences of their decision and the way their personal and medical data are treated. This is a realistic fear, as research indicates only 0.1–0.2% of consumers access end-user license agreements at all (Bakos et al. 2009). Investigators in the mPower study successfully implemented eConsent after experimenting with several ways to test subject comprehension, such as forcing potential participants to obtain a perfect score on a series of questions regarding the study. Eventually, an assessment in which every incorrect answer led to more education appeared to be the most sensible approach (Wilbanks 2018). While this is a promising approach for the implementation of eConsent, one should also not underestimate the proportion of patients who do not completely comprehend the current paper consent form format, which a recent study found to be a mere 54% (Manta et al. 2016). In fact, the development of eConsent applications and accompanying formats should lead investigators and medical ethical committees to reappraise the necessity of the boilerplate language present in current informed consent forms and to focus on presenting concise and relevant information for potential study participants.

9 Electronic Patient-Reported Outcomes (ePROs) Represent Value for Study Participants

A bigger role of the subjective experience of patients in clinical trials also has potential to add value, as this directly reflects the value patients allocate to their treatment. It also helps investigators to define what it is their patients care about. This is already done by the reintroduction of questionnaires as a patient-relevant endpoint in clinical trials, possibly in the form of an ePRO. Daily questionnaires were largely abandoned in trial design, and rightly so, after studies showed that actual compliance for completing a paper symptom diary is as low as 11%, much lower than participants tend to report voluntarily (Stone et al. 2002). Concurrently, questionnaire assessments using a larger interval between measurements generally suffer from recall bias (Coughlin 1990). However, with the introduction of electronic diaries and ePROs, higher compliance up to 94% can be reached (Stone et al. 2002). While study participants historically received a device from investigators for electronic data capture in clinical trials, the emergence of *bring your own device* (BYOD) in ePRO design has basically made every subject's smartphone a digital diary. This has obvious advantages, such as reduced costs, reduced administrative burden for clinical site, and a reduced burden for study participants (Coons et al. 2015). PROs can clearly demonstrate priorities of patients. A 2012 study comparing rheumatoid arthritis disease activity scores reported by patients and their physicians

showed significant discordance (Khan et al. 2012). The authors demonstrated that priorities for patients were general health outcomes such as fatigue and pain, where physicians relied more on sedimentation rates and joint counts, which are endpoints that regularly feature in rheumatoid arthritis trials. Outcomes gathered via ePROs also have additional value when they are combined with objective data that is gathered concurrently. For example, it is tempting to assume that studies investigating statin therapy would have caught an effect on objectively measured physical activity in those patients complaining of muscle aches via a daily questionnaire. We expect future studies will utilize combined assessments such as these more often.

10 Frequent Measurements May Allow for Precision and Personalized Medicine

There is now ample evidence that not all patients benefit from health-care interventions in the same manner. This could be due to variability in the patient, for instance in pharmacokinetics or in presentation of the disease. The precision medicine approach requires individualized treatment (Rebhan et al. 2018; Bardakjian and Gonzalez-Alegre 2018). In practice, precision medicine has mainly been utilized in oncology research and pharmacogenomics. Other fields could also benefit from a more personalized approach, but, for most, individualized treatment is only possible if there are individual treatment outcomes. However, the probability of a major health event occurring cannot be used for individual treatment decisions in many diseases, considering the fact that the event will not occur for the majority of patients. When such endpoints are the sole basis of the evidence, it is impossible to individualize treatments. Precision medicine therefore requires parametric endpoints that can in some manner be related to treatment success or failure.

The innovations described in this chapter have in common that they allow investigators to increase the frequency of measurements without significantly increasing the burden for participants. Wearables and smartphones allow for continuous monitoring, which basically leads to investigators obtaining a high-resolution overview of the variability and day-to-day activities of patients. This will make it possible to create a profile of interindividual differences between patients, which could be an important factor for the introduction of precision medicine. With such a detailed individual profile, a deviation from the normal individual pattern of several sensor, device, and ePRO measurements may lead to detection of treatment benefits and early detection of health-care problems and events. This may not only improve clinical trials but also health care in general.

11 Validation of New Endpoints

It is tempting to incorporate innovative new endpoints in clinical trials. However, all new value-based endpoints, including but not limited to endpoints generated by devices, should undergo proper validation procedures. First, there is the analytical validation; an analyte or device sensor is compared against the technical gold standards to investigate whether the endpoint actually measures what is claimed. This should always be included in the validation procedure and, considering the state of our current technological capabilities in wearable technology, may reveal discrepancies compared to gold standards. Then, it will be up to the investigator to interpret whether the discrepancies are disqualifying or not. As mentioned, the advantages of continuous monitoring may very well outweigh the disadvantages of small measurement errors.

The next step of validation should focus on demonstrating an association between the device output and disease activity, disease severity, or another area of interest. This is a process of clinical validation comparable to the fit-for-purpose validation in laboratory biomarker research (Cummings et al. 2010). Pilot studies should be conducted where the relationship of the new endpoint with existing measures of disease activity. When pilot studies conducted during this process yield positive results, e.g., a correlation between new endpoint and old endpoint, the new endpoints could be incorporated in existing trials for comparison against clinical gold standards in larger groups.

When the novel endpoint does not correspond perfectly with existing disease activity scores, one may assume this is because of underwhelming performance of the novel endpoint. However, one should be open to the possibility that new endpoints may capture the disease severity better or in another way than the existing scores and scales estimating disease severity. They are often based on subjective clinical observations or questionnaires, subject to interrater variability (Tuijn et al. 2012). Therefore, it is often difficult to speak of a clinical "gold" standard. Investigators should critically review the underlying hypotheses behind the novel endpoint and compare this against the evidence of the reliability and practical usability of the old endpoints in clinical practice. This may lead to the development of new clinical gold standards.

12 Hype, Hope, and the Drawbacks of Continuous Innovation

While innovative new wearables and devices can revolutionize clinical trial design and health care in general, innovation should always focus on questions arising from the field itself. In the last 10 years, many wearables and small devices have been announced that never reached clinical research or practice. While developers may have hoped to become one of these new gold standards, they died a quiet death not long after their unveiling or are being kept in development indefinitely (Table 3) (K'Watch Glucose 2019; Cyrcadia iTBraTM 2019; Automated Device for Asthma Monitoring and Management (ADAMM) 2017; Lee et al. 2014; Bodytrak 2019;

Table 3 Examples of devices that did not (yet) live up to the hype (K'Watch Glucose 2019; Cyrcadia iTBra™ 2019; Automated Device for Asthma Monitoring and Management (ADAMM) 2017; Lee et al. 2014; Bodytrak 2019; Samsung 2019; AmpStrip 2015; Motio 2017)

Name	Device	Health claim	Last appearance
ADAMM	Chest- or back-worn device capable of cough counting and respiration, wheeze and heart rate monitoring	Predicts asthma attack before onset of symptoms	In development since 2015
AmpStrip	3.5-in. long adhesive with single-lead ECG sensor, accelerometer, and a temperature sensor	Measures several fitness-related parameters and provides info to smartphone using a	Cancelled in 2015
Bodytrak	In-ear device with several sensors measuring parameters such as body temperature, heart rate, VO2 and motion	Measures continuously and in real-time using proprietary algorithms and machine learning libraries to provide health and well-being alerts via a simple and configurable user interface, in order to enable early intervention to improve outcomes and reduce injury	In development since early 2017
iTBra™	Temperature sensor incorporated in breast patch/bra	Identifies and categorizes abnormal circadian patterns in otherwise healthy breast tissue for early detection of breast cancer	No news since 2015
K'Watch Glucose	Wrist-worn glucose monitor	Allows for painless and discreet continuous glucose monitoring using microneedle cassette in the watch	In development since early 2017
Motio HW™	Wrist-worn device with variety of sensors	Diagnoses and monitors sleep apnea	No news since 2017
S-Skin	A microneedle patch and separate LED device	Penetrates the skin to deliver effective ingredients and enhance absorption. Measures the hydration, redness, and melanin of the skin to provide customized skincare using LED light	No news since 2016
Zensorium Tinké	Device for fingertip measurements	Measures stress, tracks activity, monitors heart rate, and provides advanced sleep measurement of to deliver a holistic assessment of your health and reduce stress	For sale, health claims not validated

Samsung 2019; AmpStrip 2015; Motio 2017). This may be because some are solutions without a problem or because developers claim exciting unproven health benefits in order to woo potential investors. Other devices have actually been released with exciting health claims but without proper validation, like the Owlet baby monitor. The manufacturer of this smart sock claimed that it was able to alert parents if their infant stops breathing. However, no independent clinical research had been performed at that point, and a subsequent validation study found worrying accuracy (Bonafide et al. 2017, 2018).

A more recent example is the Urgonight, which is a wearable headband promising to improve sleep by using a method based on "EEG neurofeedback" and already a winner of at least one innovation award (UrgoTech 2019). However, a recent randomized controlled clinical trial indicated the employed method holds little promise in improvement of sleep quality (Wislowska et al. 2017). Furthermore, home-based EEG measurements have generally been extremely difficult to carry out reliably, making us doubt the claims of efficacy even further. Devices such as these should make all researchers primed to maintain a critical approach towards the claims made by developers and the data captured by new, exciting devices.

13 Home-Based Sampling of Alternative Matrices

Moving trial assessments towards the subjects' homes may hamper the ability for frequent blood sampling, which is common practice in clinical trials. However, this also leads to opportunities of frequent, long-term sampling of alternative biological samples for biomarker analysis. Since most patients are not proficient in obtaining venous samples of themselves, investigators will have to use an alternative sampling matrix. For example, saliva and dried blood spot assays may be suitable for biomarker research and for pharmacokinetic analysis when validated correctly (Rittau and McLachlan 2012; Bista et al. 2015; Jager et al. 2014).

14 Integration of Value-Based Endpoints in the Clinical Trial of the Future

How would the clinical trial of the future look? An imaginary example from the field of asthma: let's assume a new compound with a novel mechanism of action. Traditionally, such a compound would be given to healthy subjects without asthma to study pharmacokinetics and general tolerability, then in a dose ranging study to subjects with mild asthma and after several years to a larger or more serious group of patients. Patients will be enrolled after an on-site information visit and separate screening visit and will thereafter be studied at 2 weekly intervals but eventually even less, which results in a low resolution of measurements. The measurements will be performed by trained study nurses, and drug administration will only be performed by a trained physician. Visits will include a general questionnaire on (side) effects, pulmonary function tests, multiple ECGs, and blood levels of various

exploratory biomarkers. A large trial will be conducted with a sample size based on hard endpoints such as pulmonary exacerbations. The size of the trial will result in many participating research centers. Since asthma control is quite good in Western Europe, where exacerbations are increasingly rare compared to low- and middle-income countries, the trials will also take place in countries where health-care and research practice is less advanced. Trial monitors will travel often to all centers to try to control this widely divergent group of investigators and cultures. The process will span years and cost close to a billion dollars, if successful.

In a home-based version of these trials, the patients will be informed of the study by their physician or social media. They will download an app on their smartphone with detailed but concise information about entry criteria and study assessments. Patients will have the opportunity to call or chat with a study physician before deciding to give electronic consent to enroll in the study. The subject will then come to a clinical research unit once for a screening and baseline visit and for in-person training regarding study assessments, administration of medication, and a limited amount of tests to be performed by study staff. Subjects will then leave for their homes equipped with a smartwatch that measures movement, heart rate and sleep quality, and ECG. They will also have an app on their smartphone that allows regular data collection and instructs the patients to measure blood pressure, weight, and pulmonary function with a small wireless device. The endpoints of the study will not only focus on pulmonary function or exacerbations but also on measures that indicate value for patients, such as the (daytime or nocturnal) duration a subject spends coughing, the exertional capacity, and the perceived dyspnea a subject has. They will administer the medication at home and will use the camera on their phone to provide evidence of adherence. Because the relation between salivary concentrations and blood concentrations has been studied earlier, patients will also collect saliva at preset moments or generate a dried blood spot for pharmacokinetic, biomarker, and safety analysis, which they will then store in their freezer at home. The data will come in an automatically monitored platform in the cloud and are therefore monitored on the fly rather than at preset moments. Any deviations will result in a notification for the study team and will result in a call and subsequent visit of a help team. Patients communicate through their phone app with the study physician for any advice and can record events using their phone app. Finally, the devices and stored samples will be collected by a courier at the end of the study period.

This represents our vision regarding the value-based clinical trials of the future. This trial is far from science fiction. We believe all the techniques described could be incorporated in trials today. When executed, this trial is extremely data-rich, and participants visit a research site only once or twice. The value-based endpoints will measure outcomes patients, and ultimately investigators, care about. Outcomes will be measured with high frequency, at home, with the use of innovative technologies and incorporation of ePROs. The coordination of the value-based trial will take place at limited amount of research centers and will be designed with clear predetermined questions in mind. The improved efficiency will lower trial costs, while improving health care by reporting the value of health-care interventions, both positive and

negative. This is opposed to the event-based trial, which generally focuses on value for a small proportion of patients, although a combination of value-based endpoints and hard endpoints in clinical trials may emerge as the best of both worlds.

An alternative view of future trials could focus on new, innovative general trial design. In our opinion, there is little room for opportunity there. After all, the various forms of bias present in the first clinical trials, before blinding and randomization were commonplace, are still able to negatively impact the reliability of outcomes. However, the value-based clinical trials of the future may generate high-quality evidence in the form of real-life data. This could lead to value-based, precision endpoints capturing the effect of a health-care intervention so extremely well that the amount of bias removed by the addition of blinding would be considered negligible. After all, even the father of the clinical trial, Austin Bradford Hill, deemed the use of blinding unnecessary in the streptomycin trial for pulmonary tuberculosis. His endpoint was deemed strong enough to overcome the risk of bias. This may eventually be possible with the use of value-based endpoints in the clinical trial of the future as well.

References

AmpStrip cancelled: fitness plans end as Indiegogo refunds begin (2015). https://www.wareable.com/fitness-trackers/ampstrip-cancelled-fitness-plans-end-as-indiegogo-refunds-begin-1816. Accessed 8 Apr 2019

ASCEND (2018) Effects of aspirin for primary prevention in persons with diabetes mellitus. N Engl J Med 379(16):1529–1539

Automated Device for Asthma Monitoring and Management (ADAMM) (2017). http://healthcareoriginals.com/. Accessed 8 Apr 2019

Baer L, Ivanova A (2013) When should the sequential parallel comparison design be used in clinical trials? Clin Investig 3(9):823–833

Bakos Y, Marotta-Wurgler F, Trossen DR (2009) Does anyone read the fine print? Testing a law and economics approach to standard form contracts. Work Pap 09-04, NET Institute, Revis Aug 2009

Bardakjian T, Gonzalez-Alegre P (2018) Towards precision medicine. Handb Clin Neurol 147:93–102

Bista SR, Haywood A, Norris R, Good P, Tapuni A, Lobb M et al (2015) Saliva versus plasma for pharmacokinetic and pharmacodynamic studies of fentanyl in patients with cancer. Clin Ther 37(11):2468–2475. https://doi.org/10.1016/j.clinthera.2015.11.003

Bodytrak – Smart Biometric Sensor Technology (2019). http://www.bodytrak.co/. Accessed 8 Apr 2019

Bonafide CP, Jamison DT, Foglia EE (2017) The emerging market of smartphone-integrated infant physiologic monitors. JAMA 317(4):353–354

Bonafide CP, Localio AR, Ferro DF, Orenstein EW, Jamison DT, Lavanchy C et al (2018) Accuracy of pulse oximetry-based home baby monitors. JAMA 320(7):717–719. https://doi.org/10.1001/jama.2018.9018

Broughton S, Hussein HR, Bossley CJ, Ruiz G, Gupta A, Brathwaite N (2017) A meta-analysis of montelukast for recurrent wheeze in preschool children. Eur J Pediatr 176(7):963–969

Bush A (2015) Montelukast in paediatric asthma: where we are now and what still needs to be done? Paediatr Respir Rev 16(2):97–100. https://doi.org/10.1016/j.prrv.2014.10.007

Casiglia E, Kawecka-Jaszcz K, Dolan E, Malyutina S, Li Y, Ohkubo T et al (2016) The cardiovascular risk of white-coat hypertension. J Am Coll Cardiol 68(19):2033–2043

Chaput J, Gray CE, Poitras VJ, Carson V, Gruber R, Olds T et al (2016) Systematic review of the relationships between sleep and health indicators in school-aged children and youth. Appl Physiol Nutr Metab 41(6 Suppl 3):S266–S282

Chow S (2014) Adaptive clinical trial design. Annu Rev Med 65:405–415

ClinicalTrials.gov (2019) U.S. National Library of Medicine. https://clinicaltrials.gov/. Accessed 12 Mar 2019

Cohen AF, van Gerven JMA, Burggraaf J, Moerland M, Groeneveld GJ (2014) The use of biomarkers in human pharmacology (phase I) studies. Annu Rev Pharmacol Toxicol 55 (1):55–74

Colbert E, Zhou W, Dunn J, Rego S, Schüssler-Fiorenza Rose SM, McLaughlin T et al (2017) Digital health: tracking physiomes and activity using wearable biosensors reveals useful health-related information. PLoS Biol 15(1):e2001402

Coons SJ, Eremenco S, Lundy JJ, O'Donohoe P, O'Gorman H, Malizia W (2015) Capturing patient-reported outcome (PRO) data electronically: the past, present, and promise of ePRO measurement in clinical trials. Patient 8(4):301–309. https://doi.org/10.1007/s40271-014-0090-z

Coughlin SS (1990) Recall bias in epidemiologic studies. J Clin Epidemiol 43(1):87–91

Cummings J, Ward TH, Dive C (2010) Fit-for-purpose biomarker method validation in anticancer drug development. Drug Discov Today 15(19–20):816–825. https://doi.org/10.1016/j.drudis.2010.07.006

Cyrcadia iTBra™ (2019). http://cyrcadiahealth.com/. Accessed 8 Apr 2019

Doll R (2009) Sir Austin Bradford Hill and the progress of medical science. BMJ 305 (6868):1521–1526

Donaldson I (2015) Ambroise Paré's accounts of new methods for treating gunshot wounds and burns. J R Soc Med 108(11):457–461

Dunn J, Runge R, Snyder M (2018) Wearables and the medical revolution. Per Med 15(5):429–448

Feltner DE, Morgan P, Drummond KS, Wegner CD, Arrowsmith J, Street SDA et al (2011) Can the flow of medicines be improved? Fundamental pharmacokinetic and pharmacological principles toward improving phase II survival. Drug Discov Today 17(9–10):419–424. https://doi.org/10.1016/j.drudis.2011.12.020

Fokkert MJ, Van Dijk PR, Edens MA, Abbes S, De Jong D, Slingerland RJ et al (2017) Performance of the freestyle libre flash glucose monitoring system in patients with type 1 and 2 diabetes mellitus. BMJ Open Diabetes Res Care 5(1):1–8

Fried EI (2017) The 52 symptoms of major depression: lack of content overlap among seven common depression scales. J Affect Disord 208:191–197. https://doi.org/10.1016/j.jad.2016.10.019

Gehan EA, Freireich EJ (1974) Non-randomized controls in cancer clinical trials. N Engl J Med 290 (4):198–203

Hashim NW, Wilkes M, Salomon R, Meggs J, France DJ (2017) Evaluation of voice acoustics as predictors of clinical depression scores. J Voice 31(2):256.e1–256.e6. https://doi.org/10.1016/j.jvoice.2016.06.006

Hay M, Thomas DW, Craighead JL, Economides C, Rosenthal J (2014) Clinical development success rates for investigational drugs. Nat Biotechnol 32(1):40–51

Izmailova ES, Wagner JA, Perakslis ED (2018) Wearable devices in clinical trials: hype and hypothesis. Clin Pharmacol Ther 104(1):42–52

Jager NGL, Rosing H, Schellens JHM, Beijnen JH (2014) Procedures and practices for the validation of bioanalytical methods using dried blood spots: a review. Bioanalysis 6 (18):2481–2514

Janssen I, Leblanc AG (2010) Systematic review of the health benefits of physical activity and fitness in school-aged children and youth. Int J Behav Nutr Phys Act 7:40

Jönsson B (2015) Bringing in health technology assessment and cost-effectiveness considerations at an early stage of drug development. Mol Oncol 9(5):1025–1033. https://doi.org/10.1016/j.molonc.2014.10.009. http://www.embase.com/search/results?subaction=viewrecord&from=export&id=L604093588%0A

Joseph PD, Craig JC, Caldwell PHY (2015) Clinical trials in children. Br J Clin Pharmacol 79(3):357–369

K'Watch Glucose (2019). https://www.pkvitality.com/ktrack-glucose/. Accessed 8 Apr 2019

Kamiŝalić A, Fister I, Turkanović M, Karakatiĉ S (2018) Sensors and functionalities of non-invasive wrist-wearable devices: a review. Sensors 18(6):E1714

Khan NA, Spencer HJ, Abda E, Aggarwal A, Alten R, Ancuta C et al (2012) Determinants of discordance in patients' and physicians' rating of rheumatoid arthritis disease activity. Arthritis Care Res 64(2):206–214

Lee J, Bedra M, Finkelstein J (2014) A critical review of consumer health devices for stress self-management. Stud Health Technol Inform 202:221–224

Li T, Yu T, Hawkins BS, Dickersin K (2015) Design, analysis, and reporting of crossover trials for inclusion in a meta-analysis. PLoS One 10(8):1–12

Lind J (1753) A treatise of the scurvy: in three parts, containing an inquiry into the nature, causes and cure, of that disease. Together with a critical and chronological view of what has been published on the subject. A Kincaid & A Donaldson, Edinburgh

Lipska KJ, Yao X, Herrin J, McCoy RG, Ross JS, Steinman MA et al (2017) Trends in drug utilization, glycemic control, and rates of severe hypoglycemia, 2006-2013. Diabetes Care 40(4):468–475

Lu TC, Fu C-M, Ma M, Fang CC, Turner AM (2016) Healthcare applications of smart watches. Appl Clin Inform 7(3):850–869. http://www.schattauer.de/index.php?id=1214&doi=10.4338/ACI-2016-03-R-0042

Macinnes M, Martin N, Fulton H, McLeod KA (2019) Comparison of a smartphone-based ECG recording system with a standard cardiac event monitor in the investigation of palpitations in children. Arch Dis Child 104(1):43–47

Manta CJ, Ortiz J, Moulton BW, Sonnad SS (2016) From the patient perspective, consent forms fall short of providing information to guide decision making. J Patient Saf:1–6

Martin L, Hutchens M, Hawkins C, Radnov A (2017) How much do clinical trials cost? Nat Rev Drug Discov 16(6):381–382

Mohr D, Zhang M, Schueller SM (2017) Personal sensing: understanding mental health using ubiquitous sensors and machine learning. Annu Rev Clin Psychol 13:23–47

Montgomery AA, Peters TJ, Little P (2003) Design, analysis and presentation of factorial randomised controlled trials. BMC Med Res Methodol 3:1–5

Motio HW by Neogia and Kyomed (2017). https://neogia.xyz/press/. Accessed 8 Apr 2019

MRC Patulin Clinical Trials Committee (1944) Clinical trials of patulin in the common cold. Lancet 2:373–375

Noyes AM, Thompson PD (2017) The effects of statins on exercise and physical activity. J Clin Lipidol 11(5):1134–1144

Pallmann P, Bedding AW, Jaki T, Villar SS, Weir CJ, Wheeler GM et al (2018) Adaptive designs in clinical trials: why use them, and how to run and report them. BMC Med 16(1):1–15

Parker BA, Capizzi JA, Grimaldi AS, Clarkson PM, Cole SM, Keadle J et al (2013) Effect of statins on skeletal muscle function. Circulation 127(1):96–103

Perry S, Thamer M (1999) Medical innovation and the critical role of health technology assessment. JAMA 282(19):1869–1872. https://doi.org/10.1001/jama.282.19.1869

Perry B, Herrington W, Goldsack JC et al (2018) Use of mobile devices to measure outcomes in clinical research, 2010-2016: a systematic literature review. Digit Biomark 2(1):11–30. https://doi.org/10.1159/000486347

Pion-Massicotte J, Godbout R, Savard P, Roy JF (2019) Development and validation of an algorithm for the study of sleep using a biometric shirt in young healthy adults. J Sleep Res 28(2):e12667

Porter M (2010) What is value in healthcare. NEJM 363:2477–2481
Porter ME, Kramer MR (2006) Strategy and society: the link between competitive advantage and corporate social responsibility. Harv Bus Rev 84(12):78–92. http://link.springer.com/10.1007/s40134-013-0040-x
Porter M, Hills G, Pfitzer M, Patscheke S, Hawkins E (2012) Measuring shared value: how to unlock value by linking social and business results. FSG.org, Boston, pp 1–24
Pratap A, Doerr M, Suver C, Wilbanks J, Bot BM, Klein A et al (2016) The mPower study, Parkinson disease mobile data collected using ResearchKit. Sci Data 3:160011
Pubmed – NCBI – Search for "Wearable" (2019). https://www.ncbi.nlm.nih.gov/pubmed/?term=wearable. Accessed 13 Mar 2019
Raistrick H, Scadding JG, Tytler WH, Wilson GS, Hart PDA (1948) Streptomycin treatment of pulmonary tuberculosis. Br Med J 30:770–782
Ramos Hernández C, Núñez Fernández M, Pallares Sanmartín A, Mouronte Roibas C, Cerdeira Domínguez L, Botana Rial MI et al (2018) Validation of the portable air-smart spirometer. PLoS One 13(2):e0192789. https://dx.plos.org/10.1371/journal.pone.0192789
Rebhan M, Murphy SA, Röst H, Spang R, Schuppert A, Fröhlich H et al (2018) From hype to reality: data science enabling personalized medicine. BMC Med 16(1):1–15
Rittau AM, McLachlan AJ (2012) Investigating paracetamol pharmacokinetics using venous and capillary blood and saliva sampling. J Pharm Pharmacol 64(5):705–711
Sacks FM, Pfeffer MA, Moye LA et al (1996) Effect of pravastatin on coronary events after myocardial infarction in patients with average cholesterol levels. N Engl J Med 335(14):1001–1009
Samsung S (2019) Skin analyzes and improves your skin. https://www.medgadget.com/2017/01/samsung-s-skin-analyzes-improves-skin.html. Accessed 8 Apr 2019
Scott RS, Lintott CJWM (1991) Simvastatin and side effects. N Z Med J 104(924):493–495
Scott A, Liu M, Yong J (2018) Financial incentives to encourage value-based health care. Med Care Res Rev 75(1):3–32
Simon R (2017) Critical review of umbrella, basket, and platform designs for oncology clinical trials. Clin Pharmacol Ther 102(6):934–941
Smilowitz NR, Mahajan AM, Roe MT, Hellkamp AS, Chiswell K, Gulati M et al (2017) Mortality of myocardial infarction by sex, age, and obstructive coronary artery disease status in the ACTION Registry-GWTG (Acute Coronary Treatment and Intervention Outcomes Network Registry-Get with the Guidelines). Circ Cardiovasc Qual Outcomes 10(12):1–8
Stone AA, Shiffman S, Schwartz JE, Broderick JE, Hufford MR (2002) Information in practice patient non-compliance with paper diaries. BMJ 324(73):1193–1194
Story A, Aldridge RW, Smith CM, Garber E, Hall J, Ferenando G et al (2019) Smartphone-enabled video-observed versus directly observed treatment for tuberculosis: a multicentre, analyst-blinded, randomised, controlled superiority trial. Lancet 393(10177):1216–1224. https://linkinghub.elsevier.com/retrieve/pii/S0140673618329933
The Bible – Book of Daniel (2017) Christian standard Bible, Chapter 1, Verse 1–17
Thorlund K, Haggstrom J, Park JJ, Mills EJ (2018) Key design considerations for adaptive clinical trials: a primer for clinicians. BMJ 360:k698
Tuijn S, Janssens F, Robben P, Van Den Bergh H (2012) Reducing interrater variability and improving health care: a meta-analytical review. J Eval Clin Pract 18(4):887–895
UrgoTech (2019) URGONight brain training for sleep. https://urgonight.com/
van Diemen MPJ, Akram N, Webb A, Groeneveld GJ (2017) Validation of a pharmacological model for mitochondrial dysfunction in healthy subjects using simvastatin: a randomized placebo-controlled proof-of-pharmacology study. Eur J Pharmacol 815:290–297. https://doi.org/10.1016/j.ejphar.2017.09.031
Wang R, Blackburn G, Desai M, Phelan D, Gillinov L, Houghtaling P et al (2017) Accuracy of wrist-worn heart rate monitors. JAMA Cardiol 2(1):104–106
Warren JB (2019) Translating the dose response into risk and benefit. Br J Clin Pharmacol: bcp.13949. https://onlinelibrary.wiley.com/doi/abs/10.1111/bcp.13949
Wilbanks J (2018) Design issues in e-consent. J Law Med Ethics 46(1):110–118

Wislowska M, Gnjezda M-T, Griessenberger H, Hoedlmoser K, Heib DPJ, Schabus M (2017) Better than sham? A double-blind placebo-controlled neurofeedback study in primary insomnia. Brain 140(4):1041–1052

Wong CHIH, Siah KWEI, Lo AW (2019) Estimation of clinical trial success rates and related. Biostatistics 20(2):273–286

Zung WWK (1965) A self-rating depression scale. Arch Gen Psychiatry 12(1):63–70. https://doi.org/10.1001/archpsyc1965017203100650008

Placebos and the Placebo Effect in Drug Trials

Paul Enck and Sibylle Klosterhalfen

Contents

1 Introduction 400
2 Placebo Effects and Placebo Efficacy in Drug Trials 401
2.1 Patient Contributions Towards the Placebo Effect 401
2.2 Doctor/Therapist Contributions Towards the Placebo Effect 406
2.3 The Contribution of Disease Characteristics 408
2.4 The Role of the Trial Designs and Characteristics 411
3 Traditional Concepts to Minimize Placebo Effects 413
3.1 Multiple Centers, Transnational 413
3.2 Placebo Run-Ins and Withdrawals 415
3.3 Enrichment Designs and Adaptive Designs 416
4 The Challenge of Omitting Placebos 418
4.1 Comparative Effectiveness Research (CER) 418
4.2 Waiting List Controls, Treatment as Usual, and Preference Designs 419
4.3 Open-Label ("Real-Life") Observational Studies and Registry and Cohort Studies ... 419
5 Other Challenges for Future Studies 423
5.1 E-Health and m-Health 423
5.2 Placebo Effects with Personalized Medicines 424
6 Summary 425
References 425

Abstract

In this review, we explored different ways of controlling the placebo effects in clinical trials and described various factors that may increase/decrease the placebo effect in randomized placebo-controlled trials. These factors can be subdivided into four groups, and while not all factors are effective in every study and under all clinical conditions, they show on the whole that – even

P. Enck (✉) · S. Klosterhalfen
Department of Internal Medicine VI: Psychosomatic Medicine and Psychotherapy, University Hospital Tübingen, Tübingen, Germany
e-mail: paul.enck@uni-tuebingen.de

J. E. Barrett et al. (eds.), *Concepts and Principles of Pharmacology*,
Handbook of Experimental Pharmacology 260, https://doi.org/10.1007/164_2019_269

under the ideal condition of drug therapy, where blinded placebo provision is much easier and warranted than in, e.g., psychotherapy – many factors need to be controlled to ascertain that the goal of the clinical trials, fair assessment of superiority of the drug over placebo in placebo-controlled trials and fair assessment of non-inferiority of the drug compared to another drug in comparator trials, is reached. Ignorance towards the placebo effect, which was common in the past, is no longer acceptable; instead, it should be the goal of all therapeutic trials to minimize the placebo effect in clinical trials, while utilizing and maximizing it in clinical routine.

Keywords
Clinical trials · Control conditions · Design · Drug effect · Nocebo effect · Placebo effect

1 Introduction

In both traditional and modern pharmacology, placebos are understood as tools, as research vehicles with which the true efficacy and mechanism of action of "real" drugs can be elucidated. Although this tool has been around for over a century (Jutte 2013), it did not earn its rightful place in pharmacology until very recently. We could have known better placebo research commenced as early as in the 1940s, when Henry K. Beecher (1904–1976) reasoned about the size and the mechanisms of the "placebo effect" in the first placebo-controlled clinical trials of his time (Beecher 1955) and Steward Wolf (1914–2005) promoted experimental placebo studies in his milestone paper "pharmacology of the placebo" (sic!) in the prestigious *Pharmacological Reviews* in 1959 (Wolf 1959). In 1980, the editor of *Handbook of Experimental Pharmacology*, Volume 55/I, states in the preface that "the only real psychoactive drug is the placebo: it acts directly on the psyche" (Stille 1980).

In this chapter, we will base our discussion on the content of the *Handbook of Experimental Pharmacology*, Volume No. 225 of 2014 (Benedetti et al. 2014): Although the last 5 years may have brought some new details to light about novel aspects and sophisticated features of the placebo effect and the placebo response, most of what we know today about it is summarized in this reader, as well as in a number of other collections and books published within the last 5 years (Benedetti 2014; Colloca 2018a, b; Enck et al. 2019). For those interested in single studies and papers concerning the term, we refer to the *Journal of Interdisciplinary Placebo Studies* (JIPS) literature database (www.jips.online) which, at present (2019), contains more than 4,000 genuine data papers and reviews on the placebo topic (Enck et al. 2018).

Limitations

Due to space limitations, this chapter will not discuss at length the history of the use of placebos in pharmacology (Kaptchuk 1998; Jutte 2013), nor will we refer in detail to the underlying mechanisms of the placebo effect/response, learning, and

expectations (Schedlowski et al. 2015). We will also refrain from exploring the neurophysiological and biological pathways involved in eliciting responses after placebo provision. Finally, we will abstain from discussing the placebo effects in non-drug therapies: physical therapy (Maddocks et al. 2016), psychotherapy (Enck et al. 2019), instrumental therapies (Burke et al. 2018), acupuncture (Chae et al. 2018), and surgery (Wartolowska et al. 2014) have their own specific and non-specific effects when tested against "sham" interventions, if these are feasible and acceptable. Furthermore, we do not intend to provide an answer to the question as to whether placebo pills (or equivalent medicinal preparations: drops, ointments, injections, infusions, enemas, etc.) are actually required to elicit the placebo response or whether verbal instructions alone are sufficient.

We will instead focus on issues relevant to drug development and drug testing and discuss the ways in which drug efficacy has been dealt with in clinical pharmacology in the past and present, how they may be handled in the future using placebos, and potential alternatives to its utilization. We will continue to bear in mind that the use of placebos has been questioned not only for ethical reasons. Finally, we will explore design alternatives that may be used for both experimental and clinical studies "in the real world" of the future. Albeit this constitutes an exploration of the ways in which placebo effects have affected drug testing, and not the changes of clinical trials in general during the last 25 years (May 2019), and reading through this summary will also identify many features that we discuss in the following chapter.

2 Placebo Effects and Placebo Efficacy in Drug Trials

Below, we will discuss four major factors that determine the placebo effects in drug trials: contributions from patients, contributions from doctors, the role of the disease and its characteristics, and, finally, the role of study designs and trial features. Before doing so, we like to emphasize that, whenever possible, we will distinguish placebo effects from spontaneous variation of symptoms but are aware of the fact that in both drug and placebo arms of randomized, placebo-controlled trials (RCT), the contribution of symptom variation is not always easy and sometimes impossible unless a "no-treatment" arm is included – which is hampered by ethical restraints and psychological barriers discussed later. Our basic understanding is illustrated in Fig. 1.

2.1 Patient Contributions Towards the Placebo Effect

2.1.1 Age and Sex

Among the earliest speculations that placebo effects in RCT are controlled by patient characteristics is the assumption that placebo effects are higher in younger patients

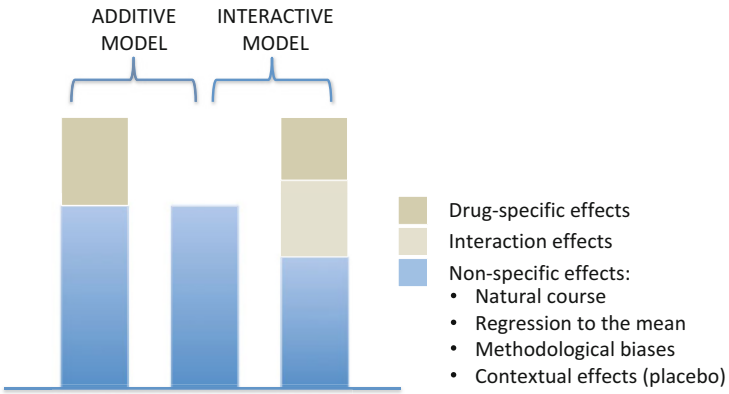


Fig. 1 The "additive model" in pharmacotherapy is the basis for all current drug therapy and its development: it assumes that by double-blinded randomization of patient to either the drug or the placebo arm of the trial, all other factors (natural course, regression to the mean, biases) are kept equally balanced between the two, and the same holds true for the contextual (placebo) effects. While this may be true in a global sense (Kirsch 2000), it has been questioned (Enck et al. 2011a, b), and evidence has been accumulated that at least in some cases, the biology of the placebo effect, e.g., release of endogenous endorphins in case of placebo analgesia, may interfere with the drug effect, e.g., of exogenous pain killers, and may either increase or decrease the placebo effect, leading to false estimation of the efficacy. This is illustrated with the "interactive model" (Enck et al. 2013a)

than in adults and in the elderly and that women show higher placebo responses than men. Both of these assumptions are, however, false.

In a systematic review of 75 meta-analyses on RCT across medicine (neurology, psychiatry, internal medicine) (Weimer et al. 2015a, b), we found only 20 in which an age effect of the placebo response was noted. In 15 analyses the response was said to be higher in younger patients, while in 5 the opposite effect was noted. This poor supportive evidence for an age effect is mainly derived from studies in children and adolescents (Weimer et al. 2013), while there are considerably more studies in adults. However, this effect may be due to specific modalities of pediatric RCT, while age effects among adults have rarely been shown. We have proposed a model (Fig. 2) that allows different developments depending on the type of disease but assumes that the overall response may be a stable pattern (type 2) once patients reach adulthood.

The situation is somewhat different with respect to gender: Again, our systematic review (Weimer et al. 2015a, b) did not support the notion that women show higher placebo effects than men, since only 3 of the 75 meta-analyses noted any gender differences at all. However, evidence from experimental placebo research, either specifically addressing the sex issue or accidentally finding sex differences, left us with a different impression: According to one systematic review using placebo (pain/analgesia) models with verbal placebo instructions (Vambheim and Flaten 2017), the summary of the results of 18 experimental approaches showed evidence of a higher placebo response in males than in females, while the females reacted

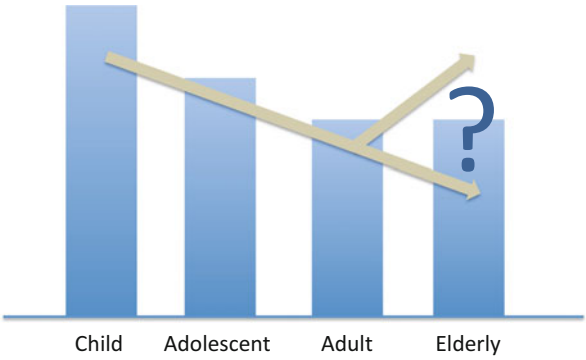


Fig. 2 The placebo effect with increasing age. Some data support that from childhood via adolescence to adulthood, the placebo effect decreases, at least in some clinical conditions (Weimer et al. 2013). We here speculate whether it further decreases at higher age due to decreased expectancy and relevance of the symptoms or whether it increases again with increased experience of effective therapy during the lifespan, based on a conditioning/learning hypothesis. Without further evidence, it is reasonable to assume that it stays stable at the level reached during adulthood

more strongly in conditioning (learning) experiments and with nocebo (symptom worsening) paradigms.

The apparent difference between experimental work on the one hand and clinical studies on the other hand lets us augment the systematic review (Enck and Klosterhalfen 2019) and hypothesize that this difference is due to the fact that, under laboratory conditions, the separation of learning (conditioning) mechanisms and verbal manipulation of expectancies is feasible and enables such differentiation. In clinical trials, however, patients are exposed to settings determining their expectations (e.g., informed consent about potential benefits and adverse effects of the treatment) but also bring their complete disease (or medicine, illness, treatment) history into this setting, thereby mixing learning and expectation mechanisms so that the net (placebo) effect does not permit the identification of the sex-specific relative contribution of each: There may be sex differences, but at the end of the day, these do not surface in RCT. And, as we will see below, this picture becomes even more distorted by the "placebo-by-proxy" effect.

2.1.2 Personality and Genes

Although there has already been much speculation over the years, the proof for a "placebo personality" (patients prone to respond to a placebo provision) remains rather weak (Kaptchuk et al. 2008). The reason is somewhat unexpected: Drug companies, when seeking approval for a novel drug in RCT during its development, do not tend to include psychometric tests to screen for personality profiles and/or specific psychometric characteristics – except in psychiatry and related areas, where psychiatric comorbidity may be part of the disease itself. This is because if the drug response depends at least partly upon psychometric scales, they are at risk of receiving a selective indication: No company would dare to do so. Furthermore, as

has been pointed out (Kaptchuk et al. 2008), to establish the existence of a behavioral response pattern "placebo responder," the response needs to be shown to be stable across different trials and with different drugs for different diseases. Since this has rarely been tested clinically (Whalley et al. 2008) and has produced conflicting results (de la Fuente-Fernandez 2012), it thus disproves the concept. Even within a setting and a RCT, placebo run-in phases were unable to eliminate placebo responses during the trial (see below).

If anything, these data indicate that specific psychological traits are associated with higher (or lower) placebo response rates, coming as they do from experimental studies, albeit involving healthy volunteers. A number of characteristics that have been subject of systematic reviews are identified (Darragh et al. 2014; Horing et al. 2014). While several of these concepts, such as dispositional optimism (Geers et al. 2010), extraversion (Kelley et al. 2009), and an external locus of control (Horing et al. 2015), have even been replicated, it is a matter of some debate as to whether this renders them applicable to patient characteristics. It is, however, important to note that – contrary to common belief – higher placebo responses are associated with an "outward" orientation (externalization), while patients with high inward orientation (high self-efficacy) are less prone to respond to placebos.

In another study with a large group of healthy volunteers ($N = 624$) undergoing placebo analgesia/nocebo hyperalgesia induction by verbal suggestion plus experimental manipulation, a multivariate analysis of somatosensory and psychological variable reveals no predictive power for placebo responses, but personality traits such as neuroticism and extraversion as well as pain modulation by distraction and sex were able to predict nocebo hyperalgesia, the somatosensory response pattern being the strongest predictor of nocebo responses (Christian Büchel, Hamburg, personal communication).

Another reason for this poor outcome of psychometric screening for placebo responders may be of a methodological nature: The significant associations of single traits (or subscales of traits) reported may have been purely random and may be due to a beta error. Many tests were carried out, but only a few subscales – precisely those reported – yielded significance. A multivariate approach with a reasonably large sample may overcome such a bias.

While it is still too early for a final conclusion, the search for genes or polymorphisms of genes predicting the placebo response makes the same mistake: For whole-genome analyses (GWAS, genome-wide association studies), the samples are usually too small to allow adjustment for multiple comparisons, and candidate gene approaches replicate only what has been found for other psychological or behavioral traits and conditions. Summary reviews (Colagiuri et al. 2015; Hall et al. 2018) propose a "placebome" list an assembly of 28 genes/SNPs in 42 studies to date (Wang et al. 2017) to which more and more studies will be added in the future, albeit probably without improving the concept to any great extent.

2.1.3 Proxies

One of the most neglected research areas in placebo research, with far-reaching effects on placebo responses, is the influence of the social environment of the patient, relatives, and friends and, specifically, of other patients with the same or with other

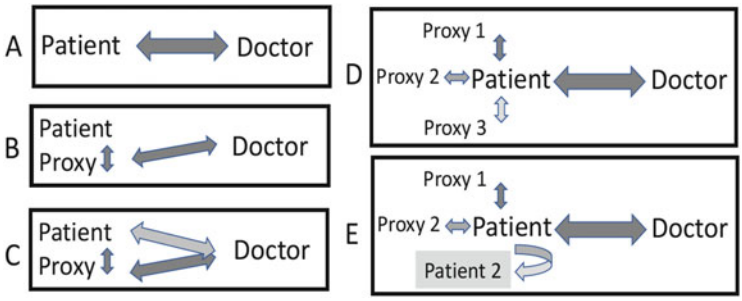


Fig. 3 The "placebo-by-proxy" concept (Grelotti and Kaptchuk 2011) illustrated in a systematic way, in which placebo responses are generated by increasing complexity of the network of interactions (different shades of gray reflect different communication intensities). (**a**) An idealized medical situation in contemporary medicine, where the (adult) patient individually communicates with the doctor and reports all relevant events in his/her medical history and environment (including family). Our understanding of the placebo effect is typically based on this constellation. (**b**) The concept illustrated reflects where the patient may experience limitations to direct communication with the doctor, due to verbal (infants, animals), social (migrant), or cognitive (intellectual disability) limitations. Proxy reports, based on either observation of the patient behavior or on (limited or special) communication strategies, are required. (**c**) Instead of exclusively communicating with the proxy, doctors may rely on additional information directly from the patient. This may generate conflicting information, e.g., higher placebo effects from proxy reports than from measures. (**d**) The social environment of a patient usually contains more than one proxy, with varying proximities to the patient, from family (parents, children, siblings) to relatives and friends/peers/colleagues. Proximity determines how much they may be involved in the medical history and its reporting and how much the doctor may be aware of this social network and its influence on disease reporting, management, and efficacy. (**e**) It is conceivable that one or more of the members of a social network may also have an impact as patient, though the timing and direction of effect may not be readily apparent but via an iterative process become contributors to the treatment effect of the index patient, either via social observation or explicit or implicit learning and vice versa

diseases. This concept has been called "placebo by proxy" (Grelotti and Kaptchuk 2011) and is observed when patients are unable to directly express their symptoms and symptom changes to their physician, instead of requiring a "proxy" to do so: these are predominantly children and mentally disabled.

We summarized this concept and developed a kind of systematic classification (Fig. 3) for future studies. For the time being, however, we are left with a few empirical examples demonstrating its clinical relevance. Our concept may also account for the differences observed between patient and proxy ratings of symptom improvement, e.g., in attention deficit hyperactivity disorder (ADHD) (Waschbusch et al. 2009).

One novel variant of the placebo-by-proxy concept will be discussed later, but the increasing use of social media and Internet fora by patients recruited for drug studies causes concern among trialists, e.g., with respect to the quality blinding in RCT (Lipset 2014); its impact on testing drug and placebo efficacy still needs to be determined.

2.2 Doctor/Therapist Contributions Towards the Placebo Effect

2.2.1 Age, Sex, and Ethnicity

Until the late 1980s, most RCTs in common diseases, where patient recruitment is not difficult to achieve, were monocentric, and thus the question as to what extent the placebo effects are attributable to the individual treating physician could not be answered: center effects on RCT outcome were simply not discernible and therefore of no consequence. This may be the real reason why everybody seems to believe that placebo responders may exist (Benedetti and Frisaldi 2014): Placebo producers, doctors who were able to push both placebo and drug effects up higher, were appreciated rather than dismissed and were rarely challenged.

However, even in individual centers, patients are often treated by different physicians. In a post hoc analysis of a RCT for treatment of irritable bowel syndrome (IBS) (Enck et al. 2005a, b), we had access to individualized patient and doctor data and ascertained that the female physician generated a better outcome than her two male colleagues in both the diet and drug and in the placebo arm of the study. Similar data resulted from an acupuncture trial in which female acupuncture therapists were more frequently believed to have administered true (as opposed to sham) acupuncture in a controlled acupuncture trial than their male counterparts (White et al. 2003): Female physicians appear to elicit more trust than their male counterparts.

While this phenomenon is well established in social psychology for most types of day-to-day communication among individuals, it had not yet been tested extensively in patient-doctor interaction. In experimental placebo research, female experimenters were observed to produce higher placebo analgesia rates (i.e., reports of less pain) in male volunteers, but not in females (Aslaksen et al. 2007); in experimental nausea and placebo/nocebo responses, we often noted sex-by-sex interactions of the outcome of respective studies (Enck and Klosterhalfen 2019), as already discussed above (Sect. 2.1.1) with regard to sex differences on placebo response in general. What is more, in a series of such nausea studies with German and Chinese volunteers (Klosterhalfen et al. 2005a, b, 2006), one female Chinese experimenter was unable to secure reliable nausea reports from her colleagues because they were (male) students, while she was a university teacher in China.

While systematic exploration of such factors in RCT is wanting, basic experiments pave the way: doctor ethnicity and gender affect patient judgment to a high degree, resulting in variable trust scores and the willingness to believe and comply (Shah and Ogden 2006). A simulation study comprised 300 UK patients who rated each one of 8 pictures of doctors of varying sex (male, female), age (young, old), and race (Asian, Caucasian) with respect to their anticipated personal manners, technical skill explanatory skills, advice, emotional aspects, and referral behavior, all of which are liable to contribute to placebo responses. They described remarkable differences – particularly between gender and race – with respect to patient expectation, but not necessarily with respect to true consulting behavior.

2.2.2 Training, Education, and Communication Skills

Little is known about how the medical training of doctors contributes to the response of patients during a RCT in general, let alone the specific response to placebo in a trial such as this. One ingenious experiment at Harvard Medical School (Jensen et al. 2014) sheds some indirect light on this question: Doctors in training were recruited for a brain imaging study in which they were told that the purpose of the study is to ascertain how the treatment of a patient effectively influences the doctor's brain. The rest is camouflage: an instructed patient-actor performed "pain relief" and "pain worsening" following button-pressing of the doctor inside the scanner that mimics successful or failed pain blockade via a sham device on the patient's arm; the doctor was able to observe the reflection of the facial response in a mirror. The perceived pain relief was directly linked to activation of the reward areas (e.g., area postrema) in the doctor's brain, and these, in turn, were the very same areas (the so-called pain matrix) that are known to mirror placebo analgesia in patients, as shown in different experiments (Legrain et al. 2011). On the basis of such data, training medical students in doctor-patient interaction may have a profound influence in future RCT.

Our final illustration of the relevance of expectations is derived from a study conducted in a Canadian hospital, in which more than 300 patients were asked to rate the empathy of the treating doctor (on a standardized scale) when attending a clinic for a common cold (Rakel et al. 2009). Patients who perceived their physician as empathic were shown to have significantly less severe symptoms, and, as even laboratory tests confirmed, the duration of their cold was almost a day shorter.

Finally, training during preparation of a RCT to better standardize patients' communication and information is required. Failure of drug trials (Kobak et al. 2007) is often associated with poor preparatory training of doctors prior to the study, inadequate conductance (e.g., recruitment and treatment in the hands of the same person), and biased evaluation of treatment outcome, particularly when based on subjective measures by the treating physician. However, standardized patient assessment by independent raters, video-recorded control, and combined doctor- and patient-reported outcomes are still not universal standards. This may well explain reported discrepancies in placebo response rates in RCTs (in depression treatment) between PRO and doctor ratings (Rief et al. 2009a).

2.2.3 Setting

In a quasi-experimental study (incidental rebuilding of a medical outpatient center), architecture, design, and service, as well as seasonal variations, were shown to have the ability to substantially improve the response to medical treatment (Rehn and Schuster 2017). This serves to illustrate that many more factors than the immediate circumstances on drug/placebo provision contribute to the overall treatment effect, of which only a few, such as those related to the empathy communication skills of the therapists, may be standardized through training, as discussed above. Such "incidental effects" (Grünbaum 1986) are difficult to control and require careful inspection of the site, time, and the staff conducting the RCT.

While we acknowledge that many of these influential factors may be averaged out by selecting many centers, each of which recruits only a small fraction of patients for the RCT, it cannot be ruled out that the known nationality-dependent effects of different placebo response rates in different regions of the world (EU versus USA) in multinational trials may be due to such effects. The time spent at the first consultation in primary care can vary substantially from country to country, even in Western countries (Irving et al. 2017).

2.3 The Contribution of Disease Characteristics

2.3.1 Disease Severity

Disease severity is one of the major driving forces for placebo effects in RCTs: Our analysis of the placebo responses in psychiatric (Weimer et al. 2015a, b) and other RCTs (Weimer et al. 2015a, b) across different clinical conditions showed that a lower disease severity in almost all meta-analyses was associated with higher placebo responses. Lower symptom severity is therefore one of the very few factors that predict the placebo effect in both adults and children.

To lend support to this statement as a more general rule, we deem it necessary to define "severity" on the basis of disease symptoms rather than of disease biomarkers: At the time of its first clinical diagnosis, a disease that initially has only very few symptoms, for example, juvenile diabetes, may not respond to placebo application at all but could well respond to metabolic interventions. On the other hand, diseases with a high symptomatic load, such as asthma, may respond stronger to placebo interventions following a drug intervention affecting forced expiratory volume (Wechsler et al. 2011). This underlines the importance of subjective measures in addition to biomarkers for many, if not for all, conditions.

2.3.2 Disease Duration

In agreement with a low disease severity at the disease onset, a short medical history and disease duration have been found to be associated with higher placebo responses in RCTs (Weimer et al. 2015a, b). Although this may well be the driving factor for higher placebo responses at younger age (see above), it has never actually been evaluated. In a meta-analysis of pediatric depression trials, the same holds true for children: the lower the severity, the higher the placebo response (Bridge et al. 2009).

At this point, drug development may run into a paradox, a kind of "trap," when selecting only mildly affected patients for treatment of a putatively chronic condition as early as possible and before the disease exacerbates: such secondary prevention trials may be at overestimation of their efficacy. The same phenomenon may occur if – for economic or marketing reasons – patients recruited for RCTs during drug development do not represent the majority of patients in clinical routine and the drug proves to be disappointing after marketing approval, as was the case with the class of serotoninergic antidepressants (Kirsch 2016).

2.3.3 Previous Treatments

It had been noted already some time ago (Rickels et al. 1966) that a preceding treatment of a disease may co-determine the success or failure of a subsequent treatment and that this applies not only to drug effects but also to placebo effects and in both directions: Treatment success may predict higher responses, and treatment failure may result in lower responses in the next trial (Colloca and Benedetti 2006). This is highly compatible with the concept that the placebo response is a conditioned response – albeit conditioning and expectancy cannot be as easily differentiated in medical treatment as in the laboratory for experimental placebo studies (Enck et al. 2008).

At the same time, for many clinical conditions, a shorter disease history and presumably a lower disease severity at least in case of chronic diseases are known to be associated with higher placebo response rates in RCT (see above). The immediate consequence of this is an apparent paradox: Testing novel drugs in patients with less severe symptoms may generate better drug responses but drives the placebo response higher, and so larger sample sizes are then required to yield significance in RCT.

This is of great relevance for drug testing in many respects: To begin with, novel drugs tested successfully in RCT often disappoint in the real world once they compete with drugs on the market and are tested on patients who have experienced both success and failure. At the same time, *The Emperor's New Clothes* (Kirsch 2014) fuels expectations and makes counter-evidence and contradictory experience likely. Finally, this calls for inclusion of the patients´ medical history, especially their previous drug treatments, into the screening procedure for RCT, including their participation in earlier drug testing RCTs. However, since the latter is at conflict with both ethical and legal rules, we will discuss a potential solution at a later stage in this paper.

2.3.4 Adverse Event Rate of Drugs

Each and every placebo-controlled trial assumes perfect blinding of the study medication, which is feasible provided that the company producing the drug is also responsible for the production of the (undistinguishable) placebo. Under these circumstances, adverse events (AE), particularly when based on subjective patient reports, occur to a similar degree in the two treatment arms (Mahr et al. 2017), and their overall incidence may not differ as long as the symptoms occurring are of a general nature (Rheker et al. 2018). This situation may change when the drug induces highly specific AE and side effects, but provided all potential AE are listed in the patient information and consent form, even those AE have a good chance of being listed under placebo conditions. A meta-analysis comparing AE reporting between different antidepressants (tricyclics, serotonin reuptake inhibitors) confirmed that in both the drug and placebo arm of the trials, it is the assessment procedure rather than the drug itself that determines the amount of AE reported and that the difference between the two drugs is also reflected in AE rates of the placebo arms in these studies (Rief et al. 2009b). This indicates that the information *provided* about AE rather than the actual occurrence of AE is the driving force of such "nocebo effects" (Enck et al. 2013a, b) in RCTs. As already illustrated in

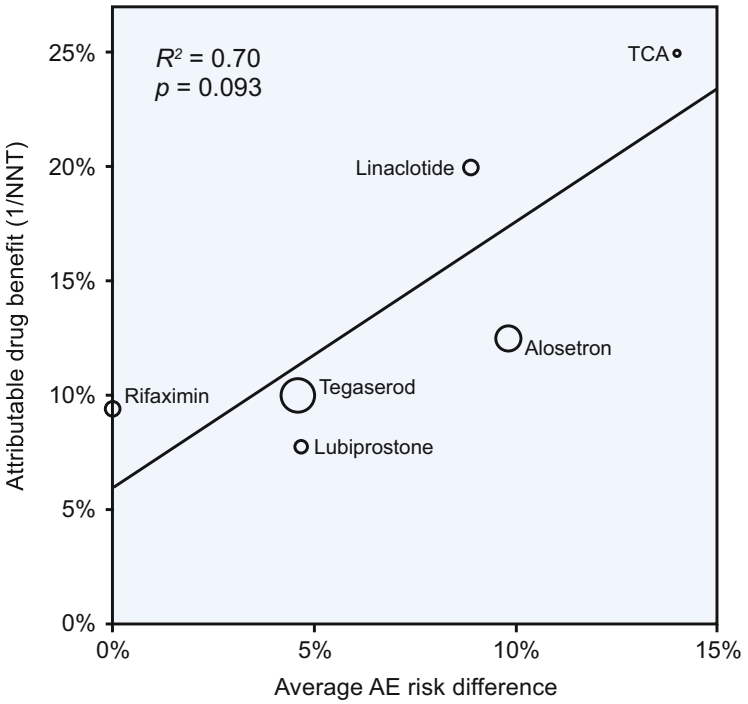


Fig. 4 Implicit unblinding of a study, based on reported adverse events (AE), as it becomes visible during a meta-analysis (Shah et al. 2014): Significant correlation between patient-reported efficacy and average adverse event risk difference in different drug therapies of irritable bowel syndrome (IBS). The size of each data point correlates with population size as a relative measure of variance for the assessment of adverse events. The positive and significant correlation indicates that with higher AE risk (difference between AE is the drug and the placebo arm of the RCT), the relative drug benefit (1/NNT) increases. (Reproduced with permission from Wiley and Sons, License No. 4627120830140)

meta-analyses, nocebo response rates determine the rates of discontinuation, e.g., in Parkinson's disease (Leal Rato et al. 2019).

Such explicit unblinding – which may also occur when patients participating in a RCT communicate via social media (see below, Sect. 5.2) – is not to be confused with another phenomenon labeled "implicit unblinding" (Shah et al. 2014), which was identified in a meta-analysis of RCT in treatment of irritable bowel syndrome (IBS): the authors analyzed 6 different IBS treatment approaches in 30 RCT, either with (serotoninergic) prokinetics (alosetron, linaclotide, tegaserod), with tricyclic antidepressants, sodium chloride channel blockers (lubiprostone), and with a locally acting antibiotic (rifaximin). In summary, they ascertained that the higher the reporting incidence of AE in the drug arm of these trials compared to placebo, the higher the reported drug-placebo difference (and, thus, the drug benefit). Figure 4 shows the correlation between the two, indicating implicit unblinding even if the individual patient in any of the trials is not aware of it.

2.4 The Role of the Trial Designs and Characteristics

2.4.1 Crossover Versus Parallel Group

In the early phases of drug development (the second half of the twentieth century), crossover trials were quite common – patients received either placebo or drug in a double-blinded manner in a first phase and then, following a washout period, the alternate application for the same duration. The advantage is each patient served as his/her "own" control, thus reducing data variance and enabling smaller numbers of patients to achieve statistical significance of drug over placebo. It also complied with an ethical stipulation that all patients should receive effective treatment, either immediately or after the placebo period.

The disadvantage is an effective treatment with the drug during the first phase affected the second treatment period – while the drug may have been washed out, conditioning effects are not, unless they are extinct (Suchman and Ader 1992). They increase the placebo effects over the "placebo-first" group. Similarly, if the drug was ineffective in the first phase, this had consequences for the placebo treatment that ensued. In consequence, treatment effects (drug and placebo, respectively) could be merged only if they were equipotent, irrespective of their order of provision; otherwise, only the first phase of treatment could be used for efficacy evaluation, and the advantage of the crossover would be lost, since it then would become a parallel-group designed study.

Nowadays, crossover designs are usually used to meet ethical requirements and to improve patient recruitment in cases in which leaving a patient with a placebo treatment only might be seen as unacceptable – for reasons medical, ethical, or psychological.

To wash out a conditioning effect in a crossover design study, it may be advisable to provide a placebo during the washout phase, as well as a kind of randomized withdrawal strategy, so that individual patients are switched from drug to placebo or vice versa, double-blinded, and with different timing (Moore et al. 2015) (Fig. 5). To the best of our knowledge, this has never been tested for feasibility in a crossover design study; it would still be necessary to control for equal starting out conditions in the two arms.

2.4.2 Trial Duration

In older textbooks of clinical pharmacology, you often will find the statement that placebo effects diminish all the more, the longer the trial lasts. In many RCTs in the last decade of the twentieth century, a conventional trial length lasted between 4 and 8 weeks, e.g., for acute conditions where a life-long intervention was not deemed necessary. In conditions prone to produce high placebo responses, e.g., in functional bowel disorders of IBS type (Elsenbruch and Enck 2015), it was proposed that trials lasting 8 weeks, which was common in the 1990s, should be extended to 12 weeks. The prediction was that this would result in lower placebo response rates in RCT (Spiller 1999) (Fig. 6).

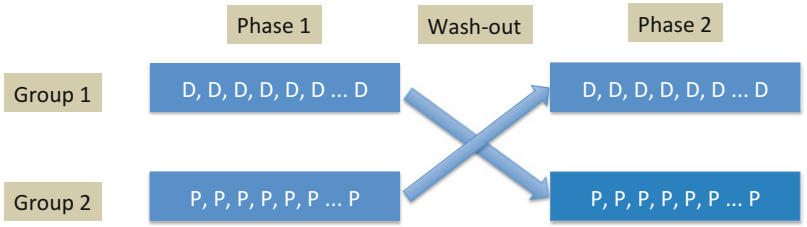


Fig. 5 A learning theory view on crossover trials with washout between drug and placebo phases. The unconditioned stimulus (US) is the drug (D), and the conditioning stimulus (CS) is the pill (shape, size, color, etc. = placebo). Groups 1 and 2 differ in the sequence they receive D and P; the washout phase may be of arbitrary length. In Group 1, the patient is conditioned in Phase 1 – by pairing the US and the CS – to respond to the CS alone in Phase 2: the washout period may eliminate the drug level, but it does not extinct the conditioned response unless a placebo (CS) is provided without the US. Thus, extinction will only gradually occur in Phase 2. In Group 2, the patient is initially exposed to the CS alone, a learning strategy which is called "latent inhibition" (Klosterhalfen et al. 2005a, b) that will minimize the conditioned response in Phase 2 – the washout phase does not serve any purpose. Therefore, while the two D phases may be comparable, the two P phases are not, and the calculation of the global drug efficacy based on intraindividual D-P differences is not an adequate estimation of it

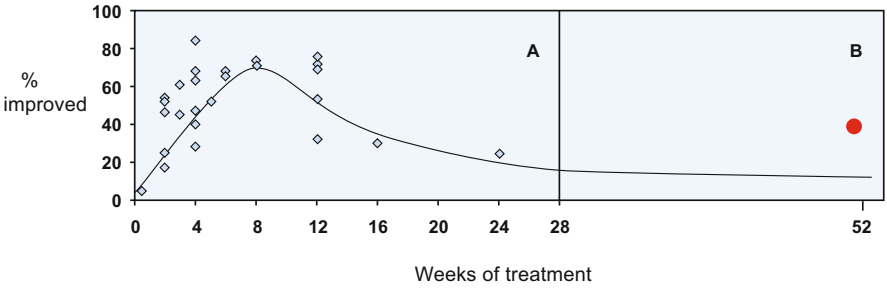


Fig. 6 Based on 26 randomized, placebo-controlled trials in irritable bowel syndrome (IBS) available at that time, it was argued (Spiller 1999) that with trial length over 24 week, placebo effects should reach a low level of 20% after half a year and decrease further afterwards (**a**). The extension of the plot beyond 28 weeks (**b**) was added to include the first 1-year study (Chey et al. 2004; red dot) in IBS with a stable 40% placebo effects for 1 year. (**a**) (Reproduced with permission from Excerpta Medica Inc., License No. 4627130163661)

However, when the first 12-week and longer trials were implemented, it became evident that placebo response could remain as high as 40% throughout such studies, and examples are available of 12-month trials with stable and high placebo response rates across the entire period (Khan et al. 2008; Quessy and Rowbotham 2008), not only in IBS (Chey et al. 2004).

The reason for this paradoxical prediction is that with a 4-week treatment trial, it may be possible to limit doctor-patient contacts to two – one at the beginning of the study and one at the end of trial – while for 12 weeks one would plan intermediate visits for motivation, compliance control, drug provision, and others.

By manipulating patients' expectancies; this increased number of contacts would reinforce the placebo effect (Enck et al. 2005a, b). And as with long-term, e.g., 1-year trials (Chey et al. 2004), the recording of symptoms and treatment effects would generally take the form of daily diary entries, phone calls from study nurses, and other measures. All these measures are liable to enhance the placebo effect, which is known to be driven by the extent of doctor-patient communication (Ford and Moayyedi 2010), irrespective of the nature of the disease (Jairath et al. 2016).

2.4.3 Randomization Ratio

If expectancy is another major driving force of the placebo effect in RCTs in addition to conditioning, the likelihood of receiving drug rather than placebo should affect the size of the placebo effect. A 50:50 randomization scheme is most common, but there are many reasons to deviate from it and to increase the percentage of patients in the drug arm of the study: for motivational reasons ("better than chance"), for ethical reasons (less patients without treatment), or to test different drug dosage in equally powered study arms against one placebo group.

It was first noted in a systematic review of migraine trials that increasing the chances of receiving active treatment causes the extent of the placebo effect to increase in a near-linear fashion (Diener et al. 1999). Subsequent analyses have confirmed this effect of "unbalanced randomization" in depression, in schizophrenia, and in other neurological and psychiatric conditions (Papakostas and Fava 2009; Mallinckrodt et al. 2010; Agid et al. 2013) (for a review see Weimer et al. 2015a, b). Interestingly, and for still unknown reasons, we were unable to confirm this phenomenon in the analysis of more than 100 RCTs in IBS (Elsenbruch and Enck 2015) (Fig. 7). Furthermore, unbalanced randomization does not influence the placebo effect in pediatric depression (Rutherford et al. 2011).

An easily conceivable endpoint of such study planning is reached when all (100%) patients receive active treatment and no placebo whatsoever is provided, such as with "comparative effectiveness research" (CER) or head-to-head trials, where a novel drug therapy is tested against another drug that is already available. We will discuss this further below, but a meta-analysis comparing efficacy of various antidepressants in placebo-controlled trials to CER studies using the same types of drugs revealed a 15% *higher* drug response in CER trials than in placebo-controlled trials (where the placebo response is on average 40%, Rutherford et al. 2009), which is solely attributable to the 100% anticipation of receiving active treatment.

3 Traditional Concepts to Minimize Placebo Effects

3.1 Multiple Centers, Transnational

Until the late 1990s, single-center studies were quite common in clinical drug testing, and there may still be a number of good reasons to maintain this tradition, e.g., in mechanistic studies in Phase II development or in the case of highly specific intervention strategies and modes, but definitely not for drug intervention. Center effects are thus avoided; they may be responsible for many drug failures once a drug

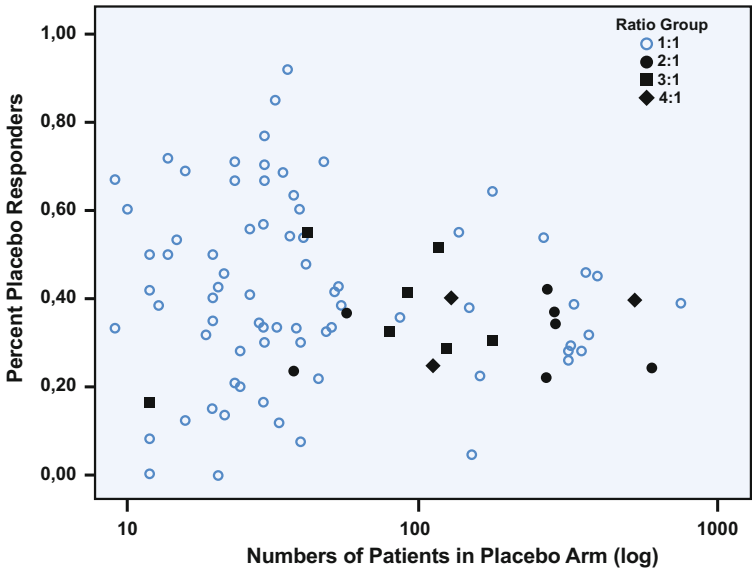


Fig. 7 The placebo effect in irritable bowel syndrome (IBS) trials as a function of the number of patients recruited: With higher patient numbers, the variance of the placebo effect between studies decreases and approximates 40% which has been found to be the global placebo response rates across all IBS trials (Ford and Moayyedi 2010). At the same time, the unbalanced randomization ratio (more patient assigned to drugs than to placebo) seems to not affect the placebo rates in IBS, while this has been found to be the case in depression, schizophrenia, and other conditions (Weimer and Enck 2014). (Reproduced with permission from Springer-Nature, License Number 4627150201527)

comes onto the market or even reaches the Phase III trials (Kobak 2010). Today's standards, multicenter trials with equal sample sizes and block randomization, may prevent overestimation of the drug-placebo difference to a considerable extent, albeit not completely: A higher number of study sites and a lower number of patients per study site were associated with higher placebo response (but not the drug response) in a meta-analysis of pediatric antidepressant trial (Bridge et al. 2009).

Extending multicenter trials across different countries is yet another option but one that bears many risks: Treatment of specific clinical conditions may be organized in very specific ways; hence, RCT results conducted in different countries could not be easily compared – and certainly not planned without taking the specifics of country, healthcare system, reimbursement policy, and alike into account. Cultural differences in the understanding (of the rationale for placebo-controlled trials) or interpretation (is it good to respond to placebo?) do exist (Ventriglio et al. 2018). Therefore, comparing placebo response rates – in meta-analyses – across different continents (Europe versus the USA) is crucial and has shown that overall European studies may generate higher placebo response, at least in some conditions (Stein et al. 2006). However, since neither Europe nor the USA is homogeneous cultural entity, subtle differences may sneak into individual RCT, depending on the range and location of recruitment centers.

3.2 Placebo Run-Ins and Withdrawals

The idea of ideally identifying putative placebo responders at an early point in a trial, or even before during recruitment of patients, is as logical as it is false: it assumes that being placebo responsive is a stable intraindividual characteristic that does not bear much empirical evidence (Kaptchuk et al. 2008). However, it bears another inherent risk: Being responsive to placebo does not rule out also being responsive to the drug, so by excluding responsive patients from the study, we may be preselecting the population, thereby introducing a selection bias; placebo responsiveness may thus indicate a subgroup of patients (such as those with lower symptom severity) and excluding these may put the requested indication for the drug at risk. A recent meta-analysis (Munkholm et al. 2019) indicates that placebo run-ins may also lead to false interpretation of drug efficacy: Participants treated with an antidepressant before recruitment and subsequently randomized to the study drug might experience withdrawal symptoms during the placebo run-in that are subsequently alleviated by the study drug.

We have already argued (above) that stable personality traits for placebo responsiveness do not exist. On the empirical-experimental side, the same person may be seen to respond to placebo provision in one trial, but not to another one in a different setting (Whalley et al. 2008). Furthermore, an effective treatment at one point in time may co-determine the response to any treatment (drug or placebo) on another occasion, both with experimental approaches (Colloca et al. 2010) and under clinical conditions (de la Fuente-Fernandez 2012), but is not warranted. The time frames for such "carry-over effects" have not been established, nor is it known how often a successful experience is required for it, how long it may last, and whether this also applies to negative (noneffective) treatment experiences ("nocebo"). The literature on Pavlovian learning is full of rules that may apply but that have yet to be explored.

In Fig. 5 (above), we have applied one such rule (extinction) to the test of carry-over effects in crossover trials. This resembles some similarities with randomized withdrawal studies, where patients are taken off the drug (or placebo) at the end of the trial in a blinded, randomized fashion (Fig. 8) to avoid conditioned rebound (nocebo) effects, i.e., effects that are due not to the pharmacologic withdrawal but to psychological effects such as disappointment at having reached the end of the study. This effect can be profound, as is shown in another example of the IBS literature (Chey et al. 2004): Having reaching the end of a 1-year study with persistent 40% placebo response and a stable 15% benefit above placebo in the respective arms, both drug and placebo recipients showed a dramatic recurrence of symptoms – a randomized withdrawal in the drug arm would presumably have shown a slower symptom worsening than in the placebo arm, which could be evaluated in terms of drug efficacy.

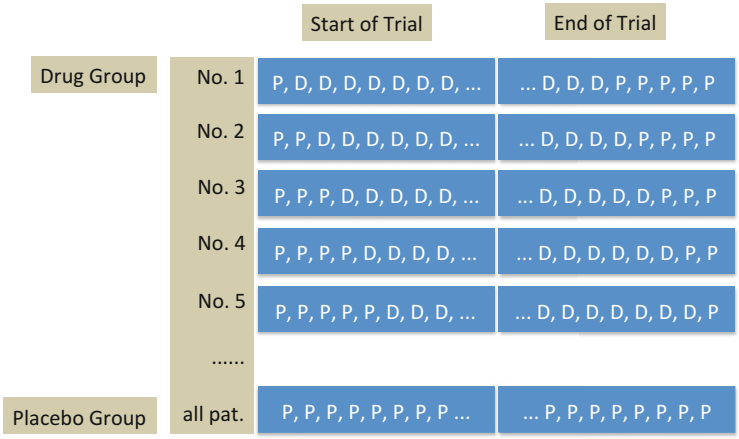


Fig. 8 The concept of randomized run-in and withdrawal in a clinical trial. To cover the true start and end of a trial, patients can be randomized to double-blinded run-in as well as withdrawal, where the true start of drug provision is hidden among days with placebo application instead. (Reproduced with permission from Springer-Nature, License Number 4627150201527)

3.3 Enrichment Designs and Adaptive Designs

Instead of removing putative or verified placebo responder, it was proposed that the group of drug responders be enriched during the course of study but without unblinding the study prematurely. The sequential parallel comparison design (SPCD) according to Fava et al. (2003) is quite an elegant attempt to overcome high placebo response rates, particularly in depression trials, for which it was originally developed. It operates in two phases, in which drug and placebo are unbalanced in favor of placebo, e.g., 1:2 or 1:3. Responders during this phase are removed to continue in an open fashion with whatever they had received. Non-responders are re-randomized to switch to the alternative (placebo and drug, respectively) and finish a second phase of the same length. At the end of the trial, data from both phases are pooled for statistical comparison in a conventional way for superiority of drug over placebo (Ivanova and Tamura 2015). This strategy (Silverman et al. 2018) allows better drug-placebo discrimination, even with placebo response rates as high as 40%, as is common in many clinical conditions with patient-reported outcomes (PRO). It is, to the best of our knowledge, the only patented design strategy that seeks to minimize placebo response and improve drug-placebo differences (assay sensitivity).

There are now many variants of the SPCD. A two-way enrichment design (Ivanova and Tamura 2015; Liu et al. 2019) re-randomizes drug responders and placebo non-responders during the first phase to a 50:50 drug: placebos are in a second phase – to maintain blinding until the very end – only the data from both the drug responders and the placebo non-responders from phase I are included in the analysis (Fig. 9). This is also thought to enrich the drug responders and can be combined with a randomized withdrawal strategy.

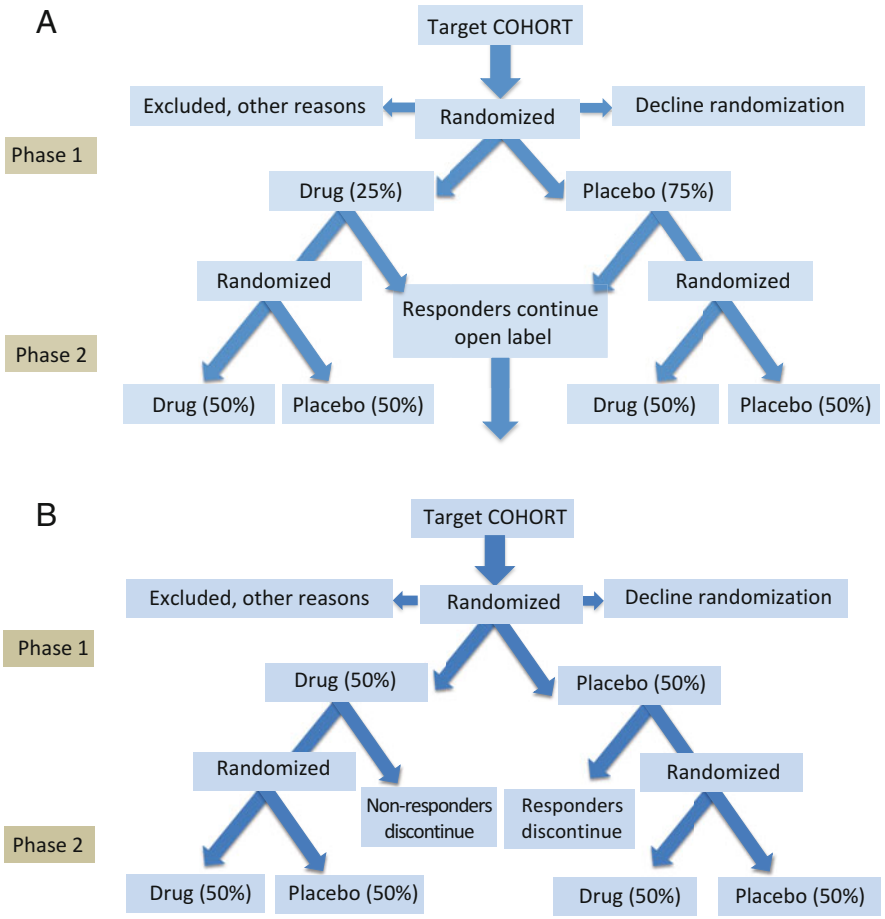


Fig. 9 Two enrichment designs to overcome increased placebo effects in RCT, especially in depression. (**a**) The original sequential parallel comparison design (SPCD) (Fava et al. 2003). The responders in both arms discontinue, and the non-responders are re-randomized to drug or placebo. Note that in Phase 1 more patients are randomized to placebo than to drug (2,1), while in the Phase 2 the randomization ratio is 1:1. (**b**) The two-way enrichment design (TED) (Ivanova and Tamura 2015) where the responders in the placebo arm and the non-responders in the drug arm are excluded while the respective others are re-randomized. Both strategies imply that – as long as both phases are equally long – the data of both phases can be merged to calculate the drug efficacy, but the results of Phase 1 are kept blinded until the very end

Other static or adaptive designs towards the same goal, such as the use of active placebos (Moncrieff et al. 2004; Jensen et al. 2017), have either been forgotten or are described in the literature but still await their clinical validation, e.g., the free-choice paradigm developed by our group (Enck et al. 2012), and our balanced crossover design (Enck et al. 2011a, b) eliminating limitations of conventional balanced

placebo design (Enck et al. 2013a, b). However, not all are suitable for validation in clinical trials, being predominately applicable predominantly in laboratory tests and trials.

4 The Challenge of Omitting Placebos

4.1 Comparative Effectiveness Research (CER)

Above (Sect. 2.4.3), we have already discussed the effects of increasing the likelihood of received active medication in placebo-controlled trials. While its extreme form – all patients receive active medication, either the drug under development or a comparator already on the market, thus having a 100% certainty of being treated by an active drug – may be favored by patients, ethics board, and approval authorities, it raises serious concerns among trialists: Omitting the placebo arm does not eliminate the placebo response but serves only to render it invisible and, therefore, uncontrollable. While we acknowledge its political and ethical intention, it is not without risk of seriously violating ethical and political rules at the same time. This is why:

- From a statistical standpoint, CER studies cannot hypothesize superiority of the novel compound over its comparator but can only claim (null hypothesis) non-inferiority (FDA 2016). However, non-inferiority requires an up to fourfold patient sample (for statistical reasons, see (Flight and Julious 2016)) and therefore violates the Declaration of Helsinki position that the least number of patients should be recruited for clinical trials, while all others should receive active medical care and treatment and not be exposed to medical research.
- CER studies require a comparator, but the choice among all *possible* comparators may co-determine the subsequent statistical testing and thereby the number of patients required to prove non-inferiority. Whether to select the best comparator on the market or an average comparative drug cannot, at the same time, be in the hands of the company developing the new drug nor in those of patients or patient representatives alone, as they may have divergent interests. It is therefore presumably an ethical issue to be decided by ethics boards or legal approval entities.
- Even if the requirement is to select the "best available comparator" on the market, this leaves a hole in the argument: should this be the best available drug on the market where the study is planned, or the best drug available globally, even if it is not available under these specific circumstances (country, healthcare system, clinic, or clinical condition), and who makes this decision? And what if the scientific community cannot even decide on account of different views on the evidence – should this again be decided by ethics board?
- And even if all these questions are answered: The drug under development needs to be indistinguishable from its comparator to allow a double-blinded assessment, and so both need to be produced by the same company, even if one is not its intellectual property. And who will force a company with a drug on the market that happens to be "the best comparator" to voluntarily provide its drug to a

competitor for such a testing that may turn out to be to its disadvantage? Is legal enforcement for such a policy required? Until these issues are solved, CER studies will not become pharmaceutical routine but are greatly dependent on a voluntary agreement among companies, ethics boards, and approval authorities. As was shown, most currently available (2019) non-inferiority trials are not appropriately designed to declare non-inferiority "even if it was worse than either placebo or another historic control" (Tsui et al. 2019).

4.2 Waiting List Controls, Treatment as Usual, and Preference Designs

One of the key issues of most, if not all, RCT designs is the fact that part of what occurs as placebo effect may be the consequence of spontaneous symptom variation and recovery – and it is generally assumed that the contribution of this factor to the overall effects (in both study arms) may be similar and can therefore be neglected when estimating the drug-placebo difference. This may also hold true in a similar way for CER studies.

However, with open-label observational studies, this becomes a factor of the utmost importance, since we are now dealing with one group only, and drug effects tend to be overestimated if non-specific contributions cannot be identified and enumerated. Conventional tools to overcome this limitation are waiting list controls and "treatment as usual," but without proper randomization, they are subject to selection bias, either by the treating physician or by patients who have to agree to "treat or wait" or to novel versus conventional therapy. At the same time, symptom changes during waiting have been described in both directions (for the better, and for the worse) (Hesser et al. 2011; Furukawa et al. 2014). These were not the result of spontaneous symptom variation but were rather due to expectations and disappointment, respectively (Zhu et al. 2014). Waiting lists are generally used in psychotherapy where a blinded application of a sham intervention appears impossible (Gold et al. 2017) but are also used in some three-arm drug trials to control for spontaneous symptom variation (Krogsboll et al. 2009).

If treatment as usual and waiting list are used in RCT, however, they tend to reduce the non-specific effects due to disappointment and overestimate the efficacy of the therapy in the treatment arm (Fig. 10) (Enck and Lackner 2019). Rather than a single waiting list, a step-wedged waiting list (Fig. 11) may add value to this strategy by enabling to calculate a dose-response function for waiting. Patient motivation can be improved by preference designs when more than one type of treatment is available (Fig. 12), and the PD can also be applied to the CER strategy.

4.3 Open-Label ("Real-Life") Observational Studies and Registry and Cohort Studies

Open-label observational studies were usually regarded as Phase IV marketing instruments of the drug industry, since their poor methodology provided little

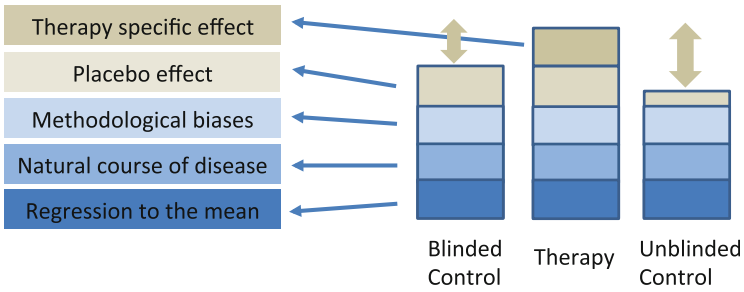


Fig. 10 The effect of unblinding in a RCT, such as with treatment as usual (TAU) and waiting list (WL) controls where blinding is impossible, e.g., in psychotherapy (Enck and Zipfel 2019), or where blinding is broken, e.g., due to AE reporting: The response in the control arm decreases and leads to overestimation of the efficacy in the treatment arm

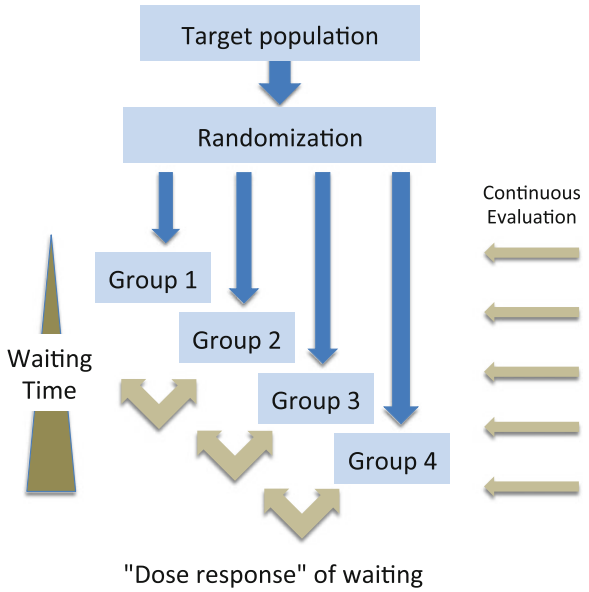


Fig. 11 A modified waiting list (WL) control strategy, where instead of one waiting list, two or more are implemented that reduce disappointment in patients randomized to WL (De Allegri et al. 2008) and allow the calculation of a waiting effect (as dose-response function) that can be separated from the placebo effect. (Reproduced with permission from Springer-Nature, License Number 4627150201527)

additional insight beyond what was known about drug efficacy at the time of approval and because they tended to substantially overestimate drug efficacy due to the lack of controlled conditions. This view changed once it became evident that patient selection during Phase III trials may also be biased – see our above arguments with respect to higher placebo response rates due to lower symptom severity in many

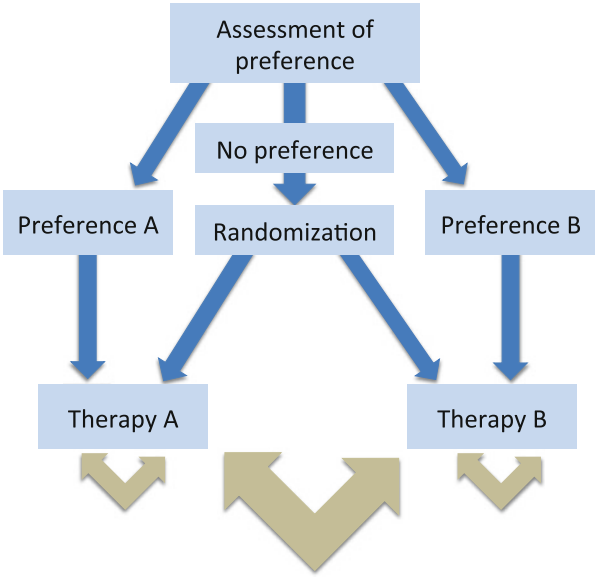


Fig. 12 A variant of a "preference design" where patients choose among true alternative treatments, and only those that do not report a preference are randomized to one of them. This can be applied to comparative effectiveness research (CER) studies where patient will receive a new drug or one already on the market or where true alternatives are to be compared, e.g., drug therapy versus surgery. It also allows comparison of efficacy between randomized and preference-assigned therapies. (Reproduced with permission from Springer-Nature, License Number 4627150201527)

clinical conditions. These may not represent those patients seen in private practices that do not participate in RCTs (the "real-world" patients) (Dal-Re et al. 2018).

In a bid to overcome these limitations, registry or cohort studies have been found helpful; at the same time, they make it possible to control spontaneous symptom variation in a very elegant way and without affecting patient motivation. An early design called "Zelen design" (Zelen 1979) was applied to all randomized placebo-controlled trials (Relton et al. 2010); here, we apply it to observational, Phase IV studies, to the best of our knowledge for the first time (Fig. 13).

Its basic idea is to recruit as many patients as possible for a "pure" observational study, either from a larger existing cohort or even a patient registry; the observational period needs to be defined and justified but can be of any length and recording frequency. The larger the cohort, the better it enables us to identify subgroups, e.g., with specific sociographic or clinical characteristics, specific treatment history, etc. These patients are asked to agree to a symptom monitoring for an extended period of time, but no interference with their ongoing therapy is envisioned (Phase I).

Once the recruitment is settled, the patients who have agreed to participate in the monitoring study are asked again whether they would consider volunteering for an interventional study (Phase II). This can be either a placebo-controlled trial or a CER trial. In both cases, the remaining observation-only group can serve as no-treatment

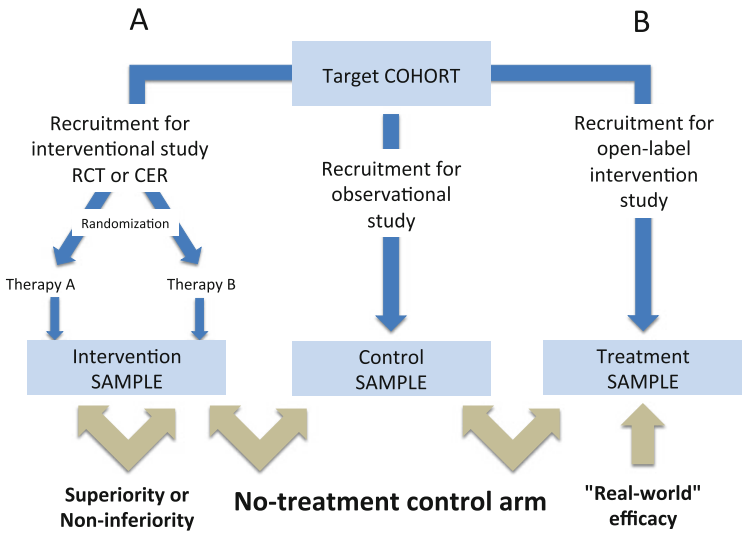


Fig. 13 A design alternative to address and calculate the effect of spontaneous symptom variation on drug and placebo effects which otherwise would require to randomize a patient to a "no-treatment" control, e.g., to a waiting list. The basic idea is to recruit a large number of patients to an "observation-only" study with fixed conditions (e.g., duration, number of observations, clinical conditions) (Zelen 1979; Relton et al. 2010), preferentially from a large patient cohort (registry). In a second step, those that have agreed to take part are subsequently asked whether they would participate in a conventional placebo-controlled or comparative study (**a**) or in an open-label study (**b**), and those not agreeing stay in the "observation-only" group. The larger the initially recruited cohort, the better a match between patients is feasible. Note that this allows a control group even in "real-world" studies, where otherwise controls are impossible to implement – we propose this "controlled open-label trial" (COLT) to overcome the limitations of pure observational studies

control and, if large enough, may even be matched to the treatment cohort with respect to sociographic or clinical criteria.

The "cohort multiple randomized controlled trial" (CMRCT) could even be applied to an observational study and would be the first of its kind to allow proper control of spontaneous symptom variation in observational studies without randomizing patients to a "no-treatment control"; we might call this the "controlled open-label trial" (COLT). Although this would at least give us some idea of the size of the "true" drug effect, we would still need to estimate the size of the contributing placebo effect, e.g., the difference between drug effect sizes in RCTs and in COLT-type studies.

5 Other Challenges for Future Studies

5.1 E-Health and m-Health

As discussed above, increasing the amount and intensity of study center (nurse, doctor) communication with patients is one of the driving factors of higher placebo response rates in some RCTs across medicine, with specific tools such as electronic symptom diaries, app-based reminders, random assessment of treatment effects, and chat rooms for patients to speak to their doctor or nurse when specific problems such as AEs arise.

At the same time, the vast amount of medical information available over the Internet (fact or faked) has dramatically changed the patient-doctor communication in daily practice and in RCTs: AE reporting is now highly correlated with the amount of websites discussing AE, e.g., of biosimilars versus biologics (Macaluso et al. 2018) and statins (Khan et al. 2018). This controls the (expectancy-mediated) "nocebo effect" of drugs and lowers patients' willingness to participate in switch trials (Bakalos and Zintzaras 2018). The same holds true for the switch from branded to non-branded, generic products (Faasse et al. 2013).

More than a quarter of a million medical apps are currently available for various purposes. These include monitoring of treatment success/failure in placebo-controlled trials and in medical routine (FDA 2015), but a systematic evaluation of media-driven placebo effects is still lacking, even in laboratory settings and experiments. However, media-assisted provision e.g., of psychotherapy (by telephone, Internet, computer programs), can be as successful as face-to-face therapy, thus underlining that "digital placebo effects" (Torous and Firth 2016) are at least in a similar range, if not higher in those akin to these media – with more to come in the future: Just imagine having virtual doctors/nurses (Horing et al. 2016), patient avatars, and telemetric, wearable diagnostic and therapeutic tools.

The very same tools, specifically social media, interest groups, and chat rooms, have been found to be ideal if patients wish to exchange views with other patients recruited for the same study. Once this has been established, it may allow them to easily break any blinding code simply by accumulating AE and their frequency, provided that enough patients partake in the discussion. Unblinding, as we have shown (see above, Fig. 11), may not increase, but actually decrease the response to placebo, leading to overestimation of the drug effect as long as the source and size of unblinding remain undisclosed. Instead of fearing such development, doctors and researchers should take an active role to control such effects in the future.

Since patient recruitment has been professionalized over the past decade, social networks and media have taken over the recruitment of patients, e.g., by websites such as "Just Another Lab Rat!™" (www.jalr.org). This further supports what has been called "guinea-pigging", uncontrolled participation of semi-professional volunteers as well as patients in more than one study at a time, or to overrule restrictions for further participation after completion of one study for the next 3, 6, or more months. Until there is a legal basis for a "study patient/volunteer registry" –

controlling recruitment and preventing its misuse but protecting both patient and drug company interests at the same time – the rapid technological development that we currently experience will leave traditional RCT methodology far behind.

5.2 Placebo Effects with Personalized Medicines

One of the promises of high-end medicine, or at least its current vision, is to provide personalized medicine, drugs developed for just one individual patient (or a subgroup of patients) whose genome has been used to develop and design the therapy. Whether this hails the end of the current mode of drug therapy testing is just one open question – with regard to the potential of placebo effects, it may certainly be seen as regressing back into the late nineteenth century: Individualization of therapy was, and still is, the premise of homeopathy and other complementary and alternative medicinal approaches, whether rational and justified or not (Mathie et al. 2018). We therefore expect, as in homeopathy, rising placebo response rates, at least for patient-reported outcomes; fortunately, most personalized therapies are developed initially for diseases with strong biomarkers (such as cancer) that are much less prone to placebo effects.

At the same time, personalized therapy – by definition – prevents the therapy from being controlled for placebo effects, e.g., against a standardized, nonindividual therapy (treatment as usual, for instance, or best therapy available). Even if groups of patients with common genomic markers, identified for a specific, personalized therapy, were to undergo such therapy, it is hard to think of a placebo or otherwise controlled condition that could be justified in terms of ethics, motivation, and costs. One way out of this dilemma would be the revitalization of $N = 1$ methodology (Kronish et al. 2018) that has developed its own strategies of proof-of-principle studies and statistical evaluation using, for example, time series analysis (Shaffer et al. 2018) to prove efficacy for one patient.

One completely different way of avoiding placebo controls was recently described by a drug company (Desai et al. 2013): they screened their entire archive of previously performed RCTs for studies where patients were recruited into a placebo arm of pain trials. After screening and merging the data (which had been stored in different databases) and screening for core data available in all studies, they were left with 203 studies with "historic" controls (called ePlacebo patients) treated with placebo. The idea is that these historic controls be used as a database rather than recruiting future patients into placebo arms of RCTs with novel compounds. The feasibility of such an approach, however, still needs to be verified prospectively, and has recently been questioned, as it may require substantially larger sample sizes as controls, especially with low effect sizes (Schoenfeld et al. 2019).

Table 1 A list of factors that have been found to be associated with the size of the placebo effect in RCT and meta-analyses

Patient characteristics	Doctor characteristics	Disease characteristics	Study characteristics
Age	Age	Baseline severity	Trial duration
Sex	Sex	Duration of illness	Number of sites
Personality	Race	Symptom load	Number of patients
Genes	Education	Previous therapy	Randomization ratio
Proxies	Empathy	Change during run-in	Number of visits
Education	Setting	Outcome measure	Type of assessment
Race	Behavior	Type of intervention	Industry support
Compliance			Type of control
Smoking status			Country/location
			Year of study

Note that the direction of change, e.g., higher placebo effects with higher or lower age, is ignored because of conflicting data and that some factors may have only been identified in only a few trials, e.g., industry sponsorship, while others, e.g., lower symptom severity at baseline, have been found in many RCTs to be associated with higher placebo effects (summary, based on Weimer et al. 2015a, b)

6 Summary

In this review, we have explored different ways of controlling the placebo effects in clinical trials and have described various factors that may increase/decrease the placebo effect in RCTs. As illustrated in Table 1, these factors can be subdivided into four groups, and while not all factors are effective in every study and under all clinical conditions, they show on the whole that – even under the ideal condition of drug therapy, where blinded placebo provision is much easier and warranted than in, e.g., psychotherapy (Enck et al. 2019) – many factors need to be controlled to ascertain that the goal of the clinical trials, fair assessment of superiority of the drug over placebo in RCTs and fair assessment of non-inferiority of the drug compared to another drug in CER trials, is reached. Ignorance towards the placebo effect, which was common in the past, is no longer acceptable; instead, it should be the goal of all therapeutic trials to minimize the placebo effect in clinical trials, while utilizing and maximizing it in clinical routine (Enck et al. 2013a).

Acknowledgments The assistence of Astrid Bahun from Bahun + Design for the artwork is gratefully acknowledged.

References

Agid O, Siu CO, Potkin SG, Kapur S, Watsky E, Vanderburg D, Zipursky RB, Remington G (2013) Meta-regression analysis of placebo response in antipsychotic trials, 1970-2010. Am J Psychiatry 170(11):1335–1344

Aslaksen PM, Myrbakk IN, Hoifodt RS, Flaten MA (2007) The effect of experimenter gender on autonomic and subjective responses to pain stimuli. Pain 129(3):260–268
Bakalos G, Zintzaras E (2018) Drug discontinuation in studies including a switch from an originator to a biosimilar monoclonal antibody: a systematic literature review. Clin Ther 41(1):155–173
Beecher HK (1955) The powerful placebo. J Am Med Assoc 159(17):1602–1606
Benedetti F (2014) Placebo effects. Oxford University Press, Oxford
Benedetti F, Frisaldi E (2014) Creating placebo responders and nonresponders in the laboratory: boons and banes. Pain Manag 4(3):165–167
Benedetti F, Enck P, Frisaldi E, Schedlowski M (eds) (2014) Placebo, Handbook of experimental pharmacology, vol 225. Springer, Heidelberg
Bridge JA, Birmaher B, Iyengar S, Barbe RP, Brent DA (2009) Placebo response in randomized controlled trials of antidepressants for pediatric major depressive disorder. Am J Psychiatry 166(1):42–49
Burke MJ, Kaptchuk TJ, Pascual-Leone A (2018) Challenges of differential placebo effects in contemporary medicine: the example of brain stimulation. Ann Neurol 85(1):12–20
Chae Y, Lee YS, Enck P (2018) How placebo needles differ from placebo pills? Front Psych 9:243
Chey WD, Chey WY, Heath AT, Dukes GE, Carter EG, Northcutt A, Ameen VZ (2004) Long-term safety and efficacy of alosetron in women with severe diarrhea-predominant irritable bowel syndrome. Am J Gastroenterol 99(11):2195–2203
Colagiuri B, Schenk LA, Kessler MD, Dorsey SG, Colloca L (2015) The placebo effect: from concepts to genes. Neuroscience 307:171–190
Colloca L (2018a) Preface (part I). In: Neurobiology of the placebo effect part I. Academic Press, Cambridge, pp xv–xx
Colloca L (2018b) Preface (part II). In: Neurobiology of the placebo effect part II. Academic Press, Cambridge, pp xvii–xxiii
Colloca L, Benedetti F (2006) How prior experience shapes placebo analgesia. Pain 124(1–2):126–133
Colloca L, Petrovic P, Wager TD, Ingvar M, Benedetti F (2010) How the number of learning trials affects placebo and nocebo responses. Pain 151(2):430–439
Dal-Re R, Janiaud P, Ioannidis JPA (2018) Real-world evidence: how pragmatic are randomized controlled trials labeled as pragmatic? BMC Med 16(1):49
Darragh M, Booth RJ, Consedine NS (2014) Investigating the 'placebo personality' outside the pain paradigm. J Psychosom Res 76(5):414–421
De Allegri M, Pokhrel S, Becher H, Dong H, Mansmann U, Kouyate B, Kynast-Wolf G, Gbangou A, Sanon M, Bridges J, Sauerborn R (2008) Step-wedge cluster-randomised community-based trials: an application to the study of the impact of community health insurance. Health Res Policy Syst 6:10
de la Fuente-Fernandez R (2012) The powerful pre-treatment effect: placebo responses in restless legs syndrome trials. Eur J Neurol 19(10):1305–1310
Desai JR, Bowen EA, Danielson MM, Allam RR, Cantor MN (2013) Creation and implementation of a historical controls database from randomized clinical trials. J Am Med Inform Assoc 20(e1): e162–e168
Diener HC, Dowson AJ, Ferrari M, Nappi G, Tfelt-Hansen P (1999) Unbalanced randomization influences placebo response: scientific versus ethical issues around the use of placebo in migraine trials. Cephalalgia 19(8):699–700
Elsenbruch S, Enck P (2015) Placebo effects and their determinants in gastrointestinal disorders. Nat Rev Gastroenterol Hepatol 12(8):472–485
Enck P, Klosterhalfen S (2019) Does sex/gender play a role in placebo and nocebo effects? Conflicting evidence from clinical trials and experimental studies. Front Neurosci 13:160
Enck P, Lackner JM (2019) Cognitive behavioural therapy for IBS: results or treatment as usual? Nat Rev Gastroenterol Hepatol. In press. https://doi.org/10.1038/s41575-019-0174-2
Enck P, Zipfel S (2019) Placebo effects in psychotherapy: a framework. Front Psych 10:456
Enck P, Klosterhalfen S, Kruis W (2005a) Determination of placebo effect in irritable bowel syndrome. Dtsch Med Wochenschr 130(34–35):1934–1937

Enck P, Klosterhalfen S, Kruis W (2005b) Factors affecting therapeutic placebo response rates in patients with irritable bowel syndrome. Nat Clin Pract Gastroenterol Hepatol 2(8):354–355
Enck P, Benedetti F, Schedlowski M (2008) New insights into the placebo and nocebo responses. Neuron 59(2):195–206
Enck P, Klosterhalfen S, Weimer K, Horing B, Zipfel S (2011a) The placebo response in clinical trials: more questions than answers. Philos Trans R Soc Lond Ser B Biol Sci 366(1572):1889–1895
Enck P, Klosterhalfen S, Zipfel S (2011b) Novel study designs to investigate the placebo response. BMC Med Res Methodol 11:90
Enck P, Grundy D, Klosterhalfen S (2012) A novel placebo-controlled clinical study design without ethical concerns – the free choice paradigm. Med Hypotheses 79(6):880–882
Enck P, Bingel U, Schedlowski M, Rief W (2013a) The placebo response in medicine: minimize, maximize or personalize? Nat Rev Drug Discov 12(3):191–204
Enck P, Weimer K, Klosterhalfen S (2013b) Balanced placebo design, active placebos, and other design features for identifying, minimizing and characterizing the placebo response. In: Colloca L, Flaten MA, Meissner K (eds) Placebo and pain: from bench to bedside. Elsevier/Academic Press, Amsterdam, pp 159–173
Enck P, Horing B, Broelz E, Weimer K (2018) Knowledge gaps in placebo research: with special reference to neurobiology. Int Rev Neurobiol 139:85–106
Enck P, Weimer K, Colloca L, Doods S (eds) (2019) Placebo and nocebo effects in psychiatry and beyond. Front Psychiatr. Research Topic https://www.frontiersin.org/research-topics/7727/placebo-and-nocebo-effects-in-psychiatry-and-beyond#articles
Faasse K, Cundy T, Gamble G, Petrie KJ (2013) The effect of an apparent change to a branded or generic medication on drug effectiveness and side effects. Psychosom Med 75(1):90–96
Fava M, Evins AE, Dorer DJ, Schoenfeld DA (2003) The problem of the placebo response in clinical trials for psychiatric disorders: culprits, possible remedies, and a novel study design approach. Psychother Psychosom 72(3):115–127
FDA (2015) Mobile medical applications – guidance for industry and food and drug aministration staff. Federal Drug Administration, Silver Spring
FDA (2016) Non-inferiority clinical trials to establish effectiveness – guidance for industy. Federal Drug Administration, Silver Spring
Flight L, Julious SA (2016) Practical guide to sample size calculations: non-inferiority and equivalence trials. Pharm Stat 15(1):80–89
Ford AC, Moayyedi P (2010) Meta-analysis: factors affecting placebo response rate in the irritable bowel syndrome. Aliment Pharmacol Ther 32(2):144–158
Furukawa TA, Noma H, Caldwell DM, Honyashiki M, Shinohara K, Imai H, Chen P, Hunot V, Churchill R (2014) Waiting list may be a nocebo condition in psychotherapy trials: a contribution from network meta-analysis. Acta Psychiatr Scand 130(3):181–192
Geers AL, Wellman JA, Fowler SL, Helfer SG, France CR (2010) Dispositional optimism predicts placebo analgesia. J Pain 11(11):1165–1171
Gold SM, Enck P, Hasselmann H, Friede T, Hegerl U, Mohr DC, Otte C (2017) Control conditions for randomised trials of behavioural interventions in psychiatry: a decision framework. Lancet Psychiatry 4(9):725–732
Grelotti DJ, Kaptchuk TJ (2011) Placebo by proxy. BMJ 343:d4345
Grünbaum A (1986) The placebo concept in medicine and psychiatry. Psychol Med 16(1):19–38
Hall KT, Loscalzo J, Kaptchuk T (2018) Pharmacogenomics and the placebo response. ACS Chem Neurosci 9(4):633–635
Hesser H, Weise C, Rief W, Andersson G (2011) The effect of waiting: a meta-analysis of wait-list control groups in trials for tinnitus distress. J Psychosom Res 70(4):378–384
Horing B, Weimer K, Muth ER, Enck P (2014) Prediction of placebo responses: a systematic review of the literature. Front Psychol 5:1079
Horing B, Weimer K, Muth ER, Enck P (2015) Prediction of symptom change in placebo versus no-treatment group in experimentally induced motion sickness. Appl Psychophysiol Biofeedback 40(3):163–172

Horing B, Newsome ND, Enck P, Babu SV, Muth ER (2016) A virtual experimenter to increase standardization for the investigation of placebo effects. BMC Med Res Methodol 16(1):84
Irving G, Neves AL, Dambha-Miller H, Oishi A, Tagashira H, Verho A, Holden J (2017) International variations in primary care physician consultation time: a systematic review of 67 countries. BMJ Open 7(10):e017902
Ivanova A, Tamura RN (2015) A two-way enriched clinical trial design: combining advantages of placebo lead-in and randomized withdrawal. Stat Methods Med Res 24(6):871–890
Jairath V, Zou G, Parker CE, Macdonald JK, Mosli MH, Khanna R et al (2016) Systematic review and meta-analysis: placebo rates in induction and maintenance trials of ulcerative colitis. J Crohns Colitis 10(5):607–618
Jensen KB, Petrovic P, Kerr CE, Kirsch I, Raicek J, Cheetham A, Spaeth R, Cook A, Gollub RL, Kong J, Kaptchuk TJ (2014) Sharing pain and relief: neural correlates of physicians during treatment of patients. Mol Psychiatry 19(3):392–398
Jensen JS, Bielefeldt AO, Hrobjartsson A (2017) Active placebo control groups of pharmacological interventions were rarely used but merited serious consideration: a methodological overview. J Clin Epidemiol 87:35–46
Jutte R (2013) The early history of the placebo. Complement Ther Med 21(2):94–97
Kaptchuk TJ (1998) Intentional ignorance: a history of blind assessment and placebo controls in medicine. Bull Hist Med 72(3):389–433
Kaptchuk TJ, Kelley JM, Deykin A, Wayne PM, Lasagna LC, Epstein IO, Kirsch I, Wechsler ME (2008) Do "placebo responders" exist? Contemp Clin Trials 29(4):587–595
Kelley JM, Lembo AJ, Ablon JS, Villanueva JJ, Conboy LA, Levy R, Marci CD, Kerr CE, Kirsch I, Jacobson EE, Riess H, Kaptchuk TJ (2009) Patient and practitioner influences on the placebo effect in irritable bowel syndrome. Psychosom Med 71(7):789–797
Khan A, Redding N, Brown WA (2008) The persistence of the placebo response in antidepressant clinical trials. J Psychiatr Res 42(10):791–796
Khan S, Holbrook A, Shah BR (2018) Does Googling lead to statin intolerance? Int J Cardiol 262:25–27
Kirsch I (2000) Are drug and placebo effects in depression additive? Biol Psychiatry 47(8):733–735
Kirsch I (2014) The emperor's new drugs: medication and placebo in the treatment of depression. Handb Exp Pharmacol 225:291–303
Kirsch I (2016) The placebo effect in the treatment of depression. Verhaltenstherapie 26:55–61
Klosterhalfen S, Kellermann S, Pan F, Stockhorst U, Hall G, Enck P (2005a) Effects of ethnicity and gender on motion sickness susceptibility. Aviat Space Environ Med 76(11):1051–1057
Klosterhalfen S, Kellermann S, Stockhorst U, Wolf J, Kirschbaum C, Hall G, Enck P (2005b) Latent inhibition of rotation chair-induced nausea in healthy male and female volunteers. Psychosom Med 67(2):335–340
Klosterhalfen S, Pan F, Kellermann S, Enck P (2006) Gender and race as determinants of nausea induced by circular vection. Gend Med 3(3):236–242
Kobak KA (2010) Inaccuracy in clinical trials: effects and methods to control inaccuracy. Curr Alzheimer Res 7(7):637–641
Kobak KA, Kane JM, Thase ME, Nierenberg AA (2007) Why do clinical trials fail? The problem of measurement error in clinical trials: time to test new paradigms? J Clin Psychopharmacol 27(1):1–5
Krogsboll LT, Hrobjartsson A, Gotzsche PC (2009) Spontaneous improvement in randomised clinical trials: meta-analysis of three-armed trials comparing no treatment, placebo and active intervention. BMC Med Res Methodol 9:1
Kronish IM, Hampsey M, Falzon L, Konrad B, Davidson KW (2018) Personalized (N-of-1) trials for depression: a systematic review. J Clin Psychopharmacol 38(3):218–225
Leal Rato M, Duarte GS, Ferreira AN, Alves M, Mainoli B, Teodoro T, Mestre TA, Costa J, Ferreira JJ (2019) Nocebo response in Parkinson's disease: a systematic review and meta-analysis. Parkinsonism Relat Disord. In press. https://doi.org/10.1016/j.parkreldis.2019.04.015

Legrain V, Iannetti GD, Plaghki L, Mouraux A (2011) The pain matrix reloaded: a salience detection system for the body. Prog Neurobiol 93(1):111–124
Lipset CH (2014) Engage with research participants about social media. Nat Med 20(3):231
Liu Y, Rybin D, Heeren TC, Doros G (2019) Comparison of novel methods in two-way enriched clinical trial design. Stat Med. In press. https://doi.org/10.1002/sim.8288
Macaluso FS, Maida M, Ventimiglia M, Renna S, Cottone M, Orlando A (2018) Factors affecting clinical and endoscopic outcomes of placebo arm in trials of biologics and small molecule drugs in ulcerative colitis: a meta-analysis. Inflamm Bowel Dis 25(6):987–997
Maddocks M, Kerry R, Turner A, Howick J (2016) Problematic placebos in physical therapy trials. J Eval Clin Pract 22(4):598–602
Mahr A, Golmard C, Pham E, Lordache L, Deville L, Faure P (2017) Types, frequencies, and burden of nonspecific adverse events of drugs: analysis of randomized placebo-controlled clinical trials. Pharmacoepidemiol Drug Saf 26(7):731–741
Mallinckrodt CH, Zhang L, Prucka WR, Millen BA (2010) Signal detection and placebo response in schizophrenia: parallels with depression. Psychopharmacol Bull 43(1):53–72
Mathie RT, Ulbrich-Zurni S, Viksveen P, Roberts ER, Baitson ES, Legg LA, Davidson JRT (2018) Systematic review and meta-analysis of randomised, other-than-placebo controlled, trials of individualised homeopathic treatment. Homeopathy 107(4):229–243
May M (2019) Twenty-five ways clinical trials have changed in the last 25 years. Nat Med 25(1):2–5
Moncrieff J, Wessely S, Hardy R (2004) Active placebos versus antidepressants for depression. Cochrane Database Syst Rev 1:Cd003012
Moore RA, Wiffen PJ, Eccleston C, Derry S, Baron R, Bell RF, Furlan AD, Gilron I, Haroutounian S, Katz NP, Lipman AG, Morley S, Peloso PM, Quessy SN, Seers K, Strassels SA, Straube S (2015) Systematic review of enriched enrolment, randomised withdrawal trial designs in chronic pain: a new framework for design and reporting. Pain 156(8):1382–1395
Munkholm K, Paludan-Muller AS, Boesen K (2019) Considering the methodological limitations in the evidence base of antidepressants for depression: a reanalysis of a network meta-analysis. BMJ Open 9(6):e024886
Papakostas GI, Fava M (2009) Does the probability of receiving placebo influence clinical trial outcome? A meta-regression of double-blind, randomized clinical trials in MDD. Eur Neuropsychopharmacol 19(1):34–40
Quessy SN, Rowbotham MC (2008) Placebo response in neuropathic pain trials. Pain 138(3):479–483
Rakel DP, Hoeft TJ, Barrett BP, Chewning BA, Craig BM, Niu M (2009) Practitioner empathy and the duration of the common cold. Fam Med 41(7):494–501
Rehn J, Schuster K (2017) Clinic design as placebo-using design to promote healing and support treatments. Behav Sci 7(4):E77
Relton C, Torgerson D, O'Cathain A, Nicholl J (2010) Rethinking pragmatic randomised controlled trials: introducing the "cohort multiple randomised controlled trial" design. BMJ 340:c1066
Rheker J, Rief W, Doering BK, Winkler A (2018) Assessment of adverse events in clinical drug trials: identifying amitriptyline's placebo- and baseline-controlled side effects. Exp Clin Psychopharmacol 26(3):320–326
Rickels K, Lipman R, Raab E (1966) Previous medication, duration of illness and placebo response. J Nerv Ment Dis 142(6):548–554
Rief W, Nestoriuc Y, Weiss S, Welzel E, Barsky AJ, Hofmann SG (2009a) Meta-analysis of the placebo response in antidepressant trials. J Affect Disord 118(1–3):1–8
Rief W, Nestoriuc Y, von Lilienfeld-Toal A, Dogan I, Schreiber F, Hofmann SG, Barsky AJ, Avorn J (2009b) Differences in adverse effect reporting in placebo groups in SSRI and tricyclic antidepressant trials: a systematic review and meta-analysis. Drug Saf 32(11):1041–1056
Rutherford BR, Sneed JR, Roose SP (2009) Does study design influence outcome? The effects of placebo control and treatment duration in antidepressant trials. Psychother Psychosom 78(3):172–181

Rutherford BR, Sneed JR, Tandler JM, Rindskopf D, Peterson BS, Roose SP (2011) Deconstructing pediatric depression trials: an analysis of the effects of expectancy and therapeutic contact. J Am Acad Child Adolesc Psychiatry 50(8):782–795

Schedlowski M, Enck P, Rief W, Bingel U (2015) Neuro-bio-behavioral mechanisms of placebo and Nocebo responses: implications for clinical trials and clinical practice. Pharmacol Rev 67(3):697–730

Schoenfeld DA, Finkelstein DM, Macklin E, Zach N, Ennist DL, Taylor AA, Atassi N (2019) Design and analysis of a clinical trial using previous trials as historical control. Clin Trials. https://doi.org/10.1177/1740774519858914

Shaffer JA, Kronish IM, Falzon L, Cheung YK, Davidson KW (2018) N-of-1 randomized intervention trials in Health Psychology: a systematic review and methodology critique. Ann Behav Med 52(9):731–742

Shah R, Ogden J (2006) 'What's in a face?' The role of doctor ethnicity, age and gender in the formation of patients' judgements: an experimental study. Patient Educ Couns 60(2):136–141

Shah E, Triantafyllou K, Hana AA, Pimentel M (2014) Adverse events appear to unblind clinical trials in irritable bowel syndrome. Neurogastroenterol Motil 26(4):482–488

Silverman RK, Ivanova A, Fine J (2018) Sequential parallel comparison design with binary and time-to-event outcomes. Stat Med 37(9):1454–1466

Spiller RC (1999) Problems and challenges in the design of irritable bowel syndrome clinical trials: experience from published trials. Am J Med 107(5a):91s–97s

Stein DJ, Baldwin DS, Dolberg OT, Despiegel N, Bandelow B (2006) Which factors predict placebo response in anxiety disorders and major depression? An analysis of placebo-controlled studies of escitalopram. J Clin Psychiatry 67(11):1741–1746

Stille G (1980) Preface. In: Handbook of experimental pharmacology, vol 55(I). Springer, Berlin, p V

Suchman AL, Ader R (1992) Classic conditioning and placebo effects in crossover studies. Clin Pharmacol Ther 52(4):372–377

Torous J, Firth J (2016) The digital placebo effect: mobile mental health meets clinical psychiatry. Lancet Psychiatry 3(2):100–102

Tsui M, Rehal S, Jairath V, Kahan BC (2019) Most non-inferiority trials were not designed to preserve active comparator treatment effects. J Clin Epidemiol 110:82–89

Vambheim SM, Flaten MA (2017) A systematic review of sex differences in the placebo and the nocebo effect. J Pain Res 10:1831–1839

Ventriglio A, Magnifico G, Borraccino L, Rinaldi A, Bellomo A (2018) Placebo and cultural responses. Nord J Psychiatry 72(supp 1):S33–S35

Wang RS, Hall KT, Giulianini F, Passow D, Kaptchuk TJ, Loscalzo J (2017) Network analysis of the genomic basis of the placebo effect. JCI Insight 2(11):93911

Wartolowska K, Judge A, Hopewell S, Collins GS, Dean BJ, Rombach I, Brindley D, Savulescu J, Beard DJ, Carr AJ (2014) Use of placebo controls in the evaluation of surgery: systematic review. BMJ 348:g3253

Waschbusch DA, Pelham WE Jr, Waxmonsky J, Johnston C (2009) Are there placebo effects in the medication treatment of children with attention-deficit hyperactivity disorder? J Dev Behav Pediatr 30(2):158–168

Wechsler ME, Kelley JM, Boyd IO, Dutile S, Marigowda G, Kirsch I, Israel E, Kaptchuk TJ (2011) Active albuterol or placebo, sham acupuncture, or no intervention in asthma. N Engl J Med 365(2):119–126

Weimer K, Enck P (2014) Traditional and innovative experimental and clinical trial designs and their advantages and pitfalls. Handb Exp Pharmacol 225:237–272

Weimer K, Gulewitsch MD, Schlarb AA, Schwille-Kiuntke J, Klosterhalfen S, Enck P (2013) Placebo effects in children: a review. Pediatr Res 74(1):96–102

Weimer K, Colloca L, Enck P (2015a) Age and sex as moderators of the placebo response – an evaluation of systematic reviews and meta-analyses across medicine. Gerontology 61(2):97–108

Weimer K, Colloca L, Enck P (2015b) Placebo effects in psychiatry: mediators and moderators. Lancet Psychiatry 2(3):246–257

Whalley B, Hyland ME, Kirsch I (2008) Consistency of the placebo effect. J Psychosom Res 64(5):537–541

White P, Lewith G, Hopwood V, Prescott P (2003) The placebo needle, is it a valid and convincing placebo for use in acupuncture trials? A randomised, single-blind, cross-over pilot trial. Pain 106(3):401–409

Wolf S (1959) The pharmacology of placebos. Pharmacol Rev 11:689–704

Zelen M (1979) A new design for randomized clinical trials. N Engl J Med 300(22):1242–1245

Zhu Z, Zhang L, Jiang J, Li W, Cao X, Zhou Z, Zhang T, Li C (2014) Comparison of psychological placebo and waiting list control conditions in the assessment of cognitive behavioral therapy for the treatment of generalized anxiety disorder: a meta-analysis. Shanghai Arch Psychiatry 26(6):319–331

Pharmacoepidemiology

Nicholas Moore, Patrick Blin, and Cécile Droz

Contents

1 Introduction 434
2 Data Sources in Pharmacoepidemiology 435
2.1 Primary Data Collection 435
2.2 Secondary Sources of Data 436
3 Methods and Designs in Pharmacoepidemiology 438
3.1 Ascertainment of Exposure 439
3.2 Ascertainment of Events and Diseases 439
3.3 Selection of Participants: Exposure-Based 440
3.4 Comparative Effectiveness or Safety Studies 441
3.5 Matching 441
3.6 Analysis 442
3.7 Selection of Participants: Disease-Based 443
3.8 Selection of Participants: Event-Based 443
3.9 The Comparator 445
3.10 Biases 445
4 Conclusion 446
References 447

Abstract

At the time of their marketing authorization, the effects of drugs and especially their efficacy have been mostly studied in randomized controlled clinical trials (RCT), comparing them to placebo or to existing drugs. However, RCT are by nature limited in their extent, and the often stringent inclusion and exclusion criteria destined to provide for homogeneous study populations reduce the generalizability of RCT results.

N. Moore (✉) · P. Blin · C. Droz
Bordeaux PharmacoEpi, INSERM CIC1401, Université de Bordeaux, Bordeaux, France
e-mail: nicholas.moore@u-bordeaux.fr

J. E. Barrett et al. (eds.), *Concepts and Principles of Pharmacology*,
Handbook of Experimental Pharmacology 260, https://doi.org/10.1007/164_2019_318

The post-authorization evaluation of drugs (pharmacoepidemiology or real-world evidence (RWE)) covers the description of drug utilization and population risks or benefits of these drugs after they have been marketed and provided to their target populations. Though field studies have existed for a long time, modern pharmacoepidemiology has been made possible essentially by the emergence of large population databases compiled from claims data or electronic health records. The methods can be exposure or disease-based cohorts or event-driven case-based studies, tailored to the specific questions to be answered. They rely on scrupulous analysis and execution of impeccable methodology, to ensure the most reliable results possible.

Pharmacoepidemiology requires knowledge of the pharmacology of drugs, of the clinical aspects of diseases and disease management, and of the epidemiological methods that can apply.

Keywords
Pharmacoepidemiology · Population databases

1 Introduction

At the time of their marketing, the effects of drugs and especially their efficacy have been studied mostly in randomized controlled clinical trials (RCT), comparing them to placebo or to existing drugs. However, these RCT are by nature limited in their extent. Stringent inclusion and exclusion criteria are destined to provide for homogeneous study populations and reduce response variability. These features reduce the representativeness of RCT to the future user population (Steg et al. 2007; Blin et al. 2017). Once the drugs have proven efficacy and a measure of safety, and are on the market, they will be prescribed to patients with concomitant diseases and medication or other risk factors that have usually been excluded from RCT (Blin et al. 2017). When several new drugs are marketed within a short time frame, as is often the case with new drug classes (e.g. direct-acting anti-anticoagulants), there is no comparative RCT. It is very unlikely that any pharmaceutical company will devise at great cost a directly comparative RCT, comparing their drug to other direct competitors. In addition, the introduction of new drugs or therapeutic options to the market may shift user populations of previously marketed drugs and modify their benefit–risk balance.

There is therefore a need to study the interactions drugs with their target populations, within a real-life environment. This includes the description of how it is used (drug utilization studies), how it compares to similar drugs within the same disease environment (comparative effectiveness), whether any new safety concerns arise, or quantify previously identified concerns (post-authorization safety studies). In addition, even before a drug is marketed, its future environment and place on the market can be anticipated and modelled, as well, once it has effectively been marketed, as its real impact on health economics (health technology assessment).

By definition, pharmacoepidemiology studies are non-interventional, i.e., there is no influence or there should be no influence on the choice of therapeutic options studied, in contrast with interventional studies (RCT) where treatment is assigned to each patient.

Pharmacoepidemiology has long been limited to field studies such as case–control studies (Pierfitte et al. 2001) or simple cohort studies describing drug utilization, though some early databases such as Saskatchewan in Canada, VAMP research (now CPRD) in the UK, Medicare or Health maintenance organizations (HMO) such as Kaiser-Permanente in the USA paved the way for the large population resources now available.

Data resources such as countrywide healthcare systems databases have become readily available, and hospital-based data repositories or electronic health records are opening new possibilities, including multi-database and multi-country efforts involving very large populations of dozens to hundreds of million patients.

Pharmacoepidemiological studies require knowledge of epidemiological and statistical methods as well as the resources available to study the drugs and their specificities, but also knowledge of the pharmacology of the drugs being studied, and of the diseases involved as indications, efficacy or safety outcomes.

Finally, pharmacoepidemiology, as the study of drug effects in large populations, might be seen as just another method in experimental pharmacology, on a larger scale, much like the observation of drug effects in individual animals or persons in traditional experimental pharmacology.

2 Data Sources in Pharmacoepidemiology

Pharmacoepidemiological studies can involve primary collection of data (field studies) or secondary use of data previously collected for other ends (claims databases, electronic health records (EHR)).

2.1 Primary Data Collection

In primary data collection, studies are devised ad hoc, much as clinical trials, and specific information, such as quality of life, lifestyle data not present in medical records, blood or DNA samples can be acquired.

Field studies may be obligatory when the data needed is not readily found in the claims or EHR databases, such as the site of an ocular injection, or the presence of lifestyle characteristics, or again the reasons for which a drug may have been prescribed or stopped. Because these studies involve contact with patients and the generation of primary data, they are subject to patient safety requirements and informed consent. The rules for reporting adverse events will also not be the same as for secondary data [https://www.ema.europa.eu/en/human-regulatory/post-authorisation/pharmacovigilance/good-pharmacovigilance-practices#final-gvp-modules-section. Guideline on good pharmacovigilance practices (GVP) Module VI – Collection, management and submission of reports of suspected adverse reactions to medicinal products (Rev 2)].

Studies may also combine data from an ad-hoc field study and claims databases, either directly where patients are identified, characterized and recruited by

prescribers, but then followed in claims databases including after patient randomization (Mackenzie et al. 2016; MacDonald et al. 2013, 2014; Flynn et al. 2014) or indirectly, by verifying in a field study potential associations of confounders with prescribing. In the absence of an association (e.g. a drug is not preferentially prescribed in smokers), then that potential confounder is just a risk modifier and can be neglected in database studies.

An alternative is to identify patients in a database, then return to the patient and/or prescriber to complete the data. Since this may infringe on patient confidentiality protection laws, this design, which could be thought optimal to identify and enrol patients in highly targeted field studies, may not be easily feasible (Depont et al. 2007a, b).

The benefits of primary data collection (field studies) are their great flexibility, since the data acquisition is tailored to the needs of the study. Their main drawback is cost: since pharmacoepidemiological studies generally require large numbers of patients, identifying, recruiting and following large numbers of patients is usually difficult and expensive.

This is true whatever the study design. In some specific cases, there is no option, such as for rare genetic diseases, where patients are often included in registries and more easily available. Patients may also be recruited through disease-based associations, with a clear risk of recruitment bias: patients who participate in disease associations may not be representative of the whole patient population.

In some cases, especially when expensive drugs are studied, these may be on specific dispensing registries, and serve to identify patients (e.g. for targeted cancer therapies), which will allow characterization of users, and specific follow-up including, for instance, reasons for drug discontinuation (if recorded) and/or progression-free survival (Fourrier-Reglat et al. 2014a, b; Noize et al. 2017; Rouyer et al. 2018). If medical records are complete enough, it may not even be necessary to interact with the patient.

A basic principle is to include patients only after the drug has been prescribed, without interfering with the prescription process. There may be some interventions in the study, such as blood or DNA sampling, or recording of QOL variables, but these should not interfere with the free choice by the prescriber of the therapeutic options, and do not alter the observational status of the study (Guiard et al. 2019).

2.2 Secondary Sources of Data

In secondary data sources, the data is usually already present at the time of the study, and it would generally not be possible to enrich the dataset, though new developments in clinical data repositories might change this in the near future. These data sources might be medical records (electronic health records (EHR)) or data derived from healthcare insurance systems (claims data).

Electronic Health Records

These are databases or repositories of medical records, from general practitioners (GP) or hospitals. They are based on the voluntary recording by participating

physicians of the clinical details of the patients they follow, often within patient management software. This will include outpatient diagnoses and prescriptions (not dispensing), results of lab tests and other exams (if entered) and results of specialist visits, or hospital discharge summaries, as well as lifestyle characteristics. The quality and completeness of the data depends on the heath care professional's input. Ideally this will be done by healthcare records management software, the data being transmitted to the database after anonymization. In this case the data will actually be used for patient management. Quality and completeness of the data needs to be regularly verified, and missing data may be an issue. The completeness of data may also be an issue and depend on the healthcare system. If the GP is the overall curator of all the patient's healthcare, the data may be presumed complete, though hospital data may not always be fully transcribed, nor might some specialist visits (Jick et al. 2003). Though these data are in principle anonymized, it is possible under certain circumstances to return to the originator GP to obtain precisions on specific points, or for quality control.

Claims Databases

Claims databases contain recordings of all healthcare encounters that are covered by the healthcare system or insurance company. This may include outpatient medical consultations, drugs or devices dispensed, lab test or imaging, paramedical interventions, but also hospital admissions including diagnoses and procedures. Often there is the recording that this lab test or exam has been done, but not always the results of such tests. In some countries there are outpatient diagnoses, or records of chronic conditions, and linkage to national lab test or pathology repositories. This is especially true in Nordic countries. These data might also be linked to specific registries such as cancer, diabetes or rare disease registries, or death registries including or not its cause.

Depending on the data sources, such databases may contain much information on medical expenses than can provide direct or indirect information on various potential confounders. For instance, if they do not contain such lifestyle information as smoking or BMI, they do contain information on their medical consequences, such as chronic bronchitis, sinus infections, including use of antibiotics, peripheral arterial disease, tobacco cessation aids or devices, specialist consultations, etc. Increased BMI may be related to diabetes (identified directly or through its treatment), osteoarthritis and procedures such as knee or hip replacement, and the use of drugs for these indications, but also bariatric surgery and other procedures related to obesity, or the use of assistance such as walkers or canes, and use of spas and weight-reducing programs. All these variables can be included in modern statistical analyses such as high-dimensional propensity scores and disease risk scores (Schneeweiss et al. 2009; Neugebauer et al. 2015).

The data in claims databases are collected systematically and prospectively. They concern all the information for all the patients covered by the healthcare system. This might be lifelong for the whole population as in France or in Nordic Countries, or limited to specific areas (in Germany, Italy or Canada), or to specific ages, social status and resources (as in the USA), which may limit the usefulness or

representativeness of such claims databases (Trifiro et al. 2009; Coloma et al. 2011; Bezin et al. 2017).

Chart Reviews

A third approach to secondary use of data is the concept of chart reviews, where patient files are examined for the presence of specific events such as indicators of cancer progression, which will usually not be found in the claims data (Fourrier-Reglat et al. 2014a, b; Rouyer et al. 2018). Chart reviews may concern patients treated with specific drugs, or recorded drug exposures before events such as liver transplantation (Gulmez et al. 2013a, 2015).

Finally different data sources may be combined in datahubs or data repositories that aggregate information from claims databases and from clinical records, in-hospital or outpatient, data from registries, including results of lab tests or description of DNA sequencing or target information, as well as information from the emerging wearable devices (Dhainaut et al. 2018). These multisource data linkages mutually enrich all the datasources.

3 Methods and Designs in Pharmacoepidemiology

Pharmacoepidemiology is based on the study of the conjunction of subjects, exposures and events. One may consider several approaches in pharmacoepidemiology: event-based or exposure-driven methods.

Exposure-driven methods describe the use of a drug in the population, the drivers for that use, and the consequences of such use. This can represent cross-sectional studies, simply describing the user population. Possibly including past or concomitant events, prescriptions or diseases. If the subject of interest is what happens in drug users, then these studies are based on cohort methods, where patients are usually included at the time of the first prescription or dispensing of a drug and followed forward in time. Cohorts are by definition prospective.

One might be interested in the occurrence of a specific event and what happened before it that might be causally involved. These case-based studies are by definition retrospective studies, since subjects are included at the time of the event, and what is studied is what happened before.

Finally, the description of disease healthcare burden or management profiles, where patients are identified by a disease or event, and followed thereafter to describe management and costs, and the consequences of management represents what is called health technology assessment (HTA). In this field, the subject of interest in not so much the health consequences of drug utilization, but its economic consequences within a healthcare system. The main methods used here would be meta-analyses exploring the magnitude of effects compared to the cost of the drug, using all available information from clinical trials and from observational studies, which are the input to mathematical models. A very well-known source of HTA information is the National Centre for Clinical Excellence in the UK, also known as NICE. Most countries and/or health care and health insurance systems have HTA

activities, to optimize the use of resources tailored to their specific healthcare systems.

3.1 Ascertainment of Exposure

Ascertainment of exposure is fundamental to any pharmacoepidemiological study, by definition. Exposure might be ascertained from the prescriber of the patient in ad-hoc field studies, or from medical records of prescription of the drug (EHR), indicating an intent of exposure, or from claims databases indicating the dispensing of the drug, i.e. that the subject was actually in possession of the drug.

Asking patients about exposures is often uncertain, and specific methodologies are needed. Knowing what drugs were actually prescribed, or better yet bought, will focus on suspect drugs. In some cases, such as over-the-counter medicines that can be freely bought, querying the patient may be the only source of information, unless there are exhaustive pharmacy records (and drug sales are restricted to pharmacies only, notwithstanding internet sales) (Moore et al. 1993; Noize et al. 2009, 2012). In some rare cases, exposure can be ascertained by biomarkers such as plasma drug concentrations (Moore et al. 2001).

Knowledge of exposure may also vary depending on the source: medical records may cite only the drugs prescribed by the keeper of the records: GP prescriptions, or in-hospital prescriptions. Claims databases include only reimbursed dispensing of drugs covered by the insurance scheme, usually not including OTC drugs. Drugs dispensed in the hospital may be included in the hospital cost, and not itemized, except perhaps very expensive drugs that are covered separately.

3.2 Ascertainment of Events and Diseases

Events can be considered as outcomes, as patient selection criteria, or as background variables. Events as outcomes in cohorts usually consider the first occurrence of the event during follow-up (e.g. bleeding in a cohort of patients on anticoagulants, myocardial infarction (MI) in coronary prevention studies, or cancer progression), and of course death. These will most often be identified from hospital diagnoses and ICD10 coding (Duong et al. 2018). Events may also be the initiator for the inclusion of patients in a study, for instance in disease-based follow-up studies such as current practice for the management of a disease (for instance secondary prevention post MI (Blin et al. 2017; Bezin et al. 2018)). Events are also the point of entry in case-based studies, studying causes of such events, e.g., drugs associated with the onset of hepatic injury or heart failure (Gulmez et al. 2013a, 2015; Moore et al. 2019). Finally diagnoses serve as indicators of diseases that may be potential confounders or indicators of risks, used to devise risk or prognostic scores, existing at the time of or prior to inclusion, or arising during the course of follow-up.

The identification of events may also include more information than just an ICD-10 code, such as a stay in intensive care for myocardial infarction (Blin et al.

2019a). These diagnoses can be the object of external validations, comparing codes to the actual patient files (Bezin et al. 2015; Bosco-Levy et al. 2019), or internal validation using adjudication committees with a complete patient health utilization history to the code itself, when it is not possible to link claims data to medical records (Pladevall-Vila et al. 2019; Wentzell et al. 2018; Czwikla et al. 2017).

Identification of diseases as previous history will rely on patient or physician interrogation in field studies as in clinical trials, with uncertainties (Fourrier-Reglat et al. 2010a, b), or on previously registered diseases, procedures or treatments indicative of such diseases in EHR or claims databases. One issue may be the depth (duration) of the database or the previous history one may wish to explore. Quite often only a few years are available, especially for commercial claims databases, when patients may change their healthcare provider.

3.3 Selection of Participants: Exposure-Based

Subjects can be selected on the exposure of interest. Such exposure-based studies may have several types of main objectives: drug utilization, non-comparative or comparative outcomes studies.

Drug Utilization Studies

Generally, new users of a drug will be selected, and described in a cross-sectional study of drug utilization, or included in a cohort, and followed for ulterior events. This would traditionally be done in field studies using so-called registries or phase-IV (post-marketing) studies. In such studies with primary data collections, new users of a new drug are identified by the prescriber and described for factors associated with the prescription, including lifestyle factors, and followed for common events, such as drug cessation and the reasons thereof, or common adverse reactions. Except in very high-risk patients such as in oncology, the event rates for serious events would be too low to allow quantification, in these studies of limited size. They may however provide valuable information on less-serious common events that would not warrant hospitalization, and would not be captured in claims databases, and on potential confounders not included in population databases.

Using large population databases will provide for the detection, description and follow-up of the very first users of the drug, and how this use might change over time.

Outcomes Cohorts

These will provide event rates for events resulting in hospital admissions (serious adverse reactions), or that may have therapeutic markers (such as the use of antidepressant drugs to identify depression) using prescription symmetry analysis (Hallas 1996; Petri et al. 1988; Idema et al. 2018). These cohort studies can be very large and will provide unbiased whole-population event rates (Miranda et al. 2017). These event rates may be considered in the absolute, for instance in the absence of non-drug related occurrences of the event, or compared to those observed in pivotal

clinical trials (Blin et al. 2017) especially in cancer (Fourrier-Reglat et al. 2014a, b; Noize et al. 2017; Rouyer et al. 2018).

Among the benefits of these new-user cohort studies are the possible comparisons with pivotal clinical trials for outcomes that were identified in these trials, including efficacy outcomes so that the applicability and representativeness of these trials can be appreciated (Garbe et al. 2013).

Another benefit is that these studies will allow the evaluation of the risks associated with the drugs, and their possible benefits, especially with the final arbiter, all-cause death. This will inform the assessment of benefit–risk for drugs used in serious conditions, or to prevent serious outcomes. Outcomes can also be compared to those historically observed with other drugs in the same indication in different studies, or to these other drugs in comparative effectiveness studies.

3.4 Comparative Effectiveness or Safety Studies

Comparative effectiveness or safety studies compare event rates in similar populations treated with different drugs. These will be conducted according to a "new user" cohort design, comparing two or more marketed drugs, reducing comparison biases with adapted methods. The inclusion only of new users is necessary to avoid selection biases, such as depletion of susceptibles, whereby patients who remain on a drug (prevalent users) are those who tolerate the drug or benefit from it (Moride and Abenhaim 1994). It might also be that the use of a new drug indicates failure of a previous drug, or poor tolerability of older drugs, or again that the new drug is indicated in a specific subset of the population with a different baseline risk. Only new users in the same indication and same patient groups are at equal risk of positive or negative events. It might be difficult to find new users in chronic diseases with a low incidence of new cases, where most patients have a long history of previous treatments. Because no randomization is possible in these purely observational studies, much care is taken to ensure comparability of patient populations and avoid confounding. Confounding can be reduced by adjustment methods, taking advantage of these large populations or by matching. When more than two drugs are compared, matching may be more complex, and adjustment preferable, with or without weighing.

3.5 Matching

The comparability of cohort groups may be enhanced by matching on variables known to be associated with the outcome, and with exposure. These variables will be most likely to be confounding variables, or to be associated with confounding variables. A confounding variable is one that is associated both with the exposure and the outcome, and which can explain some or all of the association of the outcome with the exposure. Presently the most extreme matching methods use high-dimensional propensity scores (hdPS), which are determined from several

hundred variables among the thousands present in the datasets. They result in groups that are identical or very close on a large number of variables, including variables that are not part of the hdPS itself (Schneeweiss et al. 2009; Neugebauer et al. 2015; Rassen et al. 2011; Wang et al. 2017; Schneeweiss 2018). Some call these highly matched cohort studies virtual-clinical trials or pseudo-randomized studies, but the absence of real randomization for exposure allocation cannot exclude residual confounding, and makes these studies simply indicative, within a wider context of many studies with different methods and biases.

3.6 Analysis

Typically, analysis of cohort studies is the determination of the relative risk, comparing event rates in exposed and comparator groups:

	Events	No event	All
Treated	a	b	a + b
Control	c	d	c + d

Relative risk is a/(a + b)/c/(c + d)

Variants of this very simplistic approach use time-dependent or survival models that define hazard ratios, based on person-time exposed or followed rather than the absolute per-person event rates above.

In these studies, any comparative analysis would be on treatment, as exposed. Intent to treat (ITT) analyses are justified in clinical trials, where the majority of treatments will be continued until the end of the follow-up, so that any untreated period will be only a small part of overall study time. In observational studies, the duration of treatment is not imposed, and the observation time might be very long, so that most of the observation time may be off treatment, resulting in major unexposed time bias. ITT is therefore essentially meaningless when the initial treatment period is short compared to potentially unlimited observation time. It would also be meaningless when stopping treatment materially alters the outcomes. For example, anticoagulants decrease clotting (thrombotic events) but increase bleeding, whereas stopping them increases the risk of clotting but the risk of bleeding disappears. ITT is also difficult to interpret if the diseases are spontaneously reversible, or if the main objective is the timing of outcomes, such a death in cancer patients.

ITT can however be used when the duration of follow-up is limited to the expected duration of treatment. Most analyses of observational data will be on treatment, comparing event rates while on treatment in exposed patients.

Analysis will commonly use time-dependent variables in survival analysis methods with Kaplan–Meier curves and Cox proportional hazards analyses. When death is a common occurrence that may act as a competing risk (patients who die are no longer at risk of another event), specific analyses such as Fine and Gray competing risk model should be used for the other, non-fatal outcomes (Fine and

Gray 1999). Cohort studies can provide absolute risks and added risks, which may be important for regulatory decisions.

There are many other methodological approaches or considerations that are amply discussed each year during the annual International Conference on PharmacoEpidemiology (ICPE) meeting (www.pharmacoepi.org), such as methods for using big data, or the enrichment of claims data with data from patient data warehouses or new data sources that include not only health expenditures, but also clinical, pathology, societal or genetic information.

3.7 Selection of Participants: Disease-Based

These studies are typically used to describe disease management, and often to prepare for other designs, by specifying the expected event rates in a given disease population that may be the indication for future drugs of interest. Patients are selected on the disease of interest (e.g. diabetes, myocardial infarction or metastatic cancer) to describe disease management and prepare for HTA studies, to model the impact of an as yet unmarketed drug. Disease-based studies are also used in the post-marketing arena to test the impact a new drug or intervention has had on the disease management and cost. The analysis of these studies will be the same as the cohort studies above. One major use of such disease-based cohorts is as a source for nested case–control studies: in contrast with exposure-based studies, where only one or two exposures are studied, in these disease-based cohorts, all health interventions will be included and can be used as potential exposures in nested case–control studies, which will in addition provide information on potential interactions between exposures.

3.8 Selection of Participants: Event-Based

In this approach the event is the main driver, and case-based methods will be applied. These studies are always retrospective, in that the patients are included once the event has occurred and previous exposures are identified. In some variants, the cases and controls are identified within a cohort, in a nested case–control design. A control at a given moment might later become a case if an event occurs. Cases are usually excluded from further studies and are not used as controls.

The general approach is that cases of a given event of interest are identified, and exposures prior to that event are compared to exposures in patients or periods without an event. The comparators might be the patients themselves in case-crossover methods or self-controlled case series; cases may be matched to selected controls in the classical case–control methods, with matching that may be more or less complex including (high-dimensional) disease risk scores, or at the simplest in case-population approaches, which consider the whole source population as the control population. Case-population studies however require identification of all the cases in that population. This might entail events that are easily recognized

and circumscribed to a very specific treatment environment, such as transplantation centres or intensive care units, so that exhaustive identification of all cases in the population can be obtained (Gulmez et al. 2013a). If a sample only is studied, this sample must be representative of the complete case-population. An alternative is the use of whole-population databases where all the cases of a given event may be identified, providing the case specifications are consistent with such identification: for instance, all cases of liver injury or of MI admitted to hospital can be identified in a national healthcare system (Moore et al. 2019). Less severe or serious events that are not hospitalized might not be identified.

Such a method might be useful for surveillance of exposures associated with a given event, for instance hospital admissions for acute liver injury or liver transplantation. Using national claims databases, or national transplantation networks, these events are easy to identify. Because all events in a given territory are captured, one can compare drug exposure in cases to the countrywide exposure to the same drugs (sometimes limited to the age group that might be transplanted), using either person-time or persons (Moore et al. 2013).

The important issue in case-based studies is that the cases and the controls should come from the same cohort (population): this is easy for the self-controlled and population controls, or in the nested case–control design, it might be more difficult for traditional field based case–control studies (Pierfitte et al. 2001).

In most case–control methods, exposures in cases are compared to exposures in controls.

	Cases	Controls
Exposed	a	c
Unexposed	b	d
	a + b	c + d

The usual measure of association here is the odds ratio (ad/bc)

Most commonly exposure in cases and controls is compared to non-exposure (i.e. users to non-users). However, this presupposes that exposure is random and has no link with the event, whereas exposure to drugs is not random, but determined by a disease that causes the prescription and might be associated with the event. For instance, one might expect that patients using NSAIDs have pain or inflammation, more than persons not using these drugs. Pain and inflammation may indicate an underlying disease. One would therefore expect patients using NSAIDS to be sicker and therefore die more than non-users, which indeed is the case (Fosbol et al. 2009). This bias will be common to all drugs that are given to sick patients to treat diseases, especially if they may be associated with the event under consideration (confounding by indication). A related bias is confounding by contraindication, where a drug is avoided in patients at risk of the event of interest (e.g. late at night, passengers will often be more at risk of being drunk, and less at risk of causing an accident as the designated driver who did not drink). When drugs are given to a healthy patient for disease prevention, the comparison of users with non-users may be valid. In other

cases, certainly the use of active controls, drugs with the same indications would be preferable (extreme restriction) (Secrest et al. 2019).

Going from self-controlled case series to case-population is just changing the nature of the controls, from full matching in self-controlled methods to more or less tight matching in the case–control methods, to little or no matching in case-population approaches.

Case-based approaches would be useful in a pharmacovigilance setting when there is a suspicion or signal of the association of a drug with an event, where the first step would be to verify how the association compares to similar drugs with similar indications. Case-based studies and especially the very simple case-population approach may also be used for systematic surveillance of known indicators of drug-related risks, such as the WHO critical terms lists or the more common reason for removing drugs from the market. As a first approximation, these might be liver injury, renal failure, myocardial infarction, sudden death, cytopenia, gastro-intestinal bleeding. Such systematic surveillance might be automated in the future.

3.9 The Comparator

Comparing the use of a drug to non-user is irrelevant in real life. If a drug is used there is a reason, and that reason may be associated with adverse outcomes that will be found only in treated (sick) persons, and not in untreated (healthy) persons. Most drugs when used in sick persons will be associated with a higher risk of disease. For instance, persons identified by the use of low-dose aspirin will have a much higher rate of cardiovascular events than non-users of low-dose aspirin (Duong et al. 2018). Users of NSAIDs will have a higher death rate than non-users, because NSAIDs are not used randomly, but because of pain or inflammation, which are associated with possibly fatal diseases. This is a typical indication bias. Clinical trials will typically use a placebo to negate the indication bias, since all patients have the same initial disease state. In real-life studies, to find patients with similar disease-related risk of events, one must choose comparator drugs with the same indications, i.e. active comparators. This is obviously true for cohort studies, but also in case-based analyses. A comparison with no treatment or untreated periods may simply measure the effect of the indication, not of the exposure. It is therefore imperative to use active comparators in pharmacoepidemiological studies. Ideally a comparator would be another new drug marketed within the same timeframe, and sharing similar pharmacological characteristics (mode of action, target) and indications. Using standard of care as comparator may lead to a biased comparison, where patients put on a new drug because of poor tolerance or lack of efficacy of the standard treatment. At the very least only new users of each not previously exposed to the other should be selected.

3.10 Biases

Biases in pharmacoepidemiology are for the most part common with traditional epidemiology, and can be divided into a few major categories:

1. Selection biases, where the wrong subjects are chosen: e.g. in a case–control study the controls are not from the same population as the cases (e.g. cases are detected in the emergency room and compared to hospitalized controls, or to controls hospitalized for other diseases); in a comparative cohort study one group may consist of incident users of a drug, whereas the other group may consist of prevalent users of the comparator, or again new users of a recently marketed drug may be compared to a historical cohort of patients followed at a time when disease management and outcomes might have been quite different.
2. Ascertainment biases, where the data are not collected in the same way in the different study groups. This might be the fact for historical cohorts, where the data available may vary over time.
3. Analysis biases, including the confounding biases, where the association between exposure and outcomes is in fact related to a third factor that is associated with both the exposure and the outcome.

In pharmacoepidemiology, where the exposure is not random as in clinical trials or externally determined (e.g. place of living) but related to patient conditions and history, there are some very specific biases, such as:

1. The protopathic bias (reverse causality), where the exposure is related to early symptoms of the outcome, rather than the other way around. For instance antibiotics may be prescribed for fever related to undiagnosed agranulocytosis. When agranulocytosis is later diagnosed, it may be mistakenly attributed to the antibiotics. This bias is identified by knowledge of early symptoms and causes of events, and careful determination of the time of onset of the event (index date) as that of the very first symptoms rather than its diagnostic date (Feinstein and Horwitz 1981; Horwitz and Feinstein 1980; Gulmez et al. 2013b).
2. Depletion of susceptibles (or healthy survivor bias), where patients remaining on long-term treatment (prevalent users) have a lower risk of having an event related to the use of the drug than patients initiating the drug. This is one of the reasons that new users designs are preferred (incident users) rather than prevalent users, who have been self-selected for good tolerability or effectiveness of the drug (Moride and Abenhaim 1994).
3. Immortal time bias: In this bias, patients are included in a study arm after having had a period of observations. Only those patients that have not died or had an event during that time are included. If the start of observation time is counted from the initial consideration, then that first time is immortal time. There are many variations on immortal time bias (Suissa 2008; Levesque et al. 2010).

4 Conclusion

Over the last 30 years or so, pharmacoepidemiology has changed considerably, from a field dominated mostly by case–control studies of severe adverse events such as upper gastro-intestinal bleeding with NSAIDs (Henry et al. 1996) or hip fractures

with benzodiazepines (Pierfitte et al. 2001) to routine post-authorization surveillance and assessment of new drugs, made possible by the development of large population databases of electronic health records or claims data, which can cover many millions of patient-lives. These databases allow the description of usage patterns of new drugs as they are marketed, or of older drugs (Duong et al. 2014, 2016).

Pharmacoepidemiology has also contributed to better understanding and quantification of risks initially suspected from clinical trials, such as the cardiovascular risk related to rofecoxib (Graham et al. 2015), or from experimental data, such as bladder cancer with pioglitazone (Neumann et al. 2012). It has confirmed in real life the results of clinical trials in many studies for instance for direct-acting anticoagulants confirming the superior safety to warfarin (Graham et al. 2015) or antiplatelet agents confirming superior effects of ticagrelor to clopidogrel in similar patients (Blin et al. 2017, 2019a). They can also provide data on comparative effectiveness of drugs when no clinical trial exists (Blin et al. 2019b, c, d, e). Pharmacoepidemiological studies can also be used for systematic and comparative risk detection or quantification for selected adverse events such as hepatotoxicity (Gulmez et al. 2013a, 2015; Moore et al. 2019), or myocardial infarction (Duong et al. 2018).

Pharmacoepidemiology studies allow more precise evaluation of the association of exposure to drugs of interest with specific events of interest, using scientifically validated methods. The development of these resources provides considerably more leeway in the exploration of drug-related information, and studies can be tailored exactly to the question raised, and to the precise specifications of each data source. On the other hand, this requires knowledge of each database's specificities, in addition to knowledge of pharmacology and therapeutics. Future developments will include the enrichment of claims or EHR databases with other information such as genetic or imaging, and the use of artificial intelligence or machine learning methods.

However powerful these tools may become, pharmacoepidemiology still requires an understanding of drug characteristics such as drug targets, mechanisms of actions and pharmacokinetics, in addition to drug safety and drug efficacy, and of the underlying diseases and disease characteristics and, of course, the statistical methods needed to explore these data sources, so as to avoid or obviate biases as much as possible.

References

Bezin J, Girodet PO, Rambelomanana S, Touya M, Ferreira P, Gilleron V et al (2015) Choice of ICD-10 codes for the identification of acute coronary syndrome in the French hospitalization database. Fundam Clin Pharmacol 29(6):586–591

Bezin J, Duong M, Lassalle R, Droz C, Pariente A, Blin P et al (2017) The national healthcare system claims databases in France, SNIIRAM and EGB: powerful tools for pharmacoepidemiology. Pharmacoepidemiol Drug Saf 26(8):954–962

Bezin J, Klungel OH, Lassalle R, Dureau-Pournin C, Moore N, Pariente A (2018) Medications recommended for secondary prevention after first acute coronary syndrome: effectiveness of treatment combinations in a real-life setting. Clin Pharmacol Ther 103(6):1038–1046

Blin P, Dureau-Pournin C, Lassalle R, Jove J, Thomas-Delecourt F, Droz-Perroteau C et al (2017) Outcomes in patients after myocardial infarction similar to those of the PEGASUS-TIMI 54 trial: a cohort study in the French national claims database. Br J Clin Pharmacol 83 (9):2056–2065

Blin P, Dureau-Pournin C, Benichou J, Bonello L, Dallongeville J, Danchin N et al (2019a) Secondary prevention of acute coronary events with antiplatelet agents (SPACE-AA): one-year real-world effectiveness and safety cohort study in the French nationwide claims database. Atherosclerosis 281:98–106

Blin P, Dureau-Pournin C, Benichou J, Cottin Y, Mismetti P, Abouelfath A et al (2019b) Comparative real-life effectiveness and safety of dabigatran or rivaroxaban vs. vitamin K antagonists: a high-dimensional propensity score matched new users cohort study in the French National Healthcare Data System SNDS. Am J Cardiovasc Drugs. e-pub ahead of print 2019/06/30; https://doi.org/10.1007/s40256-019-00359-z

Blin P, Dureau-Pournin C, Cottin Y, Benichou J, Mismetti P, Abouelfath A et al (2019c) Effectiveness and safety of 110 or 150 mg dabigatran vs. vitamin K antagonists in nonvalvular atrial fibrillation. Br J Clin Pharmacol 85(2):432–441

Blin P, Fauchier L, Dureau-Pournin C, Sacher F, Dallongeville J, Bernard MA et al (2019d) Effectiveness and safety of rivaroxaban 15 or 20 mg versus vitamin K antagonists in nonvalvular atrial fibrillation. Stroke 50(9):2469–2476

Blin P, Dureau-Pournin C, Cottin Y, Benichou J, Mismetti P, Abouelfath A et al (2019e) Comparative effectiveness and safety of standard or reduced dose dabigatran vs. rivaroxaban in nonvalvular atrial fibrillation. Clin Pharmacol Ther 105(6):1439–1455

Bosco-Levy P, Duret S, Picard F, Dos Santos P, Puymirat E, Gilleron V et al (2019) Diagnostic accuracy of the International Classification of Diseases, Tenth Revision, codes of heart failure in an administrative database. Pharmacoepidemiol Drug Saf 28(2):194–200

Coloma PM, Schuemie MJ, Trifiro G, Gini R, Herings R, Hippisley-Cox J et al (2011) Combining electronic healthcare databases in Europe to allow for large-scale drug safety monitoring: the EU-ADR project. Pharmacoepidemiol Drug Saf 20(1):1–11

Czwikla J, Jobski K, Schink T (2017) The impact of the lookback period and definition of confirmatory events on the identification of incident cancer cases in administrative data. BMC Med Res Methodol 17(1):122

Depont F, Fourrier A, Merliere Y, Droz C, Amouretti M, Begaud B et al (2007a) Channelling of COX-2 inhibitors to patients at higher gastrointestinal risk but not at lower cardiovascular risk: the Cox2 inhibitors and tNSAIDs description of users (CADEUS) study. Pharmacoepidemiol Drug Saf 16(8):891–900

Depont F, Fourrier A, Merliere Y, Droz C, Amouretti M, Begaud B et al (2007b) The CADEUS study: methods and logistics. Pharmacoepidemiol Drug Saf 16(5):571–580

Dhainaut JF, Huot L, Pomar VB, Dubray C, Participants of Round Table of Giens X, contributeurs (2018) Using connected objects in clinical research. Therapie 73(1):53–62

Duong M, Salvo F, Pariente A, Abouelfath A, Lassalle R, Droz C et al (2014) Usage patterns of 'over-the-counter' vs. prescription-strength nonsteroidal anti-inflammatory drugs in France. Br J Clin Pharmacol 77(5):887–895

Duong M, Gulmez SE, Salvo F, Abouelfath A, Lassalle R, Droz C et al (2016) Usage patterns of paracetamol in France. Br J Clin Pharmacol 82(2):498–503

Duong M, Abouelfath A, Lassalle R, Droz C, Blin P, Moore N (2018) Coronary events after dispensing of ibuprofen: a propensity score-matched cohort study versus paracetamol in the French nationwide claims database sample. Drug Saf 41(11):1049–1058

Feinstein AR, Horwitz RI (1981) An algebraic analysis of biases due to exclusion, susceptibility, and protopathic prescription in case-control research. J Chronic Dis 34(8):393–403

Fine JP, Gray RJ (1999) A proportional hazards model for the subdistribution of a competing risk. J Am Stat Assoc 94:496–509

Flynn RW, MacDonald TM, Hapca A, MacKenzie IS, Schembri S (2014) Quantifying the real life risk profile of inhaled corticosteroids in COPD by record linkage analysis. Respir Res 15:141

Fosbol EL, Gislason GH, Jacobsen S, Folke F, Hansen ML, Schramm TK et al (2009) Risk of myocardial infarction and death associated with the use of nonsteroidal anti-inflammatory drugs (NSAIDs) among healthy individuals: a nationwide cohort study. Clin Pharmacol Ther 85 (2):190–197

Fourrier-Reglat A, Lacoin L, Pariente A, Lassalle R, Robinson P, Droz-Perroteau C et al (2010a) When patients report diseases that prescribers seem unaware of: discordance between patient and physician reporting of risk-related previous history in NSAID users from the CADEUS study. Clin Pharmacol Ther 88(5):668–675

Fourrier-Reglat A, Cuong HM, Lassalle R, Depont F, Robinson P, Droz-Perroteau C et al (2010b) Concordance between prescriber- and patient-reported previous medical history and NSAID indication in the CADEUS cohort. Pharmacoepidemiol Drug Saf 19(5):474–481

Fourrier-Reglat A, Noize P, Facon T, Fermand JP, Fitoussi O, Marit G et al (2014a) Real-life patterns of use and effectiveness of bortezomib: the VESUVE cohort study. Leuk Lymphoma 55 (4):848–854

Fourrier-Reglat A, Smith D, Rouyer M, Benichou J, Guimbaud R, Becouarn Y et al (2014b) Survival outcomes of bevacizumab in first-line metastatic colorectal cancer in a real-life setting: results of the ETNA cohort. Target Oncol 9(4):311–319

Garbe E, Kloss S, Suling M, Pigeot I, Schneeweiss S (2013) High-dimensional versus conventional propensity scores in a comparative effectiveness study of coxibs and reduced upper gastrointestinal complications. Eur J Clin Pharmacol 69(3):549–557

Graham DJ, Reichman ME, Wernecke M, Zhang R, Southworth MR, Levenson M et al (2015) Cardiovascular, bleeding, and mortality risks in elderly Medicare patients treated with dabigatran or warfarin for nonvalvular atrial fibrillation. Circulation 131(2):157–164

Guiard E, Lelievre B, Rouyer M, Zerbib F, Diquet B, Megraud F et al (2019) Bismuth concentrations in patients treated in real-life practice with a bismuth subcitrate-metronidazole-tetracycline preparation: the SAPHARY study. Drug Saf 42:993

Gulmez SE, Larrey D, Pageaux GP, Lignot S, Lassalle R, Jove J et al (2013a) Transplantation for acute liver failure in patients exposed to NSAIDs or paracetamol (acetaminophen): the multinational case-population SALT study. Drug Saf 36(2):135–144

Gulmez SE, Larrey D, Pageaux GP, Lignot-Maleyran S, de Vries C, Sturkenboom M et al (2013b) Methodology for a multinational case-population study on liver toxicity risks with NSAIDs: the Study of Acute Liver Transplant (SALT). Eur J Clin Pharmacol 69(3):605–616

Gulmez SE, Larrey D, Pageaux GP, Bernuau J, Bissoli F, Horsmans Y et al (2015) Liver transplant associated with paracetamol overdose: results from the seven-country SALT study. Br J Clin Pharmacol 80(3):599–606

Hallas J (1996) Evidence of depression provoked by cardiovascular medication: a prescription sequence symmetry analysis. Epidemiology 7(5):478–484

Henry D, Lim LL, Garcia Rodriguez LA, Perez Gutthann S, Carson JL, Griffin M et al (1996) Variability in risk of gastrointestinal complications with individual non-steroidal anti-inflammatory drugs: results of a collaborative meta-analysis. BMJ 312(7046):1563–1566

Horwitz RI, Feinstein AR (1980) The problem of "protopathic bias" in case-control studies. Am J Med 68(2):255–258

Idema DL, Wang Y, Biehl M, Horvatovich PL, Hak E (2018) Effect estimate comparison between the prescription sequence symmetry analysis (PSSA) and parallel group study designs: a systematic review. PLoS One 13(12):e0208389

Jick SS, Kaye JA, Vasilakis-Scaramozza C, Garcia Rodriguez LA, Ruigomez A, Meier CR et al (2003) Validity of the general practice research database. Pharmacotherapy 23(5):686–689

Levesque LE, Hanley JA, Kezouh A, Suissa S (2010) Problem of immortal time bias in cohort studies: example using statins for preventing progression of diabetes. BMJ 340:b5087

Macdonald TM, Mackenzie IS, Wei L, Hawkey CJ, Ford I, Collaborators SSG (2013) Methodology of a large prospective, randomised, open, blinded endpoint streamlined safety study of celecoxib versus traditional non-steroidal anti-inflammatory drugs in patients with osteoarthritis or rheumatoid arthritis: protocol of the standard care versus celecoxib outcome trial (SCOT). BMJ Open 3(1)

MacDonald TM, Ford I, Nuki G, Mackenzie IS, De Caterina R, Findlay E et al (2014) Protocol of the Febuxostat versus Allopurinol Streamlined Trial (FAST): a large prospective, randomised, open, blinded endpoint study comparing the cardiovascular safety of allopurinol and febuxostat in the management of symptomatic hyperuricaemia. BMJ Open 4(7):e005354

Mackenzie IS, Ford I, Walker A, Hawkey C, Begg A, Avery A et al (2016) Multicentre, prospective, randomised, open-label, blinded end point trial of the efficacy of allopurinol therapy in improving cardiovascular outcomes in patients with ischaemic heart disease: protocol of the ALL-HEART study. BMJ Open 6(9):e013774

Miranda S, Chaignot C, Collin C, Dray-Spira R, Weill A, Zureik M (2017) Human papillomavirus vaccination and risk of autoimmune diseases: a large cohort study of over 2 million young girls in France. Vaccine 35(36):4761–4768

Moore N, Masson H, Noblet C, Joannidès R (1993) What medicines do patients really take? A comparison of free form vs oriented questionnaires. Post Marketing Surveillance 7:355–362

Moore N, Pierfitte C, Pehourcq F, Lagnaoui R, Begaud B (2001) Comparison of patient questionnaires, medical records, and plasma assays in assessing exposure to benzodiazepines in elderly subjects. Clin Pharmacol Ther 69(6):445–450

Moore N, Gulmez SE, Larrey D, Pageaux GP, Lignot S, Lassalle R et al (2013) Choice of the denominator in case population studies: event rates for registration for liver transplantation after exposure to NSAIDs in the SALT study in France. Pharmacoepidemiol Drug Saf 22(2):160–167

Moore N, Duret S, Grolleau A, Lassalle R, Barbet V, Duong M et al (2019) Previous drug exposure in patients hospitalised for acute liver injury: a case-population study in the French National Healthcare Data System. Drug Saf 42(4):559–572

Moride Y, Abenhaim L (1994) Evidence of the depletion of susceptibles effect in non-experimental pharmacoepidemiologic research. J Clin Epidemiol 47(7):731–737

Neugebauer R, Schmittdiel JA, Zhu Z, Rassen JA, Seeger JD, Schneeweiss S (2015) High-dimensional propensity score algorithm in comparative effectiveness research with time-varying interventions. Stat Med 34(5):753–781

Neumann A, Weill A, Ricordeau P, Fagot JP, Alla F, Allemand H (2012) Pioglitazone and risk of bladder cancer among diabetic patients in France: a population-based cohort study. Diabetologia 55(7):1953–1962

Noize P, Bazin F, Dufouil C, Lechevallier-Michel N, Ancelin ML, Dartigues JF et al (2009) Comparison of health insurance claims and patient interviews in assessing drug use: data from the Three-City (3C) study. Pharmacoepidemiol Drug Saf 18(4):310–319

Noize P, Bazin F, Pariente A, Dufouil C, Ancelin ML, Helmer C et al (2012) Validity of chronic drug exposure presumed from repeated patient interviews varied according to drug class. J Clin Epidemiol 65(10):1061–1068

Noize P, Grelaud A, Bay JO, Chevreau C, Gross-Goupil M, Culine S et al (2017) Real-life patterns of use, safety and effectiveness of sunitinib in first-line therapy of metastatic renal cell carcinoma: the SANTORIN cohort study. Pharmacoepidemiol Drug Saf 26:1561

Petri H, de Vet HC, Naus J, Urquhart J (1988) Prescription sequence analysis: a new and fast method for assessing certain adverse reactions of prescription drugs in large populations. Stat Med 7(11):1171–1175

Pierfitte C, Macouillard G, Thicoipe M, Chaslerie A, Pehourcq F, Aissou M et al (2001) Benzodiazepines and hip fractures in elderly people: case-control study. BMJ 322(7288):704–708

Pladevall-Vila M, Pottegard A, Schink T, Reutfors J, Morros R, Poblador-Plou B et al (2019) Risk of acute liver injury in agomelatine and other antidepressant users in four European countries: a cohort and nested case-control study using automated health data sources. CNS Drugs 33 (4):383–395

Rassen JA, Glynn RJ, Brookhart MA, Schneeweiss S (2011) Covariate selection in high-dimensional propensity score analyses of treatment effects in small samples. Am J Epidemiol 173(12):1404–1413

Rouyer M, Francois E, Cunha AS, Monnereau A, Noize P, Robinson P et al (2018) Effectiveness of cetuximab as first-line therapy for patients with wild-type KRAS and unresectable metastatic colorectal cancer in real-life practice: results of the EREBUS cohort. Clin Colorectal Cancer 17 (2):129–139

Schneeweiss S (2018) Automated data-adaptive analytics for electronic healthcare data to study causal treatment effects. Clin Epidemiol 10:771–788

Schneeweiss S, Rassen JA, Glynn RJ, Avorn J, Mogun H, Brookhart MA (2009) High-dimensional propensity score adjustment in studies of treatment effects using health care claims data. Epidemiology 20(4):512–522

Secrest MH, Platt RW, Dormuth CR, Chateau D, Targownik L, Nie R et al (2019) Extreme restriction design as a method for reducing confounding by indication in pharmacoepidemiologic research. Pharmacoepidemiol Drug Saf. e-pub ahead of print 2019/01/11; https://doi.org/10.1002/pds.4708

Steg PG, Lopez-Sendon J, Lopez de Sa E, Goodman SG, Gore JM, Anderson FA Jr et al (2007) External validity of clinical trials in acute myocardial infarction. Arch Intern Med 167(1):68–73

Suissa S (2008) Immortal time bias in pharmaco-epidemiology. Am J Epidemiol 167(4):492–499

Trifiro G, Pariente A, Coloma PM, Kors JA, Polimeni G, Miremont-Salame G et al (2009) Data mining on electronic health record databases for signal detection in pharmacovigilance: which events to monitor? Pharmacoepidemiol Drug Saf 18(12):1176–1184

Wang SV, Schneeweiss S, Berger ML, Brown J, de Vries F, Douglas I et al (2017) Reporting to improve reproducibility and facilitate validity assessment for healthcare database studies V1.0. Pharmacoepidemiol Drug Saf 26(9):1018–1032

Wentzell N, Schink T, Haug U, Ulrich S, Niemeyer M, Mikolajczyk R (2018) Optimizing an algorithm for the identification and classification of pregnancy outcomes in German claims data. Pharmacoepidemiol Drug Saf 27(9):1005–1010

Why Are New Drugs Expensive and How Can They Stay Affordable?

Basma Hammel and Martin C. Michel

Contents

1 Introduction 454
2 How Rapidly Are Medication Expenses Increasing? 455
3 Increasing Share of Generic Prescriptions 457
4 Why Must Profitability of Pharmaceutical Companies Be High? 458
5 Rising Costs of Drug Discovery and Development 460
6 How Can Pharmaceutical Companies and Societies Deal with Increasing Cost? 461
7 Conclusions 464
References 465

Abstract

Increasing life expectancy leading to a higher median age causes an increasing need for healthcare resources, which is aggravated by an increasing prevalence of preventable diseases such as type 2 diabetes. This includes increasing expenditures for medicines, although these increases when expressed as a share of overall societal wealth are more moderate than often claimed. An increasing use of generic medicines (currently about 90% of all prescriptions) means that costs for discovery and development of innovative drugs must be recovered on a shrinking percentage of prescriptions. However, the key challenge to affordable drugs is exponentially increasing costs to bring a new medicine to the market, which in turn are largely driven by an about 90% attrition rate after start of clinical development. While many factors will be required in concert to keep innovative medicines affordable, reducing attrition appears to be the factor with the greatest

B. Hammel
Institute of Pharmacology, West German Heart and Vascular Center, University of Duisburg-Essen, Essen, Germany

M. C. Michel (✉)
Department of Pharmacology, Johannes Gutenberg University, Mainz, Germany
e-mail: marmiche@uni-mainz.de

J. E. Barrett et al. (eds.), *Concepts and Principles of Pharmacology*,
Handbook of Experimental Pharmacology 260, https://doi.org/10.1007/164_2019_301

potential to contain escalating drug development costs and thereby medication expenditures.

Keywords
Attrition · Drug affordability · Drug pricing · Generic drug prescriptions · Healthcare expenditure · Societal aging

Abbreviations

GDP Gross domestic product
HTA Health technology assessment
R&D Research and development

1 Introduction

After a long period of low numbers of new drug approvals, the number of newly approved drugs has increased again in the past 5 years reaching the highest number in the past 20 years in 2018 (Fig. 1). Although this demonstrates that pharmacology as a discipline and its outcomes in drug discovery and development are alive and productive, the public discussions have been focused less on this success but more on challenges associated with high prices of some of the newly approved drugs and overall cost for medicines.

The development of a novel drug that satisfies the requirements of regulatory authorities is a complex, interdisciplinary effort that can only be handled by dedicated organizations with broad and deep expertise and major financial resources. Theoretically, these could be governmental organizations, nongovernmental nonprofit organizations and for-profit pharmaceutical companies. Governmental organizations including state-owned pharmaceutical companies in socialist countries

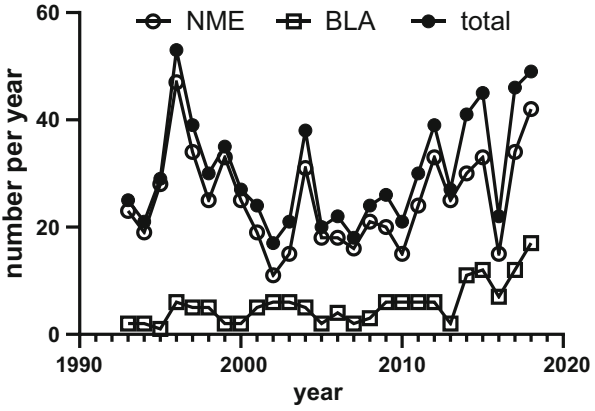


Fig. 1 Development of drug approvals by the US FDA over time for new molecular entities (NME), biologics license applications (BLA), and total values. Based on numbers from Drugs@FDA

have rarely been successful in discovering and developing breakthrough drugs; the anti-malaria drug artemisinin is one of the exceptions, perhaps because it came out of military research (Hsu 2006). There are also examples where large nonprofit organizations such as the Drugs for Neglected Diseases Initiative have successfully developed drugs for primary use in developing countries (Maxmen 2016). However, these examples are few in number, and it remains questionable whether such projects would have satisfied the regulatory authorities in developed countries. This leaves the development of new drugs largely in the hands of for-profit pharmaceutical companies, a business model that has been successful for decades. The term "successful" has a double meaning here: the business model delivered valuable new drugs for the improved treatment of diseases, and it was profitable for the companies developing such drugs. However, this model has been questioned based on increasing expenditures for healthcare in general and medication in particular in societies with increasing average age. Reported profit margins of the overall pharmaceutical industry of 15–20% (Scannell 2015), as well as some very costly medications, have boosted this debate. Against this background, we will explore how strongly novel drugs contribute to overall increases in healthcare expenditure, why newly launched drugs must be pricey, and how innovative medicines can remain affordable at the societal level.

2 How Rapidly Are Medication Expenses Increasing?

It is often claimed that healthcare cost in general and drug prices in particular are "exploding." Such claims are primarily supported by high prices of some newly launched medications such as the antiviral drug sofosbuvir for the treatment of hepatitis C. However, sofosbuvir was found to be cost-effective as compared to previous treatment options as assessed by health technology assessment (HTA) bodies, e.g., National Institute for Health and Care Excellence (NICE) in the UK (Cure et al. 2015), partly because the price for curing one patient has declined by about 1/3 with the introduction of sofosbuvir (Scannell 2015). On the other hand, many widely prescribed and previously costly drugs including the all-time best-selling medications atorvastatin (Lipitor®), clopidogrel (Plavix®), and the fluticasone/salmeterol combination (Seretide®) have gone off-patent and been replaced with much cheaper generics (Kakkar 2015), thereby reducing annual healthcare expenditures by many billions.

To better understand the impact on these counteracting trends, we have previously analyzed the development of medication and overall healthcare expenditure over a 20-year period in Germany based on data from the Organization for Economic Co-operation and Development (Michel 2017). Medication expenditure in Germany increased from 16.9 billion € in 1995 to 33.6 billion € in 2014, i.e., about doubled within a 20-year period. In the same period, overall healthcare expenditure increased from 180 to 322 billion €. While a doubling is impressive, even when occurring over 20 years, overall societal wealth generation as indicated by the gross domestic product (GDP) has increased from 1,899 to 2,916 billion € in that period. When these numbers are put in perspective, medication expenditures increased from 0.89%

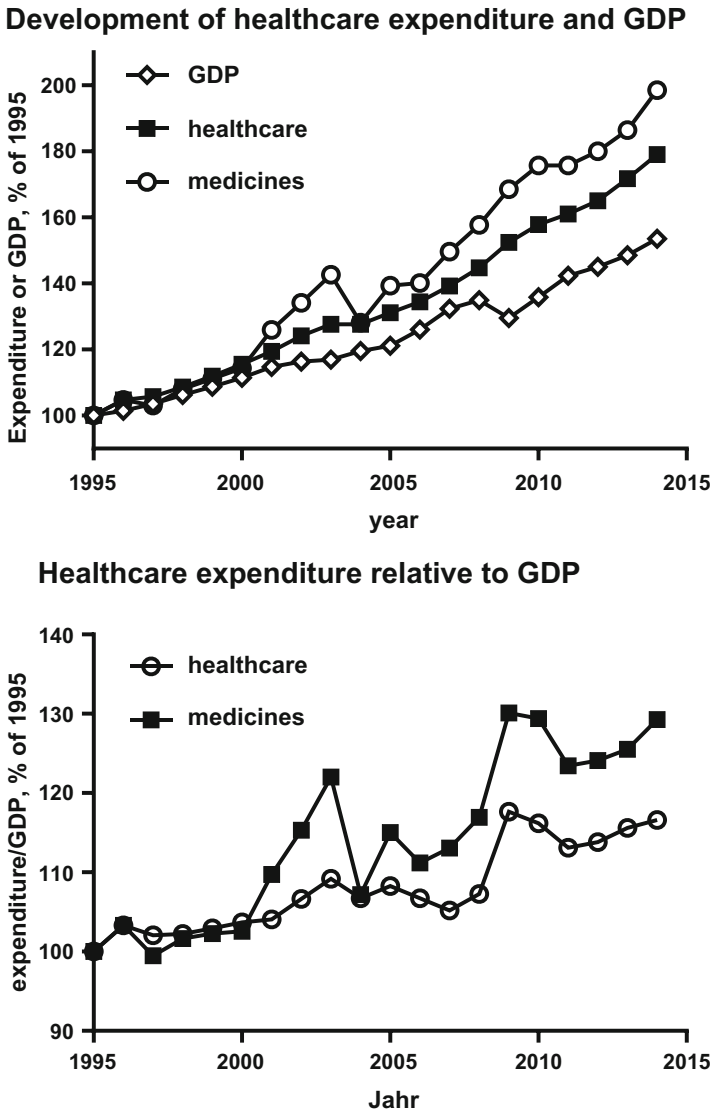


Fig. 2 Development of GDP and of expenditure for healthcare in general and for medicines in Germany 1994–2014 (upper panel) and healthcare and medicines expenditure relative to GDP (lower panel) based on OECD data (www.oecd.org). Adapted with permission from (Michel 2017). OECD data show similar trends for other member countries

in 1995 to 1.15% in 2014 and overall healthcare expenditures from 9.5 to 11.0% (Fig. 2). We feel that expenditure relative to GDP is the key figure that should be considered when discussing societal affordability. While even fractions of a percentage point of GDP are relevant for the economy, these increases are hardly what

would be called an explosion in any other field. Interestingly, related to several regulations on drug pricing, medication expenditure in Germany relative to GDP peaked in 2009 and did not reach those levels again until 2014.

When health insurance representatives and politicians talk about increasing expenditures for medicines, they often imply that this is due to increased cost per dispensed drug. However, this ignores the fact that some medications have replaced costly surgical treatment, for instance, the shift of peptic ulcer treatment from surgery to eradication of *H. pylori* by antibiotics. More importantly, other factors may have led to greater drug utilization and thereby influenced expenditure indirectly. Aging societies are a key factor among those. For instance, the percentage of people aged 65 and older in Germany increased from 22.7% in 1995 to 31.7% in 2014 according to OECD data (Michel 2017). Given this increase in elderly people, an increase of medication expenditures from 0.89% in 1995 to 1.15% in 2014 may actually be seen as surprisingly small.

While no one doubts that increasing expenditures for medications and healthcare in general are a challenge to aging societies, rational strategies for handling this can only be developed based on a sound understanding of why they are rising. We find it surprising that it is rarely voiced in the public debate how age-adjusted medication expenditures are changing over time. However, we see a big caveat: even if rising healthcare expenditures are largely driven by societal aging, societies do not and should not wish to curb aging. If a societal need exists to limit increases in healthcare expenditure, other sources for potential savings must be identified. These could, of course, include changes to the business model how new medications are discovered, developed, and commercialized. An important part of this could be a reduction of attrition rates during development (see below).

3 Increasing Share of Generic Prescriptions

For decades, it had been generally accepted that a pharmaceutical company invents and develops a new medication and, upon regulatory approval, is rewarded for its research and development (R&D) efforts by a limited period of exclusivity for marketing that drug – usually at an attractive price. However, after that period of exclusivity has elapsed, other manufacturers can market the same active compound as a generic without the associated R&D effort as long as they can show bioequivalence of their version of the drug (specific regulations on biosimilars differ somewhat but are similar in principle). As more and more innovative drugs were developed and their patents expired in due time, more and more of them became available as generics. From the viewpoint of pharmaceutical companies, this generated a "patent cliff" because major sources of income largely disappeared, for instance, global blockbuster drugs such as the lipid-lowering agent atorvastatin or the platelet aggregation inhibitor clopidogrel (Kakkar 2015). On the other hand, the market entry of generics can be a blessing for payors because generics and biosimilars typically cost only a fraction of the originator drugs. Thus, it is not surprising that healthcare systems in many countries strongly encourage physicians to preferentially

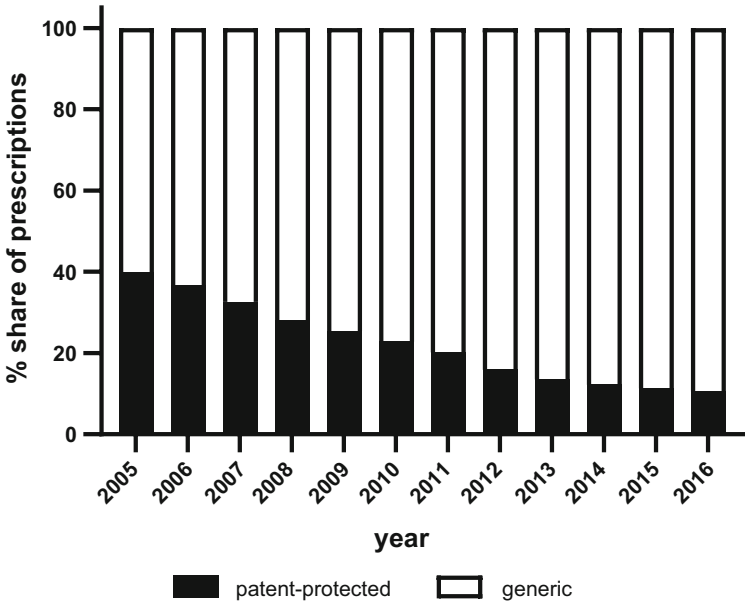


Fig. 3 Share of dispensed prescriptions in USA for patent-protected and generic medications from 2005–2016. Based on data from www.statista.com

prescribe generics – and this has been successful. For instance, only 60.1% of all prescriptions in the USA were filled with generics in 2005, but their share has risen continuously reaching 89.5% in 2016 (last year with available data; Fig. 3). In other words, almost 90% of all prescriptions are now filled with considerably less costly generics in the USA with similar trends across most developed countries. Thus, it could be concluded that innovative drugs have been pricey for a certain period but became very cheap forever after.

The availability of generics has direct implications for the business model of the pharmaceutical industry. If it is assumed that R&D expenditure stays constant (it is not, see below), a given amount of investment could be recovered by 39.9% of all prescriptions in 2005 but had to come from 10.5% in 2016. All other factors assumed to be equal, the average branded prescription in 2016 needed to be four times as expensive as in 2005 to generate the same revenue to support a constant return on R&D investment.

4 Why Must Profitability of Pharmaceutical Companies Be High?

It is estimated that the pharmaceutical industry has a profitability of 15–20% (Scannell 2015). Even when ignoring that such averages include some highly profitable and some less successful companies, this value markedly exceeds that of

most other types of enterprise. Nonetheless, share prices of publicly traded pharmaceutical companies have not been skyrocketing in recent years – even in an investment climate where low or even negative returns on bonds have been and will be around for many years. What is so special about the pharmaceutical industry?

If we decide to buy a car, we can choose between an electric Tesla, a fancy Mercedes, and a more basic Fiat. We can make a choice based on personal needs, preferences, and price; moreover, we probably can negotiate with the car dealer. If we are seriously ill, the situation is different. For many serious diseases, the number of available treatments is limited; in some cases, there may be only one treatment that is available, or one has a so much better overall efficacy/tolerability ratio than the others that it factually becomes the only option. If that is the case and the condition to be treated is serious enough, we probably would pay any price. Moreover, we cannot negotiate the price in the pharmacy as individual patients; however, our health insurance company may already have done that. Thus, we may be grateful for the therapeutic option being made available to us but have limited choice and price negotiation power. The decision to buy a car or get a prescription also differs in another way. As a patient, we need the treatment but are not qualified to choose, and in most developed societies, our health insurance will pay for it. A doctor is qualified to choose the appropriate treatment, but does neither need it nor have to pay for it. The health insurance company must pay but cannot make the treatment choice (although the influence of health insurance companies on the selection of treatments and their influence on pricing are increasing). These factors violate the fundamental assumptions of a free market economy. Nonetheless, drug pricing is based largely on the same principles as other products in a market economy.

The abovementioned considerations have ignored a fundamental player, the investor giving a pharmaceutical company the money to discover and develop new drugs by buying its shares. From an investor perspective, putting money into a pharmaceutical company means investment in a long-term, high-risk project. Projects have a long duration because it typically takes a decade from initial research to regulatory approval and is not substantially faster even in a field such as cancer where many compounds can follow an accelerated approval program (Prasad and Mailankody 2017). They have a high risk because only about 10% of drug candidates entering phase I clinical testing will lead to an approved drug, a phenomenon called attrition (Hay et al. 2014; Wong et al. 2018). The high risk (which also implies a chance of high gains) is also illustrated by the observations that small companies can become very large in a short period of time, for instance, Gilead, whereas large companies may lose leading positions and may become takeover candidates. When each of us decides how to place personal savings, we demand higher returns if the investment is tied up for a longer time and/or has higher risk; that is what investors in pharmaceutical companies do. At fairly stable market capitalization, investors apparently have decided by swarm intelligence that the profitability of 15–20% (Scannell 2015) is what they need to balance long-term, high-risk investment.

5 Rising Costs of Drug Discovery and Development

Based on recent empirical data for anticancer drugs, the median cost to develop a drug up to approvability was 757 million US $ (range: 204–2,602 million US $); taking into account cost of capital, this rises to a median of 794 million US $ (Prasad and Mailankody 2017). R&D costs may differ in other therapeutic areas but probably are in a similar range, except for orphan diseases. Another analysis, using a different approach, yielded much higher costs: when looking at overall R&D expenditure and number of newly approved drugs over a longer period and across multiple large pharmaceutical companies, it has been found that the investment to bring one drug to the market ranged from 3.7 billion US $ (Amgen) to 11.8 billion US $ (AstraZeneca) (Herper 2012). The difference between these values and the 794 million US $ calculated by others (Prasad and Mailankody 2017) is largely explained by the about 90% attrition rate between drugs entering first-in-human phase I studies and regulatory approval since the turn of the century (Hay et al. 2014; Wong et al. 2018). From an investor perspective, it is not the cost behind a single drug development but rather the total investment to bring one drug to the market that counts.

These numbers become even more worrisome if it is considered that the cost of bringing a new drug to the market has increased constantly since 1950 with no indication of a slowdown, even after inflation adjustment (Scannell et al. 2012). Importantly, such increases occur on a logarithmic scale, i.e., continue to double about every 7 years.

The reasons behind these cost escalations are complex. One factor is attrition, which at least in some therapeutic areas such as oncology appears to be higher at present than in the past (Wong et al. 2018). This is in part related to changes of the process how a drug can make it to the market. Regulatory authorities evaluate efficacy, safety, and pharmaceutical quality (the three hurdles), and if all three criteria are met, a drug is approved. The regulatory approval was sufficient to enter the market in the past. However, in many jurisdictions a fourth hurdle has been introduced, HTA. HTA bodies such as NICE in the UK mostly exist as separate institutions outside of regulatory authorities. They explore whether a new treatment is cost-effective compared to already existing treatments. Thus, even the 32nd β-adrenoceptor antagonist on the market[1] could ask for an attractive price in the past, but probably would not pass examination by HTA bodies today. HTA bodies and those in charge of making rules for reimbursements tell pharmaceutical companies that they are primarily interested in breakthrough innovation and much less in incremental improvements of efficacy and tolerability. This may make sense from a societal perspective because societies do not like to pay much higher prices for something that represents only a minor improvement. On the other hand, this

[1]When bisoprolol was launched in Germany, it was advertised with a slogan that even the 32nd β-adrenoceptor antagonist can be important, if it brings a therapeutic advantage. At that time, several other members of this drug class had already been available in generic forms.

approach may backfire in the long run: breakthrough innovation can be translated into a high-risk project. In that sense, a focus on breakthrough innovation/high-risk projects may further increase the rate of attrition and, thereby, total investments to bring one new drug to the market.

A second factor behind escalating R&D costs is an increasing complexity of clinical trials, as discussed elsewhere in this book (Kruizinga et al. 2019). For instance, it was considered sufficient for approval of a new diabetes drug to demonstrate improved glucose homeostasis in a 12-week trial. Today, several trials, including not only placebo but also active controls, with 12-month duration are expected plus a cardiovascular endpoint study, the latter typically involving thousands of patients and a duration of several years (Zinman et al. 2015). Thus, the landmark trial on streptomycin in tuberculosis from 1946 would probably cost less than 100,000 US $ today, whereas the average costs of a contemporary phase II or phase III trial are approximately 8.6 and 21.4 million US $, respectively (Martin et al. 2017). Thus, the real explosion, if any, is in cost for pharmaceutical R&D, not in total drug expenditure.

6 How Can Pharmaceutical Companies and Societies Deal with Increasing Cost?

The above shows three mega-trends: firstly, the increasing median age of developed societies creates an ever-increasing need for overall consumption of medicines; this may be further aggravated by an increasing prevalence of preventable chronic diseases such as type 2 diabetes (World Health Organization 2016). This increasing use of medicines creates a burden to healthcare systems across the globe including many high-income countries, although the increases in societal expenditures for medication and healthcare in general relative to GDP have been less in the past two decades than often assumed (Fig. 2). Second, there is an unbroken, exponential trend for increasing the cost of bringing a new drug to the market with a doubling about every 7 years (Scannell et al. 2012). Even with the incorrect assumption of a stable age structure of societies, providing an attractive return on investment for exponentially increasing R&D costs is unsustainable no matter how wealthy a country is. As investor-financed, for-profit R&D of new medicines has been the only business model providing a stream of innovative new drugs, a point may come in the not too distant future where the ratio between staggering investment and limited returns becomes unattractive to investors, which could lead to a collapse of the present model of developing new treatments to address unmet medical needs. This is aggravated by the third mega-trend, i.e., that an ever-increasing percentage of prescriptions is filled by generic drugs (Fig. 3). Although this acutely helps societies coping with covering the increasing consumption of medicines, it causes an ever-shrinking base of prescriptions from which pharmaceutical companies can generate a return on investment. As many medical needs continue to exist, e.g., in brain-related diseases, oncology, and rare diseases, a scenario can be envisioned where some of these will no longer be tackled by new drug development because the necessary

investment has become unattractive. Therefore, pharmaceutical companies and societies are considering multiple options to cope with the situation – which are not mutually exclusive.

Some of the novel options relate to pricing models. One model currently being discussed is called performance-based or value-based pricing. Under this model, a new medication gains market access at a certain price – but this price is not based on the number of patients receiving the treatment but rather on those in which a pre-specified therapeutic result has been achieved. An example of such discussions is chimeric antigen-receptor T-cell (CAR-T) therapies, for which an original price of more than 1 million US $ per patient was proposed; the claimed justification for this enormous price was a combination of high cost not only for developing such treatments but also for each administration to a patient, and a limited patient base for which this could be prescribed (a rare type of childhood leukemia). Modeling of total healthcare cost for the treatment of such a condition in the UK shows that the question of whether it is cost-effective depends critically on the specific patient selection criteria (Kefalas et al. 2018). While a society may be willing to pay for very expensive treatments if they are lifesaving for a rare and otherwise lethal condition in children, a concern exists about what the implications of the agreed-upon pricing model would be if such treatment is expanded to more common conditions, for instance, bladder cancer. A potential downside of value-based pricing is that it could lead to a situation where a company would ask the maximum price the quality-adjusted life years could justify (Scannell 2015). Another example for innovative pricing models is the world's first subscription-style payment model for innovative antibacterial drugs, which was recently announced by the UK Government and is intended to incentivize pharmaceutical companies to develop new drugs for infections resistant to existing antibiotics (Kmietowicz 2019). Such examples are interesting and may contain expenditures for new medicines in the short run. Moreover, developed societies may become willing to pay a greater share of GDP for healthcare to cope with increasing median age of their populations, but this has limits. However, it appears obvious to us that none of the above measures can keep pace with exponentially increasing cost of bringing a new medicine to the market (Scannell 2015) and the continuing trend that investments in R&D for new medicines need to be recovered on an ever-shrinking percentage of prescriptions for patent-protected drugs. Therefore, we feel that discussions should focus on how the exponential trend for increasing costs of bringing a new medicine to the market can be broken.

One potential source of making drug R&D less costly would be to reduce cost of a given development program, particularly the clinical trials required to obtain regulatory approval. As discussed elsewhere in this book (Kruizinga et al. 2019), the cost of a development program satisfying the standards of major regulatory authorities has increased markedly (see above). In many cases the additional requirements are based on the need to include both genders and all relevant ethnicities and their impact on power calculations for sample sizes; they can also relate to the need for additional safety studies such as dedicated QT studies (to address possible effects of a new medication on heartbeat) or cardiovascular outcome studies for new

antidiabetic medicines. Innovative forms of clinical trials may contribute to a leaner clinical development. The idea of adaptive trial design was developed more than two decades ago (Bauer and Köhne 1994) but only became relevant after regulatory authorities have embraced it (U.S. Food and Drug Administration 2010). However, adaptive trial design is scientifically sound only if specific conditions are met; it is not a tool to "rescue" failed trials but, if anything, requires much greater efforts on study planning (Yildirim et al. 2016). While smart study design could reduce patient numbers to some degree, regulatory authorities and societies at large would be ill advised to contain the cost of clinical development at the risk of being less representative for society at large (as compared to white males only) or of taking shortcuts when it comes to patient safety. Regulatory authorities around the globe are increasingly open to discussions about smarter study design that leads to leaner drug development without putting patients at risk.

While making individual development programs lest costly will be helpful, this approach fails to address the problem that the lion share of cost for bringing a new drug to the market comes from attrition, i.e., the programs that were unsuccessful, particularly during clinical stages of development (Hay et al. 2014; Wong et al. 2018). The trend to reward only breakthrough, but not incremental innovations with attractive pricing makes sense in the short run. However, breakthrough innovation may be seen as synonym for very high risk. This implies a greater chance of failure/attrition; moreover, the higher the risk of an investment, the higher the returns expected from investors (see above). Thus, short-term benefits of only paying handsomely for highly innovative treatments could lead to greater problems in the long run by further increasing the risks inherent in drug R&D. Conceptually a new medicine could fail during development for two reasons: it does not work (or has an unfavorable risk/benefit ratio), and the observed risk/benefit ratio holds insufficient promise for commercial viability. We refer to this as scientific and commercial attrition, respectively. We feel that the potential measures of addressing either appear distinct but may be overlapping.

The earliest step of efficient R&D is the discovery of promising novel molecular targets for the prevention, diagnosis, and treatment of disease. To this end, the pharmaceutical industry is increasingly tapping the pool of innovation coming from academia by means of public-private partnerships (Yildirim et al. 2016). These can occur in multiple constellations ranging from project-specific collaboration between a single lab and a single company to large consortia between multiple companies, academic institutions, and, in some cases, nonprofit organizations including regulatory authorities. An example for the latter, probably the largest public-private partnership related to biomedicine and healthcare on earth, is the Innovative Medicine Initiative in the EU (Laverty and Goldman 2014). While giving the pharmaceutical industry access to the innovation potential from academia, public-private partnerships also help to fund and sustain the development of academic drug discovery capabilities. Not surprisingly, public-private partnerships are increasing in number (Roehrich et al. 2014), supporting how much they are seen as a source of innovation. Partnerships can also occur at the precompetitive level between pharmaceutical companies, for instance, in the development of biomarkers and

companion diagnostics for precision medicine (Fridlyand et al. 2013). An example of this is the TransCelerate initiative in the USA (Gill 2014). Thus, most public-private partnerships focus on a better understanding of disease and the identification of novel molecular targets and biomarkers as well as corresponding companion diagnostics.

Each stage of drug R&D tends to be more expensive than the preceding one. Therefore, an effective curbing of cost related to attrition focuses on the "kill early" concept, i.e., eliminate less-promising programs at the earliest possible stage, which is applicable to both scientific and commercial attrition. Attempts to reduce scientific attrition include the early use of translational models of disease (Michel and Korstanje 2016; Michel et al. 2015; Modjtahedi et al. 2014) and major efforts to develop biomarkers that are predictive for enrichment of suitable patient groups and measurement of target engagement and proxy parameters of efficacy. For instance, programs within the Innovative Medicines Initiative have already identified more than 460 new biomarker candidates and developed or standardized more than 50 new animal models, more than 100 new in vitro models, and more than 100 new in silico models (Yildirim et al. 2016).

Recognizing the impact of attrition on affordability of drugs, regulatory authorities have long departed from evaluating a new drug only once it is submitted for approval for a yes/no decision; rather they encourage pharmaceutical companies to discuss their findings and strategy at various steps of the development process, for instance, start of first-in-human studies and end of phase II studies (European Medicines Agency 2017). Such dialogue de-risks drug R&D by providing clarity on what a new medicine is expected to deliver. While regulatory authorities have a long tradition of communicating their expectations for new treatments in major therapeutic areas, this is a more novel concept for HTA bodies. A pioneer in this field has been the NICE in the UK, which has developed clear rules on how much reimbursement a given new treatment is worth based on quality-adjusted life years. Other HTA bodies become increasingly more open to discuss their requirements with pharmaceutical companies early in the development process; interestingly, initiatives are under way to align the interaction of pharmaceutical companies with regulatory authorities and HTA bodies (European Medicines Agency and eunethta 2019).

7 Conclusions

The combination of increasing median age (i.e., a greater share of the elderly in the overall population) and exponential increases in the cost for bringing a new drug to the market creates challenges for all stakeholders in the healthcare sector. This is aggravated by a (justified!) increasing use of generic medications, which means that escalating drug development costs must be recovered from an ever-shrinking share of prescriptions – currently only about 10%. We feel that no single solution will solve this problem. Rather, many approaches must be combined. Given that an attrition rate of about 90% after start of clinical development (Hay et al. 2014;

Wong et al. 2018) is a key driver of exponentially escalating costs of drug R&D, we see reducing attrition rates despite greater focus on breakthrough innovation as key to slowing the trend for greater costs to bring a new drug to the market. Regulatory authorities and HTA bodies can contribute to reduced attrition by transparently communicating their needs, but the key responsibility for reducing attrition in drug R&D lies within the pharmaceutical industry.

Conflict of Interest BH is an employee of Boehringer Ingelheim; the opinions expressed here are solely those of the authors and not necessarily those of Boehringer Ingelheim. MCM is a past employee of Boehringer Ingelheim and presently an advisor to pharmaceutical companies (Algomedix, Apogepha, Astellas, Dr. Willmar Schwabe, Ferring, NMD, Sanofi, Velicept), venture capital (Inkef), and nonprofit organizations (Fraunhofer Society); he is also a shareholder of Velicept.

References

Bauer P, Köhne K (1994) Evaluation of experiments with adaptive interim analysis. Biometrics 50:1029–1041

Cure S, Guerra I, Dusheiko G (2015) Cost-effectiveness of sofosbuvir for the treatment of chronic hepatitis C-infected patients. J Viral Hepat 22:882–889

European Medicines Agency (2017) European Medicines Agency guidance for applicants seeking scientific advice and protocol assistance. https://www.ema.europa.eu/en/documents/regulatory-procedural-guideline/european-medicines-agency-guidance-applicants-seeking-scientific-advice-protocol-assistance_en.pdf. Accessed 30 July 2019

European Medicines Agency and eunethta (2019) Guidance for parallel consultation. https://www.ema.europa.eu/en/documents/regulatory-procedural-guideline/guidance-parallel-consultation_en.pdf. Accessed 30 July 2019

Fridlyand J, Simon RM, Walrath JC, Roach N, Buller R, Schenkein DP, Flaherty KT, Allen JD, Sigal EV, Scher HI (2013) Considerations for the successful co-development of targeted cancer therapies and companion diagnostics. Nat Rev Drug Discov 12:743–755

Gill D (2014) Re-inventing clinical trials through TransCelerate. Nat Rev Drug Discov 13:787–788

Hay M, Thomas DW, Craighead JL, Economides C, Rosenthal J (2014) Clinical development success rates for investigational drugs. Nat Biotechnol 32:40

Herper M (2012) The truly staggering cost of inventing new drugs. http://onforb.es/yNffHT. Accessed 19 January 2016

Hsu E (2006) Reflections on the 'discovery' of the antimalarial qinghao. Br J Clin Pharmacol 61:666–670

Kakkar AK (2015) Patent cliff mitigation strategies: giving new life to blockbusters. Expert Opin Ther Pat 25:1353–1359

Kefalas P, Ali O, Jørgensen J, Merryfield N, Richardson T, Meads A, Mungapen L, Durdy M (2018) Establishing the cost of implementing a performance-based, managed entry agreement for a hypothetical CAR T-cell therapy. J Mark Access Health Policy 6:1511679

Kmietowicz Z (2019) New antibiotics: NHS will test "pay for usefulness" model to stimulate research. BMJ 366:l4610

Kruizinga M, Stuurman R, Groeneveld GJ, Cohen A (2019) The future of clinical trial design: the transition from hard endpoints to value-based endpoints. Handb Exp Pharmacol. https://doi.org/10.1007/164_2019_302.

Laverty H, Goldman M (2014) The innovative medicines initiative - collaborations are key to innovating R&D processes. Biotechnol J 9:1095–1096

Martin L, Hutchens M, Hawkins C, Radnov A (2017) How much do clinical trials cost? Nat Rev Drug Discov 16:381–382

Maxmen A (2016) Busting the billion-dollar myth: how to slash the cost of drug development. Nature 536:388–390

Michel MC (2017) Sind neue Medikamente zu teuer? MMW Fortschr Med 159:65–67

Michel MC, Korstanje C (2016) ß$_3$-adrenoceptor agonists for overactive bladder syndrome: role of translational pharmacology in a re-positioning drug development project. Pharmacol Ther 159:66–82

Michel MC, Mayoux E, Vallon V (2015) A comprehensive review of the pharmacodynamics of the SGLT2 inhibitor empagliflozin in animals and humans. Naunyn Schmiedeberg's Arch Pharmacol 388:801–816

Modjtahedi H, Cho BC, Michel MC, Solca F (2014) A comprehensive review of the preclinical efficacy profile of the ErbB family blocker afatinib in cancer. Naunyn Schmiedeberg's Arch Pharmacol 387:505–521

Prasad V, Mailankody S (2017) Research and development spending to bring a single cancer drug to market and revenues after approval. JAMA Intern Med 177:1569–1575

Roehrich JK, Lewis MA, George G (2014) Are public-private partnerships a healthy option? A systematic literature review. Soc Sci Med 113:110–119

Scannell JW (2015) Four reasons drugs are expensive, of which two are false. http://onforb.es/1ZdvQbc. Accessed 14 Oct 2015

Scannell JW, Blanckley A, Boldon H, Warrington B (2012) Diagnosing the decline in pharmaceutical R&D efficiency. Nat Rev Drug Discov 11:191–200

U.S. Food and Drug Administration (2010) Guidance for industry: adaptive design clinical trials for drugs and biologics. http://www.fda.gov/downloads/Drugs/Guidances/UCM201790.pdf. Accessed 25 July 2016

Wong CH, Siah KW, Lo AW (2018) Estimation of clinical trial success rates and related parameters. Biostatistics 20:273–286

World Health Organization (2016) Diabetes fact sheet. http://www.who.int/mediacentre/factsheets/fs312/en/. Accessed 21 July 2017

Yildirim O, Gottwald M, Schüler P, Michel MC (2016) Opportunities and challenges for drug development: public-private partnerships, adaptive drug design and big data. Front Pharmacol 7:461

Zinman B, Wanner C, Lachin JM, Fitchett D, Bluhmki E, Hantel S, Mattheus M, Devins T, Johansen OE, Woerle HJ, Broedl UC, Inzucchi SE (2015) Empagliflozin, cardiovascular outcomes and mortality in type 2 diabetes. N Engl J Med 373:2117–2128