Defining and redefining its role in biology: Synthetic biology as an emerging field at the interface of engineering and biology

Wenfa Ng

Novena, Singapore, Email: ngwenfa771@hotmail.com

Abstract

Synthetic biology is often misunderstood as creation of artificial life or new biology using principles different from those of extant organisms around us. But, fundamentally, the field is about engineering biology in a more efficient and effective way, and endowing new functions in existing organisms using a more refined and predictable approach. Thus, synthetic biology as encapsulated by the field it helps defined, is enhanced recombinant DNA technology, an example of which is modular and orthogonal "standard swappable biological parts". But, as the field grows and matures, various "allied" fields are subsumed into it such as metabolic engineering, protein engineering, directed evolution, origins of life research, and systems biology, which in totality represents a new perspective of how engineering principles can be utilized to expand, in scope and depth, the realms of questions that biology ask. Two parallel approaches, directed evolution and *de novo* protein design, are frequently used to engineer new phenotypes into organisms. Similar to evolution but with purposeful use of selection pressure to elicit progressive refinement of specific traits in an efficient manner, directed evolution is a powerful methodology that generates, at the cell level, libraries of mutants of slightly different function such as differing resistance to heavy metals, that upon exertion of continued selection pressure, led to the evolution of a strain capable of thriving under a hostile environment previously inhabitable to the organism. Taking a different approach, *de novo* protein design taps on advances in biomolecule structure modeling together with bioinformatic sequence search for inserting, in a structure defined manner, specific amino acids (natural or unnatural) in a protein structure to endow desired functionality, where one highly sought function is catalysis of unnatural reactions such as the Diels-Alder reaction. Long chain length DNA synthesis, on the other hand, finds utility in enabling the synthesis of a minimal genome for a bacterium, which demonstrates the huge possibilities of having a microbe with an optimized genome (free of extraneous genes) for biotechnological applications in delivering drugs and fuel at high titer with lower cost. Having assimilated other fields, synthetic biology is again redefining its role as its seeks to use, in an ethical and responsible manner, a new way of adding new functions into organisms through genome editing. For example, CRISPR/Cas9 genome editing holds enormous potential for providing life saving gene editing capability in medical treatments, while enabling fast, easy removal of undesirable genes and prophages from a production microorganism. Synthetic biologists are asking themselves deep questions on how best to regulate this powerful

technology that could be as impactful on science and human society as recombinant DNA technology was in 1973.

Keywords: modular genetic parts, metabolic engineering, protein engineering, directed evolution, *de novo* protein design, metabolic flux analysis, systems biology, recombinant DNA technology, DNA synthesis, minimal cells,

Subject areas: chemical engineering, bioengineering, biochemistry, cell biology, molecular biology,

Extended abstract

Genetic crosses, whether of plants or animals, represent the first steps human take to improve traits of food sources as population growth increases demand for stable and more nutritious food supply. Before the advent of recombinant DNA technology, the time-consuming and tedious process of cross-breeding remains the *de facto* method for expanding the genetic repertoire of food crops and animal proteins available to humans. But, various incarnations of cross breeding occur naturally, particularly in the microbial world, where transformation, conjugation and transduction move small segments of genetic material encoding new traits and functions between microorganisms. Most prevalent in bacteria, but not limited to the prokaryotic branch of the tree of life, DNA transfer between microbes was tapped on in the first demonstration of how foreign genes could be introduced into a cell as a plasmid, and later, more stably incorporated into the host's genome through integration of foreign genes into a chromosome. Though reliable and available in the kit format in modern molecular biology manipulations, traditional articulation of recombinant DNA technology remains laborious, and most importantly, is limited to moving one or a few genes, which restricts the speed and extent in which the genetic constituents of a cell could be altered for producing a fuel molecule that may require entire pathways to be introduced for its heterologous production. To make biology easier to engineer, a relatively new field of synthetic biology aims to create "standard biological parts" which could be "connected" in different ways to endow new functions to organisms at higher speed with less trial and error experimentation. Orthogonal in function and modular in genetic part design, standard biological parts nevertheless suffer from emergent properties where the introduced parts participate in cross-talks with host cellular processes in unexpected ways. Though deficient in significantly reducing gene expression noise, these standard parts may be the way forward as we seek to build minimal genomes and cells through synthesizing entire chromosomes with standard biological parts incorporated using a bottom-up approach. In parallel to, but alternative from biological parts design, directed evolution aims to endow extant organisms with novel functions through an iterative process of random mutagenesis and exertion of selection pressure with cellular processes as readout. Enabled by high throughput screening technology capable of rapidly accessing every point of the sequence space produced by mutagenesis experiments, directed

evolution has potentiated success in generating improved enzymes for catalysis. More important than improving existing enzymes, directed evolution is one of two key pillars of protein engineering efforts seeking to confer novel functions such as silicon-carbon bond formation and Diels-Alder reactions to structurally evolved motifs of existing enzyme templates. The alternative approach of *de novo* protein design, on the other hand, uses computational biology and structural biology tools to design an enzyme with active site capable of performing a hitherto non biological reaction. Using DNA synthesis as a tool, the desired changes in nucleotide sequence and, by extension, amino acid secondary and tertiary structures could be realized. But, what happens at the protein and enzyme level does not necessarily speak of cellular function. Armed with mathematical tools that cross check steady state processes of cells with the dynamic, systems biology approaches have been used to gain a holistic understanding of cellular processes, or manifestations resultant from genetic engineering where, for example, metabolic flux analysis identifies enzymatic choke points where flux could be directed to increase the production titer of specific metabolic precursors. Nevertheless, the greatest challenge of synthetic biology in the future may lie in how the nascent field of genome editing could be usefully integrated into synthetic biology's armamentarium without incurring significant ethical concerns. With tools such as CRISPR/Cas9 capable of high precision editing of large number of genes with similar genetic identification tags, genome editing potentially opens up a Pandora's box of ethical and safety issues concerning its use. Beneficial use such as CRISPR/Cas9 inactivation of prophages in recombinant hosts that led to a more stable and productive organism for industrial use is one exciting possibility of how improved genome editing technology could lend utility to humanity. A field that is still learning and defining its role for biology, synthetic biology owns its roots to genetic engineering, but which seeks to understand cellular biology in totality from the bottom-up.

Perspective

When it was first demonstrated in 1973, genetic engineering or recombinant DNA technology was much worried as it holds potential for altering the genetic makeup of organisms, and thus, may be the trigger for infectious disease spread or release of dangerous microorganisms from labs that research on genetic engineering. Many decades on, the fear has subsided as the public gradually understood the technology and the many positive benefits it brings. Dangers of microbes picking up new genetic repertoire that may endanger the ecosystem are largely in check due to natural constraints on the transmissibility of non-essential traits.

Fast forward to the early 2000s, a new research area of synthetic biology gradually coalesce into an emerging field still in development and definition. Specifically, the desire to tinker with genes on a larger scale compared to the one gene at a time approach in established recombinant DNA

technology led to ideas of bringing an engineering approach to biological research. For example, synthetic biologists aim to alter specific genes in the genome with high specificity and fidelity; thereby, enabling desired traits to be conferred to particular microorganisms with greater ease and speed.

More importantly, the conventional approach to genetic engineering is laborious and time-consuming, where multiple iterations of adjusting gene dosage and type of promoters are necessary for heterologous DNA to be expressed in a recombinant organism. Trying to do much better than the status quo, synthetic biologists aim to create "standard biological parts" comprising modular genetic units that could be easily transferred into microorganisms for the expression of new proteins. Going beyond single genes, the longer term goal would be the facile transfer and integration of entire segments of genes into a host organism for the reconstitution of a missing metabolic or signaling pathway or expression of new functions.

But, synthetic biology is larger than genetic engineering, its most popular definition describes its role as a broad field that aims to use enhanced tools of genetic engineering and high throughput screening for conferring new and beneficial functions to organisms. Articulation of this goal would naturally mean that desired mutations must be reliably engineered into the genome. Thus, comes the enabling technology of DNA synthesis that facilitates the design of specific nucleotide sequences with point mutations precisely inserted for understanding the role of specific changes in DNA on protein function. This is a step change from the use of error prone polymerase chain reaction (error prone PCR) for random insertion of mutations in a specific stretch of DNA in terms of specificity, speed, and reliability, and with the help of DNA synthesis companies synthesizing specific DNA fragments on order, greatly accelerates genetic engineering research.

With the ability to expand the sequence search space for new functions comes the problem of identifying the phenotype of interest, to which high throughput screening approaches helps to address. Specifically, fluorescence tagged molecules expressed when certain conditions are met in the cell or more traditional biochemical assays can be used for screening desired traits conferred by sought after mutations that, for example, enable the formation of novel types of chemical bonds such as that between silicon and carbon.[1] Enlarged to a larger scale where thousands of samples can be sampled and analyzed by robotic means, high throughput screening technology searches the sequence space generated by random mutagenesis that, hopefully, would yield the desired function.

4

Random generation of mutations by chemical means, however, may not be sufficient for pointing the development of a sequence of mutations that change the function of a protein. To this end, the approach of directed evolution[2] find use in exerting selection pressure on a microbial species for evolving successively better solutions to an environmental stressor or existential threat. Usually a chance occurrence amidst a sea of thousands or millions of mutations requiring examination, high throughput screening find ready use for automating and reducing the search time for profiling the desired trait. The key in facilitating the search for the target mutation is in developing a simple yes or no response that either could be colour-coded or encapsulated in a growth response. Like cells that grow giving a "yes" answer, a high throughput screen based on a similar concept would enable a rapid targeted search of a vast genetic space.

Desire to engineer biology for beneficial uses such as producing fuel molecules from sunlight capture meant that multiple traits must coexist in a cell to enable the complex sequence of chemical transformation to occur sequentially *in situ* for the desired outcome to manifest at the population level. But, how do we fathom about the complexity of biological networks that interconnect pathways using conceptual tools that usually apply to single enzyme cascade. Enter the conceptual engine of systems biology that seeks to understand, holistically, the complicated interplay between proteins and genes as well as metabolic flows in channeling biological building blocks into energy powering movement, biological computing and chemical decision making. Beyond understanding, the mathematical tools developed for systems biology to glean aggregate information from molecular processes could also be used, in reverse, to develop strategies for facilitating enhanced production of specific metabolic precursors for final drug production in a microbe. Known broadly as metabolic engineering[3] [4] where metabolic flux analysis identifies the crucial nodes and proteins that are choke points of metabolism in a native host or recombinant organism, modulation of gene dosage or promoter strength are common tools for channeling metabolite flux through one pathway over another that, phenotypically, translates into production of a fuel molecule or enzymatically decorated glyosidic drug molecule functional in treating human diseases.

Moving forward, genome editing[5] [6] provides an alternative view of genetic engineering different from that of the additive recombinant DNA technology where new genes (and usually functions) are added to organisms. Specifically, genome editing seeks to identify and reproducibly alter a specific gene for gaining or losing a function.[7] For example, the high specificity and fidelity tool of CRISPR/Cas9 (clustered regularly interspersed short palindromic repeats/Cas 9) approach is able, with the help of a guide RNA, locate a specific stretch of target DNA (complementary to the guide RNA sequence) and perform molecular cleavage with few nucleotide resolution.[8] Doing so, a defective gene could be removed or inactivated, while combining with other approaches, a new gene could be introduced.[9] Useful for en masse removal of prophages from a

5

genome, the technique has also been used for editing specific non desirable traits in embryos, though the latter use raise serious ethical considerations.

Thus, one and a half decade on, the field of synthetic biology remains emerging and is continuously defining its role in biological research. Thrusted with the objective of making genetic engineering faster and with higher fidelity and able to orthogonally introduce new traits into established hosts without upsetting cellular processes, standard biological parts represent a first foray into making biology engineerable like an electronic circuit with swappable parts. On the other hand, new traits are left to evolve on its own in cells put through successive rounds of selection pressure in an approach known as directed evolution. Observed through high throughput screens able to quickly search a large library of mutations covering a big sequence space, directed evolution allows mutations to be selected by natural selection to survive under specific external pressure. This is diametrically different from the *de novo* protein design approach where specific mutations are designed into a sequence for conferring a desired point mutation.[10] [11] [12] But, both approaches are useful for introducing specific mutated genes into organisms that could help transform a cellular metabolite into desired compounds after the metabolic circuitry has been examined under metabolic flux analysis and other system biology approaches; thus, filling a gap in natural catalytic wizardry. The future holds real promise on the use of DNA synthesis technology for the creation of large genomes for examining the genetic and biological basis of minimal cells in understanding the origins of life, or on a more practical note, improving the efficiency and effectiveness of drug and fuel production through processes encapsulated in a minimal but optimized microbial genome. However, advent of genome editing, while promising for reproducible alteration of large numbers of genes sprouting the same genetic identification tag as a guide RNA, is also fraught with dangers of misuse for designing embryos with target traits.

References

1. Kan, S. B. J., Lewis, R. D., Chen, K. & Arnold, F. H. Directed evolution of cytochrome c for carbon–silicon bond formation: Bringing silicon to life. *Science* **354,** 1048 (2016).

2. Packer, M. S. & Liu, D. R. Methods for the directed evolution of proteins. *Nat Rev Genet* **16,** 379–394 (2015).

3. Yadav, V. G. Biosynthonics: Charting the Future Role of Biocatalysis and Metabolic Engineering in Drug Discovery. *Ind. Eng. Chem. Res.* **53,** 18597–18610 (2014).

6

4. Khosla, C. & Keasling, J. D. Metabolic engineering for drug discovery and development. *Nat Rev Drug Discov* **2,** 1019–1025 (2003).

5. Yang, L. *et al.* Engineering and optimising deaminase fusions for genome editing. *Nat. Commun.* **7,** 13330 (2016).

6. Standage-Beier, K., Zhang, Q. & Wang, X. Targeted Large-Scale Deletion of Bacterial Genomes Using CRISPR-Nickases. *ACS Synth. Biol.* **4,** 1217–1225 (2015).

7. Tsarmpopoulos, I. *et al.* In-Yeast Engineering of a Bacterial Genome Using CRISPR/Cas9. *ACS Synth. Biol.* **5,** 104–109 (2016).

8. Nuñez, J. K., Harrington, L. B. & Doudna, J. A. Chemical and Biophysical Modulation of Cas9 for Tunable Genome Engineering. *ACS Chem. Biol.* **11,** 681–688 (2016).

9. Hsu, P. D., Lander, E. S. & Zhang, F. Development and Applications of CRISPR-Cas9 for Genome Engineering. *Cell* **157,** 1262–1278

10. Tinberg, C. E. *et al.* Computational design of ligand-binding proteins with high affinity and selectivity. *Nature* **501,** 212–216 (2013).

11. Burton, A. J., Thomson, A. R., Dawson, W. M., Brady, R. L. & Woolfson, D. N. Installing hydrolytic activity into a completely de novo protein framework. *Nat Chem* **8,** 837–844 (2016).

12. Baker, M. Protein engineering: navigating between chance and reason. *Nat Meth* **8,** 623–626 (2011).

Conflicts of interest

The author declares no conflicts of interest.

Author's contribution

The author wrote the manuscript.

Funding

No funding was used in this work.

8