RCSB Protein Data Bank Advisory Committee

Meeting and Teleconference

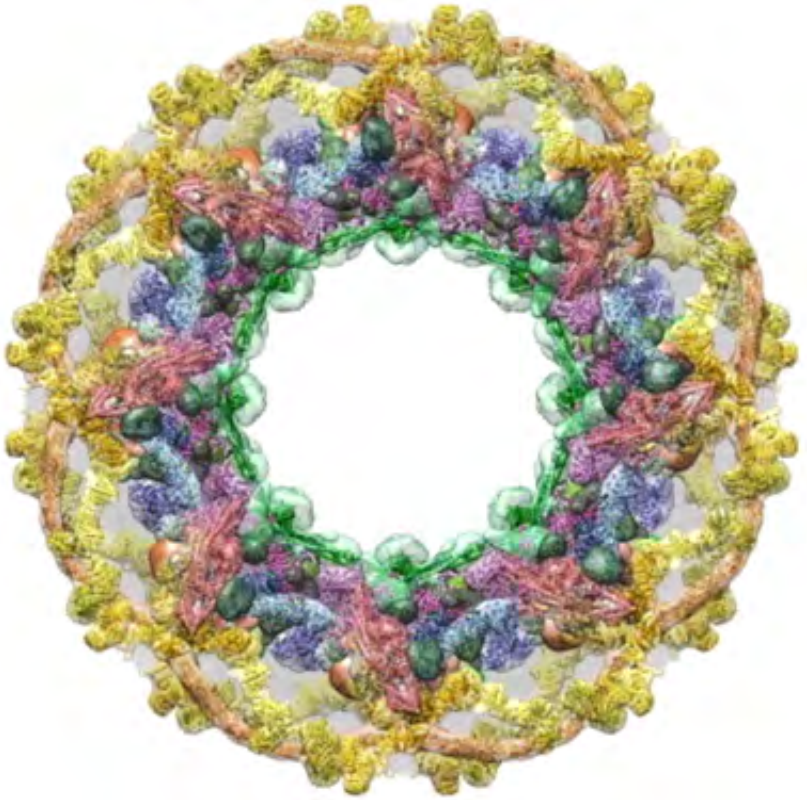
Wednesday May 8, 2019

State of the RCSB PDB

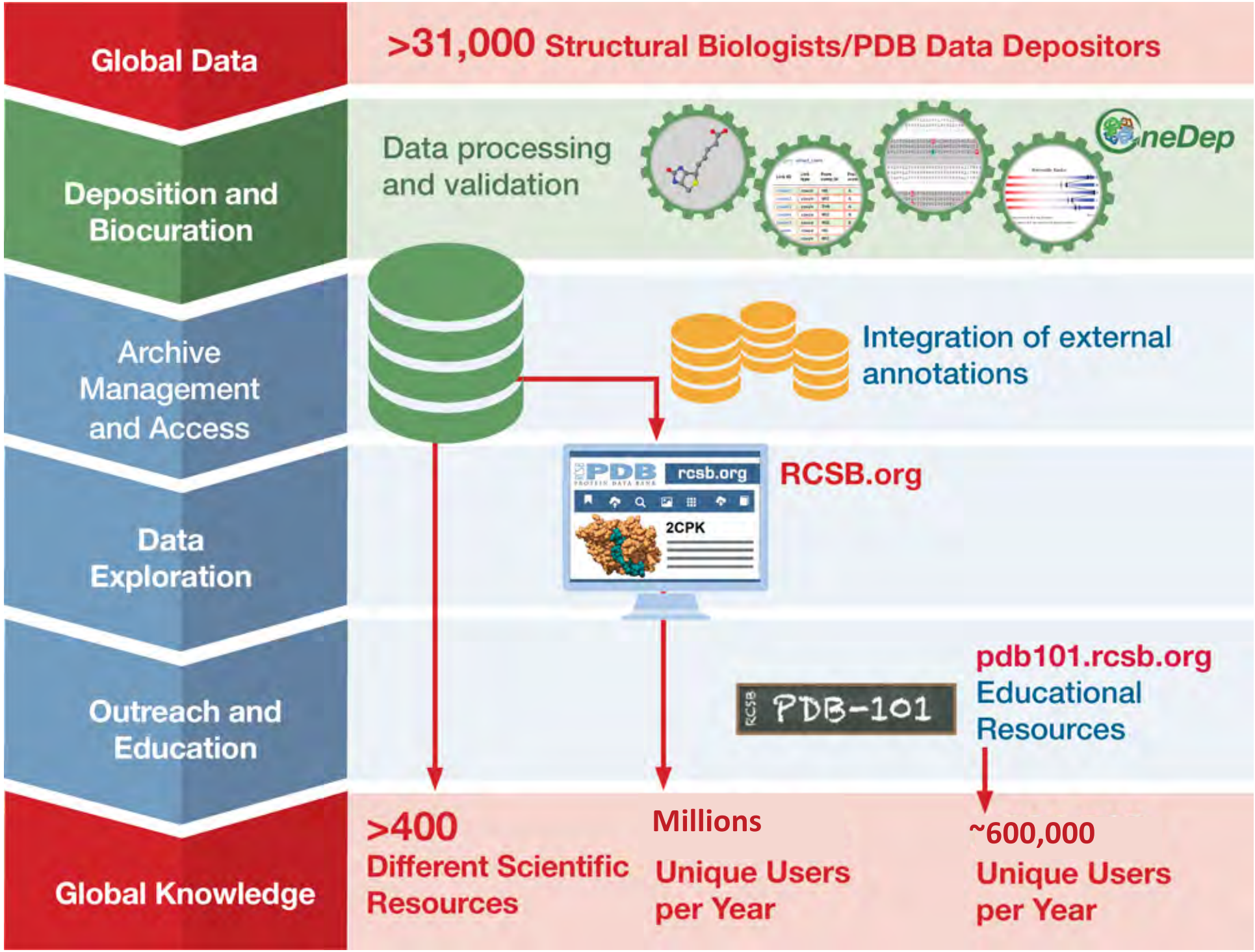
2019-2023: Meeting the Challenges Ahead

Structural biology is evolving

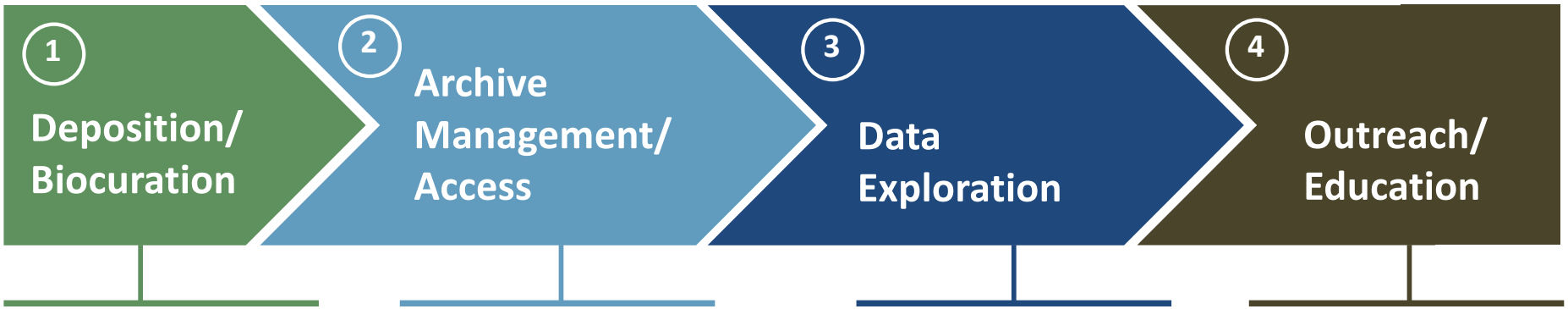
1. Growth/Complexity
2. Evolving Experimental Methods (SFX/XFEL, 3DEM)
3. Emerging Integrative/Hybrid Methods (I/HM)



I/H Methods Structures
552-protein yeast Nuclear Pore Complex
Kim et al. (2018) *Nature 555*, 475-82
PDBDEV_00000010; PDBDEV_00000011; PDBDEV_00000012


Global Data
>31,000 Structural Biologists/PDB Data Depositors
Deposition and Biocuration
Data processing and validation
neDep
Archive Management and Access
Integration of external annotations
Data Exploration
PDB
rcsb.org
2CPK
RCSB.org
Outreach and Education
PDB-101
pdb101.rcsb.org
Educational Resources
Global Knowledge
>400
Different Scientific Resources
Millions
Unique Users per Year
~600,000
Unique Users per Year

RCSB PDB: Four Interoperating Services



- Deposition
- Validation
- Biocuration

- Data standards
- Data integration
- Data storage
- Data access

- Portal
- Search
- Browse
- 3D visualization

- PDB-101

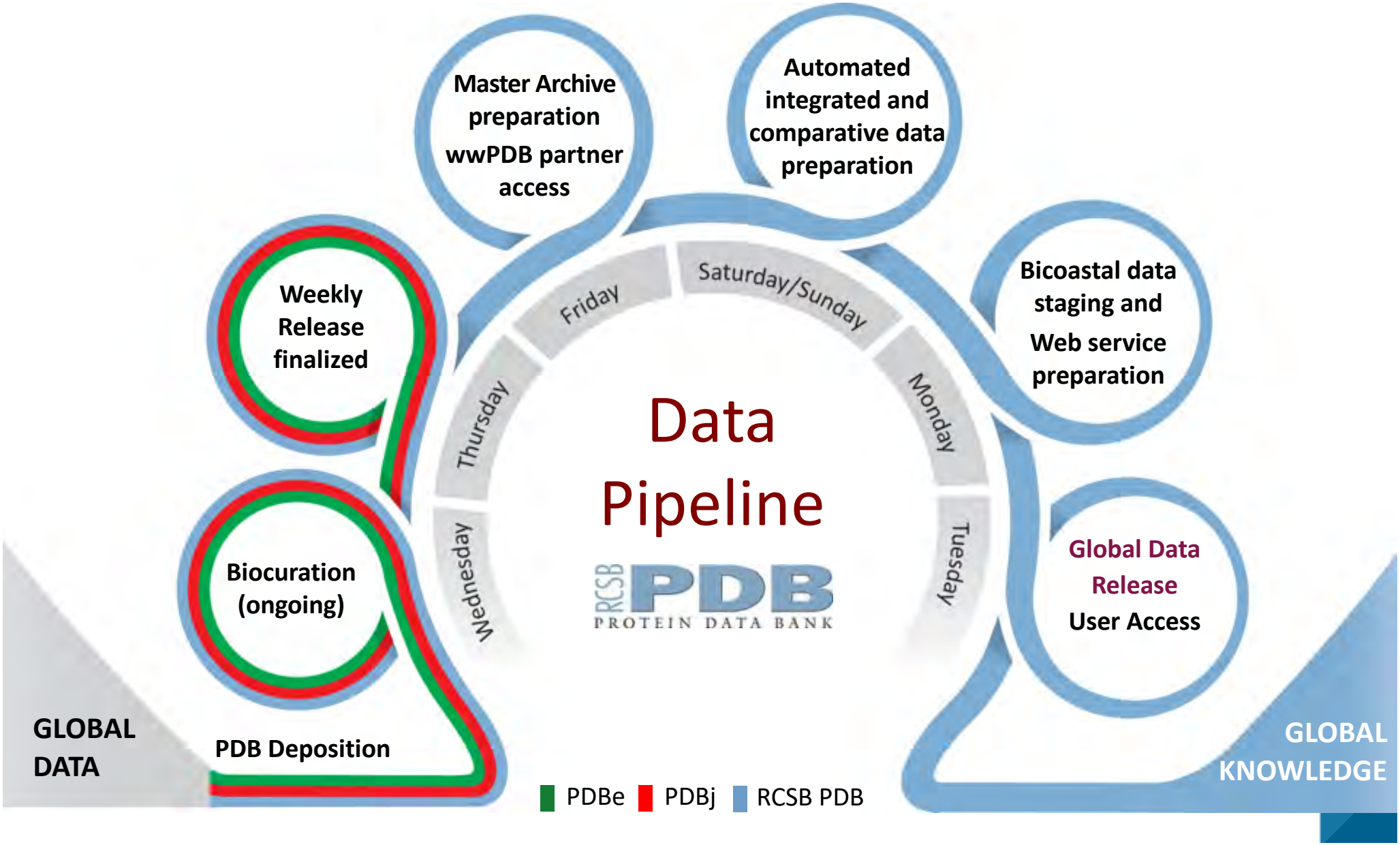
Customer Service Help Desk and IT Support

RCSB PDB Data Pipeline Assures Adherence to the *FAIR* Principles

1. **Deposition/Biocuration** supporting *Data Depositors* through deposition, validation, and biocuration. Data are well-curated and validated for scientific/technical accuracy. **(*FAIR*)**
2. **Archive Management/Access** supporting *Data Consumers* by maintaining the PDB archive and data standards, enabling global data delivery, and integrating PDB data with other data resources. **(*FAIR*)**
3. **Data Exploration** supporting *Data Consumers* through open-access tools for structure query, visualization, and analysis. **(*FAIR*)**
4. **Outreach/Education Services** support educators, students, and the general public *via* PDB-101 website. **(*FAR*)**

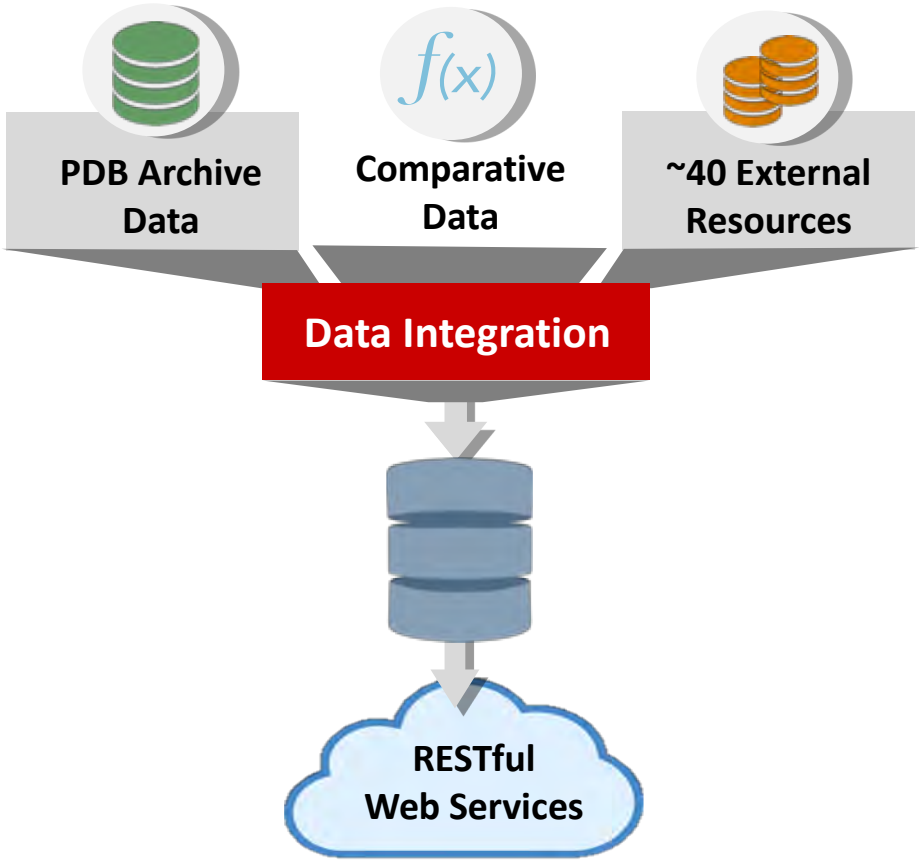
Customer Service and IT Support underpin all services

>200 New Structures Released Each Week



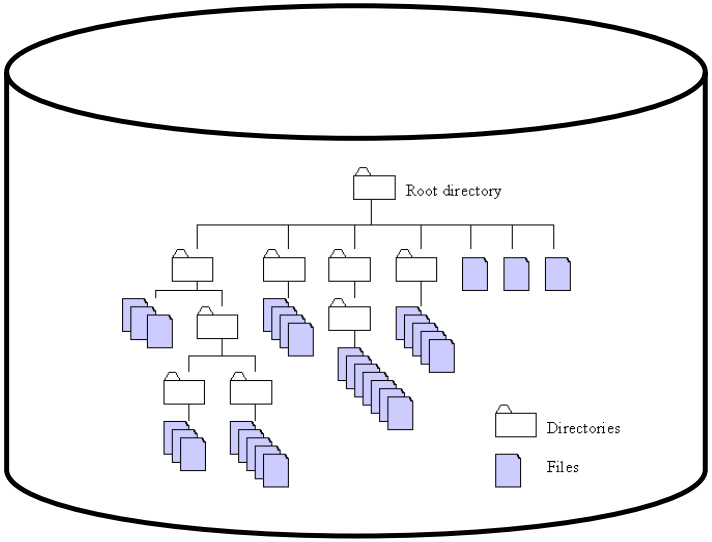
Archive Management/Access Ensures *FAIR* PDB

- Package PDB data for release
- Maintain PDB Data Standards and conduct archive-wide standardization
- Compute comparative data to support search applications
- Integrate data from across the Life Sciences ecosystem
- Support programmatic access



Updating the PDB Master Archive

- Assemble weekly data from wwPDB partners
- Package final archival data files, validation reports, reference dictionaries, and supporting data files (1GB)
- Authoritative Master Archive readied for delivery
  - Traditional archive layout
  - Versioned archive layout
- Exported to wwPDB partners

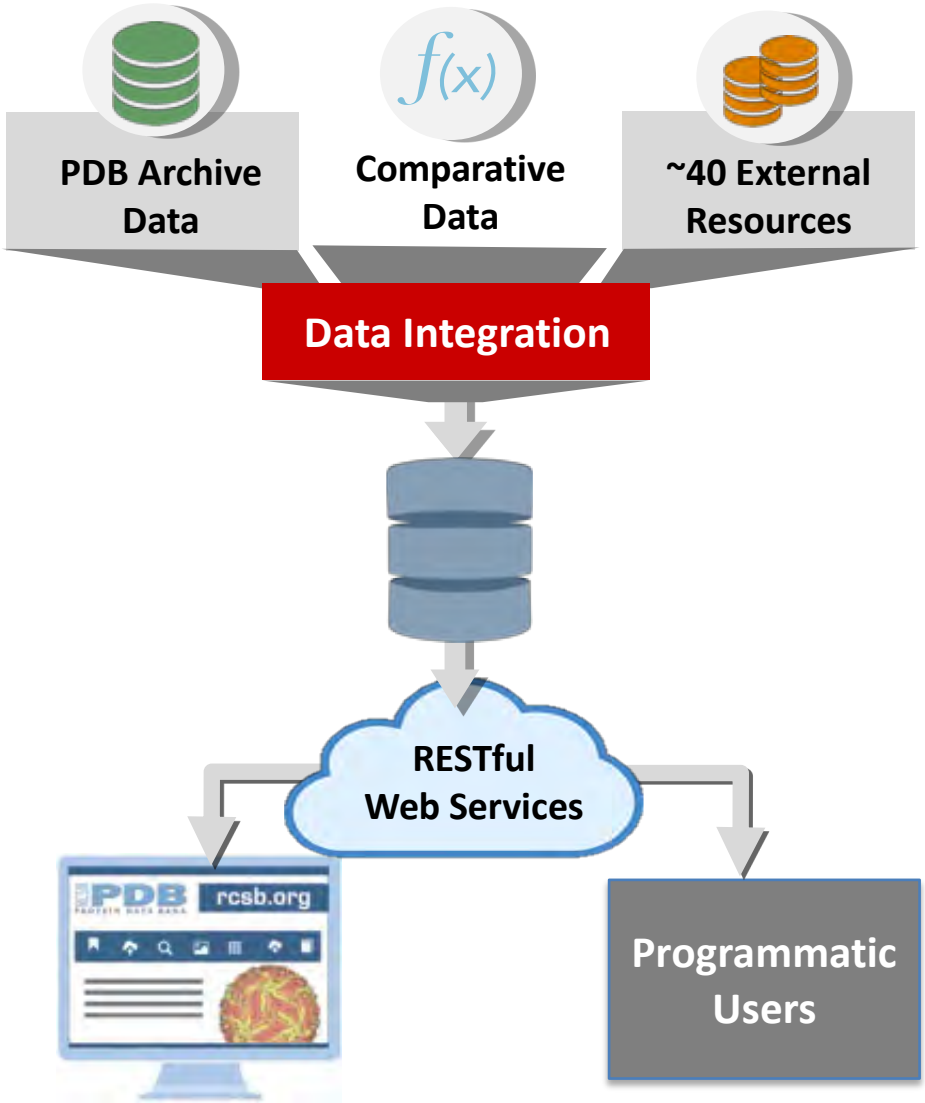


2018 Annual Snapshot
January 1, 2019

- 147,610 structures
- >1.5 TB data
- ~1.9 million unique archive files
- >1 billion 3D atomic coordinates
- Online annual and milestone ftp archive snapshots from 2005

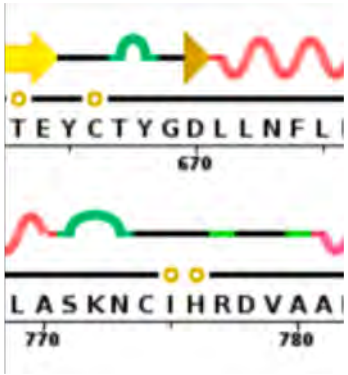
Comparative and Integrated Data for Contextual Views

- Comparative data
  - Sequence clustering
  - 3D structure clustering
- Leverage cyberinfrastructure (CI) data from ~40 key life science resources
  - Diffraction data (ProteinDiffraction.org, SBGrid, Store.Synchrotron Data Store)
  - DrugBank
  - NCBI
  - Gene Ontology (GO)
  - Sequence (UniProt, SIFTS/PDBe)
  - SCOP and CATH
- Leverage CI computing from DIBBS and Open Science Grid

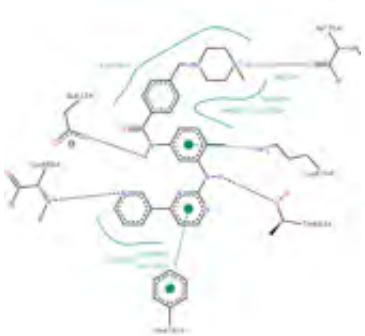


RCSB.org: Supporting the Scientific Ecosystem

- RCSB PDB services go well beyond original structure and scientific publication
- Up-to-date access to
  - Newly-released PDB structures
  - Sequence/3D structure comparisons
  - Integration with ~40 external resources
  - 3D structure/annotation visualization



2D/3D Sequence Annotations



Target-Drug Interactions

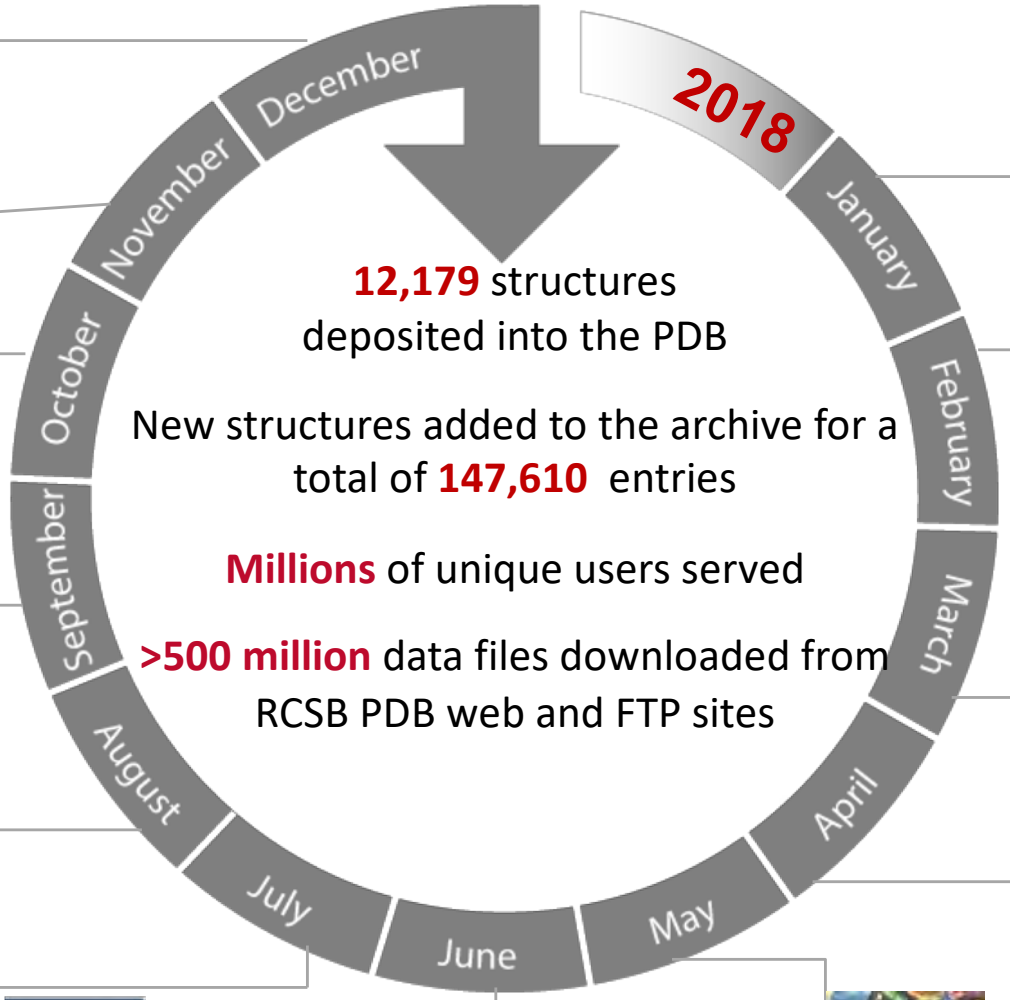
Pathways

Genetic Variations

PDB-101: Training Support for ~600K Users/Year

- Primary distribution of Outreach/Education efforts
- *Molecule of the Month:* >230 articles about Fundamental Biology, Biomedicine, and Energy
- Curricular modules on public health concerns, fundamental structural biology
- Videos, posters, PDB data user guides, and other content
- Today's students are tomorrow's PDB users

Year in the Life of the RCSB PDB Community



AAAS Biomedical Career Symposium

BioXFEL

Undergraduate Course on Antimicrobial Resistance

BPS18 62nd Annual Meeting Biophysical Society San Francisco, California February 17-21, 2018

Science Olympiad

RUTGERS DAY Experience New Jersey's State University

ISB International Society for Biocuration

Site visit

5th Annual Video Challenge

CCP4 Argonne

ICS 2018

ISMB 2018 CHICAGO JULY 6-10

The New York Structural Biology Discussion Group Summer Meeting

Superbugs: Evolutionary Insights from the Protein Data Bank

SACNAS

wwPDB and RCSB PDB AC Meetings

CASP13

Publications Supporting RCSB PDB Services

Investigation of protein quaternary structure via stoichiometry and symmetry information
(2018) *PLoS ONE* 13: e0197176. doi: 10.1371/journal.pone.0197176

RCSB Protein Data Bank: Sustaining a living digital data resource that enables breakthroughs in scientific research and biomedical education
(2018) *Protein Science* 27: 316–330 doi: 10.1002/pro.3331

Worldwide Protein Data Bank biocuration supporting open access to high-quality 3D structural biology data
(2018) *Database* 2018: bay002 doi: 10.1093/database/bay002

Recon3D enables a three-dimensional view of gene variation in human metabolism
(2018) *Nature Biotechnology* 36: 272–281 doi: 10.1038/nbt.4072

Automated evaluation of quaternary structures from protein crystals
(2018) *PLoS Comput Biol* 14: e1006104 doi: 10.1371/journal.pcbi.1006104

Outlier analyses of the Protein Data Bank archive using a probability-density-ranking approach
(2018) *Scientific Data* 5: 180293 doi: 10.1038/sdata.2018.293

Analyzing the symmetrical arrangement of structural repeats in proteins with CE-Symm (2018) *bioRxiv* doi: 10.1101/297960

From atoms to cells: Using mesoscale landscapes to construct visual narratives
(2018) *Journal of Molecular Biology* 430: 3954–3968 doi: 10.1016/j.jmb.2018.06.009

Molecular illustration in research and education: Past, present, and future
(2018) *Journal of Molecular Biology* 430: 3969-3981 doi: 10.1016/j.jmb.2018.04.043

Learning biology through molecular storytelling
(2018) *The Science Teacher* 86: 28-33

NGL viewer: web-based molecular graphics for large complexes
(2018) *Bioinformatics* 34: 3755–3758. doi: 10.1093/bioinformatics/bty419

Analysis of impact metrics for the Protein Data Bank
(2018) *Scientific Data* 5: 180212 doi: 10.1038/sdata.2018.212

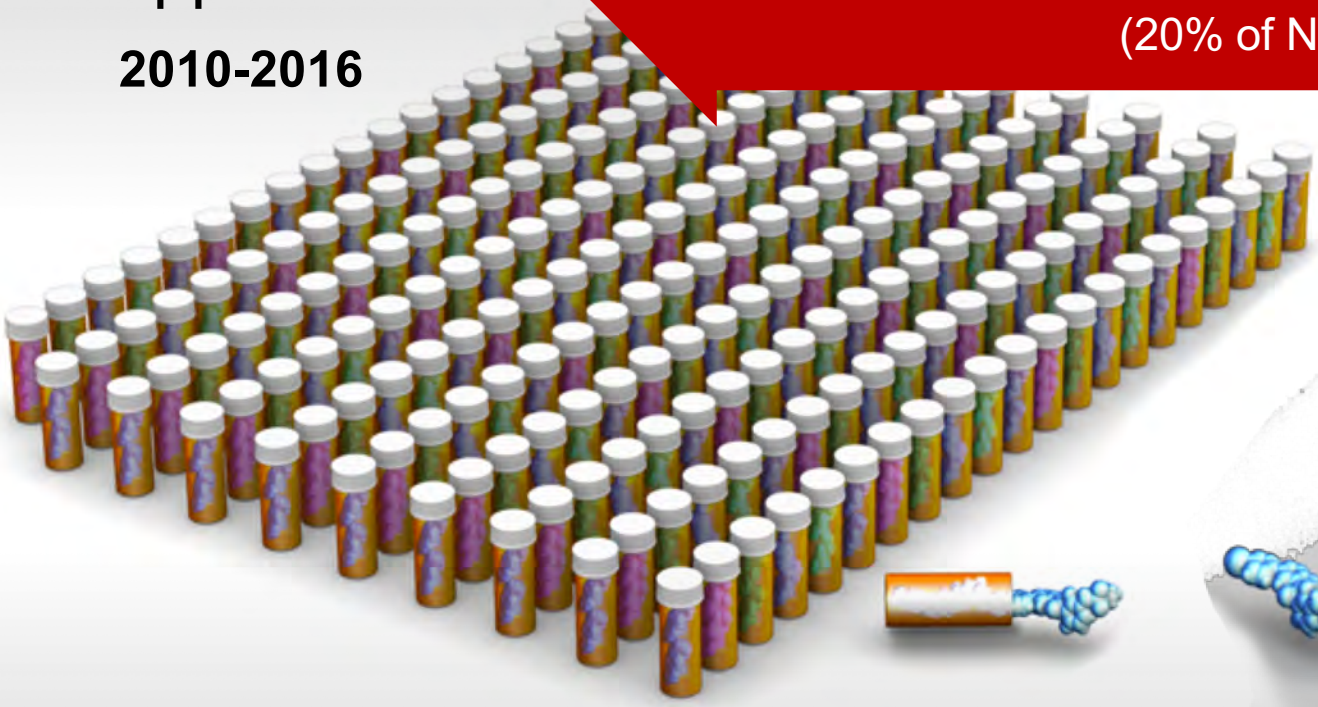
Amino acid modifications for conformationally constraining naturally occurring and engineered peptide backbones: Insights from the Protein Data Bank
(2018) *Biopolymers* 109: e23230 doi: 10.1002/bip.23230

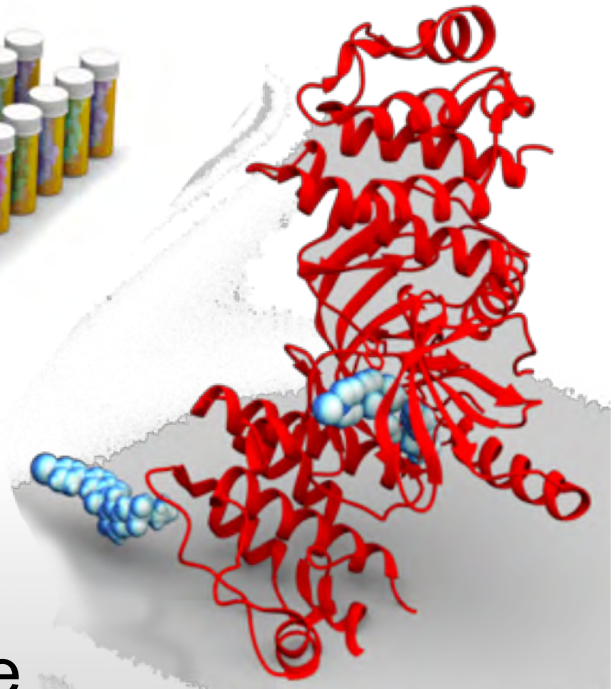
PDB Impact on 2010-2016 New Drug Approvals[1]

210 NEW DRUGS
approved
2010-2016

2000-2016

>$100 BILLION **of NIH funding**
contributed to these approvals
(20% of NIH Budget)[2]





B-Raf Kinase complex with Vemurafenib
PDB ID 3og7

5,913 PDB Structures contributed to **184** of these drug approvals

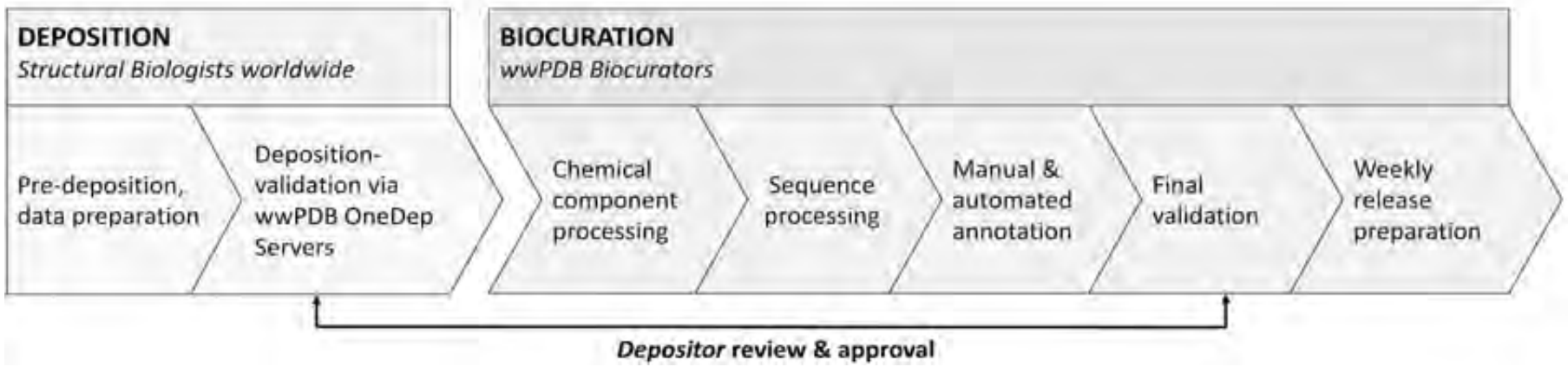
1. Westbrook & Burley (2019) *Structure 27*, 211-217.
2. Galkina Cleary *et al*. (2018); Value in 2016 US$.

Deposition/Biocuration (Service 1)

Jasmine Young

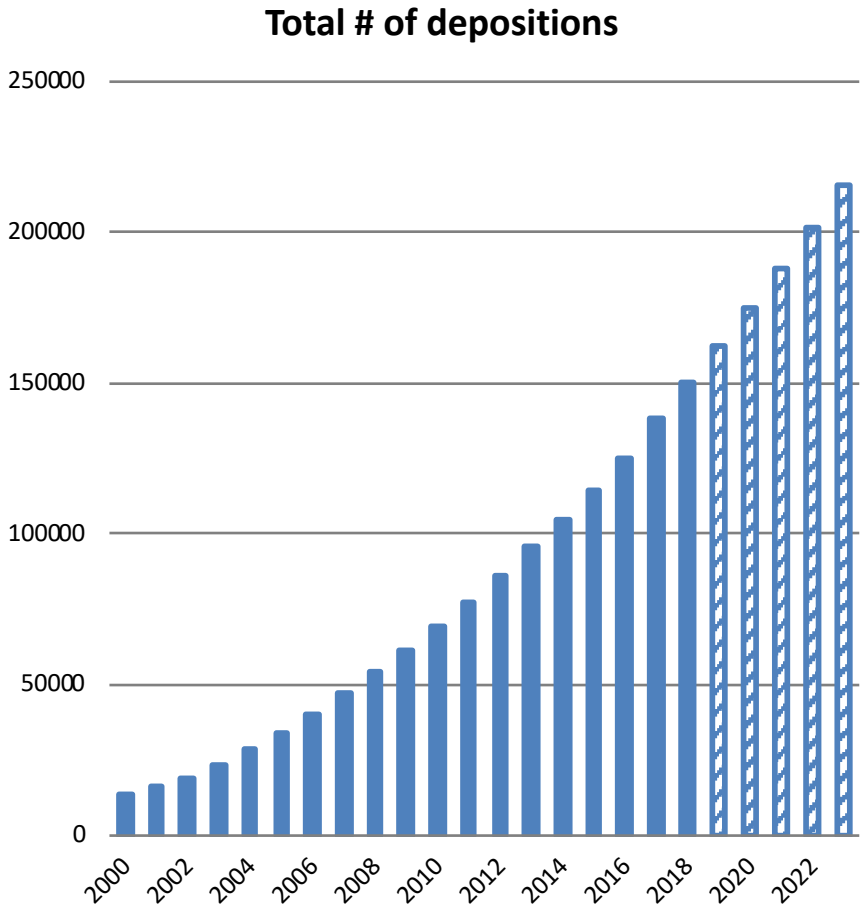
Deposition/Biocuration Ensures Well-curated and High Quality Structure Data

- Support structures determined by MX, NMR, and 3DEM methods and combinations with these techniques (e.g., NMR-SAS)
- Pre-deposition tools provide data preparation for submission
- Validation implements community Task Force recommendations
- Geographically distributed biocuration



2018 Deposition Growth

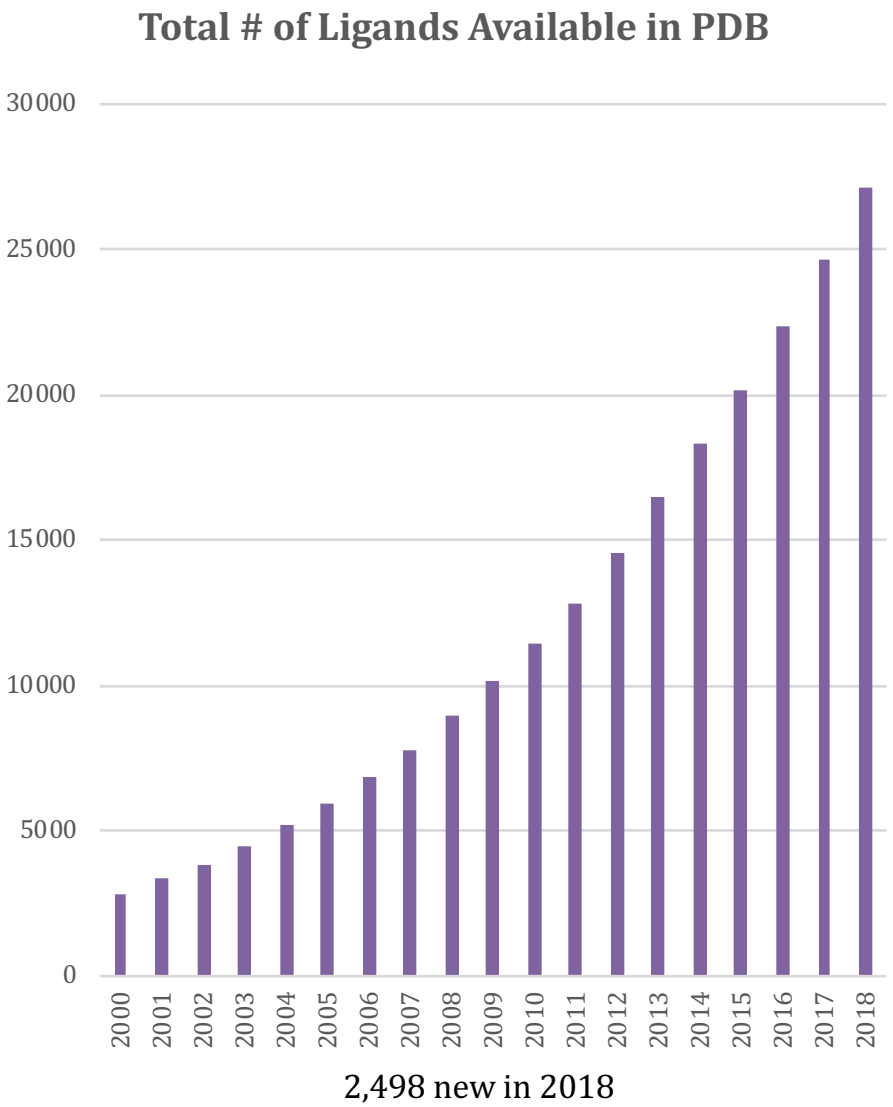
12,179 Structures

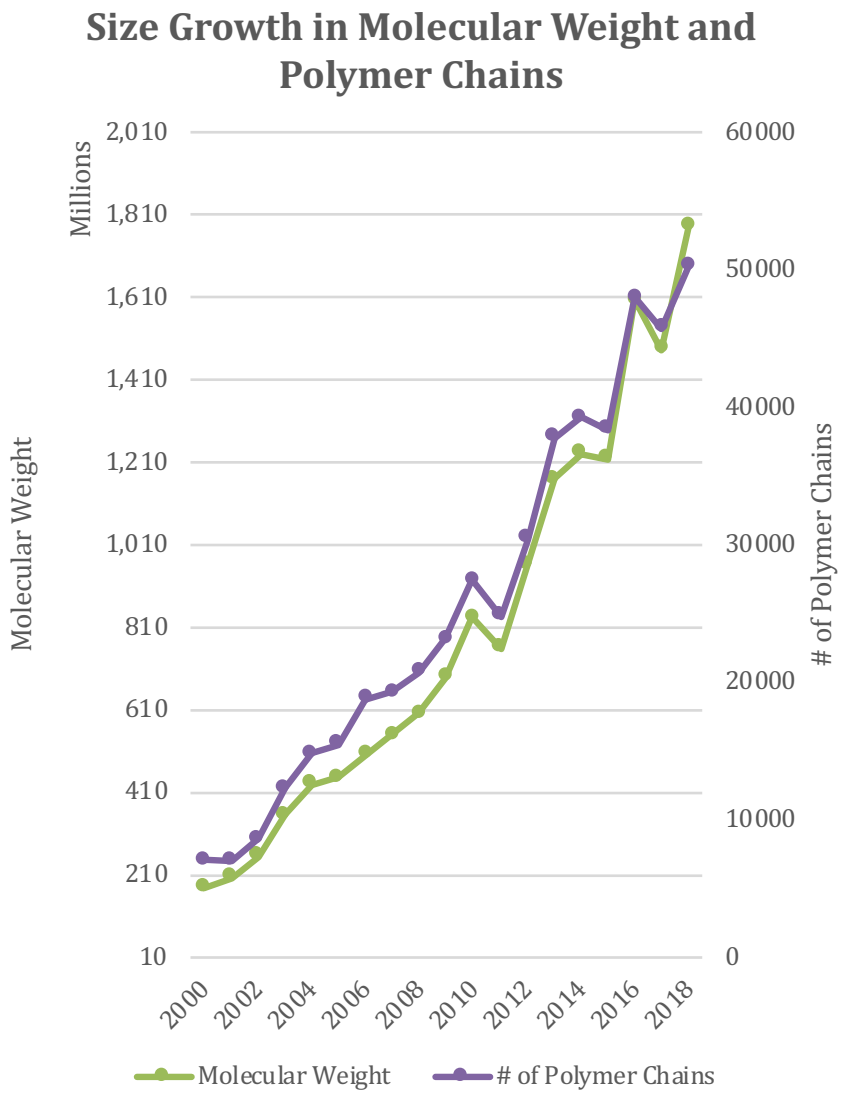


Rapid growth in 3DEM

Method	2017 Depositions	2018 Depositions
MX	11,889 (91.1%)	10594 (87.0%)
NMR	460 (3.5%)	418 (3.4%)
3DEM	**674 (5.2%)**	**1140 (9.4%)**
Other	26 (0.2%)	27 (0.2%)

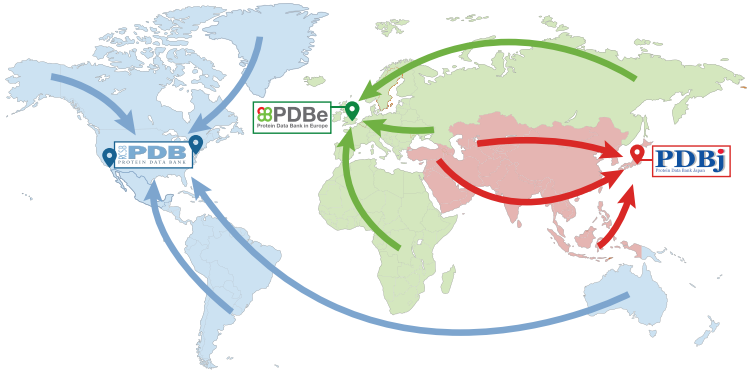
Complexity and Size Growth

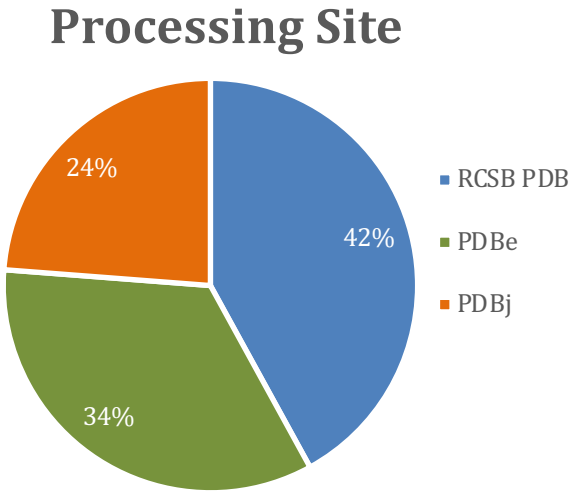


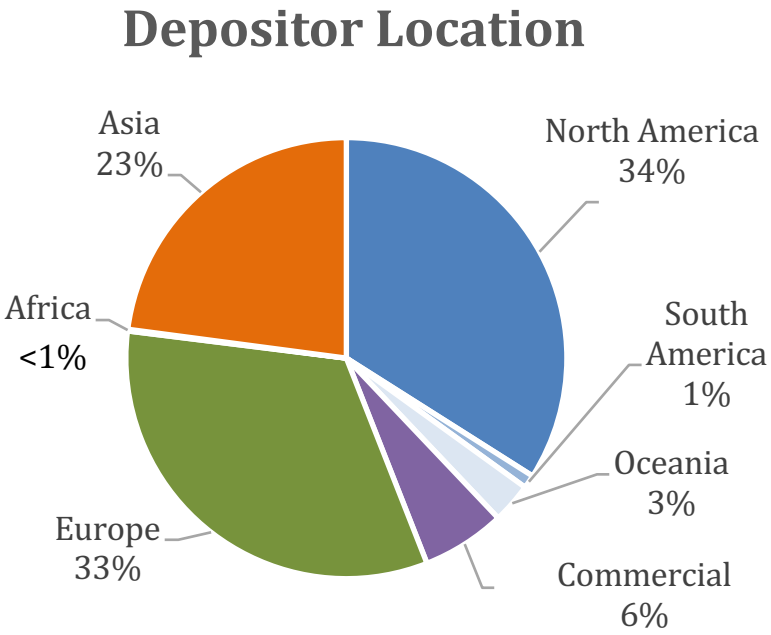


2018 Deposition/Biocuration Statistics

- 12,179 deposited
  - 5117 RCSB PDB-biocurated
- Workload balanced geographically
  - 42% Americas, Oceania
  - 34% Europe, Africa
  - 24% Asia

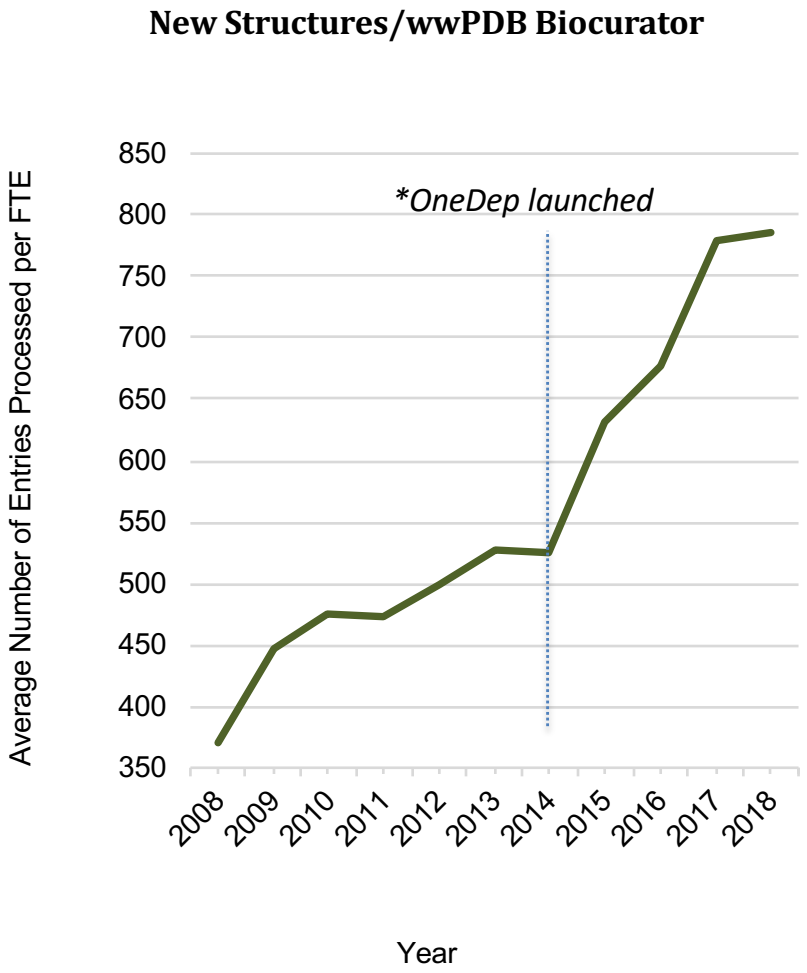






Addressing Increasing Growth and Complexity Through Biocuration Efficiency

- Continuing increased efficiency since 2009
- Ongoing improvements in Biocuration processes
- Significant increase from OneDep launch
  - Need to boost productivity in 2019 and beyond



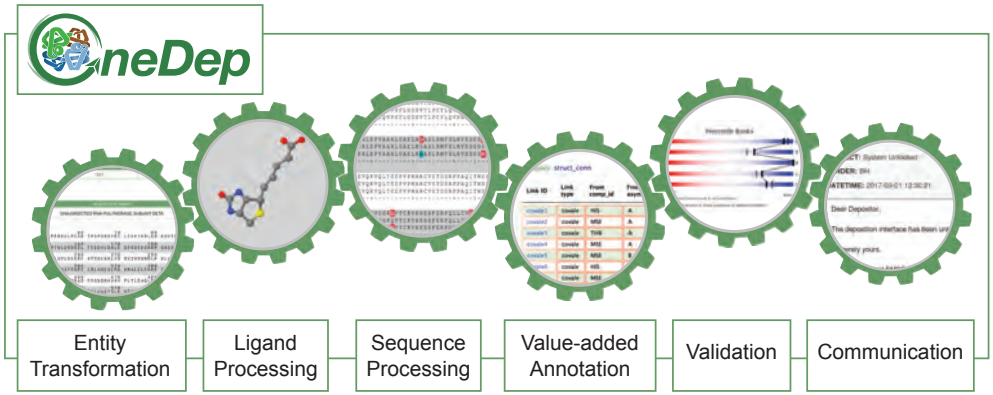
2018 Efforts to Improve Biocuration Efficiency

Biocuration

- Enable utilization of external computing for large calculations (e.g., ribosome validation)
- Re-use previous sequence annotation
- Routine tasks more automated
- Processes streamlined

Deposition

- Major issues made more prominent to depositors
- More checks and gates to deposition



2018 Milestones

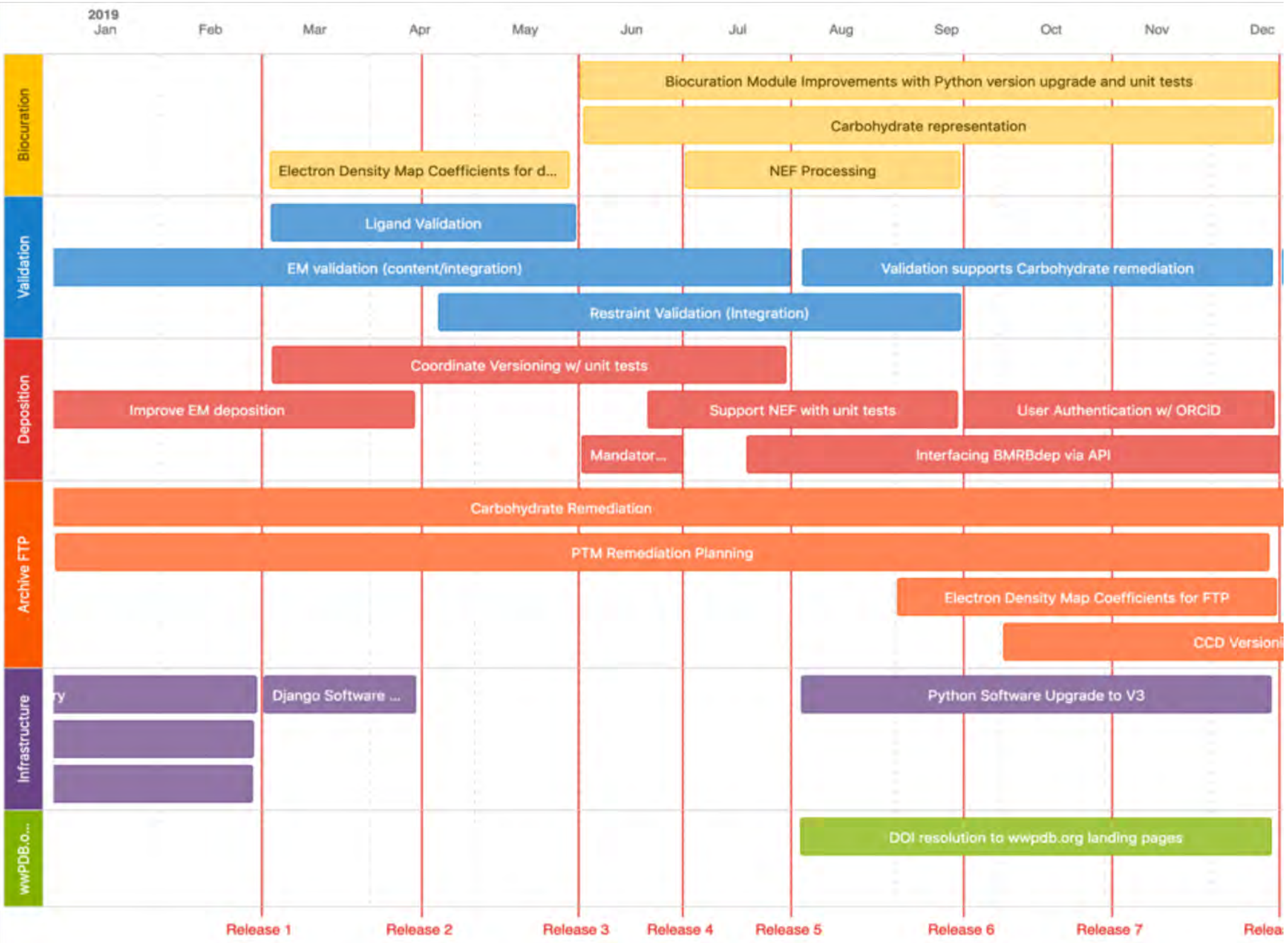
OneDep Development

- Mandatory ORCiD
  - 25% of unique depositors with ORCiD
  - 3342 unique PIs with ORCiD
- Improved biocuration efficiency
- Better software management *via* GitHub
- GDPR-compliant

PDB Archive Improvements

- Carbohydrate Remediation
  - Collaboration with Glycoscience community
  - PDBx/mmCIF dictionary extension and examples public *via* GitHub
- Validation report recalculation

ORCiD
Connecting Research and Researchers


2019
Jan
Feb
Mar
Apr
May
Jun
Jul
Aug
Sep
Oct
Nov
Dec
Biocuration
Validation
Deposition
Archive FTP
Infrastructure
wwPDB.o...
Biocuration Module Improvements with Python version upgrade and unit tests
Carbohydrate representation
Electron Density Map Coefficients for d...
NEF Processing
Ligand Validation
EM validation (content/integration)
Validation supports Carbohydrate remediation
Restraint Validation (Integration)
Coordinate Versioning w/ unit tests
Improve EM deposition
Support NEF with unit tests
User Authentication w/ ORCiD
Mandator...
Interfacing BMRBdep via API
Carbohydrate Remediation
PTM Remediation Planning
Electron Density Map Coefficients for FTP
CCD Versioni
Django Software ...
Python Software Upgrade to V3
DOI resolution to wwpdb.org landing pages
Release 1
Release 2
Release 3
Release 4
Release 5
Release 6
Release 7

2019 Goals

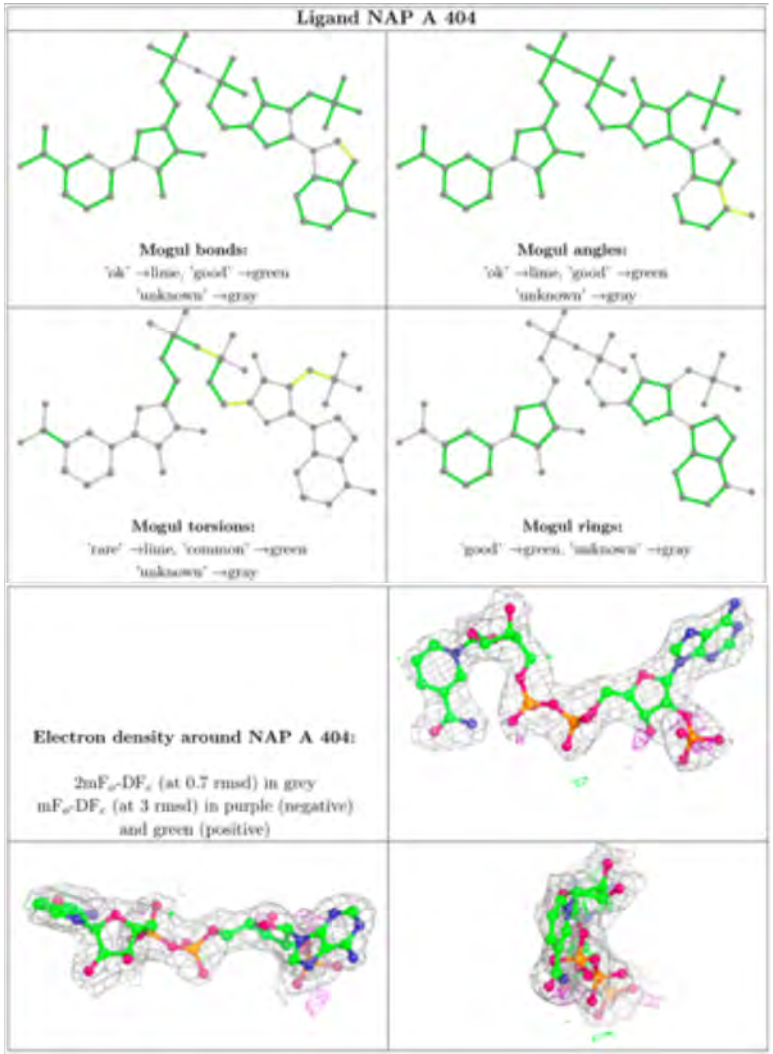
Goal	Impact/Gain
Validation enhancements: Ligands, NMR restraints, EM maps	Improve data quality Increase Biocuration efficiency
Mandatory mmCIF deposition for crystallographic structures	Capture more complete data Increase Biocuration efficiency
Author-initiated coordinate replacement	Improve data quality More automated Biocuration
Supporting NEF format from NMR technique	Enable restraint validation Improve data quality
Carbohydrate remediation	Enable FAIR Better data validation
Chemical Component versioning	Better data management Automated tracking on changes
Biocuration by Depositor and Biocuration Automation	Increase Biocuration efficiency
DOI resolution at wwpdb.org landing page	Highlight wwPDB collaboration
Provide ED map coefficients at FTP	Enable data reproducibility
Infrastructure software upgrade	More effective software testing and deployment Reduce Biocuration testing resource

Developing Next Generation Ligand Validation

- Software adapted from Global Phasing Ltd. under formal agreement
- Provides 2D depiction of geometrical quality
- Provides electron density fit for X-ray
- Now mandatory at deposition: identification of Ligand/s Of Interest (LOI)

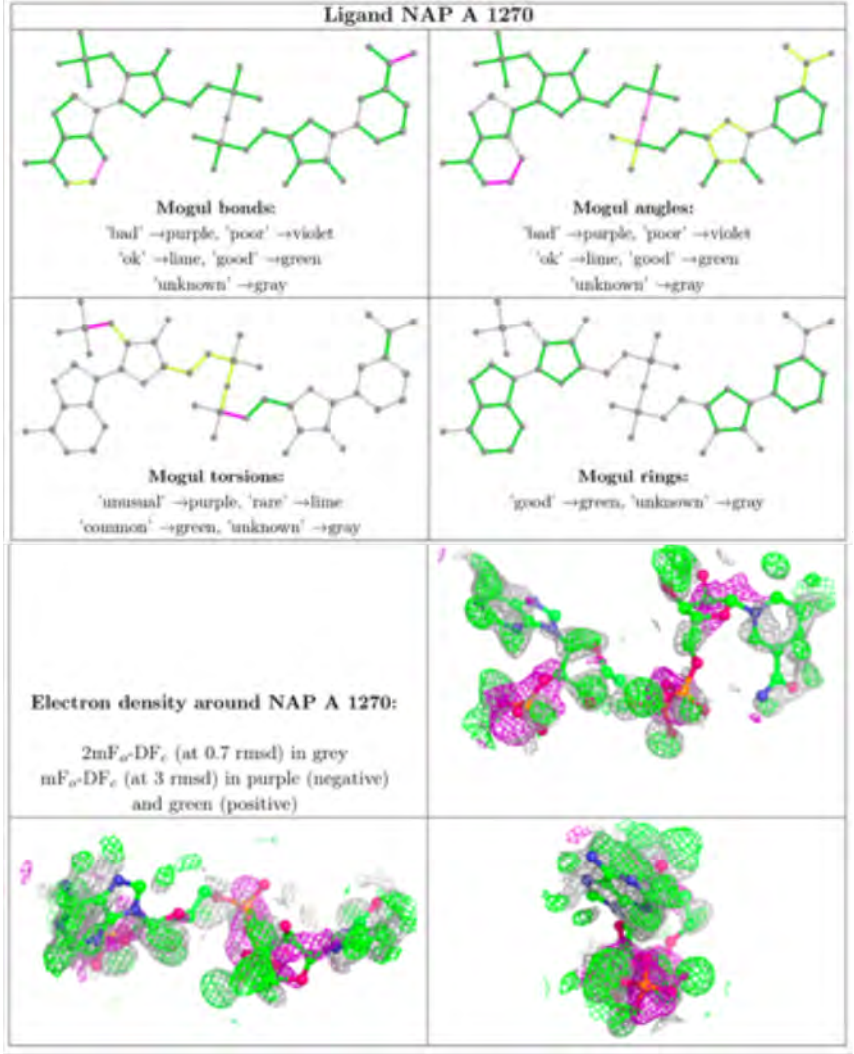
Examples of NADP

Mol	Type	Chain	Res	Atoms	RSCC	RSR	B-factors(Å^2)	Q<0.9
3	NAP	A	404	48/48	0.96	0.14	31,43,66,70	0



PDB entry 5zix (Better data quality)

Mol	Type	Chain	Res	Atoms	RSCC	RSR	B-factors(Å^2)	Q<0.9
3	NAP	A	1270	48/48	-0.06	0.67	87,96,100,100	0



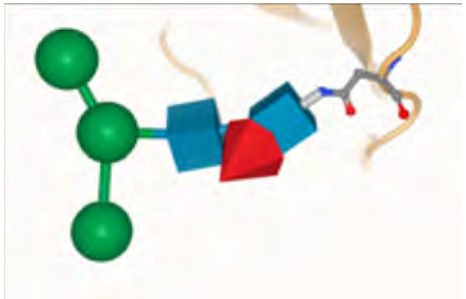
PDB entry 1zk4 (Worse data quality)

Carbohydrate Remediation

(NIGMS grant U01 CA221216)

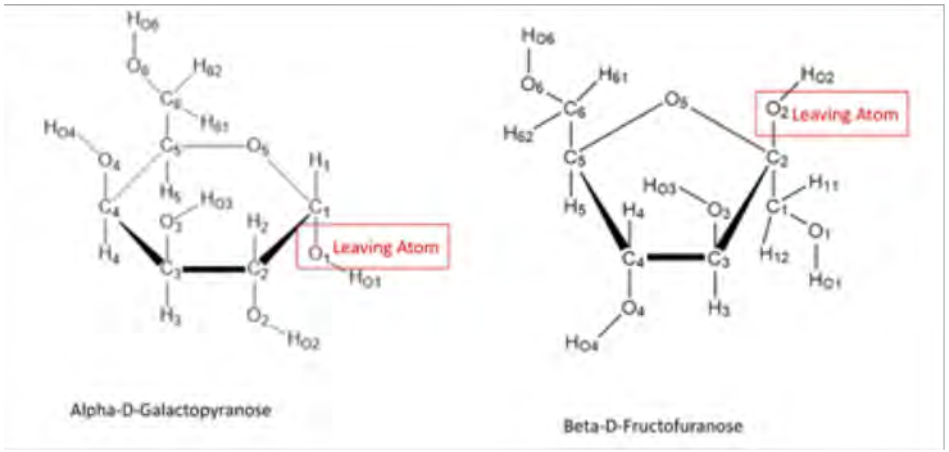
Objectives

- Standardize nomenclature following IUPAC/IUBMB
- Adopt community software for
  - standard nomenclature assignment
  - linear description for oligosaccharides
- Provide uniform representation for oligosaccharides with appropriate descriptor(s)
- Identify, validate, and biocurate glycosylation



Scope

- 1,614 monosaccharides and 369 oligosaccharides in PDB Chemical Component Dictionary
- 15,244 PDB structures
- ~20,000 oligosaccharides in 9,000 PDB structures



Interoperable Linear Representation for Oligosaccharides

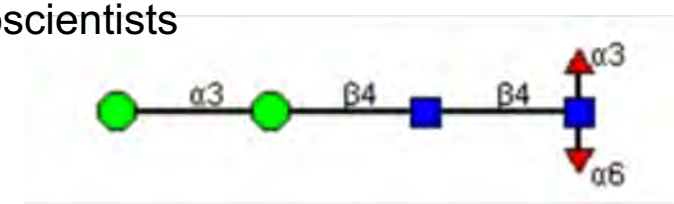
Condensed IUPAC

LFucpa1-6[DManpa1-3DManpb1-4DGlcpNAcb1-4][LFucpa1-3]DGlcpNAcb1-ASN

LINUCS

[][ASN]{[(4+1)][b-D-GlcpNAc]{[(3+1)][a-L-Fucp]{}[(4+1)][b-D-GlcpNAc]{[(4+1)][b-D-Manp]{[(3+1)][a-D-Manp]{}}}[(6+1)][a-L-Fucp]{}}}

These description can be translated into Symbolic representations used by glycoscientists



PDB ID 6cmg

IUPAC

```
                                a-L-Fucp-(1-6)+
                                              |
a-D-Manp-(1-3)-b-D-Manp-(1-4)-b-D-GlcpNAc-(1-4)-b-D-GlcpNAc-(1-4)-ASN
                                              |
                                a-L-Fucp-(1-3)+
```

Carbohydrate Project Status

- Extended data content and examples available to the public
  - Project summary:
    - https://www.wwpdb.org/documentation/carbohydrate-remediation
  - Examples of remediated data
    - https://github.com/pdbxmmcifwg/carbohydrate-extension/tree/master/examples
- Glycoscience community tools that produce oligosaccharide linear descriptors and IUPAC nomenclature tested
- Branched polymer representation software ready for integration with community tools
- Currently standardizing monosaccharide nomenclature in PDB Chemical Component Dictionary

Management

Stephen K. Burley

Managing a Global Public Good

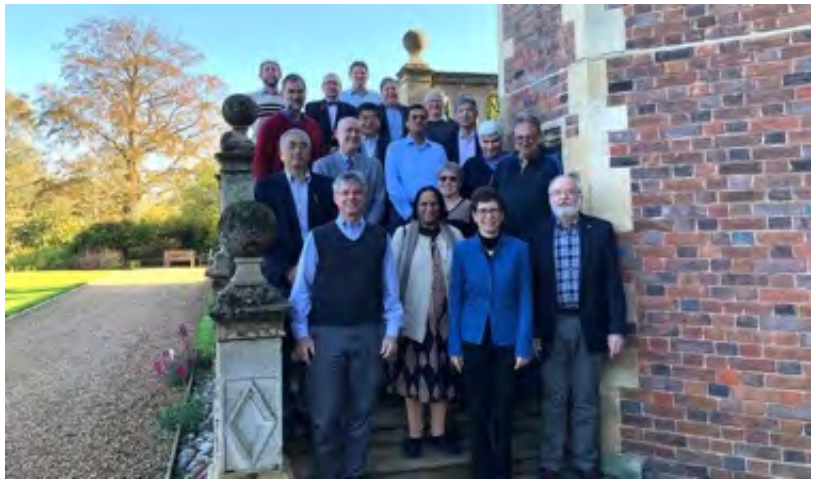
- Experienced leadership team
- Broad knowledge of basic and applied research
- Deep subject matter expertise in biomedical science and information technology
- Strong project management support
- Specialist community engagement and oversight
- Professional accreditation
- Responsible data stewardship

Organizational Considerations

- Delivery of high quality services and complex products to a diverse, global user community
- RCSB PDB staff is by design
  - Broad range of skills represented
  - Domain experts in key areas
  - Geographically distributed
- Strategic collaborations with international partners
- Scientific rigor combined with effective project management

wwPDB International Collaboration

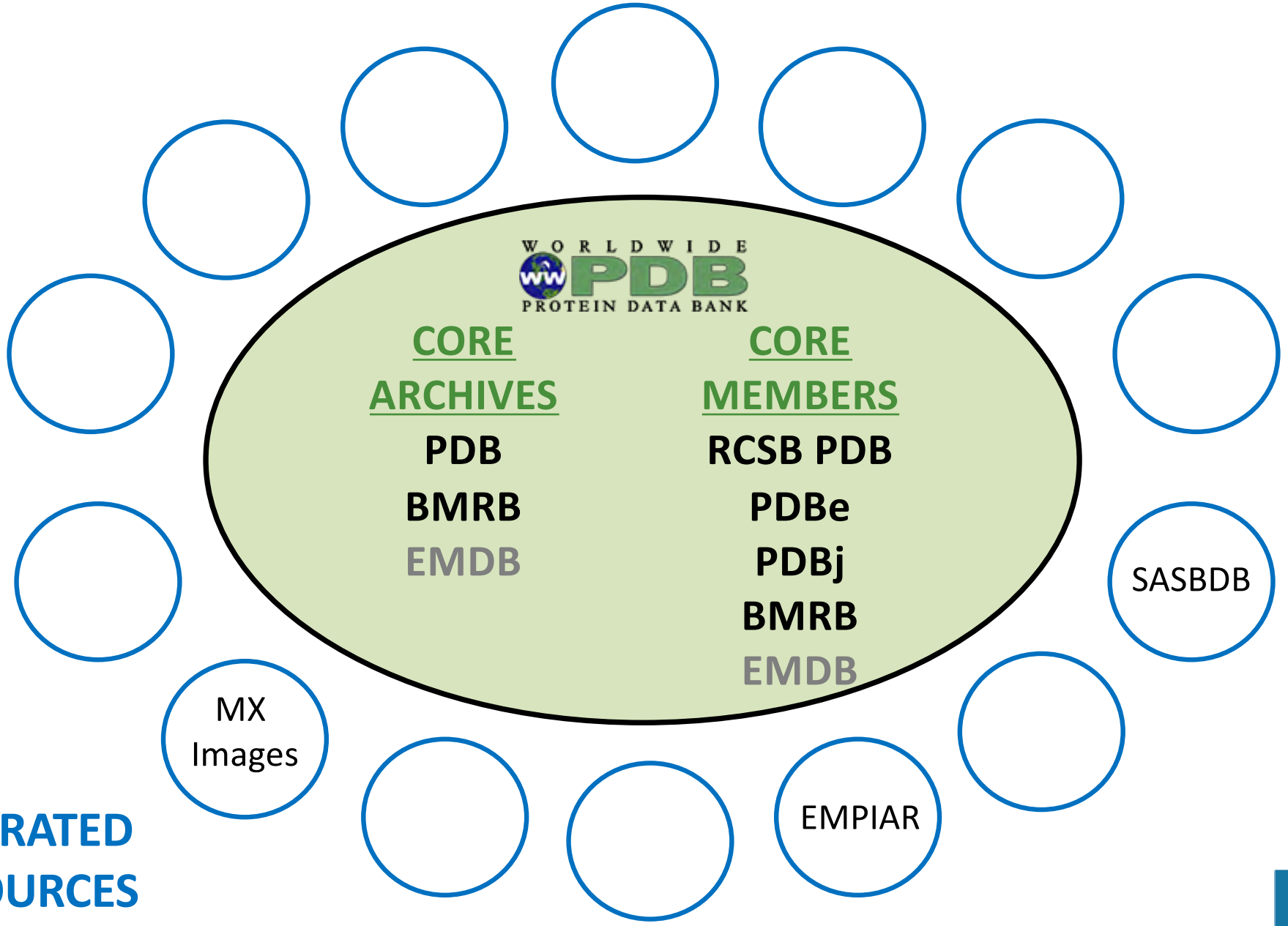
- Cost-sharing ensures PDB continues as the single Open Access archive
- Operates the global system for deposition, validation, and biocuration (OneDep)
- Defines data standards and content in concert with the scientific community (PDBx/mmCIF Dictionary)
- Enables data uniformity in the PDB archive ("Remediation")
- Synergistic sharing of resources and services (e.g., SIFTS, Mol*)
- Member-hosted websites offer complementary services and views of the data



2018 wwPDB Advisory Committee Meeting

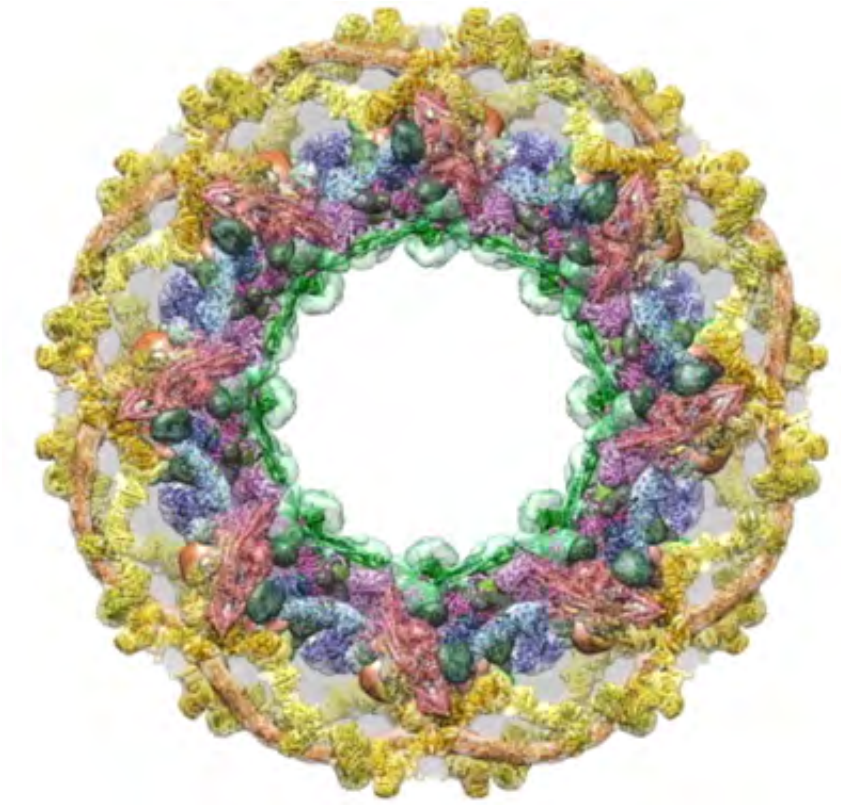
WORLDWIDE PDB PROTEIN DATA BANK — RCSB PDB PROTEIN DATA BANK — PDBe Protein Data Bank in Europe — PDBj — BMRB

New wwPDB Organizational Structure



Other Strategic Collaborations

- Integrative/Hybrid Methods Working Groups
- EMDB Archive Team
- UniProt
- CCDC: Cambridge Crystallographic Data Centre
- NCBI PubChem
- Other external data resources



I/H Methods Structures
552-protein yeast Nuclear Pore Complex
Kim et al. (2018) *Nature 555*, 475-82
PDBDEV_00000010; PDBDEV_00000011; PDBDEV_00000012

Staff Recruitment and Advancement

- Ongoing recruitment with assistance from host institution Human Resources
  - Diversity and Inclusion Planning
- Professional development
  - Mentoring
  - In-service training
  - Professional society involvement
  - Co-authoring scientific papers and proposals
  - Science communication outreach and teaching
- Healthy turnover of ~10%/year
- Where do RCSB PDB staff go when they leave?
  - Private sector jobs (*e.g.*, Disney, Google, Invitae)
  - Academic faculty positions

RCSB PDB Team

RCSB PDB PROTEIN DATA BANK

RCSB.ORG

info@rcsb.org

Funding

RCSB PDB is funded by the
National Science Foundation (DBI-1832184),
the National Cancer Institute,
the National Institute of General Medical Sciences,
and the US Department of Energy (DE-SC0019749)

Management

RCSB PDB is hosted by:

RUTGERS

UC San Diego

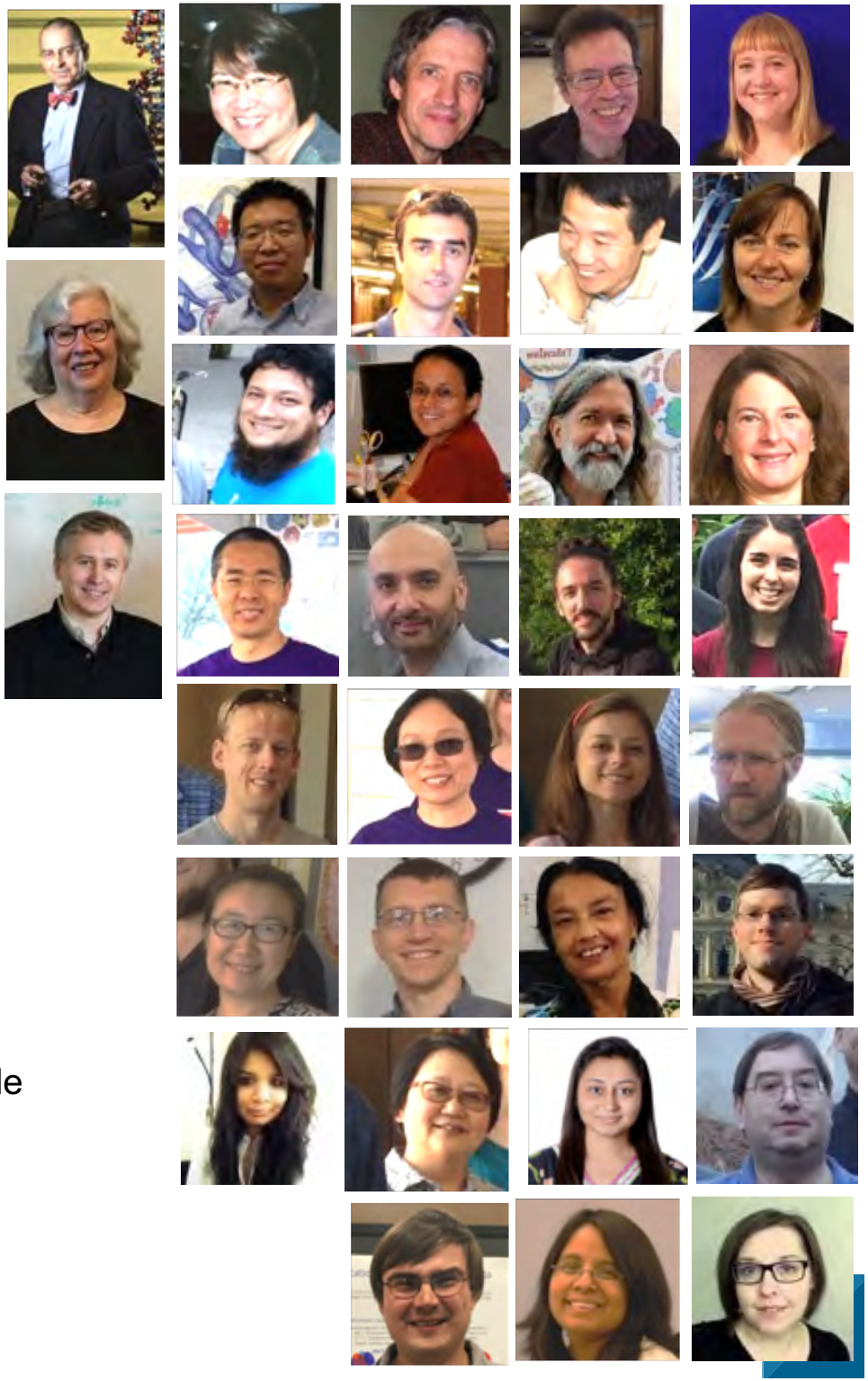
SDSC SAN DIEGO SUPERCOMPUTER CENTER

UCSF University of California San Francisco

WORLDWIDE PDB PROTEIN DATA BANK

RCSB PDB is a member of the Worldwide Protein Data Bank partnership (wwPDB; **wwpdb.org**)

Follow us



SOFTWARE DEVELOPERS AND BIOCHEMISTS

Join the RCSB Protein Data Bank Team at
Rutgers, The State University of New Jersey

Open positions:

Biochemical Information & Annotation Specialist (Biocurator)

Curate, validate, and standardize macromolecular structures from the PDB community.

Knowledge and skills:

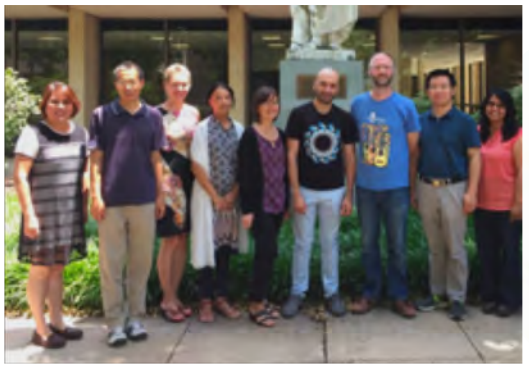
- PhD in Biological chemistry
- Background in 3DEM, small molecule crystallography, or macromolecular crystallography
- Experience with metalloprotein and small molecule data
- Knowledge of Linux computer systems and biological databases preferred

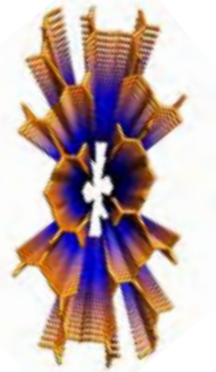
Front End Web Developer

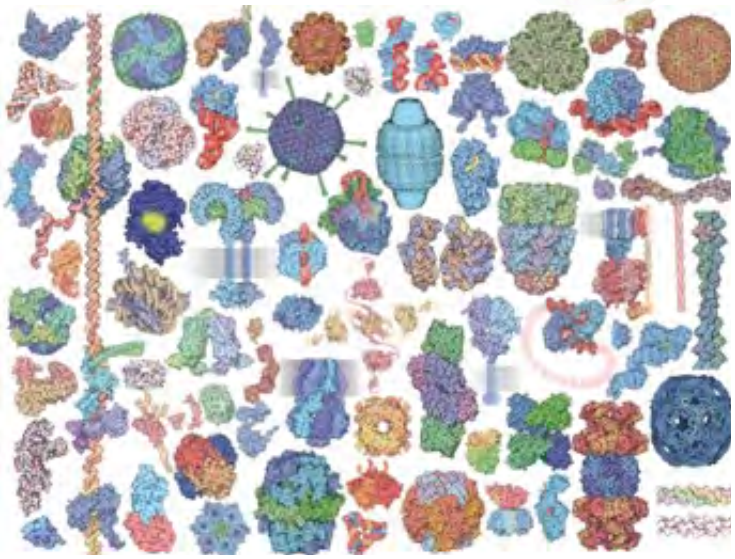
Develop and maintain web applications, from design to deployment.

Knowledge and skills:

- Familiarity with responsive, adaptive design practices using HTML, CSS, Bootstrap
- Experience with JavaScript, JavaScript frameworks and libraries
- Any experience with backend services such as databases (MongoDB), REST, or GraphQL a plus
- Experience with TypeScript and WebGL a plus

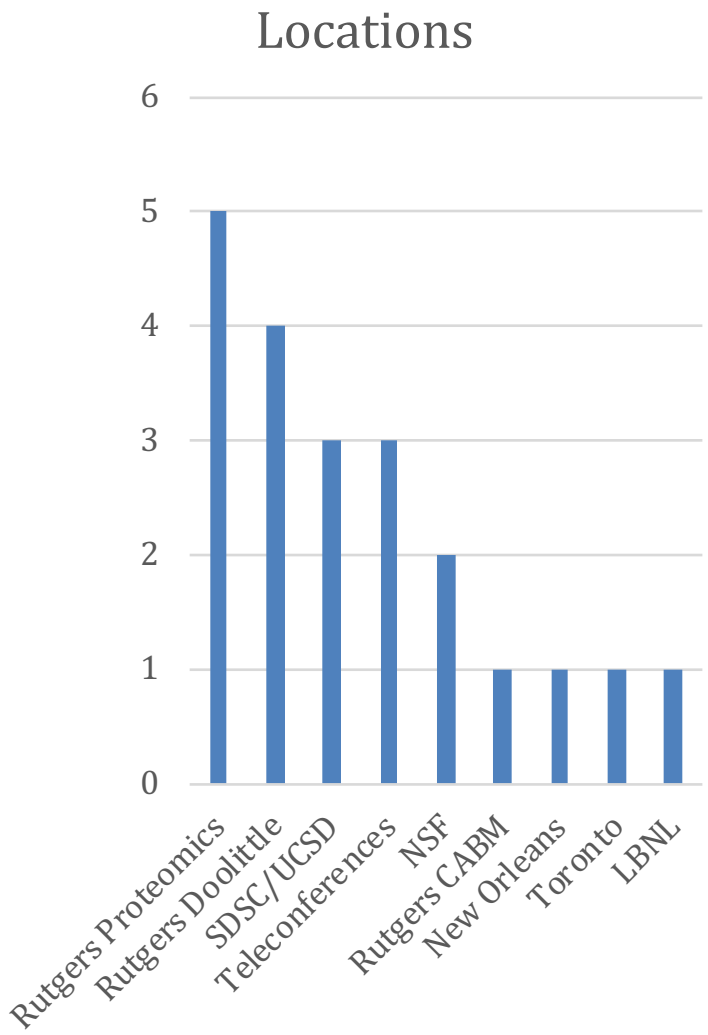






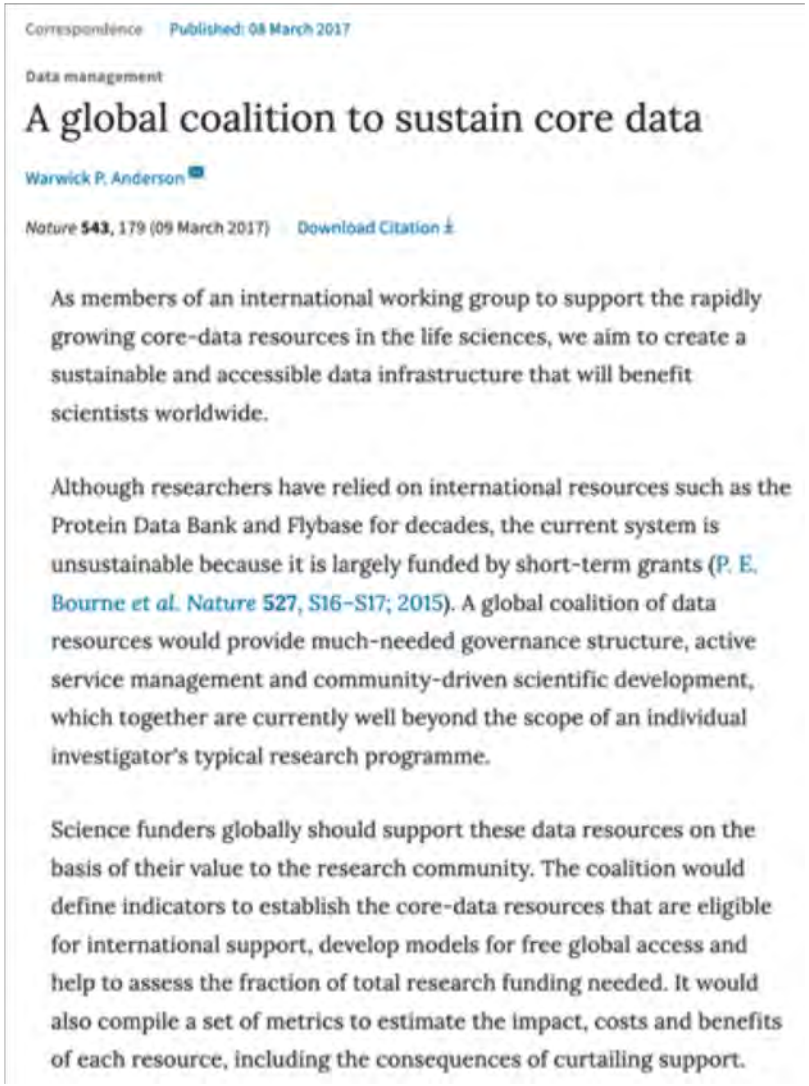
More information: http://www.rcsb.org/pages/jobs
Questions? info@rcsb.org

Celebrating 21 Advisory Committee Meetings



Sustaining Open Access Biological Data

- Perceptible, but slow, movement towards finding a global solution for funding data resources like PDB
  - Global Life Science Data Resources Working Group
  - NIH Scientific Data Council
  - US Interagency Working Group on Biological Data Sharing
  - EU ELIXIR



Correspondence Published: 08 March 2017

Data management

A global coalition to sustain core data

Warwick P. Anderson

Nature **543**, 179 (09 March 2017) Download Citation ±

As members of an international working group to support the rapidly growing core-data resources in the life sciences, we aim to create a sustainable and accessible data infrastructure that will benefit scientists worldwide.

Although researchers have relied on international resources such as the Protein Data Bank and Flybase for decades, the current system is unsustainable because it is largely funded by short-term grants (P. E. Bourne *et al. Nature* **527**, S16–S17; 2015). A global coalition of data resources would provide much-needed governance structure, active service management and community-driven scientific development, which together are currently well beyond the scope of an individual investigator's typical research programme.

Science funders globally should support these data resources on the basis of their value to the research community. The coalition would define indicators to establish the core-data resources that are eligible for international support, develop models for free global access and help to assess the fraction of total research funding needed. It would also compile a set of metrics to estimate the impact, costs and benefits of each resource, including the consequences of curtailing support.

doi: 10.1038/543179a

Do not anticipate change 2019-2023

Mol* 3D Visualization Demonstration

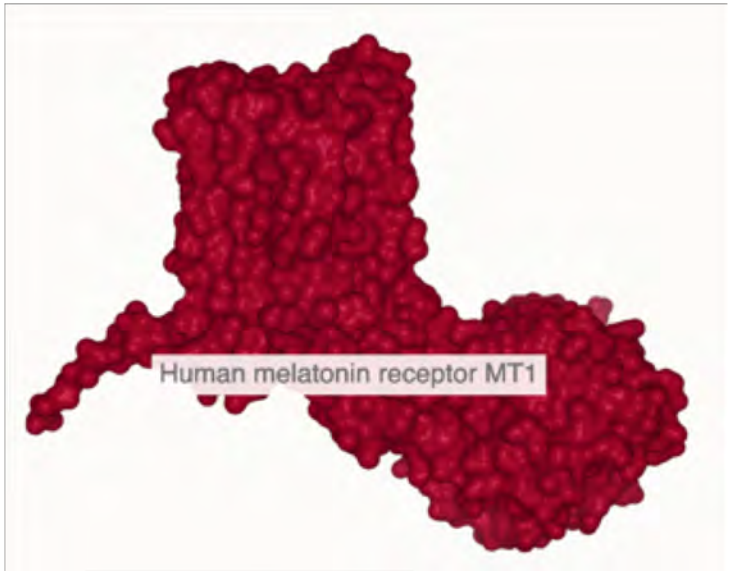
Alex Rose

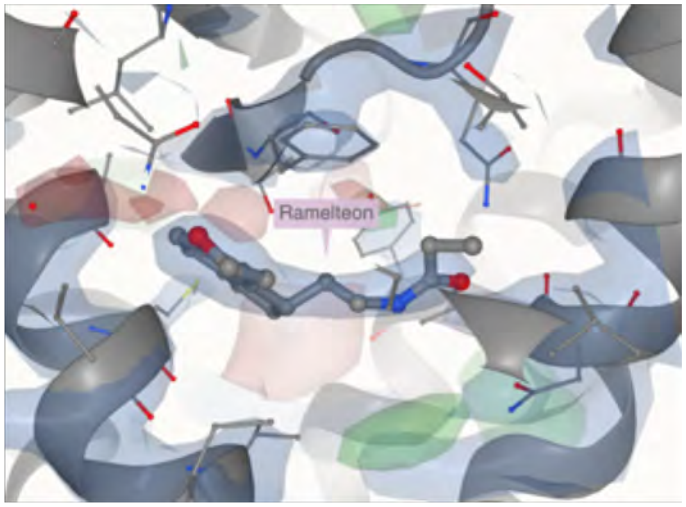
Mol* Overview

- What
  - Web Molecular Graphics + UI/State + Data Delivery Services
- Who
  - Collaborative Project with PDBe (and others, open to everyone)
  - Successor to NGL (RCSB PDB) and LiteMol (PDBe)
- Status
  - Currently working on core capabilities
  - Soon focus on making it more user friendly
  - Try: https://molstar.org/viewer
  - Develop: https://github.com/molstar

Macromolecular Rendering

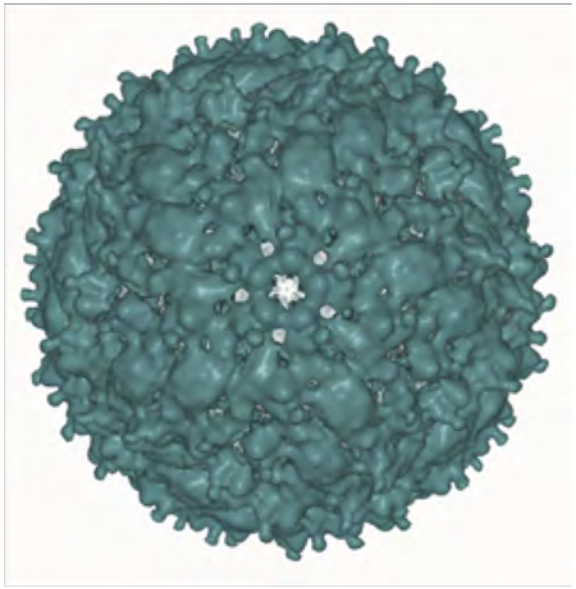
- Basic 3D Representations
  - Cartoon, Spacefill, Ball & Stick, Molecular Surface
- Demo
  - XFEL crystal structure of human melatonin receptor in complex with Ramelteon (6ME2)



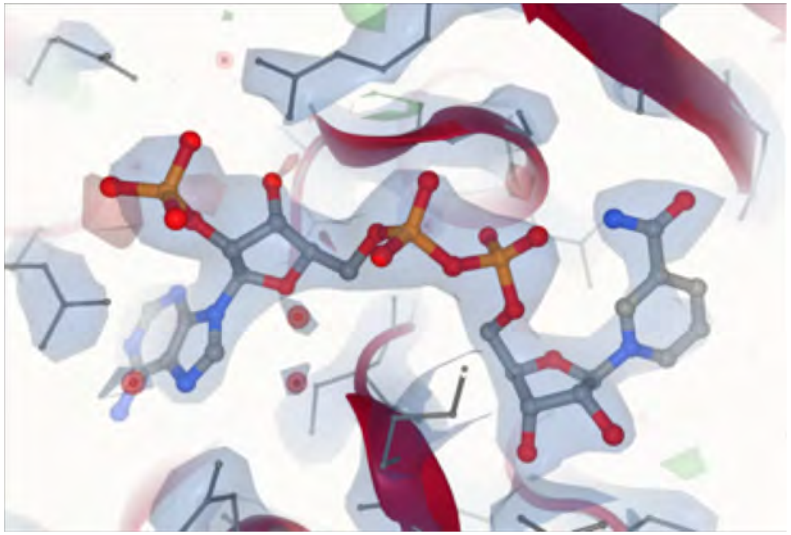


Volume Streaming and Rendering

- VolumeServer
  - Efficient access to small and very large volumetric data sets
  - Evolved from LiteMol's DensityServer
- Demos
  - Zika virus EM density at different resolutions from same dataset
  - X-ray density of selection (around NADP ligand in 5ZIX)



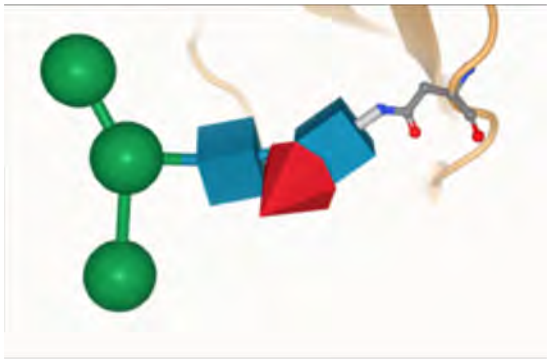
Zika Virus (5IRE)



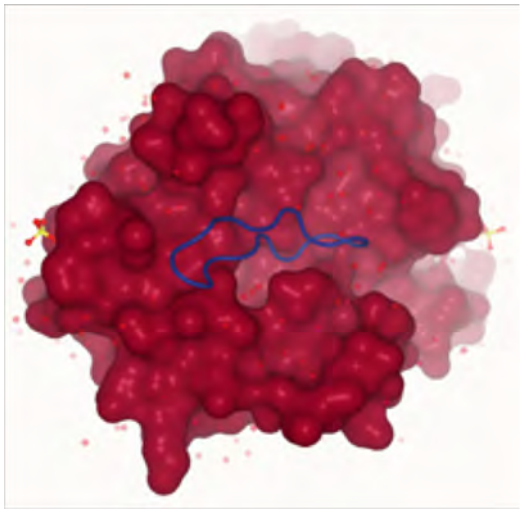
Ketopantoate reductase bound to NADP+ (5ZIX)

Bonus

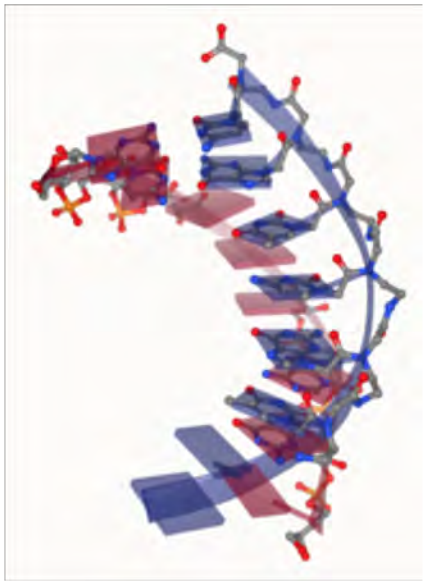
- Carbohydrate Symbols (3D-SNFG)
- Special Cases
  - Cyclic peptides
  - Peptide nucleic acid (PNA)



Carbohydrate in Cardosin (1B5F)

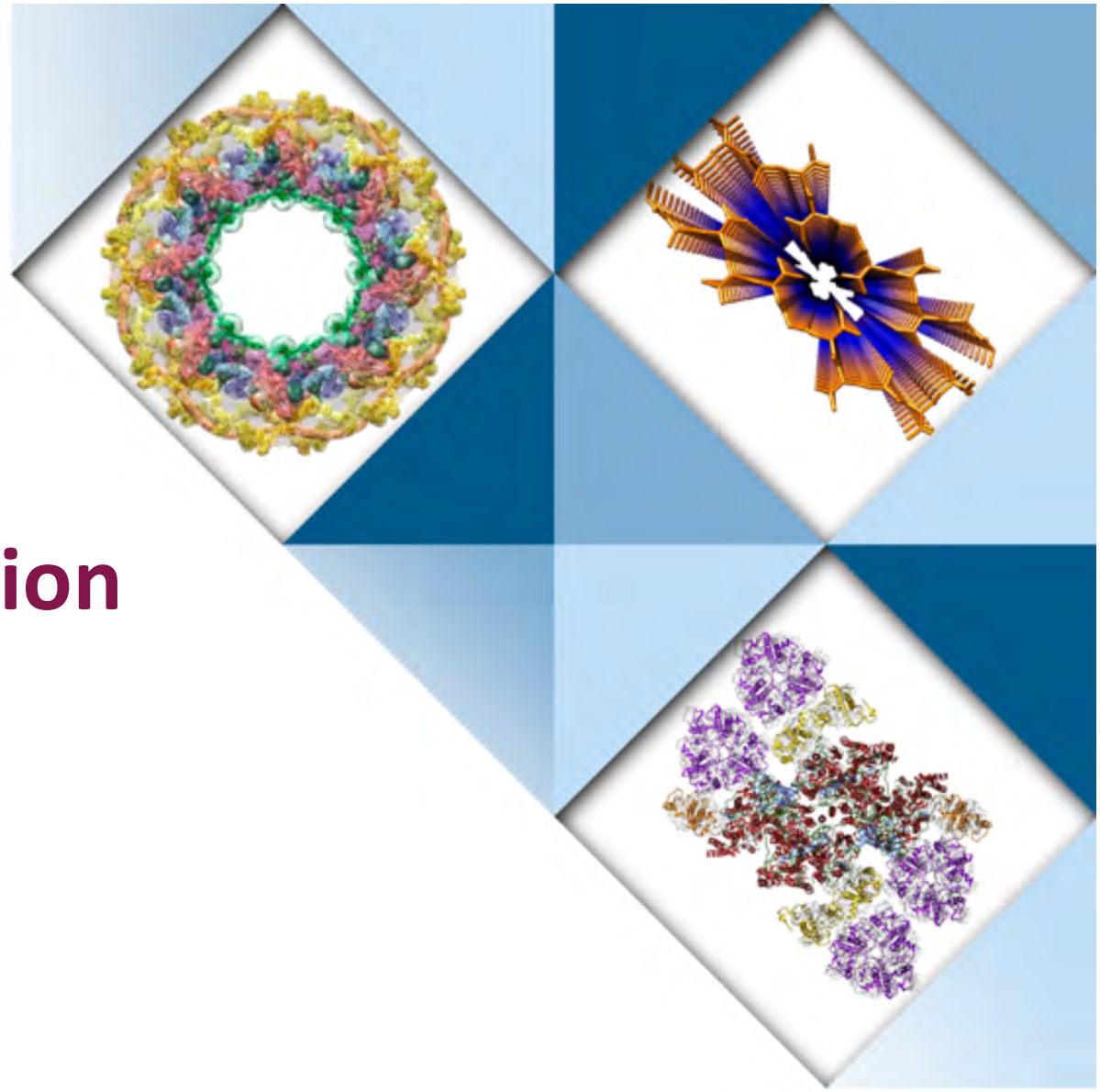


Cyclic protease inhibitor (1SFI)



RNA/PNA complex (5EME)

Demonstration

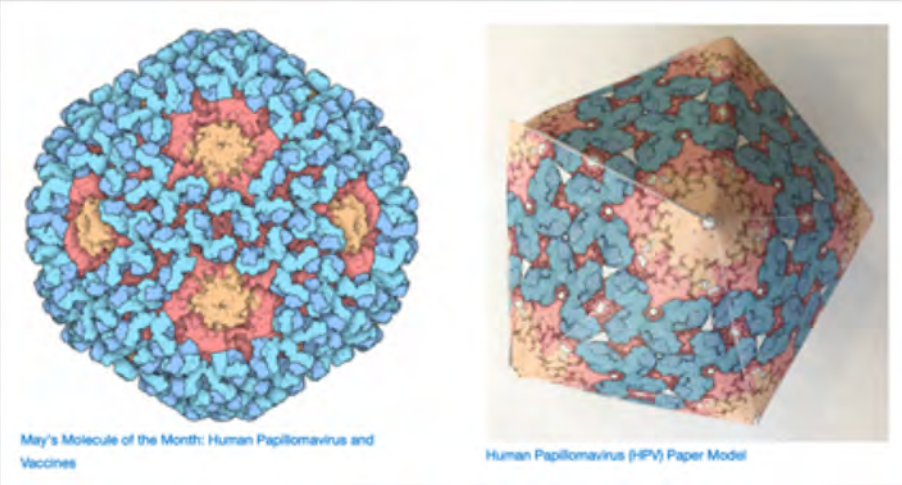


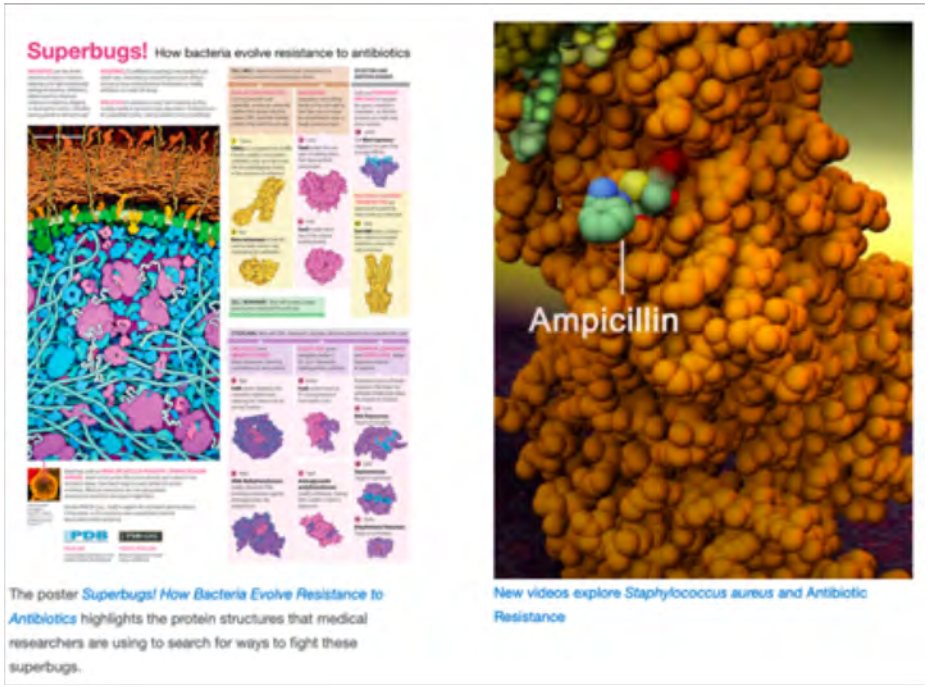
Outreach/Education (Service 4)

Christine Zardecki

Goal: Promote Structural Views of Biology

- Enable open-access exploration of PDB highlights
- Enable education of undergraduate, graduate and professional students, postdoctoral fellows, and researchers in academe, government, and industry
- Provide training materials for PDB Users
- Expose the public to global health topics through the lens of 3D structure





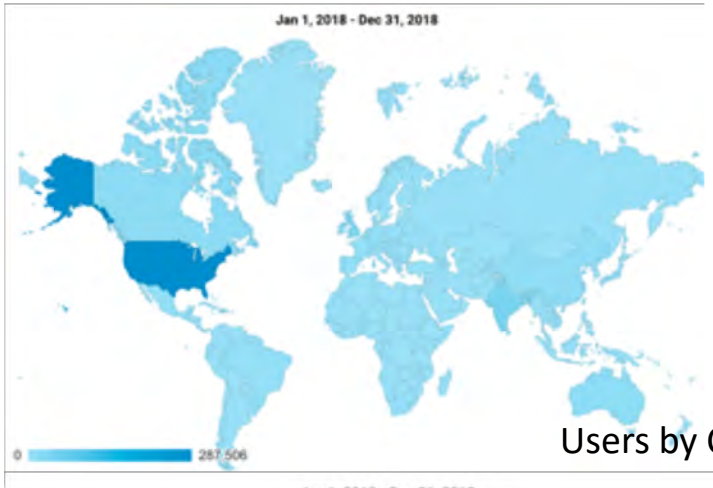
PDB-101 Website Serves as Primary Vehicle

Key Performance Indicators:

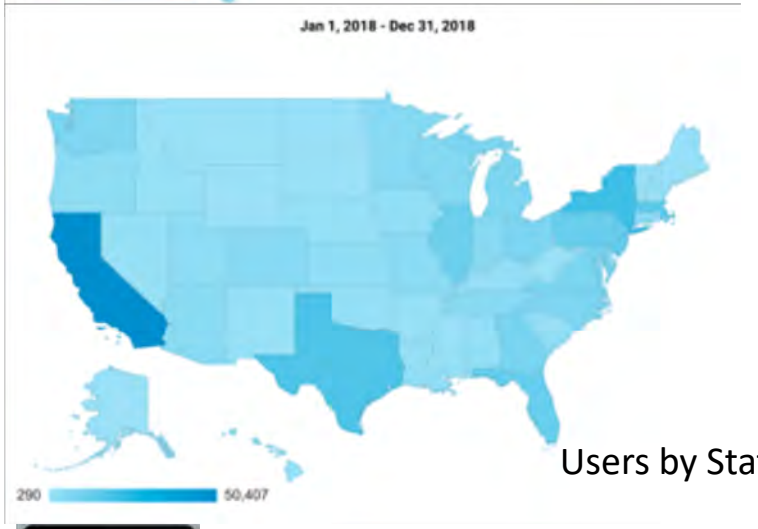
- 1,816,972 pageviews in 2018
  - 1,750,456 in 2017
- 594,073 Users in 2018
  - 620,784 in 2017

Top accessed features

- Molecule of the Month (hemoglobin, catalase, GFP, carbonic anhydrase)
- Guide to Understanding PDB Data
- Paper Models
- Content Browser



Users by Country



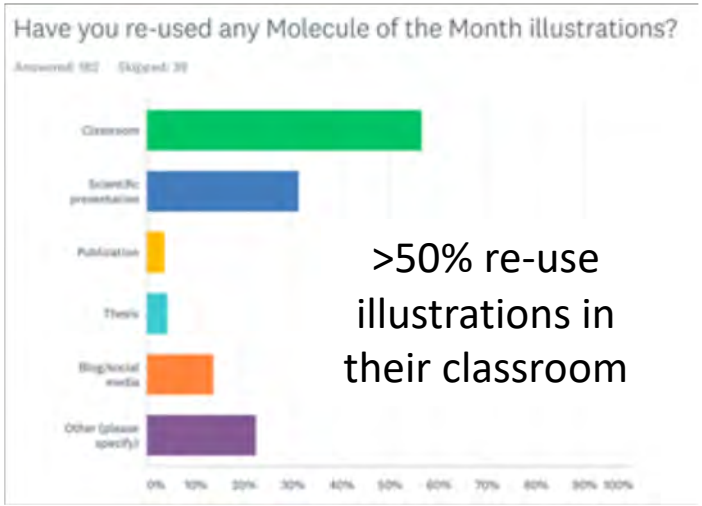
Users by State

80% Desktop,
20% mobile or tablet

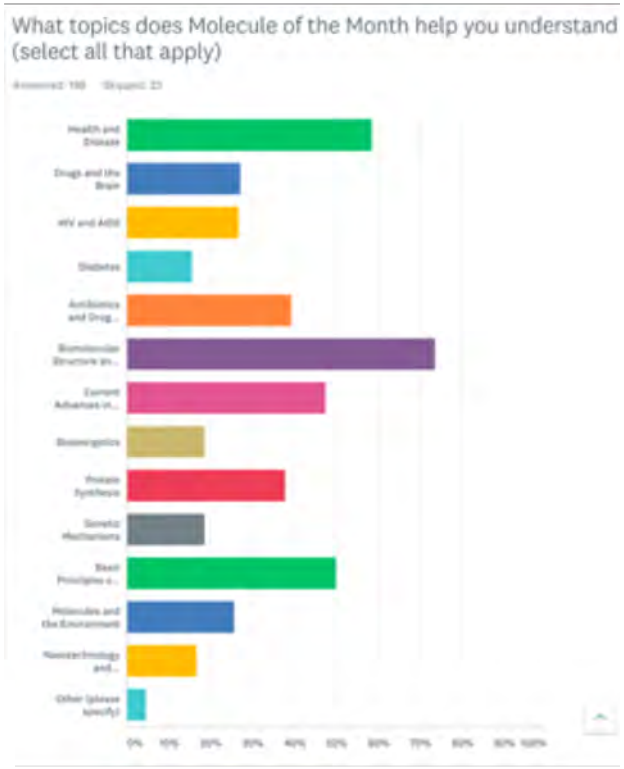
Molecule of the Month Survey

2020: 20 Years of Molecule of the Month

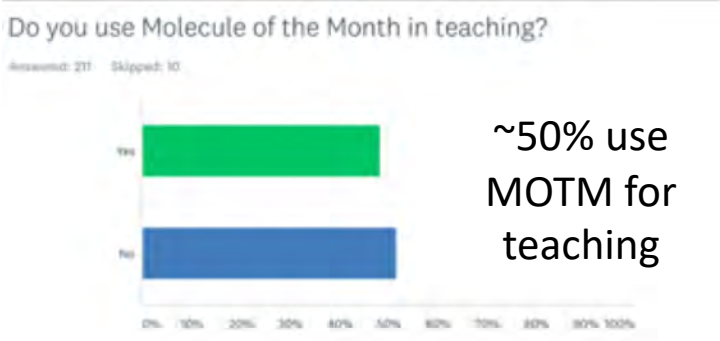
Based on ~240 Responses as of April 30



>50% re-use illustrations in their classroom



>60% located at a College/University



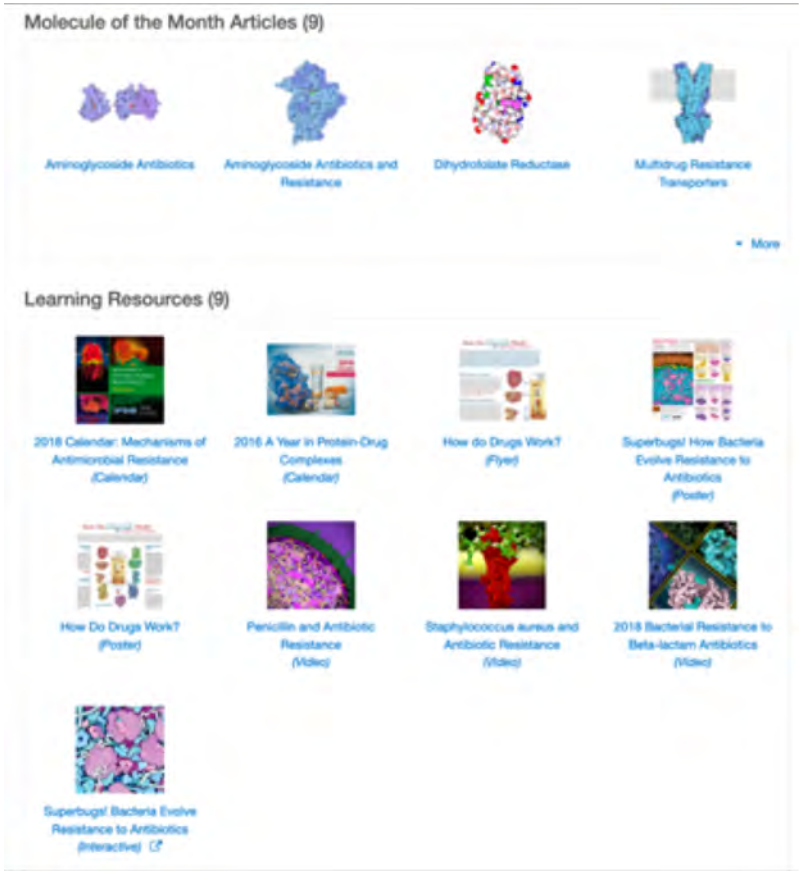
~50% use MOTM for teaching

MOTM helps ≥50% understand
Health and Disease
Biomolecular Structure and Function
Basic Principles of Molecular Biology

Public Health Focus: Antimicrobial Resistance (2018-2019)

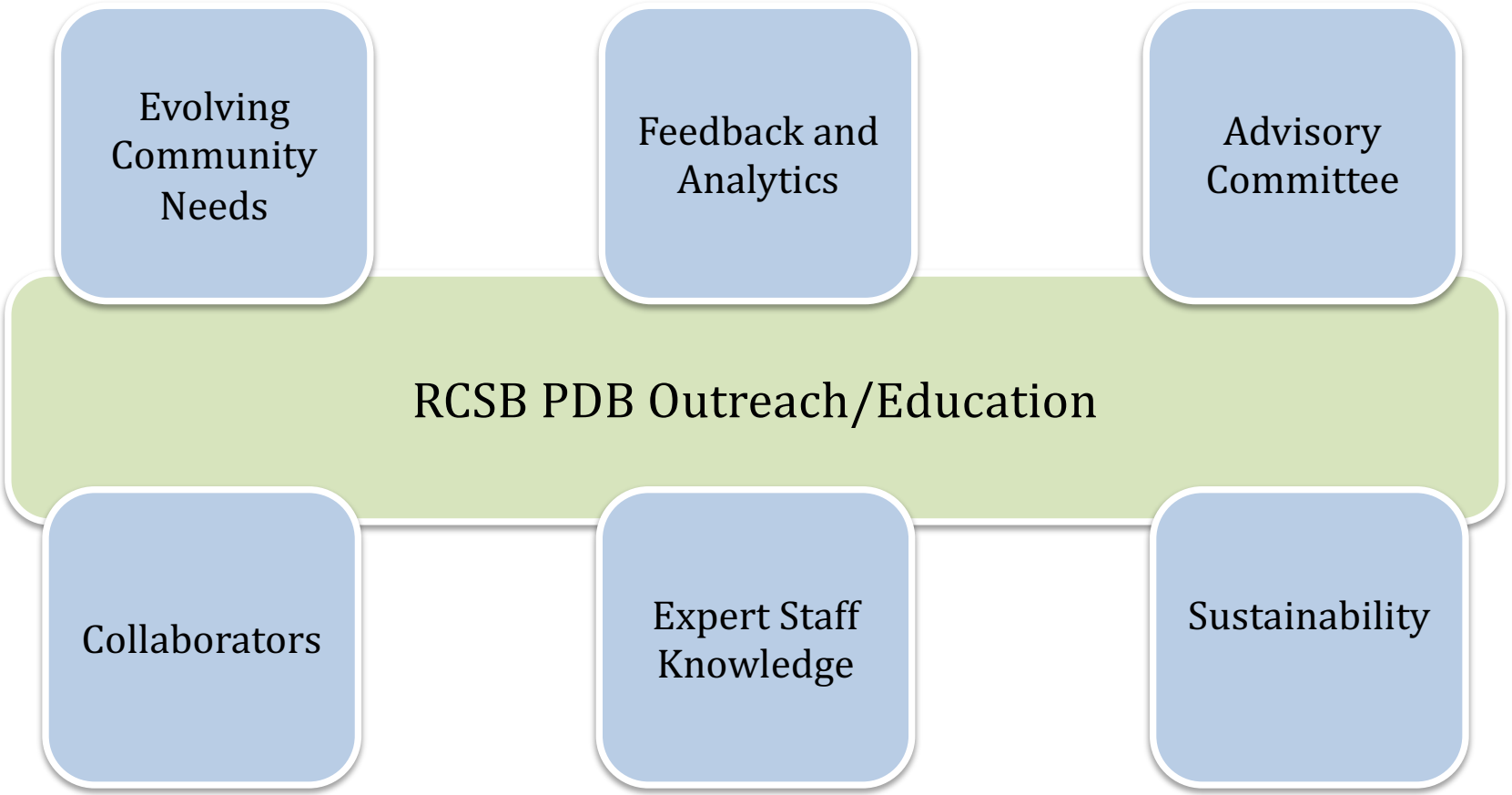
- PDB-101
  - Molecule of the Month
  - Videos
  - 2018 calendar
  - Interactive Poster
- Undergraduate course
- Global Health resource development
- High school Video Challenge

Next focus: Drugs and the Brain



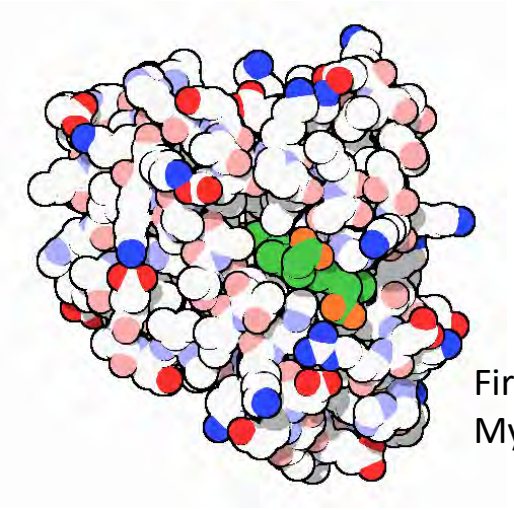
High School Video Challenge Winners

Factors Influencing Decision Making

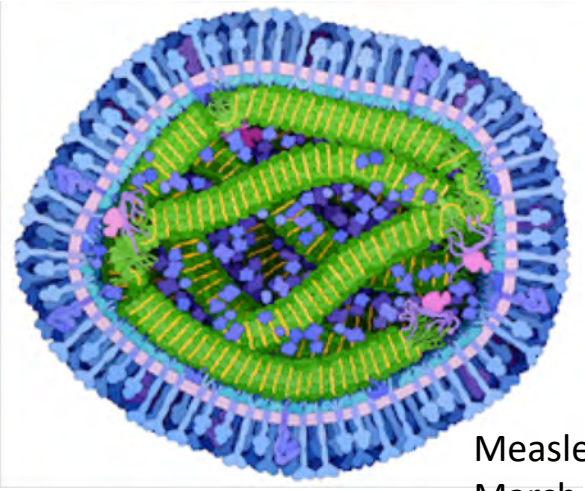


2019 Goals: Growth and Preparation

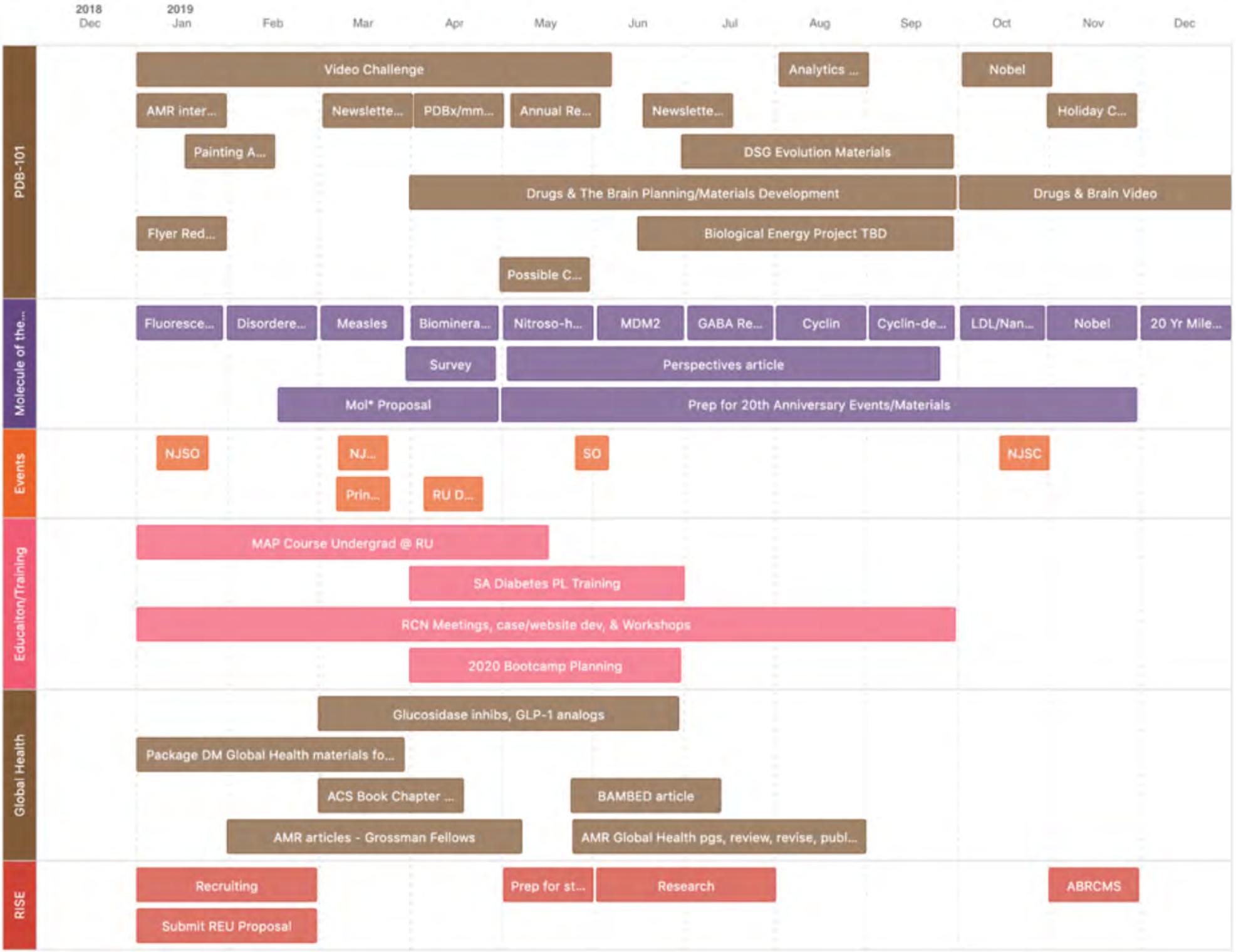
- New PDB-101 materials to drive traffic (users and visits)
  - Global Health: diabetes, AMR
  - Curricular materials
  - Molecule of the Month articles on Fundamental Biology, Biomedicine, and Energy
- Support RCSB.org development with training materials
- Maintain PDB-101 uptime
- Plan for the future
  - Develop materials for Drugs and the Brain health focus (2020-2021)
  - Initial PDB50 discussions
  - Plan materials and events to leverage 20 years of Molecule of the Month



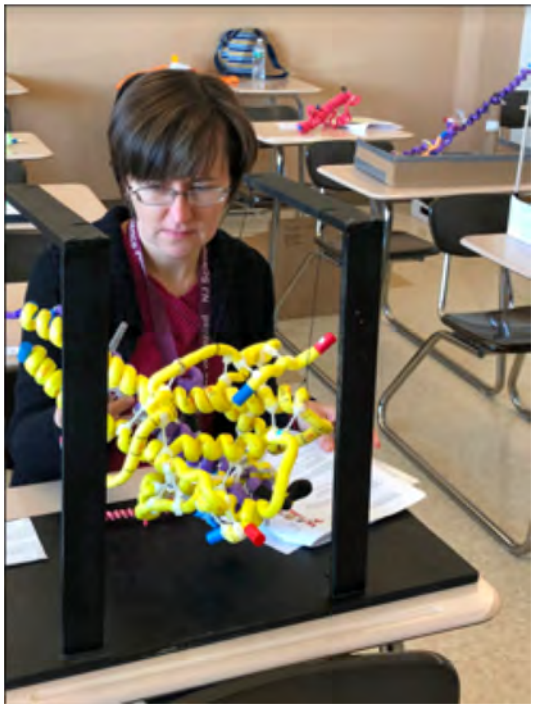
First Molecule of the Month: Myoglobin, January 2000

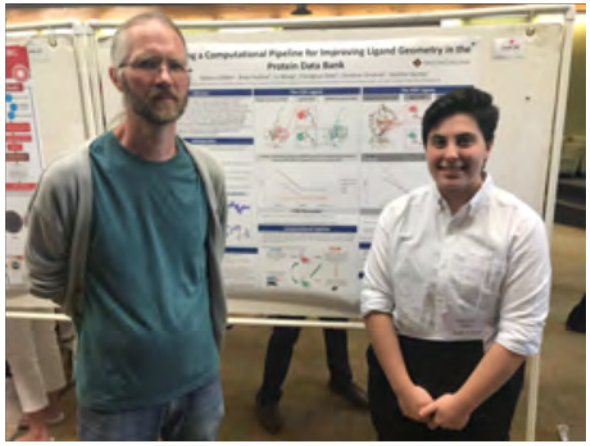


Measles Virus Proteins, March 2019

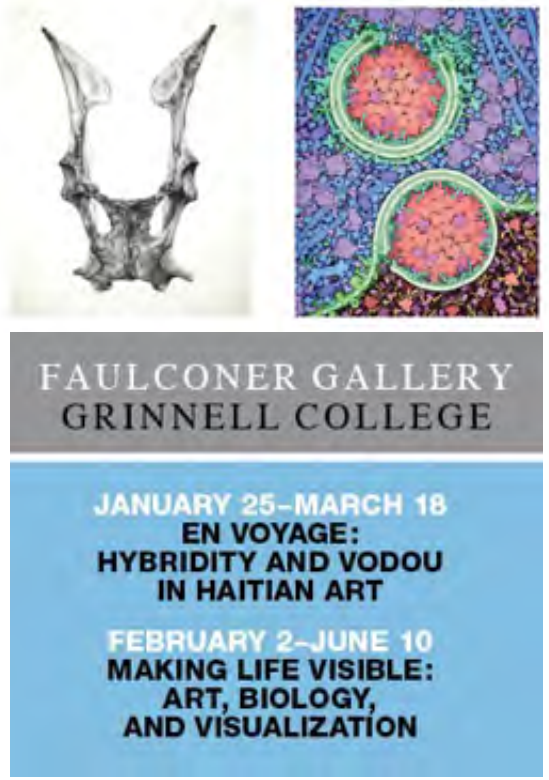

2018
Dec
2019
Jan
Feb
Mar
Apr
May
Jun
Jul
Aug
Sep
Oct
Nov
Dec
PDB-101
Video Challenge
Analytics ...
Nobel
AMR inter...
Newslette...
PDBx/mm...
Annual Re...
Newslette...
Holiday C...
Painting A...
DSG Evolution Materials
Drugs & The Brain Planning/Materials Development
Drugs & Brain Video
Flyer Red...
Biological Energy Project TBD
Possible C...
Molecule of the...
Fluoresce...
Disordere...
Measles
Biominera...
Nitroso-h...
MDM2
GABA Re...
Cyclin
Cyclin-de...
LDL/Nan...
Nobel
20 Yr Mile...
Survey
Perspectives article
Mol* Proposal
Prep for 20th Anniversary Events/Materials
Events
NJSO
NJ...
SO
NJSC
Prin...
RU D...
Educaiton/Training
MAP Course Undergrad @ RU
SA Diabetes PL Training
RCN Meetings, case/website dev, & Workshops
2020 Bootcamp Planning
Global Health
Glucosidase inhibs, GLP-1 analogs
Package DM Global Health materials fo...
ACS Book Chapter ...
BAMBED article
AMR articles - Grossman Fellows
AMR Global Health pgs, review, revise, publ...
RISE
Recruiting
Prep for st...
Research
ABRCMS
Submit REU Proposal

Outreach Depends Upon Everyone





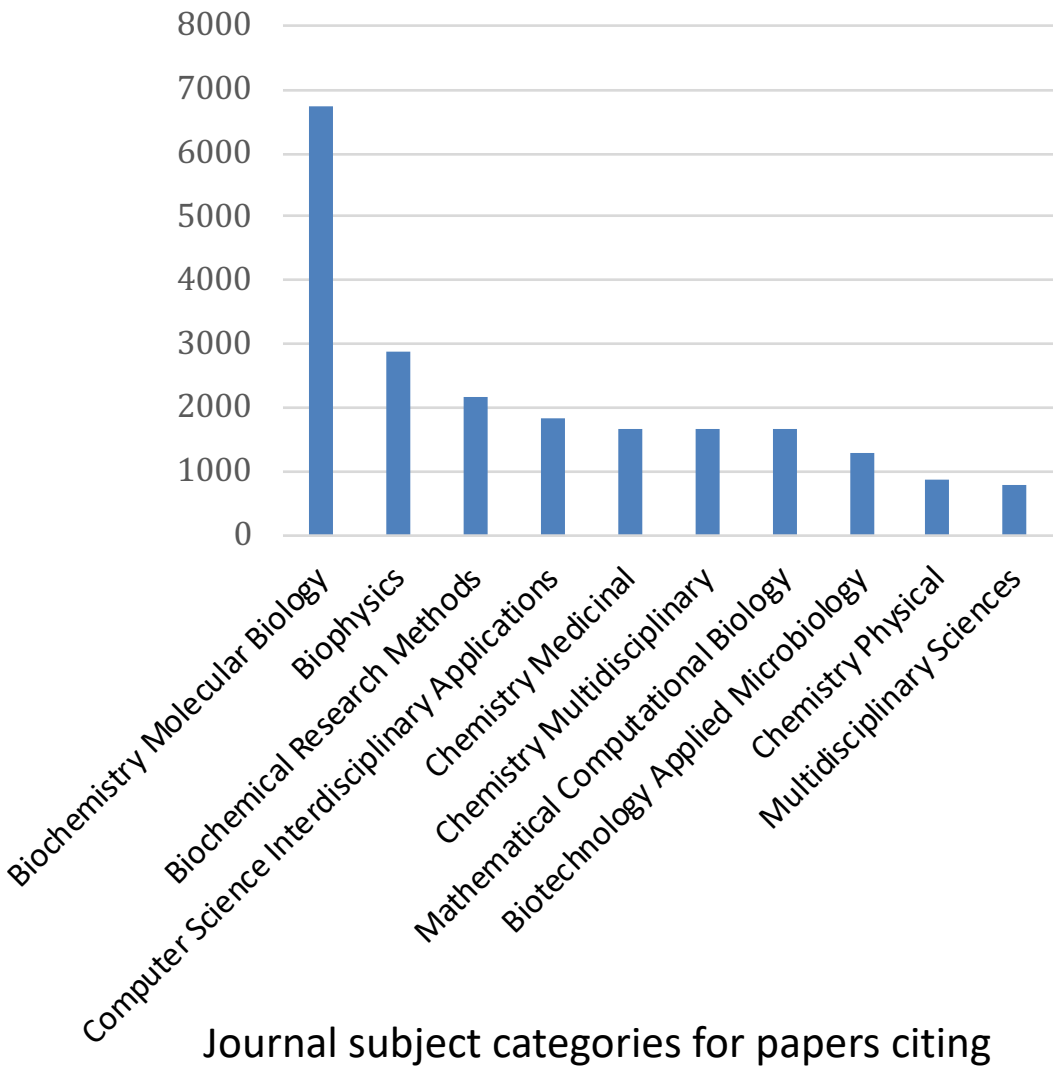
2018: Select Offline Highlights



Growing the Data Consumer Community

Communities Currently Served

- Millions of users visit RCSB.org
  - Increased 10% in 2018
  - Estimated 3.5 million unique users in 2018
  - Unable to directly track research interests
- >400 resources utilize PDB data
- ~19,000 publications cite inaugural RCSB PDB publication (Berman *et al.* 2000)
  - Predominately biology-biomedicine-chemistry



Journal subject categories for papers citing Berman *et al*. 2000

RCSB.org Integration with Key Resources

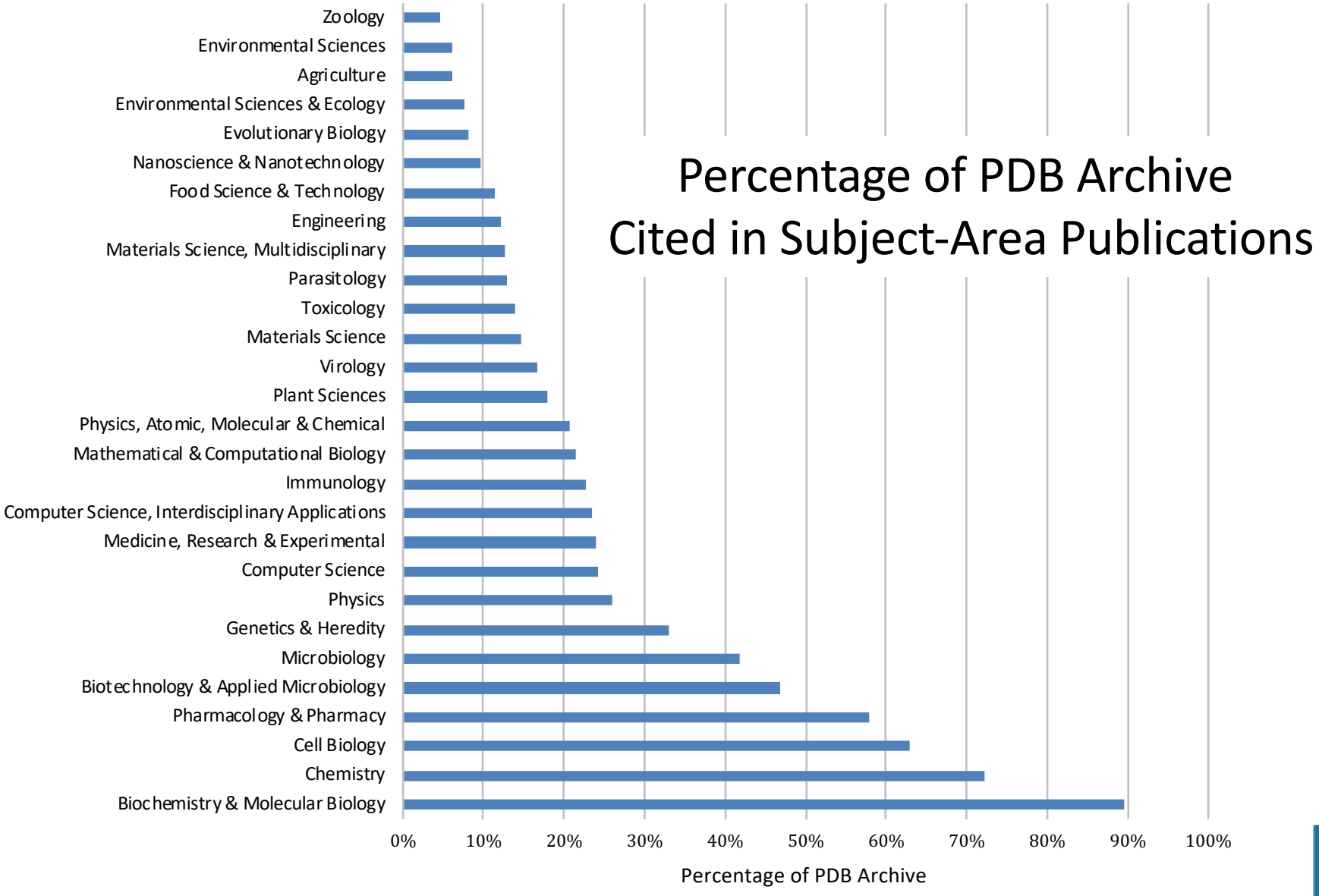
Resource	Type of Data	Resource	Type of Data	Resource	Type of Data
BindingDB	Binding affinities	Immune Epitope Database	Antibody and T cell epitopes	Protein Model Portal	Theoretical models
Binding MOAD	Binding affinities	LS-SNP	Single Nucleotide Polymorphisms	PubMed	Citation information
BiGG	Reconstruction of metabolic pathways	mpstruc	Classification of transmembrane protein structures	PubMedCentral	Open access literature
BMRB	BMRB-to-PDB mappings	NCBI Gene	Gene info, reference sequences, et al.	Recon3D	A 3-Dimensional View of Human Metabolism and Disease
Catalytic Site Atlas	Enzyme active sites and catalytic residues	NCBI Taxonomy	Organism Classification	RECOORD	NMR structure ensembles
CATH	Protein structure classification	NDB	Experimentally-determined nucleic acids and complex assemblies	RESID	Protein modifications
DrugBank	Drug and target data	OLDERADO	NMR domain composition and clustering	SBGrid	diffraction images
EMDB	3DEM density maps and associated metadata	OPM	Orientation of transmembrane proteins	SCOP	Protein structure classification
ExPASy	Enzyme classification	PDBbind-CN	Binding affinities	SIFTS (PDBe)	Structure, function, taxonomy, sequence
Gencode	Gene structure data	PDBflex	Protein structure flexibility	Store.Synchrotron Data Store	diffraction images
Gene Ontology	Biological ontologies	Pfam	Protein families	TCDB	membrane transport protein classification
HMMER3	Sequence similarity searches	PhospoSitePlus	Mammalian post-translational modifications	UniProt	Protein sequences and annotations
Human Gene Nomenclature Committee	nomenclature and genomic information	ProteinDiffraction.org	diffraction images	UCSC genome browser	human genome data

http://www.rcsb.org/pages/external-resources

Supporting and Growing Established User Communities

- Can we grow user community by
  - Improving RCSB.org tools
  - Building new tools
  - Integrating with comparative protein models
  - Integrating with additional data resources
    - PubChem, CARD, Model Archive, …
  - Develop new training materials

PDB Data Impact on Scientific Literature



Recruiting New User Communities

- Are there roadblocks to using RCSB.org?
  - Utility of 3D data for research not clear?
  - Barriers to utilizing RCSB.org tools?

- Opportunities for future growth to consider?
  - New tools to develop
  - Integration with new resources
    - PubChem, CARD, Model Archive, …
  - Training materials
  - Collaborations with scientific societies
  - …

- Should we expand the current Advisory Committee membership?

Thank you for your contributions