

# RCSB Protein Data Bank Advisory Committee

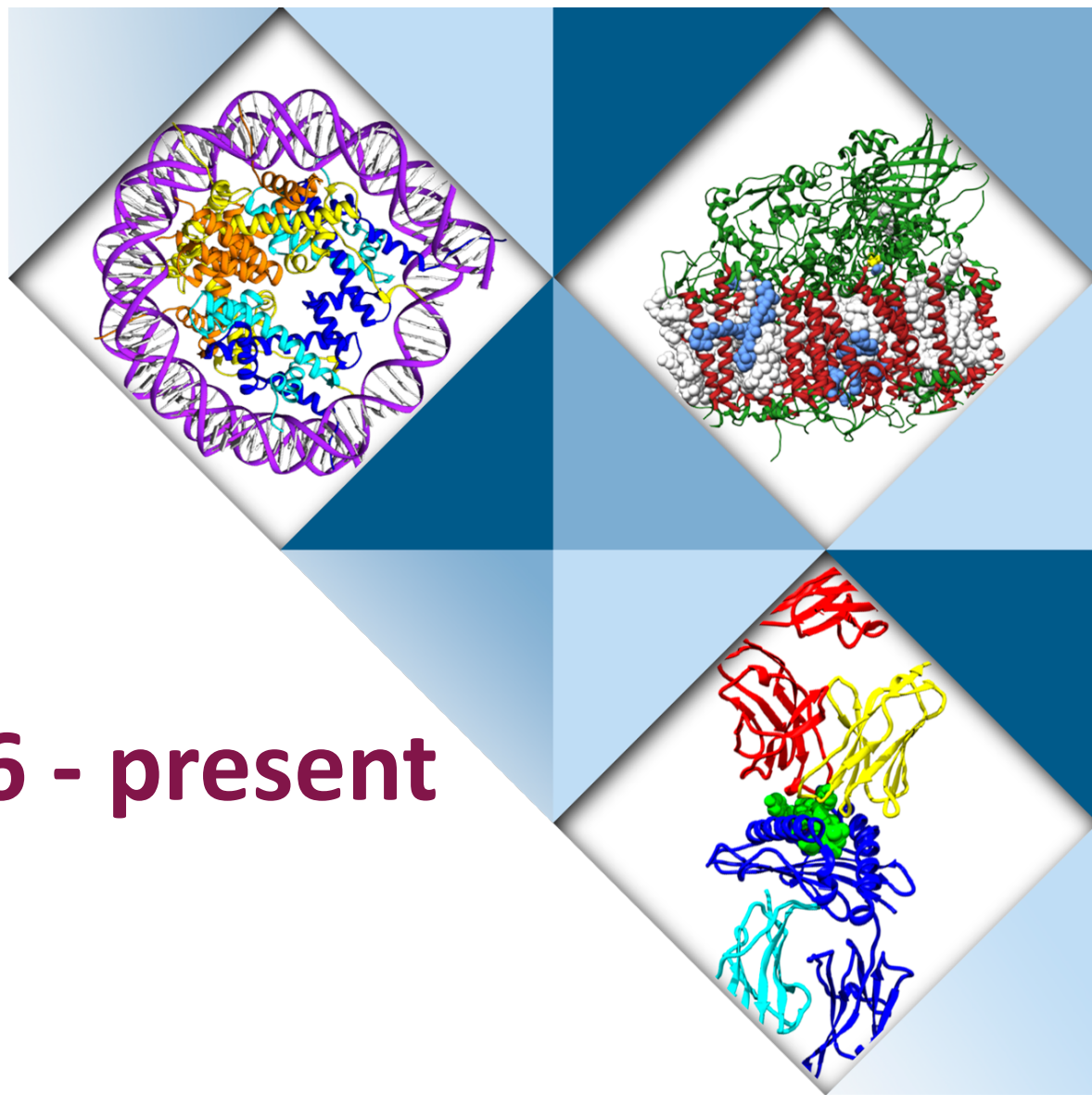
Teleconference  
October 19, 2017

# Today's Agenda

1:00pm ET	Highlights, 2016 – present	Stephen Burley
1:30	Response to 2017 Site Visit Report	Stephen Burley
2:00	Questions for the RCSB PDB AC	
2:30	Executive Session	

# Meeting Participants

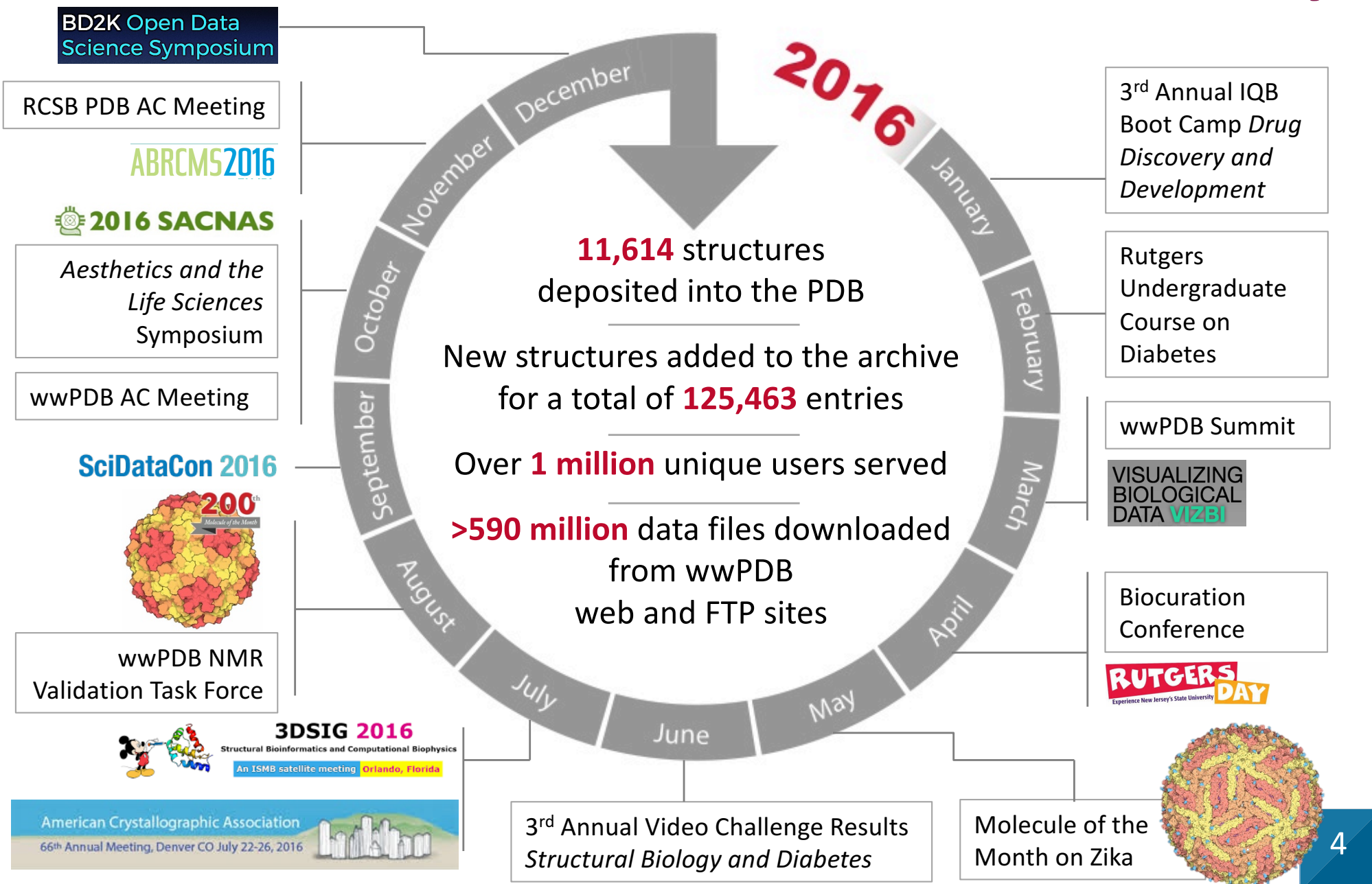
- Advisory Committee
  - Participants: Cynthia Wolberger (Chair), R. Andrew Byrd, Bridget Carragher, Wah Chiu, Kirk Clark, Paul Craig, Roland Dunbrack, Tom Ferrin, Cathy Peishoff, Sue Rhee, Andrej Sali, Torsten Schwede, Jill Trehwella
  - Not participating: Paul Adams
- RCSB PDB
  - Rutgers: Stephen K. Burley, Helen M. Berman, Zukang Feng, John Westbrook, Jasmine Young, Christine Zardecki
  - UCSD: Cole H. Christie, Jose Duarte, Tara Kalro



# Highlights: 2016 - present



# Year in the Life of the RCSB PDB Community



# Responses to 2016 AC Report

<p><i>(AC) hopes that the release of a pre-deposition server will encourage depositors to check coordinates</i></p>	<p>Link to server added to OneDep interface</p> <p>Major issues made more visible during deposition</p> <p>Pre-deposition server usage increased from 2500 -&gt; 3500 runs/month</p> <p>2015 Data Reloads: 29%</p> <p>2016 Data Reloads: 25%</p>
<p><i>Assemble comprehensive metrics that show PDB utility and impact on NSF, NIH and DOE research</i></p>	<p>Highlights in AC Google impact folder</p> <ul style="list-style-type: none"><li>• <i>Protein Science</i> preprint</li><li>• Rutgers Economic Report</li><li>• Clarivate Bibliometric Report</li></ul>

# Responses to 2016 AC Report

<i>Feature NSF, DoE research at PDB-101 and Molecule of the Month</i>	<p>XFEL highlighted in</p> <ul style="list-style-type: none"><li>• Photoactive Yellow Protein (March)</li><li>• Adenine Riboswitch (June)</li></ul> <p>Globin Evolution highlighted in February</p>
<i>Prioritize applications for additional funding</i>	<p>Proposals focused on I/HM and educational efforts</p>
<i>AC gave the (HIV Film) preview a unanimous thumbs up</i>	<p>Private screenings in Los Angeles, CA and Piscataway, NJ</p> <p>Currently on the Festival Circuit <a href="http://targetzerofilm.org/">http://targetzerofilm.org/</a></p>

# Competing for Grants to Support Development

<b>Submissions 2013 -2017</b>	<b>Number Submitted</b>	<b>Number Funded</b>	<b>Number In Review</b>	<b>Total Amount Awarded</b>
Data In	3	2	0	619,121
Data Out	4	3	0	1,674,744
Outreach/ Education/ REUs	13	8	0	350,215
Infrastructure	2	2	0	285,000
Nucleic Acid Database Collaboration	1	1	0	615,419
EM Collaboration	1	1	0	TBD
I/HM Collaboration	6	1	0	300,000
<b>Total</b>	<b>30</b>	<b>18</b>	<b>0</b>	<b>3,844,499</b>

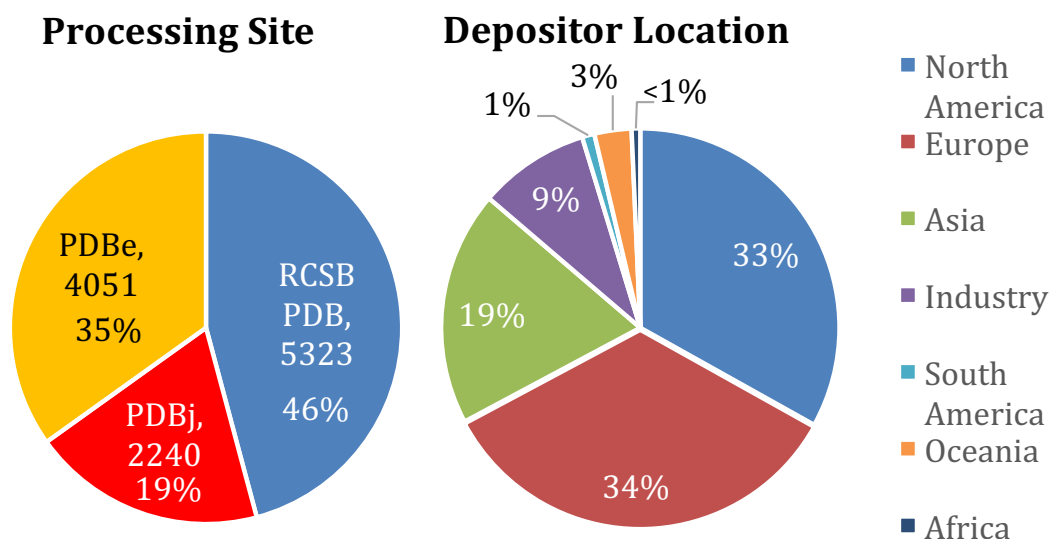
# 2016 Data In/Archive Keeping Statistics

- On track for >13,000 depositions in 2017
- More EM structures submitted in 2016 than NMR structures

Method	2016 Depositions	2015 Depositions
X-ray	10583 (91%)	10167
NMR	473 (4%)	510
EM	531 (4.6%)	255
Other	27 (0.2%)	25

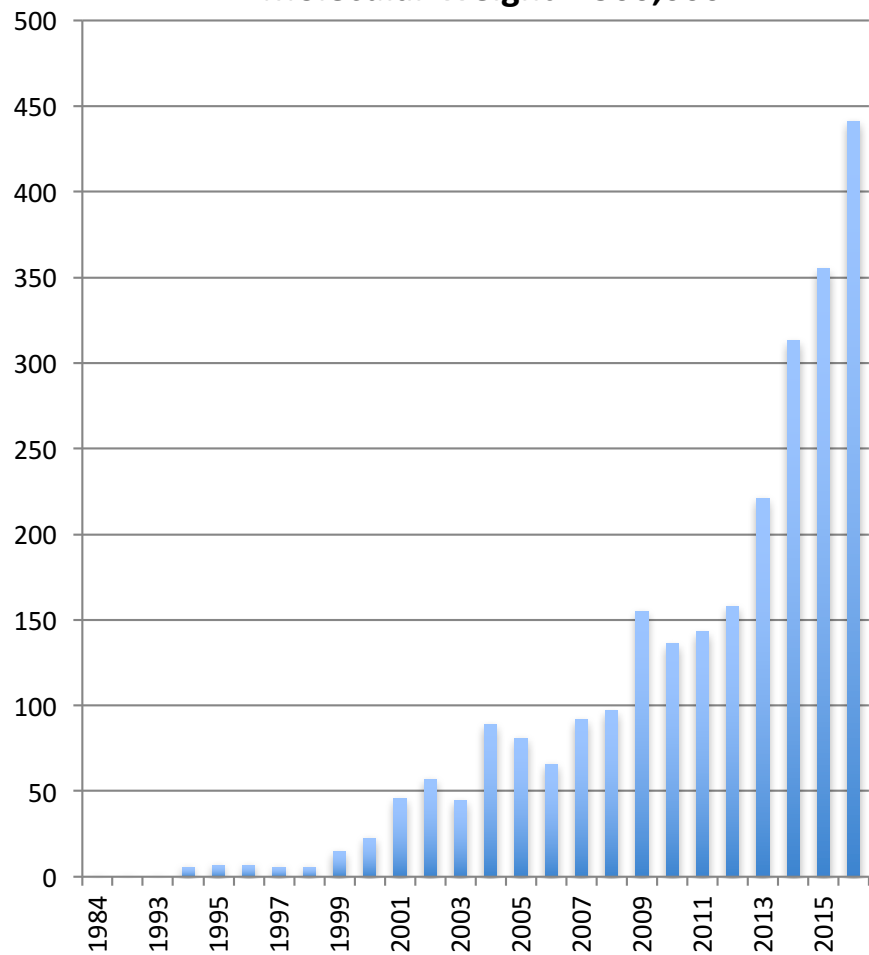
- Increase of structures from XFEL

- 2013: 5
- 2014: 30
- 2015: 45
- 2016: 66

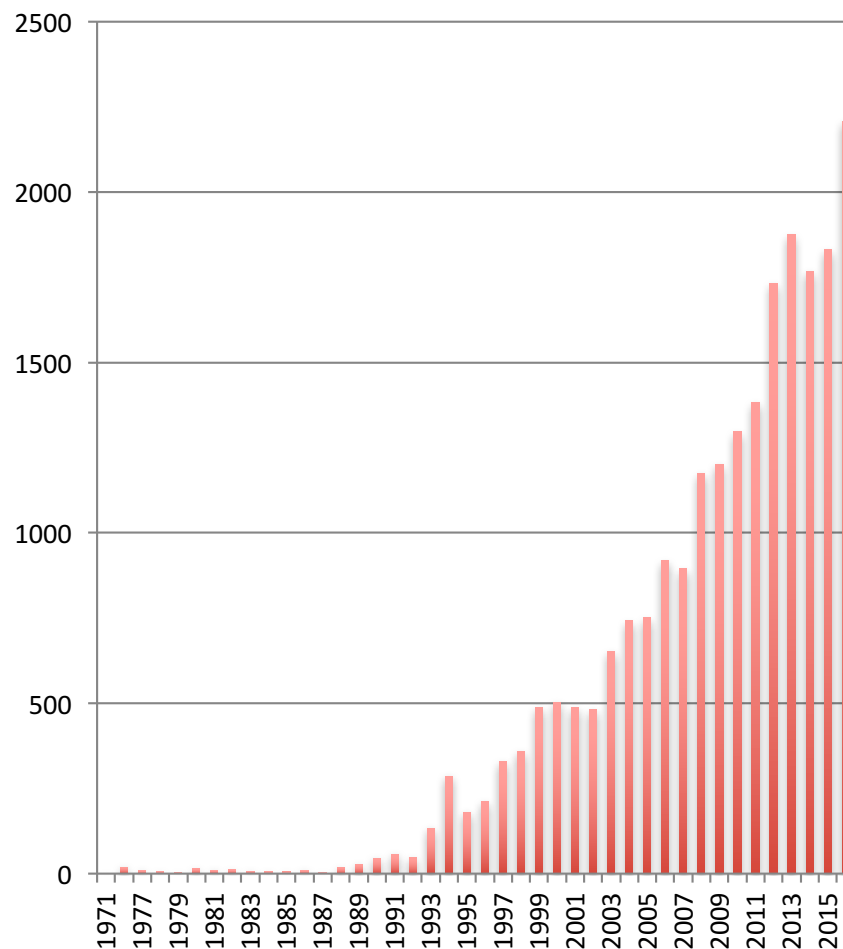


# Complexity Continues to Increase

**Annual Releases with Asymmetric Unit  
Molecular Weight > 500,000**

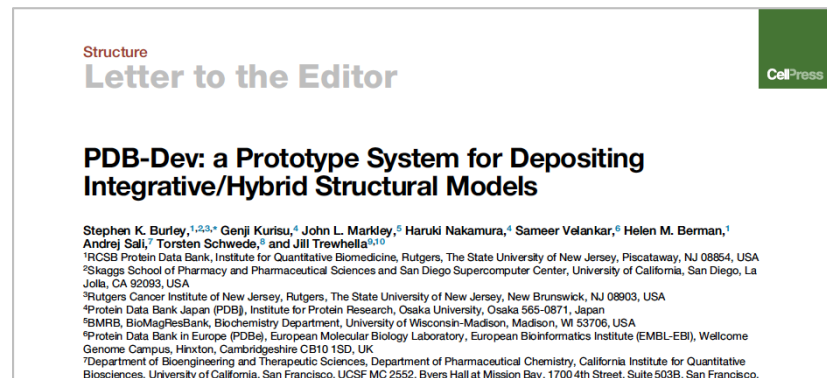


**New and Unique Ligands Added to  
Chemical Component Dictionary Annually**

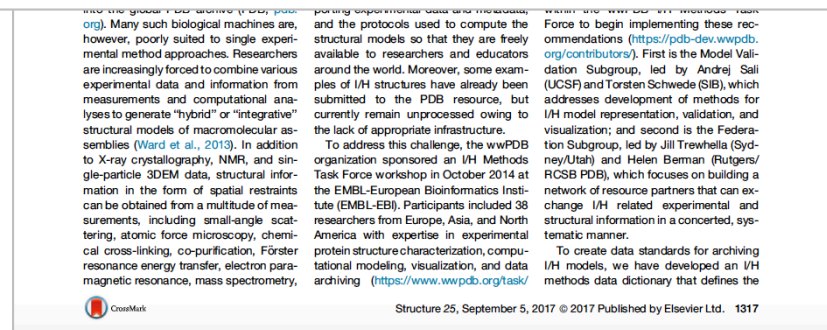


# PDB-Dev: A Prototype System for Integrative/Hybrid Structural Models

- Prototype system for depositing I/H structural models
- Announced September 5, 2017



The screenshot shows the top section of the PDB-Dev website. On the left, it says "Welcome to PDB-Dev". On the right, the URL "pdb-dev.wwpdb.org" is displayed in a large, bold font. Below the URL, there is a paragraph of text explaining that PDB-Dev is a prototype deposition and archiving system for structural models obtained through integrative/hybrid (I/H) methods. At the bottom of this section, there is a red-bordered box containing the text: "Deposit new structures [here](#). More information of how to deposit structures can be found in the [FAQ](#) section."







# Recent *Data In/Archive Keeping* Highlights

- Improved collection experimental support for assemblies
- Improved ligand data collection and validation
- Improved Archival content and management
- Established third data storage site

## Experimental evidence for the assembly

Experiments performed to support the assembly

- SAXS
- assay for oligomerization
- cross-linking
- equilibrium centrifugation
- gel filtration
- homology
- immunoprecipitation
- isothermal titration calorimetry
- light scattering
- mass spectrometry
- microscopy
- native gel electrophoresis
- none
- scanning transmission electron microscopy

R1-1 R1-2 R1-3 T1  
R2-1 R2-2 R2-3 T2  
R3-1 R3-2 R3-3 T3



Additional information about the assembly



N  Y

ing matrices?:  N  Y

in the format:



# 2016 *Data Out* at a Glance

## RCSB.org Users

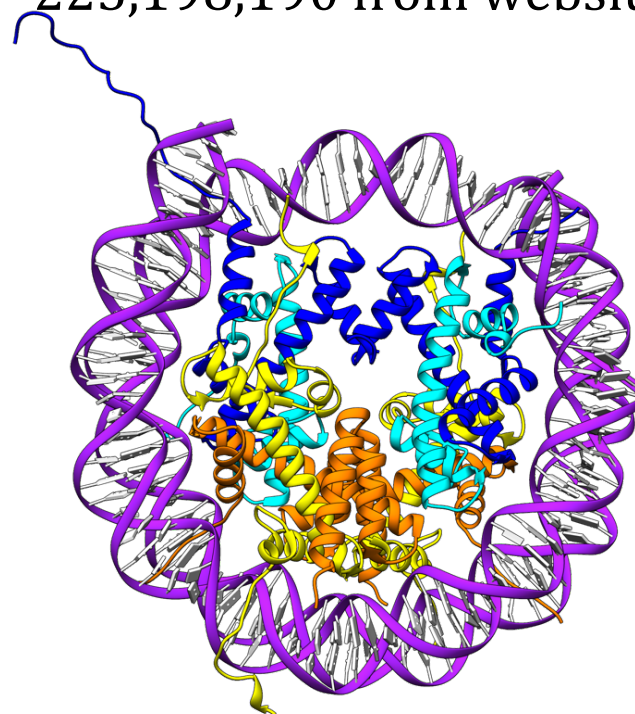
- >395,000 monthly, >1 million annually
- 3% annual growth in non-bounce unique users

## Sessions at RCSB.org

- 35% growth since 2010
- High average session duration (~6 minutes)
- Low fraction of 0-second “bounce” sessions

## wwPDB Data Downloads

- 591,876,087 total
  - 366,677,897 from FTP sites
  - 225,198,190 from websites



Nucleosome (PDB 1aai)  
Luger *et al.* (1997) *Nature* 389, 251-260

Frequently access structure—Structure data downloaded  
~257K times since 2007

# Data Downloaded Annually

Year	Total	Total FTP Archive	Total Website	RCSB PDB Total	PDBe Total	PDBj Total
2009	328,362,536	271,116,934	57,245,602	276,492,545	31,616,455	20,253,536
2010	294,326,976	213,180,966	81,146,010	223,817,872	48,400,568	22,108,536
2011	383,131,048	276,952,286	106,178,762	286,499,504	59,475,613	37,155,931
2012	376,944,070	255,837,735	121,106,335	303,948,848	45,583,904	27,411,318
2013	441,262,210	296,176,290	145,085,920	312,881,488	81,447,346	46,933,376
2014	512,227,251	339,193,721	173,033,530	347,283,931	100,393,784	64,549,536
2015	534,339,871	368,244,766	166,095,105	367,149,527	89,671,549	77,518,795
<b>2016</b>	<b>591,876,087</b>	<b>366,677,897</b>	<b>225,198,190</b>	<b>454,856,822</b>	<b>74,707,114</b>	<b>62,312,151</b>

# Recent *Data Out* Highlights

- All services running in our private cloud
- Qualitative website improvements
  - User Interface design
  - Search algorithms
- New modular architecture, including new REST Web Services
- Now responsible for EPPIC (Evolutionary Protein-Protein Interface Classifier, [eppic-web.org](http://eppic-web.org))
  - New funding opportunity

Search

Currently showing 1 - 13 of 13 Displaying 25 Results

View:  Reports:  Sort:

**1IAX**  
**CRYSTAL STR**  
[Huai, Q., Xia, Y., C...](#)  
(2001) J Biol Chem  
**Released:** 4/4/2001  
**Method:** X-ray Diffraction  
**Resolution:** 2.8 Å  
**Residue Count:** 856

**Unique Ligands:** PLP, SO4  
**Search term match score:** 469.21

**Matched fields in 1IAX.cif:**

- **\_citation.title:** Crystal structures of 1-aminocyclopropane-1-carboxylate (ACC) synthase in complex with aminoethoxyvinylglycine and pyridoxal-5'-phosphate provide new insight into catalytic mechanisms
- **\_entity.pdbx\_description:** 1-AMINOCYCLOPROPANE-1-CARBOXYLATE SYNTHASE 2, SULFATE ION, PYRIDOXAL-5'-PHOSPHATE
- **\_entity\_name\_com.name:** ACC SYNTHASE 2
- **\_struct.title:** CRYSTAL STRUCTURE OF ACC SYNTHASE COMPLEXED WITH PLP

# Recent Team Advancements and Transitions

- Data In
  - Promotions to Biocuration Leadership positions (Irina Periskova, Yuhe Liang)
- Archive Keeping
  - Promotion to Infrastructure Team Lead (Vladimir Guranović)
- Data Out
  - Promotion to UCSD Site Manager (Cole Christie)
  - Promotion to Scientific Team Lead (Jose Duarte)
  - Promotion to Derived Data Team Lead (Tara Kalro)
  - Promotion to Front End Team Lead (Alex Rose)
- Transitions
  - Peter Rose and Andreas Prlic exited
    - Ongoing recruitment for Senior Scientist, postdocs at UCSD
  - 2 layoffs announced (Rutgers), effective 12.31.17



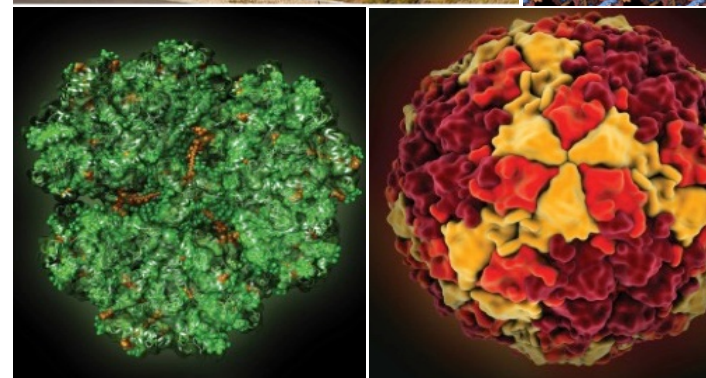
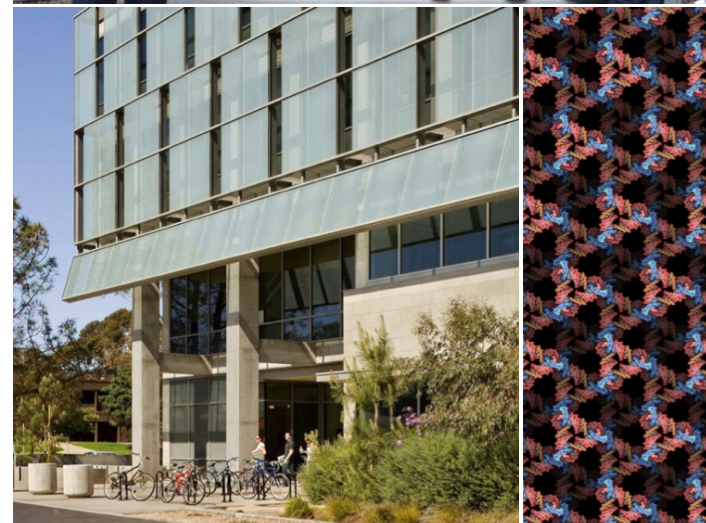
# Join the RCSB Protein Data Bank Team at the University of California San Diego

Open Positions:

## Scientific Software Developer Postdoctoral Fellows

### The Challenge:






Develop innovative analysis, integration, query, and visualization tools for 3D biomolecular structures to help accelerate research and training in biology, medicine, and related disciplines.





# Matrix Team of Expert Developers

## Builds and Maintains RCSB PDB and wwPDB Tools

	Roadmap Leadership Team	Functional Teams			
Project Coordination Role		Infrastructure 	Primary Data 	Derived Data 	Front End 
Team Lead Coordinator	Christine	Vladimir	Jose	Tara	Alex
Architecture Coordinator	John	Chris	Zukang John	Wendy	Jesse
Plan Coordinator	Cole	Harry*	Chenghua	Ezra Huanwang	Raul
Dependency Coordinator	Jasmine	Li, Ken*	Chunxiao	Yana*	Rob

\*supported by non-PDB funding

# Recognized Educational Impact




Announced March 2017

**GEN** Best of the Web

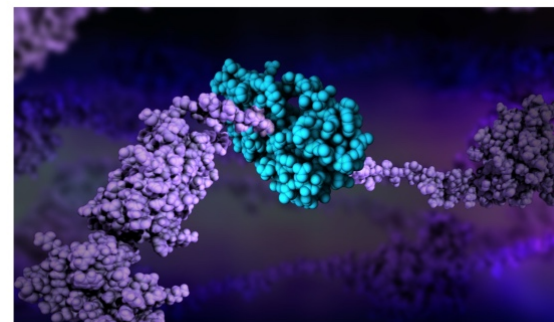
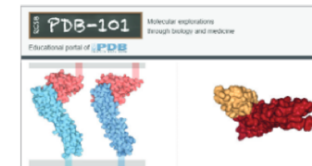
PDB-101 ★★★★★

Feb 15, 2017

[pdb101.rcsb.org](http://pdb101.rcsb.org)

 Beautiful protein structures, wealth of educational resources

Most biologists are probably familiar with the Protein Data Bank (PDB). However, fewer are likely aware of the accompanying educational portal, PDB-101. Let's remedy that, since PDB-101 is a fantastic (and incredibly informative) website that is sure to pique the interests of both experienced structural biologists and students. "Molecular explorations through biology and medicine," boasts the tagline of the site, and in fact, there is much to explore here. Front-and-center on the homepage is the "Molecule of the Month," and alongside that is a navigation bar that invites visitors to browse resources by category. There are four categories in total: health and disease, molecules of life, biotech and nanotech, and structures and structure determination. Alternatively, visitors can browse content either by student resources (under the "learn" tab of the menu bar) or teacher resources (under the "teach" tab). Learning resources include paper models, posters, and interactive animations, while teaching resources include three entire curriculum modules.



*HIV Enzymes*  
FASEB BioArt Winner



# wwPDB AC Meeting October 13, 2017



# Any Questions About Recent Milestones?



# Planning Ahead for PDB 2021



**CRYSTALLOGRAPHY**

## Protein Data Bank

A repository system for protein crystallographic data will be operated jointly by the Crystallographic Data Centre, Cambridge, and the Brookhaven National Laboratory. The system will be responsible for storing atomic coordinates, structure factors and electron density maps and will make these data available on request. Distribution will be on magnetic tape in machine-readable form whenever possible. There will be no charge for the service other than handling costs. Files will be updated as new material is received. The total holding will be announced annually in the organic bibliographic volumes of the reference series "Molecular Structures and Dimensions" published for the Crystallographic Data Centre and the International Union of Crystallography by Oosthoek's, Utrecht.

The success of the proposed system will depend on the response of the protein crystallographers supplying data. These will be accepted either "raw" or refined, in machine-readable form or as manuscripts. Laboratories intending to join the scheme should communicate with Mrs Olga Kennard or Dr D. G. Watson at the University Chemical Laboratories, Lensfield Road, Cambridge, who are responsible for the organization of the system. Data can be submitted to Cambridge or to Dr W. C. Hamilton at the Brookhaven National Laboratory, Upton, New York 11973, where the data will be computer processed.

The two centres will maintain identical files and both will provide data services. The new data bank is intended to supplement existing publication media so that depositing material in this form is not a substitute for the publication of the results of structural investigations in a scientific journal.

October 20, 1971  
*Nature New Biology*