RCSB Protein Data Bank
Advisory Committee
Teleconference
October 19, 2017
## Today’s Agenda

<table>
<thead>
<tr>
<th>Time</th>
<th>Session</th>
<th>Speaker</th>
</tr>
</thead>
<tbody>
<tr>
<td>1:00pm ET</td>
<td>Highlights, 2016 – present</td>
<td>Stephen Burley</td>
</tr>
<tr>
<td>1:30</td>
<td>Response to 2017 Site Visit Report</td>
<td>Stephen Burley</td>
</tr>
<tr>
<td>2:00</td>
<td>Questions for the RCSB PDB AC</td>
<td></td>
</tr>
<tr>
<td>2:30</td>
<td>Executive Session</td>
<td></td>
</tr>
</tbody>
</table>
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 Trewhella
  - 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
11,614 structures deposited into the PDB

New structures added to the archive for a total of 125,463 entries

Over 1 million unique users served

>590 million data files downloaded from wwPDB web and FTP sites
## Responses to 2016 AC Report

| (AC) hopes that the release of a pre-deposition server will encourage depositors to check coordinates | Link to server added to OneDep interface  
| Major issues made more visible during deposition  
| Pre-deposition server usage increased from 2500 -> 3500 runs/month  
| 2015 Data Reloads: 29%  
| 2016 Data Reloads: 25%  
| Assemble comprehensive metrics that show PDB utility and impact on NSF, NIH and DOE research | Highlights in AC Google impact folder  
| • Protein Science preprint  
| • Rutgers Economic Report  
| • Clarivate Bibliometric Report |
## Responses to 2016 AC Report

| Feature NSF, DoE research at PDB-101 and Molecule of the Month | XFEL highlighted in  
| --- | --- |
|  | • Photoactive Yellow Protein (March)  
|  | • Adenine Riboswitch (June)  
|  | Globin Evolution highlighted in February  

| Prioritize applications for additional funding | Proposals focused on I/HM and educational efforts  
| --- | --- |

| AC gave the (HIV Film) preview a unanimous thumbs up | Private screenings in Los Angeles, CA and Piscataway, NJ  
| --- | --- |
|  | Currently on the Festival Circuit  
|  | http://targetzerofilm.org/  

---

[61x511]Responses to 2016 AC Report

Feature NSF, DoE research at PDB-101 and Molecule of the Month

XFEL highlighted in

• Photoactive Yellow Protein (March)

• Adenine Riboswitch (June)

Globin Evolution highlighted in February

Prioritize applications for additional funding

Proposals focused on I/HM and educational efforts

AC gave the (HIV Film) preview a unanimous thumbs up

Private screenings in Los Angeles, CA and Piscataway, NJ

Currently on the Festival Circuit

http://targetzerofilm.org/
## Competing for Grants to Support Development

<table>
<thead>
<tr>
<th>Submissions 2013 –2017</th>
<th>Number Submitted</th>
<th>Number Funded</th>
<th>Number In Review</th>
<th>Total Amount Awarded</th>
</tr>
</thead>
<tbody>
<tr>
<td>Data In</td>
<td>3</td>
<td>2</td>
<td>0</td>
<td>619,121</td>
</tr>
<tr>
<td>Data Out</td>
<td>4</td>
<td>3</td>
<td>0</td>
<td>1,674,744</td>
</tr>
<tr>
<td>Outreach/Education/REUs</td>
<td>13</td>
<td>8</td>
<td>0</td>
<td>350,215</td>
</tr>
<tr>
<td>Infrastructure</td>
<td>2</td>
<td>2</td>
<td>0</td>
<td>285,000</td>
</tr>
<tr>
<td>Nucleic Acid Database</td>
<td>1</td>
<td>1</td>
<td>0</td>
<td>615,419</td>
</tr>
<tr>
<td>Collaboration</td>
<td></td>
<td></td>
<td></td>
<td>TBD</td>
</tr>
<tr>
<td>EM Collaboration</td>
<td>1</td>
<td>1</td>
<td>0</td>
<td>TBD</td>
</tr>
<tr>
<td>I/HM Collaboration</td>
<td>6</td>
<td>1</td>
<td>0</td>
<td>300,000</td>
</tr>
<tr>
<td>Total</td>
<td>30</td>
<td>18</td>
<td>0</td>
<td>3,844,499</td>
</tr>
</tbody>
</table>
2016 Data In/Archive Keeping Statistics

- On track for >13,000 depositions in 2017
- More EM structures submitted in 2016 than NMR structures
- Increase of structures from XFEL
  - 2013: 5
  - 2014: 30
  - 2015: 45
  - 2016: 66

### Method

<table>
<thead>
<tr>
<th>Method</th>
<th>2016 Depositions</th>
<th>2015 Depositions</th>
</tr>
</thead>
<tbody>
<tr>
<td>X-ray</td>
<td>10583 (91%)</td>
<td>10167</td>
</tr>
<tr>
<td>NMR</td>
<td>473 (4%)</td>
<td>510</td>
</tr>
<tr>
<td>EM</td>
<td>531 (4.6%)</td>
<td>255</td>
</tr>
<tr>
<td>Other</td>
<td>27 (0.2%)</td>
<td>25</td>
</tr>
</tbody>
</table>

### Processing Site

- PDBj, 2240 (19%)
- RCSB PDB, 5323 (46%)
- PDB, 4051 (35%)
- PDBe, 4051 (35%)

### Depositor Location

- North America: 33%
- Europe: 34%
- South America: 19%
- Asia: 9%
- Industry: 3%
- Oceania: <1%
- Africa: 1%
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

Creation funded by NSF EAGER grant DBI-1519158
Recent *Data In/Archive Keeping* Highlights

- Group Deposition system (GroupDep) supports automated depositions of X-ray structures in bulk
  - >1300 depositions in 2017
- Implemented initial PDB archive versioning
- Prototyped a loosely-coupled *Federation of Structural Biology Data Archives* with SASBDB

April 24, 2017: Record-breaking 860 PDB Structures From New Pan-Dataset Density Analysis Method Released
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
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 1aoi)

Frequently access structure—Structure data downloaded ~257K times since 2007
## Data Downloaded Annually

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

*supported by non-PDB funding
Recognized Educational Impact

Announced March 2017

GEN Best of the Web

PDB-101 ★★★★★

Feb 15, 2017

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

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