Macromolecular structure archiving

Sameer Velankar
Team leader
Protein Data Bank in Europe
Data archiving

- CLEM
- 3DSEM
- SXT

- ET
- SAXS
- EM

- X-ray
- NMR
- EM

- Neutron
- XFEL

Organism

Infected Cell

Virus

Assembly

Complex

Chains

Molecule

Chemical Entity

PDB (1971)

EMDB (2002)

EMPIAR (2014)

SASBDB and BIOISIS archive SAXS data
wwPDB mission

• Manage and provide access to the PDB archive
  • Follows FAIR* data principle (findable, accessible, interoperable and reproducible)
  • Deposition and Biocuration
  • Establish common data standards
• Data validation
• Develop policies and procedures

*Wilkinson et al. (2016) Scientific Data doi: 10.1038/sdata.2016.18
wwPDB initiatives

- Data standards
- Method specific validation taskforces
- OneDep – common deposition, biocuration and validation system
  - Develop policies for data archiving
Data standards

- **PDBx/mmCIF Dictionary**
  - Dictionary for the PDB archive format - essential for large/complex structures
  - Extensible to new and integrative methods

- **PDBx/mmCIF Working Group**
  - Community of developers support dictionary development
  - Improving deposition
  - Deposition of un-merged intensities

- **NEF - NMR Exchange Format**
  - Data exchange format - NEF will include raw, software-specific data
Method specific validation task force

- **wwPDB X-ray VTF 2.0 Meeting - Nov 2015 at EMBL-EBI**

- **wwPDB NMR VTF-NEF Meeting – Aug 2016 at PDBj**

- **SAS task force virtual meeting – May 2017**

- **Hybrid/Integrative Methods – monthly meeting**
OneDep- Unified Deposition Portal for PDB and EMDB

- **Deposition**: deposit.wwpdb.org
- **Validation server**: validate.wwpdb.org
- **Biocuration**
- **File Preparation**
- **Public Release**
New wwPDB Policy for 3DEM Data

- Effective Sep 6, 2016, deposition to the PDB of atomic models determined by 3DEM requires prior or simultaneous deposition of the associated 3DEM electric potential maps to EMDB

- For joint PDB/EMDB depositions, the hold period is the same for both map(s) and model(s)
File Versioning: Objectives

Current Issues:

- Loss of connection between PDB ID and Publication under current wwPDB Obsolete/Supersede Policy
- Current wwPDB Policy represents a non-trivial barrier to revisions by the Depositor of Record

Objectives:

- Introduce new procedure to manage revision of atomic coordinates by the Depositor of Record
- Establish a robust extensible framework for versioning of all archival data
Looking ahead
“Resolution Revolution”

- 1.8Å structure in 2016 (PDB ID 5K12; EMD-8194)
- Increasing number of 3DEM structures at 2-4Å resolution (75 in 2015 and 80 in first 7 months of 2016)
Archiving 3DEM data

- Electric potential maps archived in EMDB
- Model Coordinates archived in PDB
- Work with EMDB and EMDataBank to determine a consensus path forward
Archiving 3DEM Entries in the PDB

- Diverse refinement strategies in use
- Masked refinement – each masked region is refined independently
- Make identification of model coordinates that have not been refined mandatory (i.e., docked as rigid bodies)?

Nguyen et al., 2016, *Nature*, 530, 298-302
Hybrid Methods Task Force summary

- Be as inclusive as possible in collecting data from many different experimental methods
- Accommodate many types of structural representations
- Create a federated system to collect/curate data
- Use a common interface to collect data
- wwPDB should play a leadership role
Describing integrative modeling


- Dictionary definitions
  - Multi-scale models
  - Multi-state models
  - Time ordered models

- Links to existing archives for experimental data
Federated archive system

- The coordinates are deposited to the PDB
- Each specialised data resource collects experimental data
- Deposition of data coordinated between PDB and data resources

PRIDE archives– CX-MS data
SAS archives

bioisis.net

sasbdb.org
SASBDB

- Maintained by EMBL-Hamburg (2014)
- Archive for biological SAS
  - SAXS, SANS, WAXS data
  - Non-atomistic models (beads)
  - Atomistic models determined solely by SAS methods

sasbdb.org
Thank you
Questions?

pdbhelp@ebi.ac.uk
proteindatabank
@PDBEurope