PDBj
Protein Data Bank Japan
http://pdbj.org/

Protein Data Bank Japan (PDBj):
Maintaining structural data archive and integration of structure data with other life sciences data resources by semantic web technologies

WORLDWIDE PDB
PROTEIN DATA BANK
http://wwpdb.org/

Haruki Nakamura, Akira R. Kinjo, Hirofumi Suzuki, Reiko Yamashita, Yasuyo Ikegawa, Takahiro Kudou, Gert-Jan Bekker, Reiko Igarashi, Yumiko Kengaku, Hasumi Cho, Junko Sato, Nahoko Haruki, Daron M. Standley, Atsushi Nakagawa

Abstract

The Protein Data Bank Japan (PDBj, http://pdbj.org/) is a member of the worldwide Protein Data Bank (wwPDB: http://wwpdb.org/) [1, 2]. PDBj accepts and processes experimental macromolecular structure data determined by X-ray crystallography, SAXS/SANS, NMR, and Electron Microscopy as a Data-in service. While maintaining the archive in collaboration with other wwPDB partners, PDBj also provides a wide range of Data-out services and tools for analyzing structures and functions of proteins, as shown in Table 1.

To facilitate the integration of structure data with other life sciences data resources, we have developed data in formats compatible with semantic web technologies: PDB/RDF (http://rdf.wwpdb.org/pdb/) and Chem_Comp/RDF (http://rdf.wwpdb.org/cc/), PDB and chemical compounds data in the Resource Description Framework (RDF) format, along with its ontology in Web Ontology Language (OWL) based on the PDB mmCIF Exchange Dictionary [1, 2]. NMR experimental data in BMRB database are also being described in RDF as BMRB/RDF, and those data are integrated through the linked-data technology.

Protein Data Bank Japan (http://pdbj.org/)

Since 2001, PDBj has been managed at Institute for Protein Research, Osaka University as a member of the wwPDB, to curate, edit and process the deposited data for an open, public, and single archive of the wwPDB. *the PDB archive contains more than 100,000 entries now.*
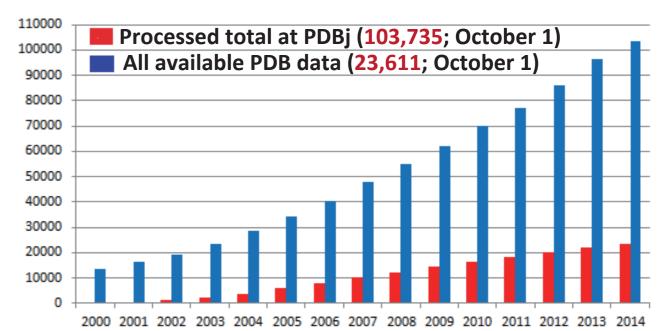
wwPDB.org

PDBj staffs (April, 2014)

wwPDB members and their heads
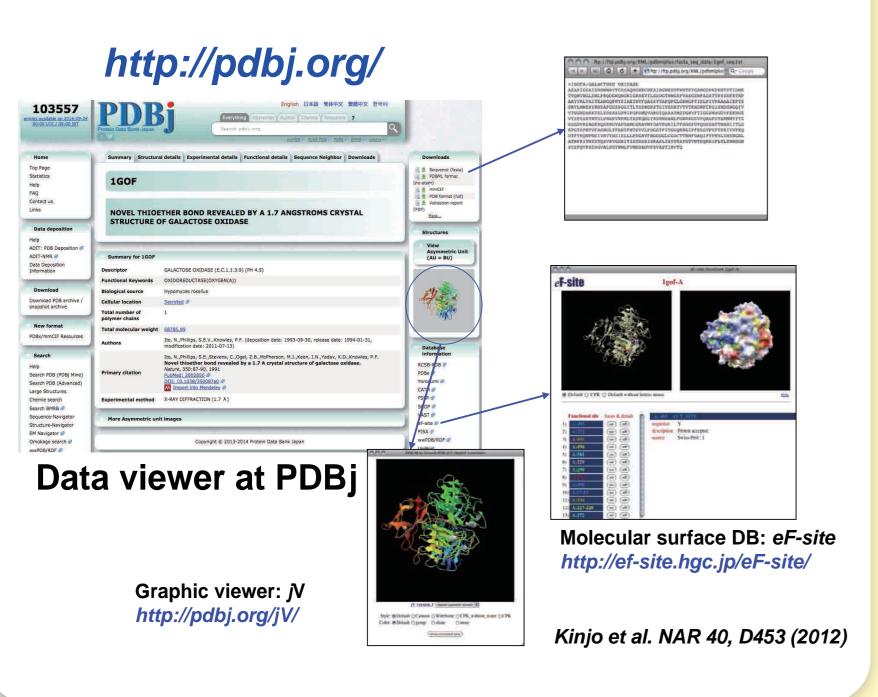
Activities/Services of each member of the wwPDB

● "Data-in" activitiy, common in all the wwPDB members with high quality control. For that purpose, new format, data deposition, and validation system are developed.

● "Data-out" services, characteristic at each wwPDB member site except common ftp-site data.

Data in at PDBj and wwPDB

PDBj curates and processes about a quarter of the deposited data, mainly from Asian and Oceania regions.

Processed total at PDBj (103,735; October 1)
All available PDB data (23,611; October 1)

Summary page

http://pdbj.org/

PDBj
103357
1GOF
NOVEL THIOETHER BOND REVEALED BY A 1.7 ANGSTROMS CRYSTAL STRUCTURE OF GALACTOSE OXIDASE

Data viewer at PDBj

Graphic viewer: jV
http://pdbj.org/jV/

Molecular surface DB: eF-site
http://ef-site.hgc.jp/eF-site/

Kinjo et al. NAR 40, D453 (2012)

PDBj services

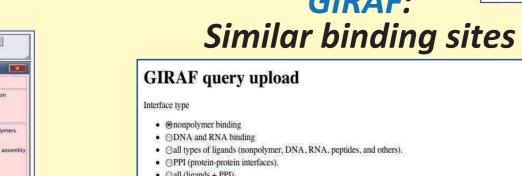### Molmil Viewer for eF-site (β-version)

HIV-1 Capsid 3J3Q, 1356 chains, >2M atoms

Molmil: *Molcular Viewer based on WebGL*
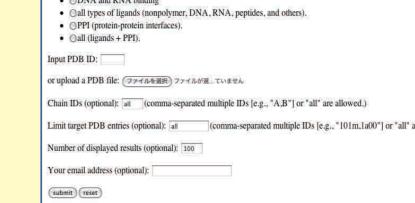- Based on JavaScript/WebGL.
- Supports PDB, mmCIF, PDBML formats.
- Links to PDB, chem_comp (Compound), and ProMode Elastic.
- Outputs screenshots.

eF-seek: Similar surfaces
eF-seek

eF-site: Mol. Surface DB

GIRAF: Similar binding sites
GIRAF query upload

ProMode: Normal Modes

EM Navi: *EMDB browser*

Yorodumi: *PDB & EMDB*

Integration of EMDB: Search for similar SHAPE

Query: human RNA polymerase II with RNA (EMDB: 2190)
Similar shapes from 195,658 images

Omokage search

PDB/RDF format for Semantic Web
Service from wwPDB by Akira R. Kinjo (PDBj) & Tom Oldfield (PDBe)

http://rdf.wwpdb.org/

Kinjo et al. (2012) Nucl. Acids Res. 40, D453-D460.

WORLDWIDE PDB
PROTEIN DATA BANK
Welcome to the Worldwide Protein Data Bank

PDB/RDF

About PDB/RDF
PDB/RDF _ chem_comp/RDF

PDB ID: (e.g. '7RSA') PDB ID
property: (e.g. 'PDBo:entity.pdbx_description')
keywords: (e.g. 'alcohol')

submit   reset

Download XSLT stylesheet for converting PDBML to RDF: PDBML2rdf.xsl.gz (gzipped 22KB)

In UniProt RDF:
```
<rdf:Description rdf:about="http://rdf.wwpdb.org/pdb/1BY4">
  <rdf:type rdf:resource="http://purl.uniprot.org/core/Structure_Resource"/>
  <database rdf:resource="http://purl.uniprot.org/database/PDB"/>
  <method rdf:resource="http://purl.uniprot.org/core/X-Ray_Crystallography"/>
  <resolution rdf:datatype="http://www.w3.org/2001/XMLSchema#float">2.10</resolution>
</rdf:Description>
```
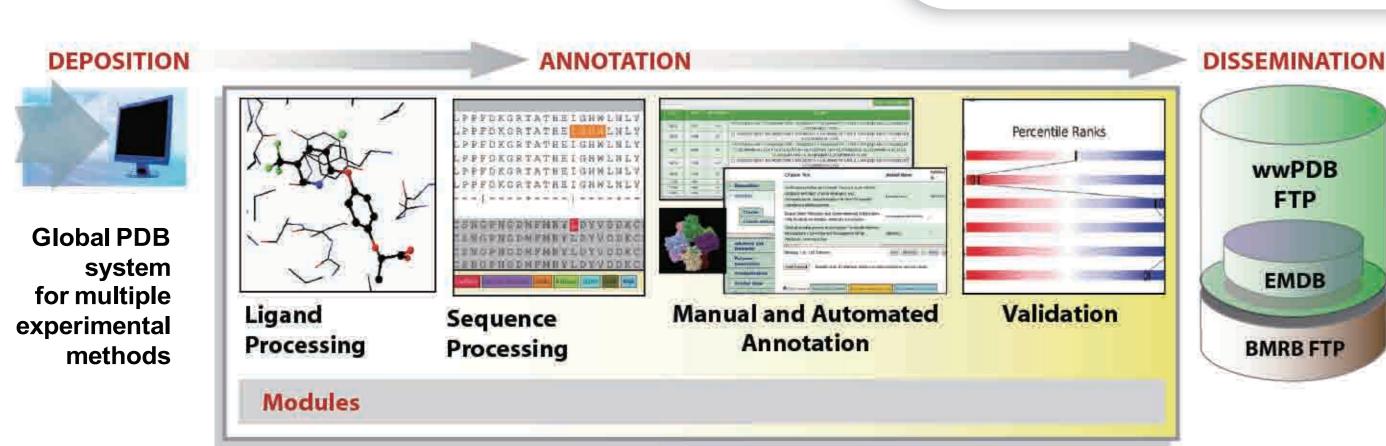
New format PDB: PDBx/mmCIF

- PDB format is almost 40 years old and does not support today's science.
- PDB Record format limitations
  - Max. 62 chains
  - Max 99.999 atoms
  - No bond orders or chirality specified for ligands
  - No support for NMR, EM hybrid methods,...
  - Meta-data specification cumbersome and inflexible

- Preserve backward compatibility where possible
- web service to create PDB format data files
- Start in early 2014

PDB format

mmCIF

New Annotation System
For data increase and high quality data management

DEPOSITION → ANNOTATION → DISSEMINATION

Global PDB system for multiple experimental methods

Ligand Processing
Sequence Processing
Manual and Automated Annotation
Validation
Percentile Ranks

Modules

wwPDB FTP
EMDB
BMRB FTP

- Enables workload balancing and increased productivity
- Better quality assurance of ligand chemistry and polymer sequences
- PDBx/mmCIF is the master file format
- Validation suites based on recommendations from experts task forces; X-ray validation pipeline is available as a stand-alone server
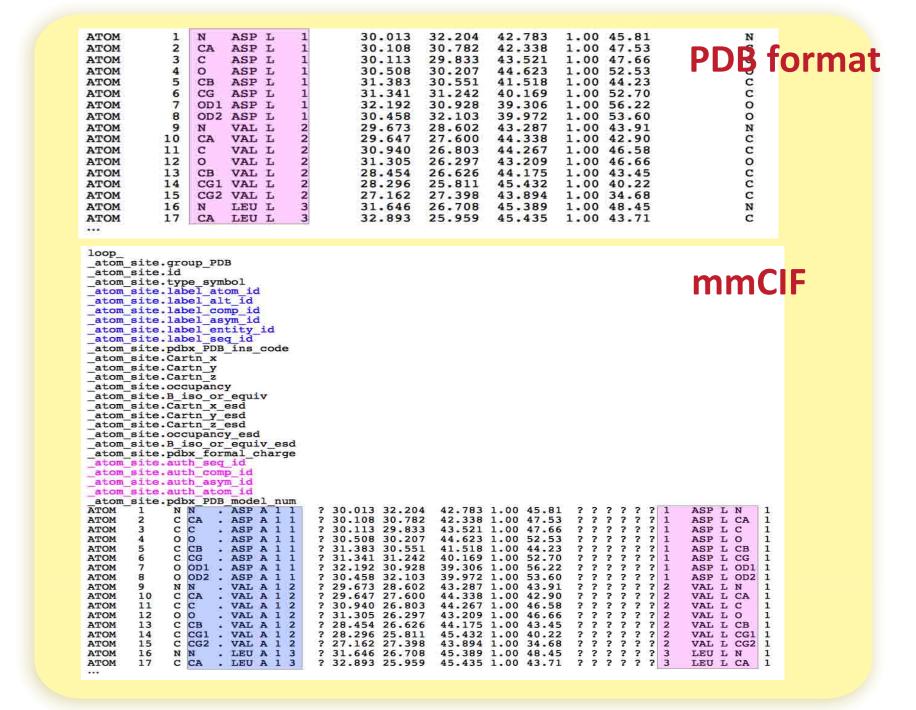- System will support all accepted experimental methods

Validation Report

wwPDB X-ray Structure Validation Report

- Version 1.0 in production use since August 2013
  - http://www.wwpdb.org/validation.html
  - Fixing occasional bugs
  - Collecting feedback to inform possible changes
  validation@mail.wwpdb.org

- January 2014 - validation data for all X-ray structures has made publicly available through the wwPDB ftp sites

PDBx/mmCIF Software Support
http://mmcif.wwpdb.org/

- **Phenix and Refmac** – produce native PDBx files for deposition
- **MMDB** - macromolecular object library in CCP4
- **iotbx.cif/ucif** - CCTBx C++/Python IO library with dictionary validation
- **CCIF** – CCP4 C++ library with FORTRAN support and dictionary validation
- **CBFLib** – ANSI-C library for CIF & imgCIF files
- **mmLIB** - Python toolkit supporting CIF & mmCIF
- **BioPython** - Python toolkit for computational biology
- **PyCifRW** - Python CIF/mmCIF parsing tools
- **BioJava** - Java mmCIF IO package
- **STAR::Parser** – Perl mmCIF parser and molecular object library
- **RCSBTools** - C++/Python parsing and dictionary validation tools plus many other supporting format conversion and data management applications
- **Visualization** - Chimera, Jmol, OpenRasMol