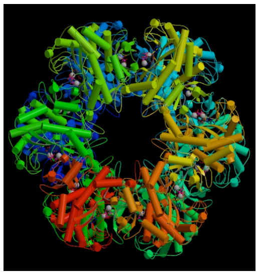
ImgCIF issues for Archivists



John Westbrook
RCSB/PDB
Rutgers University

ACA 2006

Current WORLDWIDE PDB PROTEIN DATA BANK Data Dictionaries

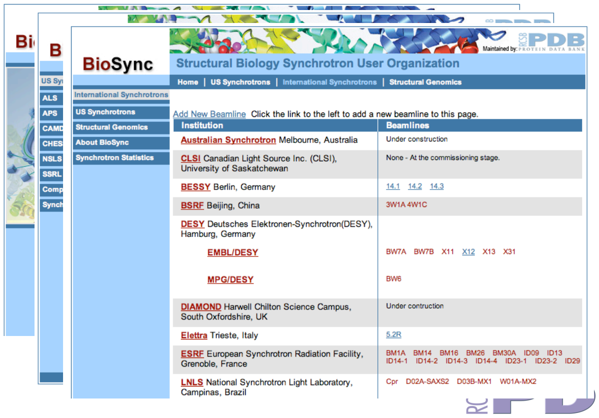
http://mmcif.pdb.org/

- **PDB data exchange (XML Schema/CIF)**
  - Including extensions for structural genomics, automated data extraction, and e-publishing

- DDL2
- mmCIF
- Image data
- Ligand data
- NMR
- Cryo-EM
- Target Registration
- Protein Production

- Modeling
- Crystallization
- Symmetry
- RNAML
- BIOSYNC

BioSync
http://biosync.pdb.org



BioSync

Structural Biology Synchrotron User Organization

Maintained by: RCSB PDB PROTEIN DATA BANK

Home | US Synchrotrons | International Synchrotrons | Structural Genomics

International Synchrotrons

US Synchrotrons
Structural Genomics
About BioSync
Synchrotron Statistics

Add New Beamline Click the link to the left to add a new beamline to this page.

Institution	Beamlines
Australian Synchrotron Melbourne, Australia	Under construction
CLSI Canadian Light Source Inc. (CLSI), University of Saskatchewan	None - At the commissioning stage.
BESSY Berlin, Germany	14.1 14.2 14.3
BSRF Beijing, China	3W1A 4W1C
DESY Deutsches Elektronen-Synchrotron(DESY), Hamburg, Germany	
EMBL/DESY	BW7A BW7B X11 X12 X13 X31
MPG/DESY	BW6
DIAMOND Harwell Chilton Science Campus, South Oxfordshire, UK	Under contruction
Elettra Trieste, Italy	5.2R
ESRF European Synchrotron Radiation Facility, Grenoble, France	BM1A BM14 BM16 BM26 BM30A ID09 ID13 ID14-1 ID14-2 ID14-3 ID14-4 ID23-1 ID23-2 ID29
LNLS National Synchrotron Light Laboratory, Campinas, Brazil	Cpr D02A-SAXS2 D03B-MX1 W01A-MX2

PDBML Schemas

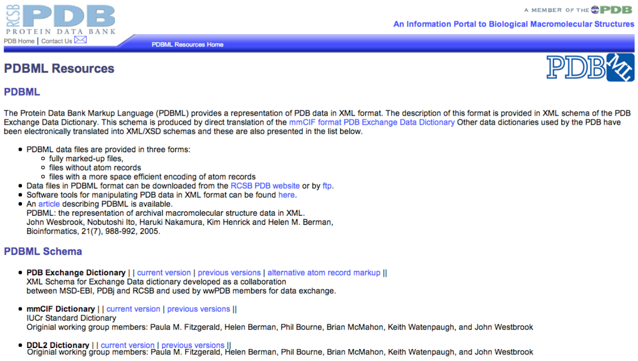
http://pdbml.pdb.org/

- New PDBML website
- Schemas provided for the PDB Exchange dictionary and component dictionaries
- Schemas and data dictionaries are updated synchronously

PDBML: the representation of archival macromolecular structure data in XML. John Wesbrook, Nobutoshi Ito, Haruki Nakamura, Kim Henrick and Helen M. Berman, Bioinformatics, 21(7), 988-992, 2005.

Dictionaries and Schemas

http://mmcif.pdb.org/ - http://pdbml.pdb.org/



A MEMBER OF THE wwPDB

An Information Portal to Biological Macromolecular Structures

PDB Home | Contact Us

PDBML Resources Home

PDBML Resources

PDBML

The Protein Data Bank Markup Language (PDBML) provides a representation of PDB data in XML format. The description of this format is provided in XML schema of the PDB Exchange Data Dictionary. This schema is produced by direct translation of the mmCIF format PDB Exchange Data Dictionary Other data dictionaries used by the PDB have been electronically translated into XML/XSD schemas and these are also presented in the list below.

- PDBML data files are provided in three forms:
  - fully marked-up files,
  - files without atom records
  - files with a more space efficient encoding of atom records
- Data files in PDBML format can be downloaded from the RCSB PDB website or by ftp.
- Software tools for manipulating PDB data in XML format can be found here.
- An article describing PDBML is available.
  PDBML: the representation of archival macromolecular structure data in XML.
  John Wesbrook, Nobutoshi Ito, Haruki Nakamura, Kim Henrick and Helen M. Berman,
  Bioinformatics, 21(7), 988-992, 2005.

PDBML Schema

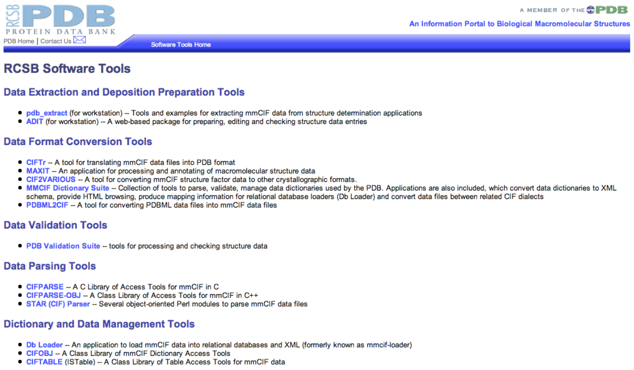
- **PDB Exchange Dictionary** | | current version | previous versions | alternative atom record markup ||
  XML Schema for Exchange Data dictionary developed as a collaboration
  between MSD-EBI, PDBj and RCSB and used by wwPDB members for data exchange.

- **mmCIF Dictionary** | | current version | previous versions ||
  IUCr Standard Dictionary
  Originial working group members: Paula M. Fitzgerald, Helen Berman, Phil Bourne, Brian McMahon, Keith Watenpaugh, and John Westbrook

- **DDL2 Dictionary** | | current version | previous versions ||
  Originial working group members: Paula M. Fitzgerald, Helen Berman, Phil Bourne, Brian McMahon, Keith Watenpaugh, and John Westbrook

Software Tools

http://sw-tools.pdb.org/



A MEMBER OF THE wwPDB

An Information Portal to Biological Macromolecular Structures

PDB Home | Contact Us

Software Tools Home

RCSB Software Tools

Data Extraction and Deposition Preparation Tools

- pdb_extract (for workstation) -- Tools and examples for extracting mmCIF data from structure determination applications
- ADIT (for workstation) -- A web-based package for preparing, editing and checking structure data entries

Data Format Conversion Tools

- CIFTr -- A tool for translating mmCIF data files into PDB format
- MAXIT -- An application for processing and annotating of macromolecular structure data
- CIF2VARIOUS -- A tool for converting mmCIF structure factor data to other crystallographic formats.
- MMCIF Dictionary Suite -- Collection of tools to parse, validate, manage data dictionaries used by the PDB. Applications are also included, which convert data dictionaries to XML schema, provide HTML browsing, produce mapping information for relational database loaders (Db Loader) and convert data files between related CIF dialects
- PDBML2CIF -- A tool for converting PDBML data files into mmCIF data files

Data Validation Tools

- PDB Validation Suite -- tools for processing and checking structure data

Data Parsing Tools

- CIFPARSE -- A C Library of Access Tools for mmCIF in C
- CIFPARSE-OBJ -- A Class Library of Access Tools for mmCIF in C++
- STAR (CIF) Parser -- Several object-oriented Perl modules to parse mmCIF data files

Dictionary and Data Management Tools

- Db Loader -- An application to load mmCIF data into relational databases and XML (formerly known as mmcif-loader)
- CIFOBJ -- A Class Library of mmCIF Dictionary Access Tools
- CIFTABLE (ISTable) -- A Class Library of Table Access Tools for mmCIF data

PDB EXTRACT

http://pdb-extract.pdb.org/
http://sw-tools.pdb.org/

PDB EXTRACT is a data extraction and validation tool which automates the assembling mmCIF data files from the outputs of X-ray and NMR structure determination applications.

Automated and accurate deposition of structures solved by X-ray diffraction to the Protein Data Bank Huanwang Yang, Vladimir Guranovic, Shuchismita Dutta, Zukang Feng, Helen M. Bermana* and John D. Westbrook Acta Cryst (2004). D60, 1833-1839

Things to consider …

- Don't let implementation issues stand in the way of developing data dictionaries (logical vs. physical)…
- It is useful to take advantage of dictionary validation as early as possible…
- Don't underestimate legacy data issues…
- Most data collection details are currently at best 'user verifiable' and details fall through the cracks..
  - Please agree on what additional details need to be archived and make this electronic.
- BioSync/PDB can collect warehouse standard features and possibly *test data sets*
- We will also be happy to adapt any of our parsing, loading, transforming and extraction tools as needed…

PDB EXTRACT

http://pdb-extract.pdb.org/
http://sw-tools.pdb.org/

- Data collection and reduction
  - HKL, SCALEPACK, d*TREK,SAINT, SCALA
- Molecular replacement
  - CNS,CNX,Amore,Morep,EPMR
- Heavy atom phasing
  - CNS,CNX,SOLVE,MLPHARE,SHARP/autoSHARP,SHELXD,SHELX, PHASES,SnB,BnP,Phaser
- Density modification
  - CNS,CNX,DM,Solomon,RESOLVE,SHELXE
- Structure refinement
  - CNS,CNX,REFMAC,RESTRAIN,SHELXL,TNT,WARP,PHENIX

RCSB PDB PROTEIN DATA BANK

http://www.pdb.org/

Operated by the RCSB members:
Rutgers, The State University of New Jersey and the
San Diego Supercomputer Center at the University of California, San Diego

The RCSB PDB is supported by funds from the National Science Foundation (NSF), the National Institute of General Medical Sciences (NIGMS), the Office of Science, Department of Energy (DOE), the National Library of Medicine (NLM), the National Cancer Institute (NCI), the National Center for Research Resources (NCRR), the National Institute of Biomedical Imaging and Bioengineering (NIBIB), and the National Institute of Neurological Disorders and Stroke (NINDS).

Access

- RCSB Protein Data Bank Site
  - http://www.pdb.org/
- RCSB/PDB Dictionary Resource Site
  - http://mmcif.pdb.org /
- PDBML site
  - http://pdbml.pdb.org/
- RCSB/PDB Software Download Site
  - http://sw-tools.pdb.org /
- PDB Extract site
  - http://pdb-extract.pdb.org/
- BioSync
  - http://biosync.pdb.org/