Worldwide Protein Data Bank Advisory Committee Meeting

September 30, 2011
Agenda

Welcome
Janet Thornton

Introductions and Overview
Gerard Kleywegt

Common Deposition and Annotation Tool
Martha Quesada

Method- and Molecule-specific Activities
John Markley
Helen Berman

PDBj Update
Haruki Nakamura
Welcome

Janet Thornton
Overview

Gerard Kleywegt
Introductions…

wwPDB – a global organisation
wwPDB
October 2010 - September 2011

- Continued growth of archive
- Increased use of data
- Funding updates
- Release of PDB archive version 4.0
- Substantial progress in Common Tool project
- Format discussions with software developers
- Task Force activities
- wwPDB Foundation
- PDB40
- Continued intensive staff interactions
- wwPDB activities at IUCr
- UAB update
- Planning of next archive remediation
Breaking News!

- wwPDB/CCDC Memorandum of Understanding signed 29 Sept 2011
  - wwPDB gets to use Mogul for ligand validation and to generate refinement dictionaries for compounds in the PDB
  - wwPDB gets to incorporate CSD coordinates for compounds in the PDB
2010 wwPDBAC Recommendations

Common D&A Tool
- Make time estimates of speed & throughput once software in place for contingency planning
  
  Processing time for ligands cut by up to 70% with new interface; benchmarking to continue

Remediation
- Endorsed plan for B-factors
  
  Released July 2011

Task Forces
- Publish white papers
  
  X-ray paper in press
  NMR, EM in preparation
  SAS, to meet in 2012
  Hybrid methods, 2013

Format issues
- Input from stakeholders in 2011, full implementation in 2012
  
  Meeting held Sept 26-27, 2011
PDB Depositions

By deposition and processing site

By experimental method

(Updated 14 Sep 2011; * projection for 2011)
PDB FTP Downloads

* Version 3.0/3.1 files released
* Version 3.15 files released
* Version 4.0 files released
2010 FTP Traffic

- RCSB PDB: 159 million entry downloads
- PDBe: 34 million entry downloads
- PDBj: 16 million entry downloads
Funding

- RCSB PDB competitive renewal funded by NSF
  - January 2009 - December 2013
- PDBe competitive grant from Wellcome Trust
  - January 2010 - December 2014
  - Stable core of ~15 EMBL posts by 2013 (up from 6 in 2008)
- PDBj competitive renewal funded by JST (Japan Science & Technology Agency)
  - April 2011 - March 2014
- BMRB competitive renewal funded from the National Library of Medicine
  - September 2009 – August 2014 (parent grant)
  - September 2009 – August 2011 (admin supplement – US recovery act funding)
  - September 2009 – August 2011 (competitive renewal – US recovery act funding)
  - NLM will no longer fund BMRB after 2014
Remediation

- Focus
  - Antibiotics and peptide inhibitors
  - Representation of biological assemblies
  - Residual B-factors
  - Entries in a non-standard crystal frame

- Released July 13, 2011
Common Tool for Deposition and Annotation

- Sequence-annotation module v1.0 completed with enhanced user interface capability
- Ligand-annotation module v1.0 including new features for small polymer molecules completed
- Workflow engine and management system running with annotation modules
- Validation module on track
- Deposition system in active development
- Cross-site data-sharing architecture in place
Method-specific (Validation) Task Forces have been convened to collect recommendations and develop consensus on method-specific issues, including validation checks that should be performed and identification of validation software applications

**X-ray Validation**
- 2008 Workshop on Next Generation Validation Tools for the wwPDB
- White paper in press in *Structure*
- Chair: Randy J. Read (University of Cambridge)

**3DEM Validation**
- Meeting September 2010
- Chairs: Richard Henderson (Maps, MRC-LMB), Andrej Sali (Models, UCSF)
- White paper in progress

**NMR Validation**
- Meetings held September 2009, January 2011
- Chairs: Gaetano Montelione (Rutgers), Michael Nilges (Institut Pasteur)
- Report in progress

**Small-Angle Scattering**
- Members: Jill Trewhella (University of Sydney), Dmitri Svergun (EMBL Hamburg), Andrej Sali (UCSF), Mamoru Sato (Yokohama City University), John Tainer (Scripps)
Established to support specific wwPDB activities
- Advisory committee meetings
- Outreach and education activities, including seminars and workshops

501(c)3 organization
- American, tax-exempt association dedicated to scientific, literary, charitable, and educational purposes

Fundraising on-going
Confirmed Speakers

- Cheryl Arrowsmith, University of Toronto, Canada
- David Baker, University of Washington
- Ad Bax, NIH/DHHS/NIDDK/LCP
- Axel Brunger, Stanford University/HHMI
- Stephen K. Burley, Eli Lilly & Co.
- Wah Chiu, Baylor College of Medicine
- Johann Deisenhofer, UT Southwestern Medical Center
- Angela Gronenborn, University of Pittsburgh
- Richard Henderson, MRC Lab. of Molecular Biology
- Wayne Hendrickson, Columbia University
- Mei Hong, Iowa State University
- So Iwata, Imperial College London
- Louise Johnson, University of Oxford
- Brian Matthews, University of Oregon
- Jane Richardson, Duke University Medical Center
- Michael Rossmann, Purdue University
- Andrej Sali, University of California, San Francisco
- David Searls, Independent Consultant
- Susan Taylor, University of California, San Diego
- Janet Thornton, EMBL, Hinxton,
- Soichi Wakatsuki, IMMS-KEK
- Kurt Wüthrich, The Scripps Research Institute, ETH Zürich

meetings.cshl.edu/meetings/pdb40.shtml
wwPDB Interactions

- wwPDB leadership
  - Monthly wwPDB Foundation phone meetings
  - Additional Skype and phone meetings
  - Yearly visits
- Common Tool for Deposition & Annotation Project
  - Weekly VTC meetings
  - Quarterly in-person meetings
  - Daily phone, email and Skype meetings
- Regular annotator exchange visits
- NMR
  - Weekly phone/VTC meetings
- EMDB
  - Biweekly phone/VTC meetings
IUCr Participation
August 22-30, 2011 in Madrid, Spain

- Joint wwPDB exhibition booth
- Q&A forum with the wwPDB PIs
- Talks
  - Gerard Kleywegt, Validation and Errors in Protein Structures
  - Swanand Gore and Marina Zhuravleva, Validation of small molecule and macro-molecular X-ray structures
  - John Westbrook, The wwPDB Working Format
- Posters
  - Martha Quesada, wwPDB Common Tool for Deposition and Annotation
  - Akira Kinjo, Protein Data Bank on the semantic web
Format Discussions

- New format needed to address limitations in molecular size and complexity and extensibility of existing PDB format
- Format proposal circulated to key developers for review by February 2011
- Revisions and simplifications based on preliminary review
- Format workshop with selected developers held September 26-27, 2011
  - Surprise outcome…
New Activities for the Coming Year

- Planning of next archive remediation
  - Issues to be addressed include:
    - Carbohydrates
    - Post-translational modifications
    - Non-standard linkages
    - Apply symmetry if this yields a more sensible biological assembly
    - Fix partial B-values (TLS issue)
    - Non-standard coordinate frames
  - Analysis ➔ Recommendations ➔ Review ➔ Decisions ➔ Remediation

- Improve wwPDB “corporate image”
  - Confusion about PDB – wwPDB – partners
  - Explore domain name change to pdb.org
PDB.org

- Goal - improve visibility of the wwPDB
- Possible option - change wwPDB URL to **pdb.org**
  - Home page for PDB and wwPDB, to launch deposition sessions, go to one of the member sites, or access wwPDB materials
  - Site-agnostic page for each PDB entry (DOI) with basic information and links to the entry pages at the partner sites, *e.g.* [pdb.org/entry/1xyz](http://pdb.org/entry/1xyz)
- Investigating technical solutions for various issues
  - **pdb.org** is owned by the RCSB PDB
  - Can’t break any existing URLs (bookmarks, URLs used by third-party software, existing Google links)
  - Can’t break programmatic access to current services
Common Deposition & Annotation (D&A) Tool

Martha Quesada
Evolving User Needs

- Larger and more complex biological molecules
- New methods
- Expanded annotation
- Improved quality
  - New validation pipelines
- Higher throughput
  - Automation and validation of routine submissions
Common D&A Project Team March 2011

Experience, expertise and diverse skills representing the broad interests of wwPDB
The Vision

Common Deposition Interface
- Accession ID
- Validation Report
- Other

Data Harvesting Tools

Author info, Citations
- Restraints
- Chemical shifts
- NMR exp details
- X-ray exp details
- X-ray SF
- EM exp details
- EM maps

Integrating Data Capture

Integrated Data Capture

Author info, Citations
- Restraints
- Chemical shifts
- NMR exp details
- X-ray exp details
- X-ray SF
- EM exp details
- EM maps
What’s In It For…

- **Depositors**
  - Interactive and informative deposition interface
  - Enhanced processing functionality to support new methods
  - Value-added validation feedback and annotation during deposition
  - Faster processing

- **Annotators**
  - Improve efficiency, freeing time for more advanced annotation

- **Data users**
  - Higher quality archive
wwPDB Common Deposition and Annotation Pipeline

Deposition Pipeline
- Deposition Interface
- Data upload, harvesting, 1° test and verification
- Client side Editor
- Validation
- Standalone and integrated
- Submission
- Progress Tracking/Status

Communication System

Workflow-Automation System

Annotation Pipeline
- Sequence Processing
- Ligand Processing (ID, Edit, Build)
- Validation
- Calculated annotations (PISA, SITE & LINK records, cross references, metal coordinates)
- Corrections
- Release Processing
- Progress Tracking/Status

Calculated annotations (PISA, SITE & LINK records, cross references, metal coordinates)
Development Status as of August 2011

- Review of old system
- Requirements definition
- Core infrastructure
- Annotation system
- Deposition system
- Communication system
- Internal testing
- Release

Yellow denotes degree of completion
Deposition Interface Design and Community Input

- wwPDB partner groups: initial requirements and design
- Introduction to community at ACA 2010
- Deposition user interface: initial feedback at IUCr 2011
  - Iterative evolution
- Interface review by targeted external user representatives (September, October 2011)
  - Iterative evolution
- Broader review by community experts (December 2011)
  - Iterative evolution
- Community beta testing to begin Q3 2012
Interface Features for Depositors

- Automated batch data uploads
- Flexible manual data entry
- Restart deposition and re-upload data without loss of general information
- Build new submissions on previous depositions
- Easily view percentage complete
- Visually review data
- Structure validation reports
Interface Look and Feel: Deposition

05Aug2011 Depositor: How can I provide additional sequence information?

05Aug2011 PDB Staff response: Add sequence information to the sequence details box.
Workload Balance

- Depositions will be distributed taking into account:
  - Expertise: relevant expertise in the experimental methodology
  - Grant-agency guidelines
  - Time zone: facilitate “help” and communication
  - Load balance: even distribution with respect to each site’s local capacity (e.g., taking into account local holidays)

- Single, wwPDB-branded, point of contact for all new depositions (e.g., wwpdb.org/deposit)
Annotation: Modules in Hand

- Sequence module - unit tested
- Ligand module in production benchmark testing
  - Ligand processing
  - “Chopper”
  - Editor
- Workflow Manager operational with ligand and sequence modules
  - Load testing (tested via shared server: RCSB PDB)
  - Batch workflow processing working at both sites
  - Functionality (tested via shared server: RCSB PDB, PDBe)
Graphical Interface for Sequence Processing

Sequence Alignment Tool
Identifier: rcsb067042  Instance: RCSB

<table>
<thead>
<tr>
<th>POSITION</th>
<th>AUTH PDB:A</th>
<th>ALIGNED SEQUENCE</th>
<th>RESIDUE</th>
<th>ANNOTATION DETAILS</th>
</tr>
</thead>
<tbody>
<tr>
<td>15</td>
<td>ALA</td>
<td>UNP:Q52L64</td>
<td>VAL</td>
<td>engineered mutation</td>
</tr>
<tr>
<td>31</td>
<td>ASN</td>
<td>UNP:Q52L64</td>
<td>TYR</td>
<td>engineered mutation</td>
</tr>
<tr>
<td>33</td>
<td>ARG</td>
<td>UNP:Q52L64</td>
<td>TYR</td>
<td>Shift/Click to edit</td>
</tr>
<tr>
<td>34</td>
<td>THR</td>
<td>UNP:Q52L64</td>
<td>ASN</td>
<td>Shift/Click to edit</td>
</tr>
</tbody>
</table>

The table shows the alignment of sequences with positions, aligned sequences, residues, and annotation details. The graphical interface includes a 3D viewer with a color legend for conflicts, deleted residues, undo, replace, insert, DNA, and RNA.

The sequences are shown with graphical representations of amino acids, highlighting specific residues and positions.
Graphical Interface for Ligand Processing
Peptide Ligand Chopper

- Annotator-directed bond breaks
- Add leaving groups (\(-\text{OH}, -\text{H}, -\text{Cl}, \text{etc.}\))
- Residues atom numbers according to CCD
- Residue atom numbering mapped to full ligand

- In production—significantly improving efficiency
- Integrated views—enabled by new reusable visualization and editing capabilities

CHOP

PRO PHE GLU 6CW LEU ASP TRP GLU PHE DPR
EM Integration

- Functional requirements 90% completed
  - Dictionary for incorporation into D&A
  - Interface requirements underway
- Large data file requirements to be supported in V1.0 of the deposition module
- Additional visualization, data harvesting to be supported in V1.X
- Validation requirements from EM VTF to be supported in VN.0
NMR Integration

- Dictionary data items supporting NMR have been defined
- Data requirements defined for chemical shifts
- Integration of software for PDB atom nomenclature correspondence to NMR experimental data
- Implement Common D&A and ADIT-NMR data exchange
Timeline

- Common Tool released for public use late 2012
- Full integration testing of the pipeline modules to begin in Q2 2012
- All modules completed and integrated into the pipelines by end of Q1 2012
- Deposition Interface – External user testing to begin mid-November 2011
Method- and Molecule-specific Activities
NMR

John Markley
Support for Chemical Shifts

- Deposition mandatory as of December 6, 2010
- Pre-validation of correspondence between nomenclature in coordinates and chemical shifts
- Training of annotators at PDBj-BMRB
- Coordination of annotator work flows
  - RCSB PDB ⇔ BMRB ⇔ PDBj/PDBj-BMRB
  - PDBe has separate system
  - Common Tool will unify these workflows
- Chemical shift files archived at PDB and BMRB are consistent with final annotated coordinates
2011 Depositions

- 516 new BMRB depositions
  - >60% associated with coordinates
- 372 new combined PDB and BMRB entries
  - 40 new coordinate sets associated with earlier BMRB depositions
- 318 new depositions through BMRB
- 14 new depositions through PDBe
Restraint Processing

- Restraints go to PDB and are sent to BMRB for processing
- Software used in restraint processing was developed in collaboration with Wim Vranken, Jurgen Doreleijers, Geerten Vuister & Gert Vriend
BMRB and the Common D&A Tool Development

- Staff members engaged in planning and initial software development
- BMRB implemented NetApp hardware ($42,000) and software for use in exchanging ADIT-NMR deposition data with RCSB PDB
- Working to set up snap mirroring of all data exchanged with RCSB PDB, PDBj, and PDBj
New Features Used by Depositors and Annotators

- Support for SAXS data and restraints for NMR structures
- Chemical shift validation reports generated by software (LACS, AVS, Sparta & PANAV) are sent to depositors
- Visualization of restraints and restraint violations (from CING software) as aids to depositors and annotators
NMR Validation Task Force

Gaetano Montelione
wwPDB NMR Structure Validation Task Force

Gaetano Montelione
Michael Nilges

Ad Bax
Peter Guentert
Torsten Herrmann
John Markley
Jane Richardson
Charles Schwieters
Wim Vranken
Geerten Vuister
David Wishart

Helen Berman
Andy Byrd
Aleksandras Gutmanas
Yuanpeng Janet Huang
Gerard Kleywegt
Naohiro Kobayashi
Cathy Lawson
Haruki Nakamura
Roberto Tejero
Eldon L. Ulrich
John Westbrook
Scope of Work for Jan 15, 2011 Workshop

It was decided to focus the Jan 15 meeting on Validation of Ordered Regions of Protein NMR Structures.

Discussion of validation of other biomolecular structures, including nucleic acids and disordered regions of biomolecules, will be deferred until we have consensus on validating the well-ordered regions of protein NMR structures.
The NMR-VTF recommends development of NMR structure validation tools by the PDB in three phases.

Phase 1.  Tasks which could be implemented by PDB in 2011 using largely existing software

Phase 2.  Tasks for which software / methods are available, but which need more assessment before defining standard validation conventions for PDB

Phase 3.  Tasks requiring further research over the coming years
Phase 1: To be implemented in Ver 1 of PDB NMR Validation Software

Validation Reports to be generated for all NMR structures submitted to the PDB.

1. Chemical Shift Validation
All NMR structures submitted to PDB must include chemical shift data
Methods already in place in BioMagResDB; AVS, LACS, Sparta

2. Defined vs Undefined regions / atoms
It is necessary to consider this issue in validation

Define standard for “well defined” vs “not-well-defined” regions.
   residue ranges
   well-defined vs not-well-defined atoms

PDB should support user-defined residue ranges and/or atoms. PDB coordinate file should have this user-defined information.

Should be implemented as a standard convention
   dihedral circular variance
   distance variance matrix – preferred as convention

Can use this convention (well-defined vs not-well-defined) to define standard convention for rmsd computation
3. Knowledge-Based Protein Structure Validation

Adopt the Crystallography VTF recommendations for Knowledge-based Validation: Geometry, Packing, Underpacking

Exclude “not well defined” regions from validation

Report Z scores relative to set of high resolution crystal structures

- same set used for X-ray statistics
- set of reference NMR structures
- vs “information content” - Phase 3
- membrane proteins
- intrinsically disordered proteins - Phase 2

4. Validation of Structures Against Constraints

Cross-check Constraint Analysis results provided by CING, PDBStat and other methods

Finalize methods for interpreting constraints involving prochiral sites

Define standard Constraint Violation Report
X-ray VTF, Remediation and Format Issues

Helen M. Berman
X-ray VTF recommendations

- Integrated battery of quality checks
  - Pool together validation code from community software into a single pipeline
- Percentile scores
  - Put validation scores in context, helping non-expert users in judging quality
- Access to validation analyses
  - Cater to all classes of users: depositors, reviewers, expert and non-expert end users, programmers

From A new generation of crystallographic validation tools for the Protein Data Bank
Structure, in press
July 2011 Remediation
PDB File Format Version 3.3
PDB Exchange Dictionary (PDBx) Version 4.0

- Entries containing residual B-factors labeled (7.3K entries)
- Antibiotics and peptide inhibitors standardized (1K entries)
- Entries in the nonstandard crystal frame labeled (148 entries)
- Biological assemblies corrected (5.8K entries)
- Added support for polymers containing non-standard polymer linkages (58 entries)
- Added support for hybrid X-ray/neutron diffraction experiments (54 entries)
- Added new revision logging to PDBx/PDBML entries (all)
July 2011 Remediation
Revision log

- New External Reference File containing detailed revision information for all remediated entries will be provided at wwpdb.org
- Content changes (remediated and ongoing) tracked in PDBx and PDBML data files, including
  - Revision date
  - Version number (e.g., 4.0001)
  - Nature of revision (e.g., atom nomenclature, sequence database correspondence, citation, …)
  - Revision details
- Revision tracking in PDB format files to continue using REVDATs
Peptide Reference Dictionary

- Provides uniform representation of small peptide inhibitors and antibiotics
- Developing infrastructure for integration with current data deposition systems
  - Preliminary standalone searching available for data processing
  - Building processes for updating and maintaining PRD
Why Create a New PDB Format?

- Problem: PDB format is almost 40 years old and does not support today’s science
  - Let alone tomorrow’s science…

- Some of the limitations
  - Max 62 chains
    - and that’s stretching it
  - Max 99,999 atoms
    - 5 ribosomes in ASU=10 PDB entries!
  - Very short chain, residue and atom names
    - 1, 3, 4 characters, respectively
  - No bond orders or chirality specified for ligands
  - No support for NMR, EM, hybrid methods, …
  - Meta-data specification cumbersome and inflexible
Why Create a New PDB Format?

- wwPDB archival/exchange format is PDBx (mmCIF)
  - No uptake in community despite libraries
  - Good for machines, not so good for humans

- Pragmatic solution needed
  - Specify new working format for data exchange between software used in labs
    - Molecular replacement, refinement, model-building, graphics, validation, deposition, ...
  - Also requires specification of a new “human-readable report” format for meta-data
Proposed new PDB Format

New Format in the Lab → Structure Determination Pipeline → wwpdb Processing and Annotation

wwpdb Deposition

New Format in wwpdb ftp Archive

Round Trip
Format meeting 26/27 Sept 2011
Meeting outcome

- And the “New PDB Format” is …

- PDBx
Plan

- PDBx as working and deposition format
  - Commitments from CCP4, Phenix and Global Phasing (i.e., ~85% of all PDB depositions)
  - Agreement on managing development between these software providers and wwPDB
    - Projected completion – January 2013
- New, simplified, future-proof PDB format
  - Think “wide-PDB”
  - Will be developed by wwPDB with input from stakeholder communities
3D Electron Microscopy
EMDataBank

- Collaborative project between RCSB PDB, PDBe, and Baylor-NCMI funded by NIH, BBSRC, and EMBL
- Unified tool for collecting model coordinates and map files in a *one-stop shop*
- Merge deposition and annotation with PDB as part of Common D&A Tool
- Joint map + coordinate deposition service
- News, software list, information about dictionaries, conventions, FAQ, community links
- Search by ID, author, sample type, keyword, deposition date
- Map and map+model 3D java viewers
Recent Progress

- Map archive now >1000 entries
- Metadata remediated to improve uniformity
- Improved web-based map+model viewer
- Requirements developed for EM in Common Tool

OpenAstexViewer adapted for EM maps and models

September 21, 2011:
1140 map entries, 407 PDB coordinate entries
Planned Archive Integration

- Current PDB archive: 130 GB
- Current EMDB archive: 50 GB
- Maps indexed by EMDB code and PDB code
- Completion in 2012


Archaeal group II chaperonin
Map: EMD-5247
Model: 3izk
EM Modelling Challenge and Workshop

- Challenge held Jul-Dec 2010
- 6 target structures
- 136 models were submitted by 10 different research groups
- Initial review of results at January 2011 workshop
- Results will be published in a special issue of *Biopolymers*
- Challenge to be repeated biannually
EM Validation Task Force

- EM VTF: Advise on approaches to validate EM maps and models
- Initial meeting was Sept 28-29, 2010
- White paper describing initial recommendations is being prepared for publication
- July 2011 renewal application to NIH: work with the EM community to carry out VTF recommendations

Co-chairs:
Richard Henderson
Andrej Sali
SAXS/SANS
wwPDB Proposed Requirements for a SAXS/SANS PDB Entry

- Model is derived and fully defined by the experimental data
- Model is a folded chain of residues with directionality
- COMPND, SOURCE, SEQRES and external sequence reference (DBREF) are included
- \( x,y,z \) coordinates per atom. \( \text{Ca} \) or \( \text{P} \) model allowed
- Has acceptable geometry (bond lengths, bond angles, torsion angles, non-bonded contacts, etc.)
- Experimental and refinement details recorded in appropriate REMARK records
- Parameters directly derived from the scattering profile should be supplied and appropriately recorded (radius of gyration, \( D_{\text{max}} \) in distance distribution function, mass, etc.)
- Reduced 1D experimental profile
- Family of models should be superimposed
SAXS/SANS Task Force

- Meeting to be held Q2 2012
- Members
  - Jill Trewhella (University of Sydney)
  - Dmitri Svergun (EMBL Hamburg)
  - Andrej Sali (UCSF)
  - Mamoru Sato (Yokohama City University)
  - John Tainer (Scripps)
- Questions to the Task Force
  - Should the PDB accept (some types of) models based on SAS studies (an alternative could be, for instance, to capture such models in a separate database)?
  - If so, which types of models should be included (and which should not)?
  - What are the minimum requirements for these models?
  - What are the requirements regarding the supporting experimental data that need to be deposited?
  - What validation procedures should be applied in the deposition and annotation process (pertaining to the quality of the model, the quality of the data, and the quality of the fit of the model to the data)
PDBj Update

Haruki Nakamura
National Bioscience Database Center in Japan and PDBj Funding

- **August 2000** Proposal from CSTP (Council for Science and Technology Policy) was issued in Japanese Government to promote Genome Informatics Research Area

- **April 2001** BIRD (Institute for Bioinformatics Research and Development) was founded in JST (Japan Science and Technology Agency). Since then, PDBj has been supported by BIRD.

- **April 2005 - March 2008** Investigation for *Integration of Life Science Databases* as a Project promoted by Cabinet Office, Japanese Government

- **September 2006 - March 2011** Integrated Database Project by MEXT (Ministry of Education, Culture, Sports, Science and Technology)

- **April 2006 - March 2011** Database Center for Life Science (DBCLS) at Research Organization of Information and Systems (ROIS)

- **April 2011** New National Bioscience Database Center in Japan was founded

- **April 2011** PDBj renewal grant was approved for three years
Roadmap for Foundation of National Bioscience Database Center (NBDC)

Preparation
FY2008

1st Stage
FY2011

Task Force in CO
CSTP

Integrated Database Project
BIRD-JST

National Bioscience Database Center
(as an organization of JST governed by MEXT)

PDBj
FY2010: 119MJY

Other Life Science DBs from METI, MHLW, and MAFF

2nd Stage
FY2014

National Bioscience Database Center
(as an organization of Cabinet Office)

PDBj
FY2011: 70 (+ 27 from Osaka U.) MJPY
Organization of National Bioscience Database Center (NBDC)

CSTP in Cabinet Office

Life Science Project Team in CSTP
(Soichi Wakatsuki)

Headquarter of NBDC in CSTP
(Haruki Nakamura)

Adv. Committee

Director of National Bioscience Database Center

JST

National Bioscience Database Center

DBs
PDBj is a member of NBDC, Japan

http://biosciencedbc.jp/

About the Director-General (Michio Oishi, Ph.D.)

Education:
1958 B.S. Dept. of Biology, Faculty of Sciences, The University of Tokyo
1963 Ph.D. Dept. of Biochemistry and Biophysics, Faculty of Sciences, The University of Tokyo
Organization of the New Project of PDBj with NBDC

PDBj: Database for X-ray and EM structure and experimental data

Haruki Nakamura (Professor)
Atsushi Nakagawa (Professor)
Rei Kinjyo (Associate Prof.)
Daron Standley (Associate Prof. at iFREc)
5 annotators, 2 programmers, 2 research scientists

PDBj-BMRB: Database for NMR structure and experimental data

Toshimichi Fujiwara (Professor)
Chojiro Kojima (Associate Prof.)
1 annotator, 1 programmer
1 research scientist

Project director
Haruki Nakamura

3-year project
April 2011-March 2014
Launching New PDBj Activities

- PDB and BMRB Data-in, collaborating with other wwPDB members
- PDB on the Semantic Web for Database Integration in NBDC
- Development of a Validation Tool for PDB Data Description
- Development of a new pipeline service from Sequence to Biological Function Through Structure
- Development of new service tools for BMRB Data-in and NMR Data Analysis
PDB/RDF is a collection of PDB data in the Resource Description Framework (RDF) format. The RDF format is the standard format for the Semantic Web. An ontology defined in the Web Ontology Language (OWL) is also provided for the PDB/RDF, which is a straightforward translation of the PDB mmCIF Exchange Dictionary.
PDB/RDF example

By accessing http://pdbj.org/rdf/1GOF, a list of category holders for the PDB entry 1GOF can be retrieved in the RDF/XML format.

Then, a list of category elements can be retrieved (again in the RDF/XML format).

Finally, for a particular category element, the list of properties of that element is retrieved.
Example of an RDF graph

The network of RDF resources for the PDB entry 1GOF.

A subgraph of the left network augmented with literal objects.
Validation of PDBx and PDBML description

Validation Tool for Master Format File

Upload Master Format file (PDBx/mmCIF)

Run Validation

reset

Validation Results

Validated file: 1hpi.xml

<table>
<thead>
<tr>
<th>No.</th>
<th>Error code</th>
<th>XPath</th>
<th>Message</th>
<th>line</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>Error 1826</td>
<td>/datablock/atom_siteCategory/atom_site/pdbx_formal_charge</td>
<td>'1.5' is not a valid value of the local union type.</td>
<td>1721</td>
</tr>
<tr>
<td>2</td>
<td>Error 1826</td>
<td>/datablock/atom_siteCategory/atom_site/pdbx_formal_charge</td>
<td>'1.07' is not a valid value of the local union type.</td>
<td>1754</td>
</tr>
<tr>
<td>3</td>
<td>Error 1826</td>
<td>/datablock/atom_siteCategory/atom_site/pdbx_formal_charge</td>
<td>'0.0' is not a valid value of the local union type.</td>
<td>1787</td>
</tr>
<tr>
<td>4</td>
<td>Error 1824</td>
<td>/datablock/atom_siteCategory/atom_site/label_req_id</td>
<td>'is not a valid value of the atomic type 'xinteger'.</td>
<td>18113</td>
</tr>
<tr>
<td>5</td>
<td>Error 1824</td>
<td>/datablock/atom_siteCategory/atom_site/label_req_id</td>
<td>'is not a valid value of the atomic type 'xinteger'.</td>
<td>18146</td>
</tr>
<tr>
<td>6</td>
<td>Error 1824</td>
<td>/datablock/atom_siteCategory/atom_site/label_req_id</td>
<td>'is not a valid value of the atomic type 'xinteger'.</td>
<td>18179</td>
</tr>
<tr>
<td>7</td>
<td>Error 1824</td>
<td>/datablock/atom_siteCategory/atom_site/label_req_id</td>
<td>'is not a valid value of the atomic type 'xinteger'.</td>
<td>18212</td>
</tr>
<tr>
<td>8</td>
<td>Error 1824</td>
<td>/datablock/atom_siteCategory/atom_site/label_req_id</td>
<td>'is not a valid value of the atomic type 'xinteger'.</td>
<td>18245</td>
</tr>
<tr>
<td>9</td>
<td>Error 1824</td>
<td>/datablock/atom_siteCategory/atom_site/label_req_id</td>
<td>'is not a valid value of the atomic type 'xinteger'.</td>
<td>18278</td>
</tr>
<tr>
<td>10</td>
<td>Error 1840</td>
<td>/datablock/atom_siteCategory/atom_site/group_PDB</td>
<td>[facet 'enumeration'] The value HETATOM is not an element of the set {'ATOM', 'HETATM'}.</td>
<td>18303</td>
</tr>
<tr>
<td>11</td>
<td>Error 1824</td>
<td>/datablock/atom_siteCategory/atom_site/group_PDB</td>
<td>'HETATOM' is not a valid value of the local atomic type.</td>
<td>18305</td>
</tr>
<tr>
<td>12</td>
<td>Error 1824</td>
<td>/datablock/atom_siteCategory/atom_site/label_req_id</td>
<td>'is not a valid value of the atomic type 'xinteger'.</td>
<td>18311</td>
</tr>
<tr>
<td>13</td>
<td>Error 1824</td>
<td>/datablock/atom_siteCategory/atom_site/label_req_id</td>
<td>'is not a valid value of the atomic type 'xinteger'.</td>
<td>18344</td>
</tr>
<tr>
<td>14</td>
<td>Error 1824</td>
<td>/datablock/atom_siteCategory/atom_site/label_req_id</td>
<td>'is not a valid value of the atomic type 'xinteger'.</td>
<td>18377</td>
</tr>
</tbody>
</table>
Developments of tools and servers for analysis of NMR experimental data
Discussion Points
Advice requested

- Encourage deposition of additional experimental data
  - Unmerged intensities for X-ray
  - Peak lists for NMR

- Funding for wwPDB Foundation

- Improve visibility of wwPDB