
Worldwide Protein Data Bank Advisory Committee Meeting

September 30, 2011



wwpdb.org

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



wwpdb.org

Overview

Gerard Kleywegt



wwpdb.org

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

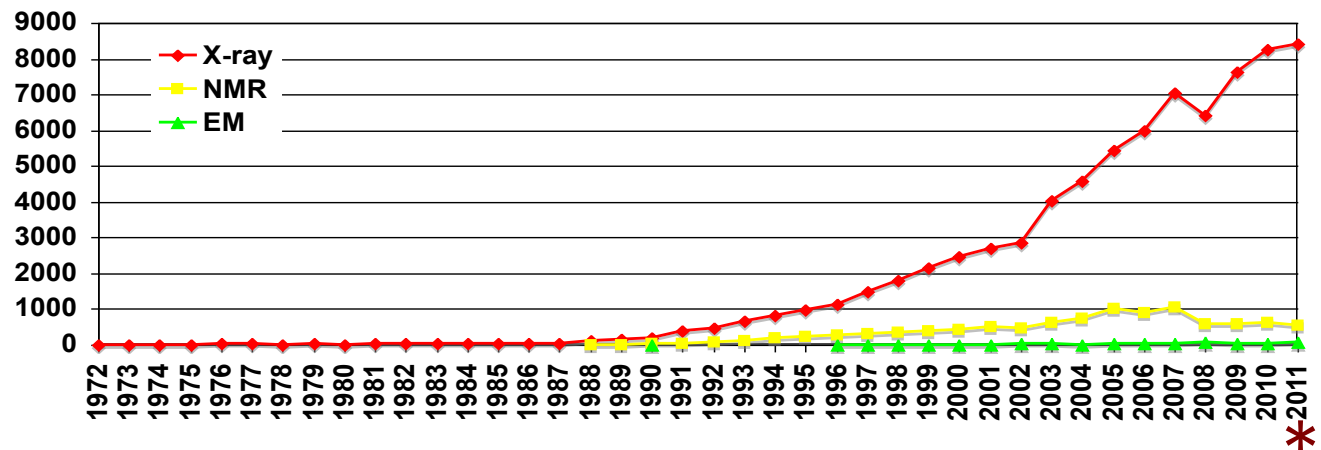
Meeting held Sept 26-27, 2011

PDB Depositions

By deposition and processing site

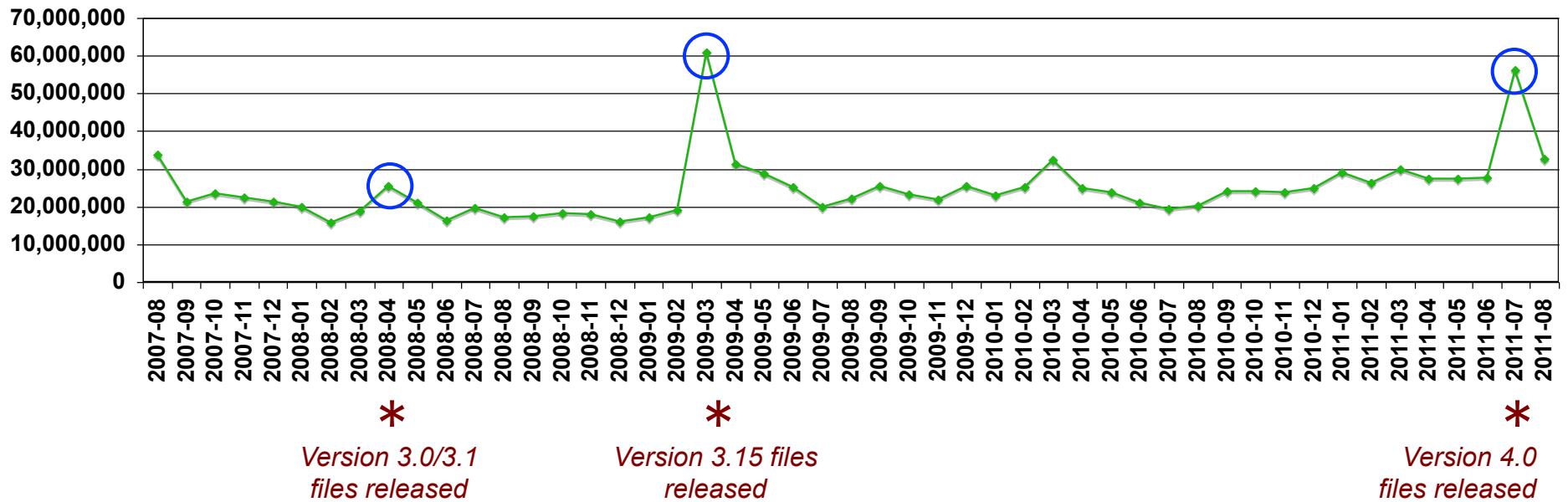
		RCSB PDB	PDBj	PDBe	RCSB PDB	PDBj	PDBe
2000	2983	2445	10	528	2297	158	528
2001	3287	2673	118	496	2408	383	496
2002	3565	2769	289	507	2401	657	507
2003	4830	3488	673	669	3135	1026	669
2004	5508	3796	900	812	3082	1614	812
2005	6678	4507	1166	1005	3563	2110	1005
2006	7282	5145	1052	1085	4252	1945	1085
2007	8130	5399	1603	1128	4703	2299	1128
2008	7073	5452	648	973	4106	1994	973
2009	8300	6715	527	1058	5069	2173	1058
2010	8878	6912	593	1373	5464	2041	1373
2011	6355 (*9088)	4939	392	1024	4159	1171	1024
TOTAL	72869	54240	7971	10658	44639	17571	10658

By experimental method

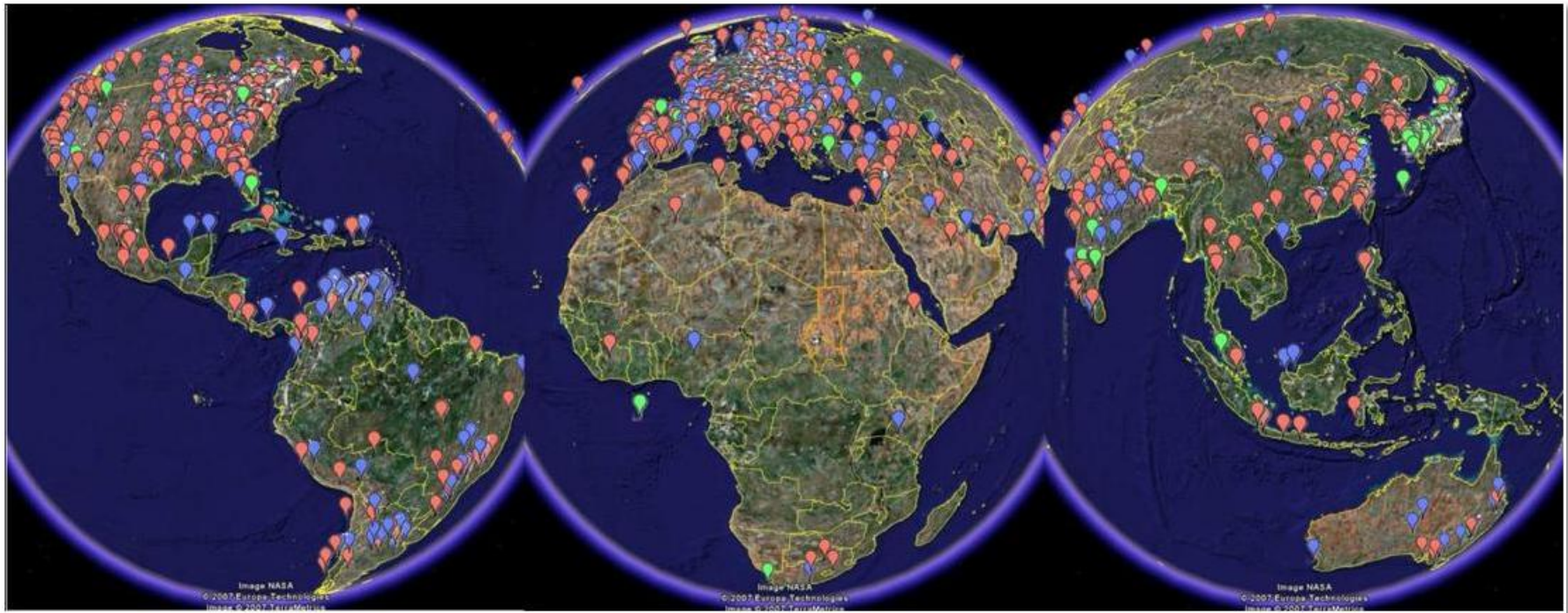



(Updated 14 Sep 2011;
* projection for 2011)

PDB FTP Downloads




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

wwPDB Task Forces

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 Trehwella (University of Sydney), Dmitri Svergun (EMBL Hamburg), Andrej Sali (UCSF), Mamoru Sato (Yokohama City University), John Tainer (Scripps)



Worldwide
Protein Data Bank
Foundation

- 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

PDB40 Symposium

October 28 - 30, 2011
Cold Spring Harbor Laboratory

**Come celebrate four
decades of innovation in
structural biology**

- 215 registered
- 34 travel awards
- 100 posters

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**
- 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

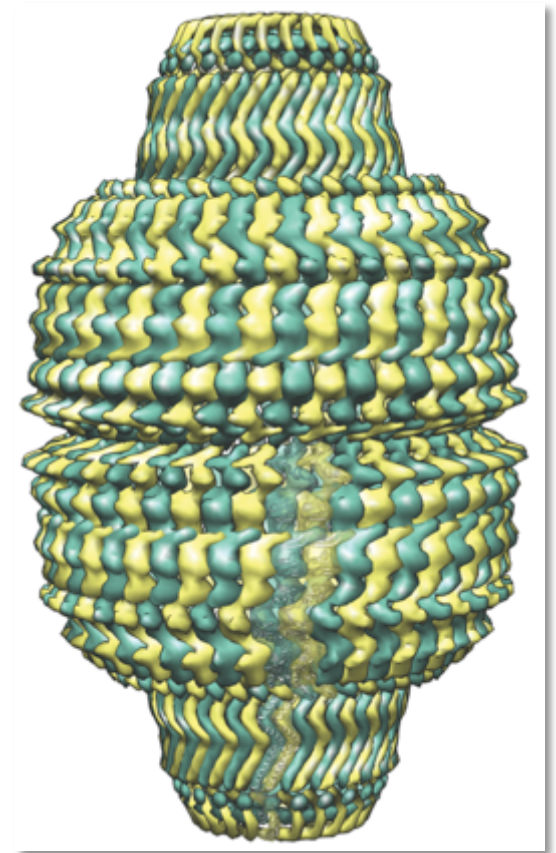
Martha Quesada



wwpdb.org

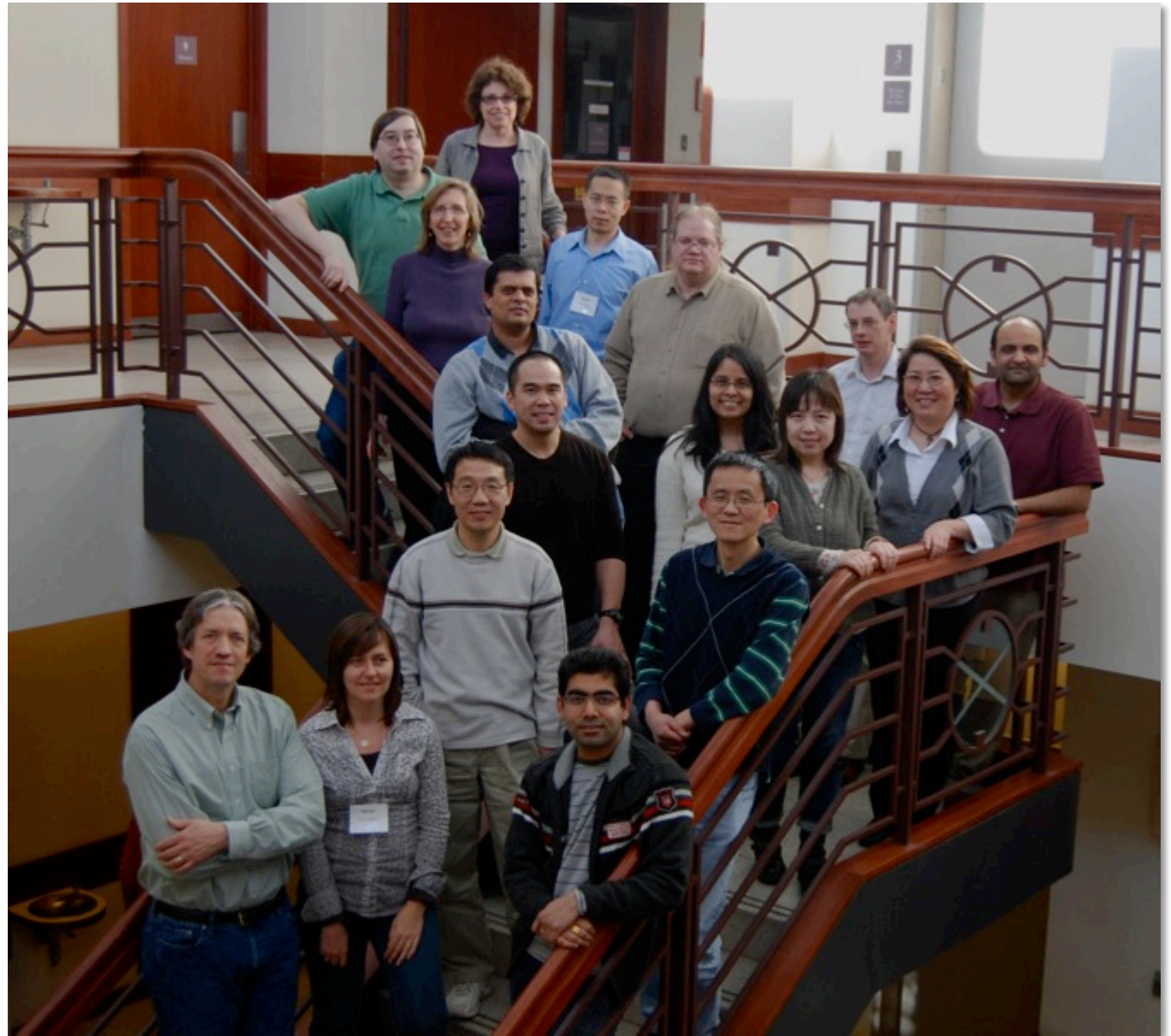
Evolving User Needs

- Larger and more complex biological molecules
- New methods
- Expanded annotation
- Improved quality
 - New validation pipelines
- Higher throughput
 - Automation and validation of routine submissions

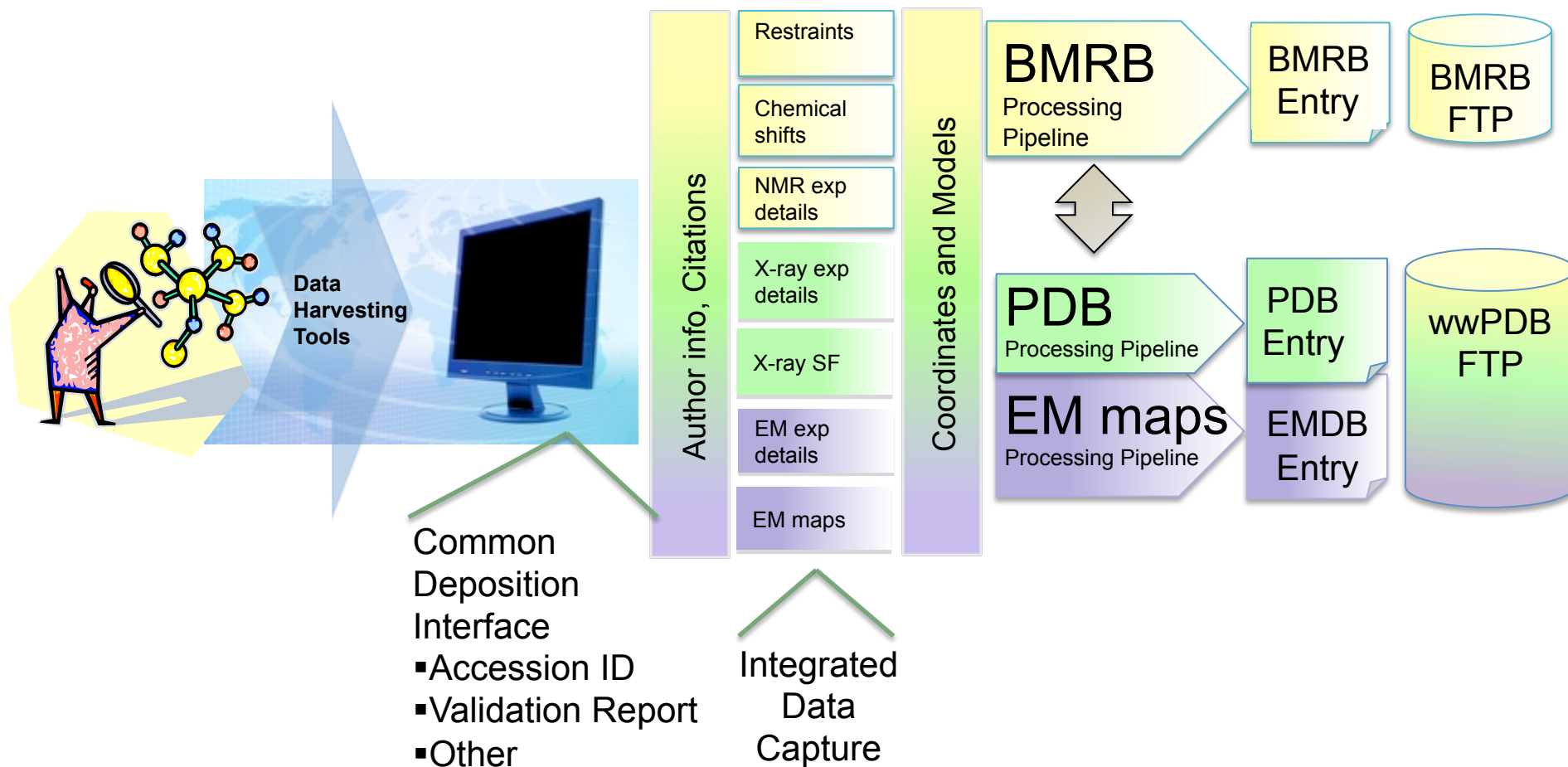


**Common
D&A
Project Team
March 2011**

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



The Vision



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

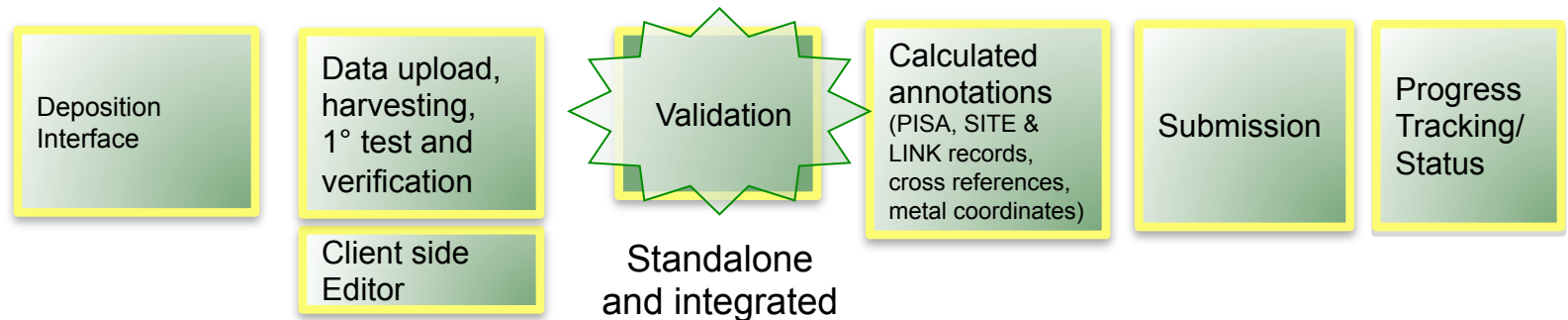
■ Data users

- Higher quality archive



wwPDB Common Deposition and Annotation Pipeline

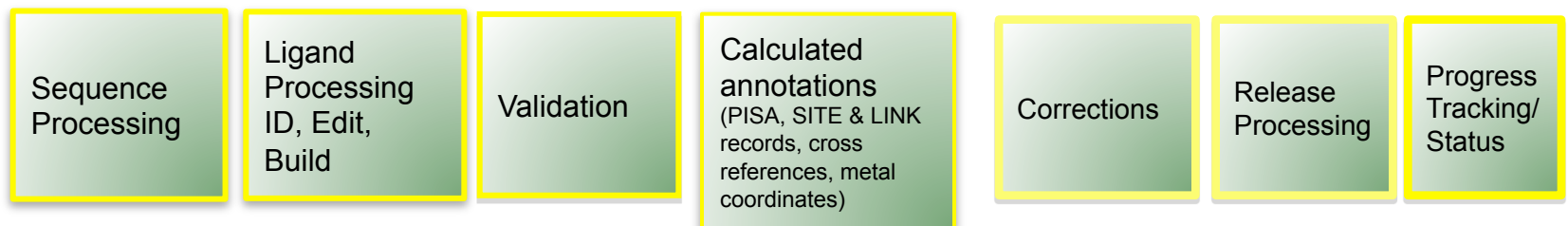
Deposition Pipeline



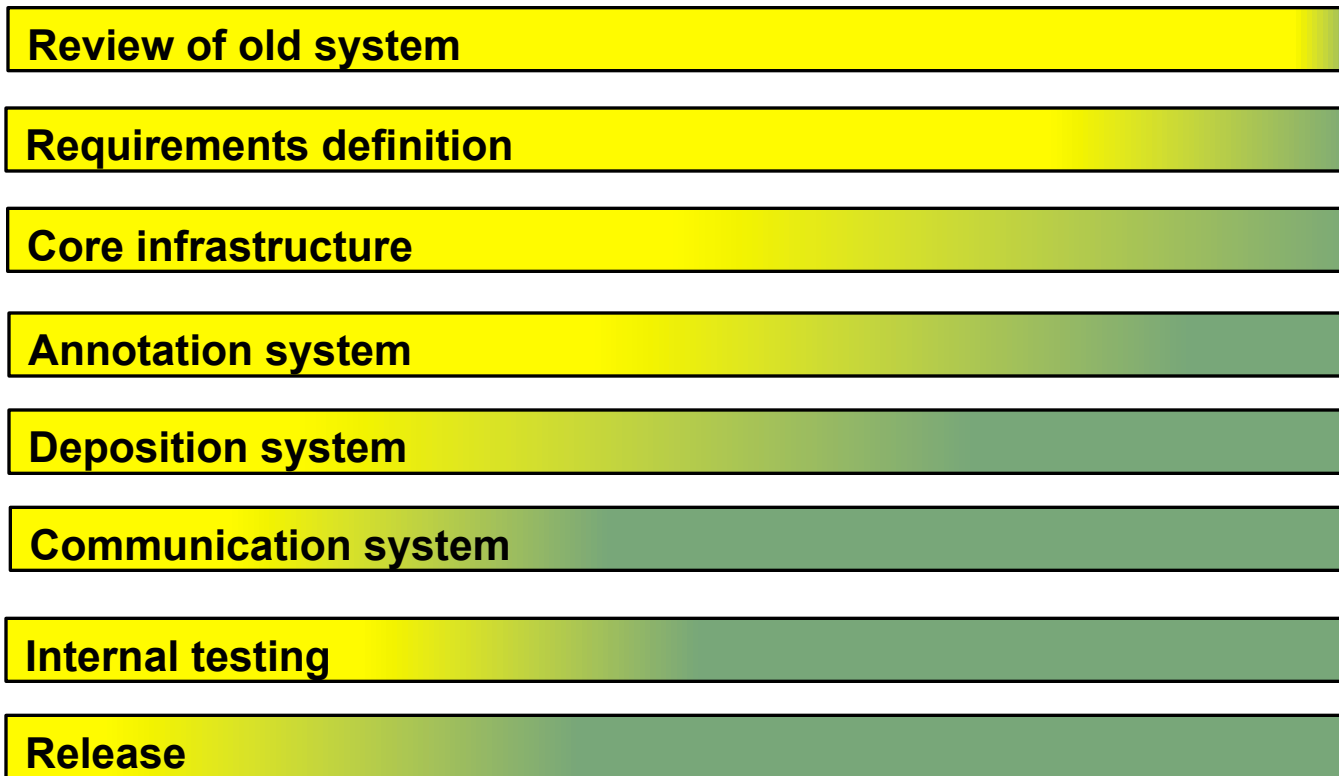
Communication System

Workflow-Automation System

Annotation Pipeline



Development Status as of August 2011



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

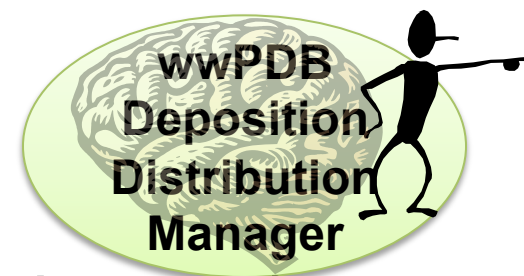
- Automated batch data uploads
- Flexible manual data entry
- Restart deposition and re-upload data without loss of general information
- Build new submissions on previous depositions
- Easily view percentage complete
- Visually review data
- Structure validation reports

Interface Look and Feel: Deposition

The screenshot displays the wwPDB Deposition Tool interface. The main window is titled "wwPDB Deposition Tool" and contains several panels:

- Navigation Panel:** Located on the left side, it features a tree view under "Content/object listing:" with categories like Admin, Files, Reports, Samples, and Sample 1. Sample 1 is expanded to show sub-items: Compound, Sequence, Taxonomy, Expression system, Sample 2, Widget demo, Ligands, Experiments, Other annotation, and Deposition summary.
- Data-entry Panel:** The central area contains multiple form sections for data entry:
 - Compound 2:** Includes "Compound details" with radio buttons for polymer types (Protein, Peptide, Virus, DNA, RNA, Polysaccharide). Fields include "Molecule Name" (lysyl-tRNA synthetase), "Enter chain name(s)", "EC number" (1.1.1.1), and "Compound details" (phosphorothioate link between A 7 ε).
 - Sequence 2:** Includes "Polymer Sequence and cross reference" with a "Yes/No" question, "N-terminal expression tag", "Sequence of the polymer", "Domain or fragment information" (anticodon-binding domain), "Sequence details" (phosphorothioate link between A 7 ε), and "Reference to other database" (Sequence database: UniProt).
 - Taxonomy 2:** Includes "Polymer source" with fields for "Scientific name of organism" (Saccharomyces cerevisiae), "Common name of organism" (Baker's Yeast), "Strain" (gm3c2), "Variant" (BRU isolate), "Cell line" (4-4-20 murine-murine hybridoma), "ATCC number" (ATCC 27355), "Organ" (Heart), "Tissue" (Muscle), "Cell" (B-lymphocyte), and "Other details" (German collection of microorganisms (DSMZ)).
 - Expression system 2:** Includes "Expression system details" with radio buttons for "Made using recombinant techniques", "Purified from natural source", and "Chemically synthesised". Fields include "Expression system organism" (Escherichia coli, or Saccharomyces), "Expression system strain" (BALB/C), "Expression system variant" (sv189), "Expression system cell line" (BG2087), "Expression system vector" (Baculovirus), and "Expression system" (yep213/leu58hiscyc1).
- Communication Panel:** Located on the right side, it contains a "Communication/news" section with two entries:
 - 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

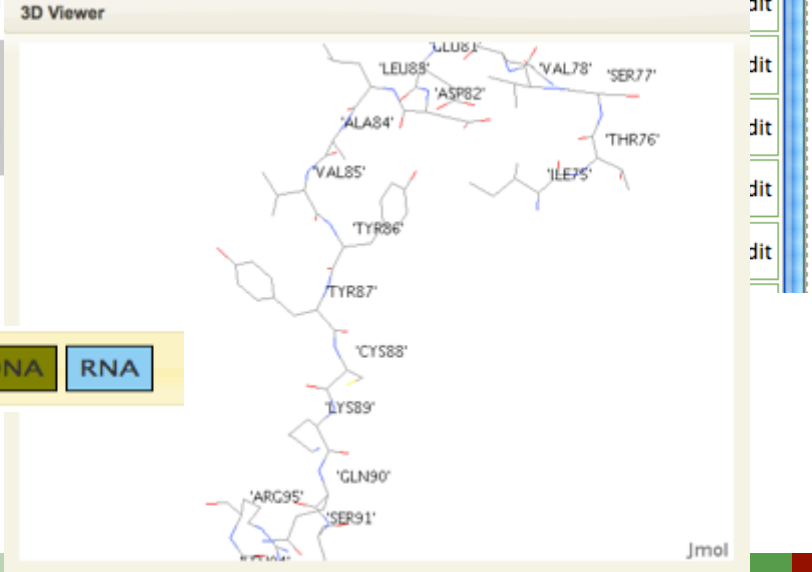
- Sequence module - unit tested
- Ligand module in production benchmark testing
 - Ligand processing
 - “Chopper”
 - Editor
- Workflow Manager operational with ligand and sequence modules
 - Load testing (tested via shared server: RCSB PDB)
 - Batch workflow processing working at both sites
 - Functionality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

Graphical Interface for Sequence Processing

```

AUTH PDB:A V(1)  DIVMSQSPSSLAVSAGEKVTMSCKSSQSLNSTRTRKNYLAWYQQKPGQSPKLLIYWASTR
XYZ PDB:A V(1)  DIVMSQSPSSLAVSAGEKVTMSCKSSQSLNSTRTRKNYLAWYQQKPGQSPKLLIYWASTR
UNP:Q52L64 (R1,V1) DIVMSQSPSSLAVSAGEKVTMSCKSSQSLNSTRTRKNYLAWYQQKPGQSPKLLIYWASTR
1 |---+---|---+---|---+---|---+---|---+---|---+---|
AUTH PDB:A V(1)  ESGVPRFRFTGSGSGTDFTLTIITSVQAEDLAVYYCKQSYN..LRTFGGTTKLEIKRADAAP
XYZ PDB:A V(1)  ESGVPRFRFTGSGSGTDFTLTIITSVQAEDLAVYYCKQSYN..LRTFGGTTKLEIKRADAAP
UNP:Q52L64 (R1,V1) ESGVPRFRFTGSGSGTDFTLTIITSVQAEDLAVYYCKQSYN..LRTFGGTTKLEIKRADAAP
61 |---+---|---+---|---+---|---+---|---+---|---+---|
AUTH PDB:A V(1)  TVSIFPPSSEQLTSGGASVVCFLNFFPKDINVKWKIDGSERQNGVLSWTDQDSKDY
XYZ PDB:A V(1)  TVSIFPPSSEQLTSGGASVVCFLNFFPKDINVKWKIDGSERQNGVLSWTDQDSKDY
UNP:Q52L64 (R1,V1) TVSIFPPSSEQLTSGGASVVCFLNFFPKDINVKWKIDGSERQNGVLSWTDQDSKDY
121 |---+---|---+---|---+---|---+---|---+---|---+---|
AUTH PDB:A V(1)  SMSSTLTTLTKDEYERHNSYTCEATHKTSTSPIVKSFNRE
XYZ PDB:A V(1)  SMSSTLTTLTKDEYERHNSYTCEATHKTSTSPIVKSFNRE
UNP:Q52L64 (R1,V1) SMSSTLTTLTKDEYERHNSYTCEATHKTSTSPIVKSFNRE
181 |---+---|---+---|---+---|---+---|---+---|
    
```

POSITION	AUTH PDB:A	ALIGNED SEQUENCE	RESIDUE	ANNOTATION DETAILS
15	ALA	UNP:Q52L64 (R1,V1)	VAL	engineered mutation
31	ASN	UNP:Q52L64 (R1,V1)	TYR	engineered mutation
33	ARG	UNP:Q52L64 (R1,V1)	TYR	Shift/Click to edit
34	THR	UNP:Q52L64 (R1,V1)	ASN	Shift/Click to edit
		UNP:Q52L64		dit



Graphical Interface for Ligand Processing

Instance: 1_L_B12_1_

TOP CANDIDATE RESULTS FOR: 1_L_B12_1_			
CANDIDATE ID	ASSIGN AS:	COMPOSITE SCORE	COMPARE
B12	<input type="radio"/>	76 / 100 / 92	<input checked="" type="checkbox"/>
CNC	<input type="radio"/>	76 / 97 / 92	<input checked="" type="checkbox"/>
COB	<input type="radio"/>	76 / 98 / 92	<input type="checkbox"/>
COY	<input type="radio"/>	82 / 85 / 77	<input type="checkbox"/>
NON-CANDIDATES FOR COMPARISON			
HEM	Must Force Assign	n.a.	<input checked="" type="checkbox"/>

Force Assign as:

Show Rerun Search Form

Edit / Create New Ligand

Assign

Enter non-candidate chem component ID to add to Comparison Panel below: Add

Comparison Panel

2D

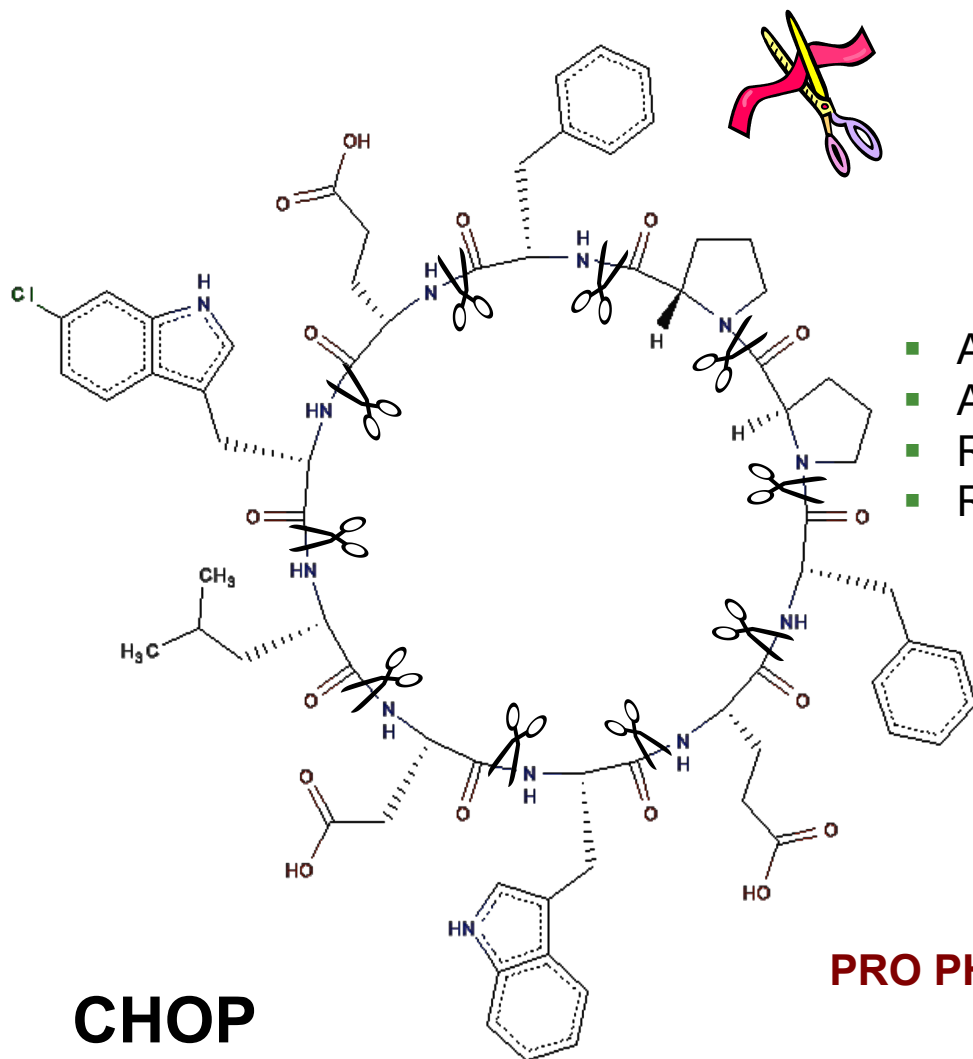
3D

Atom Map

Auth Instance ID:	1_L_B12_1_	Top Dictionary Hit:	B12	Dictionary ID:	CNC	Dictionary ID:	HEM
Name:	None	Name:	COBALAMIN	Name:	CO-CYANOCOBALAMIN	Name:	PROTOPORPHYRIN IX CON...
Formula:	C62 H89 Co N13 O14 P	Formula:	C62 H88 Co N13 O14 P	Formula:	C63 H88 Co N14 O14 P	Formula:	C34 H32 Fe N4 O4

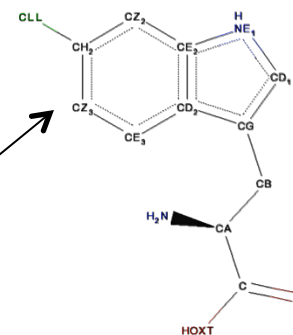
Peptide Ligand Chopper

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



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

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



wwpdb.org

NMR

John Markley



wwpdb.org

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 \leftrightarrow BMRB \leftrightarrow 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

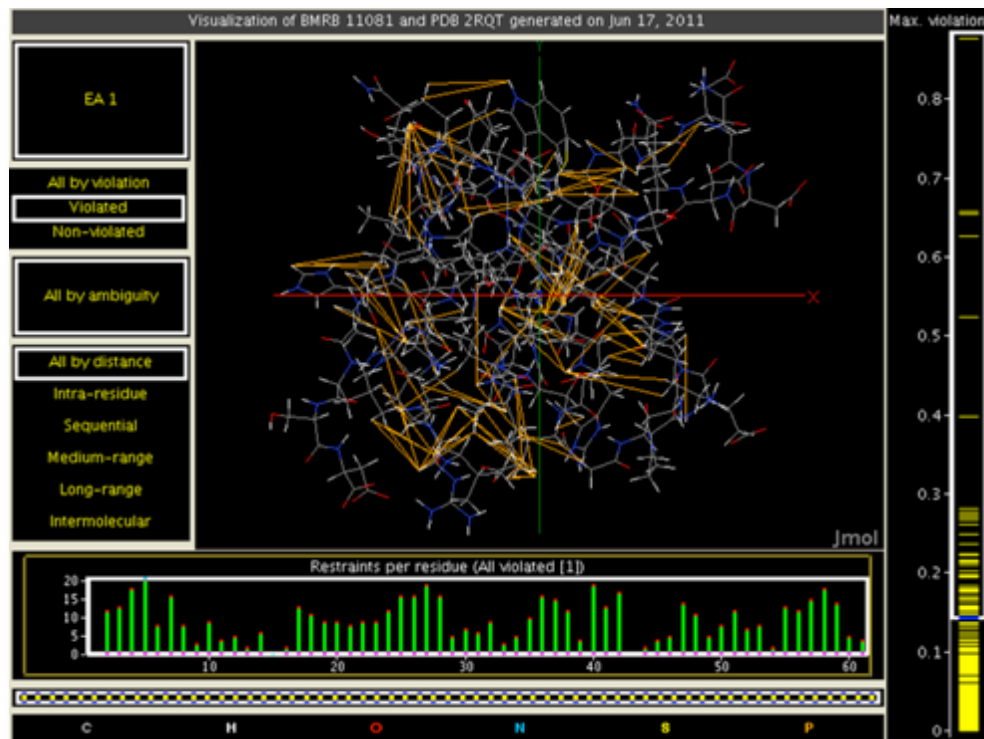
- Restraints go to PDB and are sent to BMRB for processing
- Software used in restraint processing was developed in collaboration with Wim Vranken, Jurgen Doreleijers, Geerten Vuister & Gert Vriend

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, PDBe, 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

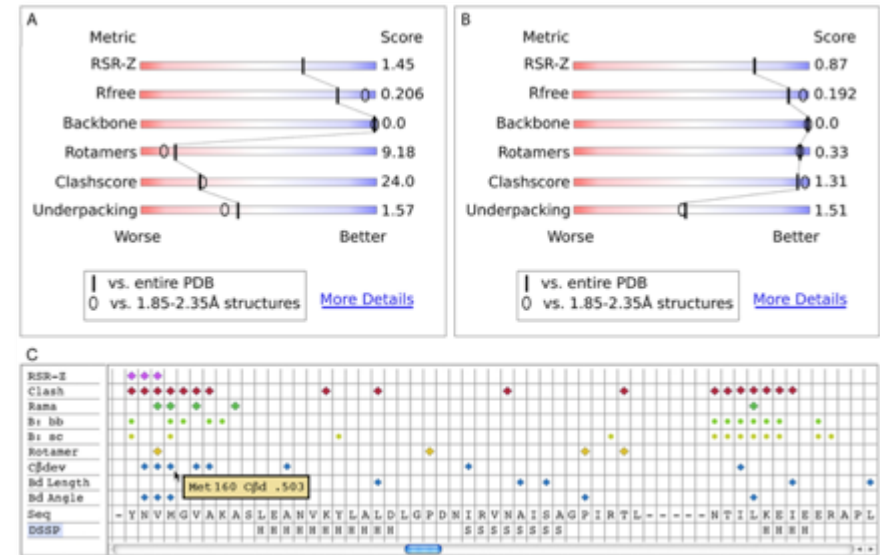
Helen M. Berman



wwpdb.org

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*

Randy J. Read, Paul D. Adams, W. Bryan Arendall III, Axel T. Brunger, Paul Emsley, Robbie P. Joosten, Gerard J. Kleywegt, Eugene B. Krissinel, Thomas Lütkeke, Zbyszek Otwinowski, Anastassis Perrakis, Jane S. Richardson, William H. Sheffler, Janet L. Smith, Ian J. Tickle, Gert Vriend and Peter H. Zwart
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

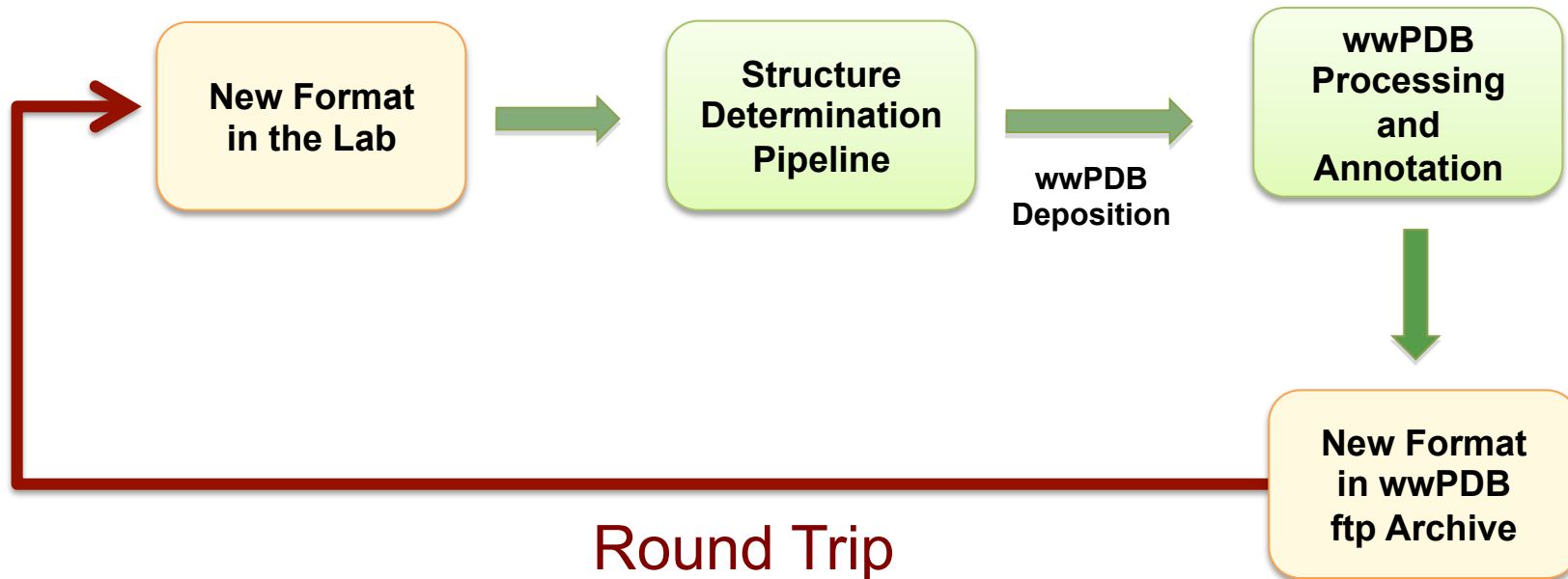
Why Create a New PDB Format?

- 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



Format meeting 26/27 Sept 2011



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

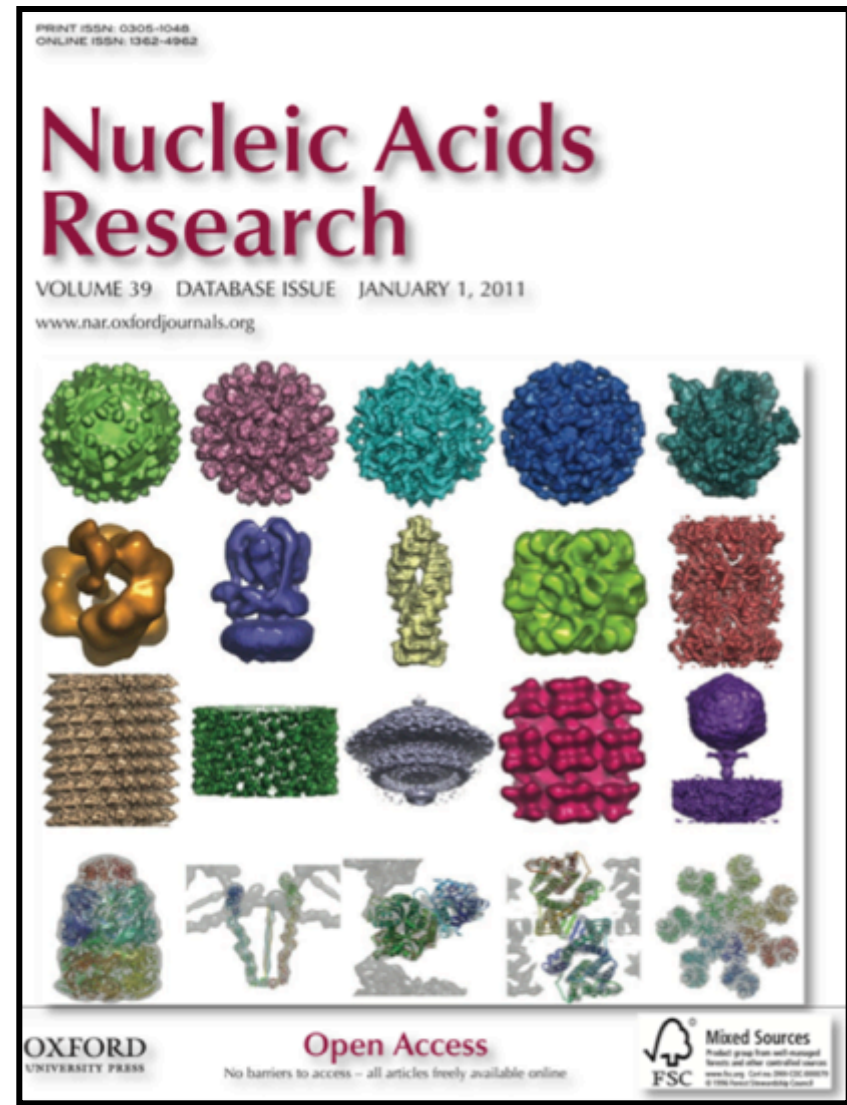
3D Electron Microscopy



wwpdb.org

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



EMDataBank
Unified Data Resource for 3DEM

One stop shop for 3DEM deposition and retrieval
2011-07-27: 1109 EMDB map entries, 402 PDB coordinate entries

Home About Deposit Search Tools Events News Links Help

The Unified Data Resource for 3-Dimensional Electron Microscopy

EMDataBank.org is a joint effort of the Protein Databank in Europe (PDBe), the Research Collaboratory for Structural Bioinformatics (RCSB), and the National Center for Macromolecular Imaging (NCMI) to create a global deposition and retrieval network for cryoEM map, model and associated metadata, as well as a portal for software tools for standardized map format conversion, map, segmentation and model assessment, visualization, and data integration.

Recently released entries

1807 (PDBe | RCSB) 2010-10-21 RELEASED ON 2011-07-19 singleParticle 28Å **NEW**
Structural basis for allosteric regulation of human ribonucleotide reductase by nucleotide-induced oligomerization.
Fairman JW, Wijerathna SR, Ahmad MF, Xu H, Nakano R, Jha S, Prendergast J, Welin RM, Flodin S, Roos A, Nordlund P, Li Z, Walz T, Dealwis CG

5200 (PDBe | RCSB) 2010-05-13 RELEASED ON 2011-07-18 Icosahedral 5.4Å **NEW**
5.4-Angstrom cryoEM structure of the Bordetella Bacteriophage capsid
L.Jin, A.Hodes, W.H.Hui, X.Zhang, X.Zhang, X.Yu, J.F.Miller, Z.H.Zhou

News **RSS**

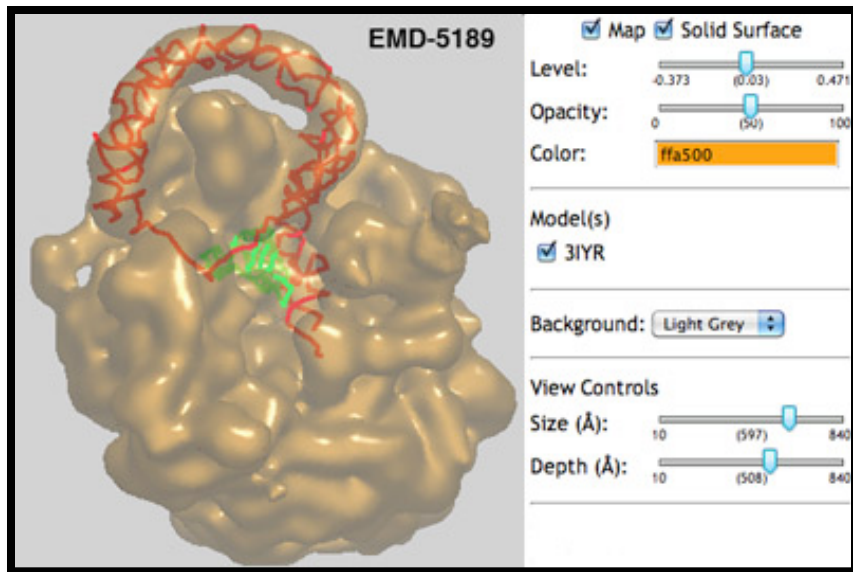
Improved 3D Viewing for EM Maps and Models

We are pleased to announce improved 3D web-based viewing of maps and associated coordinate models from EMDatabank atlas pages. The molecular visualization program OpenAstexViewer has been adapted by the EMDatabank team to display EM maps and their associated coordinate models.

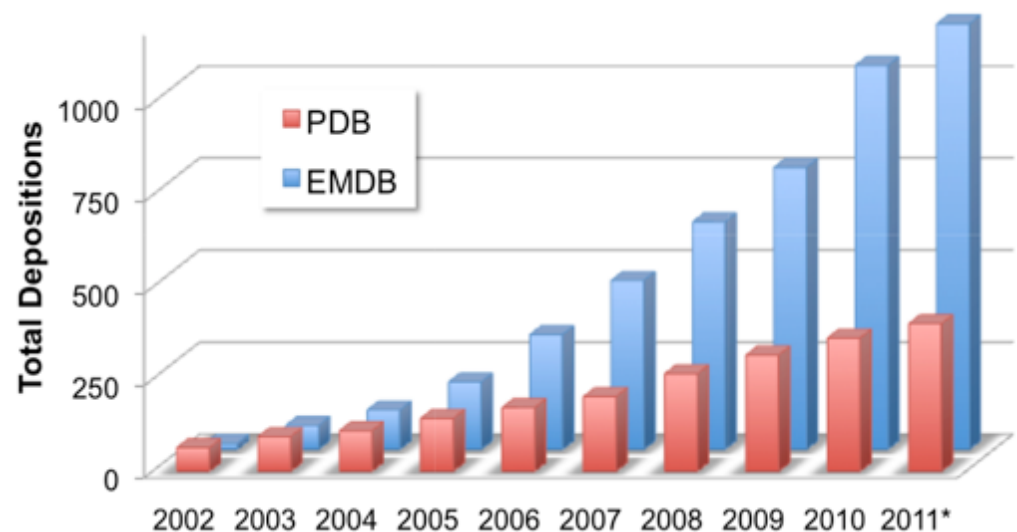
- **Joint map + coordinate deposition service**
- News, software list, information about dictionaries, conventions, FAQ, community links
- Search by ID, author, sample type, keyword, deposition date
- Map and map+model 3D java viewers

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

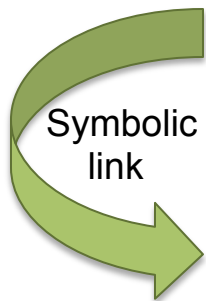


September 21, 2011:

1140 map entries, 407 PDB coordinate entries

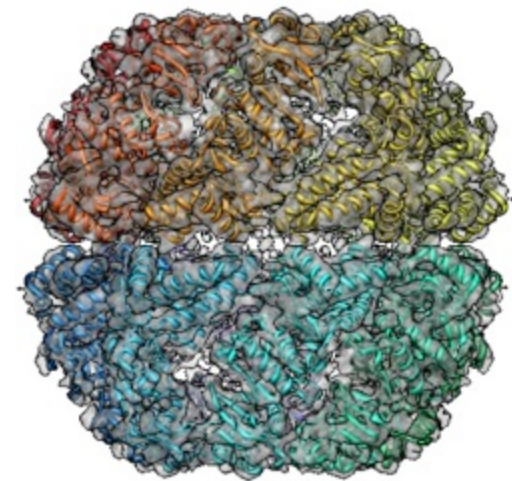
Planned Archive Integration

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



Index of [ftp://ftp.wwpdb.org/pdb/
data/structures/divided/em/iz/3izk/
3izk-map.ccp4.gz](ftp://ftp.wwpdb.org/pdb/data/structures/divided/em/iz/3izk/3izk-map.ccp4.gz)

Index of [ftp://ftp.wwpdb.org/emdb/
structures/EMD-5247/map/
emd_5247.map.gz](ftp://ftp.wwpdb.org/emdb/structures/EMD-5247/map/emd_5247.map.gz)



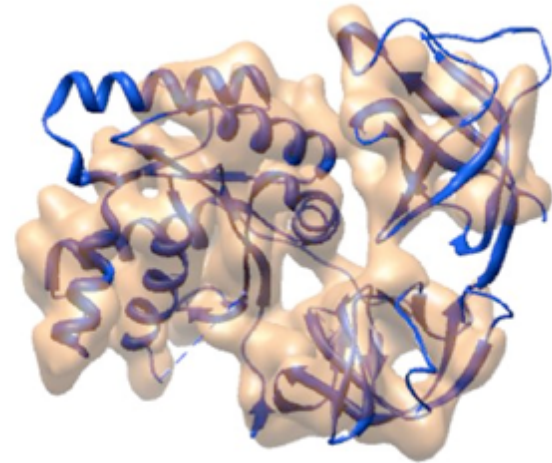
Archaeal group II chaperonin

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.org

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. C α or 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_{\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 Trehwella (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

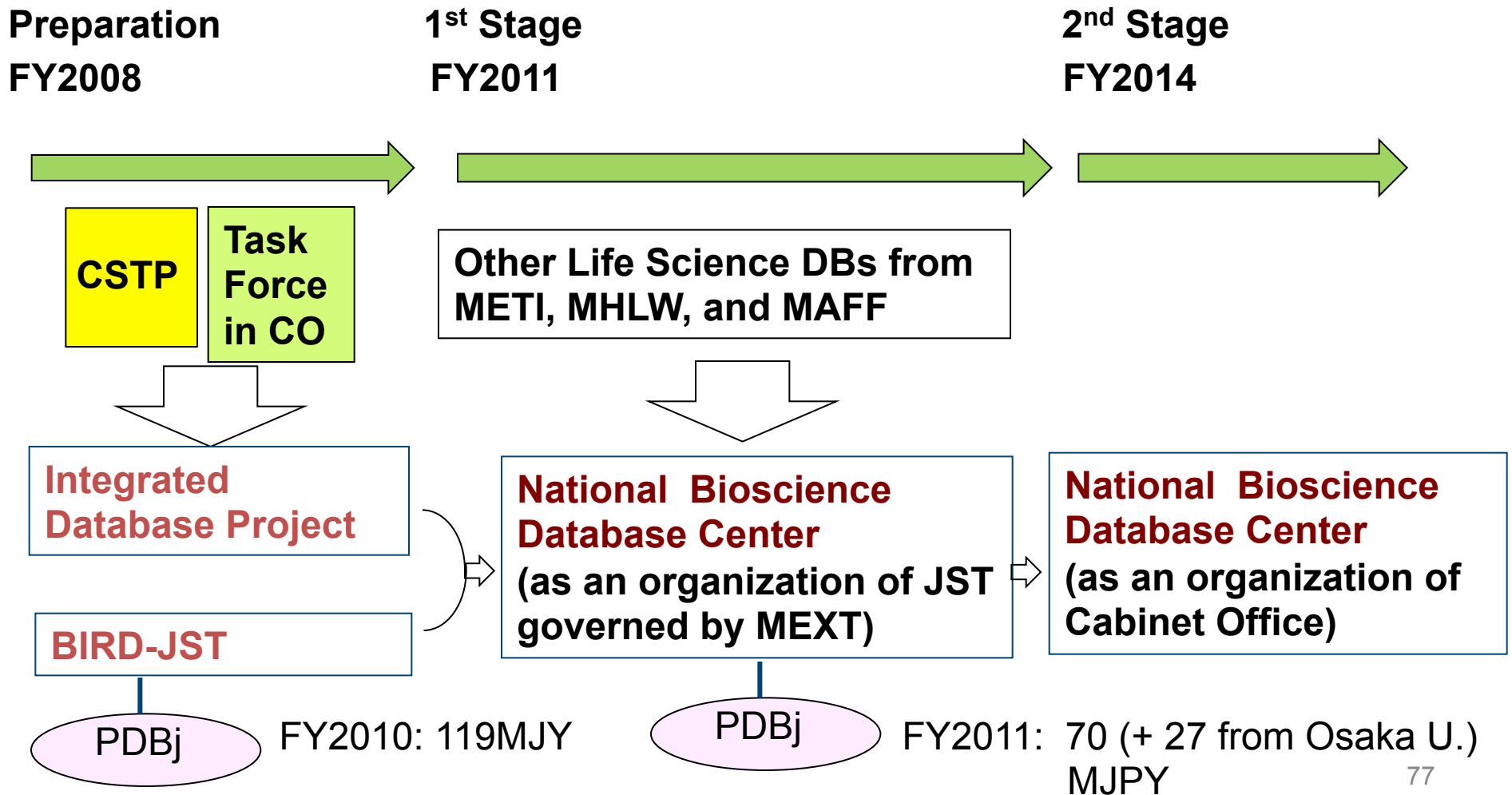


wwpdb.org

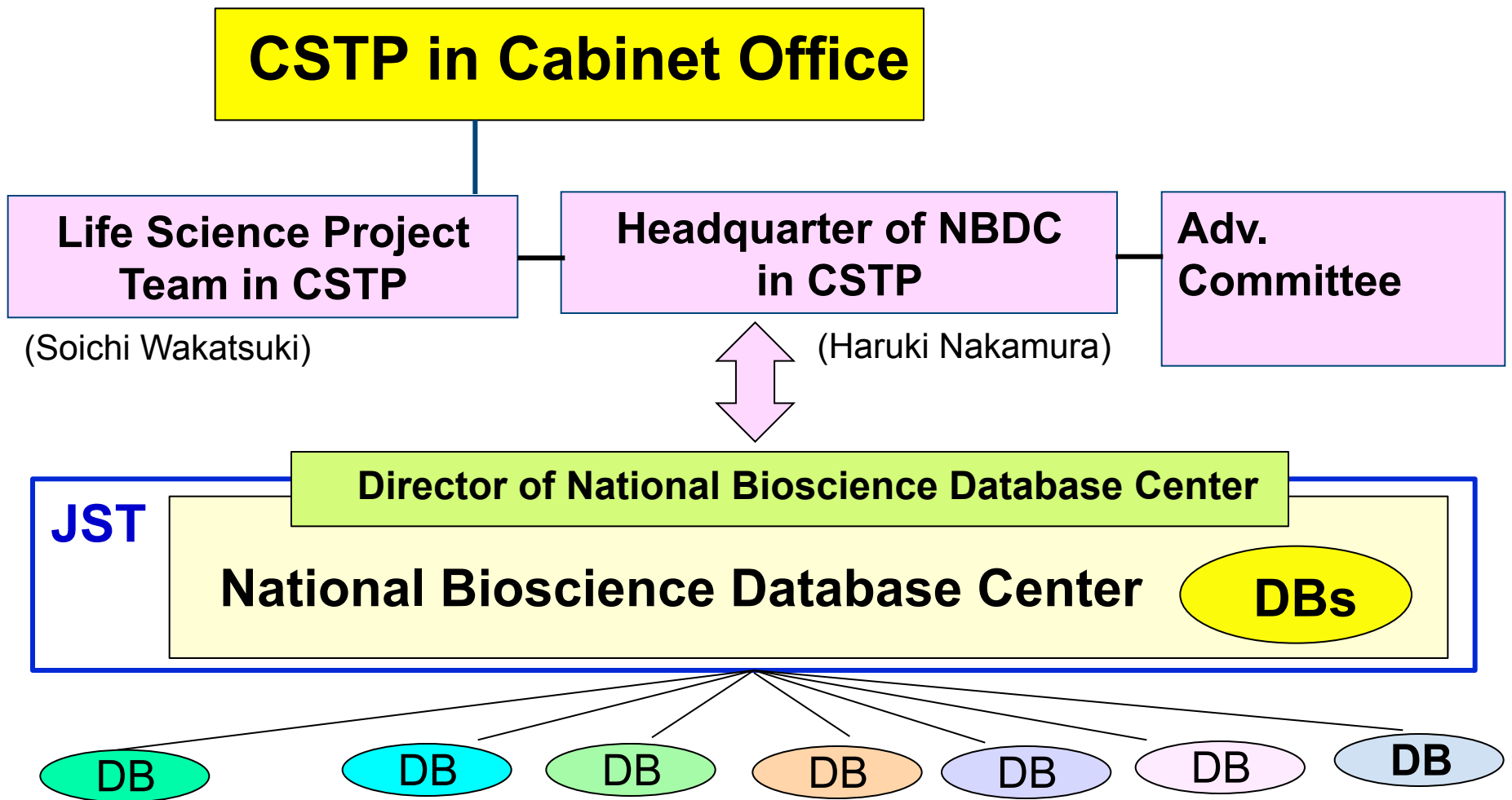
National Bioscience Database Center in Japan and PDBj Funding

- **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)



Organization of National Bioscience Database Center (NBDC)



PDBj is a member of NBDC, Japan

<http://biosciencedbc.jp/>

The screenshot shows the NBDC website interface. At the top, the NBDC logo is on the left, and the text "National Bioscience Database Center" is in the center. To the right, there is a logo for "JST 科学技術振興機構" and a font size selector (L, M, S). Below the header, there are navigation links: "Home", "About Us", "Contact Us", and "Link". A search bar is also present with the text "Japanese | Sitemap | Site search" and a "search" button.

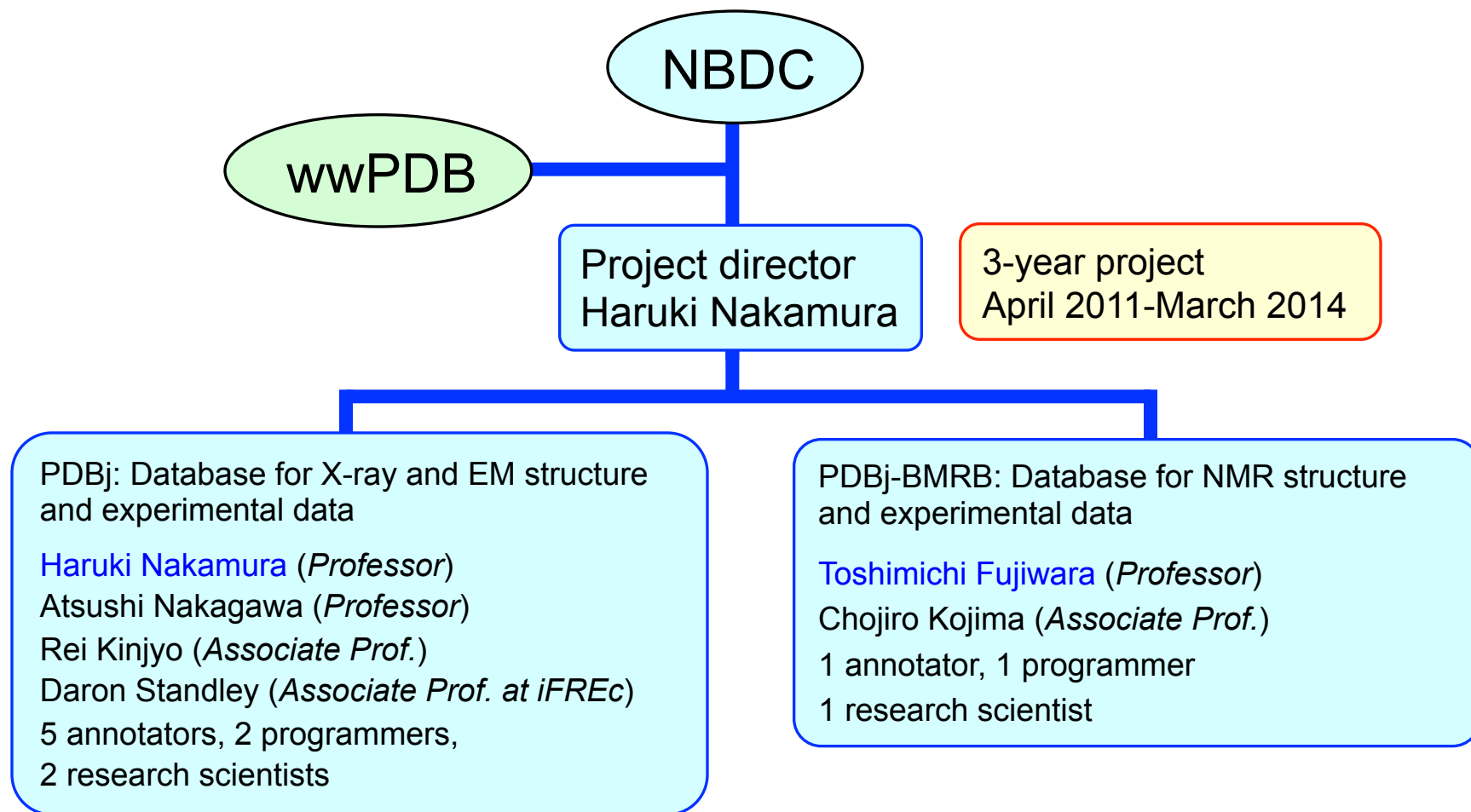
The main content area is titled "Organization". It features a hierarchical organizational chart for the National Bioscience Database Center (NBDC). At the top of the chart is the "Council for Science and Technology Policy (CSTP)", which oversees the "Life Science Project Team" and the "DB Integration Promotion Task Force". Below these is the "National Bioscience Database Center (NBDC)". The NBDC is led by the "Director-General Michio Oishi, Ph.D." and the "Deputy Director-General Toshihisa Takaai, Ph.D.". A "Steering Committee" is shown with a bidirectional arrow connecting it to the Director-General. The NBDC is divided into several departments and programs:

- Department of Planning and Management** (green box)
- Researchers** (green box)
- Program Concerning Technology Development for Database Integration** (blue box), which includes:
 - Research Supervisor: Takeshi Nagase, Ph.D.
 - Research Subject
 - Three Research Advisors
- Program for Coordination Toward Integration of Related Databases** (blue box), which includes:
 - Research Supervisor: Toshihisa Takaai, Ph.D.
 - Research Subject
 - Three Research Advisors

Below the organizational chart, there is a section titled "About the Director-General (Michio Oishi, Ph.D.)". It includes a small portrait photo of Michio Oishi. Below the photo, the text reads:

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



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 for Semantic Web

(Recently developed by PDBj: Akira R. Kinjo et al.)

<http://pdbj.org/rdf>

PDB/RDF [About PDB/RDF](#)
[chem_comp/RDF Search](#)

PDB ID (e.g., '7RSA')

property: (e.g.,
'PDBo:entity.pdbx_description')

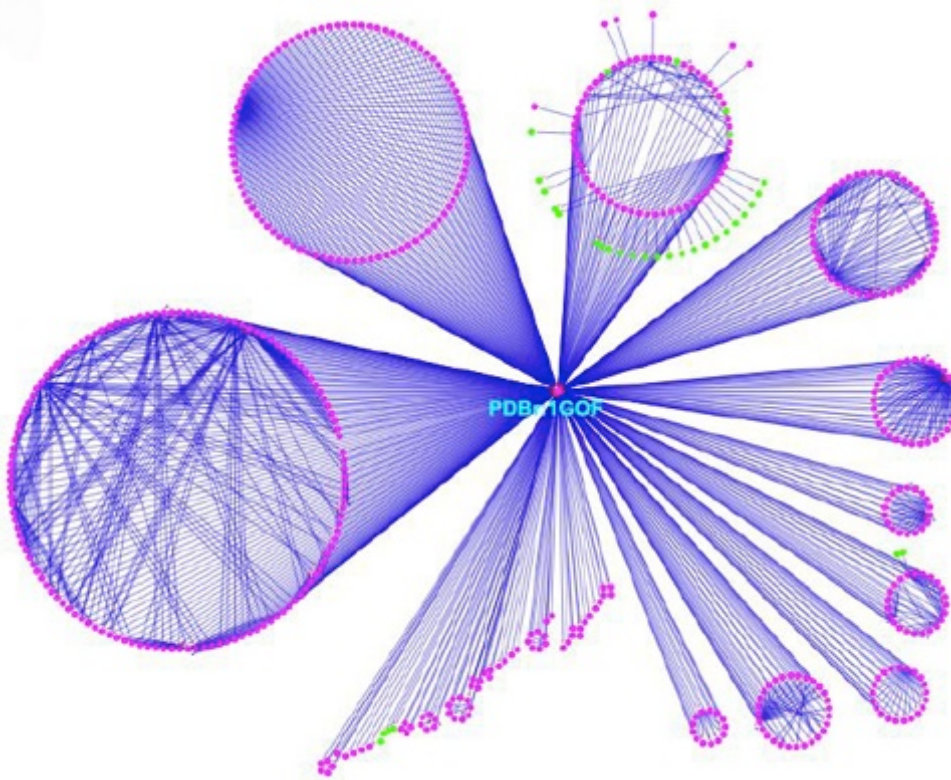
keywords: (e.g., 'alcohol')

Download XSLT stylesheet for converting PDBML to RDF: [PDBML2rdf.xsl.gz](#)
(zipped 22KB)

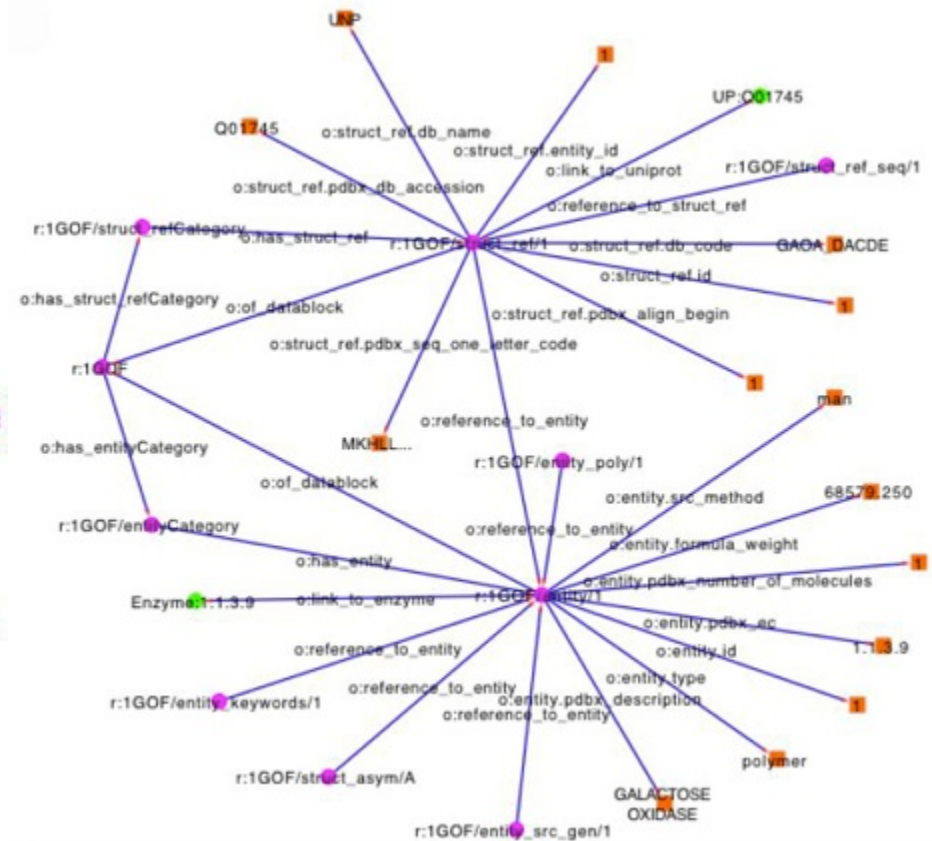
*) Some web browsers such as Internet Explorer will display the plain XML file. Please use other browsers that can handle XSLT, e.g. Firefox, Google Chrome, or Safari etc..

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**.

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)

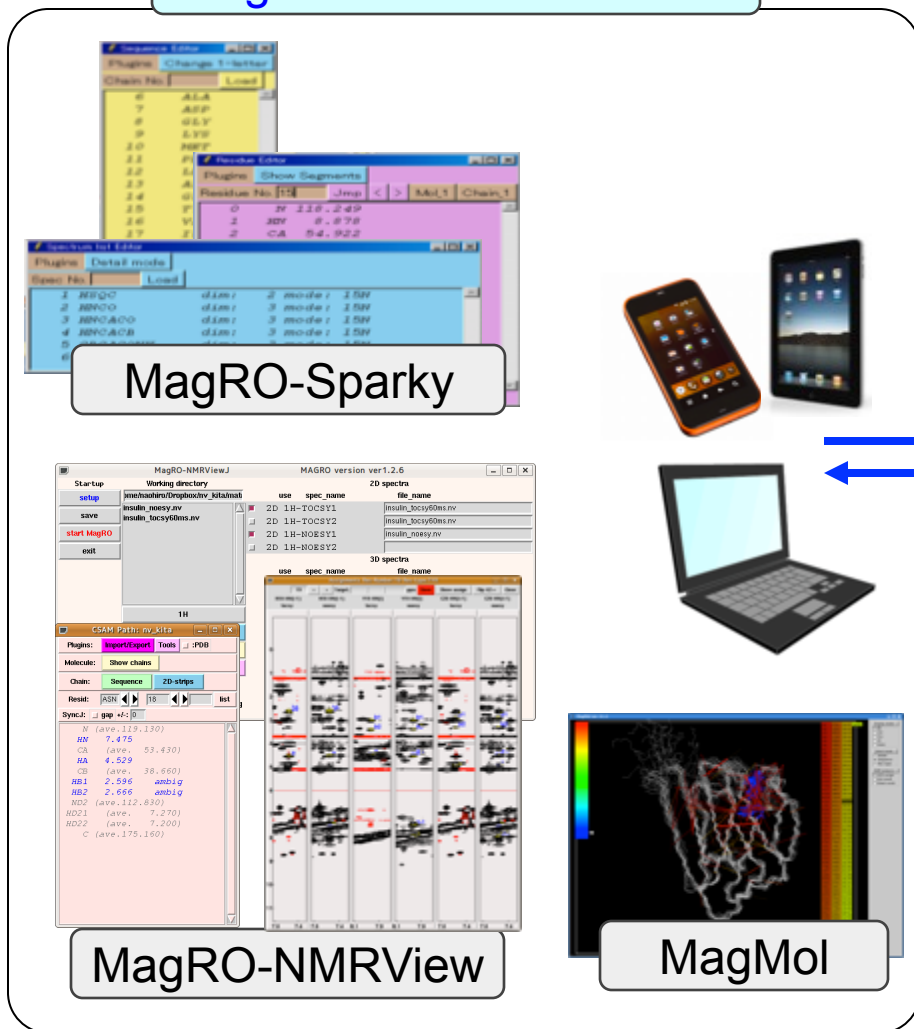
Validation Results

Validated file: 1hpi.xml

No.	Error code	XPath	Message	line
1	Error 1826	/datablock/atom_siteCategory/atom_site/pdbx_formal_charge	'1.5' is not a valid value of the local union type.	1721
2	Error 1826	/datablock/atom_siteCategory/atom_site/pdbx_formal_charge	'1.0' is not a valid value of the local union type.	1754
3	Error 1826	/datablock/atom_siteCategory/atom_site/pdbx_formal_charge	'2.0' is not a valid value of the local union type.	1787
4	Error 1824	/datablock/atom_siteCategory/atom_site/label_seq_id	* is not a valid value of the atomic type 'xs:integer'.	18113
5	Error 1824	/datablock/atom_siteCategory/atom_site/label_seq_id	* is not a valid value of the atomic type 'xs:integer'.	18146
6	Error 1824	/datablock/atom_siteCategory/atom_site/label_seq_id	* is not a valid value of the atomic type 'xs:integer'.	18179
7	Error 1824	/datablock/atom_siteCategory/atom_site/label_seq_id	* is not a valid value of the atomic type 'xs:integer'.	18212
8	Error 1824	/datablock/atom_siteCategory/atom_site/label_seq_id	* is not a valid value of the atomic type 'xs:integer'.	18245
9	Error 1824	/datablock/atom_siteCategory/atom_site/label_seq_id	* is not a valid value of the atomic type 'xs:integer'.	18278
10	Error 1840	/datablock/atom_siteCategory/atom_site/group_PDB	[facet 'enumeration'] The value 'HETATOM' is not an element of the set ('ATOM', 'HETATM').	18305
11	Error 1824	/datablock/atom_siteCategory/atom_site/group_PDB	'HETATOM' is not a valid value of the local atomic type.	18305
12	Error 1824	/datablock/atom_siteCategory/atom_site/label_seq_id	* is not a valid value of the atomic type 'xs:integer'.	18311
13	Error 1824	/datablock/atom_siteCategory/atom_site/label_seq_id	* is not a valid value of the atomic type 'xs:integer'.	18344
14	Error 1824	/datablock/atom_siteCategory/atom_site/label_seq_id	* is not a valid value of the atomic type 'xs:integer'.	18377

Developments of tools and servers for analysis of NMR experimental data

MagRO-tools for client PC



NMR machine



MagRO-Server



BioResMagBank

Discussion Points



wwpdb.org

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