Worldwide Protein Data Bank Advisory Committee Meeting

September 27, 2013

WORLDWIDE PDB
PROTEIN DATA BANK

wwpdb.org

Agenda

9:00	Welcome	Stephen Burley
9:15	Overview	Helen Berman
9:45	D&A Tool	Martha Quesada
10:30	*Break*	
11:00	Format issues, Validation & Experimental methods	Gerard Kleywegt
12:00	*Lunch*	
1:30	NMR	John Markley
2:00	Outreach	Haruki Nakamura
2:30	Matters Arising, Discussion	
3:00	*Break*	
4:00	Executive Session & Feedback	
5:00	Adjourn	

Overview

Helen Berman

WORLDWIDE PDB
PROTEIN DATA BANK

wwpdb.org

wwPDB
October 2012 - September 2013

- Continued growth of archive
- Increased use of data
- PDBx implemented in X-ray packages
- First large structures released as “non-split” PDBx files
- New validation reports since August 1, 2013
- BIRD released
- Structure-factor remediation
- Common Tool in testing
- Task Force activities
- New wwPDB Charter and Terms of Reference in place
- Funding
- wwPDB Foundation

PDB Depositions

Year	Total Depositions	Deposited To			Processed By		
		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	9250	7172	582	1496	5938	1816	1496
2012	9972	7693	603	1676	6411	1885	1676
2013	7580 *	5784	523	1273	4876	1431	1273
TOTAL	93316	69950	9287	14079	57705	21532	14079

Note: Includes theoretical models and entries later withdrawn or obsoleted
Last Updated: 18 Sep 2013

***10,700 depositions projected for 2013**
100,000th entry expected to be released Spring 2014

2012 FTP & Rsync Entry Downloads

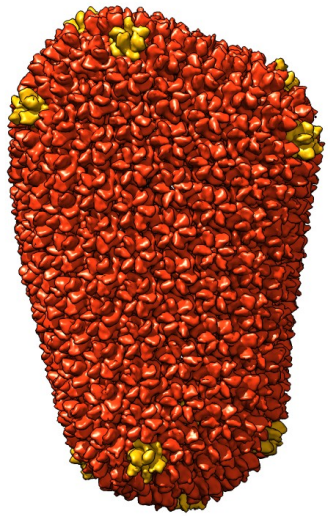
RCSB PDB
2012: 298 million
2011: 282 million
2010: 159 million

PDBe
2012: 46 million
2011: 59 million
2010: 34 million

PDBj
2012: 21 million
2011: 38 million
2010: 16 million

Format

- PDBx addresses limitations in molecular size and complexity and extensibility of existing PDB format
- Software developers implemented PDBx/ mmCIF in major X-ray packages

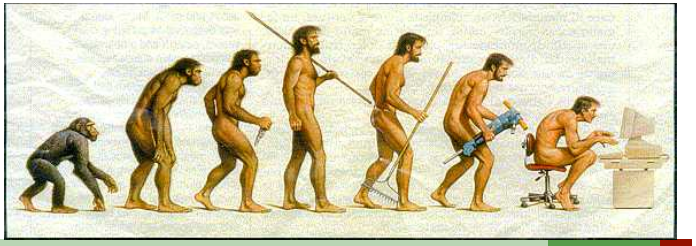


First structures deposited and released in PDBx format, May 2013.
Mature HIV-1 capsid structure by cryo-electron microscopy and all-atom molecular dynamics.
Gongpu Zhao *et al*., *Nature* 497, 643-646 (2013)

Common Tool for Deposition and Annotation

2013 Project Delivery Milestones

- ✓Deposition and Annotation pipelines in testing
  - ✓Internal pipeline testing
  - ✓Testing for developers
- ✓Public demonstration and input
- ✓Community deposition testing
- ✧Early 2014: Common D&A system in production at all sites



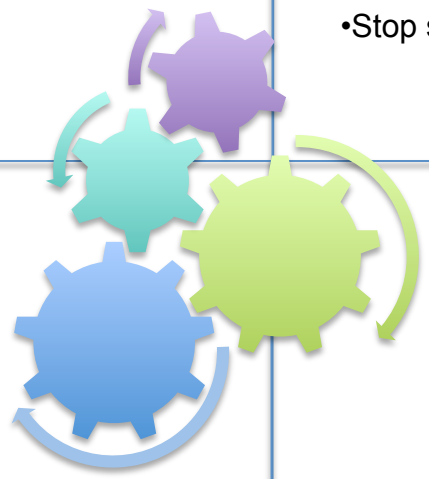
Coordinated Transition

Deposition & Annotation

✓New annotation systems being tested
✓New deposition systems being tested
✓Weekly update supporting transition
• Retire old systems

Format

✓PDBx Working Group initial implementations
✓Ribosome example: new style mmCIF files
✓Community outreach
✓Start accepting new style format depositions
•Release server to produce best-effort PDB-format files
•Stop supplying PDB files in ftp archive (future)



wwPDB Website(s)

✓ Announce upcoming changes to the world
• Unveil new pdb.org website and ftp site
• Unveil expanded wwpdb.org

Archive

✓Archive updated to support “combined” large structures
•Reformat PDBx/mmCIF data files in archive to conform to new style guidelines
•Future remediations to address carbohydrates and PTMs

wwPDB Task Forces

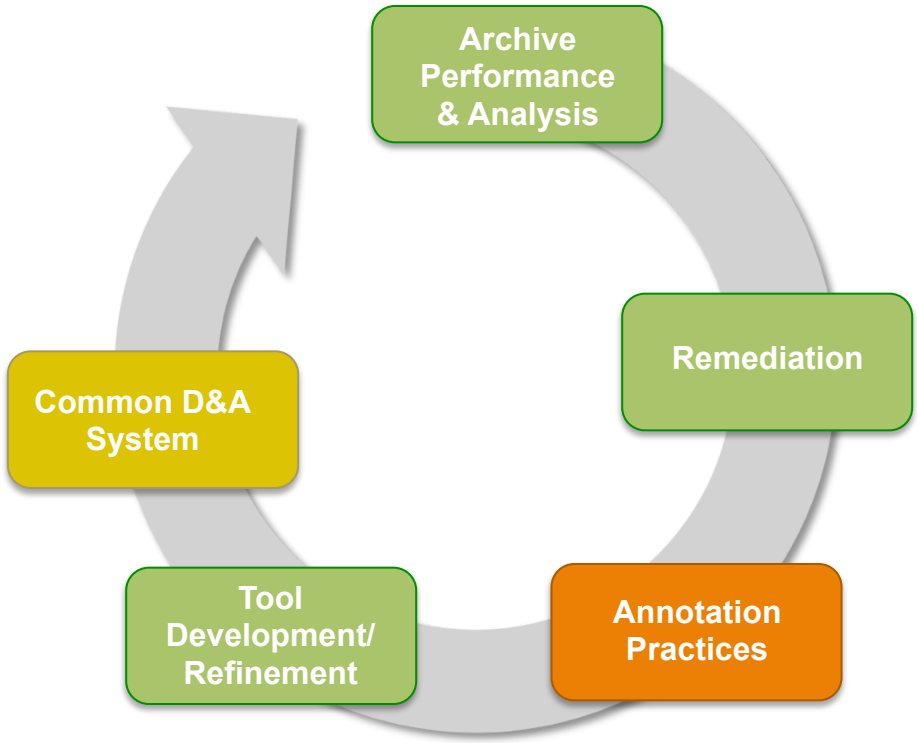
Method-specific (Validation) Task Forces have been convened to collect recommendations and develop consensus on additional validation that should be performed, and to identify software applications to perform validation tasks.

Group	Meeting/ Workshop	Chair(s)/Membership	Outcome
X-ray Validation Task Force	2008	Randy Read (Univ of Cambridge) 17 members	(2011) *Structure* 19: 1395-1412
NMR Validation Task Force	2009, 2011, 2013 (x2)	Gaetano Montelione (Rutgers) Michael Nilges (Institut Pasteur) 10 members	(2013) *Structure*, 21: 1563-1570
3DEM Validation Task Force	2010	Richard Henderson (MRC-LMB) Andrej Sali (UCSF) 21 members	(2012) *Structure* 20: 205-214
Small-Angle Scattering Task Force	2012	Jill Trewhella (Univ of Sydney) 6 members	(2013) *Structure* 21: 875-881

Remediation

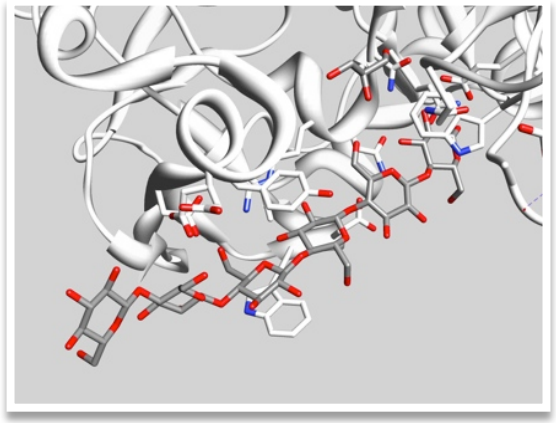
- Informs all processes
- Improves consistency in entry and archive annotation
- Enhances chemistry representation

Better query capability

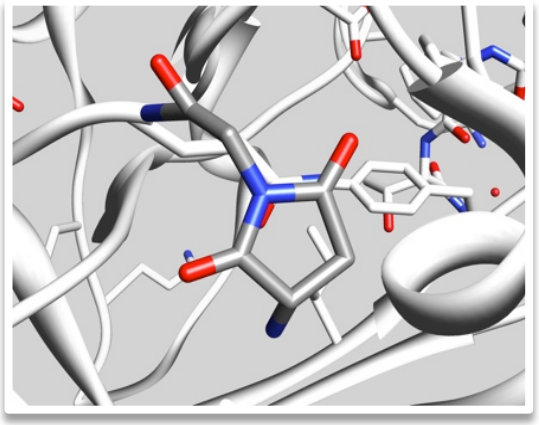


Future Plans for Remediation

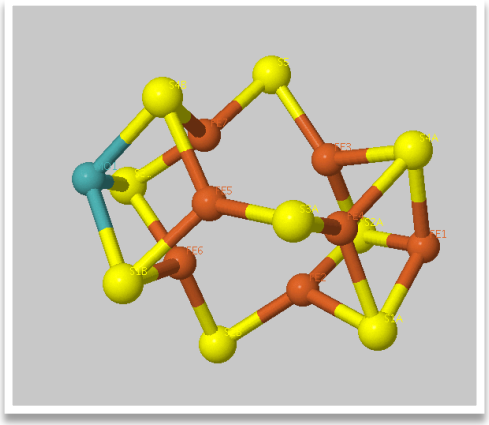
- Carbohydrates
  - Data analysis completed
- Protein modifications
  - Data analysis on-going
- Metal complexes



Carbohydrates



Protein modifications



Metal-containing ligands

wwPDB Organization

- New wwPDB charter (July 1, 2013)
  - Updated from 2003 text and circumstances
  - Signed by PIs and heads of parent institutes
  - Covers 2013-2023, with review possible in 2018
  - Technical details in appendix
- New appointments and terms of reference for wwPDB AC

Funding

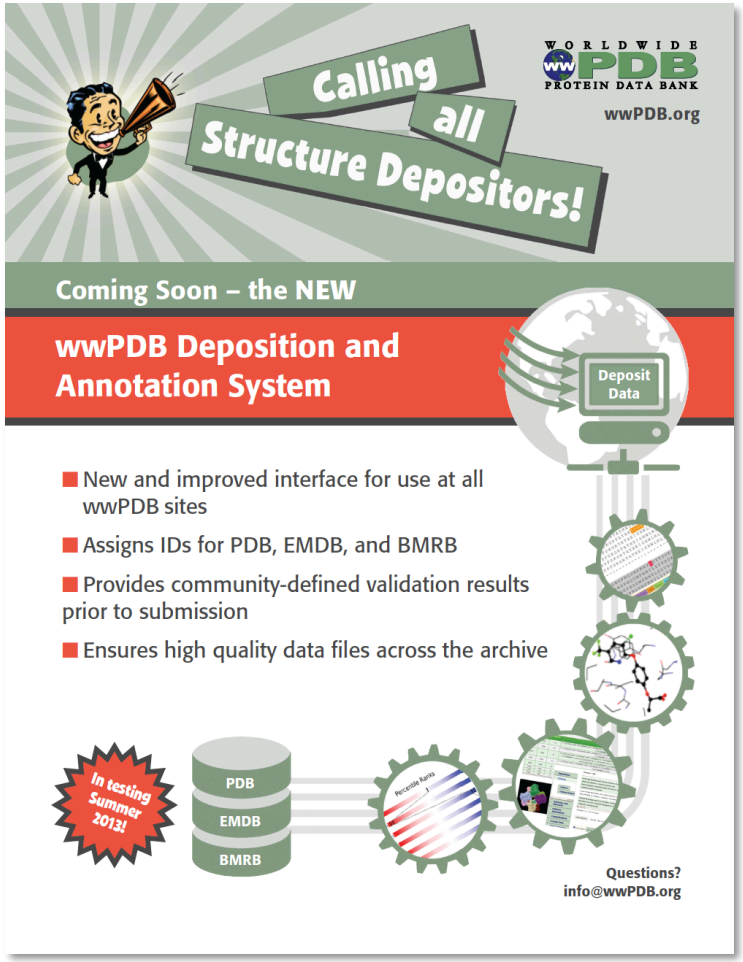
- RCSB PDB non-competitive renewal funded by NSF (2014-2018)
- PDBe main funding from EMBL and Wellcome Trust
  - WT: invited to apply for new competitive grant (pre-application due December 2013)
  - EMBL: core of ~15 posts
- PDBj competitive renewal funded by JST (Japan Science & Technology Agency) for April 2011 - March 2014
  - Next 3-5 year grant review period begins October 2013
- BMRB successfully reviewed by NIH

Worldwide Protein Data Bank Foundation

- New Chairman of the Board
  - Dr. Anthony Nicholls, OpenEye Scientific Software, Inc.
- 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

Activities for the Coming Year

- wwPDB D&A system in production
- Phasing out of PDB format
- Archive remediation
- Continued Task Force activity
- International Year of Crystallography



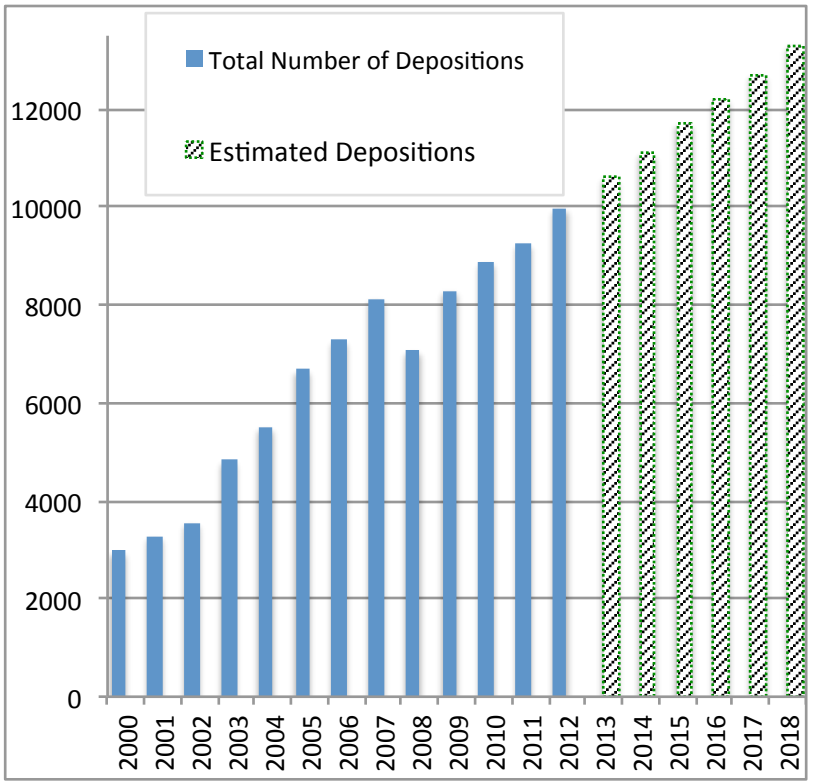
Common Deposition & Annotation (D&A) Tool: Status Report

Martha Quesada

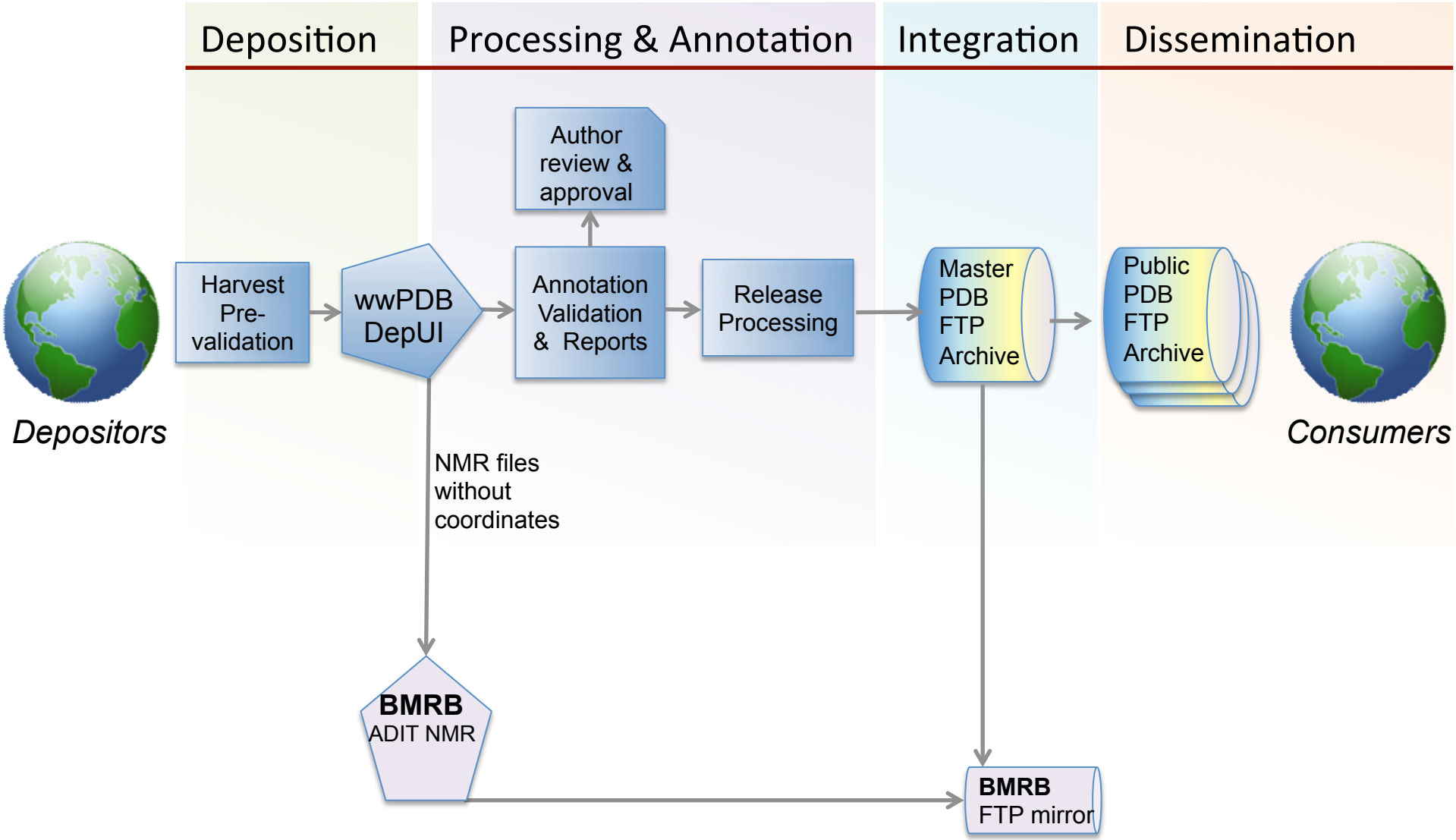
Vision and Delivery

Standardization, Quality and Efficiency

- ✓ Larger and more complex biological molecules
- ✓ Expanded annotation
- ✓ Increased throughput: Automation and validation of routine submissions



New wwPDB Data Deposition and Annotation Pipeline



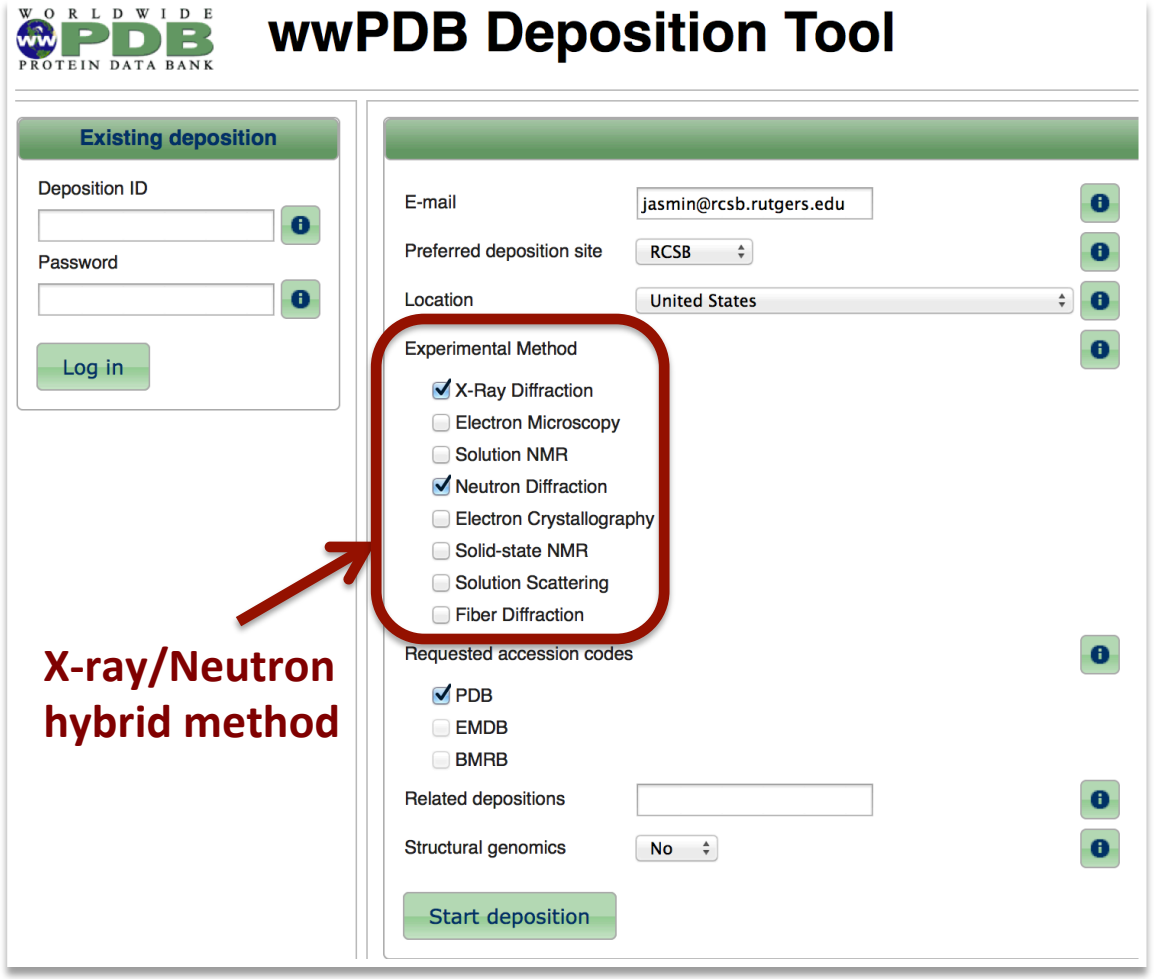
New Deposition Pipeline – In Public Testing

Depositors can:

- Base new depositions on a public PDB entry or on their own unreleased depositions
- Replace files (e.g., a re-refined model) during deposition and post annotation (prior to release) with retention of appropriate data
- Preview final PDB files (PDB, PDBx formats) prior to submission
- Communicate with expert annotators using web-based tools

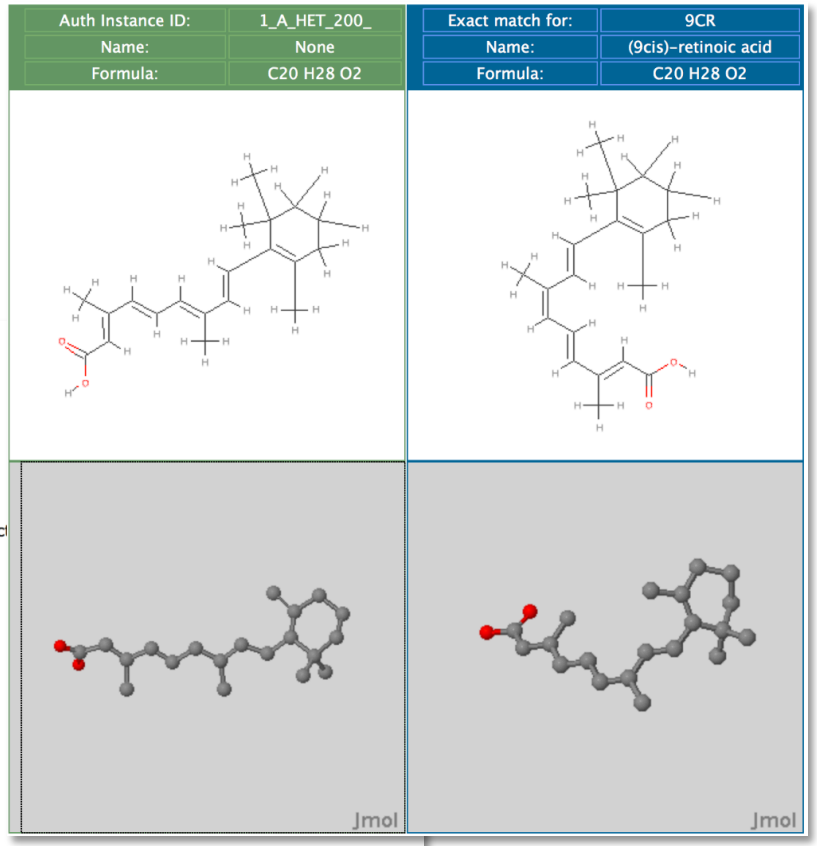
New Deposition Interface

- Single point of entry (i.e., wwpdb.org/deposit)
- Supports multiple methods
- Workload balancing based on resource capacity and geography



Deposition Interface

- Support improved data quality and processing efficiency by facilitating Depositor review of critical data: Ligand & Sequence consistency checks



Sample sequence in one letter code*:

```
MTQQITLIKDKILSDNYFTLHNITYDLTRKDG
EVIRHKREVYDRGNGATILLYNTKKKTVVLIR
QFRVATWVNGNESGQL
IESCAGLLDNDEPEVCIRKEAIEETGYEVGEV
RKLFELYMSPGGVTELIHFFIAEYSDNQRANA
GGGVEDEDIEVLELPF
```

e.g. : HHHH(MSE)AKQRSG or AUCGGAAU

The following is the alignment between the biological sample sequence and sequence from the coordinates

Important : Please address any discrepancy between the sample sequence and coordinate sequence by either providing a correct

Refresh the alignment from sample sequence provided

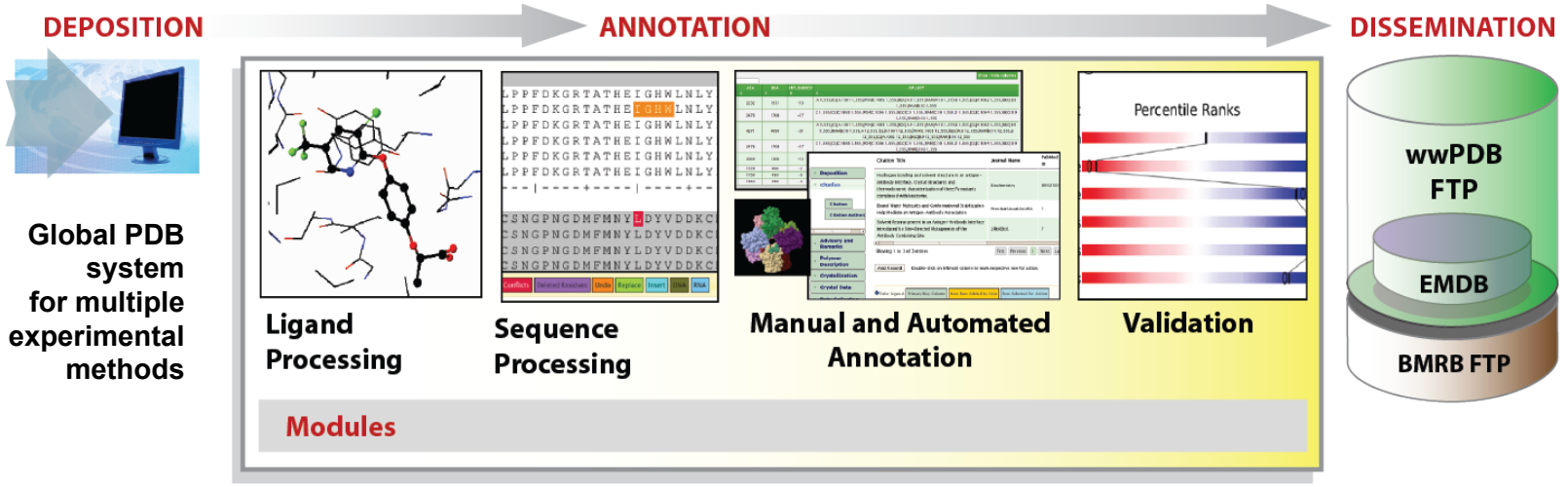
ok : Sample sequence aligned with Coordinates - all chains :

```
Molecule : 1
Chain : A
--------10--------20--------30--------40--------50--------60--------70--------80--------90-------100
MTQQITLIKDKILSDNYFTLHNITYDLTRKDGEVIRHKREVYDRGNGATILLYNTKKKTVVLIRQFRVATWVNGNESGQLIESCAGLLDNDEPEVCIRKE
-||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
.TQQITLIKDKILSDNYFTLHNITYDLTRKDGEVIRHKREVYDRGNGATILLYNTKKKTVVLIRQFRVATWVNGNESGQLIESCAGLLDNDEPEVCIRKE

-------110-------120-------130-------140-------150-------160-------170-------180-------190-
AIEETGYEVGEVRKLFELYMSPGGVTELIHFFIAEYSDNQRANAGGGVEDEDIEVLELPFSQALEMIKTGEIRDGKTVLLLNYLQTSHLMD
|||||-|||||||||||||||||||||||||||||||||||--------||-|||||||||||||||||||||||||||||||||||||||
AIEET.YEVGEVRKLFELYMSPGGVTELIHFFIAEYSDNQR........DEAIEVLELPFSQALEMIKTGEIRDGKTVLLLNYLQTSHLMD

Chain : B
--------10--------20--------30--------40--------50--------60--------70--------80--------90-------100
MTQQITLIKDKILSDNYFTLHNITYDLTRKDGEVIRHKREVYDRGNGATILLYNTKKKTVVLIRQFRVATWVNGNESGQLIESCAGLLDNDEPEVCIRKE
--|||||||||||||||||||||||||||||-|||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
..QQITLIKDKILSDNYFTLHNITYDLTRKDG.VIRHKREVYDRGNGATILLYNTKKKTVVLIRQFRVATWVNGNESGQLIESCAGLLDNDEPEVCIRKE

-------110-------120-------130-------140-------150-------160-------170-------180-------190-
AIEETGYEVGEVRKLFELYMSPGGVTELIHFFIAEYSDNQRANAGGGVEDEDIEVLELPFSQALEMIKTGEIRDGKTVLLLNYLQTSHLMD
||||||||||||||||||||||||||||||||||||||||||||||||||||||-||||||||||||||||||||||||||||||||||||||
AIEETGYEVGEVRKLFELYMSPGGVTELIHFFIAEYSDNQRANAGGGVEDEAIEVLELPFSQALEMIKTGEIRDGKTVLLLNYLQTSHLMD
```

There is a mismatch at residue 152 between the sample and coordinate sequence - this is not allowed

You must either correct the sample sequence in the text box above and click the button to refresh the sequence alignment or correct the coordinates by uploading a new file.

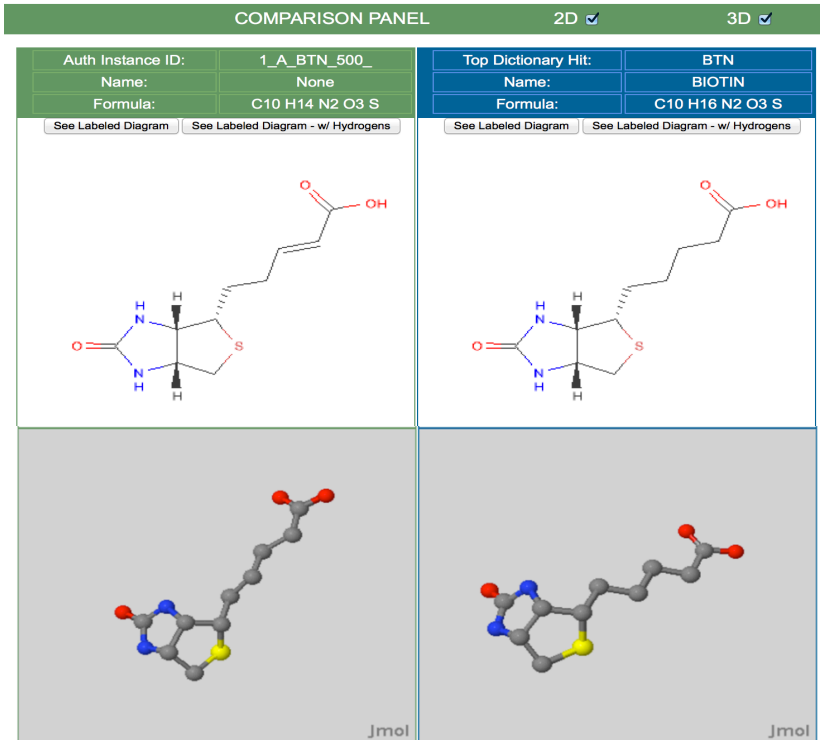
Annotation System



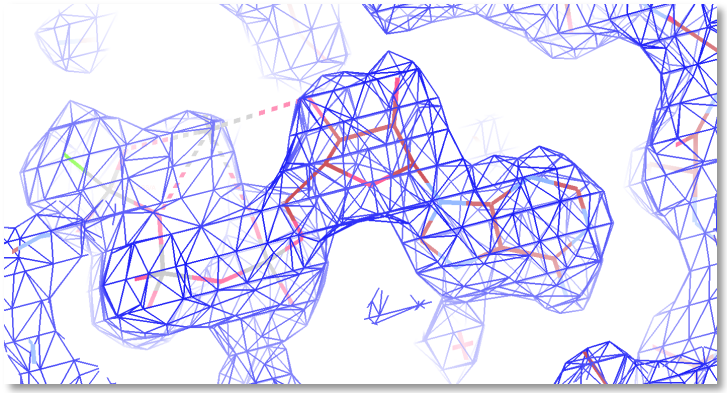
- PDBx/mmCIF is the master file format
- Better quality assurance of ligand chemistry and polymer sequences
- Validation suites based on recommendations from expert task forces; X-ray validation pipeline is available as a stand-alone server
- System will support all accepted experimental methods

Ligand Processing

- Search Chemical Component Dictionary with automated ligand ID assignment
- 2D and 3D views of ligand for review
- View author-provided chemical information
- Electron density map can be inspected by annotators

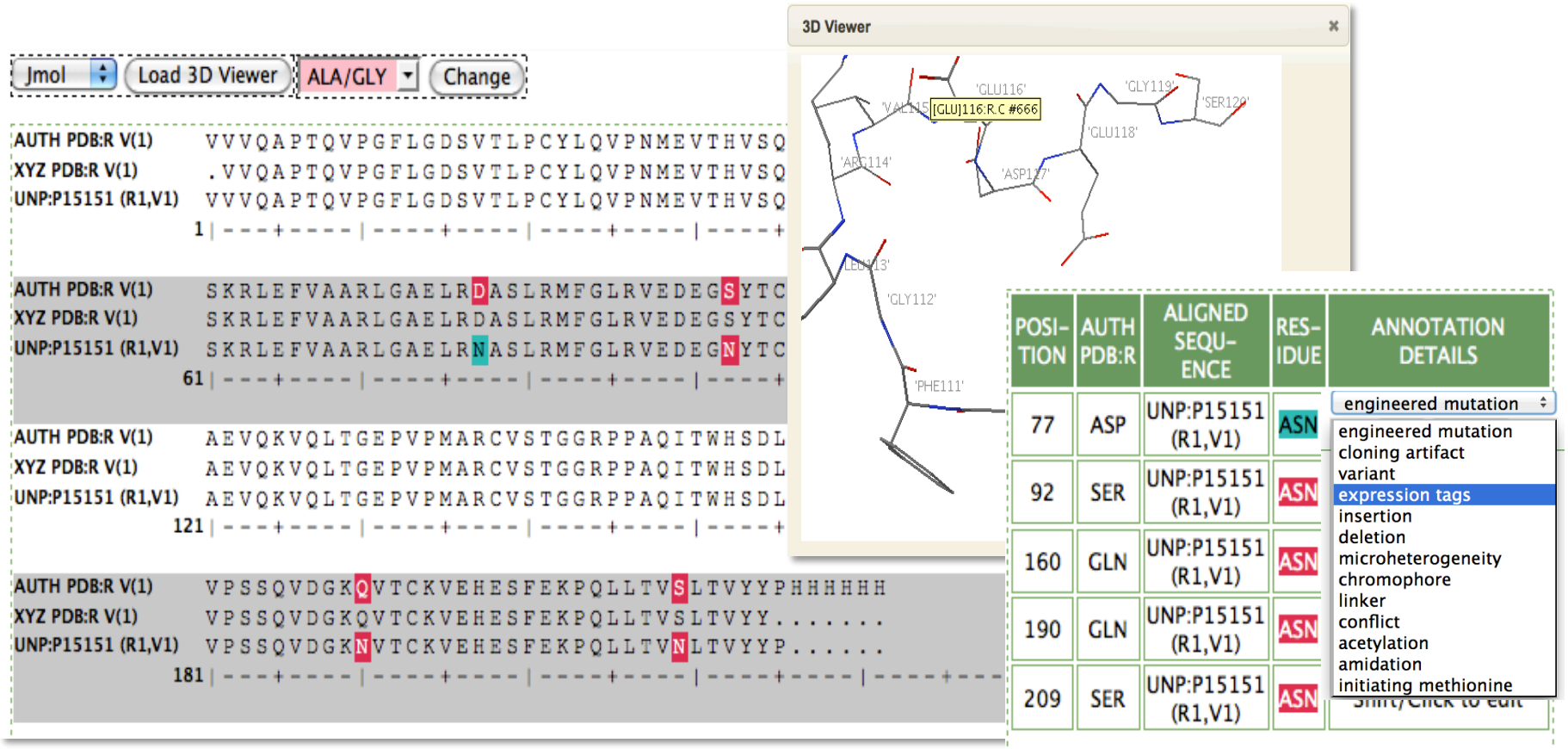


Author instance from coordinates (left) and the closest match in the dictionary (right)



Sequence Module: Consistency Check

- Biological sequence checked against sample and coordinate sequences and cross-referenced to UniProt/GenBank
- 3D structure view
- Sequence discrepancy annotation



Implementation, Testing and Evolution

- Testing at all wwPDB sites has included
  - Unit and integration testing for each functional component (sequence, ligand, etc.)
  - Pipeline testing: internal acceptance testing
    - System evolution: regression testing of all updates
  - Infrastructure testing: servers and data transfer
- Weekly Project Team meetings by Video Conferencing
  - System demonstrations and training
  - Testing feedback and design evolution
- Alpha testing: limited external testing (July, Aug)
- Beta testing: external acceptance testing (Sept-Dec)

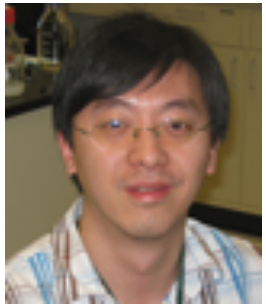
First Depositors to Test the System

Andrea Mattevi
Pavia, Italy
First unassisted depositor

- "I found it as easy as with the standard software (used today for other two entries)"
- *Ligand-lite* module needs to be explained a bit better
- "Very good job"

John Rose
University of Georgia

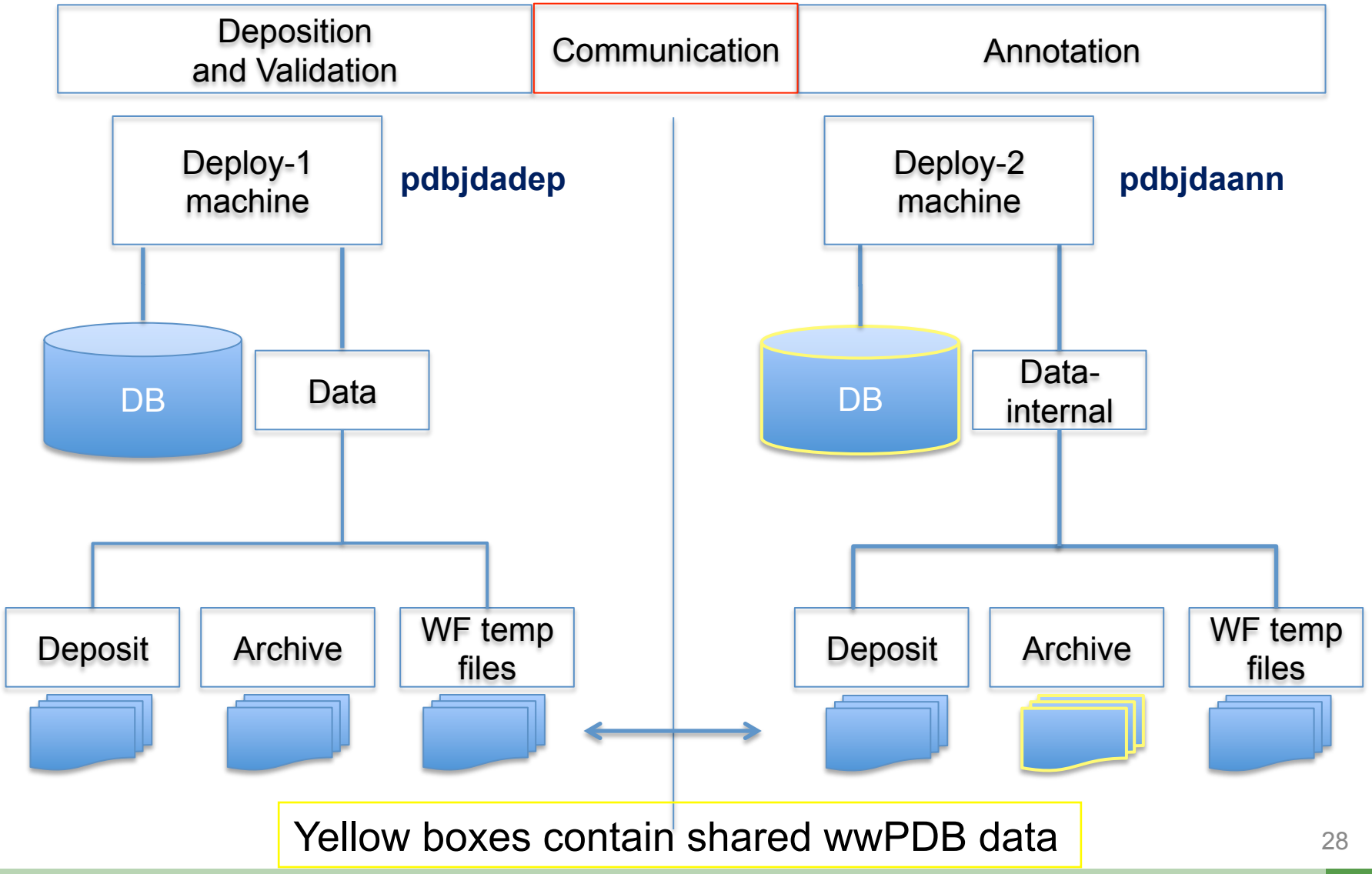
- Navigation pages are straightforward
- Clarification required for sequence mutation handling
- Particularly liked the ability to upload a refinement dictionary file as part of the ligand module



Ming Yan
Johns Hopkins
Cynthia Wolberger, PI

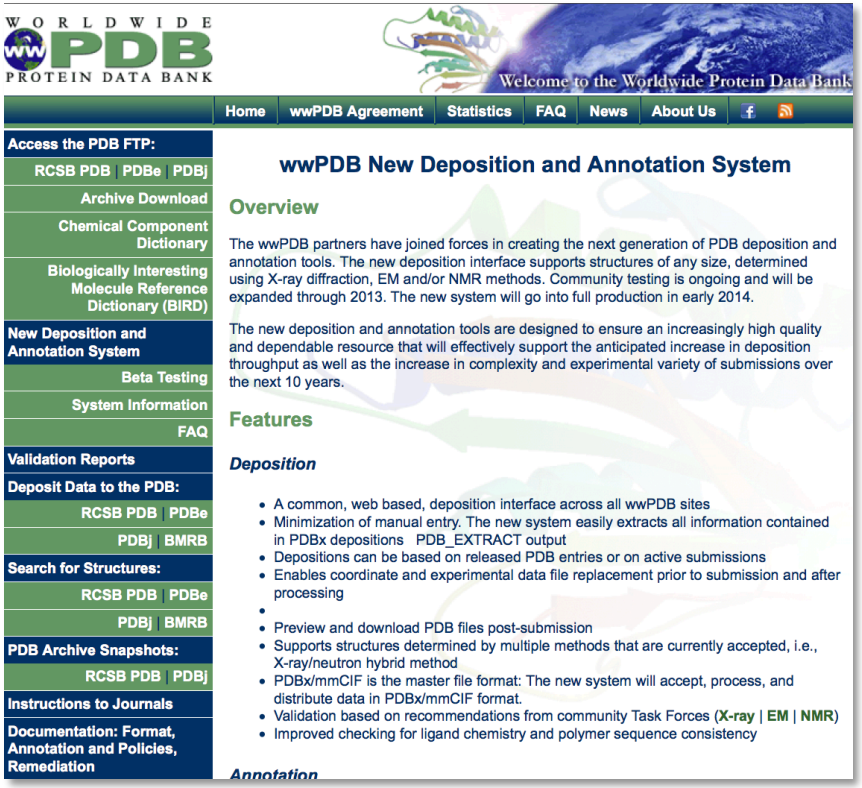
- Look and feel is similar to the outgoing system
- Likes the fact that validation will be available within a deposition session
- Comparison view in ligand module should be visible by default

Deployment of D&A Infrastructure at wwPDB Member Site (PDBj)



2013 Project Milestones

- ✓ May: Internal pipeline testing begins at all sites
- ✓ June 22: Deposition system introduced at CCP4/APS Summer School at Argonne
- ✓ July 1: Testing Begins for Developers
- ✓ July 20: D&A pipeline demo and tester solicitation at ACA meeting
- ✓ September: Initial community deposition testing begins
- ■ January 2014: Common D&A system in production and continued testing at all sites



Format issues, Validation & Experimental methods

Gerard Kleywegt

Format issues

Update on “New PDB Format”

- PDBx/mmCIF Deposition Working Group:
  - Paul Adams, Chair
  - Result of seminal workshop in 2010
  - Goal: support deposition of X-ray structures in PDBx format
  - Participants: developers of major X-ray software packages and wwPDB staff

Pragmatic Principles

- Preserve backward compatibility where possible
- Changes that do not fit within the current PDB format will be implemented only if needed (e.g., to represent a large molecule)
  - Atom serial numbers
  - Chain identifiers
  - Residue names and numbers
- Continue to assign residue-level 3-letter codes even if more descriptive identifiers are adopted (e.g., for monosaccharides)

```
ATOM      1  N   GLN A  39      24.690 -27.754  24.275  1.00 60.76           N
ATOM      2  CA  GLN A  39      23.581 -26.768  24.416  1.00 60.98           C
ATOM      3  C   GLN A  39      23.990 -25.379  23.905  1.00 59.98           C
ATOM      4  O   GLN A  39      25.070 -25.209  23.330  1.00 60.25           O
ATOM      5  CB  GLN A  39      23.136 -26.685  25.878  1.00 60.69           C
ATOM      6  N   VAL A  40      23.115 -24.395  24.122  1.00 59.58           N
ATOM      7  CA  VAL A  40      23.342 -23.010  23.690  1.00 57.26           C
ATOM      8  C   VAL A  40      24.000 -22.152  24.778  1.00 56.00           C
```

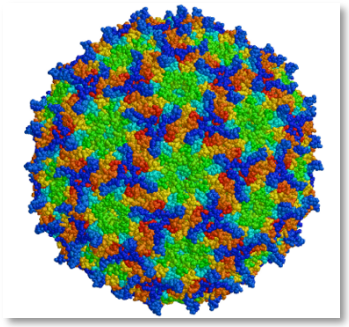
PDB

PDBx/mmCIF

```
loop_
_atom_site.group_PDB
_atom_site.id
_atom_site.auth_atom_id
_atom_site.type_symbol
_atom_site.auth_comp_id
_atom_site.auth_asym_id
_atom_site.auth_seq_id
_atom_site.Cartn_x
_atom_site.Cartn_y
_atom_site.Cartn_z
_atom_site.pdbx_PDB_model_num
_atom_site.occupancy
_atom_site.pdbx_auth_alt_id
_atom_site.B_iso_or_equiv
ATOM       1  N    N  GLN  A   39   24.690  -27.754   24.275  1  1.00  .  60.76
ATOM       2  CA   C  GLN  A   39   23.581  -26.768   24.416  1  1.00  .  60.98
ATOM       3  C    C  GLN  A   39   23.990  -25.379   23.905  1  1.00  .  59.98
ATOM       4  O    O  GLN  A   39   25.070  -25.209   23.330  1  1.00  .  60.25
ATOM       5  CB   C  GLN  A   39   23.136  -26.685   25.878  1  1.00  .  60.69
ATOM       6  N    N  VAL  A   40   23.115  -24.395   24.122  1  1.00  .  59.58
ATOM       7  CA   C  VAL  A   40   23.342  -23.010   23.690  1  1.00  .  57.26
ATOM       8  C    C  VAL  A   40   24.000  -22.152   24.778  1  1.00  .  56.00
```

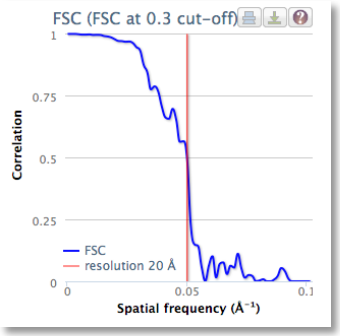
Status

- Major X-ray refinement packages can now produce PDBx/mmCIF files for deposition (Refmac, Phenix)
- Other X-ray software increasingly adding support for PDBx/mmCIF (CCP4, Phenix, MolProbity)
- Several large structures already released as “un-split” PDBx/mmCIF files
- Support for software developers
  - CCP4/APS workshop
  - Test files available from wwPDB website
  - Discussions on CCP4 bulletin board
  - COMCIFS Satellite Symposium at ECM (University of Warwick, Aug 2013)
  - Workshop on Theoretical Model Validation and PDBx/mmCIF (Rutgers, Oct 2013)
  - Workshop for UK & European developers (EBI, Nov 2013)



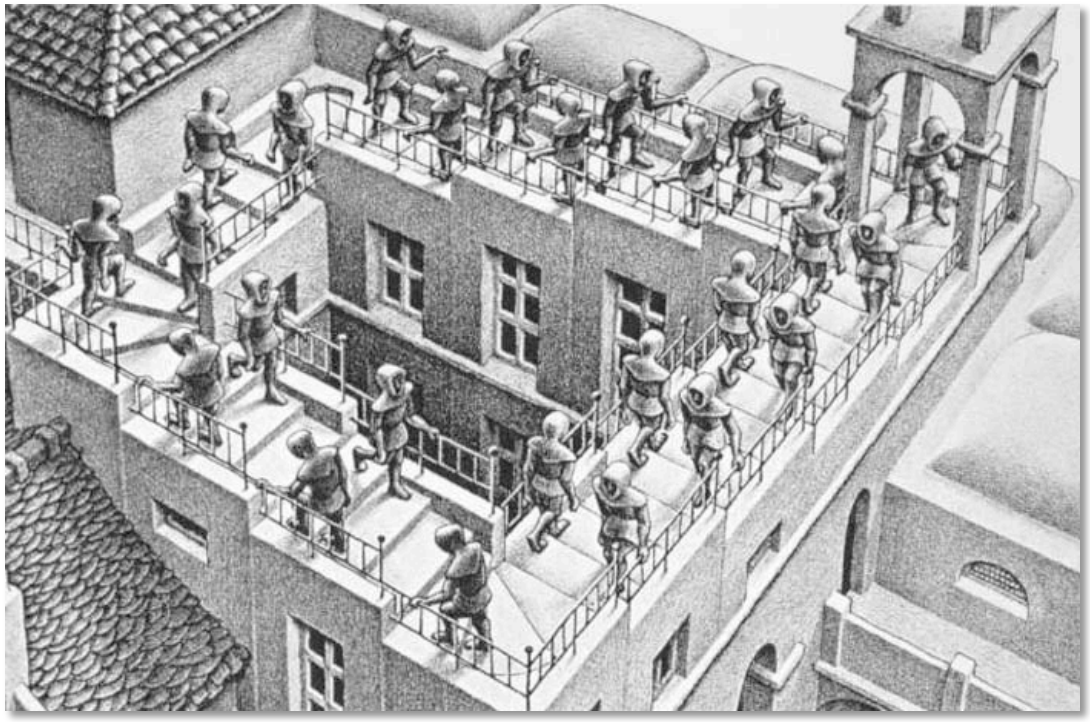
Other Format Issues

- EM-related
  - EMX
    - Information exchange between processing programs
    - Coordinated by Madrid
  - Validation-related
    - FSC curves, tilt-pair results, SAXS curves, …
    - Coordinated by EMDataBank
  - Segmentations and annotations
    - Meta-data to be defined; support existing formats for actual segmentations
    - Coordinated by EMDataBank



- NMR restraint data
  - Workshop at EBI in November with key developers to discuss a unified format for NMR restraints

Validation



History

- 1970s & 1980s – happy if people deposited models (rarely data), which were taken as holy gospel
- Late 1980s – several cases of completely wrong models
- Early 1990s – first efforts to validate models
- Late 1990s – common mandatory model deposition
- Mid 2000s – still reluctance to deposit experimental data
- Then – Fraud! Panic! End of the world!
- February 2008 – data deposition mandatory (chemical shifts in December 2010)
- April 2008 – wwPDB X-ray Validation Task Force (VTF) established

History

- September 2009 – wwPDB NMR VTF
- September 2010 – EMDataBank EM VTF
- July 2012 – wwPDB SAS Task Force
- Today: wwPDB validation pipelines being developed for **X-ray**, **NMR** and **EM** based on VTF guidelines

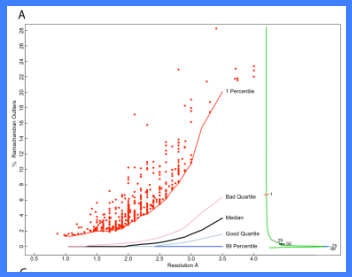
- ***X-ray pipeline – reports sent out to depositors since 1 August; stand-alone server soon to be released***

- Early 2014 – new Deposition & Annotation system
- Late 2014 – wwPDB Task Force for Hybrid Methods (workshop at EBI)

wwPDB X-ray Validation Pipeline

External reference files (*e.g.*, Engh & Huber)

Distributions

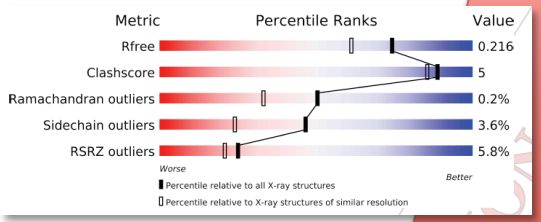


Validation XML file

Validation pipeline 1.0

MolProbity | Xtriage | EDS | Mogul | Percentiles | PDF maker

Deposited data (coordinates & reflections)



PDF report for depositor & referees - Statistics and plots for the entry, per chain, per residue, and list of unusual features

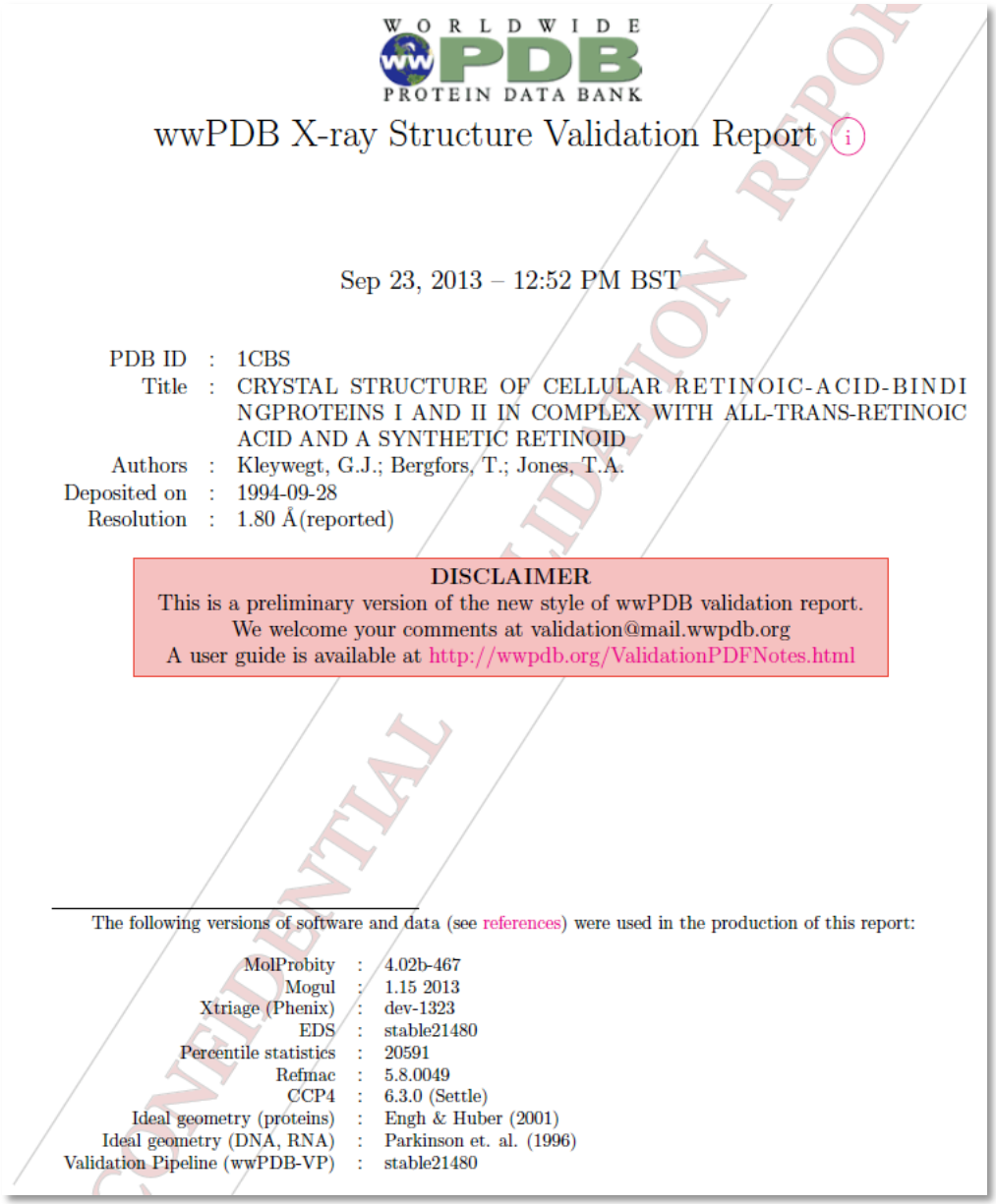
Gore *et al.*, *Acta Cryst.* **D68**, 478 (2012)

wwPDB X-ray Validation Pipeline

- Version 1.0 in production use since August
  - http://wwpdb.org/validation.html
  - Stand-alone server about to be released
  - Collecting feedback to identify bugs and inform possible changes
    - validation@mail.wwpdb.org
- January 2014 – validation data for all X-ray structures will be made publicly available through the wwPDB ftp sites
- After version 1.0:
  - WhatCheck (e.g., DACA, unusual backbone)?
  - pdbcare (carbohydrates)?
  - LabelIt (spacegroup errors)?
  - DDQ (e.g., uninterpreted density)?
  - Better real-space fit criterion for non-standard entities?
- In a few years' time – reconvene X-ray VTF for evaluation & update

Validation Reports

- Front cover
  - Deposition info
  - Software info



WORLDWIDE PDB
PROTEIN DATA BANK

wwPDB X-ray Structure Validation Report (i)

Sep 23, 2013 – 12:52 PM BST

PDB ID	:	1CBS
Title	:	CRYSTAL STRUCTURE OF CELLULAR RETINOIC-ACID-BINDI NGPROTEINS I AND II IN COMPLEX WITH ALL-TRANS-RETINOIC ACID AND A SYNTHETIC RETINOID
Authors	:	Kleywegt, G.J.; Bergfors, T.; Jones, T.A.
Deposited on	:	1994-09-28
Resolution	:	1.80 Å(reported)

DISCLAIMER
This is a preliminary version of the new style of wwPDB validation report.
We welcome your comments at validation@mail.wwpdb.org
A user guide is available at http://wwpdb.org/ValidationPDFNotes.html

The following versions of software and data (see references) were used in the production of this report:

MolProbity	:	4.02b-467
Mogul	:	1.15 2013
Xtriage (Phenix)	:	dev-1323
EDS	:	stable21480
Percentile statistics	:	20591
Refmac	:	5.8.0049
CCP4	:	6.3.0 (Settle)
Ideal geometry (proteins)	:	Engh & Huber (2001)
Ideal geometry (DNA, RNA)	:	Parkinson et. al. (1996)
Validation Pipeline (wwPDB-VP)	:	stable21480

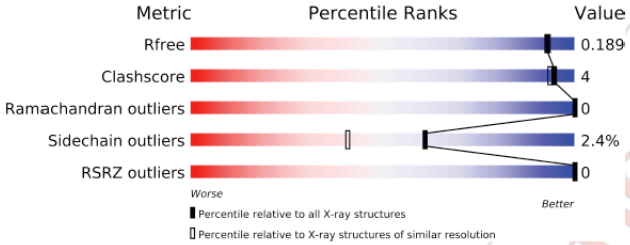
Validation Reports

- Summary
  - Quality vs. all PDB X-ray
  - Quality vs. entries at similar resolution
  - Overview of residue-based quality for every polymer
  - Table of ligands that may need attention

1 Overall quality at a glance

The reported resolution of this entry is 1.80 Å.

Percentile scores (ranging between 0-100) for global validation metrics of the entry are shown in the following graphic. The table shows the number of entries on which the scores are based.



Metric	Whole archive (#Entries)	Similar resolution (#Entries, resolution range(Å))
R_{free}	65580	5522 (1.84-1.76)
Clashscore	76988	5040 (1.82-1.78)
Ramachandran outliers	75395	6528 (1.84-1.76)
Sidechain outliers	75377	6529 (1.84-1.76)
RSRZ outliers	65576	5522 (1.84-1.76)

The table below summarises the geometric issues observed across the polymeric chains and their fit to the electron density. The red, orange, yellow and green segments on the lower bar indicate the fraction of residues that contain outliers for >=3, 2, 1 and 0 types of geometric quality criteria. The upper red bar (where present) indicates the fraction of residues that have poor fit to the electron density.

Mol	Chain	Length	Quality of chain
1	A	137	

Mol	Chain	Length	Quality of chain
1	A	371	
1	C	371	

The following table lists non-polymeric compounds that contain outliers for geometric or electron-density-fit criteria:

Mol	Type	Chain	Res	Geometry	Electron density
2	NAG	A	401	-	X
2	NAG	C	401	-	X

Validation Reports

- Entry contents
  - Inventory

2 Entry composition

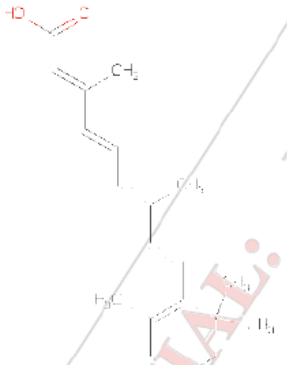
There are 3 unique types of molecules in this entry. The entry contains 1213 atoms, of which 0 are hydrogen and 0 are deuterium.

In the tables below, the ZeroOcc column contains the number of atoms modelled with zero occupancy, the AltConf column contains the number of residues with at least one atom in alternate conformation and the Trace column contains the number of residues modelled with at most 2 atoms.

- Molecule 1 is a protein called CELLULAR RETINOIC ACID BINDING PROTEIN TYPE II.

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
			Total	C	N	O	S			
1	A	137	1091	687	184	214	6	0	0	0

- Molecule 2 is RETINOIC ACID (three-letter code: REA) (formula: C20 H28 O2).



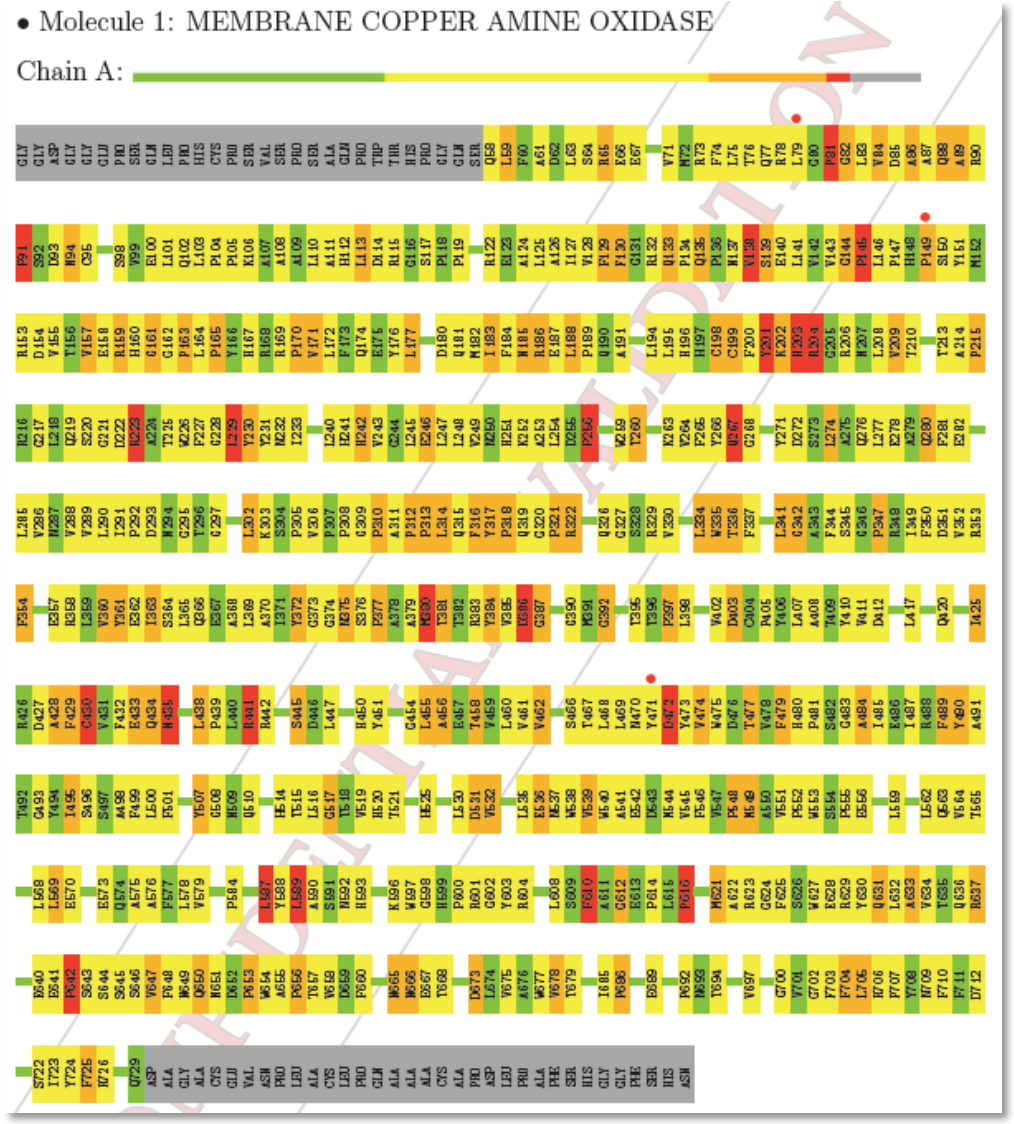
Mol	Chain	Residues	Atoms			ZeroOcc	AltConf
			Total	C	O		
2	A	1	22	20	2	0	0

- Molecule 3 is water.

Mol	Chain	Residues	Atoms		ZeroOcc	AltConf
			Total	O		
3	A	100	100	100	0	0

Validation Reports

- Residue quality
  - One plot per polymer
  - Coloured by number of ***types*** of geometric outliers
  - Grey if not modelled
  - Red dots: poor density (RSR-Z > 2, as in EDS)



Validation Reports

- "Table 1"
  - Xtriage

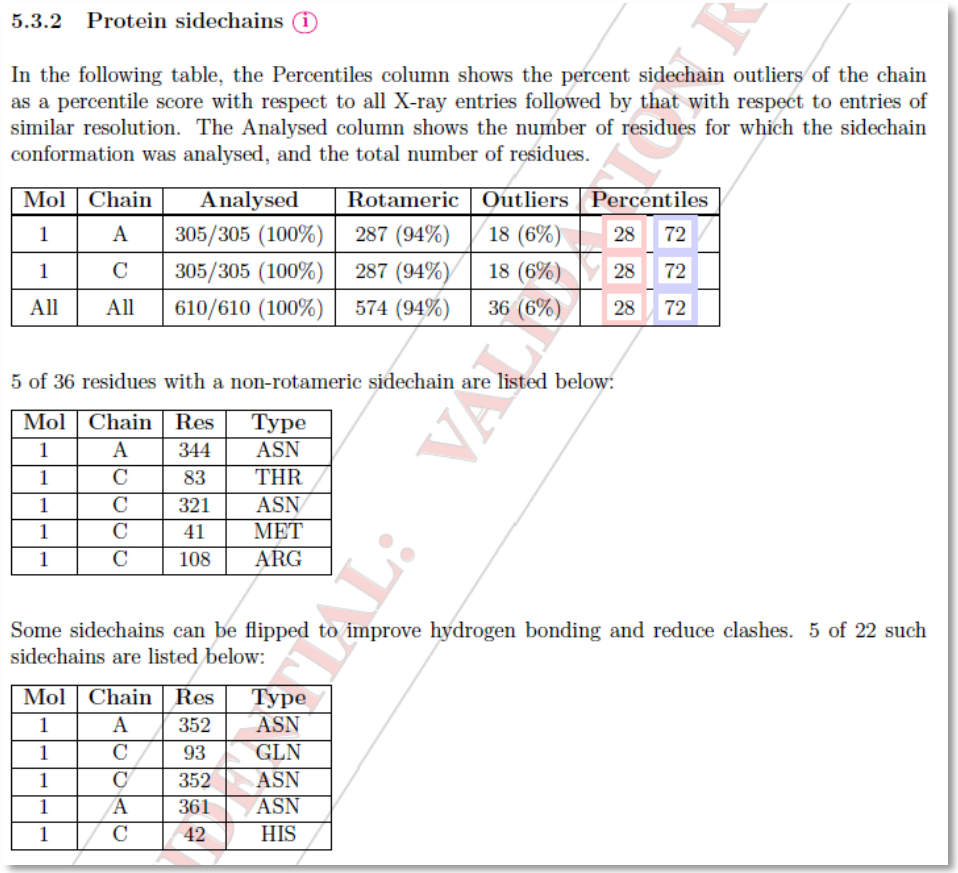
4 Data and refinement statistics (i)

Property	Value	Source
Space group	P 21 21 21	Depositor
Cell constants a, b, c, α, β, γ	45.65Å 47.56Å 77.61Å 90.00° 90.00° 90.00°	Depositor
Data completeness (%)	90.3 90.5	Depositor EDS
R_{merge}	(Not available)	Depositor
R_{sym}	(Not available)	Depositor
$< I/\sigma(I) >$ [1]	3.77 (at 1.79Å)	Xtriage
Resolution (Å)	8.00 – 1.80 14.93 – 1.80	Depositor EDS
Refinement program	X-PLOR	Depositor
R, R_{free}	0.200 , 0.237 0.184 , 0.189	Depositor DCC
Wilson B-factor (Å^2)	14.8	Xtriage
Anisotropy	0.434	Xtriage
Bulk solvent k_{sol}(e/Å^3), B_{sol}(Å^2)	0.41 , 58.87	EDS
Estimated twinning fraction	0.027 for k,h,-l	Xtriage
L-test for twinning	$< \|L\| > = 0.51$, $< L^2 > = 0.36$	Xtriage
Outliers	0 of 14678 reflections	Xtriage
F_o,F_c correlation	0.95	EDS
Total number of atoms	1213	wwPDB-VP
Average B, all atoms (Å^2)	16.0	wwPDB-VP

Xtriage's analysis on translational NCS is as follows: *The largest off-origin peak in the Patterson function is 9.26% of the height of the origin peak. No significant pseudotranslation is detected.*

Validation Reports

- Model quality
  - Bond lengths and angles
  - Torsion angles (Ramachandran, rotamers)
  - Clashes
  - Separately for standard residues, non-standard residues, ligands, carbohydrates
  - Generally: information about distribution, outlier stats, percentile scores, list of up to 5 (worst) outliers



5.3.2 Protein sidechains (i)

In the following table, the Percentiles column shows the percent sidechain outliers of the chain as a percentile score with respect to all X-ray entries followed by that with respect to entries of similar resolution. The Analysed column shows the number of residues for which the sidechain conformation was analysed, and the total number of residues.

Mol	Chain	Analysed	Rotameric	Outliers	Percentiles	
1	A	305/305 (100%)	287 (94%)	18 (6%)	28	72
1	C	305/305 (100%)	287 (94%)	18 (6%)	28	72
All	All	610/610 (100%)	574 (94%)	36 (6%)	28	72

5 of 36 residues with a non-rotameric sidechain are listed below:

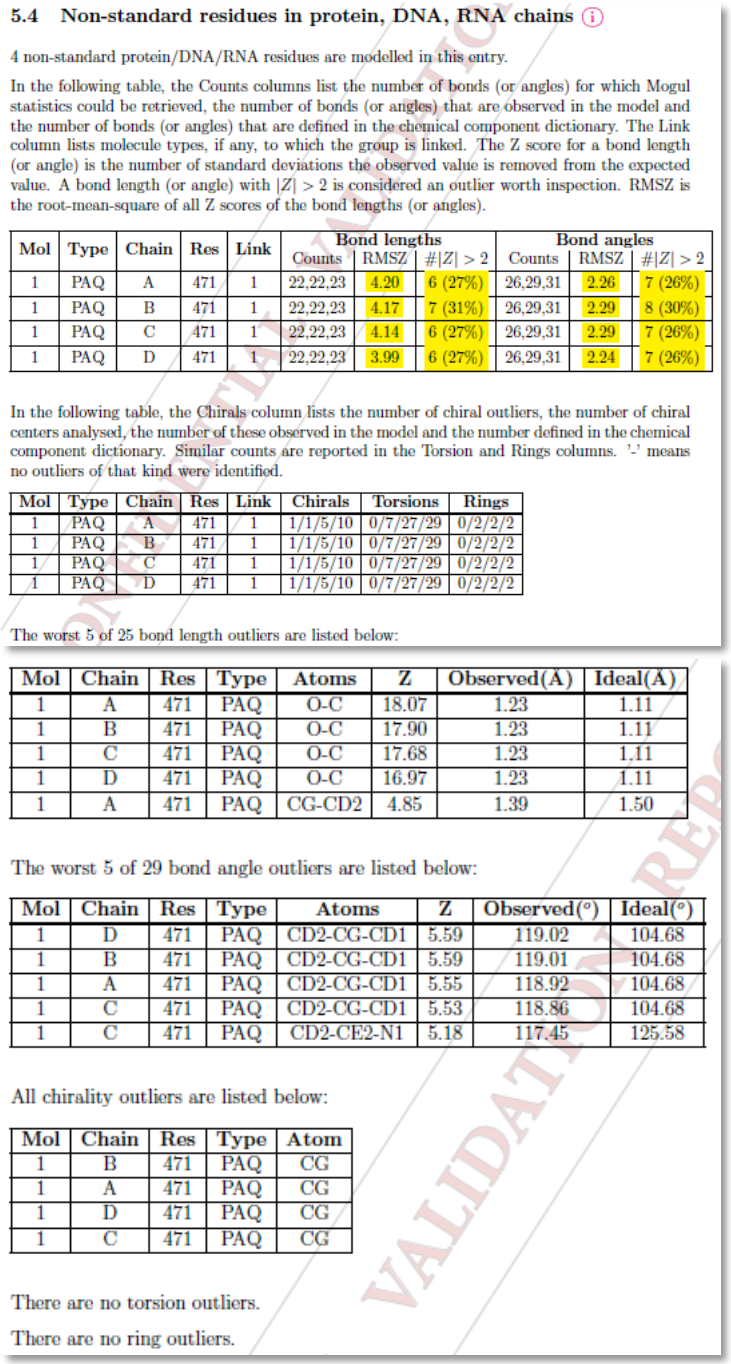
Mol	Chain	Res	Type
1	A	344	ASN
1	C	83	THR
1	C	321	ASN
1	C	41	MET
1	C	108	ARG

Some sidechains can be flipped to improve hydrogen bonding and reduce clashes. 5 of 22 such sidechains are listed below:

Mol	Chain	Res	Type
1	A	352	ASN
1	C	93	GLN
1	C	352	ASN
1	A	361	ASN
1	C	42	HIS

Validation Reports

- Geometry validation of ligands and non-standard entities
  - Mogul (CCDC)
- wwPDB will get CSD coordinates for new and existing compounds



5.4 Non-standard residues in protein, DNA, RNA chains

4 non-standard protein/DNA/RNA residues are modelled in this entry.

In the following table, the Counts columns list the number of bonds (or angles) for which Mogul statistics could be retrieved, the number of bonds (or angles) that are observed in the model and the number of bonds (or angles) that are defined in the chemical component dictionary. The Link column lists molecule types, if any, to which the group is linked. The Z score for a bond length (or angle) is the number of standard deviations the observed value is removed from the expected value. A bond length (or angle) with $|Z| > 2$ is considered an outlier worth inspection. RMSZ is the root-mean-square of all Z scores of the bond lengths (or angles).

Mol	Type	Chain	Res	Link	Bond lengths			Bond angles		
					Counts	RMSZ	$\#\|Z\| > 2$	Counts	RMSZ	$\#\|Z\| > 2$
1	PAQ	A	471	1	22,22,23	4.20	6 (27%)	26,29,31	2.26	7 (26%)
1	PAQ	B	471	1	22,22,23	4.17	7 (31%)	26,29,31	2.29	8 (30%)
1	PAQ	C	471	1	22,22,23	4.14	6 (27%)	26,29,31	2.29	7 (26%)
1	PAQ	D	471	1	22,22,23	3.99	6 (27%)	26,29,31	2.24	7 (26%)

In the following table, the Chirals column lists the number of chiral outliers, the number of chiral centers analysed, the number of these observed in the model and the number defined in the chemical component dictionary. Similar counts are reported in the Torsion and Rings columns. '-' means no outliers of that kind were identified.

Mol	Type	Chain	Res	Link	Chirals	Torsions	Rings
1	PAQ	A	471	1	1/1/5/10	0/7/27/29	0/2/2/2
1	PAQ	B	471	1	1/1/5/10	0/7/27/29	0/2/2/2
1	PAQ	C	471	1	1/1/5/10	0/7/27/29	0/2/2/2
1	PAQ	D	471	1	1/1/5/10	0/7/27/29	0/2/2/2

The worst 5 of 25 bond length outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(Å)	Ideal(Å)
1	A	471	PAQ	O-C	18.07	1.23	1.11
1	B	471	PAQ	O-C	17.90	1.23	1.11
1	C	471	PAQ	O-C	17.68	1.23	1.11
1	D	471	PAQ	O-C	16.97	1.23	1.11
1	A	471	PAQ	CG-CD2	4.85	1.39	1.50

The worst 5 of 29 bond angle outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(°)	Ideal(°)
1	D	471	PAQ	CD2-CG-CD1	5.59	119.02	104.68
1	B	471	PAQ	CD2-CG-CD1	5.59	119.01	104.68
1	A	471	PAQ	CD2-CG-CD1	5.55	118.92	104.68
1	C	471	PAQ	CD2-CG-CD1	5.53	118.86	104.68
1	C	471	PAQ	CD2-CE2-N1	5.18	117.45	125.58

All chirality outliers are listed below:

Mol	Chain	Res	Type	Atom
1	B	471	PAQ	CG
1	A	471	PAQ	CG
1	D	471	PAQ	CG
1	C	471	PAQ	CG

There are no torsion outliers.

There are no ring outliers.

Validation Reports

- Model/data fit proteins, DNA, RNA
  - RSR and RSR-Z (EDS)

6.1 Protein, DNA and RNA chains

In the following table, the column labelled '#RSRZ> 2' contains the number (and percentage) of RSRZ outliers, followed by percent RSRZ outliers for the chain as percentile scores relative to all X-ray entries and entries of similar resolution. The OWAB column contains the minimum, median, 95^{th} percentile and maximum values of the occupancy-weighted average B-factor per residue. The column labelled 'Q< 0.9' lists the number of (and percentage) of residues with an average occupancy less than 0.9.

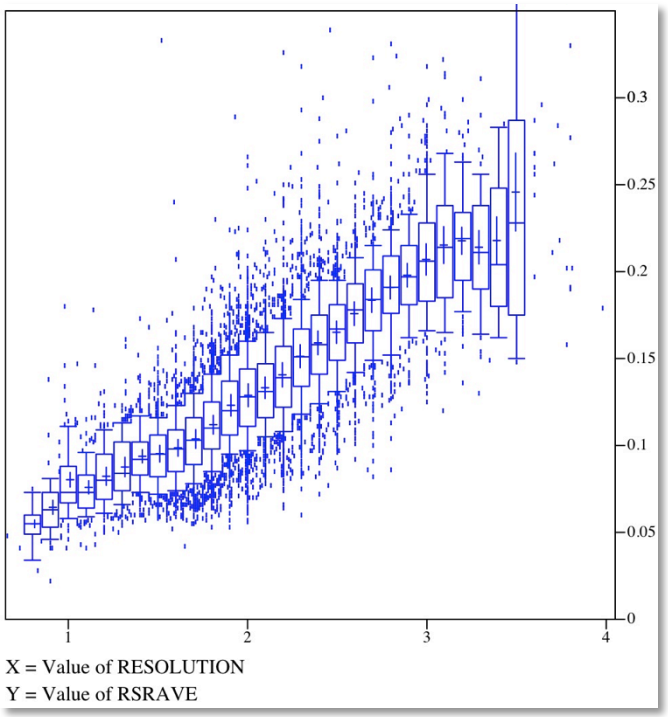
Mol	Chain	Analysed	<RSRZ>	#RSRZ>2	OWAB(Å^2)	Q<0.9
1	A	371/371 (100%)	-0.00	0 100 100	2, 37, 96, 164	0
1	C	371/371 (100%)	0.12	4 (1%) 81 65	2, 37, 96, 164	0
All	All	742/742 (100%)	0.06	4 (0%) 88 79	2, 37, 96, 164	0

All RSRZ outliers are listed below:

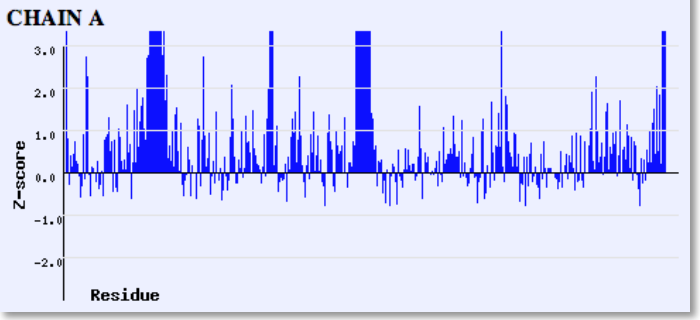
Mol	Chain	Res	Type	RSRZ
1	C	255	PHE	2.8
1	C	269	ILE	2.6
1	C	302	LEU	2.3
1	C	16	THR	2.2

RSR-Z (in EDS)

- RSR dependent on residue type and resolution
- Define RSR-Z (RSR, aa, d) = (RSR - <RSR(aa,d)>) / σ(RSR(aa,d))
  - aa = residue type
  - d = resolution (in bins of 0.2Å)
  - Calculated using 10,000s of EDS entries
- Example: Trp between 2.4 and 2.6Å:
  - 2012: 58321 observations, <> = 0.1419, s = 0.0537
  - 2008: 26794 observations, <> = 0.1602, s = 0.0660
  - RSR=0.25 ➔ RSR-Z=2.0 (2008: 1.4)



Chain	Nres	<Z>	Sigma	%<-1	%<0	%>1	%>2
A	425	0.89	2.59	0.00	30.59	20.94	9.41
B	24	8.55	9.25	0.00	0.00	100.00	95.83
C	12	7.98	9.01	0.00	0.00	100.00	100.00
All	**461**	**1.47**	**3.57**	**0.00**	**28.20**	**27.11**	**16.27**



Validation Reports

Mol	Type	Chain	Res	Atoms	RSR	LLDF	B-factors(Å²)	Q<0.9
6	CU	A	1744	1/1	0.24	5.45	74,74,74,74	0
6	CU	B	1748	1/1	0.19	2.68	66,66,66,66	0
4	NAG	D	1738	14/15	0.18	2.37	81,83,85,85	0
6	CU	B	1747	1/1	0.15	2.33	81,81,81,81	0
4	NAG	A	1735	14/15	0.16	1.74	84,86,88,89	0
6	CU	B	1750	1/1	0.20	0.87	63,63,63,63	0
6	CU	A	1737	1/1	0.20	0.72	49,49,49,49	0
6	CU	B	1742	1/1	0.20	0.64	55,55,55,55	0
4	NAG	C	1736	14/15	0.14	0.50	82,85,88,88	0
5	CA	B	1743	1/1	0.18	0.35	47,47,47,47	0
4	NAG	B	1735	14/15	0.19	0.31	69,70,73,73	0
6	CU	A	1743	1/1	0.18	0.14	84,84,84,84	0
4	NAG	B	1738	14/15	0.14	0.14	81,83,84,85	0

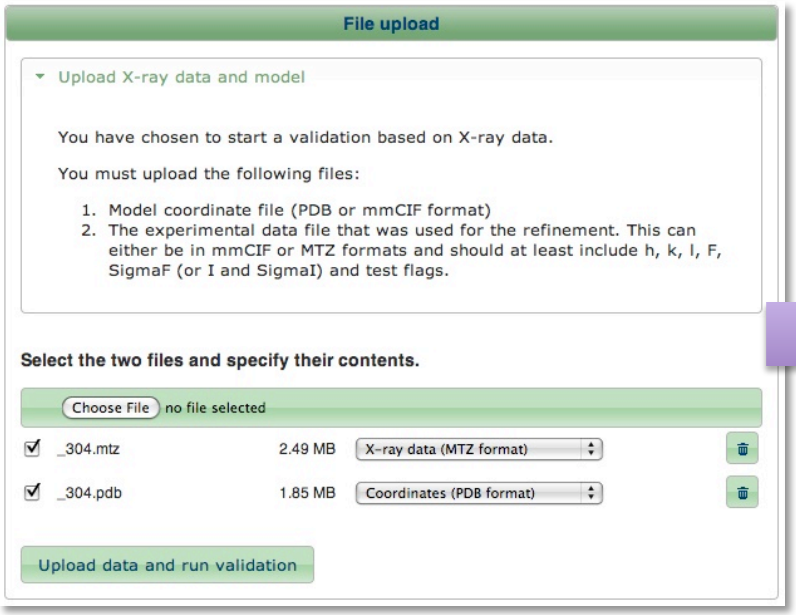
- Model/data fit ligands etc.
  - RSR as usual
  - Can't usually compute RSR-Z due to few/no occurrences in PDB

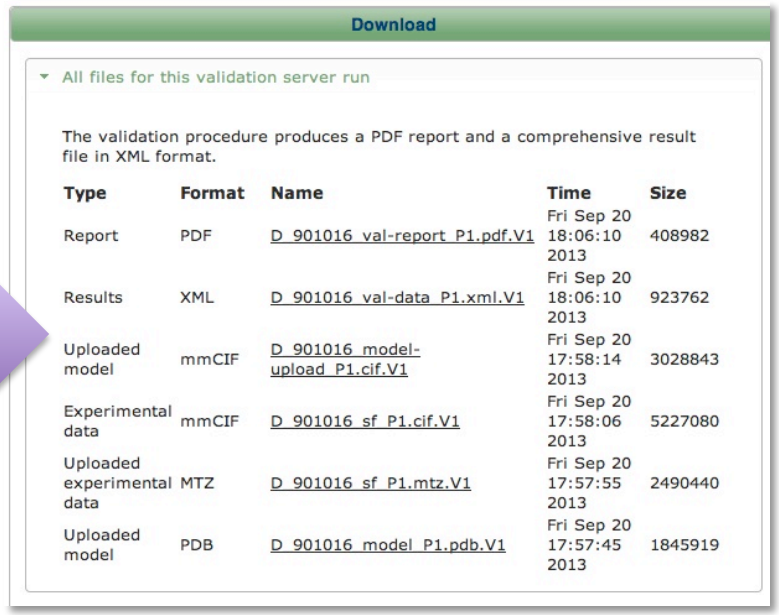
  - New: "***LLDF***" – *Local Ligand Density Fit* = Z-score of ligand RSR relative to nearby polymeric residues (incl symmetry)

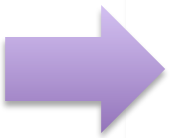
$$LLDF = (RSR(ligand) - \langle RSR(site) \rangle) / s(RSR(site))$$

Validation Server

- Stand-alone wwPDB X-ray validation server about to be released
- Input: coordinate (PDB, PDBx) and structure-factor file (CIF, MTZ)
- Output: PDF report and XML file with all details

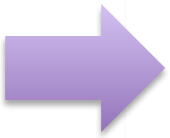






Preliminary wwPDB X-ray Structure Validation Report (i)

Sep 23, 2013 – 03:10 PM EDT



DISCLAIMER

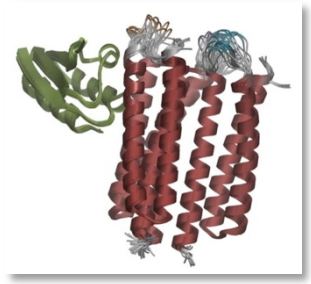
This is a preliminary version of the new style of wwPDB validation report.
This report is produced by the wwPDB validation pipeline
before deposition or annotation of the structure.
This is not an official wwPDB validation report and is not a proof of deposition.
This report should not be submitted to journals.
We welcome your comments at validation@mail.wwpdb.org
A user guide is available at http://wwpdb.org/ValidationPDFNotes.html

The following versions of software and data (see references) were used in the production of this report:

MolProbity	:	4.02b-467
Mogul	:	1.15 2013
Xtriage (Phenix)	:	dev-1439
EDS	:	stable21480
Percentile statistics	:	20591
Refmac	:	5.8.0049
CCP4	:	6.3.0 (Settle)
Ideal geometry (proteins)	:	Engh & Huber (2001)
Ideal geometry (DNA, RNA)	:	Parkinson et. al. (1996)
Validation Pipeline (wwPDB-VP)	:	stable21480

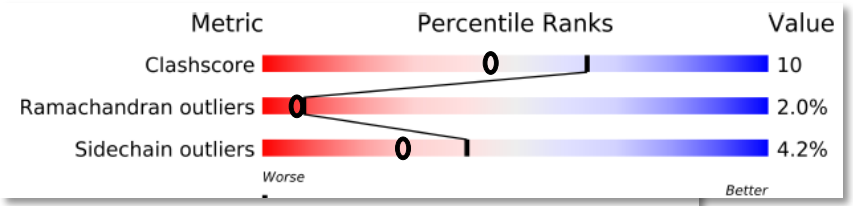
wwPDB Validation Plans for NMR

- NMR VTF recommendations published in *Structure*
- VTF recommendations will develop over time
  - "Phase 2" meeting at EBI in November 2013
- Workshop to discuss a unified NMR restraints format (crucial for automatic validation) at EBI in November 2013
- Pipeline in development
  - Currently includes MolProbity, Mogul, ensemble analysis (Cyrange)
  - To be synched with X-ray
  - BMRB will provide software for chemical-shift validation



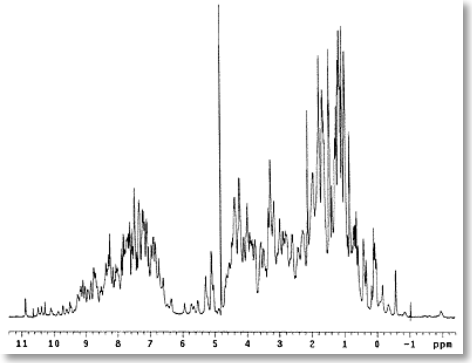
Re-using X-ray Pipeline Components

- MolProbity and Mogul for model-only validation
- Global quality scores will be reported as averages over the ensemble and for "*well-defined residues*" only
- Detailed residue-level validation scores available for all residues and all models



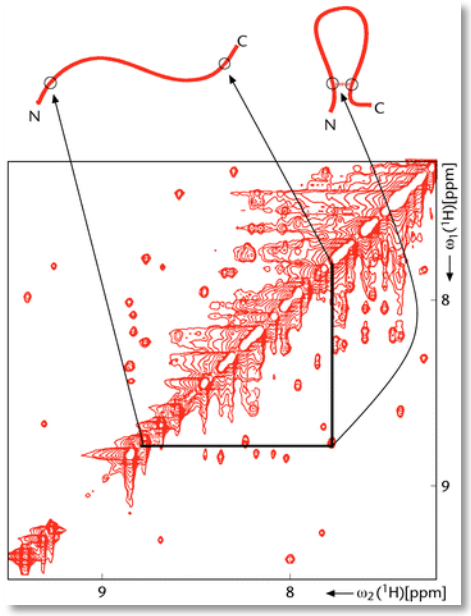
Validation of Chemical Shifts

- VTF does not recommend to use structural information for the validation of chemical shifts at this time
- Validation report to provide information on
  - Completeness of assignments
  - Referencing corrections
  - List of nuclei with statistically unusual chemical shift values
- Recommended options: AVS, LACS, PANAV and BMRB stats



Validation of Experimental Restraints

- To implement restraint-conversion software in deposition pipeline
  - Used for NMR Restraints Grid (NRG) database
  - Workshop to discuss a unified format (Nov 2013)
- Validation report will include
  - Statistics on the restraints data
  - Statistics on violations
  - List of worst violations
- Initially, only distance and dihedral angle restraints



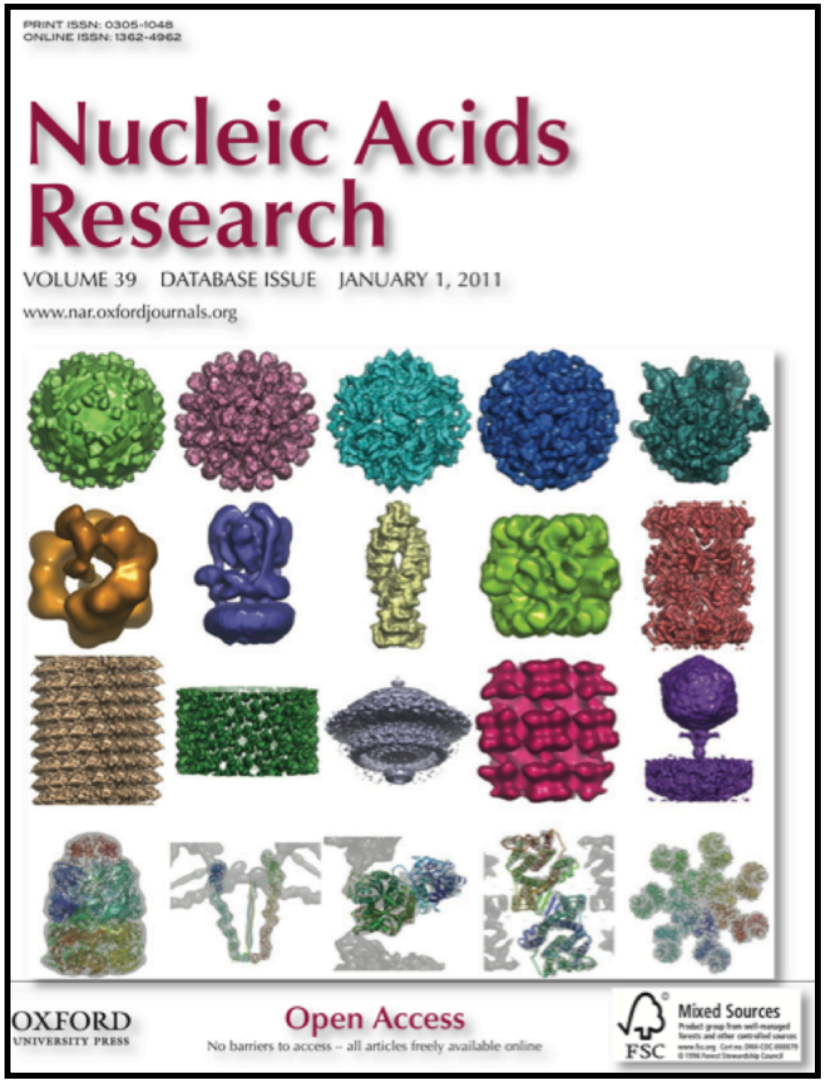
wwPDB/EMDataBank Validation Plans for EM

- Not many accepted validation standards yet
- Start with geometric checks and “sanity checks”
- Pipeline in development
  - Map visual analysis (Chimera): visual sanity check of the map and map/model overlay
  - Model validation à la X-ray (MolProbity) – clashes?
- Later
  - Harvest more validation-related data (e.g., results of tilt-pair analysis)
  - As new methods are developed and become community-accepted they can be incorporated into the validation pipeline

Experimental Methods

EMDataBank: Unified Data Resource for 3DEM

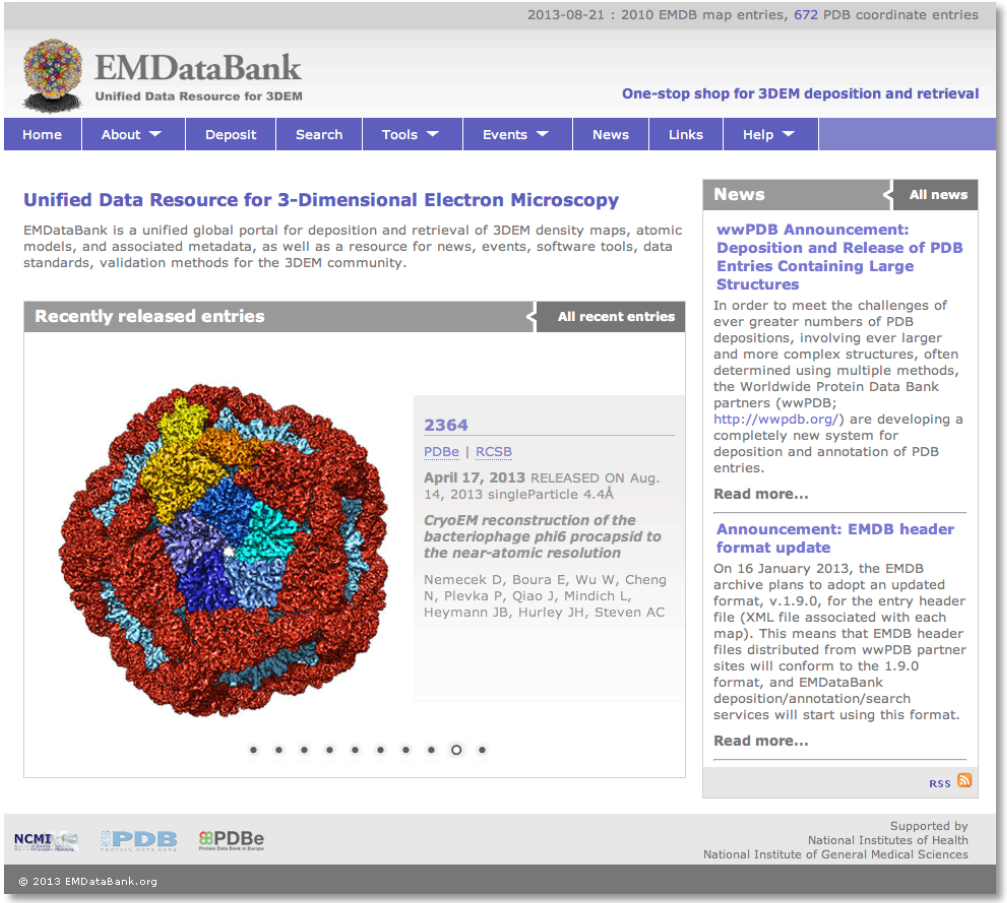
- Collaborative project between PDBe, RCSB PDB and Baylor-NCMI
- "*One-Stop Shop*" for collection of EM maps and coordinate models (to become part of D&A)
- Standardized map redistribution format
- Cross-referencing between maps and models (EMDB←→PDB)



EMDataBank Services

- News, software list, information about dictionaries, conventions, FAQ, community links, statistics, ..

- Various search mechanisms for EMDB data
- Recently released entries
- Map+model 3D Java viewer



EMDataBank: NIH grant funded

Wah Chiu PI, Helen Berman & Gerard Kleywegt co-PIs

Specific Aims

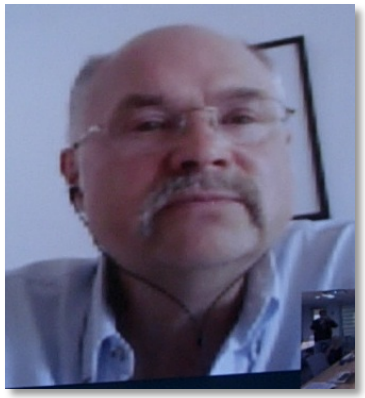
1. **Establish map-validation methods**
   - Use representative raw image datasets from both our laboratory and broad group of collaborators
2. **Establish model-validation methods**
   - Use map and model data from EMDB and PDB and community-contributed data
3. **Define standards for 3DEM data exchange and archiving**
   - Continue development of 3DEM terms in the EMDB data model and PDBx by adding metadata relevant to the validation procedures established above
   - Establish an agreed upon data exchange file format for maps, and develop or modify software converters to support the new and current data formats
4. **Facilitate the dissemination of 3DEM validation standards**
5. **Integrate 3DEM data standards and map and model validation into the wwPDB pipeline**
   - The map validation metadata and map-derived model-validation procedures developed through this project will be integrated into the wwPDB D&A system

EMDataBank EM VTF

- Main recommendations for EM maps
  - Standards for assessing resolution and accuracy of a map need to be developed
  - Structural features in a map should be in accordance with the claimed resolution
- Main recommendations for models fitted into EM maps
  - Criteria for assessing models need to be developed
  - Capability to archive coarse-grained representations of models is needed
- More research and development needed!

SAS Task Force

- Members
  - Jill Trewhella (Chair, University of Sydney)
  - Dmitri Svergun (European Molecular Biology Laboratory-Hamburg)
  - John Tainer (The Scripps Research Institute)
  - Wayne Hendrickson (Columbia University)
  - Mamoru Sato (Yokohama City University)
  - Torsten Schwede (University of Basel)



Preliminary Recommendations

- Develop an international repository for SAS data
- Standard dictionary required for definition of terms involved in data collection
- Shape and atomistic models based on SAS data should be archived
- Criteria for assessment of the uniqueness and quality of models needs to be defined
- Models derived from diverse hybrid data should be archived
- There is a need for key people involved in the different wwPDB VTFs to come together to discuss what the PDB should be archiving

Hybrid Methods

- Task force to be convened for meeting at EBI in autumn of 2014 about archiving
  - Spring 2014: scientific/methodological meeting at Baylor
- Include experts in hybrid structure determination, representatives from the various wwPDB task forces, and archiving experts
- Discussion items to include:
  - What should be archived and where?
    - Information content?
  - What information/data should be provided on deposition?
  - What validation methods can be applied?
  - What to do with models that are not atomistic or not substantially determined by the deposited data?

NMR

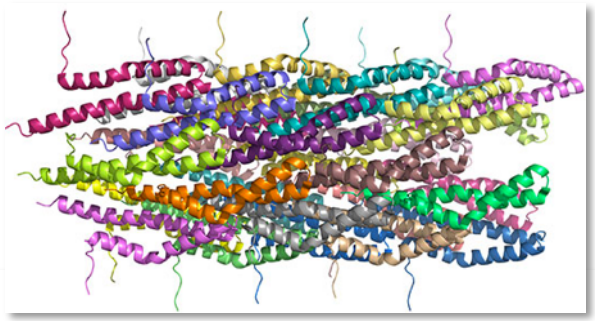
John Markley

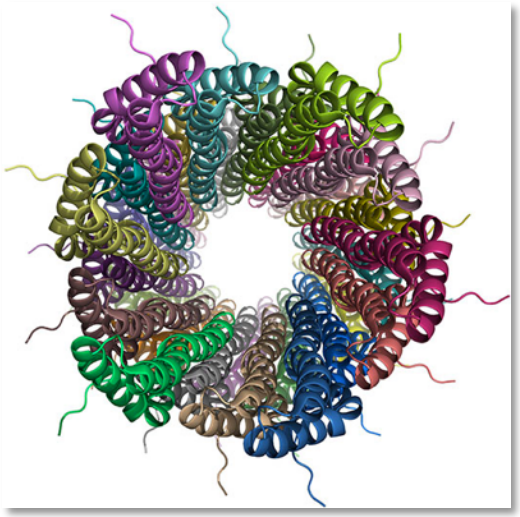
WORLDWIDE PDB
PROTEIN DATA BANK

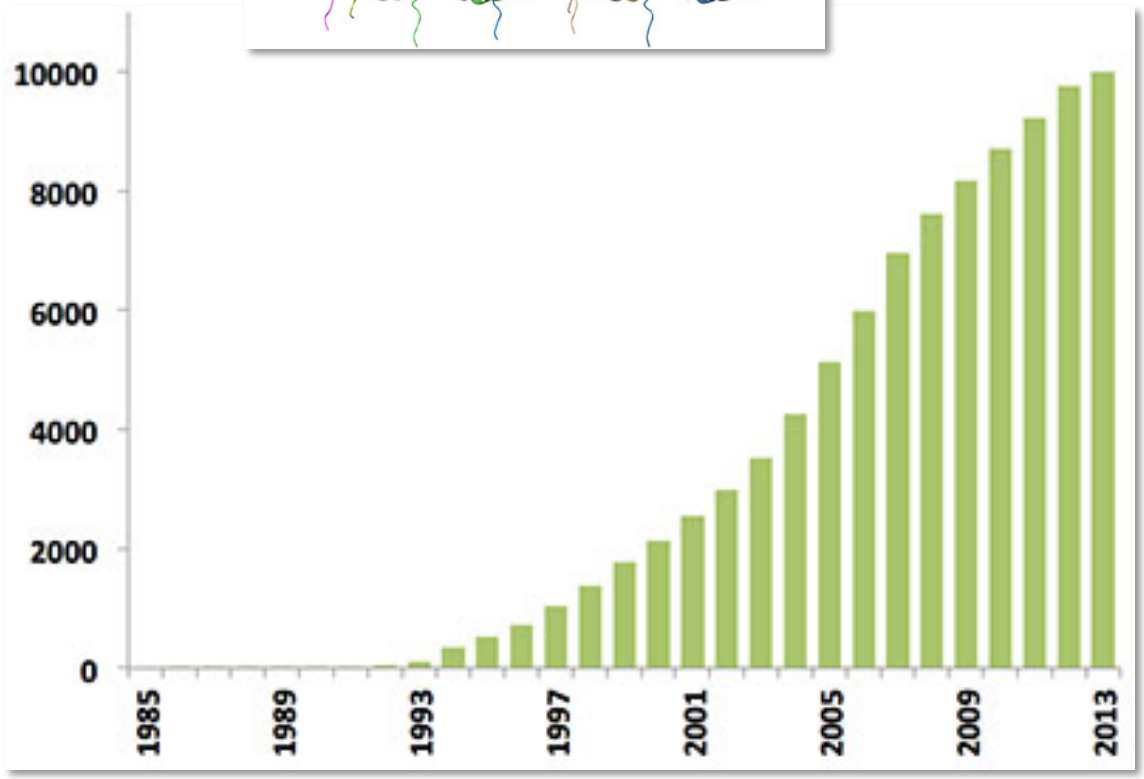
wwpdb.org

10,000th NMR structure in the PDB!

- June 2013







BMRB/XML and /RDF for Semantic Web

(Recently developed by PDBj and BMRB)

NMR-STAR v3

```
######################
#  Entry information  #
######################

save_entry_information
   _Entry.Sf_category                   entry_information
   _Entry.Sf_framecode                  entry_information
   _Entry.ID                            15400
   _Entry.Title
;
Backbone and side chain chemical shift assignments of the F1
53-to-5-flurotryptophan mutant of human cardiac troponin C
;
   _Entry.Version_type                  new
   _Entry.Submission_date               2007-07-20
   _Entry.Accession_date                2007-07-20
   _Entry.Last_release_date             .
   _Entry.Original_release_date         .
   _Entry.Origination                   author
   _Entry.NMR_STAR_version              3.0.8.100
   _Entry.Original_NMR_STAR_version     3.0.8.100
   _Entry.Experimental_method           NMR
   _Entry.Experimental_method_subtype   solution
   _Entry.Details                       .
   _Entry.BMRB_internal_directory_name
```

bmr15400.str

BMRBxTool

BMRB/XML

```
<BMRBx:entryCategory>
  <BMRBx:entry id="15400">
    <BMRBx:accession_date>2006-12-07+09:00</BMRBx:accession_date>
    <BMRBx:bmrb_internal_directory_name xsi:nil="true"/>
    <BMRBx:details xsi:nil="true"/>
    <BMRBx:experimental_method>NMR</BMRBx:experimental_method>
    <BMRBx:experimental_method_subtype>SOLUTION</BMRBx:experimental_method_subtype>
    <BMRBx:last_release_date xsi:nil="true"/>
    <BMRBx:nmr_star_version>3.0.8.100</BMRBx:nmr_star_version>
    <BMRBx:original_nmr_star_version>3.0.8.100</BMRBx:original_nmr_star_version>
    <BMRBx:original_release_date xsi:nil="true"/>
    <BMRBx:origination>author</BMRBx:origination>
    <BMRBx:sf_category>entry_information</BMRBx:sf_category>
    <BMRBx:sf_framecode>entry_information</BMRBx:sf_framecode>
    <BMRBx:submission_date>2006-12-07+09:00</BMRBx:submission_date>
    <BMRBx:title>Backbone and side chain chemical shift assignments of the F153-to-
5-flurotryptophan mutant of human cardiac troponin C</BMRBx:title>
    <BMRBx:version_type>original</BMRBx:version_type>
  </BMRBx:entry>
</BMRBx:entryCategory>
```

bmr15400.xml

BMRB/RDF

BMRBoTool

```
<BMRBo:has_entryCategory>
  <BMRBo:entryCategory rdf:about="http://bmrbpub.protein.osaka-u.ac.jp/rdf/bmr15400/entryCategory">
    <BMRBo:has_entry>
      <BMRBo:entry rdf:about="http://bmrbpub.protein.osaka-u.ac.jp/rdf/bmr15400/entry/15400">
        <BMRBo:of_datablock rdf:resource="http://bmrbpub.protein.osaka-u.ac.jp/rdf/bmr15400"/>
        <BMRBo:entry.id>15400</BMRBo:entry.id>
        <BMRBo:entry.accession_date>2006-12-07+09:00</BMRBo:entry.accession_date>
        <BMRBo:entry.experimental_method>NMR</BMRBo:entry.experimental_method>
        <BMRBo:entry.experimental_method_subtype>SOLUTION</BMRBo:entry.experimental_method_subtype>
        <BMRBo:entry.nmr_star_version>3.0.8.100</BMRBo:entry.nmr_star_version>
        <BMRBo:entry.original_nmr_star_version>3.0.8.100</BMRBo:entry.original_nmr_star_version>
        <BMRBo:entry.origination>author</BMRBo:entry.origination>
        <BMRBo:entry.sf_category>entry_information</BMRBo:entry.sf_category>
        <BMRBo:entry.sf_framecode>entry_information</BMRBo:entry.sf_framecode>
        <BMRBo:entry.submission_date>2006-12-07+09:00</BMRBo:entry.submission_date>
        <BMRBo:entry.title>Backbone and side chain chemical shift assignments of the F153-to-5-fluro
tryptophan mutant of human cardiac troponin C</BMRBo:entry.title>
        <BMRBo:entry.version_type>original</BMRBo:entry.version_type>
      </BMRBo:entry>
    </BMRBo:has_entry>
  </BMRBo:entryCategory>
</BMRBo:has_entryCategory>
```

bmr15400.rdf

Format	NMR-STAR v3	BMRB/XML	BMRB/RDF
Ontology	NMR-STAR v3 dictionary	BMRB/XML schema	BMRB/OWL
Validation	ADIT-NMR & manual	Schema validation	RDF validator
Data type	Text file (.str)	XML file (.xml)	RDF file (.rdf)

Yokochi, M. et al., *Nucleic Acids Res.* in preparation

Extensive Remediation of the BMRB Archive

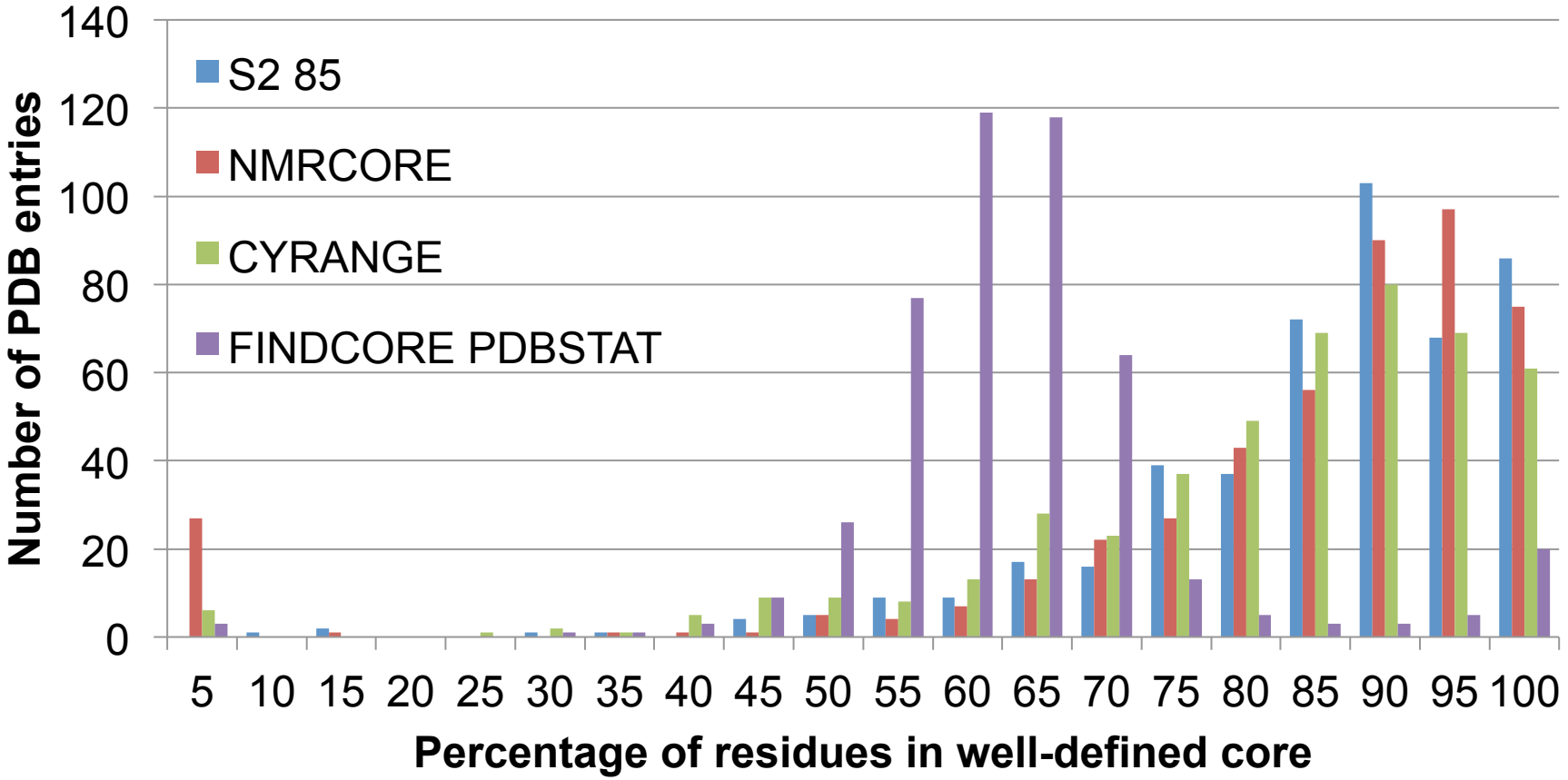
Carried out in conjunction with the development of the XML format and a relational database containing experimental data from BMRB and coordinates from PDB

- Chemical component names normalized with PDB content
- Sample component names normalized
- Clean up of entity natural source designations: scientific name, genus, and species
- Correction of software vendor information
- Older entries updated to meet the current database constraints

Activities in Preparation for the wwPDB NMR validation pipeline

- Evaluation of validation tools
  - MolProbity has been tested on NMR structures
  - Several programs to assign “well-defined regions” have been compared
  - Restraint validation methods have been compared
  - Chemical shift validation approaches have been evaluated
- wwPDB Validation Task Force will meet November 17 to discuss implementation and extension of validation tools
- wwPDB workshop for software developers on NMR nomenclature will take place November 18-19

Defining the Well-defined



After evaluation, CYRANGE was selected as the tool to be used in the wwPDB NMR validation pipeline

Validating Chemical Shifts

- AVS, LACS, PANAV, BMRB stats, and SPARTA+ currently are in use at BMRB
- AVS, LACS, PANAV, and BMRB stats have been evaluated for use in the wwPDB annotation pipeline by testing against >6,000 BMRB entries and ~100 depositor uploaded chemical shift files. CheckShift also will be evaluated if we can obtain the software code
- Current recommendation is to use LACS to check chemical shift referencing and to use BMRB stats to identify chemical shift statistical outliers
- Recommendations are based on robustness, completeness, and simplicity of results
- Final recommendations will depend on discussions with software developers and further testing
  - Results will be presented at the November 17 wwPDB VTF meeting

Funding Status

- Grant (~70% of previous award) through August 2014 from the NIH Library of Medicine
- Proposal submitted to NIH National Institute of General Medical Sciences received a top 2% rating
- Requested a September 2013 start date, but because it is going to the October council meeting, the earliest start date is December 1, 2013
- Funding contingent on the NIH budget

Outreach

Haruki Nakamura

WORLDWIDE PDB
PROTEIN DATA BANK

wwpdb.org

Worldwide Protein Data Bank Foundation

- New Chairman of the Board
  - Dr. Anthony Nicholls, OpenEye Scientific Software, Inc.
- 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

2012 Symposium: Basis for Life Science and Drug Development

October 13, 2012 in Osaka, Japan

2013 Symposium: A Celebration of Open Access in Structural Biology

Recognizing the career and achievements of Professor Helen M. Berman

Sponsors

BLAKE INDUSTRIES, INC.

BRUKER

Cambridge Isotope Laboratories

DECTRIS

EMERALD BIOSTRUCTURES

FEI

FUJITSU

Genentech
A Member of the Roche Group

HAMPTON RESEARCH

JEOL

OpenEye Scientific Software

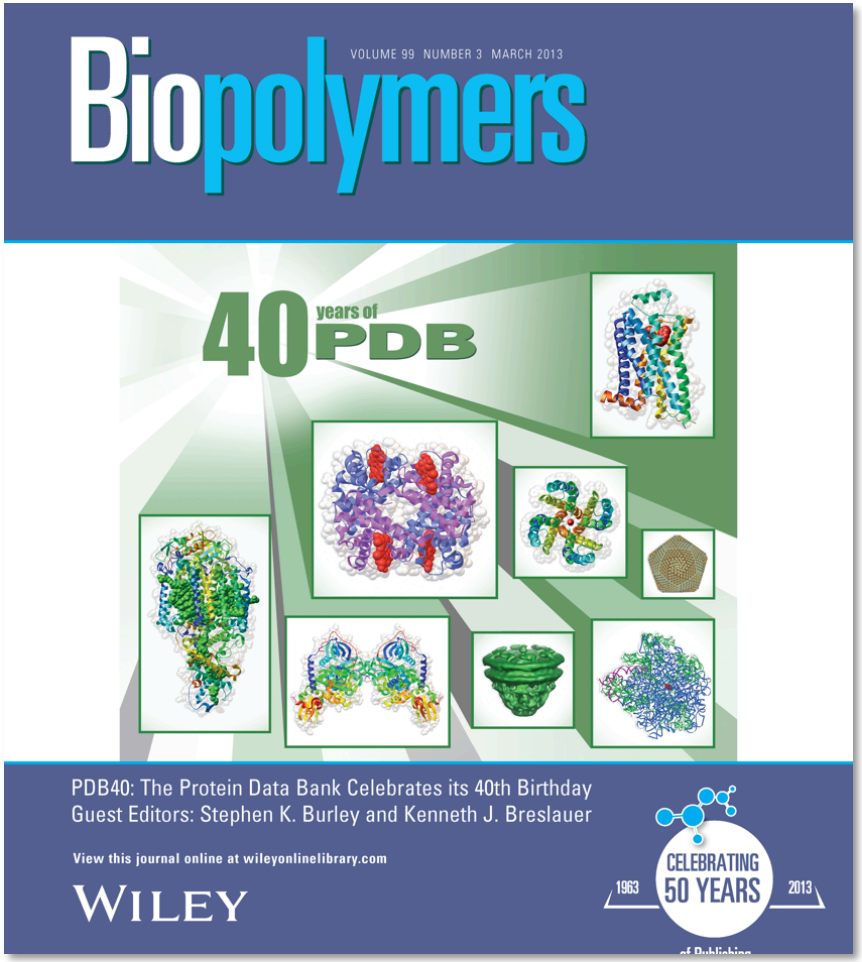
Rigaku

SCHRÖDINGER.

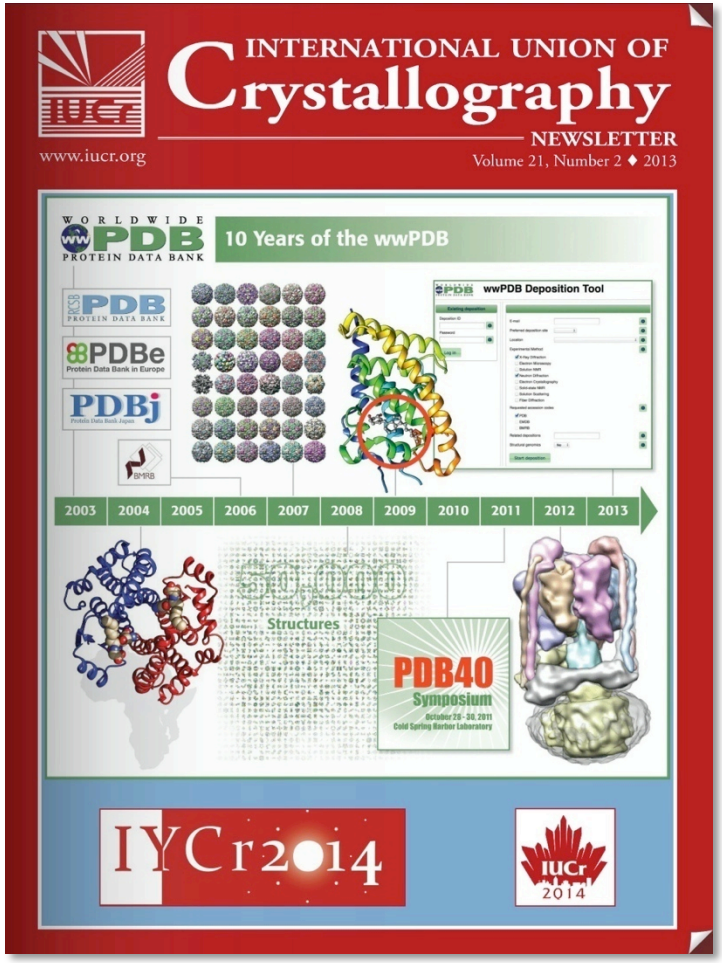
PDB40 Issue of *Biopolymers*

Volume 99, Issue 3, March 2013

- PDB40: The Protein Data Bank celebrates its 40th birthday, Stephen K. Burley
- Studying and polishing the PDB's macromolecules, Jane S. Richardson and David C. Richardson
- Abstracting knowledge from the Protein Data Bank, Nicholas Furnham, Roman A. Laskowski and Janet M. Thornton
- The impact of influenza hemagglutinin fusion peptide length and viral subtype on its structure and dynamics, Justin L. Lorieau, John M. Louis and Ad Bax
- Sweet entanglements-protein: Glycan interactions in two HIV-inactivating lectin families, Leonardus M. I. Koharudin and Angela M. Gronenborn
- A primer in macromolecular linguistics, David B. Searls
- The Future of the Protein Data Bank, Helen M. Berman, Gerard J. Kleywegt, John L. Markley, Haruki Nakamura

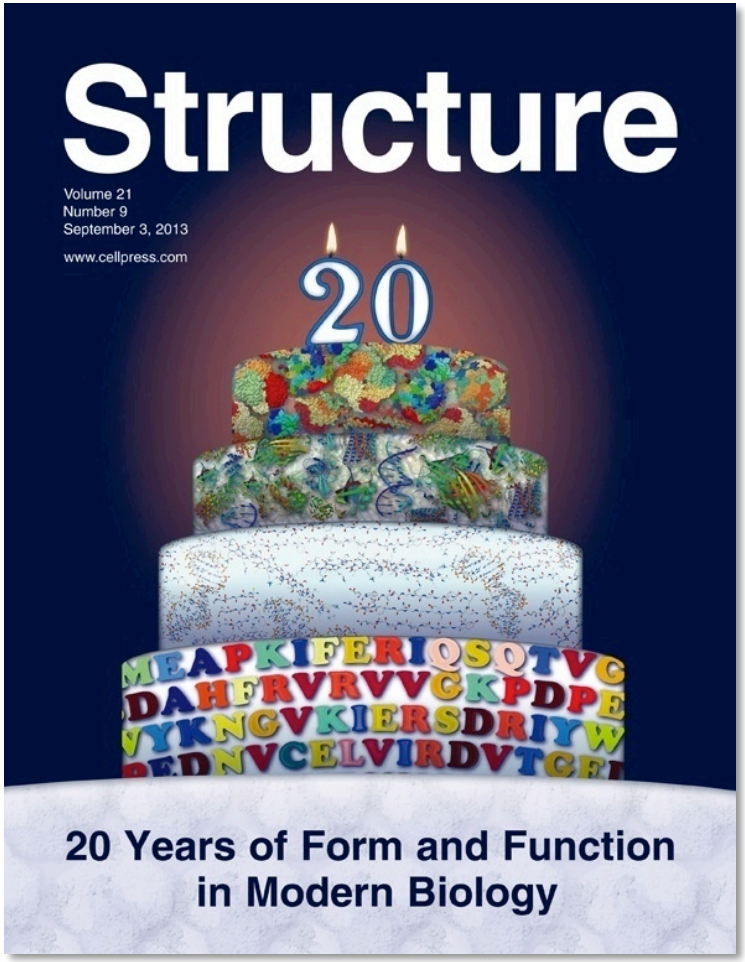


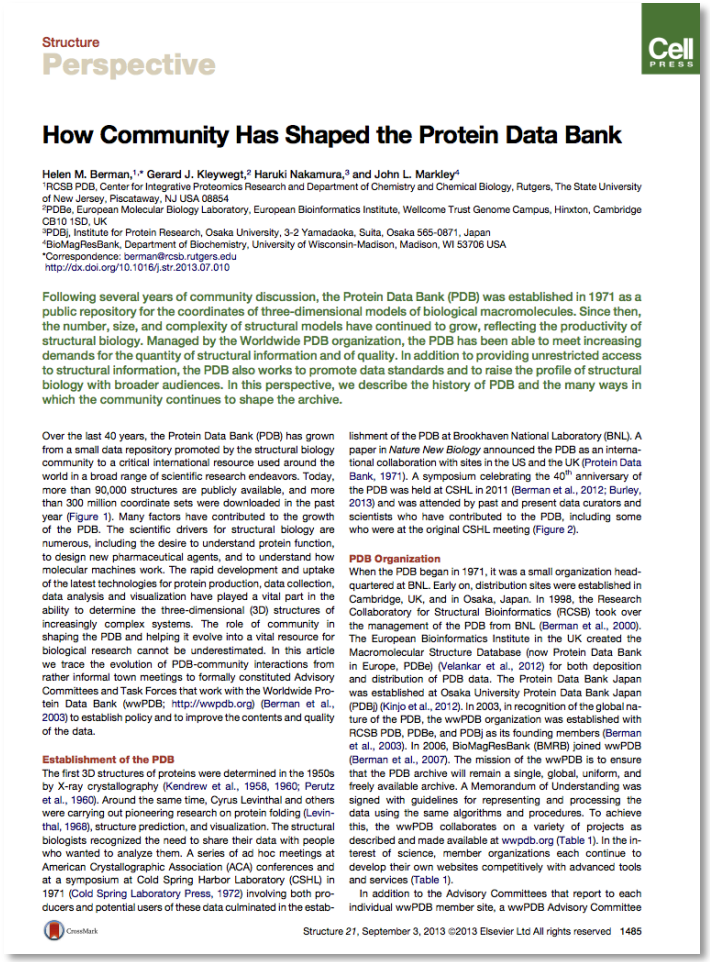
10 Years of the wwPDB



- IUCr, ACA, ECA Newsletters

20th Anniversary Issue of *Structure*





Structure
Perspective

Cell PRESS

How Community Has Shaped the Protein Data Bank

Helen M. Berman,[1,*] Gerard J. Kleywegt,[2] Haruki Nakamura,[3] and John L. Markley[4]
[1]RCSB PDB, Center for Integrative Proteomics Research and Department of Chemistry and Chemical Biology, Rutgers, The State University of New Jersey, Piscataway, NJ USA 08854
[2]PDBe, European Molecular Biology Laboratory, European Bioinformatics Institute, Wellcome Trust Genome Campus, Hinxton, Cambridge CB10 1SD, UK
[3]PDBj, Institute for Protein Research, Osaka University, 3-2 Yamadaoka, Suita, Osaka 565-0871, Japan
[4]BioMagResBank, Department of Biochemistry, University of Wisconsin-Madison, Madison, WI 53706 USA
*Correspondence: berman@rcsb.rutgers.edu
http://dx.doi.org/10.1016/j.str.2013.07.010

Following several years of community discussion, the Protein Data Bank (PDB) was established in 1971 as a public repository for the coordinates of three-dimensional models of biological macromolecules. Since then, the number, size, and complexity of structural models have continued to grow, reflecting the productivity of structural biology. Managed by the Worldwide PDB organization, the PDB has been able to meet increasing demands for the quantity of structural information and of quality. In addition to providing unrestricted access to structural information, the PDB also works to promote data standards and to raise the profile of structural biology with broader audiences. In this perspective, we describe the history of PDB and the many ways in which the community continues to shape the archive.

Over the last 40 years, the Protein Data Bank (PDB) has grown from a small data repository promoted by the structural biology community to a critical international resource used around the world in a broad range of scientific research endeavors. Today, more than 90,000 structures are publicly available, and more than 300 million coordinate sets were downloaded in the past year (Figure 1). Many factors have contributed to the growth of the PDB. The scientific drivers for structural biology are numerous, including the desire to understand protein function, to design new pharmaceutical agents, and to understand how molecular machines work. The rapid development and uptake of the latest technologies for protein production, data collection, data analysis and visualization have played a vital part in the ability to determine the three-dimensional (3D) structures of increasingly complex systems. The role of community in shaping the PDB and helping it evolve into a vital resource for biological research cannot be underestimated. In this article we trace the evolution of PDB-community interactions from rather informal town meetings to formally constituted Advisory Committees and Task Forces that work with the Worldwide Protein Data Bank (wwPDB; http://wwpdb.org) (Berman et al., 2003) to establish policy and to improve the contents and quality of the data.

Establishment of the PDB

The first 3D structures of proteins were determined in the 1950s by X-ray crystallography (Kendrew et al., 1958, 1960; Perutz et al., 1960). Around the same time, Cyrus Levinthal and others were carrying out pioneering research on protein folding (Levinthal, 1968), structure prediction, and visualization. The structural biologists recognized the need to share their data with people who wanted to analyze them. A series of ad hoc meetings at American Crystallographic Association (ACA) conferences and at a symposium at Cold Spring Harbor Laboratory (CSHL) in 1971 (Cold Spring Laboratory Press, 1972) involving both producers and potential users of these data culminated in the establishment of the PDB at Brookhaven National Laboratory (BNL). A paper in *Nature New Biology* announced the PDB as an international collaboration with sites in the US and the UK (Protein Data Bank, 1971). A symposium celebrating the 40th anniversary of the PDB was held at CSHL in 2011 (Berman et al., 2012; Burley, 2013) and was attended by past and present data curators and scientists who have contributed to the PDB, including some who were at the original CSHL meeting (Figure 2).

PDB Organization

When the PDB began in 1971, it was a small organization headquartered at BNL. Early on, distribution sites were established in Cambridge, UK, and in Osaka, Japan. In 1998, the Research Collaboratory for Structural Bioinformatics (RCSB) took over the management of the PDB from BNL (Berman et al., 2000). The European Bioinformatics Institute in the UK created the Macromolecular Structure Database (now Protein Data Bank in Europe, PDBe) (Velankar et al., 2012) for both deposition and distribution of PDB data. The Protein Data Bank Japan was established at Osaka University Protein Data Bank Japan (PDBj) (Kinjo et al., 2012). In 2003, in recognition of the global nature of the PDB, the wwPDB organization was established with RCSB PDB, PDBe, and PDBj as its founding members (Berman et al., 2003). In 2006, BioMagResBank (BMRB) joined wwPDB (Berman et al., 2007). The mission of the wwPDB is to ensure that the PDB archive will remain a single, global, uniform, and freely available archive. A Memorandum of Understanding was signed with guidelines for representing and processing the data using the same algorithms and procedures. To achieve this, the wwPDB collaborates on a variety of projects as described and made available at wwpdb.org (Table 1). In the interest of science, member organizations each continue to develop their own websites competitively with advanced tools and services (Table 1).

In addition to the Advisory Committees that report to each individual wwPDB member site, a wwPDB Advisory Committee

Upcoming Workshops

Workshop on the PDBx/ mmCIF Data Exchange Format For Structural Biology

- Rutgers, The State University of New Jersey
- October 22, 2013

wwPDB Workshop on mmCIF/PDBx for Programmers

- EMBL-EBI, Cambridge, UK
- November 20 – 21, 2013

2014: International Year of Crystallography

In distribution online and at scientific and educational meetings worldwide.

Related seminars and webinars to follow.

Facebook, Website, Mailing Lists

- Informing users, e.g.:
  - SF remediation
  - BIRD release
  - Large structures
  - SAS Task Force report
  - 10,000th NMR structure
  - 10 years of wwPDB
  - New validation reports
  - NMR VTF report
  - Conferences, workshops
  - Publications, posters