WORLDWIDE PDB
PROTEIN DATA BANK

wwpdb.org

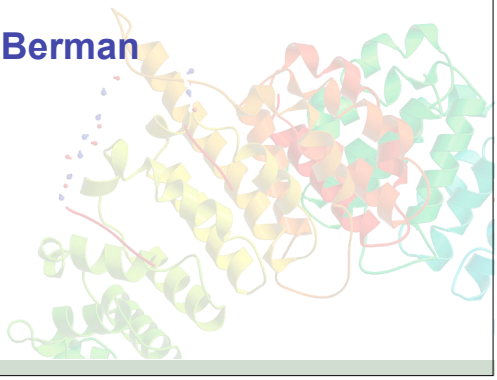
October 1, 2010

Agenda

Summary Overview	**Helen Berman**
Common Deposition and Annotation Tool	**Martha Quesada**
Method- and Molecule-specific Activities	**John Markley** **Gerard Kleywegt** **Helen Berman**
Funding Issues and Discussion	**Haruki Nakamura** **John Markley**

Summary Overview

Helen Berman



October 1, 2010

2009 wwPDBAC Recommendations

NMR

- VTF should publish a white paper
- Validation reports available for reviewers
  All journals notified
- Evaluate SMS utilization and determine what level of support is warranted

X-ray

- VTF should publish white paper
  In final stages of preparation
- Validation reports should be made available to reviewers
  All journals notified

SAXS

- Establish Task force
  Task force appointed
- Prepare white paper
- Work with journal editors

PDB Format

- Implement in Q4 2010
  Format under review

wwPDB

August 2009 - September 2010

- Funding currently stable; some long-term issues
- Continued growth of archive
- Continued intensive staff interactions
- Increased use of data
- Substantial progress in Common Tool project
- Establishment of wwPDB Foundation
- Close journal interactions
- Continued Task Force activity
- Implementation of mandatory chemical shift deposition
- Draft specification of new format

October 1, 2010

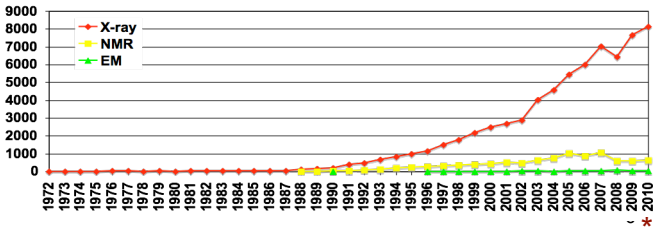
PDB Depositions

Last Updated: 11 Aug 2010

Year	Total Depositions	Deposited To			Processed By		
		RCSB	PDBj	EBI	RCSB	PDBj	EBI
2000	2983	2445	10	528	2297	158	528
2001	3286	2673	118	495	2408	383	495
2002	3563	2769	289	505	2401	657	505
2003	4830	3488	673	669	3135	1026	669
2004	5508	3796	900	812	3083	1613	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	5415 (*8800)	4291	326	798	3421	1196	798
TOTAL	63048	46680	7312	9056	38438	15554	9056

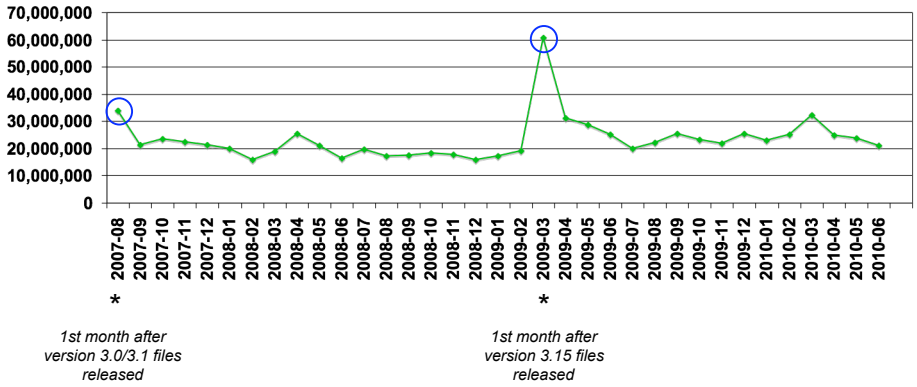
By deposition and processing site
*(projected)

By experimental method
*(projected)



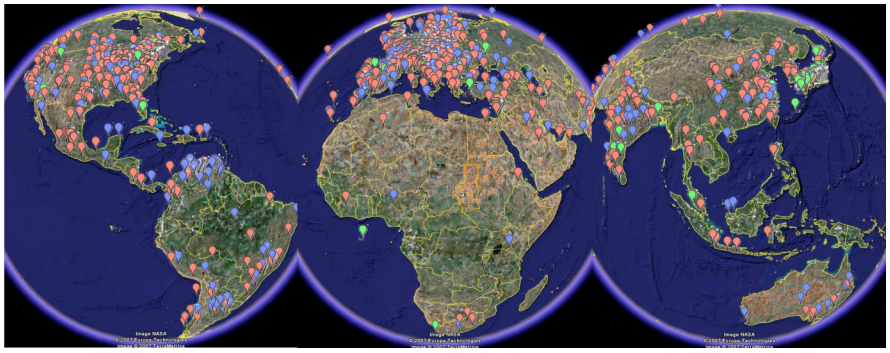
Note: NMR depositions at the RCSB PDB come through the BMRB

PDB FTP Downloads



October 1, 2010

PDB FTP Traffic (July 2009 - June 2010)



RCSB PDB
173 million
data downloads

PDBe
32 million
data downloads

PDBj
14 million
data downloads

Remediation

2010 projects

- Biological assemblies
- B-factors
- Antibiotics and peptide inhibitors

Roll-out scheduled January 2011

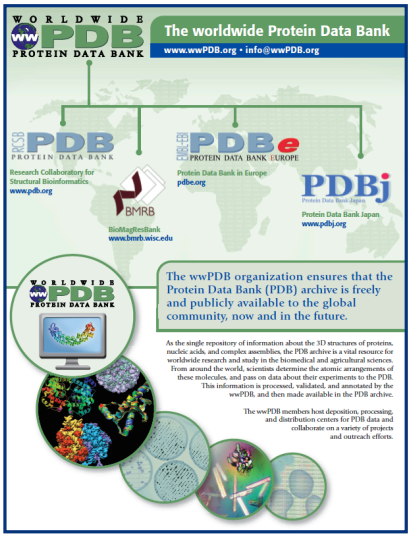
Common Tool for Deposition and Annotation

- Sequence annotation module V1.0, with enhanced user interface capability completed
- Ligand annotation module in initial testing
- Workflow engine on track
- Cross site data sharing architecture in place

Worldwide
Protein Data Bank
Foundation

wwPDB Foundation

- Board of Directors elected
- Bank account established
- Fund-raising plan being established



WORLDWIDE PDB PROTEIN DATA BANK

October 1, 2010

Is There Interest in a PDB Journal?

A proposed online, Open Access PDB Journal of articles associated with new PDB depositions would create new incentives for structure deposition, improved annotation of deposited data and generate revenue (to support task forces, advisory board meetings, *etc.*)

A survey was designed and field tested to gauge interest.

Sample questions:

- *I always publish my new structures in journal articles* (Agree/Disagree)
- *It is becoming harder to publish new structures in peer-reviewed journals* (Agree/Disagree)
- *All structures in the PDB should be properly described in peer-reviewed journal articles* (Agree/Disagree)
- *Overall, how do you rate the need for this new journal?*

Is There Interest in a PDB Journal?

Survey overview

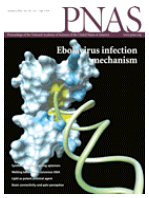
- Online survey sent to ~100 PIs with ≥10 depositions in the past 5 years (36 responses)
- Given to visitors to the ACA exhibit booth (26 responses)
- The overall response is positive, with some concerns about pricing
- Next step: survey expanded list of depositors, with a more detailed description of the product and different Open Access fees

WORLDWIDE PDB PROTEIN DATA BANK
October 1, 2010

Journal Interactions

- Coordination of *Instructions to Authors*
- Coordinating PDB release with online publication
  - Initially from *NPG* and IUCr journals
  - Now *JMB* (top PDB journal), *PNAS*, *Proteins*
  - In progress: *FEBS Journal*
- Validation reports
  - Currently required by: IUCr, under study at *Nature*

Published 454 entries in 2009

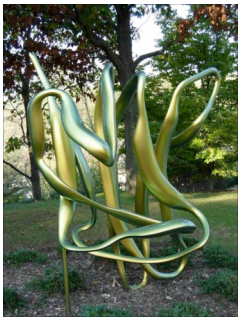


Published 663 entries in 2009

Published 123 entries in 2009

PDB 40 Symposium

- October 28-30, 2011
- Cold Spring Harbor Laboratory
  - Birthplace of PDB
  - To be held in conjunction with CSHL X-ray crystallography course
  - Main auditorium (capacity of 350)
  - CSHL handling logistics and committing $10,000 (local expenses of speakers)



wwPDB Interactions

- wwPDB Leadership Group
  - 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
- EM: Biweekly phone/VTC meetings

Common Deposition and Annotation (D&A) Tool

Martha Quesada

For the wwPDB D&A Project Team

October 1, 2010

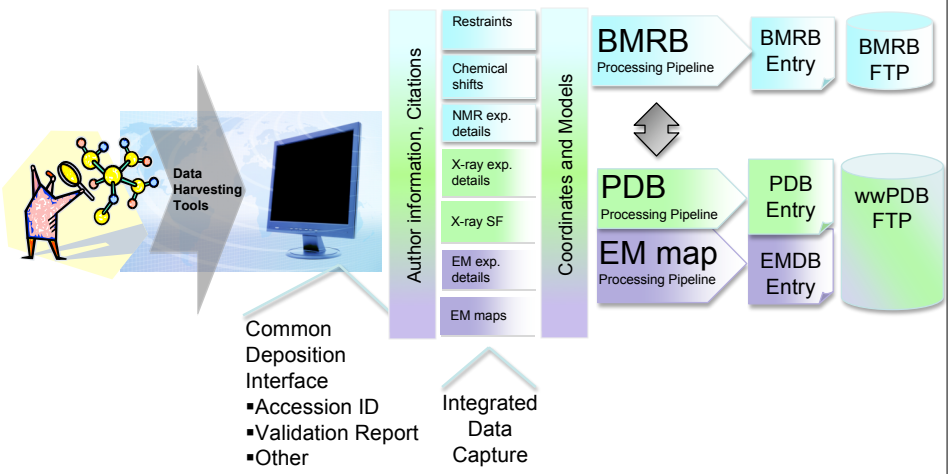
Multi-Disciplinary Project Team Representing All Four wwPDB Sites

Experts in:

- Content: annotators
- Functional applications: scientific programmers
- Graphical user interfaces
- Databases
- Application programming interfaces
- Workflow engine design
- Data sharing architecture

WORLDWIDE PDB PROTEIN DATA BANK
October 1, 2010

The Operational Vision



Project Goal

The goal is to implement a set of common deposition and annotation processes and tools that will enable the wwPDB to deliver a resource of increasingly high quality and dependability over the next 10 years.

The tools and processes will:

- Address the increase in complexity and experimental variety of submissions and the increase in deposition throughput
- Maximize the efficiency and effectiveness of data handling
- Provide for higher quality and completeness of submissions and annotation through improved use of graphical interfaces

What's in it for...

Depositors

- Uniform, interactive, and informative deposition interface
- Value-added validation input and annotation during deposition
- Faster processing

Annotators

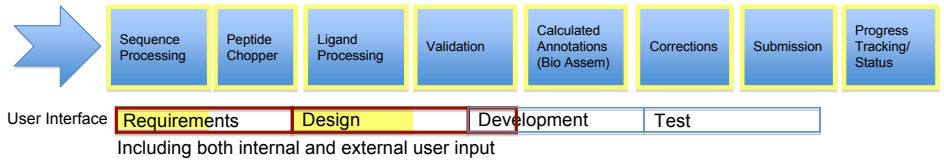
- Improved efficiency, freeing time for more advanced annotation
  - Improved quality early in the process
  - Automation of appropriate processing steps
  - Best-of-breed tools
  - Expanded functionality
  - (Shared maintenance and development effort)
- Enable system evolution through modularity

Data users

- Consistently annotated, high-quality archive

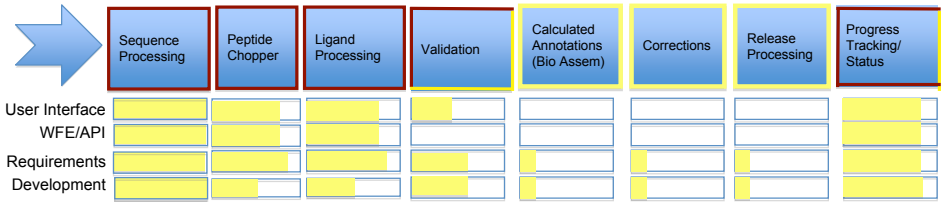
2010 Deliverables

Deposition pipeline



Including both internal and external user input

Annotation pipeline



Summary of Accomplishments

- Sequence module V1.0
- Extended archive format validation
- Ligand module
  - Initial deliverable in test
- Peptide ligand “chopper”
- Workflow manager on track
- Workflow interface on track
- Data sharing in place
- Deposition user interface
  - Requirements and mock-ups in hand

Sequence Processing Overview

Author-provided Cross-check with

```
COMPND MOL_ID: 1;
COMPND 2 MOLECULE: MYOGLOBIN;
SOURCE MOL_ID: 1;
SOURCE 2 ORGANISM_SCIENTIFIC: PHYSETER CATODON
SOURCE 4 ORGANISM_TAXID: 9755
DBREF 1MBN A 1 153 UNP P02185 MYG_PHYCA 1 153

SEQRES   1 A  153  VAL LEU SER GLU GLY GLU TRP GLN LEU VAL LEU HIS VAL
SEQRES   2 A  153  TRP ALA LYS VAL GLU ALA ASP VAL ALA GLY HIS GLY GLN
SEQRES   3 A  153  ASP ILE LEU ILE ARG LEU PHE LYS SER HIS PRO GLU THR
SEQRES   4 A  153  LEU GLU LYS PHE ASP ARG PHE LYS HIS LEU LYS THR GLU
SEQRES   5 A  153  ALA GLU MET LYS ALA SER GLU ASP LEU LYS LYS HIS GLY
SEQRES   6 A  153  VAL THR VAL LEU THR ALA LEU GLY ALA ILE LEU LYS LYS
SEQRES   7 A  153  LYS GLY HIS HIS GLU ALA GLU LEU LYS PRO LEU ALA GLN
SEQRES   8 A  153  SER HIS ALA THR LYS HIS LYS
SEQRES   9 A  153  GLU PHE ILE SER GLU ALA ILE
SEQRES  10 A  153  ARG HIS PRO GLY ASP PHE GLY ALA ASP ALA GLN GLY ALA
SEQRES  11 A  153  MET ASN LYS ALA LEU GLU LEU PHE ARG LYS ASP ILE ALA
SEQRES  12 A  153  ALA LYS TYR LYS GLU LEU GLY TYR GLN GL

ATOM      1  N   VAL A   1      -2.900  17.600  15.500  1.00  0.00           N
ATOM      2  CA  VAL A   1      -3.600  16.400  15.300  1.00  0.00           C
ATOM      3  C   VAL A   1      -3.000  15.300  16.200  1.00  0.00           C
ATOM      4  O   VAL A   1      -3.700  14.700  17.000  1.00  0.00           O
ATOM      5  CB  VAL A   1      -3.500  16.000  13.800  1.00  0.00           C
ATOM      6  CG1 VAL A   1      -2.100  15.700  13.300  1.00  0.00           C
ATOM      7  CG2 VAL A   1      -4.600  14.900  13.600  1.00  0.00           C
ATOM      8  N   LEU A   2      -1.700  15.100  16.500  1.00  0.00           N
ATOM      9  CA  LEU A   2      -0.900  14.100  17.200  1.00  0.00           C
ATOM     10  C   LEU A   2      -1.000  13.900  18.500  1.00  0.00           C
ATOM     11  O   LEU A   2      -0.900  14.900  19.000  1.00  0.00           O
```

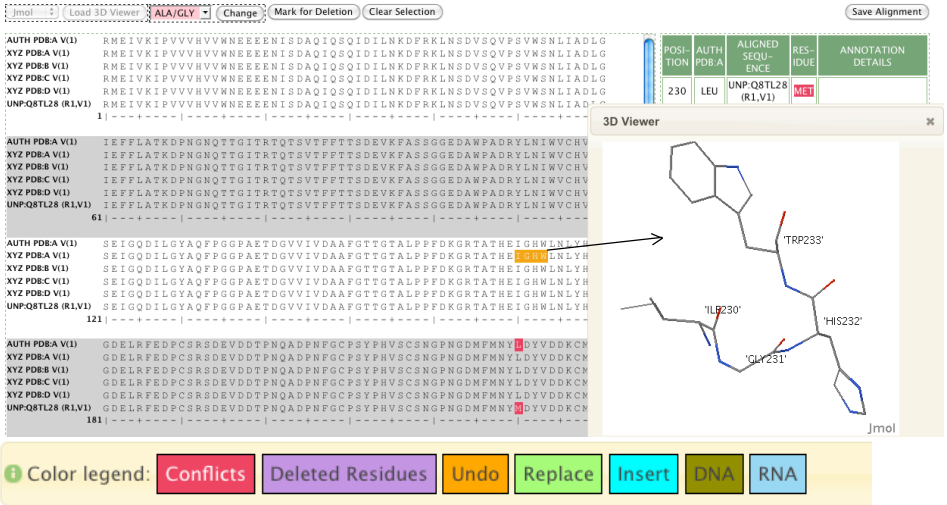
Taxonomy ⟷ Taxonomy

UniProt

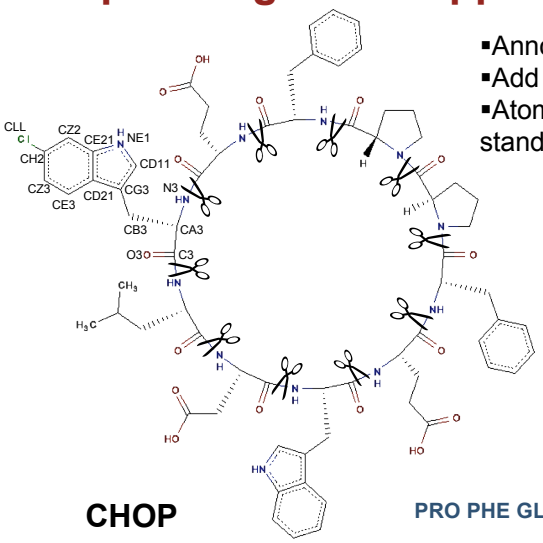
Sequence ⟷ Sequence

Atom-site records

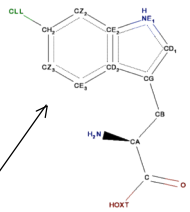
Annotator Integrated View



Peptide Ligand Chopper



- Annotator-directed bond breaks
- Add leaving groups (*e.g.* -OH, -H, -Cl)
- Atom naming and numbering standardized

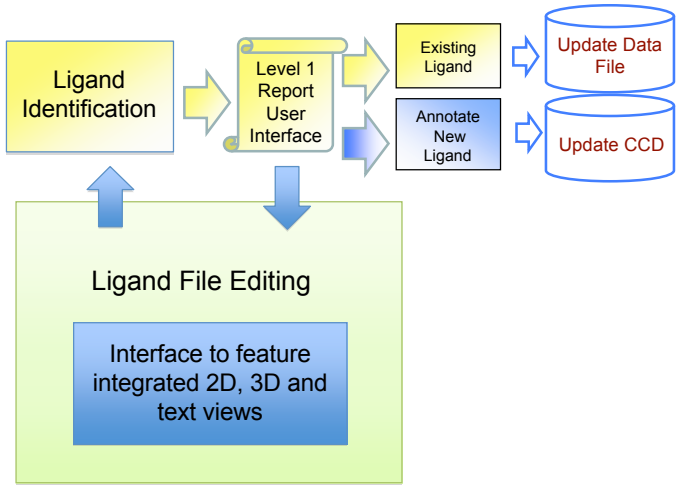


CHOP

PRO PHE GLU 6CW LEU ASP TRP GLU PHE DPR

Ligand Processing Module 09/24

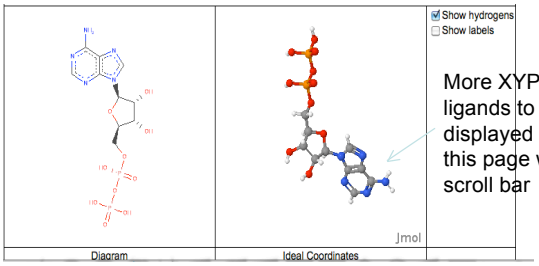
Phase 1: Simple Case. Fully-automated processing in testing



Ligand Editor Mock Up

Deposition id: D_012345

Ligand id: XYP_B_287
Name: [(2R,3S,4R,5R)-5-(6-aminopurin-9-yl)-3,4-dihydroxy-oxolan-2-yl]methyl phosphono hydrogen phosphate
Formula: C10 H15 N5 O10 P2
Formal Charge: 0



More XYP ligands to be displayed on this page with scroll bar

Undo | Save

ID	Instance	Status	Select
XYP	A503	CLOSE MATCH	☐
XYP	A504	NO MATCH	☐

Search results for Ligand instances

XYP_B_287	ID	Score (%)	Select for comparison
●	0AI	98	☐
●	1NA	97	☐
●	5AX	96	☐
●	A2G	96	☐

Create Ligand | Split/Merge | Run Search

Input new parameters here

Input your notes here

Common Tool Enhancements to Ligand Processing

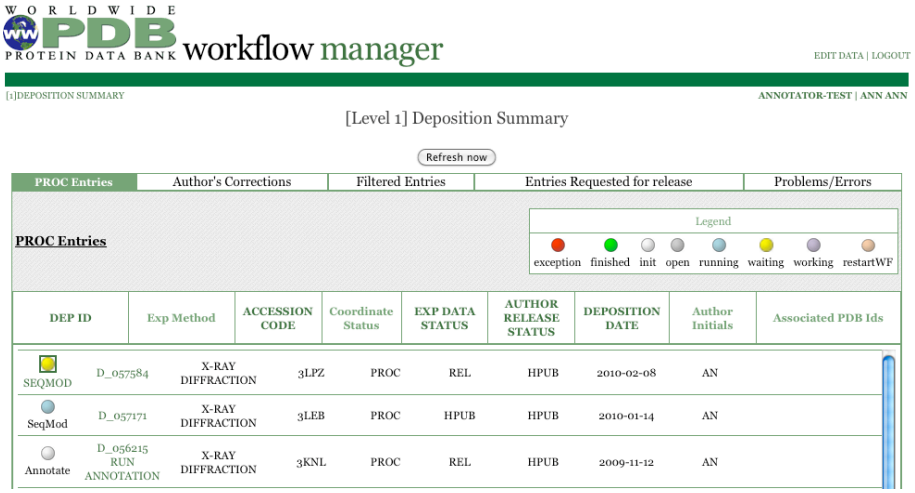
- Automated processing of “correct” existing ligands
- Better integration of process steps during annotation
- User interface to provide 2D, 3D and text views concurrently for ease of analysis
- Use of author-provided SMILES descriptor to facilitate ligand identification
- Provide ideal geometry reference for new and existing ligands

The Workflow Manager Interface

wwPDB annotators will access the new D&A workflow using the Workflow Manager interface

- Interface provides
  - Summary display of the active workflows
  - Processing status of each entry throughout the annotation process
- Action buttons
  - Launch tasks
  - Provide navigation to view details and browse output files produced by each task

Workflow Manager Example: Level 1



Deposition Interface

Goal To provide a depositor interface that improves data quality, processing efficiency and communication between the annotators and depositors

Process

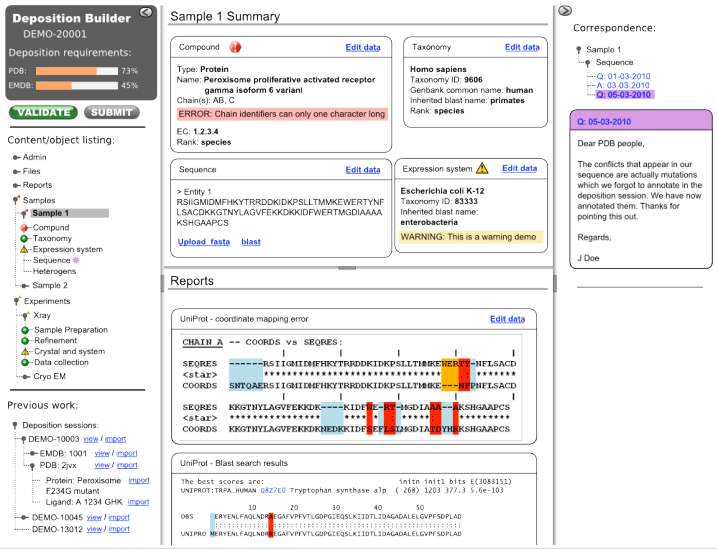
- Requirements – annotator and community driven
- Community input and feedback
  - Questionnaire distributed at ACA workshop
  - Mock-ups being prepared and community review planned

WORLDWIDE PDB PROTEIN DATA BANK
October 1, 2010

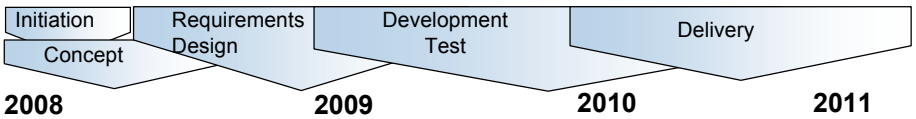
ACA 2010 - PDB Depositor Lunch

- 100 attendees
- Introduction of the D&A project goals
- Review of depositor interface questionnaire
- Answers to questionnaire itself

Deposition Interface Prototype



wwPDB Common D&A Tool Project Timeline



2008
- Concept
- Define deliverables
- Initial design
- Process definition
- Data model definition

2009
- Requirements elaboration
- Data flow documentation
- Technical Design
- Proof of Concept delivered

2010
- Sequence Module
- Ligand Chopper
- Ligand Module
- WF infrastructure
- Deposition Interface design
- Validation module in progress

2011
- D&A system delivery

Method- and Molecule-specific Activities

WORLDWIDE PDB PROTEIN DATA BANK
October 1, 2010

NMR

John Markley

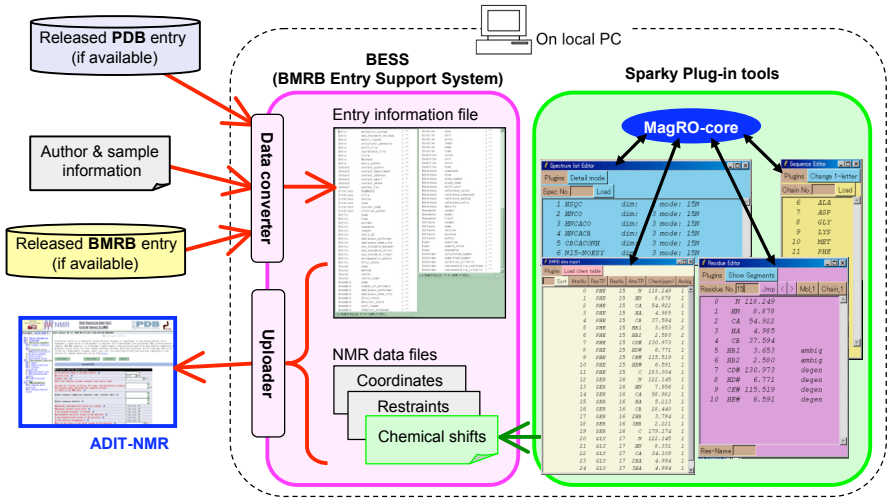
NMR Update: Topics To Be Covered

- Release of value-added files and software
- Validation reports
- Mandatory chemical shift deposition
- Small molecule structure deposition system (SMSDep)
- wwPDB NMR Validation Task Force (NMR-VTF)

Release of Value-added Files and Software

- PDBj-BMRB has processed ~300 assigned chemical shift entries in the past year from the RIKEN backlog
- Restraints have been released in the new NMR-STAR format with atom nomenclature consistent with coordinates
- Paper published on the NMR Restraints Grid (*J Biomol NMR* (2010) **45**: 389–396)
- Chemical shifts combined with restraints and coordinates are available from the BMRB FTP site; includes 5,341 entries organized by PDB ID in CCPN, CNS, CYANA & NMR-STAR formats

Development and Release of Support Tools for Depositors at BMRB-PDBj



Validation Reports

- CING coordinate and restraint validation reports are now available on the BMRB FTP site (collaboration with Jurgen Doreleijers)
- PDBe created a software tool for validating chemical shifts against a structure (*Proteins* (2010) **78**: 2482-2489)
- PDBe has published a paper on the analysis of chemical shifts and solvent accessibility (*BMC Structural Biology* (2009) **9**: 20)
- PDBe and BMRB run validation software on all NMR PDB entries

Mandatory Chemical Shift Deposition

- BMRB-developed web-service software validates atom nomenclature of chemical shift files against atom nomenclature of coordinate file
- ADIT-NMR for mandatory chemical shift depositions in beta testing
- BMRB members training annotators in the use of this software at RCSB PDB and PDBj
- RCSB PDB developed software to modify chemical shift files, if needed, to be consistent with coordinate files
- Minimal initial processing will be performed at deposition to check format and completeness and to substitute explicit atoms for pseudo-atoms and maintain nomenclature correspondence during annotation
- Targeted release of software (October 2010); implementation (December 2010)

WORLDWIDE PDB PROTEIN DATA BANK
October 1, 2010

Mandatory Chemical Shift Deposition

- CCPN now contains a software tool for validating chemical shift against coordinate file atom nomenclature; following testing, this tool will become available from the PDBe website
- Current AutoDep system accepts upload of CCPN projects that have been annotated with the ECI tool (*J Biomol NMR*, in press)
- Version of AutoDep for deposition of mandatory chemical shifts (as CCPN projects or uploaded files) in final testing at PDBe; uses the same chemical shift checking software used in ADIT-NMR
- Data files will be transferred to BMRB for further annotation
- PDB archive will contain chemical shift files in NMR-STAR format along with coordinate data files
- Download statistics for chemical shift files will be maintained for BMRB (needed for grant reporting)

Small Molecule Structure Deposition System (SMSDep)

- SMSDep was designed for depositing structures of molecular systems that fall outside the scope of the PDB
- SMSDep was developed at BMRB and is operational at PDBj-BMRB (annotation site)
- New PDB rules regarding acceptance of smaller peptides and nucleic acids need to be posted on the SMSDep website
- Current policy of accepting data only for small peptides or nucleic acids needs to be re-examined
- We plan to monitor the level of activity to determine whether this site should be maintained

wwPDB NMR Validation Task Force (NMR-VTF)

The NMR-VTF prepared an interim report along with a plan of action in December 2009

Next meeting is planned to be held at Rutgers in January 2011 following the NMR Keystone meeting

NMR-VTF Members

Gaetano Montelione (Co-Chair, Rutgers)
Michael Nilges (Co-Chair, Institut Pasteur)
Ad Bax (NIH)
Peter Guentert (University Frankfurt)
Torsten Herrmann (CNRS/ENS Lyon)
Jane Richardson (Duke University)
Charles Schwieters (NIH)
Wim Vranken (Free University Brussels) *
Geerten Vuister (Radboud University)
David Wishart (University of Alberta)

X-ray

Gerard Kleywegt

Remediation

Biological Assemblies

Problem

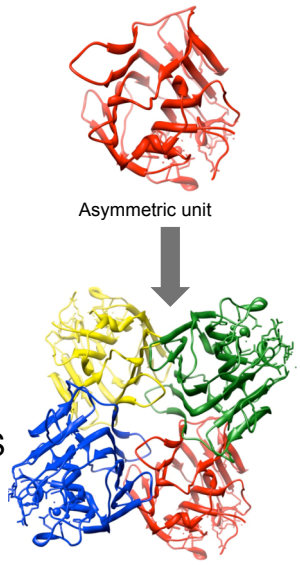
- Inconsistent and missing annotation of biological assemblies

Approach

- Compared manually curated PQS assemblies with PISA-generated assemblies
- Prefer PQS in case of discrepancies

Result

- 5800 entries updated with PISA or PQS annotation



Residual B Values

Problem

- PDB ATOM records for 7629 entries refined using TLS with REFMAC (since 2000) may contain full or residual isotropic B-values

Approach

- Back-calculation of new isotropic B-values, and comparison of refinement statistics before and after correction
- Improved statistics and closer reproduction of reported statistics used to assign full or residual B-value

Result

- 6296 entries *LIKELY* to contain residual B-values
- 154 entries determined to contain full B-values, confirmed by other information in the deposited entry
- 1179 entries require further analysis

Residual B Values – Format Details

Remediated data files for the 6296 entries identified as *likely* containing residual B-values will include the following information:

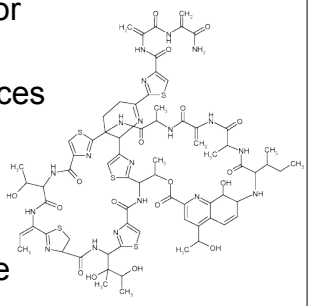
- PDB format

```
REMARK   3   B VALUES
REMARK   3   B VALUE TYPE : LIKELY RESIDUAL
```

- PDBx/mmCIF and PDBML
  In the REFINE category, a new item PDBX_ADP_TYPE will be added and assigned the value 'LIKELY RESIDUAL'

Peptide Inhibitors and Antibiotics: Challenges

- Non-standard amino acid, nucleotides or other chemical groups in sequence
- Non-linear (cyclic or branched) sequences
- Microheterogeneity
- Non-uniform annotation of the same molecule in different PDB entries
- Lack of annotation regarding the source and function of these molecules
- 300 antibiotics; 420 single component/450 polymeric peptide inhibitors:



Thiostrepton

Peptide Inhibitors and Antibiotics: Solutions

Analysis and classification

- Identify antibiotics and inhibitors and group them into polymeric molecules or single components

Dictionary updates

- Build single chemical components for appropriate cases
- Enrich dictionary with source, function and other details

Remediation and future processing

- Revise coordinate files to present chemistry in either sequence or small-molecule form
- Create a Peptide Reference Dictionary (PRD)
- Establish rules and procedures to make new annotations consistent

Peptide Inhibitors and Antibiotics: Status

- Inhibitor annotation completed
- Antibiotics nearing completion
- Annotation guideline documentation completed
- Annotator training on-going
- Load testing to be done at all sites
- To be released January 2011

Peptide Reference Dictionary (PRD)

An information resource about peptide inhibitors and antibiotics:

- Provides help in PDB data processing
- General resource for the community
- Sequence, chemical representation, source, physical, chemical, and functional information
- Links to CAS, KEGG, ChEBI, Norine, UniProt, *etc*.
- Functions extracted from these resources as well as from primary citations
- mmCIF files have been created for PRD and are being checked

wwPDB X-ray Validation Task Force

Initial meeting

- April 14-16, 2008 EBI, Hinxton, UK
- R. Read (Chair), P. Adams, A. Brunger, P. Emsley, R. Joosten, G. Kleywegt, E. Krissinel, T. Lütteke, Z. Otwinowski, A. Perrakis, J. Richardson, W. Sheffler, J. Smith, I. Tickle, G. Vriend

Goal

- Gather recommendations and consensus on validation for PDB entries, and identify software applications for these validation tasks
- Provide code/algorithms for the validation-software pipeline

Preliminary outcome

- Candidate global and local validation measures were identified
- These measures were reviewed in terms of the requirements of depositors, reviewers, and users

X-ray VTF Update – Randy Read

New Format

- PDB format defined in 1970s
  - FORTRAN (column-oriented)
  - “Small” molecules
- Limitations
  - Max 62 chains (and that’s stretching it)
  - Max 99,999 atoms (2 ribosomes in ASU? 4 PDB entries!)
  - No bond orders specified for ligands
  - Meta-data specification cumbersome and inflexible

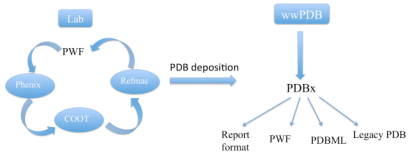
New Format

- wwPDB archival/exchange format is PDBx
  - 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
    - Refinement, model-building, graphics, validation, …
  - Define new "human-readable report" content and format for core meta-data

WORLDWIDE PDB PROTEIN DATA BANK
October 1, 2010

New Format

- PDB Working Format - PWF
  - Preserve simple style and readability of PDB format
  - Provide extensible framework for capturing larger systems and information from multiple experimental methods
  - Allow for custom extensions
  - In other words: combine best of both worlds
  - Surprisingly little blood was spilt in the discussions!
  - First reactions from X-ray developers very enthusiastic!

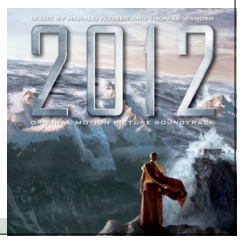


New Format

- The coordinate content in PWF is pre-defined, has a certain order and provides extensibility to add new items
- The new format will involve minor schema changes to exchange dictionary
  - Addition of new identifiers to handle multiple experimental methods
  - Rename “asym” category so that it is biology-centric instead of crystallography-centric
  - Generalized “group” concept (for TLS, NCS, sites, …)
- The “ATOM records” in PWF will use white space separators and have non-blank values for each field
- Residue names need to be wide enough (10 characters?) to accommodate carbohydrate nomenclature and variants

New Format Proposed Timeline

- First draft of well-defined PWF specification June 2010
- Bring in key software developers in Q4 2010/Q1 2011
  - Coot, Phenix, CNS, Refmac, Buster, Shelx, CCP4
  - ARIA, CYANA, UNIO, XPLOR-NIH
  - Visualization, computational biology, bioinformatics, commercial
- Finalize format specification V1, Q2 2011
- Public request for comments, test files, Q3/4 2011
- Final V2 specification and new files on ftp, Q1 2012
- 2012? – *Formageddon*
  - Start accepting new format (common tool)
  - Freeze PDB format
  - “Best-effort” PDB files from this day on!
- 2014? – stop distributing PDB files
- 2015? – stop accepting PDB files



SAXS/SANS

Gerard Kleywegt

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. Cα 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

Members

- Jill Trewhella (University of Sydney)
- Dmitri Svergun (EMBL Hamburg)
- Andrej Sali (UCSF)
- Mamoru Sato (Yokohama City University)
- John Tainer (Scripps)

Meeting will be Q2 2011; Report Q1 2012

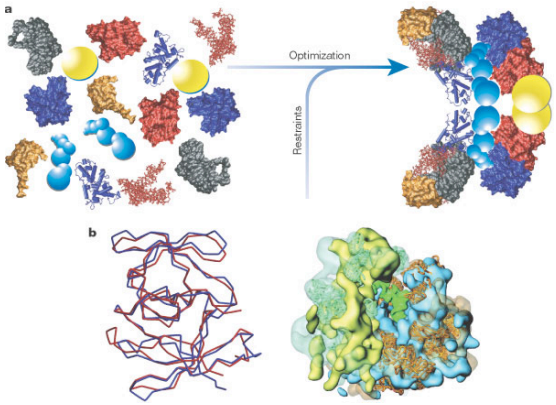
WORLDWIDE PDB PROTEIN DATA BANK
October 1, 2010

SAXS/SANS Task Force

Questions:

- Should the PDB archive SAS models?
- If "*yes*", then
  - Which types of models (and which not)?
  - Minimum requirements?
  - Minimum supporting experimental data?
  - Validation procedures?
    - Models, data, model vs. data

Hybrid Methods: Task Force Will Be Established 2011



Electron Microscopy

Helen Berman

Electron Microscopy

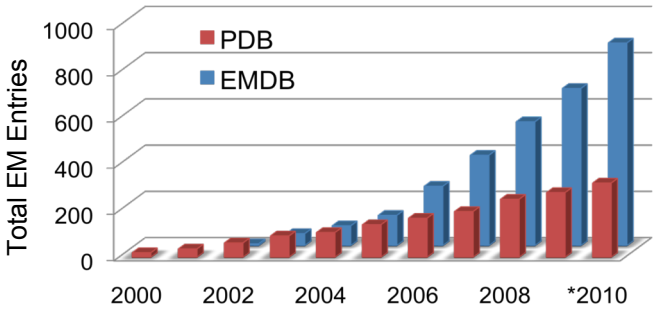
- Collaborative project between RCSB PDB, PDBe, and Baylor-NCMI is funded by NIH, BBSRC, and EMBL
- Unified tool for collecting model coordinates and map files in a *one-stop shop*
- Merge deposition and annotation with PDB as part of Common D&A Tool by 2011

EMDataBank.org

- **Joint map + coordinate deposition service**
- EMDataBank.org: news, EM software list, information about dictionaries, conventions, FAQ page, community links
- EMSEARCH: search by ID, author, sample type, keyword, deposition date
- EMViewer: simple map viewer

EM Annotation

- Remediation completed
- Requirements for EM annotation in Common Tool being developed
- Letters sent to journals about deposition requirements



*1 Sept 2010:
888 EMDB map entries
328 PDB coordinate entries

Electron Microscopy Validation Task Force

- Community input on modeling criteria for EM maps
  - First meeting January 2010, University of Houston
- First EM Validation Task Force meeting
  - September 28-29, 2010, Rutgers University

EM Map Discussion Group

- Richard Henderson (Map Chair, MRC-LMB)
- Bridget Carragher (Scripps)
- Kenneth Downing (LBL)
- Edward Egelman (U Virginia)
- Joachim Frank (Columbia)
- Niko Grigorieff (Brandeis)
- Wen Jiang (Purdue)
- Steven Ludtke (Baylor)
- Ohad Medalia (Ben-Gurion University)
- Pawel A. Penczek (UT Houston Medical School)
- Michael Schmid (Baylor)
- Alasdair Steven (NIAMSD)

EM Models Discussion Group

- Andrej Sali (Models Chair, UCSF)
- Peter Rosenthal (National Institute for Medical Research)
- Michael G. Rossmann (Purdue)
- Gunnar Schroeder (Forschungszentrum Juelich)
- Willy Wriggers (DE Shaw)

Charge to the Committee

- Advise EMDataBank.org on approaches to validate maps and models obtained from cryoEM data
- Recommend specific EM structure validation criteria and tools – can be based on existing or proposed software

Questions – Maps

1.	How should map **accuracy** be assessed?
2.	How should map **resolution** be assessed?
3.	What map **density manipulation/filtering procedures** should be specified for deposited maps? Should any procedures be disallowed?
4.	Would it be desirable to have a tool to validate map **point group/helical symmetry** and to define orientation and position?
5.	What parameters should be used to indicate reconstruction quality in **3D tomogram** and **sub-tomogram** averaged maps?

Questions – Models

1.	What **kind of structural models** do we expect based on EM maps?
2.	What general criteria should be used to describe **model quality**?
3.	What can we learn from other assessment efforts and should we strive for a **common language and framework**?
4.	How should the **fit of an atomic model into an EM map** be evaluated? (local vs. global)
5.	How can we **evaluate that a fitted model is the correct one** or that the solution is unique/optimal?
6.	How should we **"value"** stereochemistry/geometry when applied to EM models?

More Questions – Models

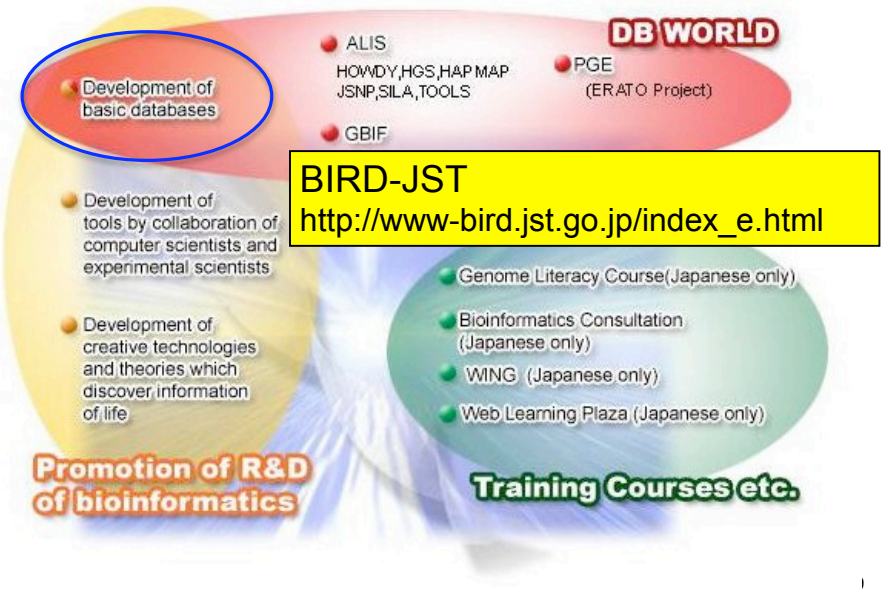
7.	How should PDB handle: **models deposited with wrong sequence(s), models deposited based on homology-modelling**?
8.	How are **errors in EM maps and errors in models coupled**? How to take into account the quality of maps when estimating the quality of models based on these maps?

Foundation of National Database Center for Bioscience in Japan and Role of PDBj

Haruki Nakamura

Brief History for Foundation of National Database Center for Bioscience in Japan

- 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 Database Center for Bioscience (temporary name) in Japan will be founded



WORLDWIDE PDB PROTEIN DATA BANK
October 1, 2010

Integrated Database Project and DBCLS (Database Center for Life Science)

The “Integrated Database Project” is a five year project (starting FY2006), funded by the Japanese MEXT, that aims to improve the accessibility and usability of life science databases in Japan.

It is lead by DBCLS/ROIS (Research Organization of Information and Systems) together with 15 other Japanese universities and institutions. Portal websites, search systems and tools useful for life science research have been developed and maintained, as part of the Project. In addition, the Project addresses issues such as how data should be shared within the research community.

WORLDWIDE PDB PROTEIN DATA BANK

October 1, 2010