

RCSB Protein Data Bank: Overview

Helen M. Berman

July 24, 2009

Vision

To provide a global resource for the advancement of research and education in biology and medicine by curating, integrating, and disseminating biological macromolecular structural information in the context of function, biological processes, evolution, pathways and disease states.

We will implement standards, and anticipate and develop appropriate technologies to support evolving science.

Structural Views of Biology and Medicine



Mission

Support a resource that is by, for, and of the community by providing

- leadership in the representation of biological structures derived via experimental methods
- data in an accurate and timely manner
- comprehensive, integrated view and unique views of the data

so as to enable scientific innovation and education

What is the PDB?

- Single international repository for all information about the structure of large biological molecules
- Archival database with hundreds of thousands of users who depend on the data

Archive Contents

- **Public archive**
 - More than 400,000 files (as of June, 2009)
 - Requires over 93 GBbytes of storage
 - Data dictionaries
 - Derived data files
- **Internal archive**
 - Depositor correspondence
 - Depositor contact information
 - Paper records
 - Documentation
 - Historical records from Day One
- **For each entry**
 - Atomic coordinates
 - Sequence information
 - Description of structure
 - Experimental data
 - Release status information

History of the PDB

- 1970s**
 - Community discusses how to establish a protein structure archive
 - Cold Spring Harbor meeting in protein crystallography
 - PDB established at Brookhaven (Oct 1971; 7 structures)
- 1980s**
 - Number of structures increases as technology improves
 - Community discussions about requiring depositions
 - IUCr guidelines established
 - Number of structures deposited increases
- 1990s**
 - Structural genomics begins
 - PDB moves to RCSB PDB
- 2000s**
 - wwPDB formed
 - 50,000th structure released (April 2008)
 - **2nd renewal of RCSB PDB management**

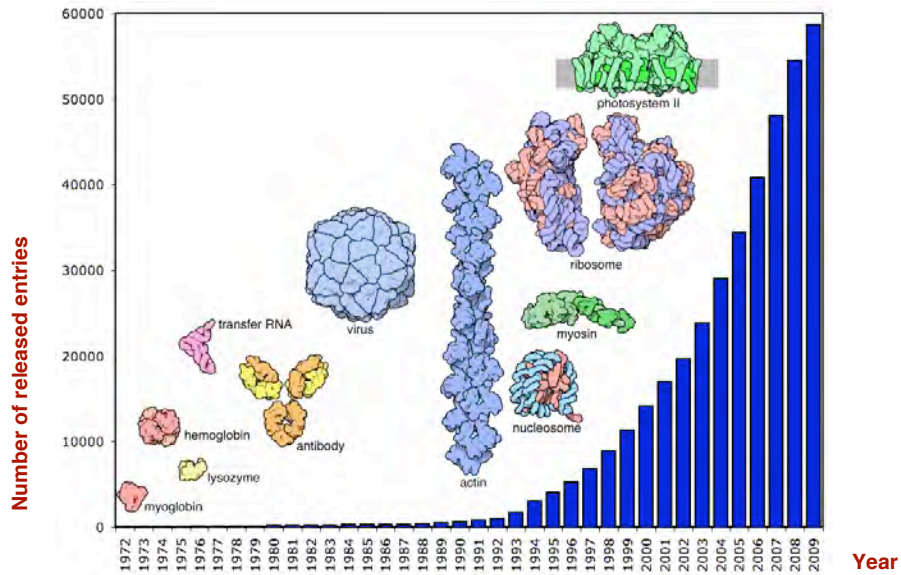
Progress

1998

- 9000 structures in archive
- 2800 depositions per year
- all depositions to US site
- <10 million FTPdownloads/year
- no clear standards
- limited annotations
- no integration with databases
- minimal harvesting
- minimal database

2009

- >57000 structures in archive
- >8000 depositions per year
- global distribution of depositions wwPDB)
- almost 300 million FTP downloads/year
- well-articulated standards
- extensive annotations
- integration with many databases
- 30% input through harvesting
- database supporting query, browsing, and visualization

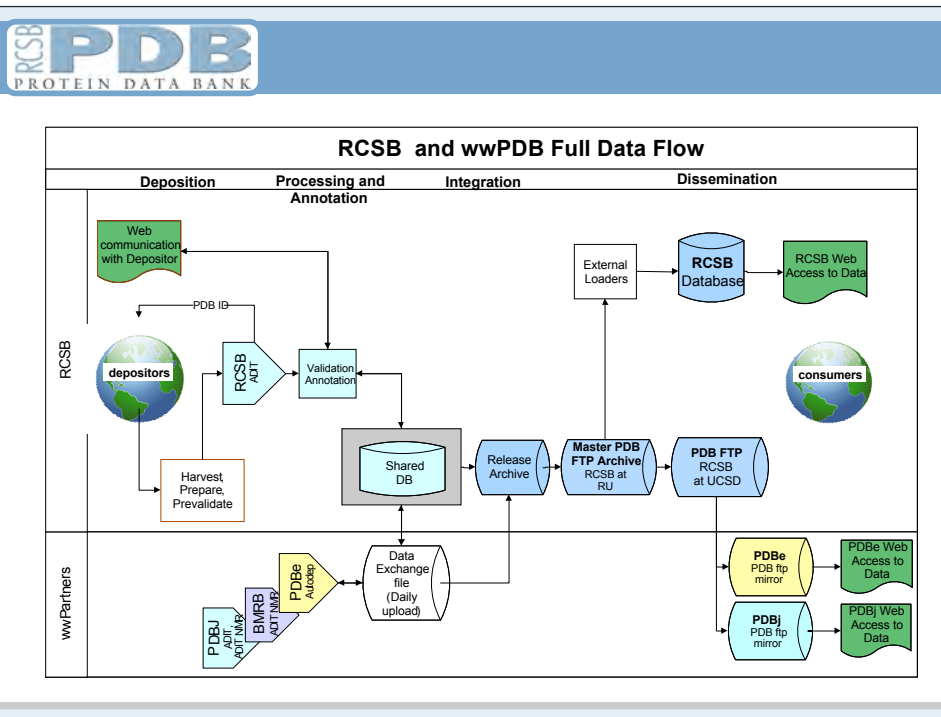


Challenges

- More structures
 - Larger
 - More complex
 - New methods
- Increased demand
 - More complex queries
 - More annotation
- Integration with other genomic and proteomic resources
- Ever-growing and diversifying user community

Major RCSB PDB Activities

- PDB grant renewal
- One-stop shops for cryo-EM and NMR
- Common wwPDB Deposition & Annotation Tool
- Remediation rollout
- New website
- Ongoing outreach



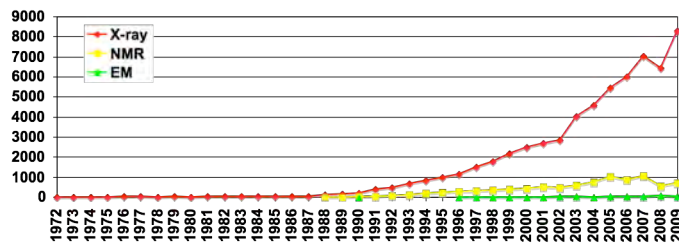
PDB Depositions

Last Updated: 14 Jul 2009

By deposition and processing site

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	4407	3597	250	560	2759	1088	560
TOTAL	53740	39271	6709	7760	32707	13273	7760

By experimental type



PDB Depositors (1999-2009)



One-stop shops



emdatabank.org



USA: <http://deposit.bmrbl.wisc.edu/bmrbl-adit/>
Japan: <http://nmradit.protein.osaka-u.ac.jp/bmrbl-adit/>

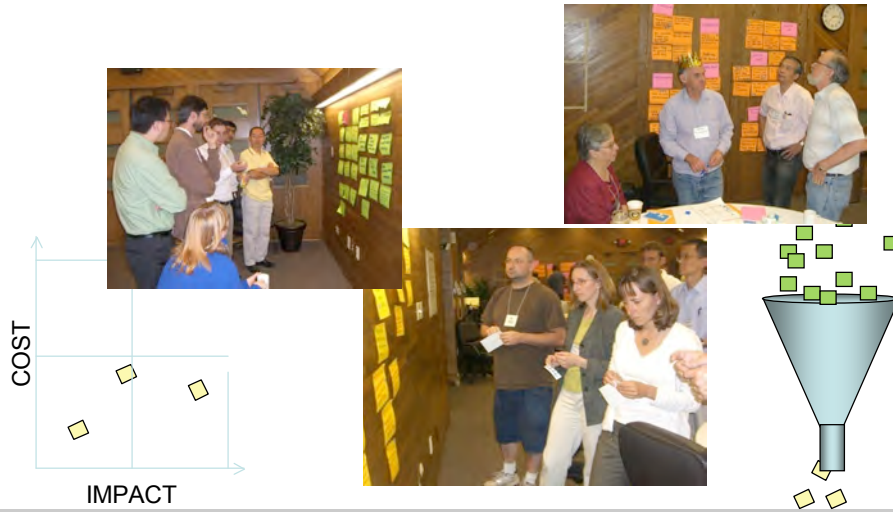


Common Tool for Deposition and Annotation

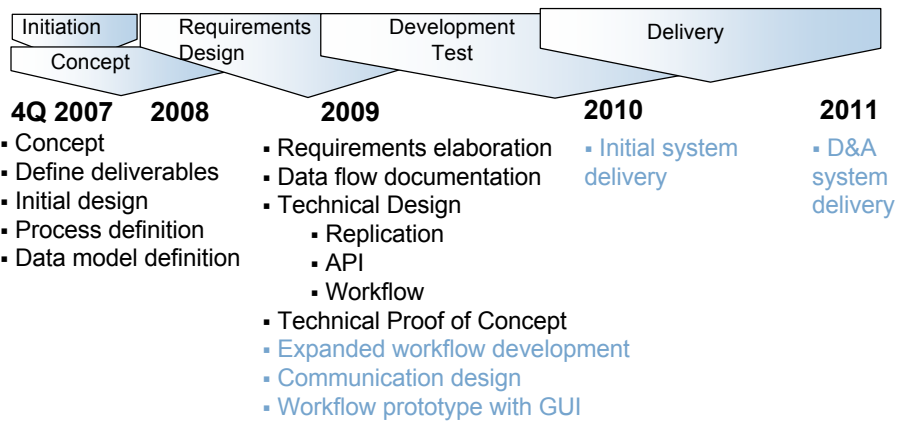
Manage increased data load without an increase in resources

- Create global deposition and annotation tools
- Distribute worldwide data load and eliminate individual points of failure
- Anticipate new developments in structural biology to keep tools up to date

Planning for a Common Tool



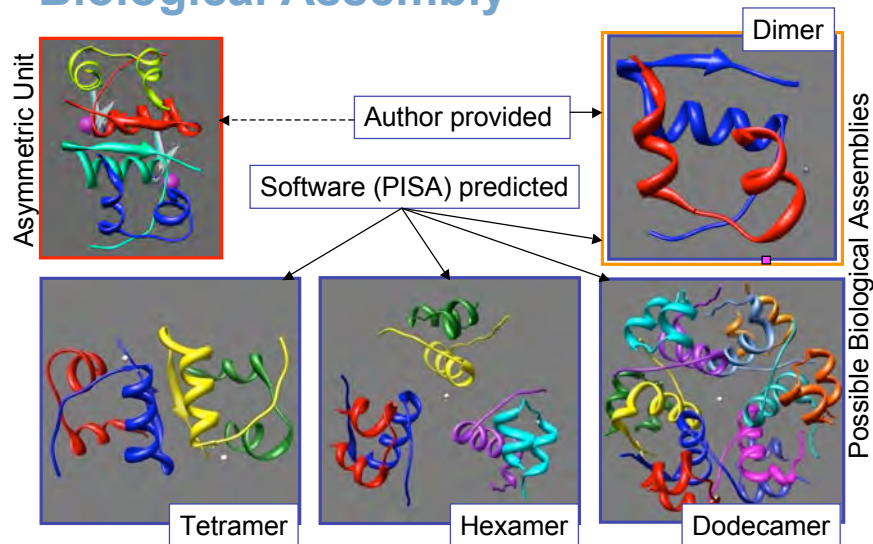
wwPDB Common D&A Tool Project Timeline



Remediation of PDB archive

- PDB format version 3.15 files released March 2009
- New records/remarks: to describe models, ligands and zero occupancy atoms/residues.
- Enhancements: assemblies, SITE records, chain IDs, taxonomy references (including PubMed, DOI IDs) added for all files
- Various corrections (including sequence, beamline, wavelength, atom connectivities, atom nomenclature, mmCIF consistency)
- Missing REMARKs restored to appropriate REMARK numbers

Biological Assembly



Validation Task Forces

X-ray VTF

- April 14-16, 2008 at EBI-EMBL, Hinxton
- Collected recommendations and developed consensus on additional validation that should be performed on PDB entries, and to identify software applications to perform validation tasks.



Chair: Randy Read; Sponsors: PDBe & RCSB PDB

NMR VTF

- Initial meeting will be held September 21, 2009 in Paris, France

Common Tool D&A and Remediation are wwPDB collaborative projects

Data Out

- Rollout of new site
- Record number of hits
- Snapshots
 - Yearly snapshot for 2009
 - March 16, 2009 archive of unremediated data
- New website features

PDB FTP Traffic (May 2008 - April 2009)



■ RCSB PDB
180546882
data downloads

■ PDBe
40650943
data downloads

■ PDBj
14506857
data downloads

RCSB PDB
PROTEIN DATA BANK

New Look & Feel

The screenshot shows the RCSB PDB website with a search bar at the top. Below the search bar, there are several features highlighted:

- Extended Top Bar Search:** A search bar with a dropdown menu showing options like "PDB ID or keyword", "Author (Structure-Citation)", "Structural Genomics Center", "Chemical Name", and "Chemical ID".
- Dedicated Print Options:** A section titled "Print Options:" with icons for a printer, a globe, and a document.
- Browser Compatibility Check:** A section titled "Browser Information: Mozilla Firefox 3.0.11" with a table showing the status of various browser features.

Description	Status	Details	Notes
JavaScript	✓ Pass	JavaScript is enabled in your browser.	
Cookies	✓ Pass	Cookies are enabled in your browser.	
Java	✓ Pass	Java version 1.5.0_16	
Popup windows	✓ Pass	Popup windows are enabled in your browser.	

RCSB PDB
PROTEIN DATA BANK

Outreach & Education Goals

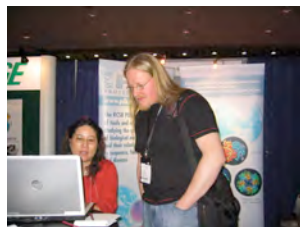
Inform and Educate

- RCSB PDB resource should meet its mission in the interest of science, medicine and education
- RCSB PDB is defined by, designed for, and owned by the communities it serves

User Communities

- **Biologists** (in fields such as structural biology, biochemistry, genetics, pharmacology)
- **Other scientists** (in fields such as bioinformatics, software developers for data analysis and visualization)
- **Students and Educators** (all levels)
- **Media** writers, illustrators, textbook authors
- **General public**

Inform



Interactions and workshops at scientific meetings

Electronic news, RSS feeds, support pages, tutorials, listserv

What's new in this release (May 26, 2009)

Changes

- New Look and Feel
- Improved Top Menu Search
- Customizable Structure Pages
- Improved Visualization Options on the Structure Summary Page
- Detailed Data View
- Literature View
- Sequence Structure View
- Visualization of Protein-Ligand Interactions
- Experimental Features: Residues in Helix Kink Viewer
- PDB Statistics

New Look and Feel

This version introduces a traditional layout scheme for the RCSB PDB that incorporates new colors and improved website features (described below). Icons are used to indicate links to help pages (H), resources (R), database searches (S), and external sites (E). Holding your mouse over a search field or link will now display more information.

Improved Top Menu Search

The Top Bar Search now includes Structural Genomics Center, Chemical Name and Chemical ID search options. Click on the "H" next to the site search for details about the search options.

RCSB is a member of the Protein Data Bank Consortium.

Publications



Educate

- **Scientists**

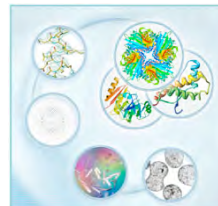
- Crystallography for Modelers (May 2009)

- **Graduate students**

- Biophysical Chemistry (2002, 2004, 2006, 2008)
- Structural Bioinformatics (2000 - 2008)
- Pharmacy Informatics (2002 - 2006, 2009)

- **Undergraduate students**

- Molecular View of Human Anatomy (2006, 2008)
- Dorothy and Linus (2000, 2002)
- Paths to and from the Double Helix (1998)



Management and Oversight

- Director, Helen M. Berman
 - Overall direction of RCSB PDB
 - Direction of Rutgers site
- Deputy Director, Martha Quesada
 - Coordination of all projects across the RCSB PDB
 - Facilitation of wwPDB initiatives
- Associate Director, Philip E. Bourne
 - Direction of UCSD site
- PDBAC and wwPDBAC
 - Stephen K. Burley, Chair

RCSB PDB 2008



Agenda

- Introduction & Overview Helen Berman
- Outreach Shuchismita Dutta
- Data In Jasmine Young
- Common Tools Martha Quesada
- Data Out Phil Bourne
- Executive Session
- General Discussion

Outreach and Education



Shuchismita Dutta, Christine Zardecki

July 24, 2009

Outreach & Education Goals

- RCSB PDB resource should meet its mission in the interest of science, medicine and education
- RCSB PDB is defined by, designed for, and owned by the communities it serves

International User Communities

- **Biologists** (in fields such as structural biology, biochemistry, genetics, pharmacology)
- **Other scientists** (in fields such as bioinformatics, software developers for data analysis and visualization)
- **Students and Educators** (all levels)
- **Media** writers, illustrators, textbook authors
- **General public**

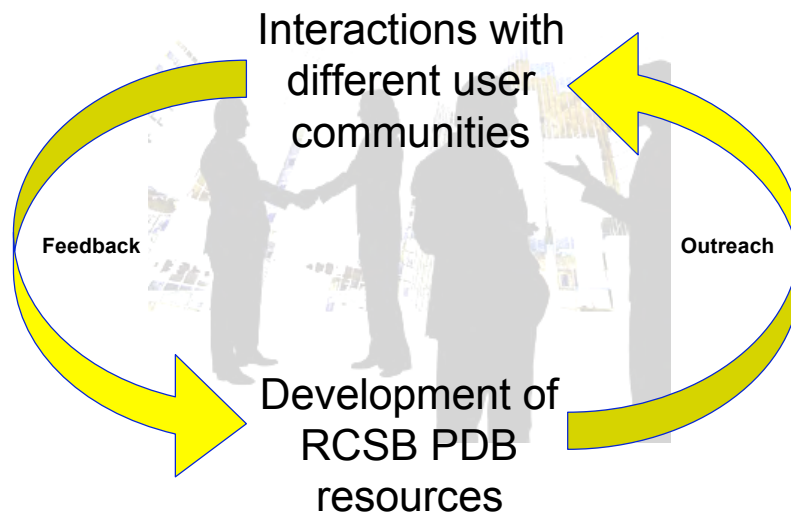
International Collaborations

- **Organizations and Advisors**
 - wwPDB
 - Advisory Committees
 - Task Forces
- **Standards and Policies**
 - Models Workshop (Nov 2005)
 - Validation Workshop (Apr 2008)
- **Resources**
 - ADIT-NMR (BMRB, RCSB PDB)
 - Joint PDB-EMDB deposition (RCSB PDB, MSD-EBI, C²BC)



wwPDB AC 2008

The Outreach Cycle



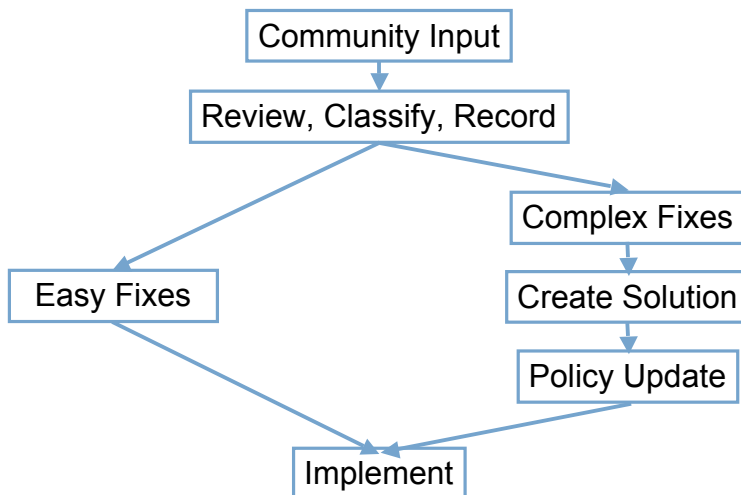
Community Interactions

- Electronic help desks, discussion groups
- Demonstrations at professional meetings
- Personal interactions
- Exhibit booths
- Workshops
- Posters



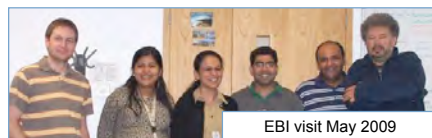
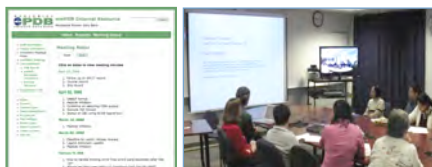
Biophysical Society Meeting, 2009

Feedback Evaluation & Response



wwPDB Internal Communications

- Intranets
- Regular VTCs
- Email
- Phone
- In-person meetings
 - Retreats, D&A
 - Development, visits



Publication Collaborations

- Structural data are released quickly and accurately when the journal articles are published
- Digital Object Identifiers (DOIs) available for released entries in the PDB archive
– <http://dx.doi.org/10.2210/pdb4hhb/pdb>



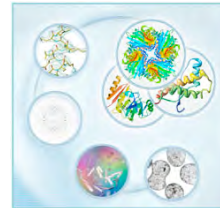
Inform and Educate

- International scientific meetings and workshops
- Electronic news, RSS feeds, support pages, tutorials, listserv
- Printed publications (annual report, newsletter, flyers, brochures)



Inform and Educate

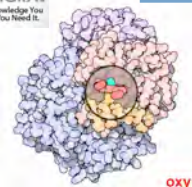
- **Scientists**
 - Crystallography for Modelers (May 2009)
- **Graduate students**
 - Biophysical Chemistry (2002, 2004, 2006, 2008)
 - Structural Bioinformatics (2000 - 2008)
 - Pharmacy Informatics (2002 - 2006, 2009)
- **Undergraduate students**
 - Molecular View of Human Anatomy (2006, 2008)
 - Dorothy and Linus (2000, 2002)
 - Paths to and from the Double Helix (1998)



Educational Resources

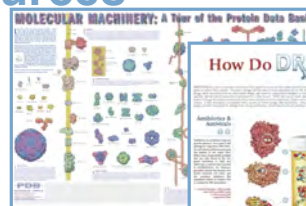


Tutorials

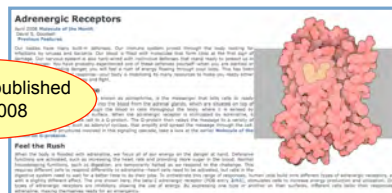


Molecule of the Month

100th feature published in April 2008



Posters



looking at structures



An Online Resource for Learning About PDB Data

Using text, images, and interactive Jmols, *Looking at Structures* intends to help researchers and educators get the most out of the PDB archive.

Where are all the hydrogen atoms in this file? Should I care about the R-factor? Why are there 20 overlapped structures in my file?

Looking at Structures

- Introduction
- Biological Units
- Dealing with Coordinates
- Methods for Determining Structure
- Missing Coordinates and Biological Units
- Molecular Graphics Programs
- Resolution
- R-value and R-free

Inspire: Activities, Visits, Workshops

▪ Teachers

- NJ Science Convention (2005 -2009)
- NJ Science Olympiad coaches workshop (2005, 2006)
- Student-Centered Education Meeting, ASBMB (2009)

▪ K-12 students

- NJSO (2006 - 2009)
- Princeton Science Expo (2006 - 2009)
- Various school visits

▪ General public

- Rutgers Day
- San Diego Science Festival



Involve and Inspire Teachers



Discussing protein structure,
Celebration of Teaching and
Learning, 2008



Building 3D viruses
New Jersey Science Convention, 2008

Involve and Inspire Students



Science and Engineering Expo, 2009



National Science Olympiad
champions, 2009 (from NJ!)

RCSB PDB poster prize awarded
to students at scientific meetings



ISMB 2007



IUCr 2008



High school visit to RCSB PDB, May 2009

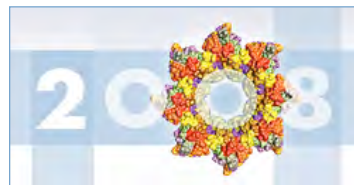
Involve and Inspire General Public



San Diego Science Festival, 2009



Birch Aquarium
2006-2007



Calendar



Kiosk Viewer: Online, Rutgers, UCSD

Molecules in Motion Kiosk Viewer

The Molecules in Motion Kiosk Viewer is a full-screen animation program that displays structures from different angles and perspectives, and focuses on chemical components within the structure.

The kiosk program runs on Mac, Windows and some versions of Linux (i.e. CentOS 5) and requires the latest version of Java. The program automatically downloads coordinate files into a folder which has access from a kiosk on an office computer. To customize the list of structures displayed in kiosk, right-click on the kiosk viewer file and save the file with a new name. Click the zip icon, go to the folder, click on the file, click the PDB IDs listed in the file and save it. Double click to launch Kiosk Viewer, press the Esc key to exit kiosk viewer.

Launch Kiosk by providing a comma separated list of PDB IDs. (Java required)

1PVS:1PVA:4HNS:1UNQ:1Q7A:1Q7A:1Q7B

By default kiosk will display biological units (if available). Uncheck here to view asymmetric units only.

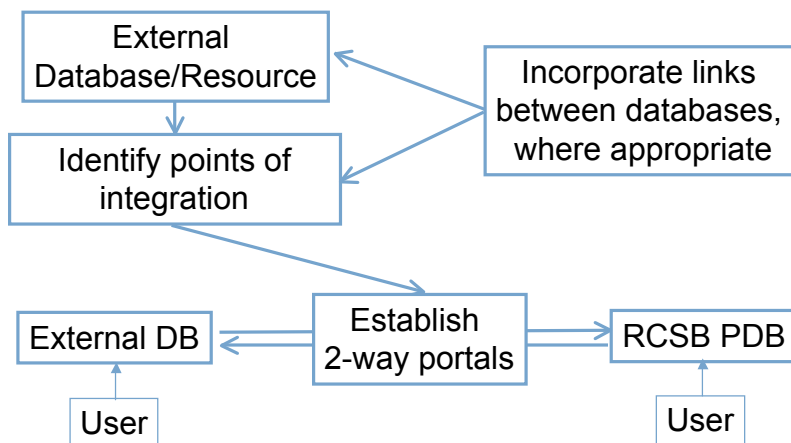
Search

Future: Community Specific Resources

- For the Scientific community
 - Structural Views of Biology and Medicine
 - Pharmacological view
 - Model organism view
- For Students and Teachers
 - Learn in 3D: A comprehensive module for promoting biological macromolecular structural literacy in undergraduates (CCLI 2009 - NSF proposal submitted)
 - Road Shows
 - Rules, Tools and Stories
- For General Audiences
 - Second Life (a new medium for promoting structural awareness)



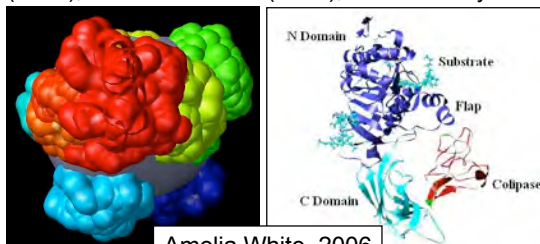
Enabling Structural Views



Molecular View of Human Anatomy

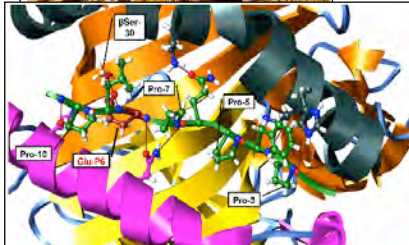
Curriculum:

- Introduction to biological molecules proteins
- Chemical bonds and interactions in molecules. Primary, secondary, tertiary and quaternary structures of proteins
- Introduction to visualization
- Introduction to the RCSB PDB
- Introduction to MAP
- Exploring Human Digestion (2006), HIV and Cancer (2008), Endocrine system (2008)



Amelia White, 2006

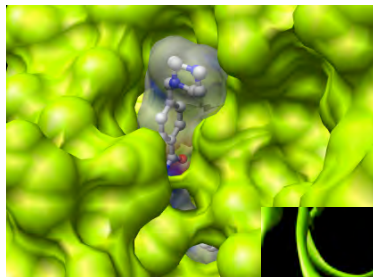
Celiac disease in MAP



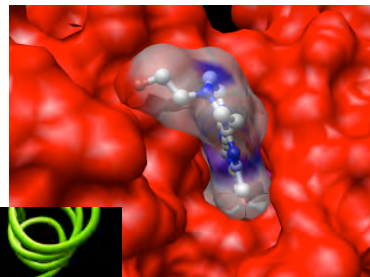
- 1 Proline rich fragments of gluten that are resistant to processing by luminal and brush-boarder enzymes survive digestion and are transported across the mucosal epithelium as polypeptides.
- 2 Gluten proteins deamidated by **tissue transglutaminase (TG2)**
- 3 **HLA-DQ2** or **HLA-DQ8** molecules on the surface of antigen presenting cells (APC) present the deamidated gluten proteins.
- 4 **CD4⁺ T cells** recognize deamidated gluten proteins
- 5 CD4⁺ T cells produce the cytokine **interferon (IFN)- γ** upon recognizing deamidated gluten proteins
- 6 Cytokine **Interleukin 15**, when stimulated by gluten causes T cells to migrate to epithelium and facilitate killing of enterocytes (epithelial cells of small intestine).
- 7 Gluten induces production of intestinal peptide **Zonulin**, which increases epithelial permeability

Basis of Celiac disease (class project 2006)

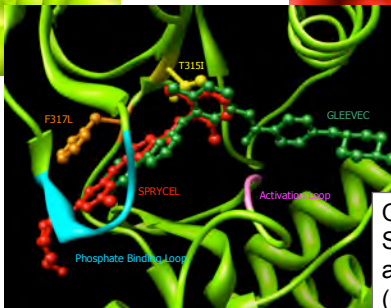
Treating Chronic Myeloid Leukemia



Gleevec
bound to Abl Kinase
PDB ID 1iep

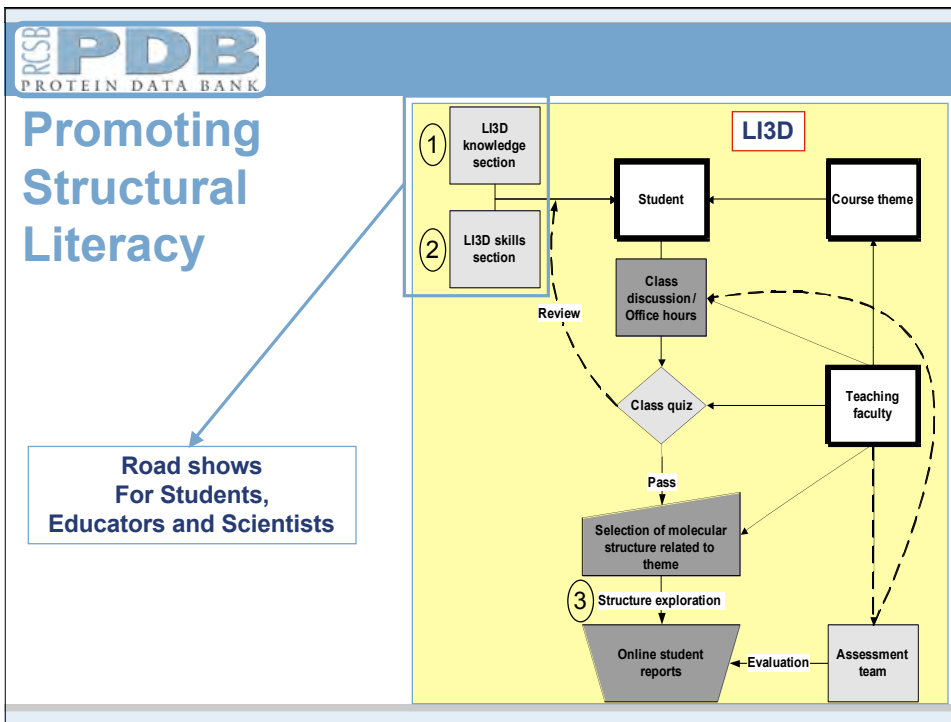


Sprycel
bound to Abl Kinase
PDB ID 2gqg



Mira Patel
Student 2008, 2009

Overlap of Gleevec &
Sprycel bound to
abl kinase
(PDB IDs 1iep, 2gqg)



Data In: Deposition, Annotation and Remediation

Jasmine Young

July 24, 2009

- **Deposition and annotation**
 - PDB data depositors and users
 - Data flow
 - Deposition statistics
- **Completed projects**
 - PDB format documentation version 3.2
 - Rollout of remediation files
- **Ongoing projects**
 - Improved tools and uniform data curation
 - Remediation and curation of files with complex chemistry
 - Peptide reference dictionary
 - Additional X-ray validation measures
 - NMR: Implementation of chemical shifts
 - EM maps
 - wwPDB common deposition and annotation tool

Deposition and Annotation

- PDB data depositors and users
- Data flow
- Deposition statistics

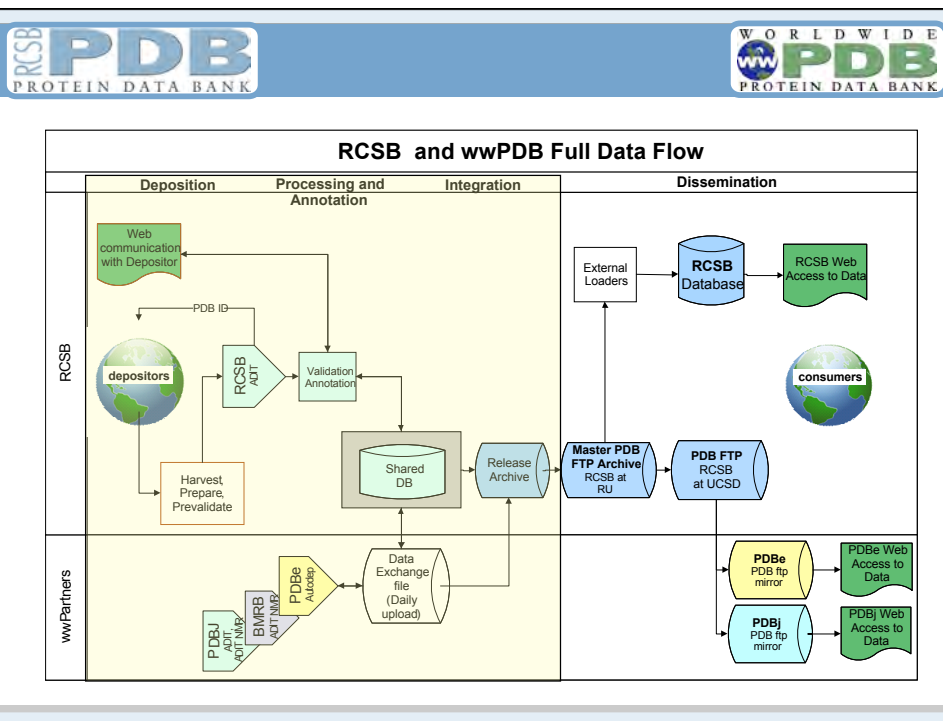
Depositor locations



Download locations

- RCSB PDB
- PDBe
- PDBj





Deposition Statistics

Month	Deposition			Process			Total depositions
	RCSB	PDBj	PDBe	RCSB	PDBj	PDBe	
Jul 2008	554	28	84	408	174	84	666
Aug 2008	459	63	75	357	165	75	597
Sep 2008	539	55	67	394	200	67	661
Oct 2008	523	64	82	409	178	82	669
Nov 2008	403	35	79	274	164	79	517
Dec 2008	410	55	112	295	170	112	577
Jan 2009	557	50	78	397	210	78	685
Feb 2009	536	40	53	395	181	53	629
Mar 2009	610	47	108	484	173	108	765
Apr 2009	540	29	89	387	182	89	658
May 2009	527	19	97	425	121	97	643
Jun 2009	633	50	98	501	182	98	781
Total	6291	535	1022	4724	2102	1022	7848

Data In: Recently Completed Projects

- Complete PDB format documentation version 3.20
- Rollout of remediation version 3.15 files



wwPDB Remediation Leadership Group

(top row): Takanori Matsuura (PDBj), Jawahar Swaminathan (PDBe), Zukang Feng, Jasmine Young (RCSB PDB)
 (bottom row): Shuchi Dutta (RCSB PDB), Miri Hirshberg (PDBe), John Westbrook (RCSB PDB), Dimitris Dimitropoulos (PDBe).

Version 3.20 Documentation and Remediation

- **Process**
 - Each record reviewed for scientific correctness
 - Some PDB records and corresponding mmCIF items revised or added
 - Community input requested and taken into account
- **Deliverables**
 - Complete **PDB File Format Contents Guide Version 3.20** released (Sep 15, 2008; www.wwpdb.org/docs.html)
 - Data files processed to version 3.20 specification since Nov 15, 2008
 - All pre-Nov 2008 files updated, corrected and released as version 3.15 (Mar 2009)

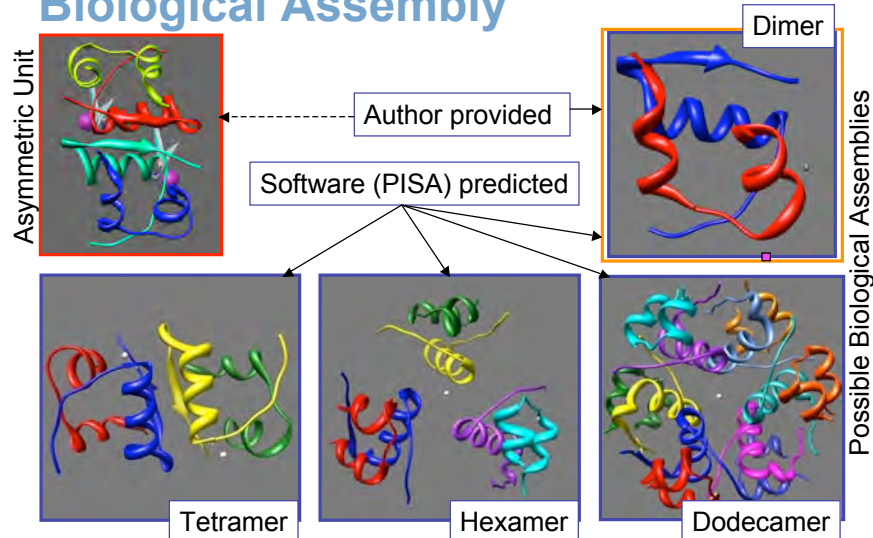
New Records

- **Model description**
 - Split, C-alpha only or phosphate backbone only, number of models in ensemble, re-refinement notice
- **Zero occupancy atoms and residues in polymer listed**
 - REMARKs 475, 480
- **Metal coordination details**
 - REMARK 620
- **Inhibitor description**
 - REMARK 630

Enhancements

- **Biological assemblies: authors and software (PQS/PISA)** ←
- **Binding sites: authors and software** ←
- **Experimental method: name standardized**
- **Ligands/waters: grouped by polymer chain (assign Chain ID)**
- **Sequence DB references deviations: chromophores, microheterogeneity**
- **Database references added/updated:**
 - Taxonomy (Tax ID added)
 - PubMed (ID added for all primary citations)
 - DOI (added for all primary citations)
 - Sequence database references with long accession code recorded in a new format (DBREF1/DBREF2)

Biological Assembly



PDB ID: 3e7y; V.I. Timofeev, A.N. Baidus, Y.A. Kisilitsyn, I.P. Kuranova. Structure of human insulin. DOI: 10.2210/pdb3e7y/pdb
 PISA: E. Krissinel and K. Henrick (2007). Inference of macromolecular assemblies from crystalline state. J. Mol. Biol. 372, 774-797

SITE Records

REMARK 800 SITE_IDENTIFIER: CAT
 REMARK 800 EVIDENCE_CODE: AUTHOR
 REMARK 800 SITE_DESCRIPTION: CATALYTIC AND SUBSTRATE BINDING SITE

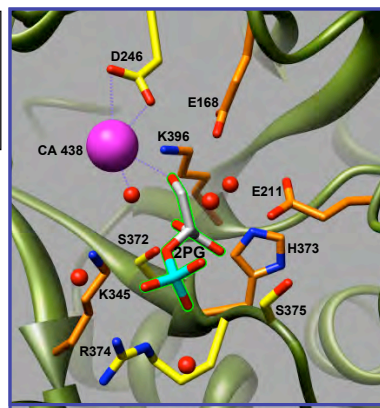
REMARK 800 SITE_IDENTIFIER: AC2
 REMARK 800 EVIDENCE_CODE: SOFTWARE
 REMARK 800 SITE_DESCRIPTION: BINDING SITE FOR RESIDUE 2PG A 442

SITE 1 CAT 5 GLU A 168 GLU A 211 LYS A 345 HIS A 373
 SITE 2 CAT 5 LYS A 396

SITE 1 AC2 15 GLU A 168 GLU A 211 ASP A 246 LYS A 345
 SITE 2 AC2 15 SER A 372 HIS A 373 ARG A 374 SER A 375
 SITE 3 AC2 15 LYS A 396 CA A 438 HOH A 514 HOH A 524
 SITE 4 AC2 15 HOH A 525 HOH A 528 HOH A 716

Author provided Site (CAT) – residue side chains shown in orange

Software calculated Site (AC2) – All residue side chains, ions and waters shown around the ligand 2-phosphoglycerate (2PG).



PDB 5enl. L. Lebioda, B. Stec, J.M. Brewer, E. Tykarska (1991) Inhibition of enolase: the crystal structures of enolase-Ca2(+)-2-phosphoglycerate and enolase-Zn2(+)-phosphoglycolate complexes at 2.2 Å resolution. Biochemistry 30: 2823-2827

Corrections

- **Sequence level corrections**
 - Restore ACE & NH2 to sequence (~800 entries)
 - Sequence DB deviations corrected (microheterogeneity, chromophores; ~370 entries)
- **Data collection details**
 - Beamline, wavelength info updated (~2400 entries)
- **Refinement details**
 - NCS and TLS information revised (~450 entries)
- **REMARKs**
 - Restore critical legacy REMARKs to appropriate current standard REMARKs (~4600 entries)
- **Atom level corrections**
 - Nomenclature of atom, residues, chemical components (~2200 entries)
 - Connectivity (~46000 entries)

Summary of Remediation Rollout

- PDB format version 3.15 files released March 2009
- New records/remarks: to describe models, ligands and zero occupancy atoms/residues
- Enhancements: assemblies, SITE records, chain IDs, database references (including taxonomy id, PubMed, DOI IDs) added for all files
- Various corrections (including sequence, beamline, wavelength, atom connectivities, atom nomenclature, mmCIF consistency)
- Missing remarks restored to appropriate current remarks

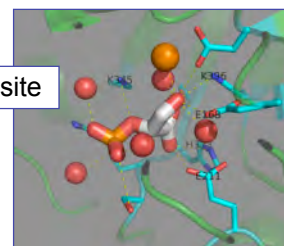
Data In: Ongoing Projects

- Improved tools and uniform data curation
- Remediation and curation of files with complex chemistry
 - Peptide-like inhibitors
 - Antibiotics
- Peptide reference dictionary
- Additional X-ray validation measures
- NMR: Implementation of chemical shifts
- EM maps
- wwPDB common deposition and annotation tool

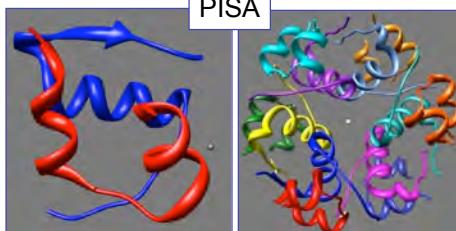
Improved Tools for Uniform Data Curation

- Virus annotation (Point Suite)
- Biological assembly (PISA)
- Getsite (SITE records)
- Ligand Expo: CC Dictionary tools

Getsite



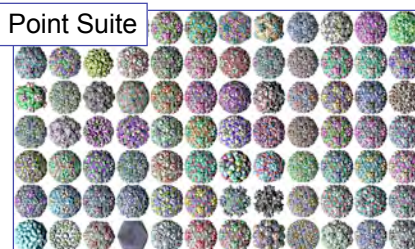
PISA



Author/PISA Dimer

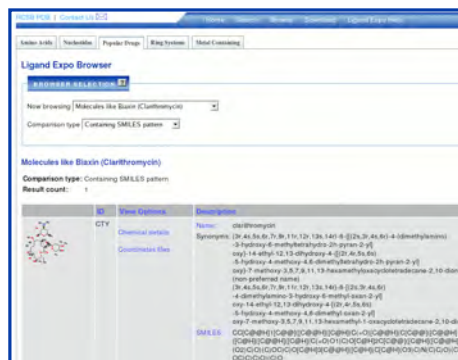
PISA Dodecamer


Point Suite




Ligand Expo and Chemical Component Dictionary Updates

- Ligand Expo
 - Search for chemical components
 - Identify structure entries containing particular small molecules
 - Download the 3D structures of the small molecule components in the PDB entry
 - Build new chemical definitions from reported PDB chemical components
- Chemical Component Dictionary
 - Chemical and systematic names made consistent
 - Various software-generated SMILES strings
 - Chirality checks between coordinates and systematic names
 - Provisions for capturing subcomponent information for peptide-like inhibitors





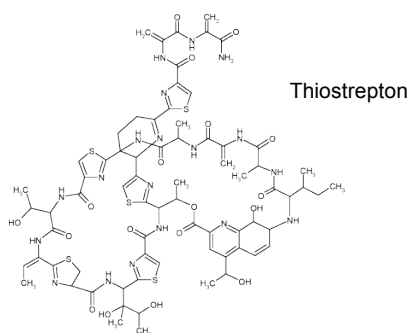
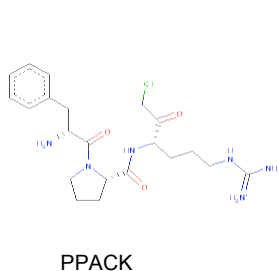


Data In: Ongoing Projects

- Improved tools and uniform data curation
- Remediation and curation of files with complex chemistry
 - Peptide-like inhibitors
 - Antibiotics
- Peptide reference dictionary
- Additional X-ray validation measures
- NMR: Implementation of chemical shifts
- EM maps
- wwPDB common deposition and annotation tool

Remediation and curation of files with complex chemistry

- Peptide-like inhibitors (e.g. PPACK)
- Antibiotics (e.g. Thiostrepton)

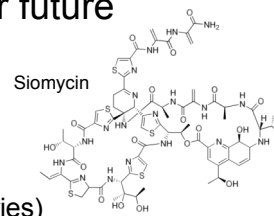


Files with Complex Chemistry: Challenges

- Non-standard amino acid, nucleotides and other chemical groups in sequence
- Non-linear (bridged or branched) sequences
- Microheterogeneity (some cases)
- Non-uniform presentation of the same molecule in different instances
- No annotation about source and function of these molecules

Files with Complex Chemistry: Solutions

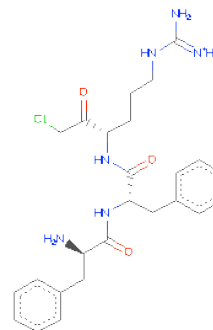
- Identification, analysis and classification of antibiotics and inhibitors to decide presentation (polymeric or single components)
- Create appropriate single chemical components
- Split appropriate components to polymers
- Remediation of files and guidelines for future processing
- Status
 - Antibiotics: 174 entries
 - Inhibitors: 739 entries
 - (366 Single component entries + 373 polymeric entries)



PPACK II

1cvr SEQRES DPN F ACL
 1dan SEQRES DPN F R CH2
 1j9c SEQRES DPN F ARM
 1qfk SEQRES DPN F R

All presented as single molecule 0Z6



Name: D-phenylalanyl-N-[(1S)-4-[[amino(iminio)methyl]amino]-1-(chloroacetyl)butyl]-L-phenylalaninamide
Synonyms: FFRCK; PPACK II
Formula: C₂₅ H₃₄ Cl N₆ O₃
Formal Charge: 1
Subcomponents: DPN PHE ARG 0QE

Sequence information

Data In: Ongoing Projects

- Improved tools and uniform data curation
- Remediation and curation of files with complex chemistry
 - Peptide-like inhibitors
 - Antibiotics
- **Peptide reference dictionary**
- Additional X-ray validation measures
- NMR: Implementation of chemical shifts
- EM maps
- wwPDB common deposition and annotation tool

Peptide Reference Dictionary (PRD)

- Purpose: To collect and organize physical, chemical, functional, commercial and other information about short peptides and peptide-like molecules that appear in the PDB archive
- Identifiers: PRD00001, PRD00002, ...
- Features
 - Links to resources with information about the molecules (CAS, KEGG, ChEBI, Norine, UNP, etc.)
 - Functions extracted from resources such as KEGG, ChEBI
 - Functions extracted from primary citations of the structures
 - Flag to denote how the molecule has been presented (polymer or single component)
- Status
 - A mmCIF framework for the PRD has been set up
 - Summer interns presently working on collecting information about the peptide-like molecules
 - All information and details will be integrated in dictionary later in the year

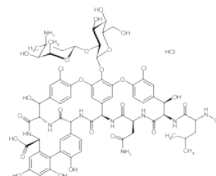
PRD Contents

■ Inhibitors

- Non-ribosomal natural peptides (leupeptin).
- Peptides with standard peptide backbone and D-amino acids or non-standard side chains.
- Peptides with minor variations in the peptide backbone but peptide bond linkages (pepstatin)
- Peptides with special groups or linkages in the peptide backbone (e.g. reduced peptide inhibitors), non-standard chemistry (e.g. PPACK), or bridges/branches. These cases are treated in the PDB as single components with substructure.

■ Antibiotics

- Ribosomal peptides (Lantibiotics like Nisin)
- Non-ribosomal natural peptides (Vancomycin)



Data In: Ongoing Projects

- Improved tools and uniform data curation
- Remediation and curation of files with complex chemistry
 - Peptide-like inhibitors
 - Antibiotics
- Peptide reference dictionary
- **Additional X-ray validation measures**
- NMR: Implementation of chemical shifts
- EM maps
- wwPDB common deposition and annotation tool

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. Luetkeke, Z. Otwinowski, T. Perrakis, J. Richardson, W. Sheffler, J. Smith, I. Tickle, G. Vriend
- Goal
 - Gather recommendations and consensus on additional validation for PDB entries, and identify software applications for these validation tasks.
- 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 Validation Task Force – Next Steps

- October 2009 - Meeting to complete report after CSHL Crystallography Course with follow-on meeting
- November 2009 - Report to be presented at wwPDB Advisory meeting in Japan
- wwPDB partners are pooling manpower to implement Task Force recommendations.
- Validation tools and procedures will also be incorporated in the new wwPDB Common Deposition and Annotation system

Data In: Ongoing Projects

- Improved tools and uniform data curation
- Remediation and curation of files with complex chemistry
 - Peptide-like inhibitors
 - Antibiotics
- Peptide reference dictionary
- Additional X-ray validation measures
- **NMR: Implementation of chemical shifts**
- EM maps
- wwPDB common deposition and annotation tool

NMR

- ADIT-NMR: collaboration with BMRB
- NMR Validation Task Force
 - 1st meeting Sept. 21, 2009, Paris, France
- Implementation of chemical shifts recommended by wwPDB advisory committee

USA: <http://deposit.bmrbl.wisc.edu/bmrbl-adit/>
 Japan: <http://nmradit.protein.osaka-u.ac.jp/bmrbl-adit/>

Data In: Ongoing Projects

- Improved tools and uniform data curation
- Remediation and curation of files with complex chemistry
 - Peptide-like inhibitors
 - Antibiotics
- Peptide reference dictionary
- Additional X-ray validation measures
- NMR: Implementation of chemical shifts
- **EM maps**
- wwPDB common deposition and annotation tool

Electron Microscopy

Collaborative project to create a unified tool for collecting **model coordinates** and **map files** in a *one-stop shop*

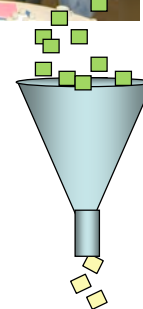
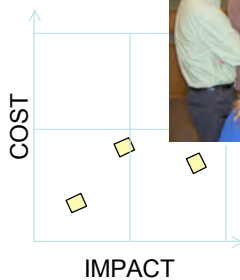
wwPDB Common Deposition and Annotation Tool

Martha Quesada

July 24, 2009

2007 wwPDB Retreat

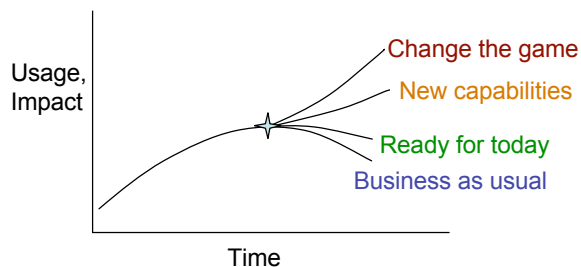
What **strategic objectives** should we start now to meet our goals in the next 5 – 10 years?



Drivers for wwPDB Common Deposition & Annotation Tool Project

- Redundant tools in use: i.e., AutoDep and ADIT
 - Require data harmonization
 - Need for better maintainability and extensibility
- New deposition types: multiple methods and new data
- New validation procedures
- Good collaborations among sites
 - Possibility of sharing of workload
 - Precedent for common tools for NMR and EM

Inflection Point



Decision to come together as the wwPDB in developing the tools that will support the shared functions of the wwPDB for the next 10 years



wwPDB Common D&A Tool Project

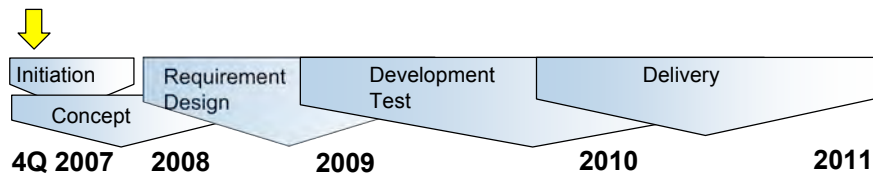
The goal is to produce a set of common deposition and annotation processes and tools that will enable the wwPDB to deliver a resource of increasingly high quality and dependability over the next 10 years.

- The tools will address: the increase in complexity and experimental variety of submissions and the increase in deposition throughput
- The processes and tools will maximize the efficiency and effectiveness of data handling and support for the scientific community

Initiation of the wwPDB Common D&A Tool Project

October 2007: Project Steering Committee

- Set Project Goal, Scope, Assumptions and Constraints
- Set initial high-level timelines
- Name leadership teams
- Approve project at each milestone



Project Scope

- **Common deposition interface and processing**
 - Coordinates (x,y,z) regardless of experiment origin (X-ray, NMR, EM)
 - NMR restraints, chemical shifts
 - X-ray structure factors
 - 3D-EM maps
 - **Data processing: Validation Tools**
 - Coordinates for polymers and ligands
 - NMR Restraints
 - Structure factor validation
 - EM map validation
- } Developed by
External Task Forces
- } To be integrated over
time

Assumptions & Constraints

- **Functional requirements**
 - Deposition tool must handle all current, agreed upon data entry and report formats from the user community
 - All data elements covered within the PDB annotation manual must be included
- **Technical requirement**
 - Design will enable flexibility for growth and evolution
 - Technical level reasonable standard, not bleeding edge or declining
 - Design must enable integration with community data capture

For example

- Deposition will capture all currently deposited experimental data for each method
- The tool will support all data formats and validation requirements for all deposition types
- The system will allow partitioning in deposition among sites
 - Workload balancing
- In the event of multi-method submissions, a path will be created to distribute the annotation effort

Common D&A Project Characteristics

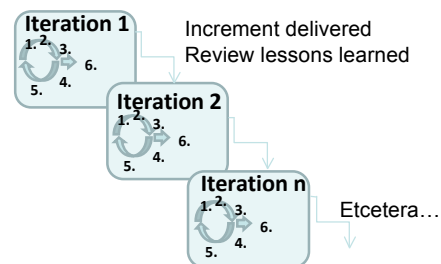
Large and complex – Break into smaller bits

Distributed developers – Establish development controls and communication

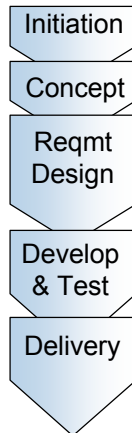
Some requirements well understood – User led

Some requirements evolving

Potentially novel designs require early experimentation



Project Phases, Structure and Roles



- **Steering Committee**
 - Governance
 - Milestone reviews and guidance
- **Concept Team**
 - Initial requirements and design
- **Core Team**
 - Plan and manage the project
- **Project Teams**
 - Design, develop and test component solutions
 - Deliver the solution

Initiation

Steering Committee (wwPDB Directors) October 2007

- Set project Project Goal, Scope, Assumptions and Constraints, initial timelines
- Approves project at each milestone



Concept

Concept Team November 2007

- Objectives, strategies and metrics
- Stakeholder analysis & risk assessment
- New system requirements
- Concept process maps



Approved May 08

Objectives & Strategies

- **Improve data quality** beginning at data capture
 - Provide for interactive feedback and value to the depositors during the deposition process
 - Employ community-driven validation methods
- **Improve efficiency** through workflow automation
- **Improve existing tools**
 - Use “best of breed” existing tools; redevelop tools as time and need dictate
- **Enable system maintenance and evolution** through system modularity

Initiation of Design and Planning

Reqmt
Design

Develop
& Test

The Core Team will lead the project through design and implementation in conversation with the Steering Committee



RCSB PDB: John Westbrook, Jasmine Young; PDBe: Tom Oldfield, Sameer Velankar; BMRB: Steve Mading; PDBj: Takanori Matssura

Project Team: Distributed

Reqmt
Design

- Quarterly face-to-face meetings
- Weekly video conferencing

Develop
& Test

- On-going teleconferences and email
- Shared web-based document and code management tools

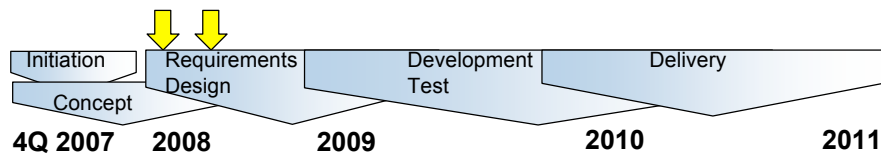
Delivery



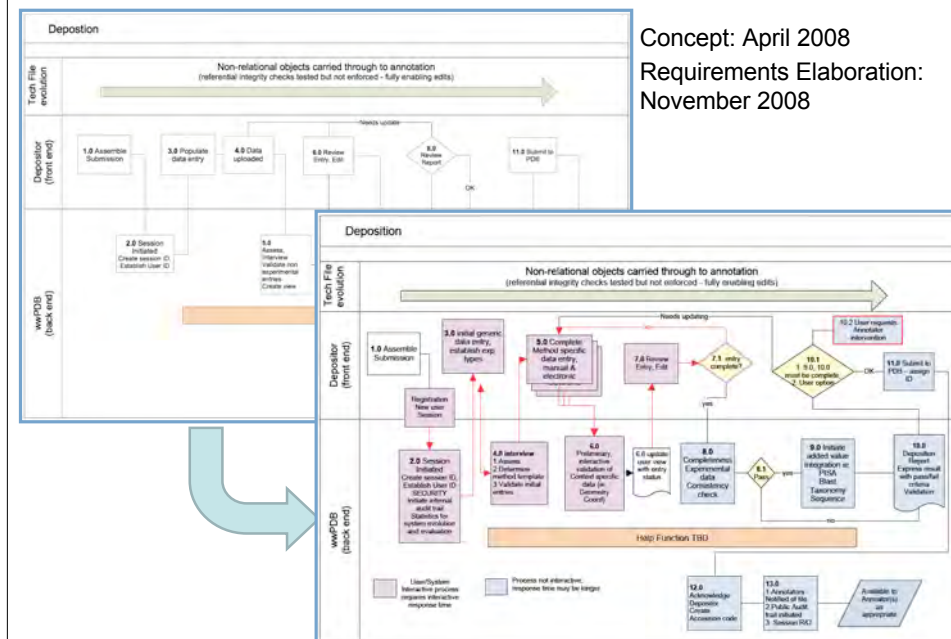
Subject and technical experts from all sites

Project Progress

- **July 2008: Core Team – Initiate Design and Development Planning**
 - Elaborate on requirements and conceptual design
 - Define project deliverables and prioritize tasks
 - Begin initial design research and testing
- **Nov 2008: First Project Team Meeting**
 - Elaborate process requirement
 - Focus on data model requirements
 - Document data flow



Design Through Iterative Analysis



March 2009 – Project Team Meeting Detailed Requirements and Design for Initial Development

Outcomes:

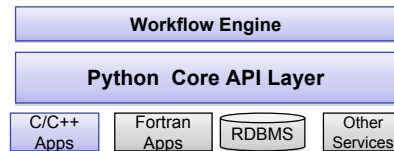
- Detailed data flow requirements by annotators
- Workflow: workflow management tech design
- Modular Design details: API will be created using Python
- Data Replication and sharing infrastructure design
- Technical principles: support distributed development
- Plans for June Proof of Concept deliverables

Proposed technical design and deliverables reviewed and approved by the Steering Committee

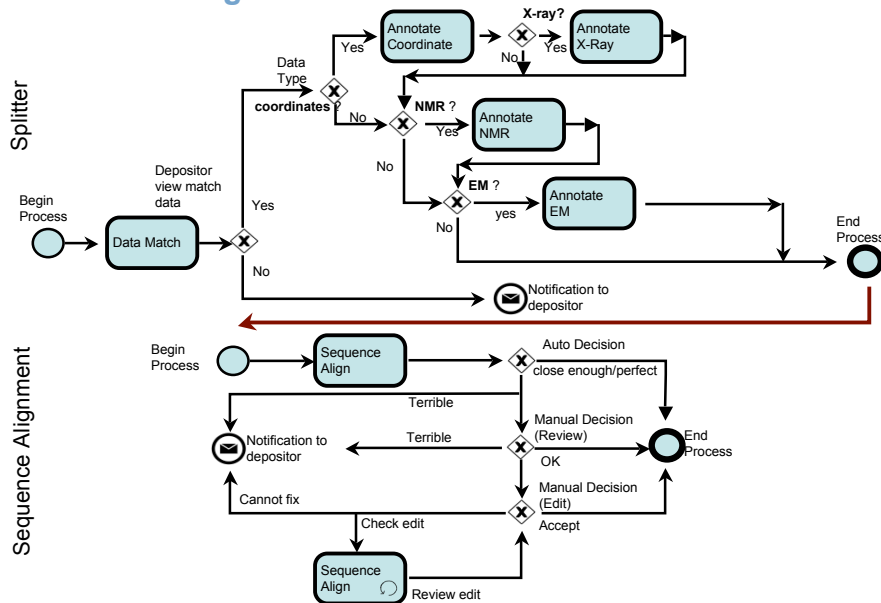
Technical Design Proof Of Concept

Potentially novel designs require early experimentation

- Modular construction through a **core API**
 - Reuse of “best of breed” existing tools; redevelop tools as time and need dictate.
 - Enable system maintenance and evolution through system modularity
- Faster processing through improved efficiency
 - **workflow automation**
- Improved collaboration
 - More flexible data sharing
 - **Snapmirror**



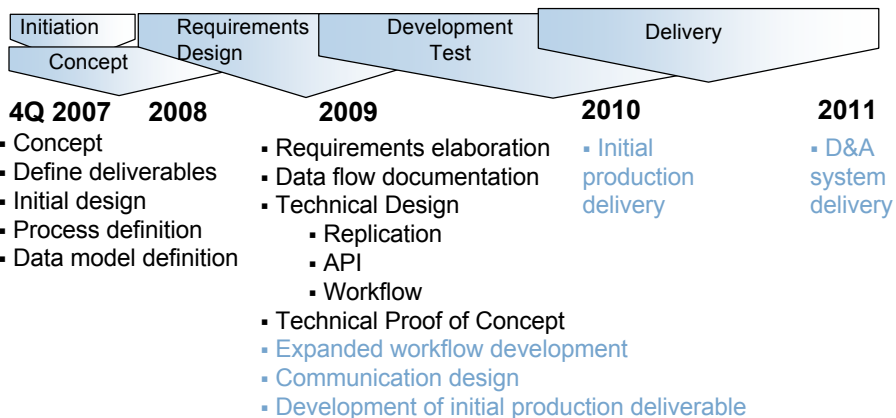
Proof of Concept: Technical Design applied to Data Flow and Processing



Proof of Concept Review Results

- July 14, 2009: Proof of Concept deliverables reviewed with Steering Committee
 - Data sharing results
 - Workflow POC implemented using Core API
 - Team recommendation to adopt proposed architecture design
- July 16, 2009: Steering Committee
 - Accepts Team findings and recommendations
 - Triggers next step: Core Team to prepare detailed proposal for January and June 2010 deliverables by mid-August
 - The proposal will include the deployment of a production ready workflow component

wwPDB Common D&A Tool Project Timeline Going Forward



Deliverables

- **For Depositors**
 - Interactive and informative deposition interface
 - Value added input during deposition
 - Faster processing
- **For Annotators**
 - Improve efficiency, freeing time for more advanced annotation
 - Improved quality early in the process
 - Automation of appropriate processing steps
 - Best of breed tools
 - Enable system maintenance and evolution through system modularity
- **For Data Users**
 - **Higher Quality Archive**

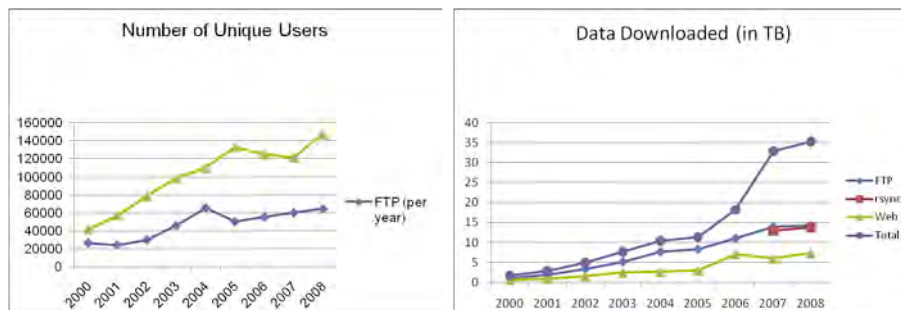
Data Distribution and Query “Data Out”

Philip E. Bourne and Peter Rose

July 24, 2009



Growth of RCSB PDB User Base



PDB FTP Traffic (May 2008 - April 2009)



■ RCSB PDB
180546882
data downloads

■ PDBe
40650943
data downloads

■ PDBj
14506857
data downloads

Website Improvements and New Features

MyPDB

Welcome Peter!

MyPDB Login - Username: admin Password: **** Login

Forgot your Password? Register a New Account!

About MyPDB stores and automatically runs your favorite queries. Email alerts are sent when matching structures are found.

MyPDB Login

[Saved Queries](#)
[Account](#)
[Logout](#)

Query Manager

Your saved queries are listed below. You can assign queries to be executed during each weekly update of the RCSB PDB. Click on email notification to turn automatic notification on or off. Notifications about new structures matching your queries are sent on either a weekly or monthly schedule based on your account settings. Email notifications are only sent if new structures match your queries.

The query name and query description are generated automatically. Click on the name or description to edit the text. Click the save button that appears after you have made your changes.

Name	Query Description	Email Notification	Last Run	Next Scheduled Run	Run Query (No Email)	Delete
Protein-serine/threonine kinases	EnzymeClassificationTree Search for 2.7.11: Protein-serine/threonine kinases.	Yes	07/14/2009	07/21/2009	▶	✕
Latest Release	This week's new structures	Yes	07/14/2009	07/21/2009	▶	✕

From: admin@rcsb.org
To: Peter
Subject: RCSB PDB MyPDB query notification 20 Jun 2009

RCSB Protein Data Bank

Your queries returned the following results:

Query	Structures
Lipid Kinase	<p>2K5A FGF-1-CTA binary complex structure: a key component in the fibroblast growth factor anti-cancer pathway</p> <p>2K3Z NMR structure of protein gp13 of bacteriophage SP1</p> <p>2K7X Structure of the N-terminal domain of human cardiac troponin C bound to caldesmon and to the substrate Y1</p>

- Framework to store user preferences
- Saves queries in a private account
- Notifies users via email when new structures match stored queries
- Users
 - 1580 registered
 - 892 actively checking notification emails

Redesigned Summary Page

Summary [Default Data](#) [Sequence](#) [Seq. Similarity](#) [Contacts](#) [Res. & Chem.](#) [Mutations](#) [Geometry](#) [Links](#)

STRUCTURAL ORIGINS OF HIGH-AFFINITY BIOTIN BINDING TO STREPTAVIDIN **1stp** [Display Files +](#) [Download Files +](#) [Print this Page](#)

DOI:10.2210/pdb1stp/pdb

Primary Citation

Structural origins of high-affinity biotin binding to streptavidin.
Weber, P.C., Ohlendorf, D.H., Wendoloski, J.J., Salemme, F.R.
(1989) Science **243**: 85-88
PubMed: 2911722 [Search Related Articles in PubMed](#)

PubMed Abstract:
The high affinity of the noncovalent interaction between biotin and streptavidin forms the basis for many diagnostic assays that require the formation of an irreversible and specific linkage between biological macromolecules. Comparison of the refined crystal structures of apo and ...
[Read More & Search PubMed Abstracts](#)

Molecular Description [Hide](#)

Classification: **Biotin Binding Protein**
Structure Weight: 14799.31
Molecule: STREPTAVIDIN COMPLEX WITH BIOTIN
Polymer: 1 Type: polypeptide(L) Length: 159
Chain: A

Source [Hide](#)

Polymer: 1 Scientific Name: **Streptomyces avidinii**

Ligand Chemical Component [Hide](#)

Identifier	Name	Formula	Links
BTN	BIOTIN	C10 H16 N2 O3 S	PDB Chemical PubChem

Derived Data [Hide](#)


- SCOP Classification v1.73 - (1 Domains)
- CATH Classification v3.2.0 - (1 Domains)
- PFAM Classification - (1 Domains)
- GO Terms - (1 Terms)

Deposition Summary [Hide](#)

Authors: Weber, P.C., Salemme, F.R.
Deposition: 1992-03-12
Release: 1992-10-15
Last Modified (REVDAT): 2009-02-24

Experimental Details [Hide](#)

Method: X-RAY DIFFRACTION
Experimental Data: N/A
Resolution(Å): 2.60



3-D Viewers: [Jmol](#) [SimpleViewer](#) [Protein Workshop](#) [Other Viewers](#)

Digimeric State: TETRAMERIC

- Information summarized in easy-to-read 2-column format
- Related information presented in customizable “widgets”
- Abstract from PubMed is displayed

Customizable Structure Pages

Primary Citation

The crystal structure of human deoxyhaemoglobin at 1.74 Å resolution
Fermi, G., Perutz, M.F., Shaanan, B., Fourme, R.
(1984) J.Mol.Biol. **175**: 159-174
PubMed: 6726807 [Search Related Articles in PubMed](#)

PubMed Abstract:
The structure of human deoxyhaemoglobin was refined at 1.74 Å resolution using data collected on film at room temperature from a synchrotron X-ray source. The crystallographic R-factor is 16.0%. The estimated error in atomic positions is 0.1 Å overall, 0.14 ... [Read More & Search PubMed Abstracts](#)

Related Citations in PDB Entry (REMARK 1) [Show](#)

Molecular Description [Hide](#)

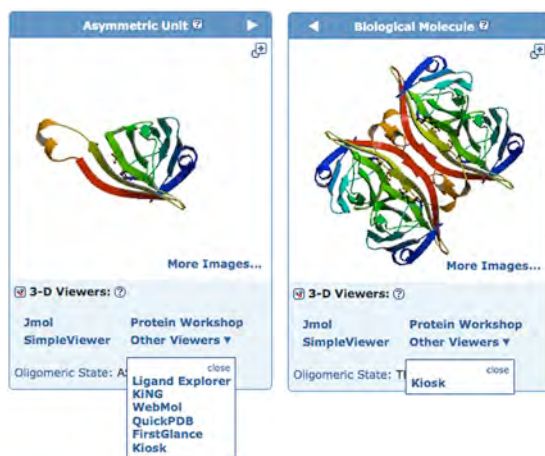
Classification: **Oxygen Transport**
Structure Weight: 64737.73
Molecule: HEMOGLOBIN (DEOXY) (ALPHA CHAIN)
Polymer: 1 Type: polypeptide(L) Length: 141
Chains: A, C
Molecule: HEMOGLOBIN (DEOXY) (BETA CHAIN)
Polymer: 2 Type: polypeptide(L) Length: 146
Chains: B, D

Ligand Chemical Component [Hide](#)

Identifier	Name	Formula	Links
HEM	PROTOPORPHYRIN IX CONTAINING FE	C34 H32 Fe N4 O4	PDB Chemical PubChem
PO4	PHOSPHATE ION	O4 P	PDB Chemical PubChem

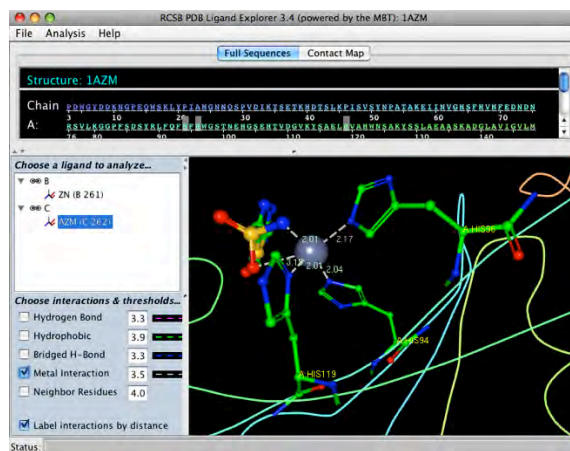
- Widgets can be
 - Re-arranged
 - Hidden
- Settings are stored in a “cookie” to keep view customized for all pages
- Layout can be reset to default
- Derived data are summarized with links to derived data page

Improved Visualization Options



- 3D Viewers are context-sensitive
 - Asymmetric unit
 - Biological molecule
- Biological molecule is displayed by default
- Presumed oligomeric state of biological molecule is displayed (for X-ray structures)

Protein-Ligand Interaction View



- Simplified user interface
- Added metal interactions
- Display of bond orders from Chemical Component Dictionary

Derived Data View

- Annotations displayed in tabular format
- Blue text indicates *Query by Example*

Summary Derived Data Sequence Seq. Similarity Ligature Res. & Chem. Methods Geometry Links

Derived Data **4hhb** [Display Files](#) [Download Files](#) [Print this Page](#) [Share this Page](#)

Derived Data: SCOP Classification (version 1.73) [G](#) [Hide](#)

Domain Entry	Class	Fold	Superfamily	Family	Domain	Species
041MB_e	All alpha proteins	Globin-like	Globin-like	Globins	Hemoglobin, alpha-chain	Human (Homo sapiens) [TaxId: 9606]
041MB_c	All alpha proteins	Globin-like	Globin-like	Globins	Hemoglobin, alpha-chain	Human (Homo sapiens) [TaxId: 9606]
041MB_b	All alpha proteins	Globin-like	Globin-like	Globins	Hemoglobin, beta-chain	Human (Homo sapiens) [TaxId: 9606]
041MB_d	All alpha proteins	Globin-like	Globin-like	Globins	Hemoglobin, beta-chain	Human (Homo sapiens) [TaxId: 9606]

Derived Data: PFAM Classification [G](#) [Hide](#)

Chain	PFAM Accession	PFAM ID	Description	Type	Class ID
D	PF00042 G	Globin	Globin	Domain	
C	PF00042 G	Globin	Globin	Domain	
B	PF00042 G	Globin	Globin	Domain	
A	PF00042 G	Globin	Globin	Domain	

Derived Data: GO Terms [G](#) [Hide](#)

Protein	Molecular Function	Biological Process	Cellular Component
HEMOGLOBIN (DEOXY) (ALPHA CHAIN) (4HHB:A,C)	<ul style="list-style-type: none"> oxygen transporter activity iron ion binding protein binding oxygen binding heme binding metal ion binding 	<ul style="list-style-type: none"> transport oxygen transport 	<ul style="list-style-type: none"> hemoglobin complex cytosolic small ribosomal subunit
HEMOGLOBIN (DEOXY) (BETA CHAIN) (4HHB:B,D)	<ul style="list-style-type: none"> oxygen transporter activity iron ion binding protein binding oxygen binding heme binding hemoglobin binding metal ion binding 	<ul style="list-style-type: none"> transport regulation of blood pressure oxygen transport nitric oxide transport positive regulation of nitric oxide biosynthetic process regulation of blood vessel size 	<ul style="list-style-type: none"> hemoglobin complex

New Views of Data

RCSB PDB PROTEIN DATA BANK

Literature View

Summary Derived Data Sequences Seq. Similarity Literature **1a0i** Misc. & Other Methods Secondary Links

COMPLEX BETWEEN NUCLEOSOME CORE PARTICLE (H3,H4,H2A,H2B) AND 146 BP LONG DNA FRAGMENT

Primary Citation and Related Literature

Primary Citation
 Crystal structure of the nucleosome core particle at 2.8 Å resolution.
 Luger, K., Mader, A.W., Richmond, R.K., Sargent, D.F., Richmond, T.J.
 (1997) Nature 389: 251-260
 PubMed: 9305377 Search Related Articles in PubMed

PubMed Abstract:
 The X-ray crystal structure of the nucleosome core particle of chromatin shows in atomic detail how the histone protein octamer is assembled and how 146 base pairs of DNA are organized into a superhelix around it. Both histone/histone and histone/DNA ... [Read More & Search PubMed Abstracts]

Information provided by BioLit:

PubMedCentral articles found to contain 1A0I

Alteration of the nucleosomal DNA path in the crystal structure of a human nucleosome core particle
 Yasuo Tsunaka, Naoko Kajimura, Shin-ichi Tate, Kosuke Morikawa
 Nucleic Acids Research 2005, 33(10):3459ff
 PubMedCentral: 1150222 PubMed: 15951514 DOI: 10.1093/nar/gki663
 Abstract Copyright

Other PDB IDs found in the above articles

Image	PDB ID	Structure Title	Sequence Similarity
	1EQZ	X-RAY STRUCTURE OF THE NUCLEOSOME CORE PARTICLE AT 2.5 Å RESOLUTION	> 95%
	11D3	CRYSTAL STRUCTURE OF THE YEAST NUCLEOSOME CORE PARTICLE REVEALS FUNDAMENTAL DIFFERENCES IN INTER-NUCLEOSOME INTERACTIONS	> 70%
	1K03	X-Ray Structure of the Nucleosome Core Particle, NCP146, at 2.0 Å Resolution	100%

- Links PDB IDs to articles in PubMedCentral
- Linkages are established by the BioLit project
- Provides abstracts, figures, figure legends, DOIs, and copyright statements
- Lists other PDB IDs found in these articles

RCSB PDB PROTEIN DATA BANK

Sequence Similarity View

Summary Derived Data Sequences **Seq. Similarity** Literature Misc. & Other Methods Secondary Links

THE CRYSTAL STRUCTURE OF HUMAN DEOXYHAEMOGLOBIN AT 1.74 ANGSTROMS RESOLUTION

4hhb Display Files Download Files Print this Page

Sequence Clustering and Redundancy Reduction Results

Documentation
 Click here for more detailed documentation on the redundancy reduction and sequence clustering procedure used by RCSB.

Sequence Clusters for the Sequence Entitle in PDB 4HHB

Entity #1: Chains: A,C - HEMOGLOBIN (DEOXY) (ALPHA CHAIN) polypeptide(L), length: 141

Cluster Sequence Similarity Cutoff	Rank in Cluster	Cluster Nr.
100%	16	1
95%	18	4
90%	18	5
70%	22	8
50%	52	3
40%	52	4
30%	149	3

Entity #2: Chains: B,D - HEMOGLOBIN (DEOXY) (BETA CHAIN) polypeptide(L), length: 146

Cluster Sequence Similarity Cutoff	Rank in Cluster	Cluster Nr.
100%	16	1
95%	18	4
90%	18	5
70%	22	8
50%	52	3
40%	52	4
30%	149	3

Cluster # 5 Members. Sequence cutoff: 90% Cluster size: 175 sequences

Rank in Cluster	PDB Id	Entity Id	Chains	Description	Details	EC Number
1	2W72	1	A	HUMAN HEMOGLOBIN A	CHAIN ALPHA, RESIDUES 2-142	
2	2W72	3	C	HUMAN HEMOGLOBIN A	CHAIN ALPHA, RESIDUES 2-142	
3	13R0	1	A	Hemoglobin alpha chain		
4	2DN2	1	A, C	Hemoglobin alpha subunit		
5	2DN3	1	A	Hemoglobin alpha subunit		
6	2DN1	1	A	Hemoglobin alpha subunit		
7	1341	1	A, C, E, G	Hemoglobin alpha Chain		
8	1340	1	A, C, E, G	Hemoglobin alpha Chain		

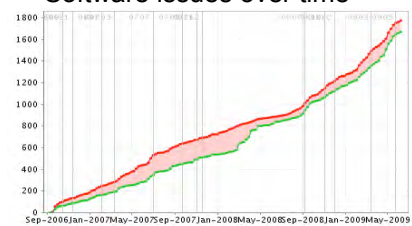
- Displays clusters of structures related by sequence (generated by BlastClust)
- Convenient way to explore related structures by sequence similarity
- If available, EC numbers are displayed

Software and Hardware

Software Releases (July 08 – May 09)

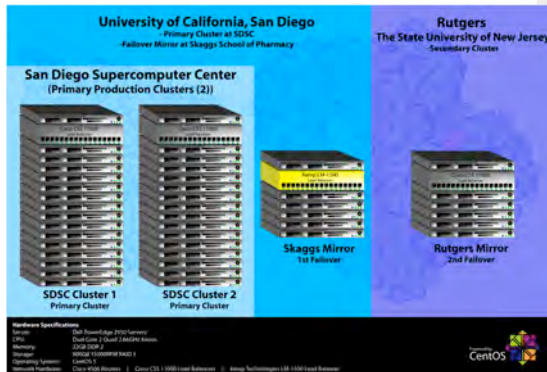
Release	Comment	New Features	Improvements	Tasks	Bugs	Total
Jul-08	Maintenance release	6	29	9	87	131
Oct-08	MyPDB beta & 3D Domain view	20	22	11	65	118
Dec-08	MyPDB	15	25	16	62	118
Mar-09	File remediation	28	43	42	89	202
May-09	New web layout & Literature view	39	95	37	78	249

Software issues over time



- Red: issued filed
- Green: issue resolved

Current Hardware & Software Infrastructure



Optimized for uptime and performance

- Commodity cluster – high level of redundancy, load balancing
- Failover from SDSC to UCSD Pharmacology & Rutgers
- Performance monitoring: Server Stalker
- Complete open source software solution using Java Enterprise Edition technology stack: Struts2, Tomcat, MySQL
- Query catching: Memcached
- Quality assurance: Selenium, Jira, JUnit tests, Load tests

Server Monitoring

Administrator Panel Login - Username: Password: Login

Server Stalker 4
Version 4.1.1

LOD's
FAILOVER

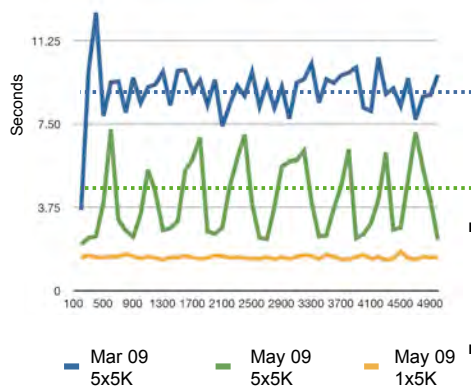
Server	IP	Availability
pdb101.sdsc.edu	198.202.122.101	Available
pdb102.sdsc.edu	198.202.122.102	Available
pdb103.sdsc.edu	198.202.122.103	Available
pdb109.sdsc.edu	198.202.122.109	Available
pdb110.sdsc.edu	198.202.122.110	Available
pdb104.sdsc.edu	198.202.122.104	Available
pdb105.sdsc.edu	198.202.122.105	Available

Test Details:

ping: Alive
Test in ~ 2 Mins

httpStatus: Status: Available, CPU: 3%, HDD: 76%, RAM: 78%, Structure Count: 58886, MySQL: ACTIVE
Test in ~ 0 Mins

Improved Server Response Time



- 5 tests run in parallel

- Retrieve 5000 random IDs
 - Structure summary page
 - Structure image
 - Sequence display
 - Sequence display with DSSP & SCOP
 - Biology & Chemistry page

- Improvements

- Parallel garbage collection
- Query caching

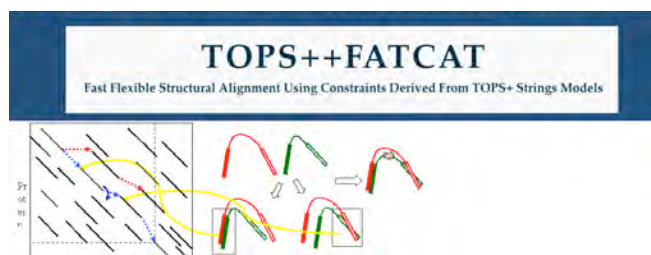
- Result

- May 2009 release is about twice as fast as March 2009 release

Proposed Features: Next 6-12 Months

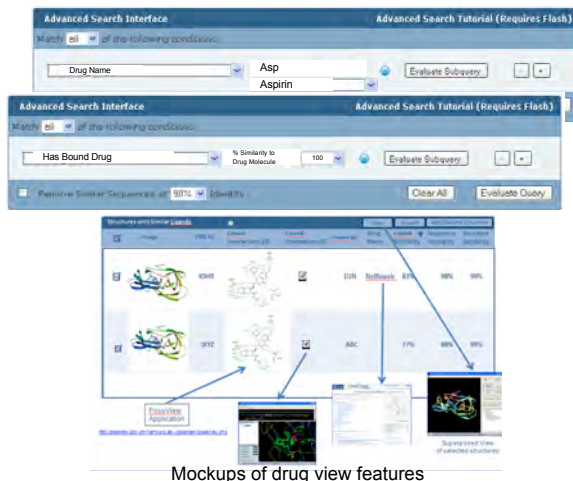
Structural Neighbors

- Flexible and rigid structural alignment
- Evaluate FATCAT method*
- Use to classify common and novel folds



* Veeramalai, Ye, Godzik *BMC Bioinformatics* 2008.

Pharmacology/Drug View



Mockups of drug view features

- Establish linkages to drug resources (FDA, PubChem, DrugBank, etc.)
- Create query capabilities for drug information
- Provide superposed views of ligand binding sites
- Analyze and display protein-ligand interactions

Current Use of External Data

- ▶ 1: Oxidoreductases.
- ▶ 2: Transferases.
- ▶ 3: Hydrolases.
- ▶ 4: Lyases.
- ▶ 5: Isomerases.
- ▶ 6: Ligases.

Browse classifications

[GO Hits](#) [SCOP Hits](#) [CATH Hits](#)

Advanced Keyword Query for: KINASE

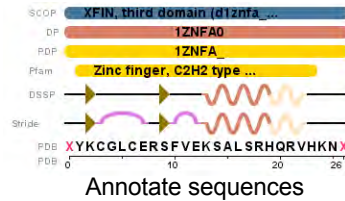
1 2 3 4 5 .. 320 ↩

Classify search results

EXPERIMENTAL DATA

- [Electron Density Server \(EDS\)](#)
- [Electron Microscopy Database \(EMDB\)](#)

Links to external resources



Annotate sequences

↓ **Derived Data: CATH Classification (version v3.2.0)** ↗

Domain	Class	Architecture	Topology
1aewA00	Mainly Alpha	Up-down Bundle	Ferritin;

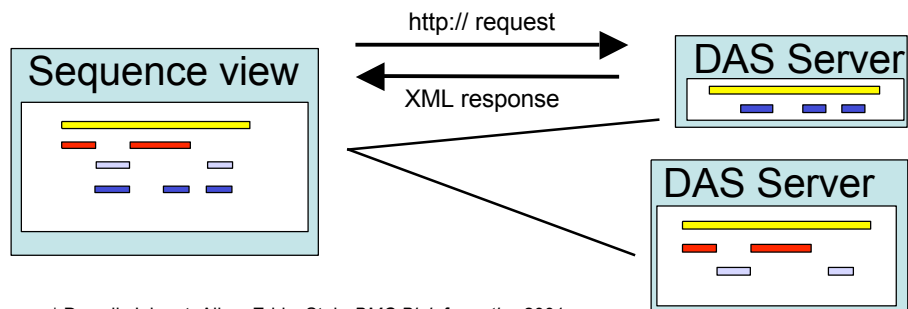
↓ **Derived Data: GO Terms** ↗

Polymer	Molecular Function	Biological Process
FERRITIN (1AEW:A)	<ul style="list-style-type: none"> ◦ binding ◦ iron ion binding ◦ ferric-iron binding ◦ oxidoreductase activity ◦ metal ion binding ◦ transition metal ion binding 	<ul style="list-style-type: none"> ◦ iron ion transport ◦ cellular iron ion homeostasis

Query PDB using derived data

Future Directions for External Data

- Data Integration via Distributed Annotation System (DAS)*
- Dynamic data exchange with external data providers
- Extending annotation while keeping derived data external



* Dowell, Jokerst, Allen, Eddy, Stein *BMC Bioinformatics* 2001.

Data Access Services

- Scalable access to PDB data
 - Programmatically (Web Services)
 - Web browser (Tabular Reports)

