

RCSB Protein Data Bank Advisory Committee Meeting

September 25, 2013



Overview

Helen M. Berman



Response to Major 2012 Recommendations

AC: Conduct a user survey in preparation of renewal proposal

Response: Conducted November-December 2012

AC: Carefully consider final testing and implementation of the Common Deposition and Annotation System

Response: Detailed transition plan being implemented

AC: Need to improve communication with depositors about biological assembly information

Response: New system captures biological assembly information, including experimental details



Response to Major 2012 Recommendations (cont.)

AC: Monitor use of the RCSB PDB by mobile devices

Response: Usage addressed in Data Out

AC: Development of online courses

Response: R25 Proposal under review

AC: Seek external funding for outreach

Response: Success with NIDA, Rutgers proposals

AC: Periodically reevaluate outreach expenditures

**Response: Print newsletter reviewed via survey;
printing discontinued after July 2013**



2013 NSF Site Visit

“Investment in the RCSB PDB is bringing huge benefits to research, education, discovery, and innovation. The return on the dollar is particularly strong. Continued funding of the RCSB PDB is very strongly recommended.”

Site Visit Team, May 10, 2013



Rationale and Vision

To provide a Structural View of Biology that enables an understanding of biological functions and processes at the molecular level in 3D

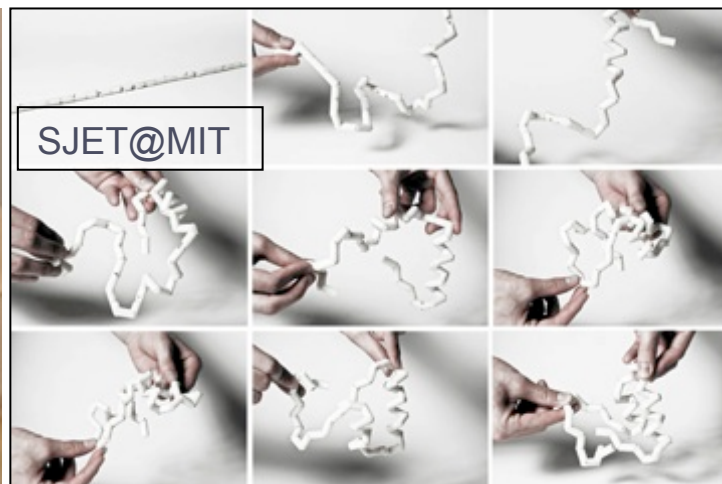
Mission

To produce a sustainable resource that is by, for, and of the community by providing

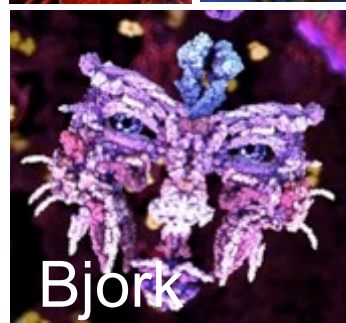
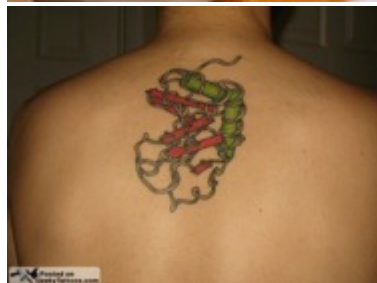
- Leadership in the representation of biological structures derived via experimental methods
- Access to data in an accurate and timely manner
- Leadership and maintenance of the unified global PDB archive
- Comprehensive, integrated and unique views of the data supporting a broad base of scientific inquiry



Impact



IYCr2014

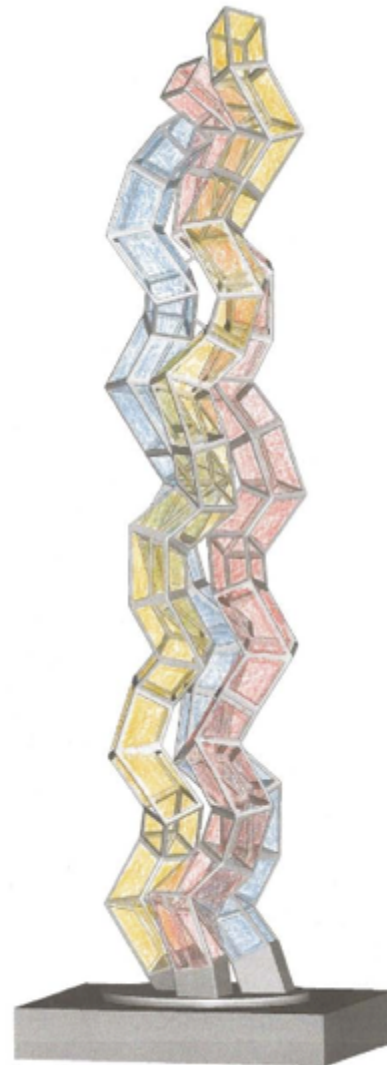


Bjork



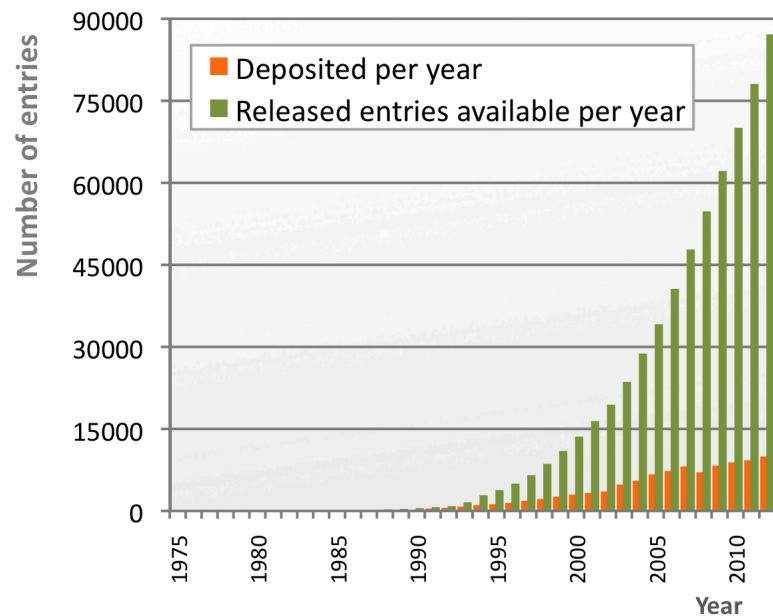
Meeting the Mission: Components

- Data In
- Data Out
- Outreach and Impact
- Management



Julian Voss-Andreae
Synergy, 2013
Sketch rendering

Data In



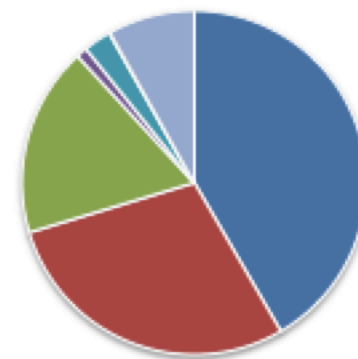
PDB Depositions

9972 depositions in 2012

>800 new entries/month



PDB Depositors by Location



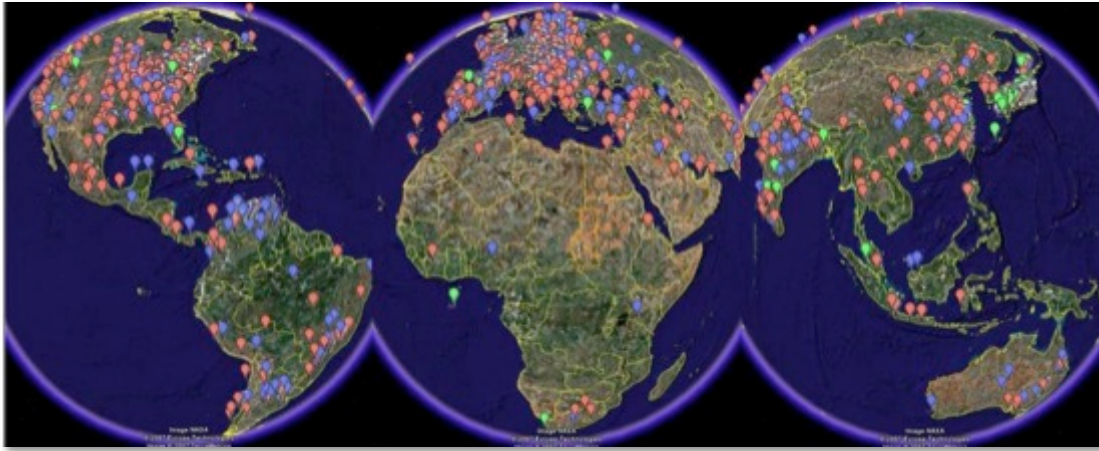
- North America (42%)
- Europe (29%)
- Asia (18%)
- South America (1%)
- Australia/New Zealand (2%)
- Africa (<1%)
- Commercial (8%)

Ensuring That Data Are Freely and Globally Available



- Partners
 - RCSB PDB (Research Collaboratory for Structural Bioinformatics, Rutgers/UCSD)
 - PDJ (Osaka University)
 - PDBe (EMBL-EBI)
 - BioMagResBank (University of Wisconsin, Madison)
- Collaborate on the guiding policies for unified data processing and annotation (Data In)
- Each partner website offers diverse services and views of the data (Data Out)

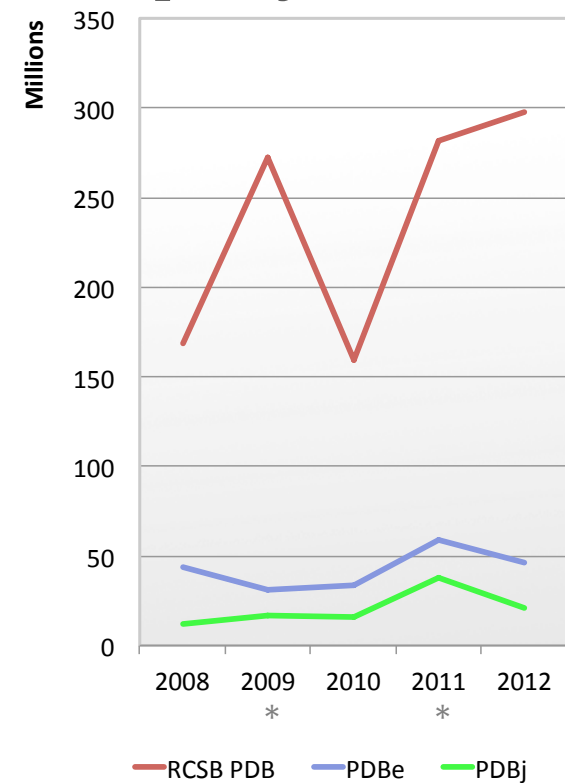
Data Out



365 million data downloads (FTP/HTTP) in 2012

- RCSB PDB: 298 million
- PDBe: 46 million
- PDBj: 21 million

Downloads per year



*Release of remediated data

RCSB PDB Data Out Services



2012 Website Usage Monthly Averages

- 250K Unique Visitors
- 600K Visits
- 854 GB Bandwidth Usage

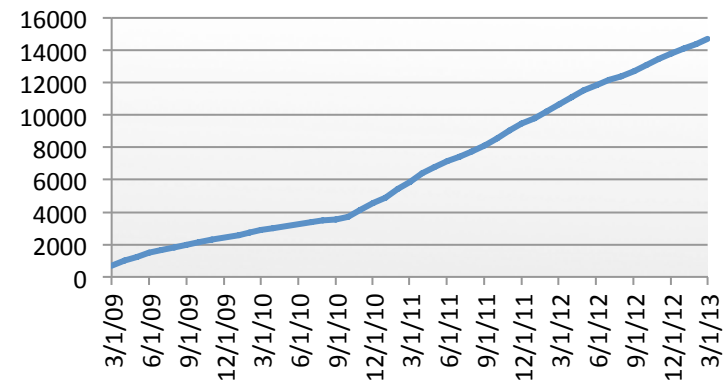
RCSB PDB *Mobile*

>9000 downloads since Aug 2012



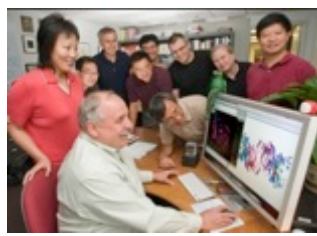
MyPDB Service

~15K Users (since Mar 2009)

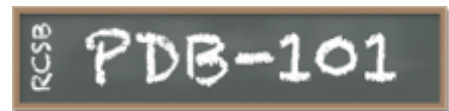


International User Communities

Who are our users?	What are they using?	How do we know?
Biologists: structural biology, biochemistry, genetics, pharmacology, ...	RCSB PDB website, deposition tools, data	Publication requests, website usage, info@rcsb.org requests, community outreach
Other scientists: bioinformatics, software developers, ...	Web Services, search engines, data	Publication requests, website usage, info@rcsb.org requests, community outreach
Students & teachers	PDB-101	Increase in web hits, email, meeting interactions
Media	Images, data, information	Publications, image requests
General public	Images, <i>Molecule of the Month</i> , information from external media	Concerts, media, Wikipedia



PDB-101



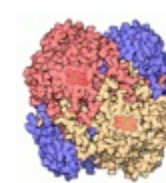
2012
PDB-101
Access:



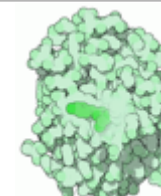
Top Molecules of the Month



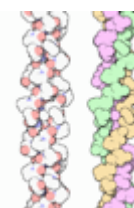
Hemoglobin
43,849 views



Catalase
31,875 views



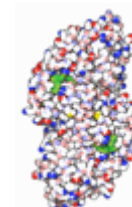
GFP
27,500 views



Collagen
25,429 views



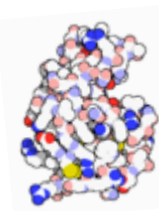
Alpha-amylase
22,389 views



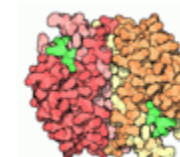
Alcohol dehydrogenase
20,921 views



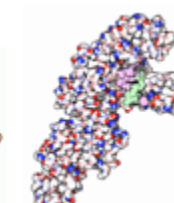
Insulin
20,071 views



Lysozyme
20,008 views



Caspases
17,543 views

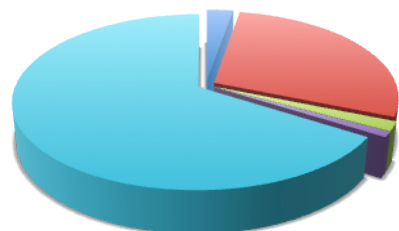


DNA polymerase
17,469 views

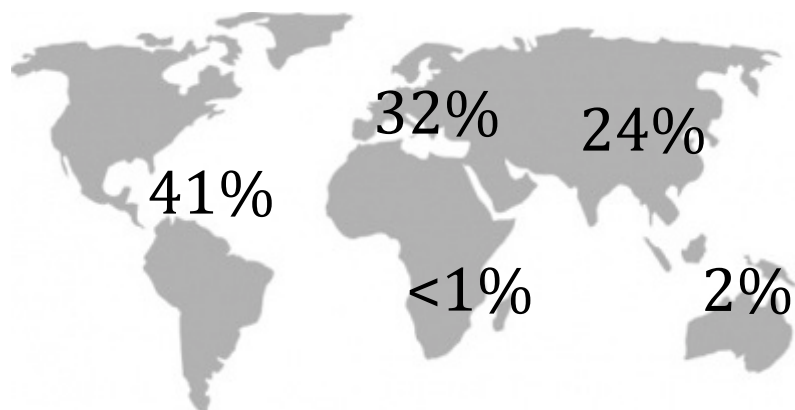
2012 Survey: Who Are Our Users?



- American Indian or Alaska Native (3%)
- Asian (28%)
- Black or African American (2%)
- Native Hawaiian or Other Pacific Islander (1%)
- White (66%)



(12% Hispanic or Latino)



Where do they work?

- College/University (70%)
- Research Institute (16%)
- Government (3%)
- Pharma/drug discovery/biotech (6%)
- K12 (2%)
- Other (3%)



College/University

Type

- University (580)
- 4-year college (54)
- Women's college (9)
- Historically Black College/University (1)

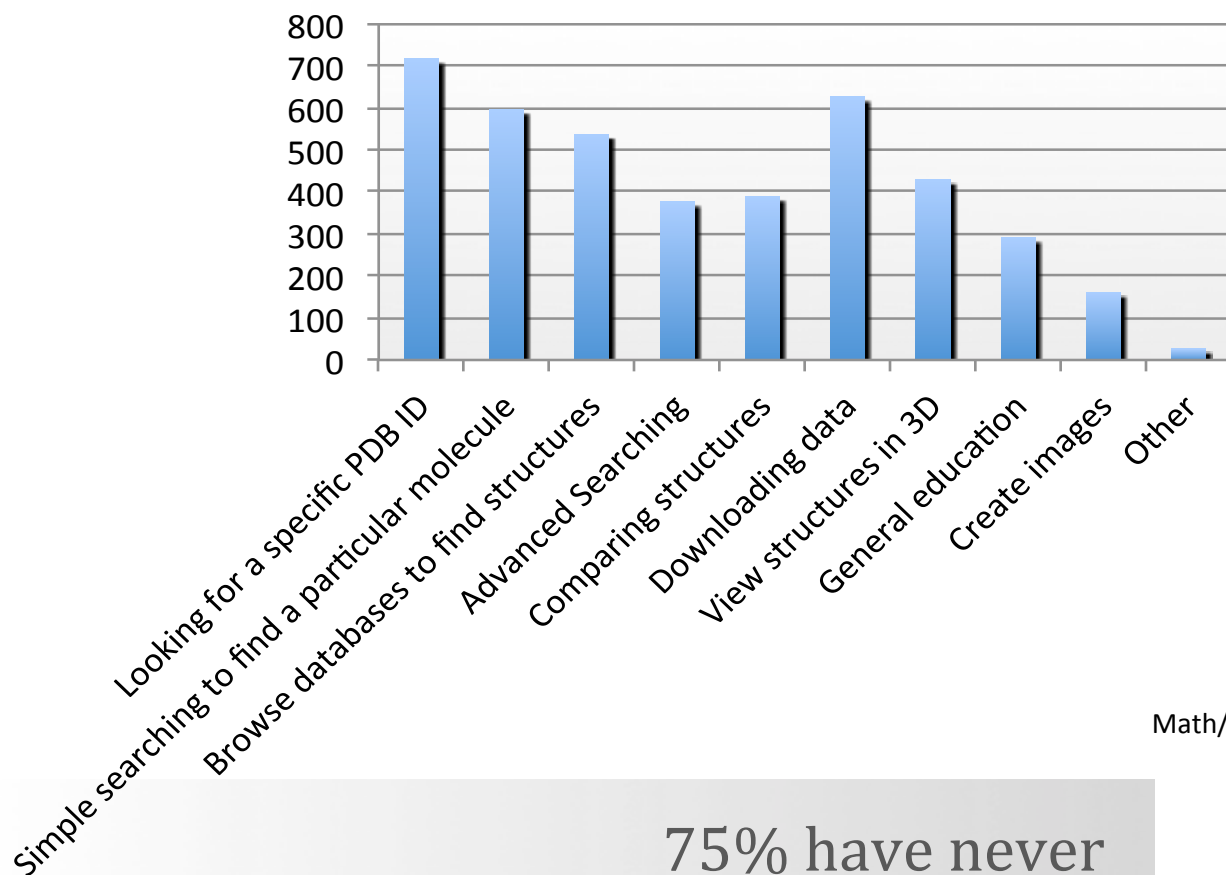
Role

- Undergrad (28%)
- Graduate (37%)
- Postdoc (10%)
- Faculty (13%)
- Staff (9%)
- Other (3%)

Based on 973 responses

What Do They Do?

Why do you use the RCSB PDB?



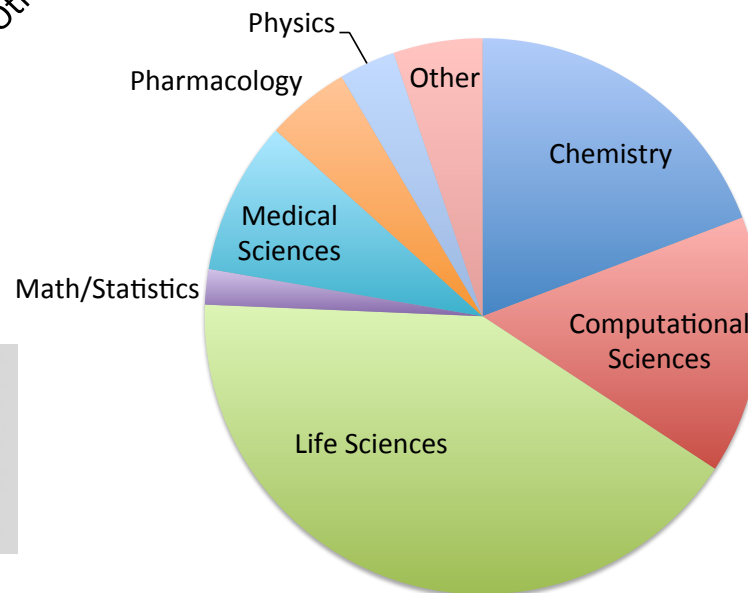
75% have never deposited a structure

Users visit...



- Daily (26%)
- Weekly (41%)
- Monthly (12%)
- Occasionally (21%)

Research Area



Who Cites the RCSB PDB?



Primary reference cited by >15,000

15 RCSB PDB references are cited >100 times

© 2000 Oxford University Press

Nucleic Acids Research, 2000, Vol. 28, No. 1 235-242

The Protein Data Bank

Helen M. Berman^{1,2*}, John Westbrook^{1,3}, Zekang Feng^{1,3}, Gary Gilliland^{1,3}, T. N. Bhat^{1,3}, Helge Weissig^{1,4}, Ilya N. Shindyalov⁴ and Philip E. Bourne^{1,4,5,6}

¹Research Collaboratory for Structural Bioinformatics (RCSB), ²Department of Chemistry, Rutgers University, 610 Taylor Road, Piscataway, NJ 08854-8087, USA, ³National Institute of Standards and Technology, Route 270, Quince Orchard Road, Gaithersburg, MD 20899, USA, ⁴San Diego Supercomputer Center, University of California, San Diego, 9500 Gilman Drive, La Jolla, CA 92093-0505, USA, ⁵Department of Pharmacology, University of California, San Diego, 9500 Gilman Drive, La Jolla, CA 92093-0500, USA and ⁶The Burnham Institute, 10901 North Torrey Pines Road, La Jolla, CA 92037, USA

Received September 20, 1999; Revised and Accepted October 17, 1999

RCSB Protein Data Bank [Edit](#)

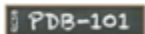


Rutgers and UCSD [Edit](#)

[structural biology - databases](#) [Edit](#)

Verified email at rcsb.org [Edit](#)

My profile is private [Edit](#) [Add homepage](#)



[Change photo](#)

Citation indices

	All	Since 2008
Citations	20555	10705
h-index	29	26
i10-index	41	37

Citations to my articles



Select: **All**, None [Actions](#)

Show: 20 [1-20](#) [Next >](#)

Title / Author	Cited by	Year
The protein data bank		
<input type="checkbox"/> HM Berman, J Westbrook, Z Feng, G Gilliland, TN Bhat, H Weissig, IN ...	15371	2000
Nucleic acids research 28 (1), 235-242		

The protein data bank	785	2002
HM Berman, T Battistuz, TN Bhat, WF Bluhm, PE Bourne, K Burkhardt, Z Feng ... Acta Crystallographica Section D: Biological Crystallography 58 (6), 899-907		
Announcing the worldwide protein data bank	693	2003
H Berman, K Henrick, H Nakamura Nature structural biology 10 (12), 980-980		
The worldwide Protein Data Bank (wwPDB): ensuring a single, uniform archive of PDB data	455	2007
H Berman, K Henrick, H Nakamura, JL Markley Nucleic acids research 35 (suppl 1), D301-D303		
The protein data bank and structural genomics	312	2003
J Westbrook, Z Feng, L Chen, H Yang, HM Berman Nucleic Acids Research 31 (1), 489-491		
The RCSB Protein Data Bank: a redesigned query system and relational database based on the mmCIF schema	267	2005
N Deshpande, KJ Address, WF Bluhm, JC Merino-Ott, W Townsend-Merino, Q Zhang ... Nucleic acids research 33 (suppl 1), D233-D237		
The Protein Data Bank and the challenge of structural genomics	264	2000
HM Berman, TN Bhat, PE Bourne, Z Feng, G Gilliland, H Weissig, J Westbrook Nature Structural Biology 7 (11; SUPP), 957-959		
The protein data bank: unifying the archive	210	2002
J Westbrook, Z Feng, S Jain, TN Bhat, N Thanki, V Ravichandran, GL Gilliland ... Nucleic Acids Research 30 (1), 245-248		
The RCSB Protein Data Bank: redesigned web site and web services	190	2011
PW Rose, B Beran, C Bi, WF Bluhm, D Dimitropoulos, DS Goodsell, A Pricl, M ... Nucleic acids research 39 (suppl 1), D392-D401		
The RCSB PDB information portal for structural genomics	168	2006
A Kouranov, L Xie, J De La Cruz, L Chen, J Westbrook, PE Bourne, HM Berman Nucleic Acids Research 34 (suppl 1), D302-D305		
The Molecular Biology Toolkit (MBT): a modular platform for developing molecular visualization applications	156	2005
JL Moreland, A Gramada, OV Buzko, Q Zhang, PE Bourne BMC bioinformatics 6 (1), 21		



Title: **The Protein Data Bank**
 Author(s): **Berman, HM ; Westbrook, J ; Feng, Z ; et al.**
 Source: **NUCLEIC ACIDS RESEARCH** Volume: **28** Issue: **1** Pages: **235-242** DOI:
10.1093/nar/28.1.235 Published: **JAN 1 2000**

This item has been cited by items indexed in the databases listed below. [\[more information\]](#)

11,486 in All Databases

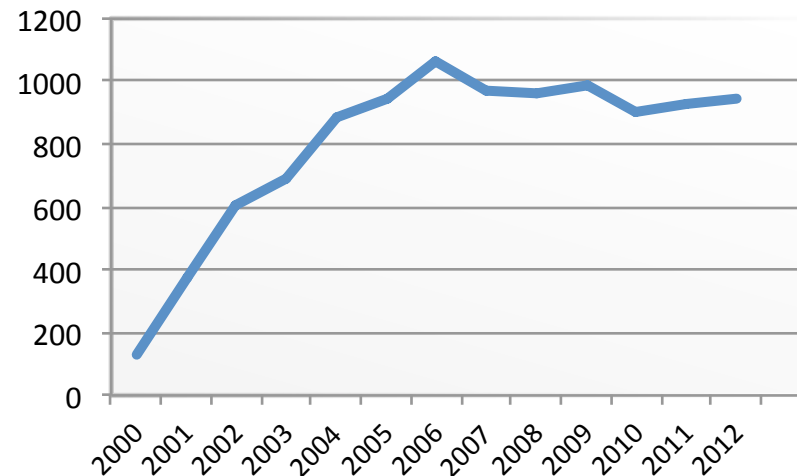
11,205 publication in Web of Science

- + **10,567** in Science Citation Index Expanded (SCIE), Social Science Citation Index (SSCI), and Arts & Humanities Citation Index (A&HCI)
- + **985** in Conference Proceedings Citation Index - Science (CPCI-S); Conference Proceedings Citation Index - Social Science & Humanities (CPCI-SSH)
- + **214** in Book Citation Index- Science (BKCI-S); Book Citation Index- Social Sciences & Humanities (BKCI-SSH)

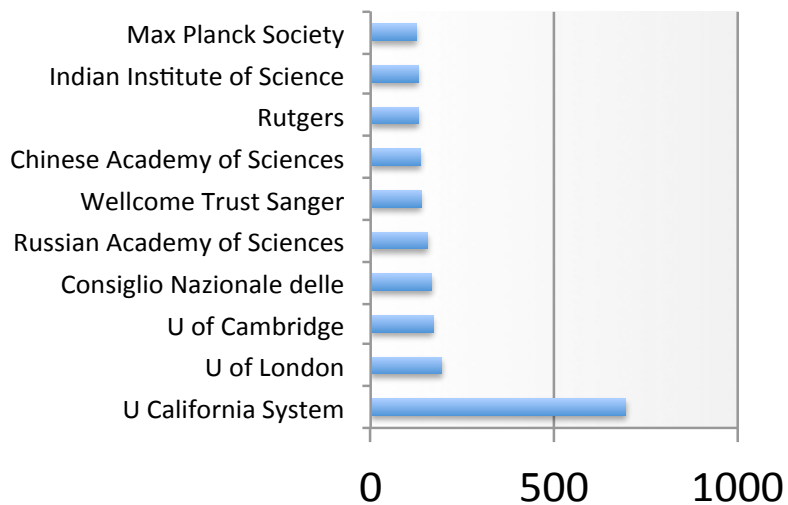
8,215 publication in BIOSIS Citation Index

104 publication in Chinese Science Citation Database

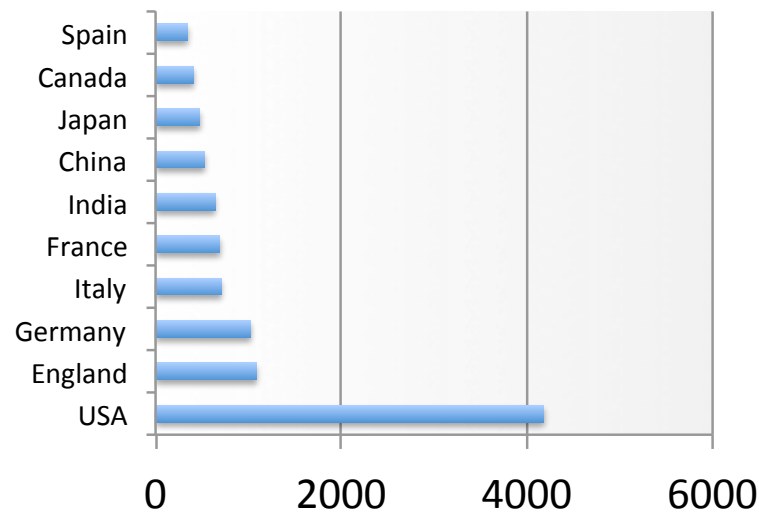
Annual Citations



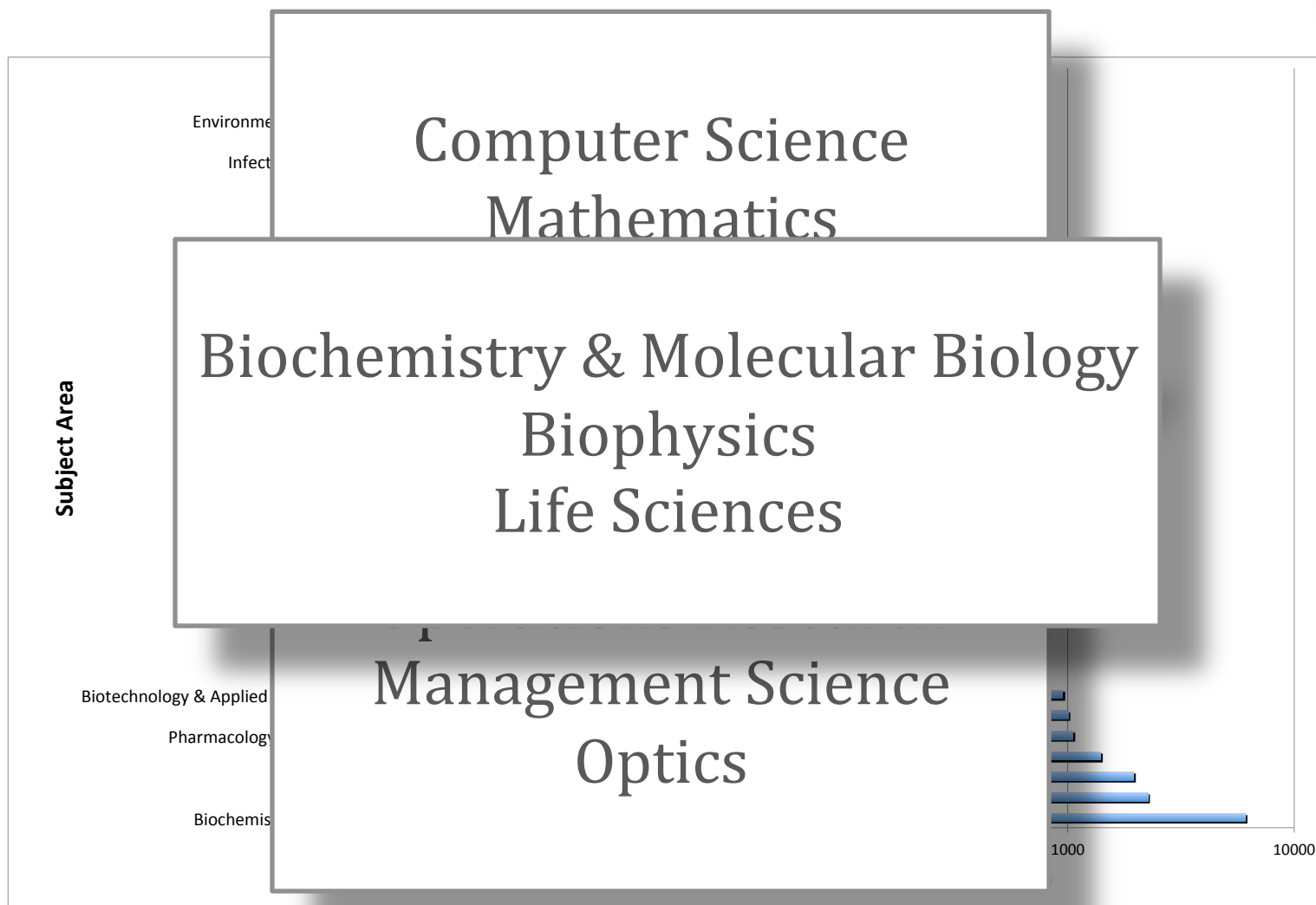
Citations by Institution



Citations by Country



Citations by Subject Area



ORIGINAL ARTICLE

A new germline VHL gene mutation in three patients with apparently sporadic pheochromocytoma

Angela V. D'Elia*, Franco Grimaldi†, Stefano Pizzolitto*, Giovanna De Maglio*, Elisa Bregant*, Nadia Passon*, Alessandra Franzoni*, Antonella Verrienti‡, Giulia Tamburrano‡, Cosimo Durante‡, Sebastiano Filetti‡, Federico Fogolari‡, Diego Russo¶ and Giuseppe Damante*§

Energy Fuels 2010, 24, 1464–1470 · DOI:10.1021/ef901132v
Published on Web 12/30/2009

energy&fuels
article

Toward Greener Carbon Capture Technologies: A Pharmacophore-Based Approach to Predict CO₂ Binding Sites in Proteins

Michael L. Drummond,* Angela K. Wilson, and Thom

Department of Chemistry, Center for Advanced Scientific Computing and Modeling (C
Denton, Texas 76201, USA

Received October 5, 2009; Revised Manuscript Received D

J. Physiol. 37, 668–676 (2001)

MINIREVIEW

ALGAL SENSORY PHOTORECEPTORS¹

Peter Hegemann,² Markus Fuhrmann, and Suneel Kateriya

Institut für Biochemie I, Universität Regensburg, 93040 Regensburg, Germany

Sunlight is the primary energy source for all life on earth, but it also plays an important regulatory role for the growth and development of living organisms. Additionally, light is used as a source of information, which enables the organism to orient and adapt to its steadily changing world. To exploit light as a complex sensory stimulus, algae and higher plants developed sophisticated networks of photoreceptors and sensory pathways that generate appropriate responses. These responses cover a time scale of seconds up to days or even years.

THE PRINCIPAL PHOTORECEPTORS IN NATURE

Nature uses a very limited number of principal or-

also used in the archaeal branch (where they serve as sensory photoreception of the cells in different light rhodopsins) or as light-driven ion transport rhodopsin and halorhodopsin). A contain all-trans,15-S-anti retinal, which sorption undergoes a concerted 180° rotation. The concomitant rotation of the retinal ion movement across the retinal membrane pumping process in bacteriorhodopsin (Kolbe et al. 2000). The function of rhodopsins is surprisingly similar. After retinal binding site is displaced after retinal isomerization: however, the proton is not released but instead

MOLECULAR ECOLOGY

Molecular Ecology (2009) 18, 4997–5017

doi: 10.1111/j.1365-294X.2009.04427.x

INVITED REVIEW

Linking genotypes to phenotypes and fitness: how mechanistic biology can inform molecular ecology

BERS† and PATRICIA M. SCHULTE*

Columbia, 6270 University Blvd, Vancouver, British Columbia, Canada V6T 1Z4
y of Calgary, 2500 University Drive N.W., Calgary, Alberta, Canada T2N 1N4

omic resources, high-throughput molecular technologies and as genome scans have made finding genes contributing to populations an increasingly feasible task. Once candidate

Clinica Chimica Acta 413 (2012) 1605–1611



ELSEVIER

Contents lists available at SciVerse ScienceDirect

Clinica Chimica Acta

journal homepage: www.elsevier.com/locate/clinchim



Characteristics and prevalence of KRAS, BRAF, and PIK3CA mutations in colorectal cancer by high-resolution melting analysis in Taiwanese population

Li-Ling Hsieh^{a,c}, Tze-Kiong Er^{a,c}, Chih-Chieh Chen^d, Jan-Sing Hsieh^e, Jan-Gowth Chang^{a,f}, Ta-Chih Liu^{a,b,f}

- ^a Division of Molecular Diagnostics, Department of Laboratory Medicine, Kaohsiung Medical University Hospital, Kaohsiung, Taiwan
- ^b Division of Hematology and Oncology, Department of Internal Medicine, Kaohsiung Medical University Hospital, Kaohsiung, Taiwan
- ^c Graduate Institute of Medicine, College of Medicine, Kaohsiung Medical University, Kaohsiung, Taiwan
- ^d Institute of Bioinformatics and Systems Biology, National Chiao Tung University, Hsinchu, Taiwan
- ^e Department of Surgery, Kaohsiung Medical University Hospital, Kaohsiung, Taiwan
- ^f Institute of Clinical Medicine, College of Medicine, Kaohsiung Medical University, Kaohsiung, Taiwan

ARTICLE INFO

Article history:
Received 17 April 2012
Received in revised form 27 April 2012
Accepted 28 April 2012
Available online 8 May 2012

Keywords:
Colorectal cancer
KRAS gene
BRAF gene

ABSTRACT

Background: The identification of KRAS, BRAF, and PIK3CA mutations before the administration of anti-epidermal growth factor receptor therapy of colorectal cancer has become important. The aim of the present study was to investigate the occurrence of KRAS, BRAF, and PIK3CA mutations in the Taiwanese population with colorectal cancer. This study was undertaken to identify BRAF and PIK3CA mutations in patients with colorectal cancer by high-resolution melting (HRM) analysis. HRM analysis is a new gene scan tool that quickly performs the PCR and identifies sequence alterations without requiring post-PCR treatment.

Methods: In the present study, DNAs were extracted from 182 cases of formalin-fixed, paraffin-embedded (FFPE) colorectal cancer samples for clinical KRAS mutational analysis by direct sequencing. All the samples were also tested for mutations within BRAF V600E and PIK3CA (exons 9 and 20) by HRM analysis.

Management

- Director, Helen M. Berman
 - Overall direction of RCSB PDB
- Deputy Director, Martha Quesada
 - Facilitation of wwPDB initiatives
- Associate Director, Philip E. Bourne
 - Direction of UCSD site
- Associate Director, Stephen K. Burley
 - Direction of Outreach & Education



Advisory Groups

Advisory Committees

- RCSB PDB AC
 - Cynthia Wolberger (Johns Hopkins/HHMI)
- wwPDB AC
 - Soichi Wakatsuki (Stanford)

Working Groups

- PDBx/mmCIF WG
 - Paul Adams (LBL)

Task Forces

- X-ray Validation TF
 - Randy Read (Univ of Cambridge)
- NMR Validation TF
 - Gaetano Montelione (Rutgers)
 - Michael Nilges (Institut Pasteur)
- 3DEM Validation TF
 - Richard Henderson (MRC-LMB)
 - Andrej Sali (UCSF)
- Small-Angle Scattering TF
 - Jill Trehwella (Univ Sydney)
- Hybrids Methods TF, *TBD*

Agenda

Overview Helen Berman

**Outreach &
Education** Christine Zardecki
Shuchismita Dutta

Data In Jasmine Young
John Westbrook

Data Out Peter Rose
Andreas Prlić



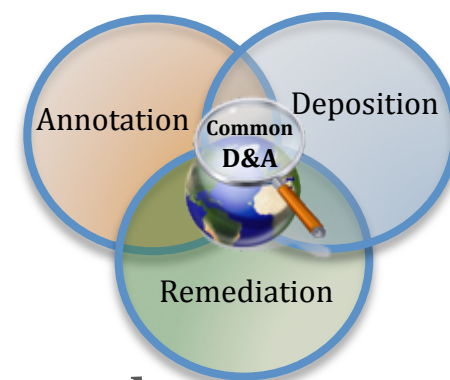
Data In: Data Deposition, Processing, and Remediation

Jasmine Young

John Westbrook



RCSB PDB provides a Structural View of Biology that enables understanding of biological functions and processes at the molecular level in 3D



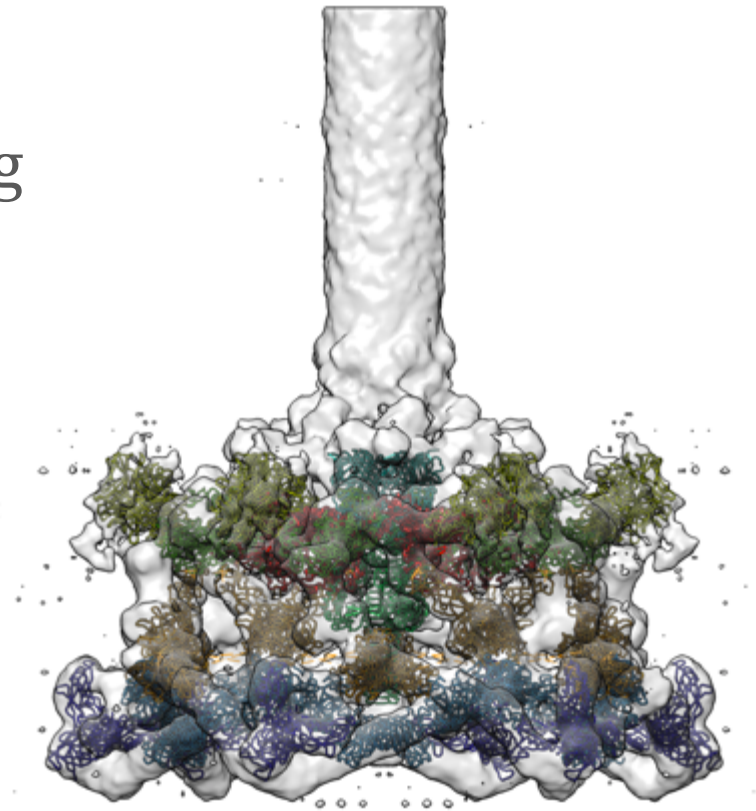
Data In

- Captures experimental data that defines the 3D structure of macromolecules
- Supports a diverse user base by maximizing the quality and completeness of the data files
- Provides leadership in data standardization, representation and validation



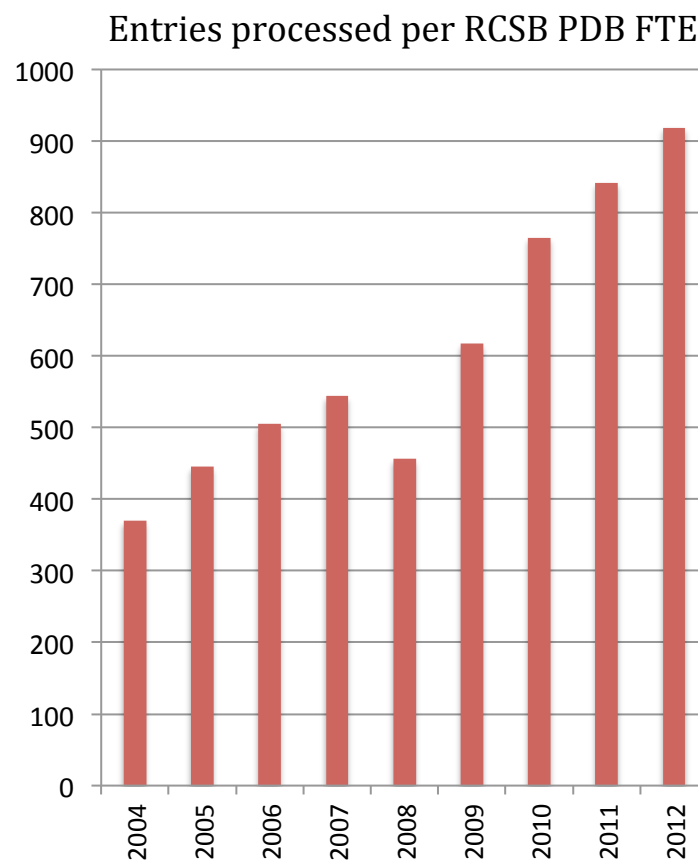
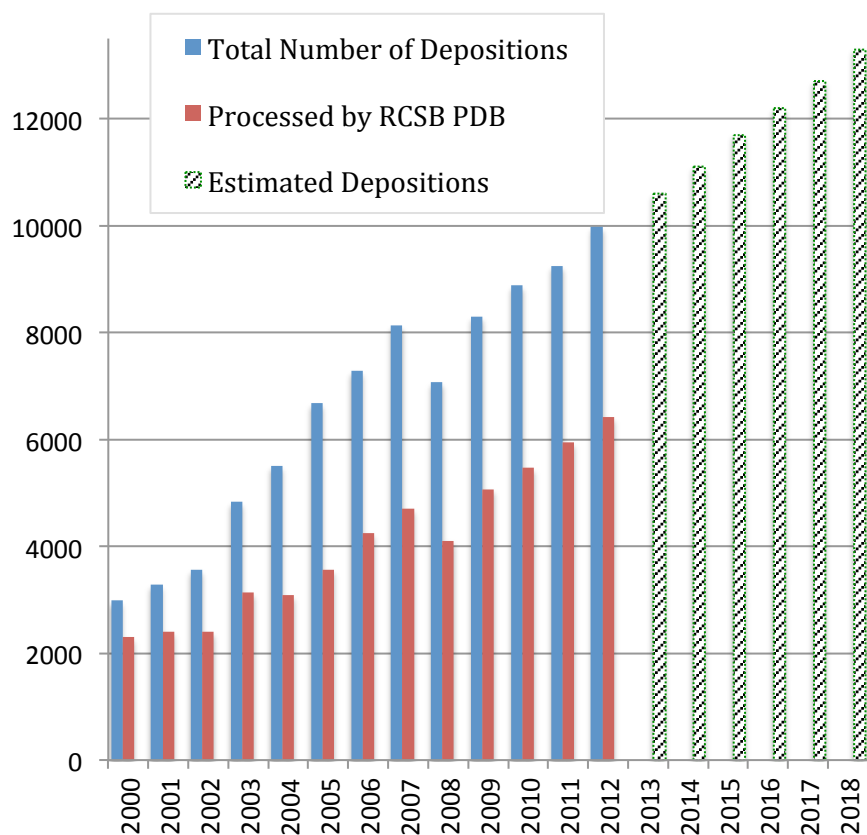
Five Year Plan

- Significantly increase processing throughput and data quality with current staffing
- Process increasingly complex structures determined by new and hybrid methods
- Enhance query capabilities
- Continue to standardize, harmonize and remediate files
- Enhance annotation through development of **External Reference Files (ERFs)**



EMD-1048; PDB IDs 1pdf, 1pdi, 1pdj, 1pdl, 1pdm, 1pdp: Three-dimensional structure of bacteriophage T4 baseplate. VA Kostyuchenko *et al. Nat. Struct. Biol.* 10, 688-93 (2003); PDB ID 2fl8: Evolution of bacteriophage tails: Structure of T4 gene product 10. PG Leiman *et al. J. Mol. Biol.* 358, 912-21 (2006); PDB ID 3h3w: The structure of gene product 6 of bacteriophage T4, the hinge-pin of the baseplate. AA Aksyuk *et al. Structure* 17, 800-808 (2009)

Increasing Productivity



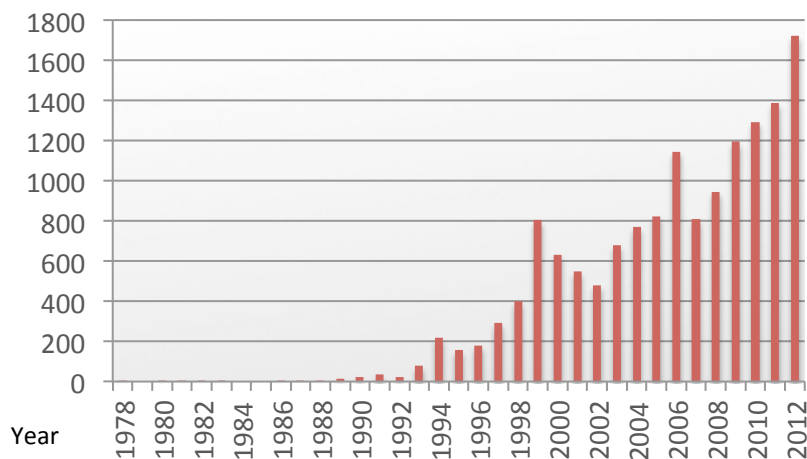
RCSB PDB annotated 64% of all depositions in 2012

Processing productivity at RCSB PDB doubled since 2008, while global deposition increased 40%

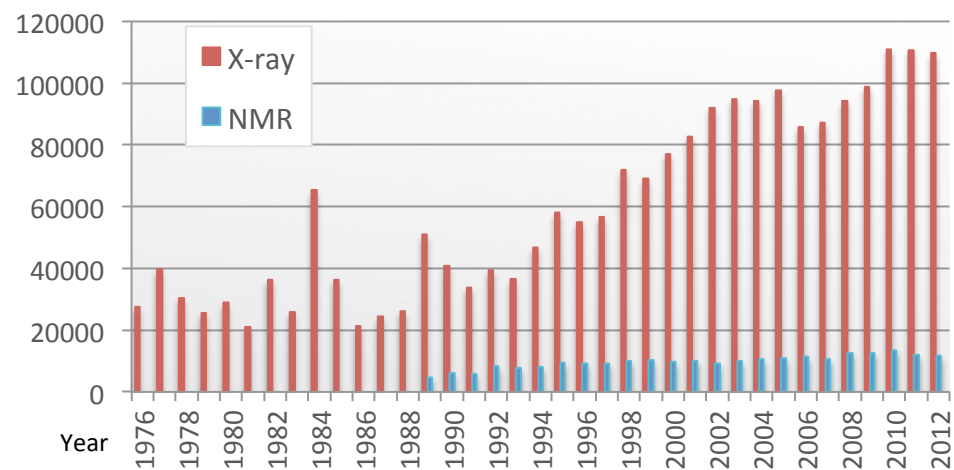


Increasing Complexity

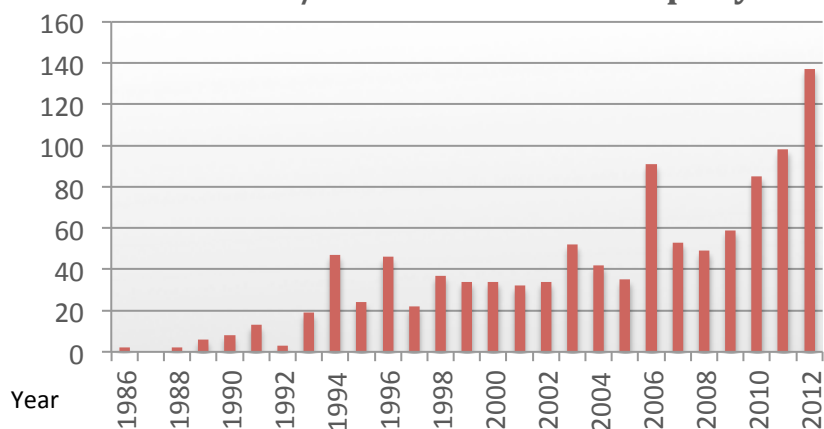
Number of ligands released per year



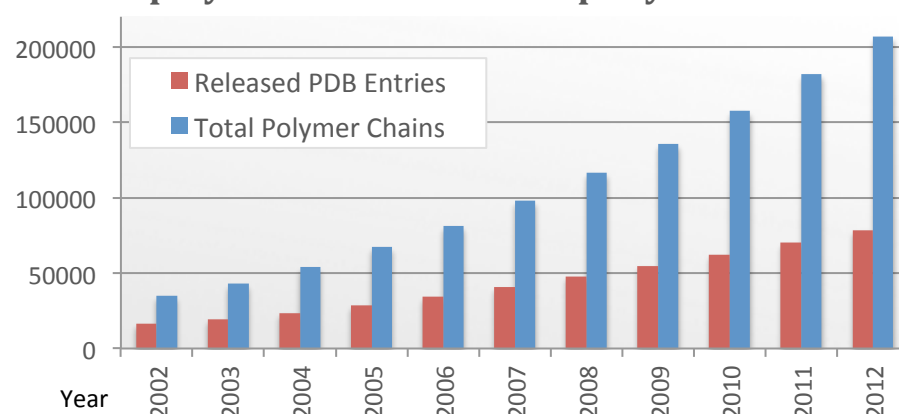
Average Molecular Weight released per year



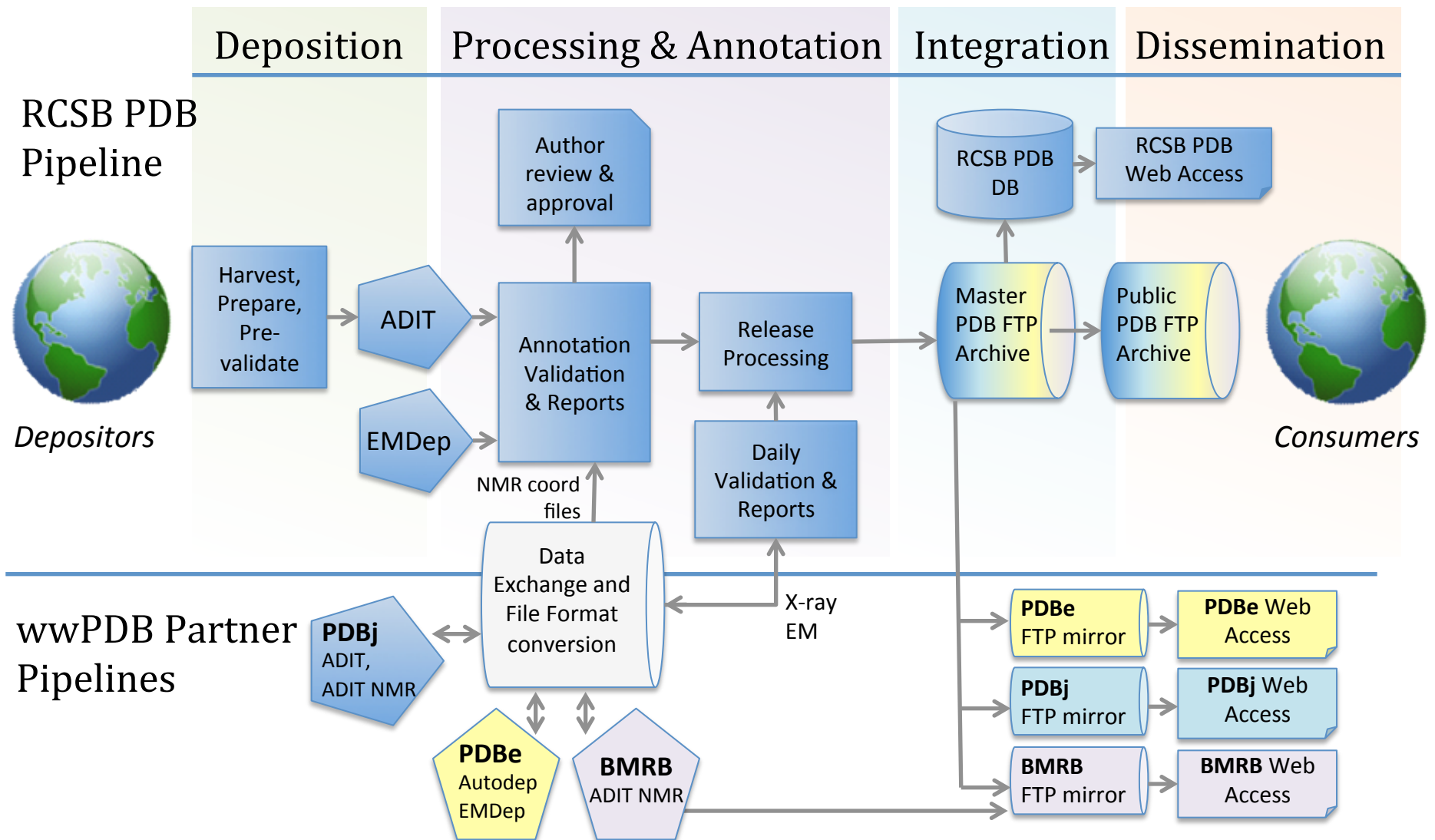
Number of entries with peptide-like inhibitors/antibiotics released per year



Number of entries and total number of polymer chains released per year



“Retiring” Data Deposition & Annotation Pipelines



Legacy System Was Not Scalable!

Challenges:

1. Larger and more complex biological molecules that stretch the limits of the PDB file format are being deposited to the archive
2. Data from hybrid experimental methods are difficult to represent in the PDB file format
3. Two different wwPDB data processing pipelines require complex data exchange and harmonization
4. Workload across wwPDB sites is not geographically balanced

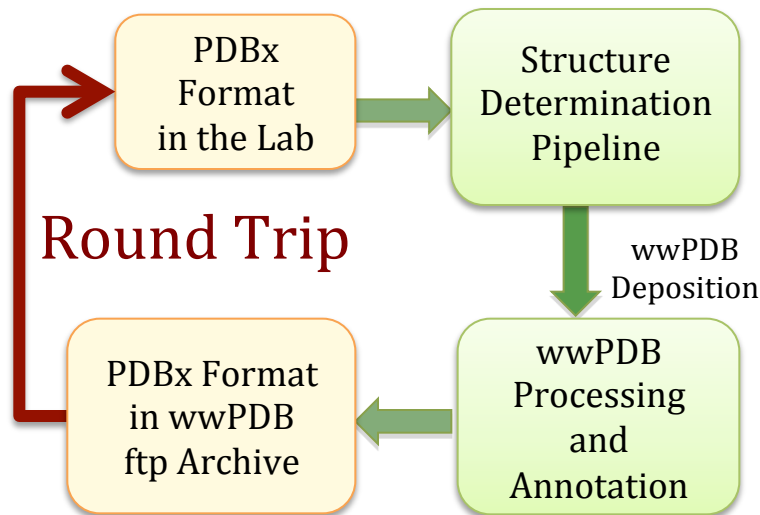


Addressing Large Structures and Hybrid Methods (Challenges 1 & 2)



Workshop Participants, September 2011

- Working group charged with finding “round trip” single format that can handle complex data not supported by the PDB file format
- Consensus reached on using stylized PDBx format
 - Follows dictionary-driven macromolecular Crystallographic Information Format (mmCIF) syntax
 - Many supporting tools available
 - Supports large and complex structures
 - Supports new and hybrid methods
 - Preserves simple style and readability of PDB format



Now embedded in major X-ray software packages (CCP4/REFMAC and Phenix)

First Non-split PDBx Large Structure

30-May-2013

Landmark HIV Capsid Structures Released in PDB

Two complete HIV-capsid structures, both of unprecedented size, are **described in this week's issue of Nature** and released in the Protein Data Bank (PDB; wwpdb.org). This represents a significant advance in the field of structural biology and a milestone for the PDB.

PDB entries 3J3Q and 3J3Y are models based on cryo-electron microscopy data and use of a molecular dynamics flexible-fitting method. They contain 1356 and 1176 protein chains, respectively, and over two million atoms each. The HIV-1 capsid is the protein envelope that encloses and protects the RNA genome of the virus. An important subject of study, the full capsid has been a difficult target for structural characterization due to its extremely large size and morphological variability.

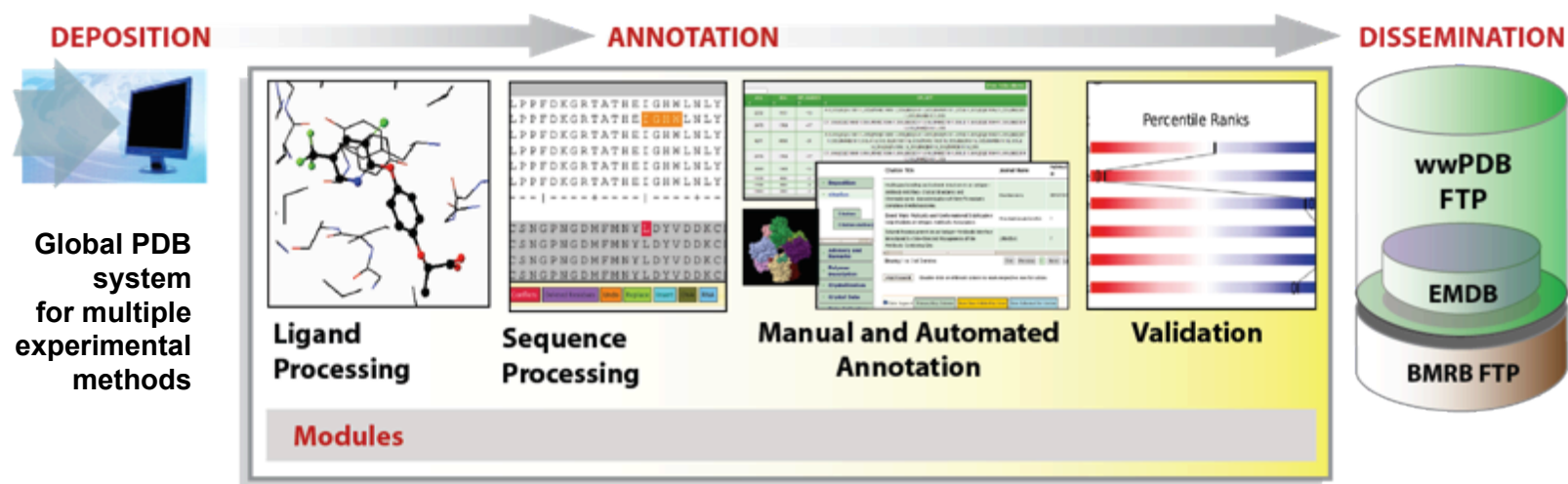


The wwPDB has anticipated structures of increasing size and complexity that exceed the limitations of the original PDB file format. These capsid structures have been curated following the **recently announced wwPDB procedures for the deposition and release of large structures**. Extremely large structures can now be deposited, annotated, and released as single files in PDBx/mmCIF and PDBML/XML formats.



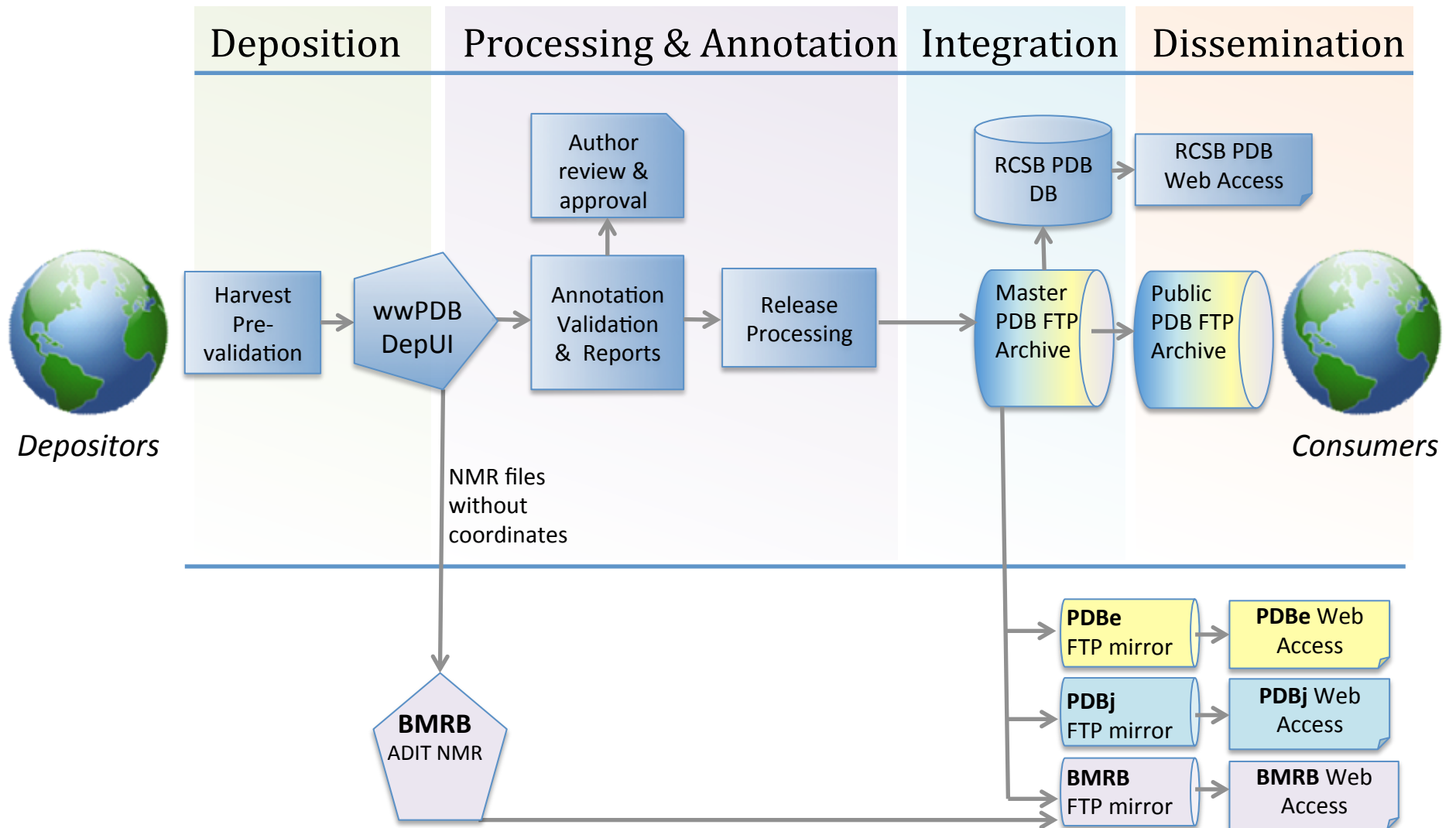
Addressing Pipeline and Workflow Issues:

A Unified wwPDB Common Deposition & Annotation System (Challenges 3 & 4)



- PDBx/mmCIF is the master file format
- System supports all experimental methods
- Validation suites based on recommendations from community Task Forces; X-ray is implemented
- Better checking for ligand chemistry and polymer sequence
- Enables workload balancing and increased productivity

New wwPDB Data Deposition & Annotation Pipeline



New Deposition Interface

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

WORLDWIDE PDB PROTEIN DATA BANK wwPDB Deposition Tool

Existing deposition

Deposition ID ⓘ

Password ⓘ

Log in

E-mail ⓘ

Preferred deposition site ⓘ

Location ⓘ

Experimental Method

X-Ray Diffraction

Electron Microscopy

Solution NMR

Neutron Diffraction

Electron Crystallography

Solid-state NMR

Solution Scattering

Fiber Diffraction

Requested accession codes ⓘ

PDB

EMDB

BMRB

Related depositions ⓘ

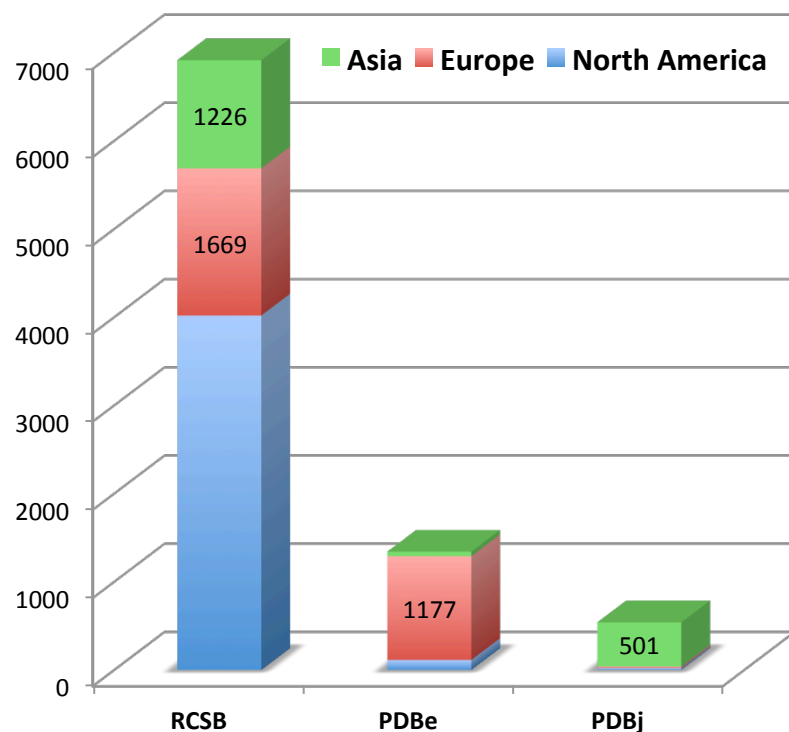
Structural genomics ⓘ

Start deposition

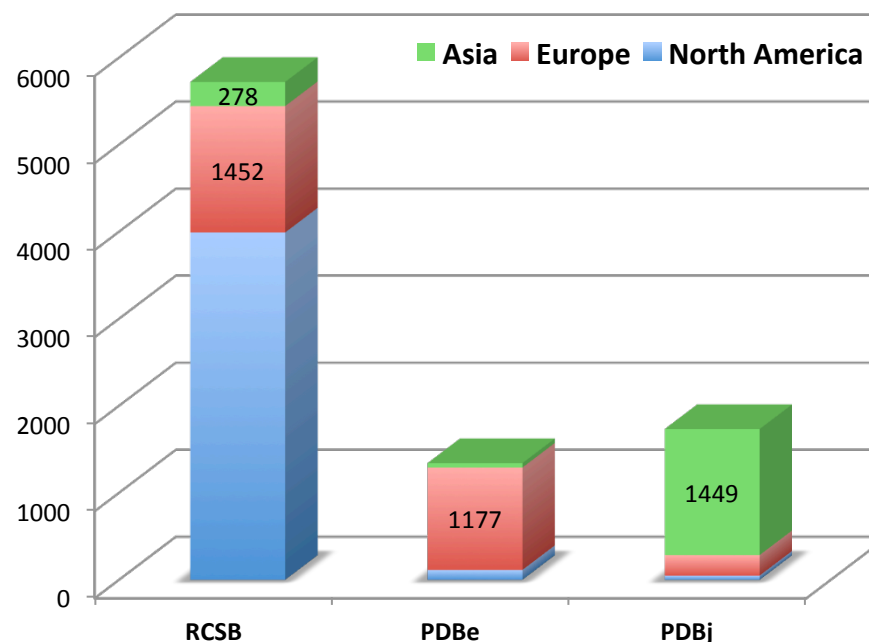
Multiple methods

2012 Workload Balancing

Deposition distribution



Processing distribution

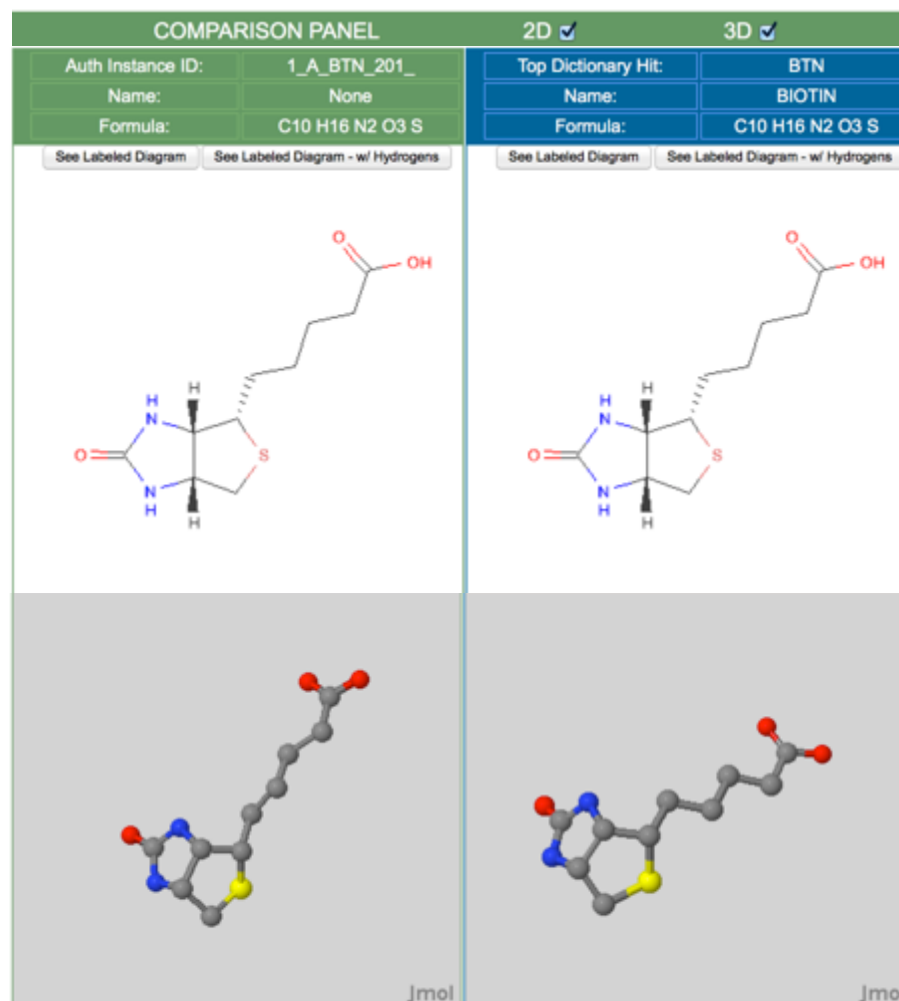


New D&A system enables workload balancing across all 3 sites
Only possible between RCSB PDB and PDBj in legacy system

Ligand Module: Checks Against Chemical Component and BIRD Dictionaries

- Batch search against CCD/BIRD and automated ligand ID assignment
- Captures and displays author-provided chemical information
- Comparison Panel
 - 2D and 3D views of ligand for review
 - ID assignment

Currently in production
at RCSB PDB



Author Coordinates (left), Closest Dictionary Match (right)

Sequence Module: Consistency Checks

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

Jmol ALA/GLY

```

AUTH PDB:R V(1)  VVVQAPTQVP GFLGDSVTLPCYLQVPNMEV
XYZ PDB:R V(1)  .VVQAPTQVP GFLGDSVTLPCYLQVPNMEV
UNP:P15151 (R1,V1) VVVQAPTQVP GFLGDSVTLPCYLQVPNMEV
1 | ---+--- | ---+--- | ---+--- |

AUTH PDB:R V(1)  SKRLEFVAARLGAE LR DASLRMFGLRVEDE
XYZ PDB:R V(1)  SKRLEFVAARLGAE LR RASLRMFGLRVEDE
UNP:P15151 (R1,V1) SKRLEFVAARLGAE LR NASLRMFGLRVEDE
61 | ---+--- | ---+--- | ---+--- |

AUTH PDB:R V(1)  AEVQKVQLTGE PVPMARCVSTGGRPPAQIT
XYZ PDB:R V(1)  AEVQKVQLTGE PVPMARCVSTGGRPPAQIT
UNP:P15151 (R1,V1) AEVQKVQLTGE PVPMARCVSTGGRPPAQIT
121 | ---+--- | ---+--- | ---+--- |

AUTH PDB:R V(1)  VPSSQVDGK QVTCKVEHESFEKPQLLTV S
XYZ PDB:R V(1)  VPSSQVDGK QVTCKVEHESFEKPQLLTV S
UNP:P15151 (R1,V1) VPSSQVDGK NVTCKVEHESFEKPQLLTV N
181 | ---+--- | ---+--- | ---+--- |
                
```

3D Viewer

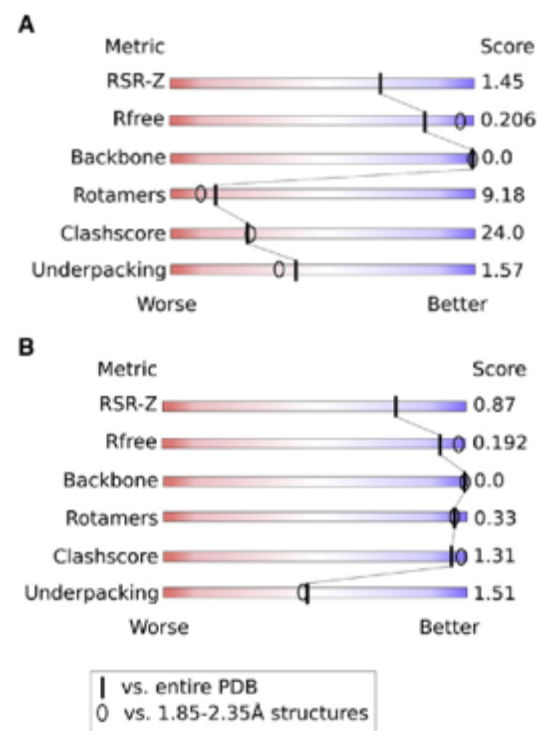
POSITION	AUTH PDB:R	ALIGNED SEQUENCE	RESIDUE	ANNOTATION DETAILS
77	ASP	UNP:P15151 (R1,V1)	ASN	engineered mutation cloning artifact variant
92	SER	UNP:P15151 (R1,V1)	ASN	expression tags
160	GLN	UNP:P15151 (R1,V1)	ASN	insertion deletion microheterogeneity chromophore linker
190	GLN	UNP:P15151 (R1,V1)	ASN	conflict acetylation amidation
209	SER	UNP:P15151 (R1,V1)	ASN	initiating methionine

X-ray Validation

X-ray Validation Task Force (VTF):
Developed consensus recommendations on
data and structure validation

- 2008 Workshop
- Chair: Randy Read (Univ. of Cambridge)
- *Structure* 19, 1395-1412 (2011)

**New X-ray validation reports are
embedded in the current system**

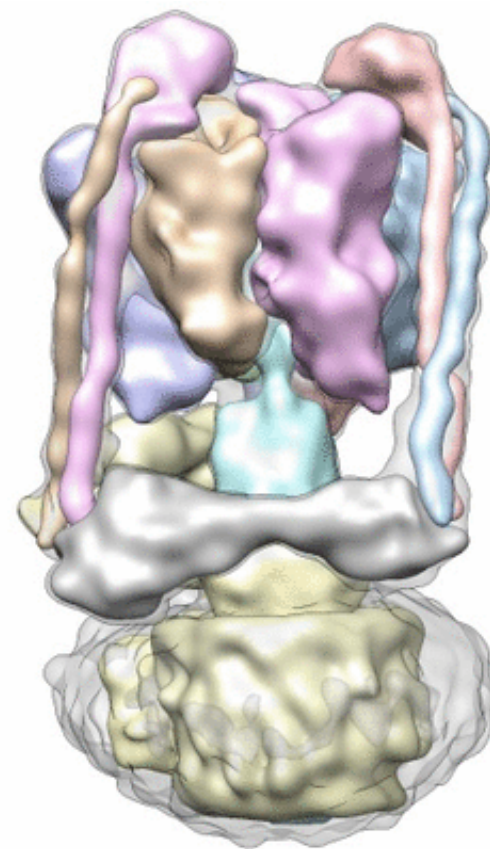


Validation Module Checks	Software Used
Protein geometry	MolProbity (Chen <i>et al. Acta Cryst.</i> D66, 12-21 (2010))
Ligand geometry	Mogul (IJ Bruno <i>et al. J. Chem. Inf. Comput. Sci.</i> 44, 2133-2144 (2004))
DNA validation	NUCheck (sw-tools.rcsb.org)
Crystallographic symmetry clashes	Maxit (sw-tools.rcsb.org)
Ligand stereochemistry & assignment	wwPDB D&A Ligand Module
Sequence	wwPDB D&A Sequence Module
Structure factor validation	SF-Tool (sw-tools.rcsb.org), Refmac (AA Vagin <i>et al. Acta Cryst.</i> D60, 2284-229 (2004)), Phenix/Xtriage (PD Adams <i>et al. Acta Cryst.</i> D66, 213-221 (2010)), EDS/Mapman (GJ Kleywegt <i>et al. Acta Cryst.</i> D60, 2240-2249 (2004))

3DEM

Collaboration with EMDataBank

- EM maps are annotated by PDB and part of PDB archive since March 2012
- 3DEM data items for wwPDB common deposition system developed
- EM Validation: EM VTF recommendations published *Structure* 20, 205-214 (2012)

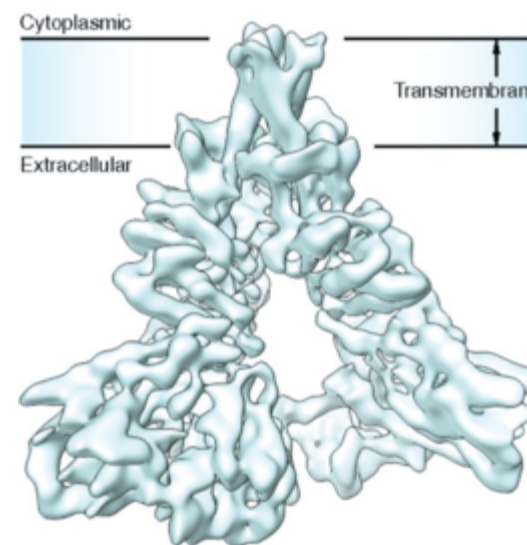


EMD-5476 ATPase from *S. cerevisiae*
S. Benlekbir *et al.* *Nature Structural & Molecular Biology* 19, 1356-1362 (2012)

3DEM: Importance of Validation

- Recent public commentaries highlight pressing need for 3DEM structure validation
- 3DEM validation pipeline under development, following EM VTF recommendations
 - 1st version will evaluate model quality using X-ray criteria
 - Validation methods for **maps** and **map-model fit**, in testing, with the 3DEM community

- J. Cohen, Is High-Tech View of HIV Too Good to Be True? *Science* 341, 443-444 (2013)
- R.M. Glaeser, Replication and validation of cryo-EM structures *J. Struct. Biol.*, in press (2013)

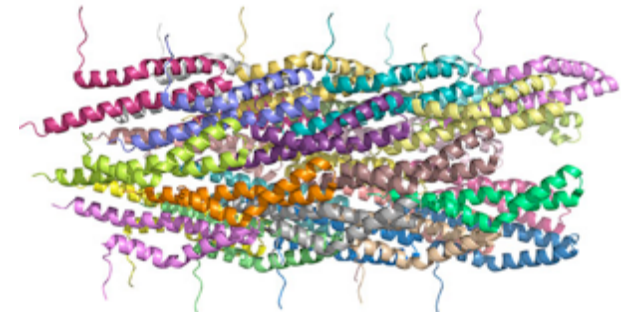
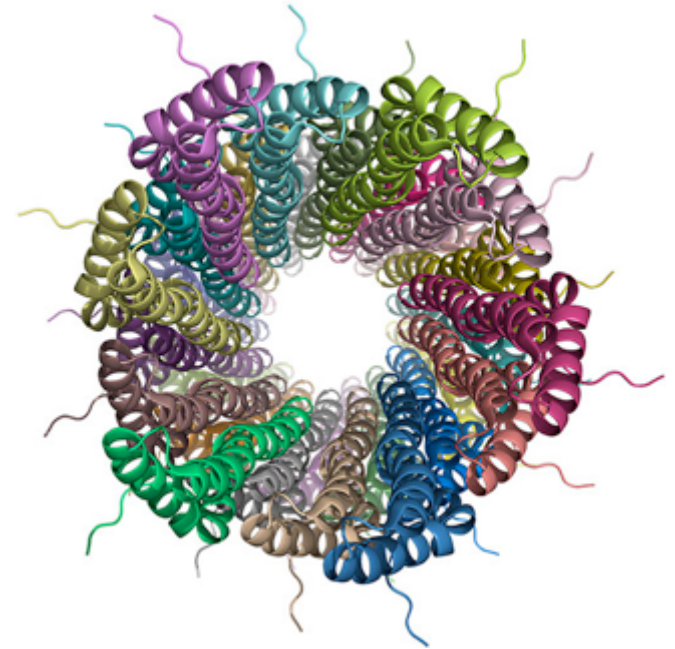


EMD-5418 JG Sodroski *et al.* Molecular architecture of the uncleaved HIV-1 envelope glycoprotein trimer *PNAS* 110, 12438-12443 (2013)

NMR



- NMR data items for wwPDB common deposition system have been developed in collaboration with BMRB
- NMR Validation
 - New validation standards recommended by NMR VTF *Structure* 21, 1563-1570 (2013)
 - NMR validation module incorporating NMR VTF recommendations under development



PDB ID 2lpz; BMRB entry 18276. A Loquet *et al.*, Structure of a bacterial type III secretion needle from *Salmonella typhimurium*. *Nature* 486, 276-279 (2012); determined using a hybrid solid-state NMR and electron microscopy (EM) approach.

Version 1.0 Testing and Deployment

- Deposition and annotation pipelines now being tested
- Both deposition systems available in parallel during transition
- January 2014
 - All incoming depositions will be processed using the new system
 - Depositions initiated in old system will have time to complete

Tasks	Date
PDBx working format announcement	May 22, 2013
D&A systems announcement	Jul 1, 2013
D&A public alpha testing by X-ray software developers	Jul 1, 2013
Produce new validation reports for depositors once entry is fully processed	Aug 1, 2013
D&A public beta testing by selected X-ray depositors	Sep 16, 2013
D&A in production at all sites	Jan 1, 2014
Old deposition systems operating in parallel	Jan 1, 2014



Data In-Data Out Synergies



Data quality (*data processing*)
Data standardization (*remediation*)
Extended annotation (*ERFs*)



Improved query functionality
Extended query options

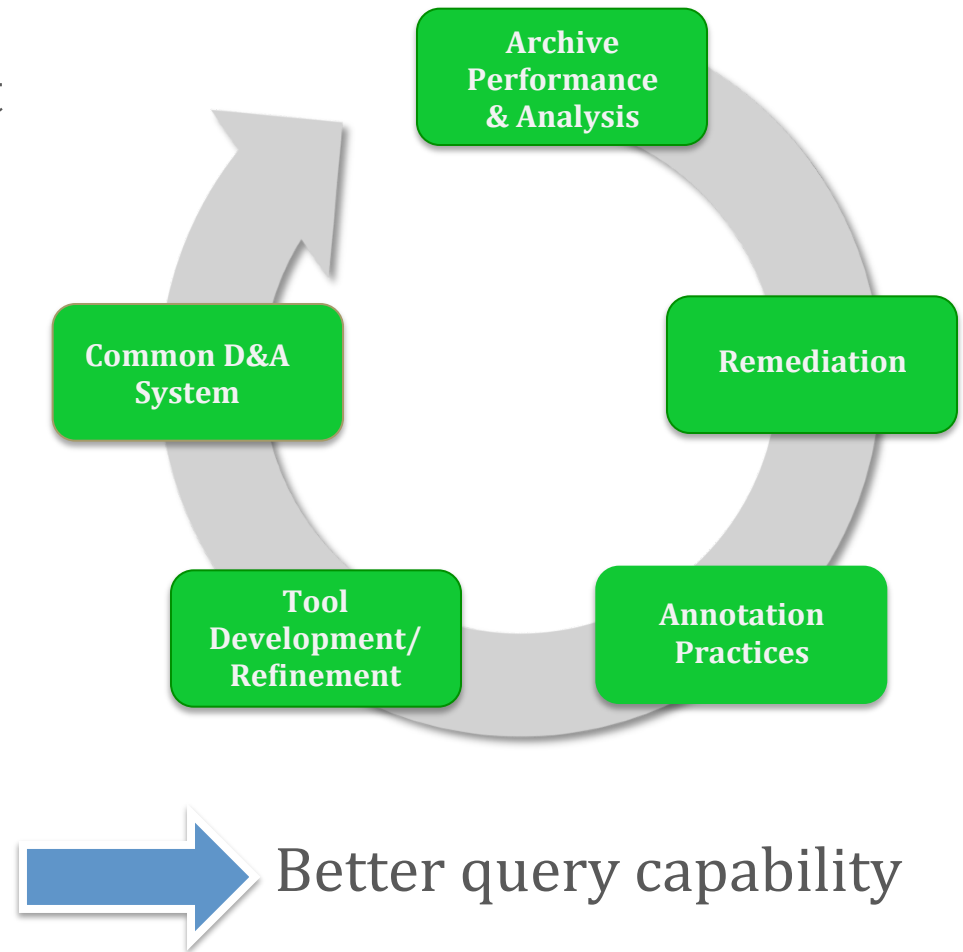


Remediating Data for a Uniform Archive

wwPDB regularly reviews the archive to identify and correct errors and inconsistencies (“remediation”)

These efforts:

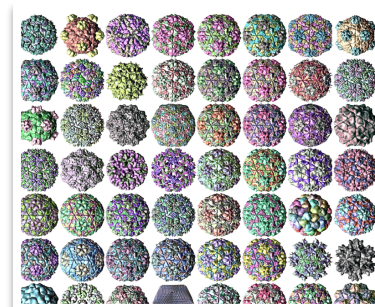
- Inform all processes
- Improve consistency in entry and archive annotation
- Enhance chemistry representation



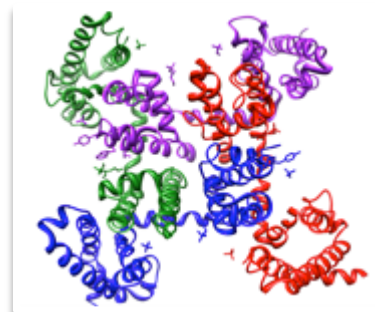
Previous Remediations

wwpdb.org/remediation.html

- 2007: Sequences, nomenclature, citation and virus representation
Nucleic Acids Research 36, 426-433 (2008)
- 2009: Biological assemblies and binding sites of ligands and metal ions
- 2011: Representation of biological assemblies, residual B factors, and peptide inhibitors and antibiotics
(manuscript in preparation)

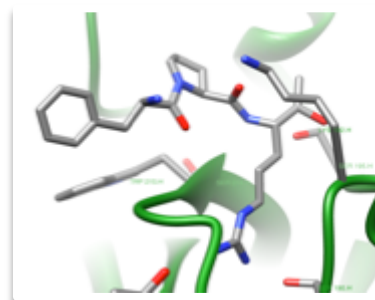


Representation of viruses in the remediated PDB archive, CL Lawson *et al. Acta Cryst. D* 64, 874-882 (2008)

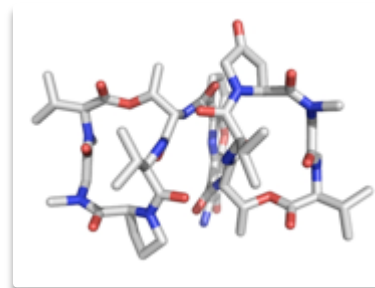


PDB ID: 4ekw, Ion Channel, J Payandeh *et al. Nature* 486, 135-139 (2012)

PISA: E Krissinel and K Henrick, *J. Mol. Biol.* 372, 774-797 (2007)



PDB ID: 2fir, PPACK inhibitor binding site, SP Bajaj *et al., J.Biol.Chem.* 281, 24873-24888 (2006)

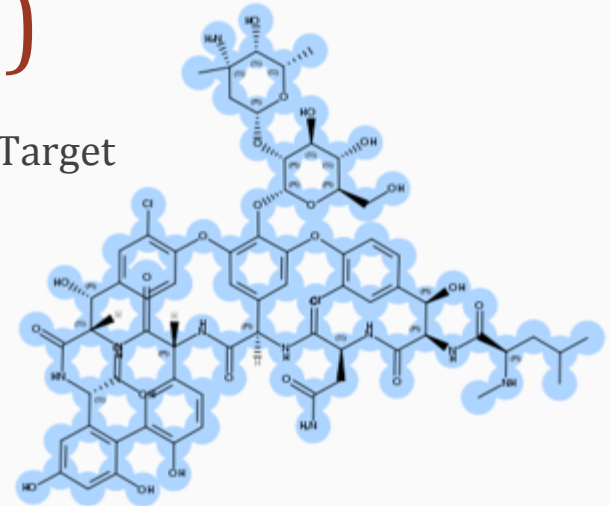


PDB ID: 1qfi, Actinomycin, A Lifferth *et al., Z.Naturforsch* 54, 681-691 (1999)

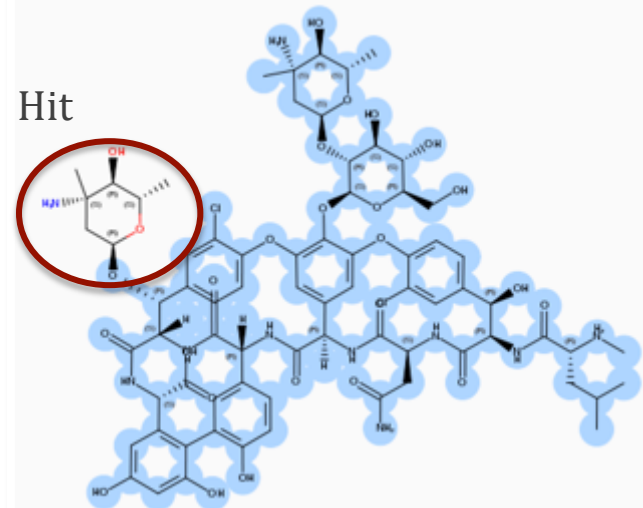
Biologically Interesting Molecule Reference Dictionary (BIRD)

- Developed as a result of the remediation of peptide-like inhibitors and antibiotics
- Contains information about the chemistry, biology, and structure of these molecules
- Integrated in current and future D&A annotation systems
- Example of an ERF

Target



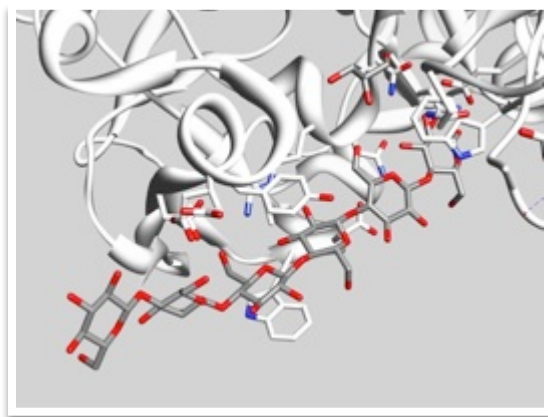
Hit



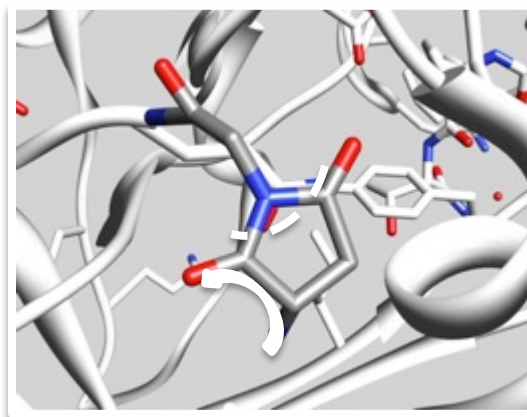
Future Remediation

Current reviews are focused on representation of

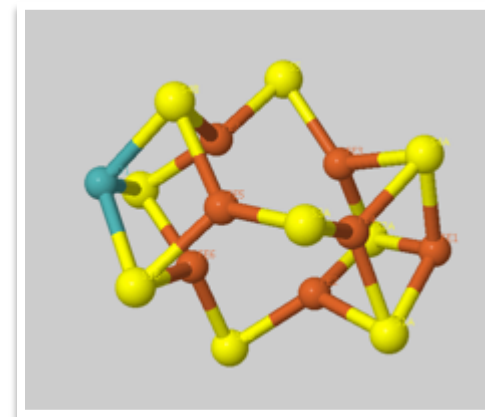
- Carbohydrates
- Protein modifications
- Metal-containing ligands



Carbohydrates



Protein modifications

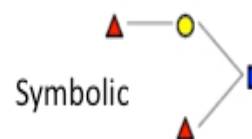
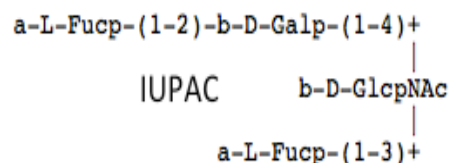


Metal-containing ligands

Carbohydrate Remediation

Issues	<ul style="list-style-type: none"> Multiple representations in naming and linking Non-standard nomenclature and incomplete linkages Representation of branched polymers
Goals	<ul style="list-style-type: none"> Represent data consistently within the archive and in agreement with glycobiology community standards Enable searches for carbohydrates in PDB archive
Plans Forward	<ul style="list-style-type: none"> Identify and analyze carbohydrate-containing entries Create standard representation for branched polymers Incorporate standard nomenclature Develop a strategy for remediation

LINUCS: `[][b-D-GlcpNAc]{[(3+1)][a-L-Fucp]{][(4+1)][b-D-Galp]{[(2+1)][a-L-Fucp]{}}`



PDB ID: 2wmg MA Higgins *et al. J.Biol.Chem.* 284, 26161-26173 (2009)

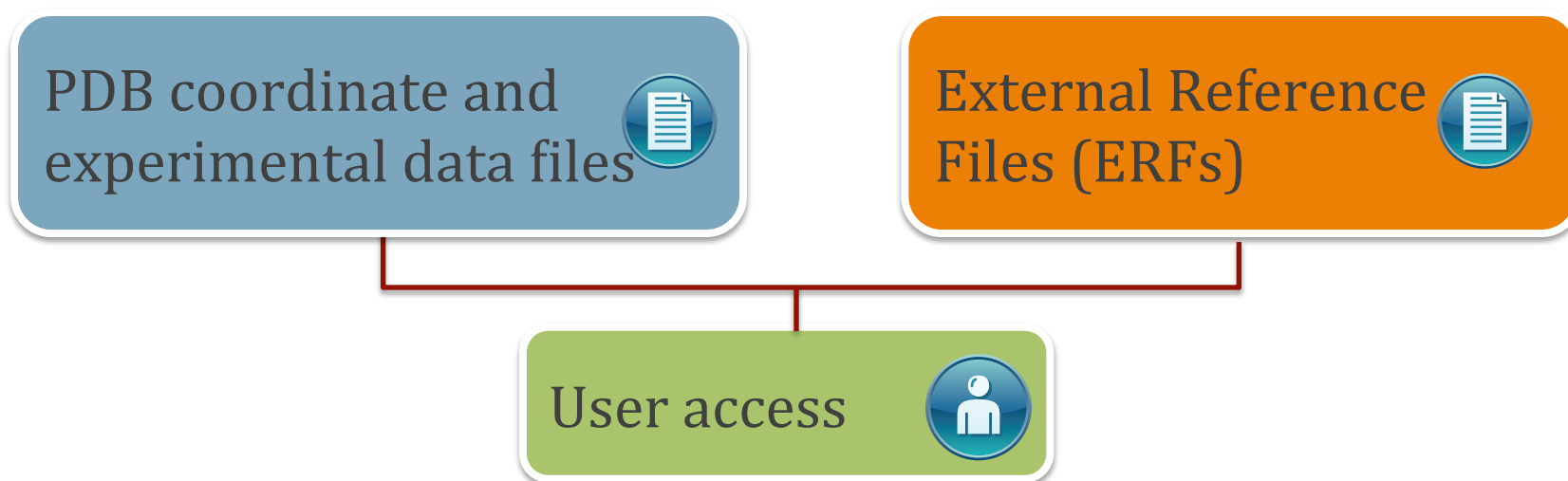
Data Enrichment Beyond Remediation

- Archive-wide remediation is limited to standardizing terminology, nomenclature and well-established annotations
- Some data within existing PDB entries constantly in flux (*e.g.*, DBREF)
- Some data required to support richer queries and analysis are not readily incorporated into PDB entries



What Do We Mean by an External Reference File (ERF)?

- A data file containing compilations of additional annotations on structure entries
- Can be extended as necessary without impacting archival data files



Current ERFs

- Chemical Reference Data (RCSB PDB/wwPDB)
 - Chemical Component Dictionary (CCD; ~18K definitions)
 - Biologically Interesting molecule Reference Dictionary (BIRD; ~1K definitions)
- Structure Integration with Function, Taxonomy and Sequence (SIFTS; PDBe and EMBL-EBI)
 - Weekly updates of mappings among PDB, UniProt, IntEnz, GO, Pfam, InterPro, SCOP, CATH



ERF Content Example

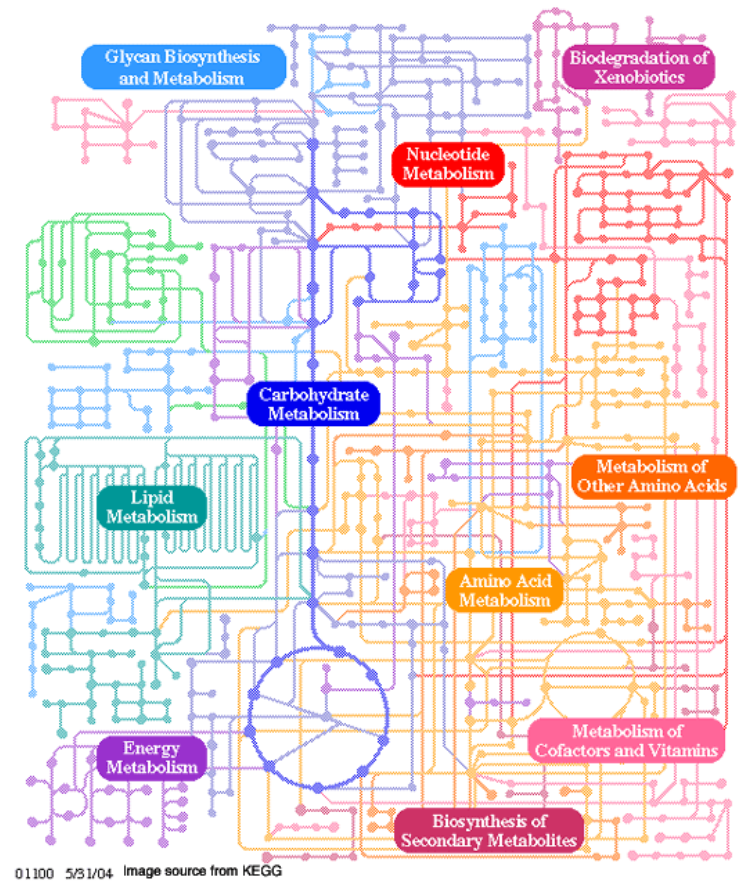
Chemical Component Dictionary (CCD)

- Molecular names and synonyms
- Chemical formula, formula weight, and formal charge
- Atom and residue nomenclature
- Polymer linking type
- Model coordinates from a PDB entry
- Computed coordinates (Corina or OpenEye)
- Connectivity and bond types
- Stereochemistry and aromaticity
- Systematic names (ACDLabs and OpenEye)
- SMILES, InChI, and InChIKey descriptors
- Release status and revision history



Future ERFs

- New keywords for structure and function
- Lists of PDB entries containing special features
 - Pharmaceutical targets
 - Membrane proteins
 - Ligand binding
 - Protein-protein and protein-nucleic acid interactions
- New mapping of data items within PDB entries to other data resources
 - Sequence and domain beyond SIFTS (*e.g.*, RefSeq)
 - Ontologies and classifications (*e.g.*, PRO)
 - Biological pathways (*e.g.*, BioCyc)
- Computed and mined annotations
 - Sequence clusters
 - Structure clusters
 - Crystallization conditions



PDB Data vs. ERF Content

- Primary data in PDB entries
 - Coordinates, supporting experimental data, sample preparation, sequence and taxonomy, chemical nomenclature, experimental metadata, secondary and tertiary structure features, citation, external database references, ...

- External Reference File content
 - Chemical reference data (*e.g.*, CCD, BIRD)
 - New keywords
 - Lists of PDB entries containing special features
 - New mappings
 - Computed and mined annotations

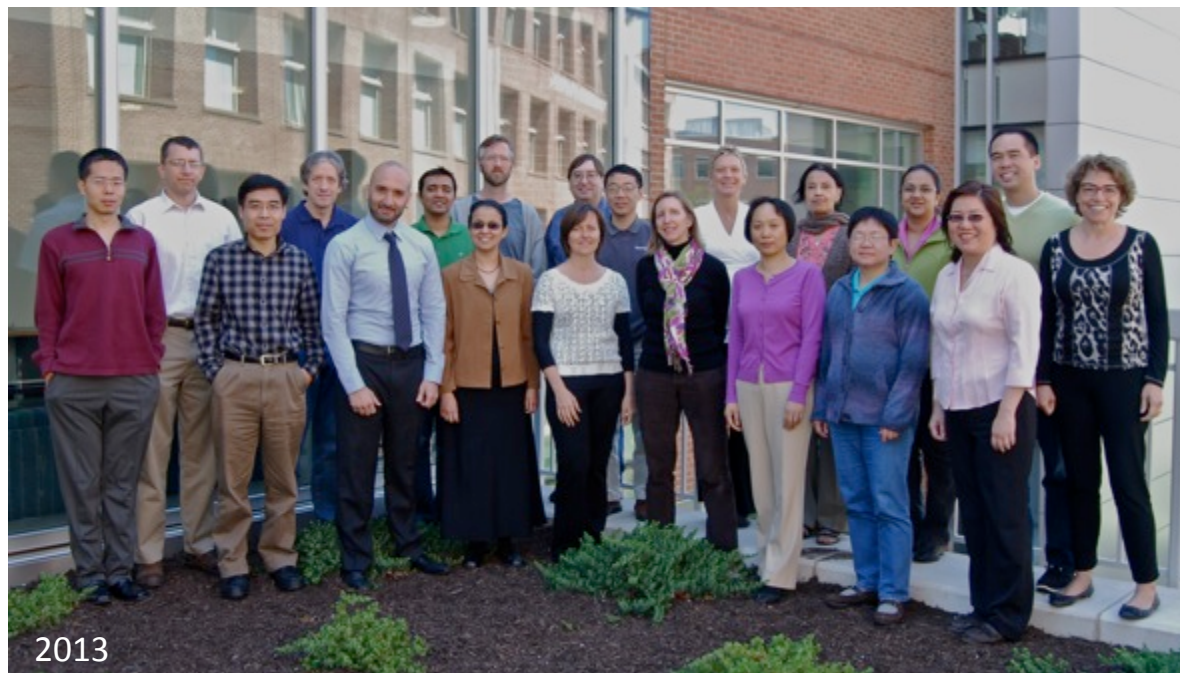


Implementation Strategy for ERFs

- Leverage RCSB PDB staff expertise to manually and computationally develop ERF content
- Work with community experts
- Package and organize data to support various delivery options
- Document provenance
- Provide ERF catalog



RCSB PDB *Data In* Team



wwPDB D&A Team



Data Out: Data Distribution, Query, and Analysis

Peter Rose

Andreas Prlić



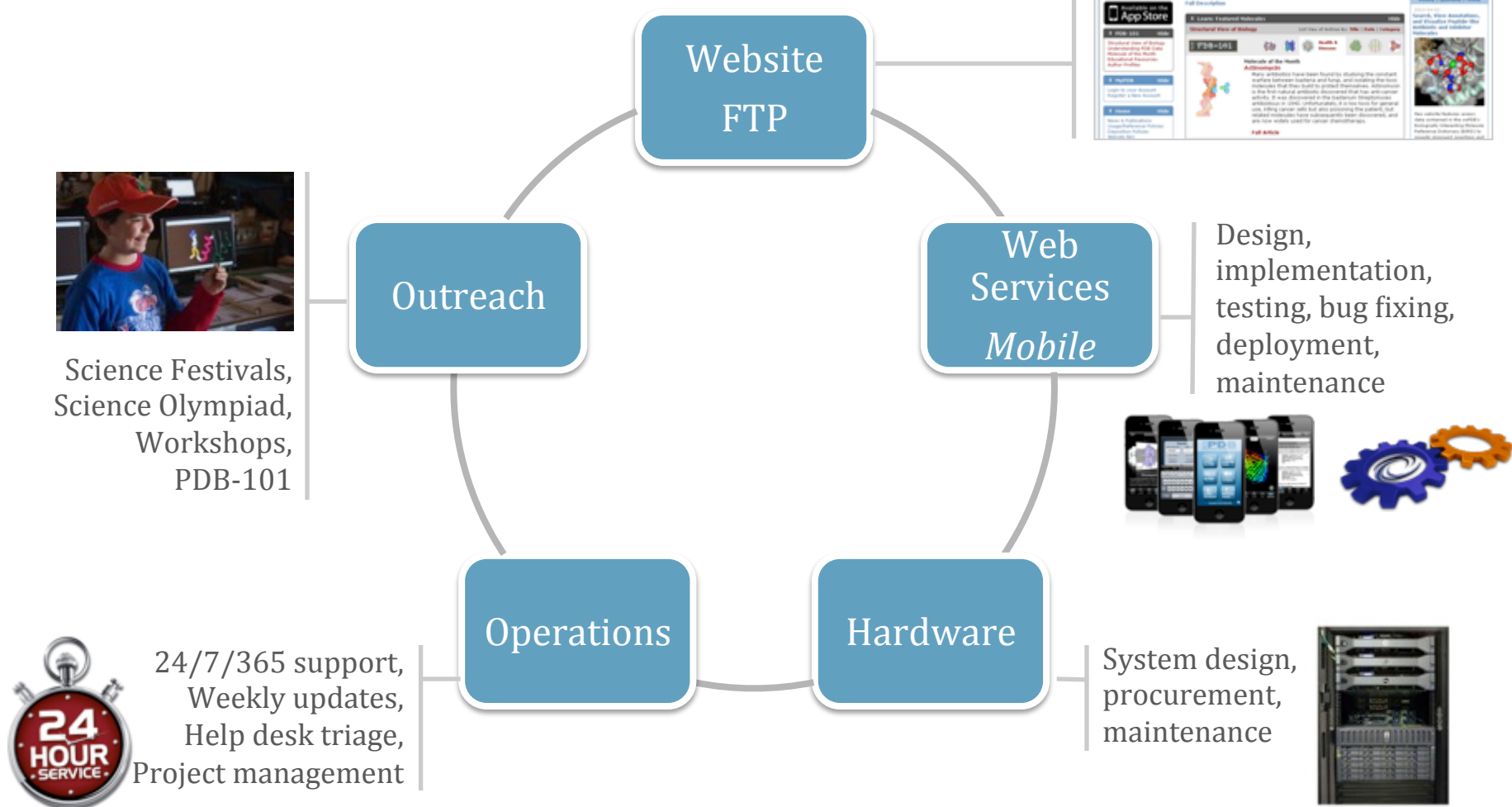
Data Out Goals

Enable research and discovery in science, medicine, and education by

- Presenting an accurate, concise, and meaningful understanding of structure and structure-function relationships
- Addressing broad biological questions where macromolecular structure is central
- Enabling computational analyses involving macromolecular structure



Data Out Responsibilities



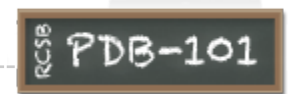
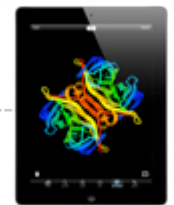
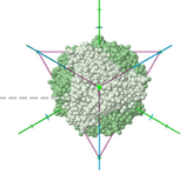
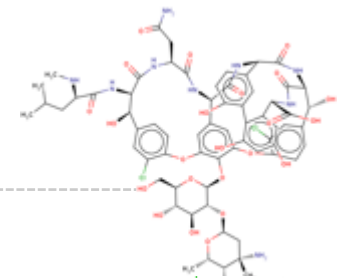
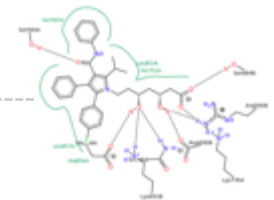
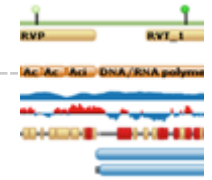
24/7/365 support,
Weekly updates,
Help desk triage,
Project management



Recent Accomplishments: Structural Views of Biology

Viewing structures in a biological context to support different user groups:

- Protein feature view
- Drug view
- Biologically Interesting molecules Reference Dictionary (BIRD) features
- Protein Symmetry view
- RCSB PDB *Mobile* app
- PDB-101 (Outreach presentation)



Protein Feature View

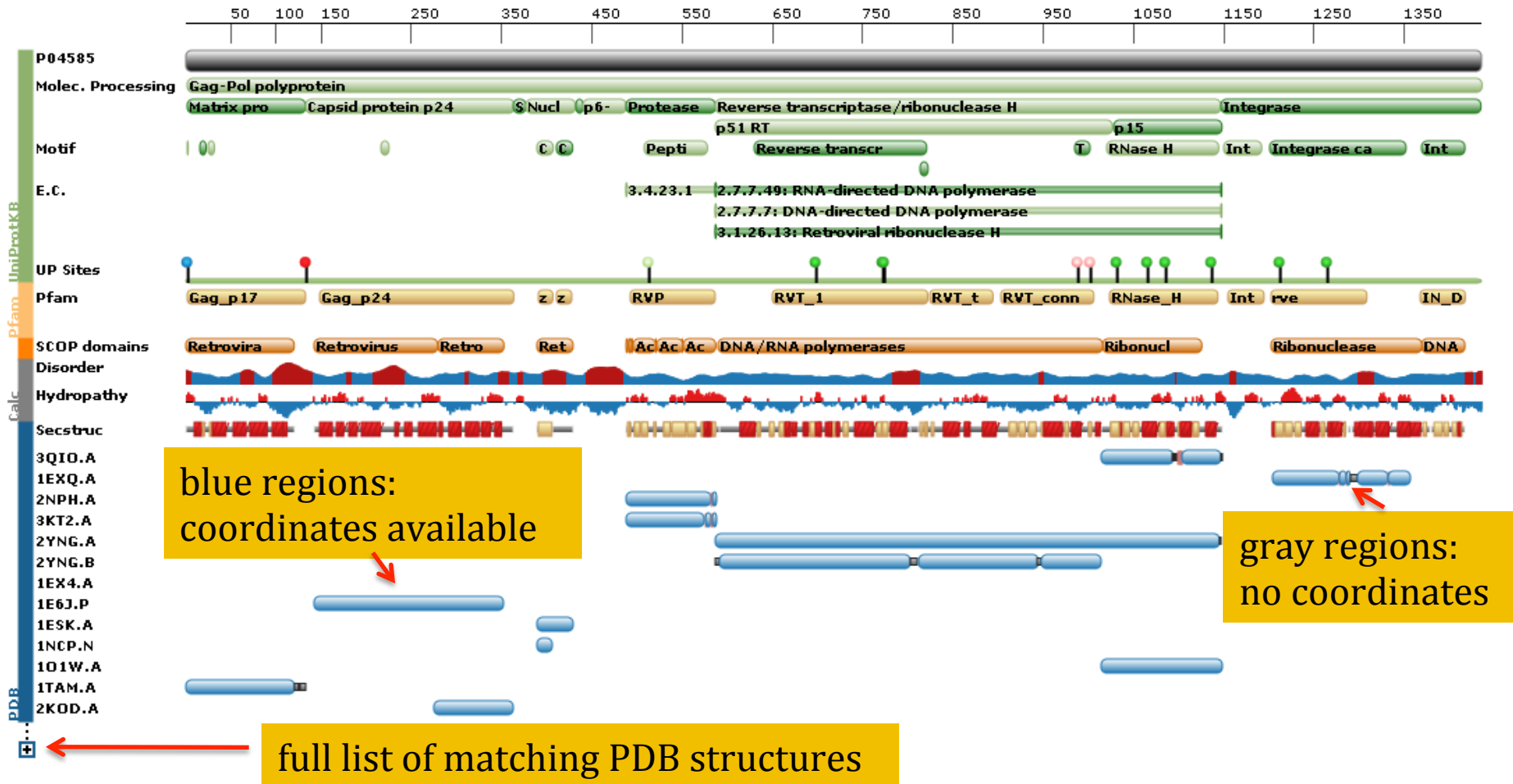
Protein Feature View of PDB entries mapped to a UniProtKB sequence ?

UniProtKB: [Search PDB](#) | [P04585](#)

Species: Human immunodeficiency virus type 1 group M subtype B (isolate HXB2)

Gene name: gag-pol

Length: 1435



Drug View

Search by Generic and Brand Name

Search

Advanced

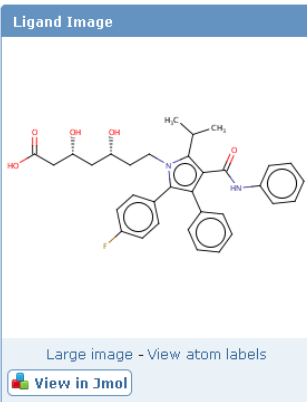
Browse

Chemical Name

- 117: Lipitor

Integration with DrugBank

Chemical Component Summary		Hide
Name	7-[2-(4-FLUORO-PHENYL)-5-ISOPROPYL-3-PHENYL-4-PHENYLCARBAMOYL-PYRROL-1-YL]-3,5-DIHYDROXY-HEPTANOIC ACID	
Identifiers	(3R,5R)-7-[2-(4-fluorophenyl)-5-(1-methylethyl)-3-phenyl-4-(phenylcarbamoyl)-1H-pyrrol-1-yl]-3,5-dihydroxyheptanoic acid (3R,5R)-7-[2-(4-fluorophenyl)-3-phenyl-4-(phenylcarbamoyl)-5-propan-2-yl-pyrrol-1-yl]-3,5-dihydroxy-heptanoic acid	
Synonyms	ATORVASTATIN	
Formula	C ₃₃ H ₃₅ F N ₂ O ₅	
Molecular Weight	558.64 g/mol	
Type	non-polymer	
Isomeric SMILES	CC(C)c1c(C(=O)Nc2ccccc2)c(c(-c2ccc(F)cc2)n1CC[C@@H](O)[C@H](O)CC(O)=O)-c1ccc(O)c1	
InChI	InChI=1S/C33H35FN2O5 /c1-21(2)31-30(33(41)35-25-11-7-4-8-12-25)29(22-9-5-3-6-10-22)32(23-13-15-24(34)16-14-23)36(31)18-17-26(37)19-27(38)20-28(39)40 /h3-16,21,26-27,37-38H,17-20H2,1-2H3,(H,35,41)(H,39,40)/t26-,27-/m1/s1	
InChI key	XUKUURHRXDUEBC-KAYWLYCHSA-N	



Drug Info: DrugBank		Hide
DrugBank ID	DB01076 (Stereoisomeric match)	
Name	Atorvastatin	
Groups	approved	
Description	Atorvastatin (Lipitor) is a member of the drug class known as statins. It is used for lowering cholesterol. Atorvastatin is a competitive inhibitor of hydroxymethylglutaryl-coenzyme A (HMG-CoA) reductase, the rate-determining enzyme in cholesterol biosynthesis via the mevalonate pathway. HMG-CoA reductase catalyzes the conversion of HMG-CoA to mevalonate. Atorvastatin acts primarily in the liver. Decreased hepatic cholesterol levels increases hepatic uptake of cholesterol and reduces plasma cholesterol levels.	
Salts	Atorvastatin calcium	
Brand names	<ul style="list-style-type: none"> Atogal Cardyl Faboxim Hipolixan Lipitor [more]	

Instances in PDB Entries [Hide](#)

As free ligands: **1 entries**

Ex: 1HWK

Related Ligands in the PDB [Hide](#)

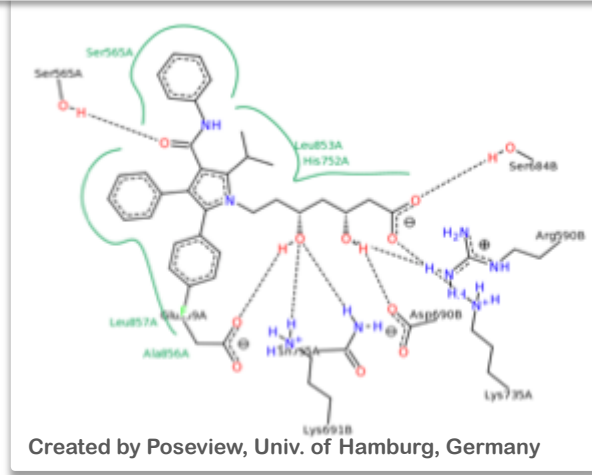
Find stereoisomers

Find similar ligands

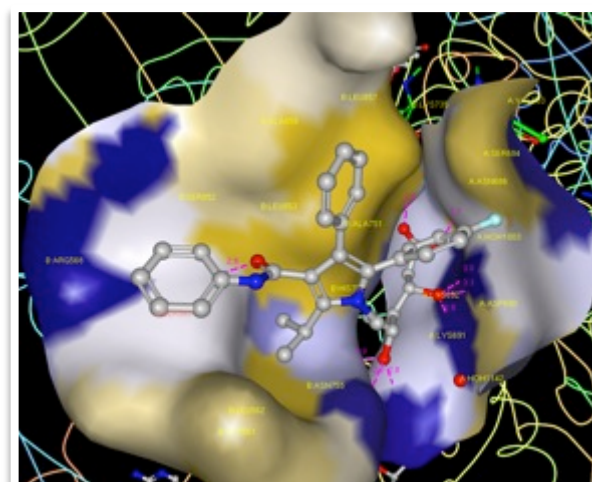
Use 117 for a chemical structure search

Chemical Details		Hide
Formal Charge	0	
Atom Count	76	
Chiral Atom Count	2	

Visualize Binding Site in 2D & 3D



~160K views over 12 months



Drug and Drug Target Mapping

Drug & Drug Target Mapping

Generic Name	Brand Name	DrugBank ID	ATC Codes	Ligand ID	Target Name	UniProt ID	PDB ID 1	Seq. Identity
					hiv			
Amprenavir	• Agenerase • Prozei • Vertex	DB00701	• J05AE05 • J05AE07	478	HIV-1 protease	O90777	1HPV	93%
Atazanavir	• Latazanavir • Reyataz • Zrivada	DB01072	• J05AE08	DR7	HIV-1 protease	O90777	2AQU	93%
Darunavir	• Prezista	DB01264	• J05AE10	017	HIV-1 protease	O90777	4HLA	93%
Indinavir	• Crivivan	DB00224	• J05AE02	MK1	HIV-1 protease	O90777	1HSG	93%
Lopinavir	• Aluviran • Kaletra	DB01601	• J05AR10	AB1	HIV-1 protease	O90777	1MUI	93%
Nelfinavir	• Viracept	DB00220	• J05AE04	1UN	HIV-1 protease	O90777	1OHR	93%
Ritonavir	• Norvir • Norvir Sec	DB00503	• J05AE03	RIT	HIV-1 protease	O90777	1HXW	92%
Saquinavir	• Fortovase • Invirase • ROC	DB01232	• J05AE01	ROC	HIV-1 protease	O90777	1HXB	94%

Browsing Drugs by ATC Tree

The **Anatomical Therapeutic Chemical (ATC)** Classification System is used for the classification of drugs. It is a hierarchical system used for drug statistics methodology. ATC names are only listed when there are corresponding PDB entries.

Here you can browse an ATC name, view the number of associated PDB structures, and search for structures.

Search in Tree

- A: ALIMENTARY TRACT AND METABOLISM - [15456 Structures]
- B: BLOOD AND BLOOD FORMING ORGANS - [7878 Structures]
- C: CARDIOVASCULAR SYSTEM - [7761 Structures]
 - C.01: CARDIAC THERAPY - [139 Structures]
 - C.02: ANTIHYPERTENSIVES - [4 Structures]
 - C.03: DIURETICS - [26 Structures]
 - C.04: PERIPHERAL VASODILATORS - [29 Structures]
 - C.05: VASOPROTECTIVES - [69 Structures]
 - C.07: BETA BLOCKING AGENTS - [6 Structures]
 - C.08: CALCIUM CHANNEL BLOCKERS - [5 Structures]
 - C.09: AGENTS ACTING ON THE RENIN-ANGIOTENSIN SYSTEM - [9 Structures]
 - C.10: LIPID MODIFYING AGENTS - [7519 Structures]
 - C.10.A: LIPID MODIFYING AGENTS, PLAIN - [7519 Structures]
 - C.10.A.A: HMG CoA reductase inhibitors - [2 Structures]
 - C.10.A.A.02: Lovastatin - [1 Structure]
 - C.10.A.A.05: Atorvastatin - [1 Structure]**
 - C.10.A.B: Fibrates - [1 Structure]
 - C.10.A.X: Other lipid modifying agents - [7493 Structures]

Click to search for structures that associated with ATC: C.10.A.A.05: Atorvastatin

Integration with Binding Affinity Databases

External Ligand Annotations

Identifier	Binding Affinity (Sequence Identity %)
117 Search Download	EC50: 2.5 nM (91) - data from BindingDB
	IC50: 3.8 - 6.2 nM (91 - 99) - data from BindingDB
	IC50: 8 nM - data from BindingMOAD

Biologically Interesting Molecules

Search, display, and visualize peptide-like molecules annotated in the Biologically Interesting molecule Reference Dictionary (BIRD)

The image displays the BIRD web interface. At the top, a table lists search results for Actinomycin D. Below the table, the chemical structure of Actinomycin D is shown. To the left is the 'Advanced Search Interface' with various filters. To the right is the 'Ligand Explorer' window showing a 3D model of Actinomycin D bound to DNA.

Identifier	Image	Name	Type	Class	Chain ID
PRD_000001 Search		Actinomycin D ?	Polypeptide ?	Antibiotic ?	E,F,G,H

Advanced Search Interface

Biologically Interesting Molecules (from BIRD) [?](#)

Find annotated biologically interesting molecules (BIRD)

Name or ID: actinomycin

Type:

Class:
 Any
 Antagonist
 Antibiotic
 Anticancer
 Anticoagulant
 Antiinflammatory
 Antimicrobial
 Antiretroviral
 Antithrombotic
 Antitumor
 Antiviral
 Caspase Inhibitor
 Enzyme Inhibitor
 Immunosuppressant
 Inhibitor
 Lantibiotic
 Metal Transport
 Thrombin Inhibitor
 Toxin
 Trypsin Inhibitor

Remove Similar Set

Match of the ab

Identity [?](#)

Ligand Explorer

Structure: 1I3W

Chain: [SATGABRC](#)
A: 2

Choose a ligand to analyze...

Choose interactions & thresholds...

- Hydrogen Bond 3.3
- Hydrophobic 3.9
- Bridged H-Bond 3.3
- Metal Interaction 3.5
- Neighbor Residues 4.0

Label interactions by distance

Surfaces: Off Transparent Opaque

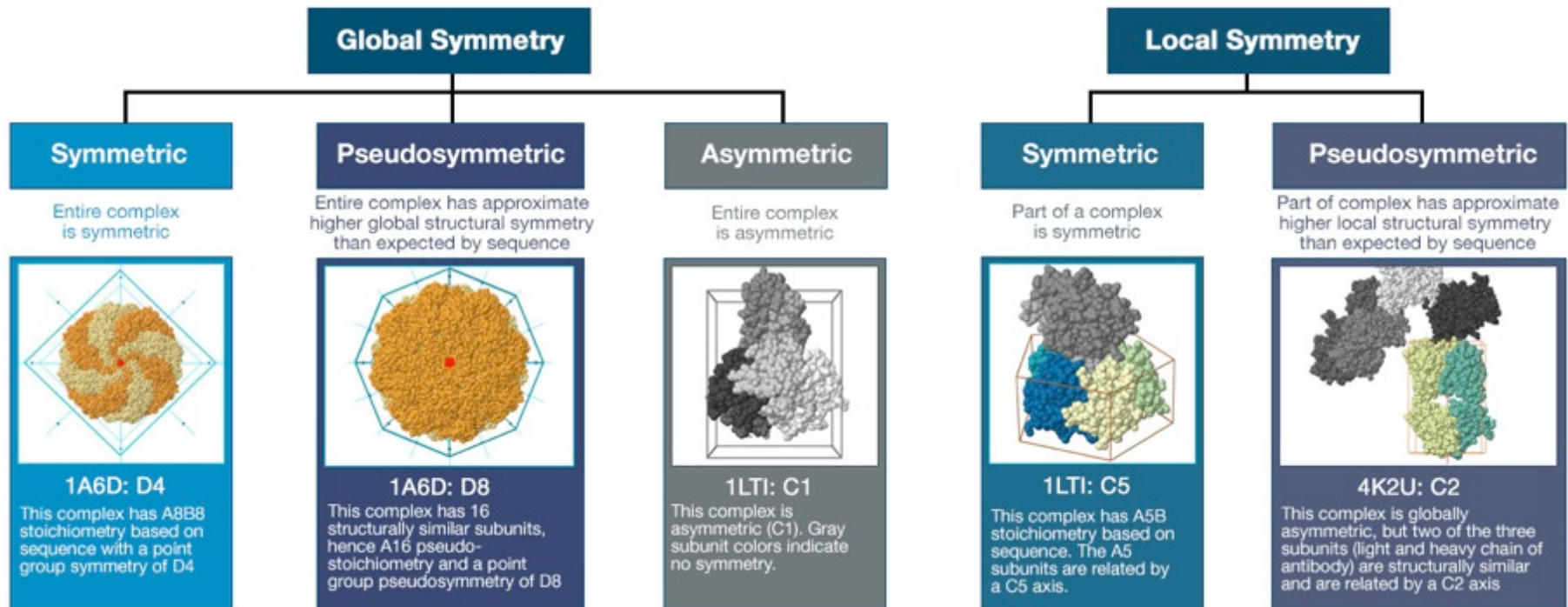
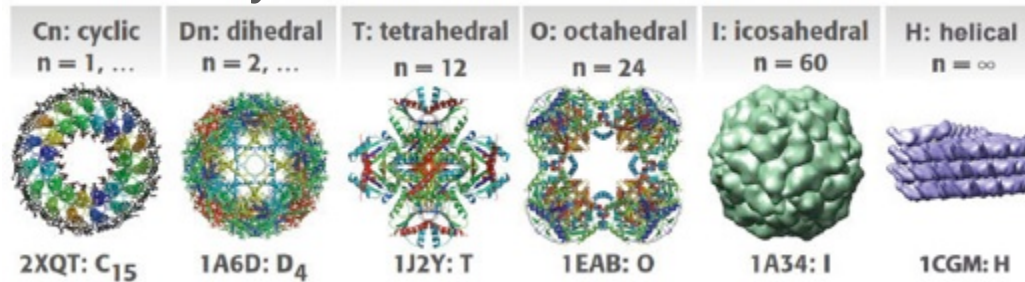
Color by: Chain

Distance: 20 Type: Solid

Actinomycin D bound to DNA,
PDB ID: 1i3w, Robinson *et al.*
Biochemistry 40, 5587-5592 (2001)

Protein Symmetry View

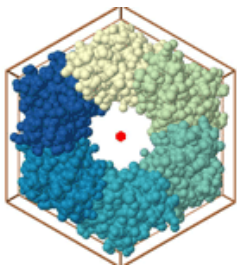
Protein symmetries observed in the PDB



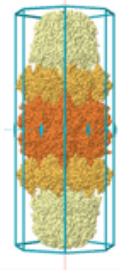
Protein Symmetry and Stoichiometry

Visualizing Symmetry in Jmol

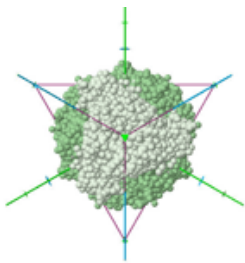
Symmetry axes, polyhedra, coloring



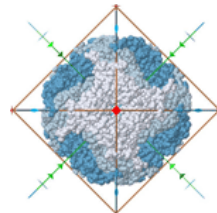
Cyclic: C6



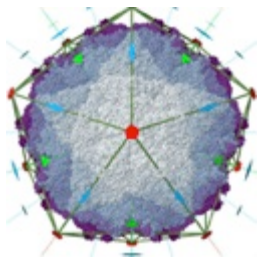
Dihedral: D7



Tetrahedral: T



Octahedral: O

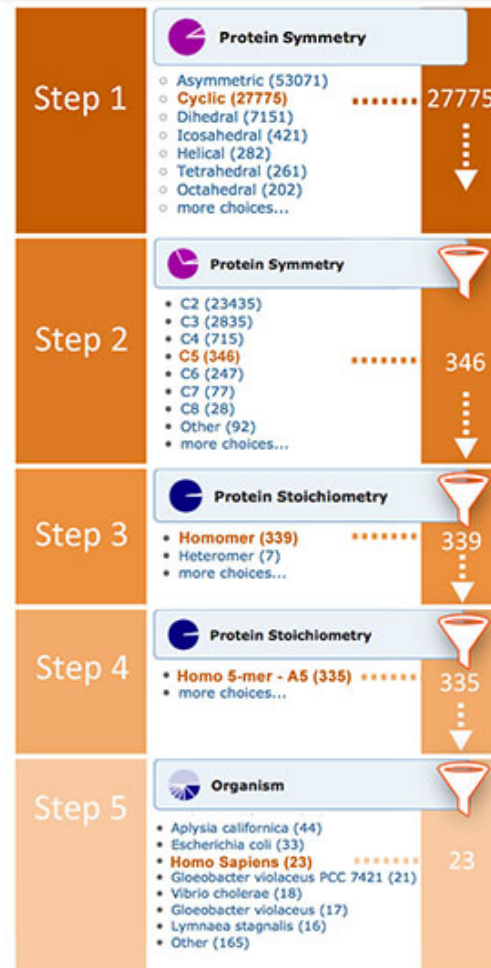


Icosahedral: I

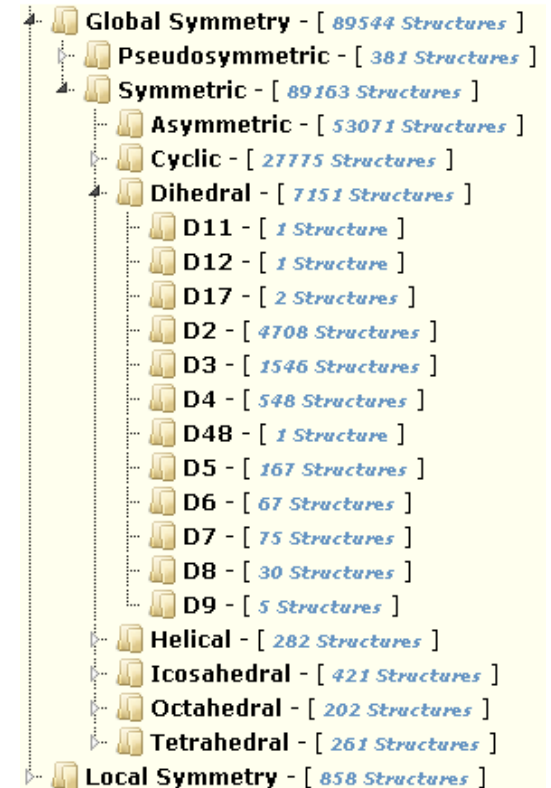
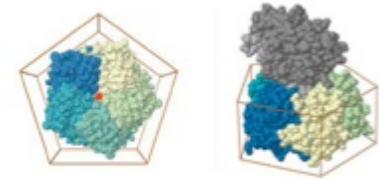


Helical: H

Drilldown by Global Protein Symmetry and Stoichiometry



Browsing by Different Symmetry Types

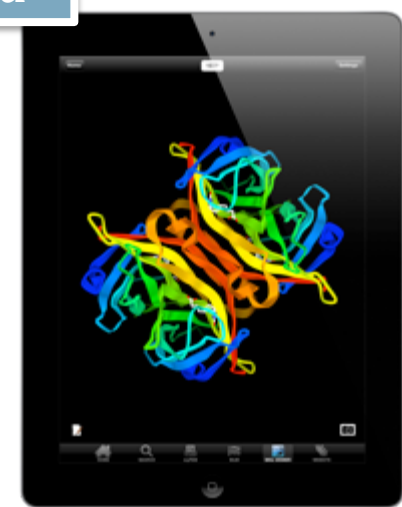


RCSB PDB *Mobile App*

Available on the App Store



iPad



Android in beta testing



iPhone
iPod

- Browse *Molecule of the Month* articles
- Simple search
- Interactive 3D viewer*
- Browse search results

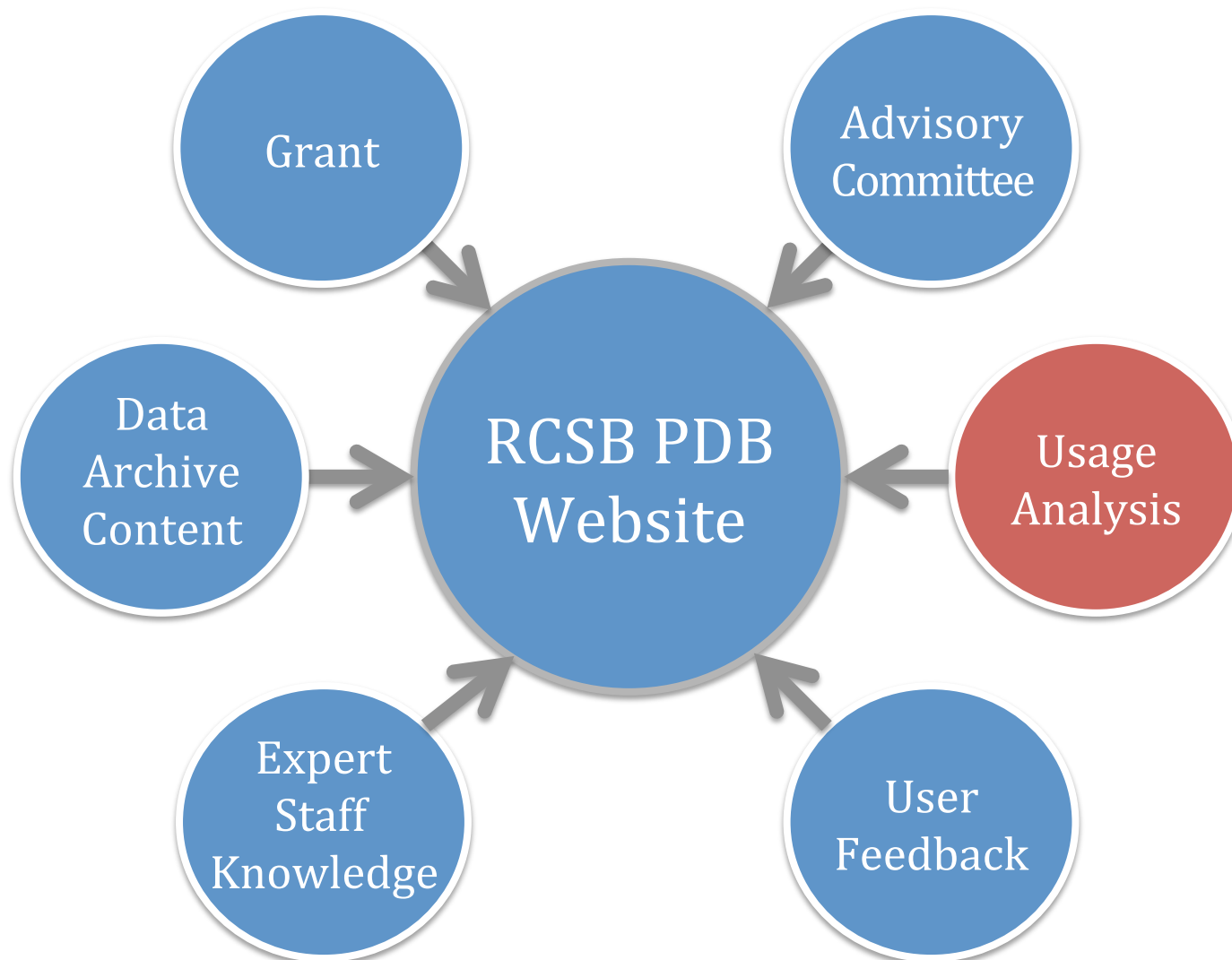
*NDKMol, Takanori Nakane, Kyoto University



How Do We Decide Which New Features to Add?



Factors Influencing Our Decisions

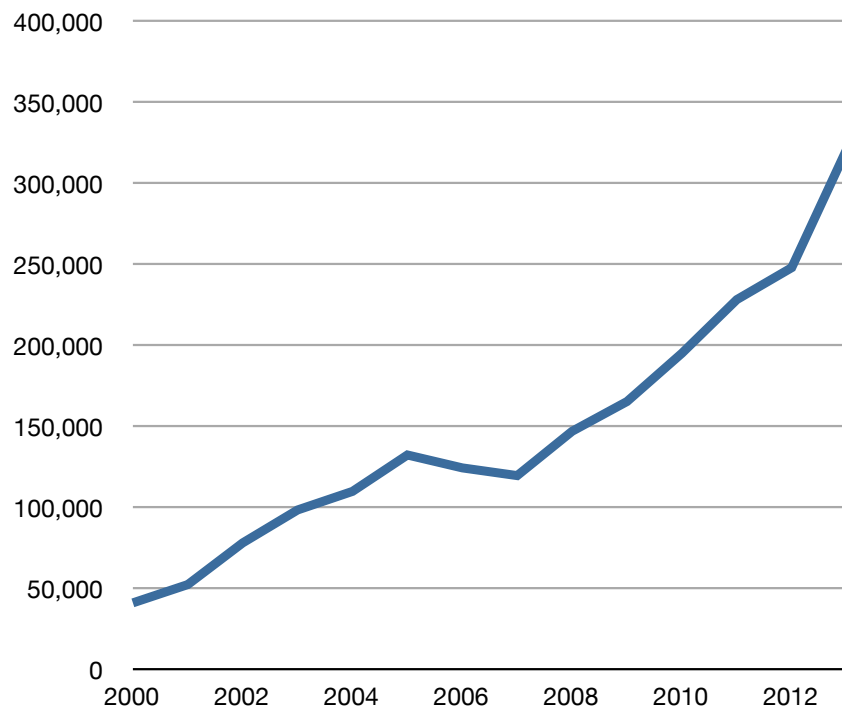


How Do We Measure the Impact of RCSB PDB Website Improvements?



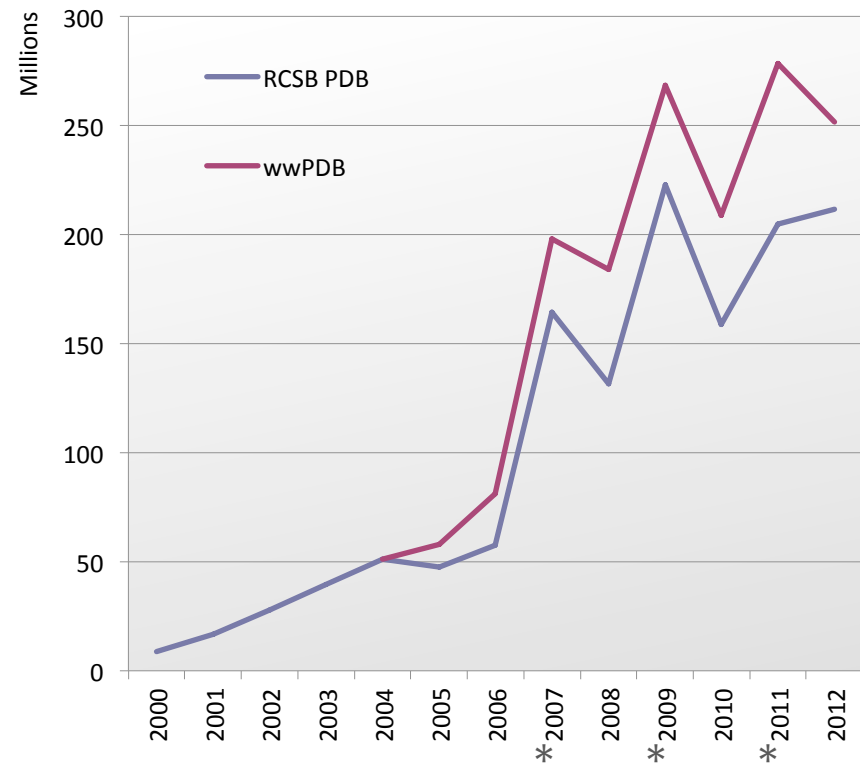
Primary Access Statistics

RCSB PDB Website Unique Visitors Per Month



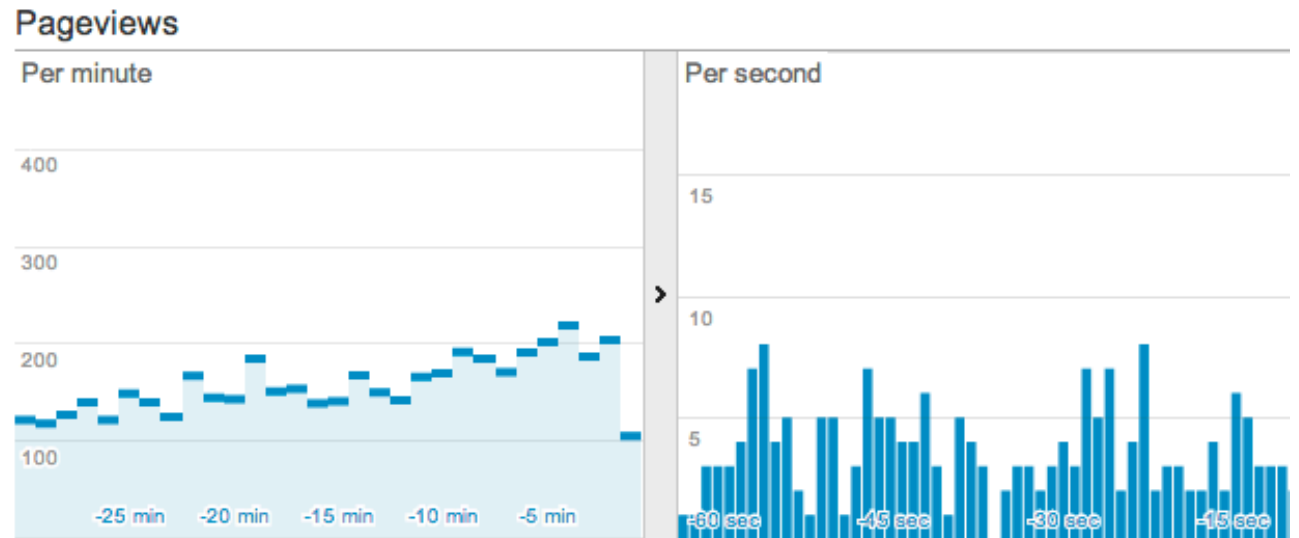
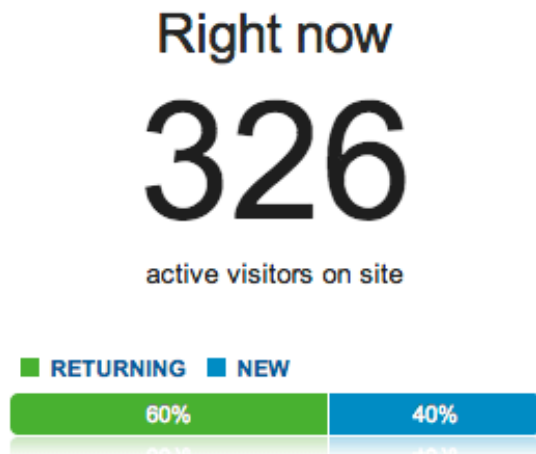
This grant period:
traffic almost doubled and growing

File Downloads (FTP)



*Release of remediated data

Website Monitoring



Google Analytics used to

- Analyze growth rates
- Monitor feature usage
- Understand user workflows
- Identify appropriate website improvements



How Users Come to Our Site and Where They Land

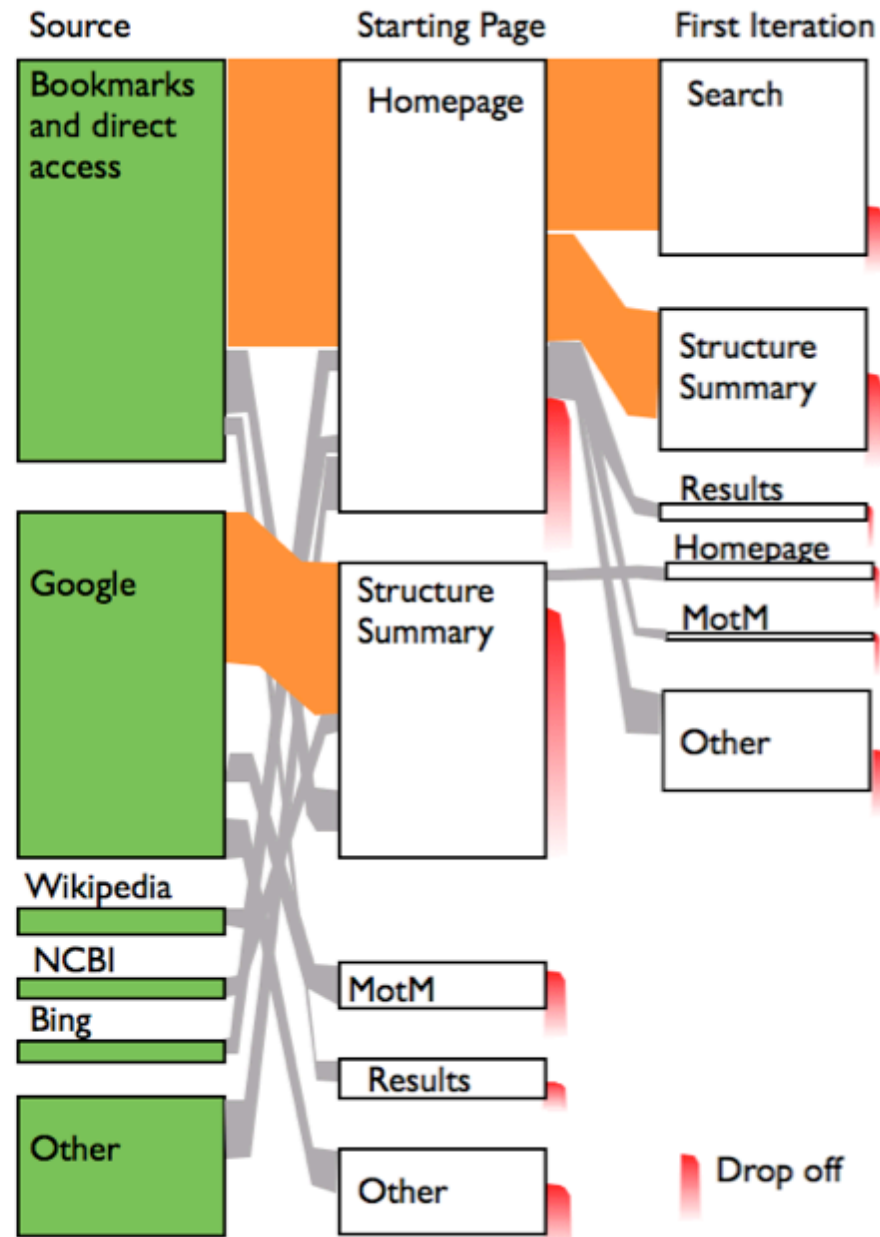
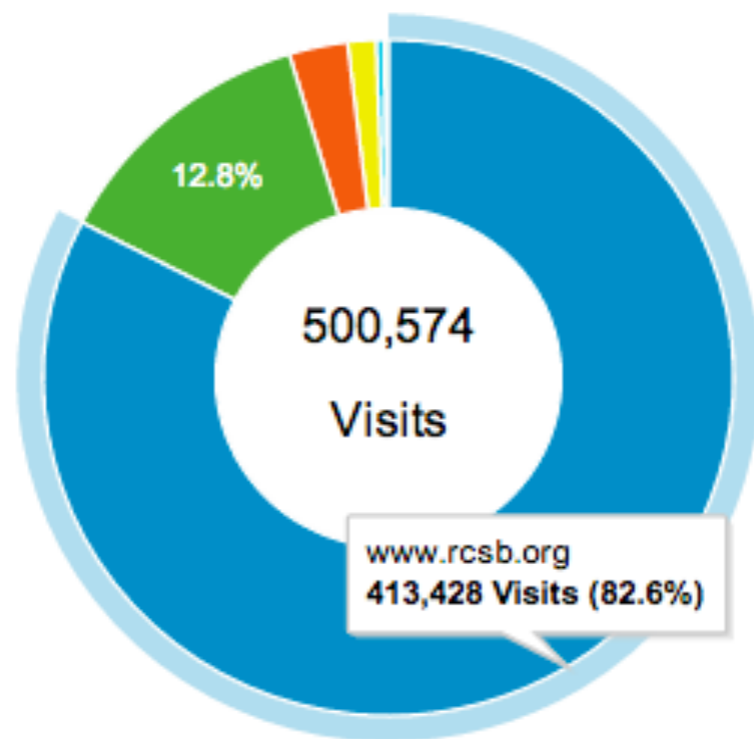


Diagram size
proportional to traffic

Visits by Hostname

■ www.rcsb.org ■ www.pdb.org ■ pdb.org
■ rcsb.org ■ pdb.rcsb.org ■ Other

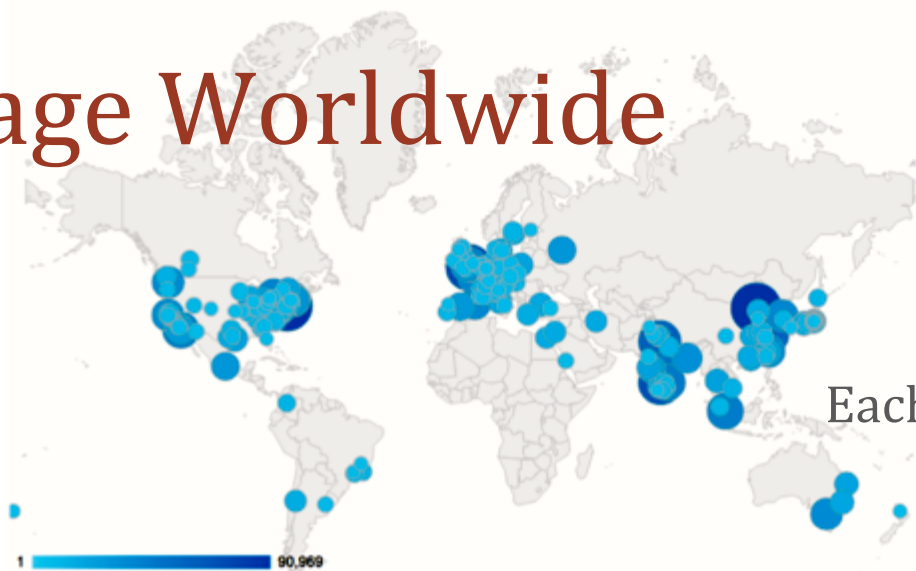


Where Is Our Growth Coming From?

Feature	Page Views May-Aug 2013	Annual Growth Rate
Drilldown and Search	2.6M	10%
Homepage	1.5M	5%
PDB-101	445k	9%
Ligand Summary	95k	5%
Protein Feature View	76k	New
Search Unreleased	61k	14%
Help	21k	37%
Chemical Search	8k	46%
News	7k	31%
MyPDB Query	5k	30%
Web Services Documents	5k	11%



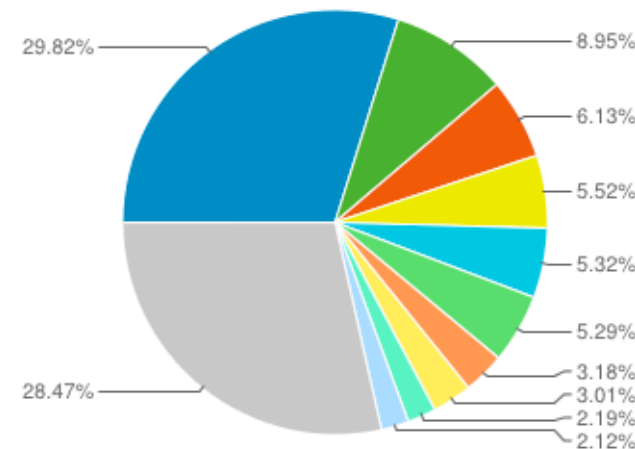
2012 Usage Worldwide



Each circle > 7k visits

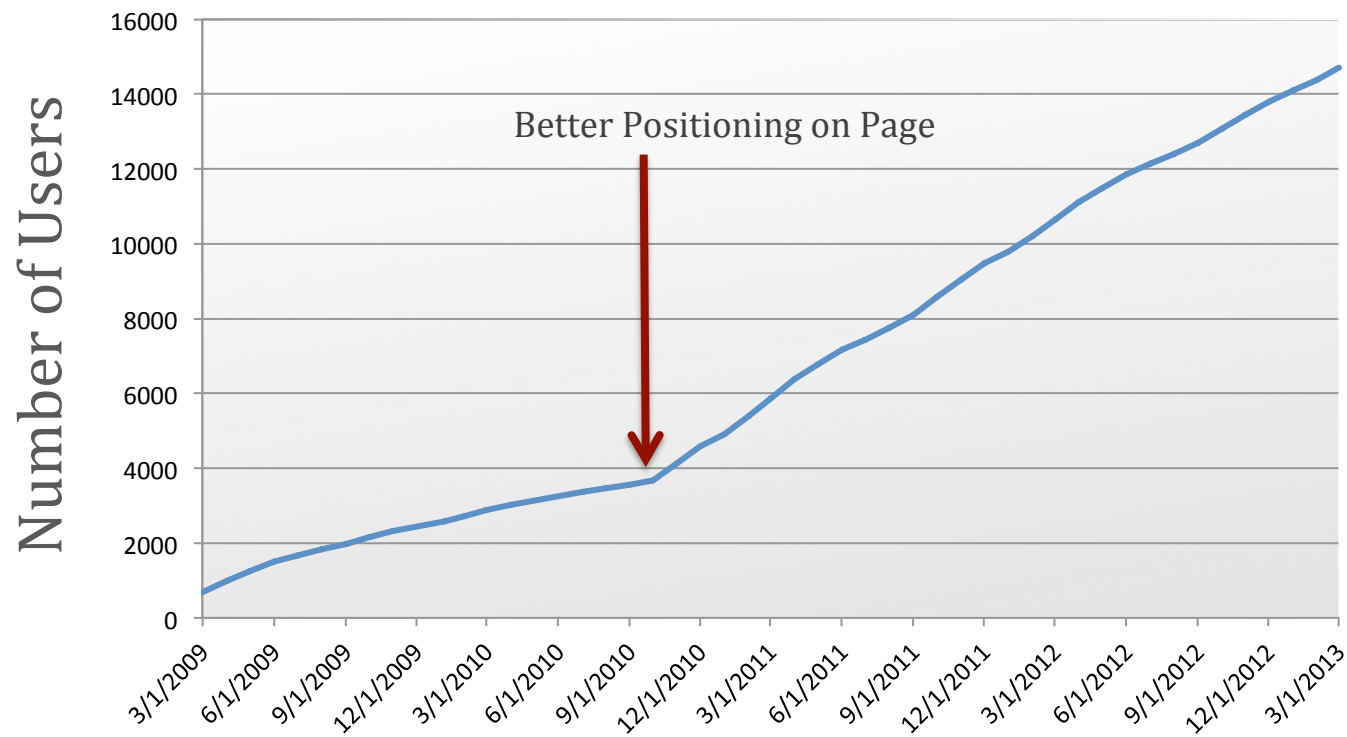
Top Countries in 2012

Country / Territory	Visits	Visits	Contribution to total:
1. United States	1,909,608	29.82%	29.82%
2. India	573,024	8.95%	8.95%
3. United Kingdom	392,663	6.13%	6.13%
4. China	353,662	5.52%	5.52%
5. Germany	340,741	5.32%	5.32%
6. Japan	338,453	5.29%	5.29%
7. Canada	203,561	3.18%	3.18%
8. France	192,786	3.01%	3.01%
9. Italy	140,080	2.19%	2.19%
10. Spain	135,536	2.12%	2.12%



MyPDB

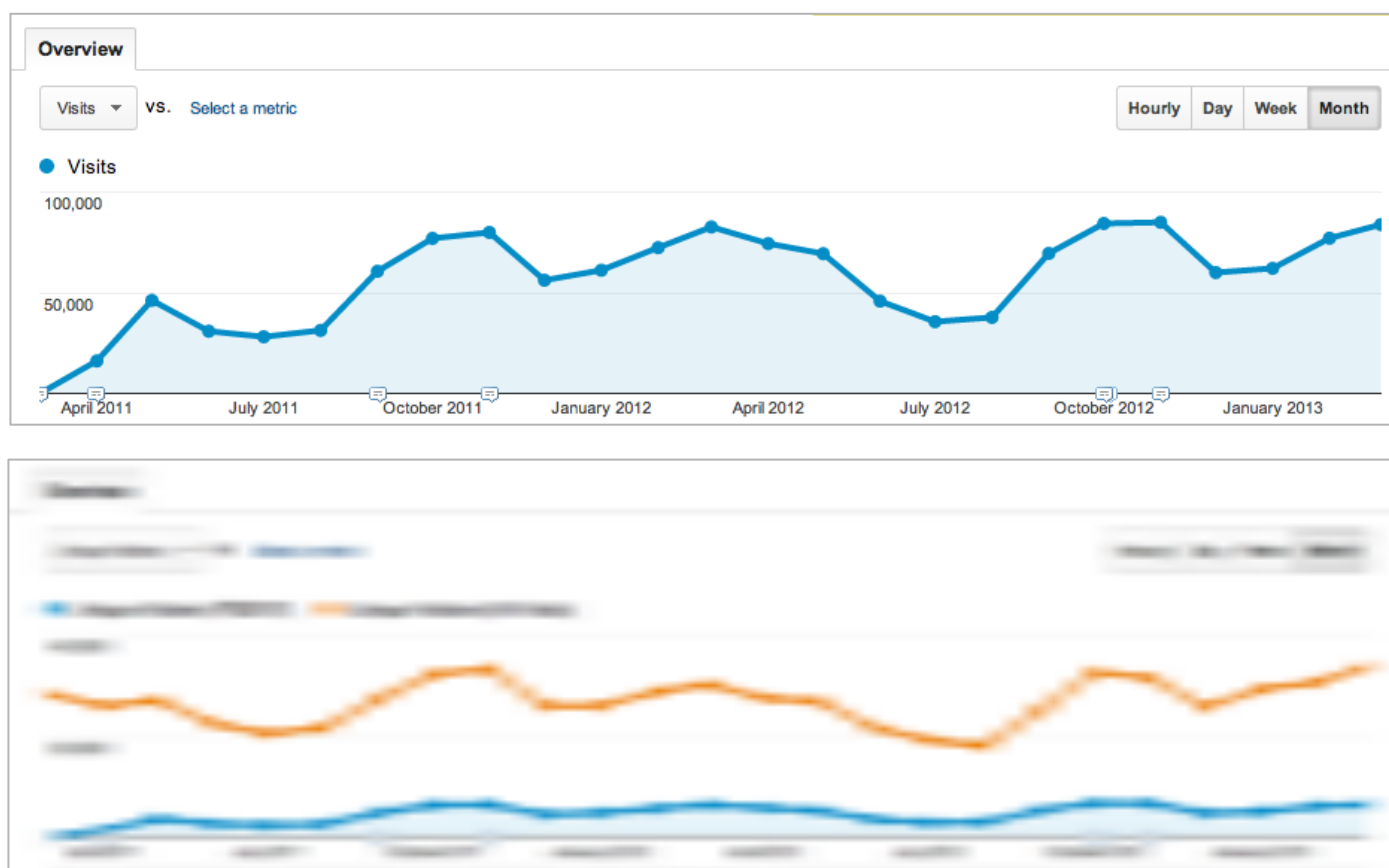
- Stores user queries and annotations
- Weekly or monthly query notifications



Growth in PDB-101 (Outreach & Education)

Usage
statistics:

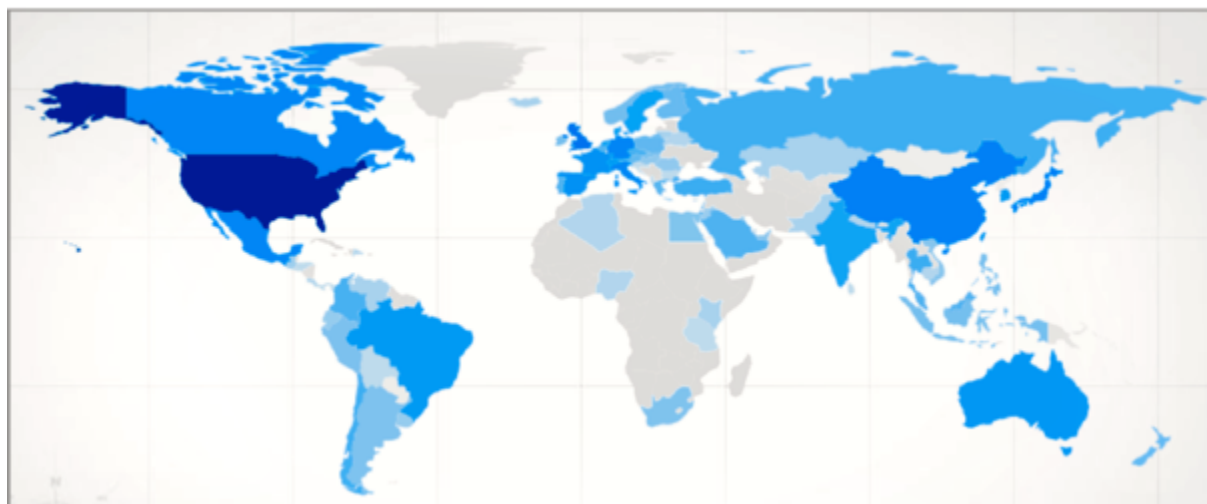
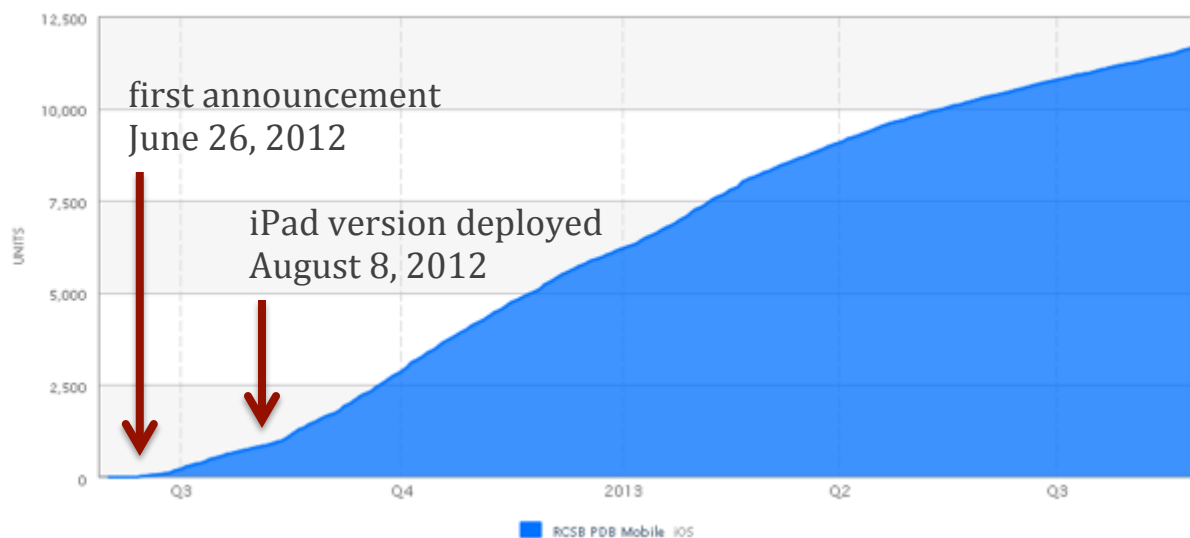
- ~75,000 visits per month
- ~12% of all site visitors
- ~17% of all page views



RCSB PDB *Mobile* App Downloads

Available in iTunes store

- iPhone/iPod: released May 29, 2012; announced June 26, 2012
- iPad: released August 8, 2012
- Android version in beta testing



Feature Usage (July 2013)



Website Usage by Device

Device Category	Month	Visits
Desktop	June 2013	525,337
	June 2012	444,596
	% Change	18.16%
Mobile	June 2013	14,801
	June 2012	7,256
	% Change	103.98%
Tablet	June 2013	6,084
	June 2012	3,498
	% Change	73.93%

← Desktop usage up 18%

← Mobile and tablet usage almost doubled

Mobile Strategy

Considerations:

- Mobile usage of website exceeds RCSB PDB *Mobile* app usage

Strategy for new grant period:

- *Responsive web design* of website to support all types of devices
- Maintain RCSB PDB *Mobile* app “as is”



Old vs. New Hardware



old



UCSD Rutgers

new

Average page load time improved ~10% since last year



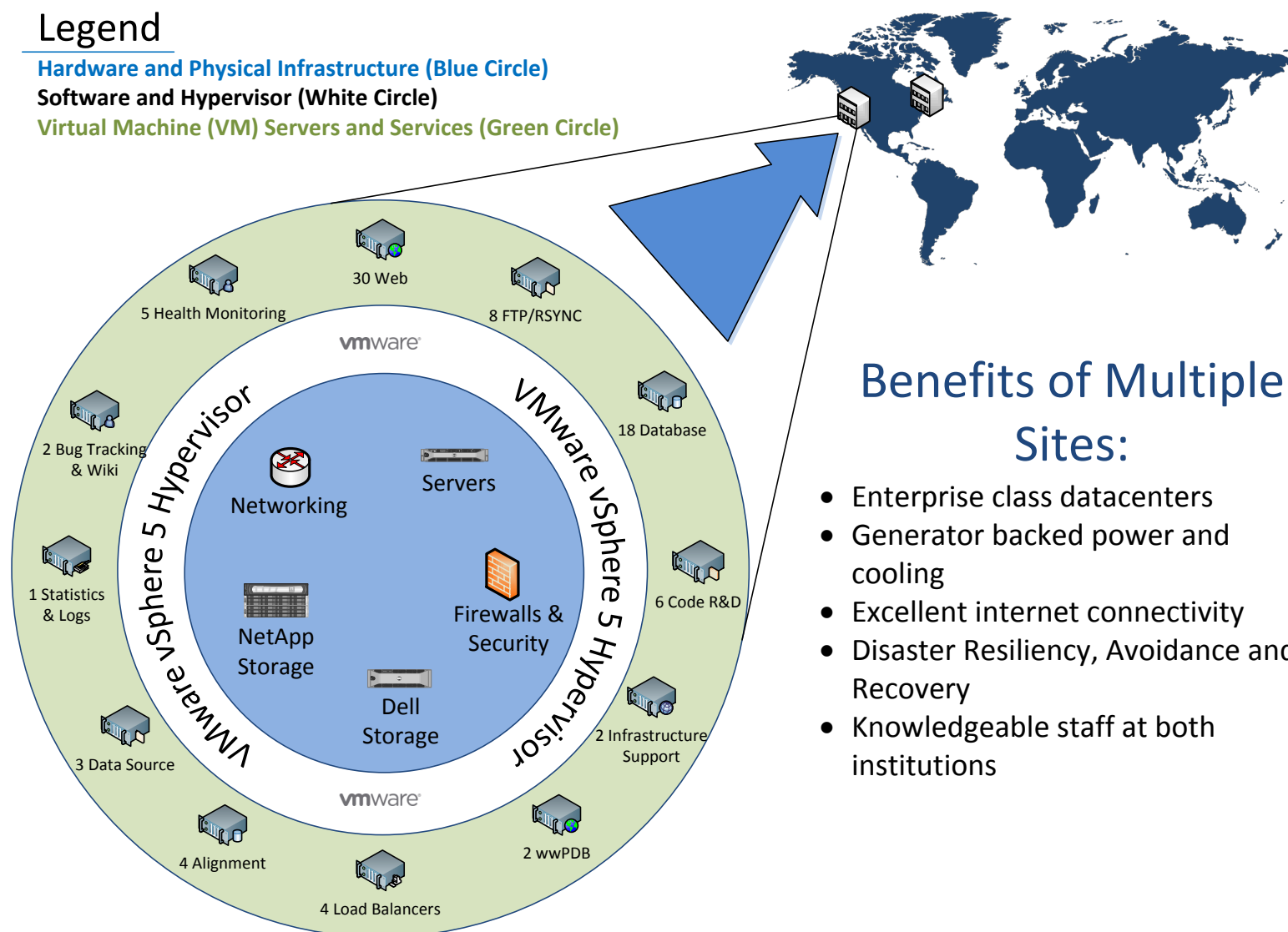
Data Out IT Infrastructure

Legend

Hardware and Physical Infrastructure (Blue Circle)

Software and Hypervisor (White Circle)

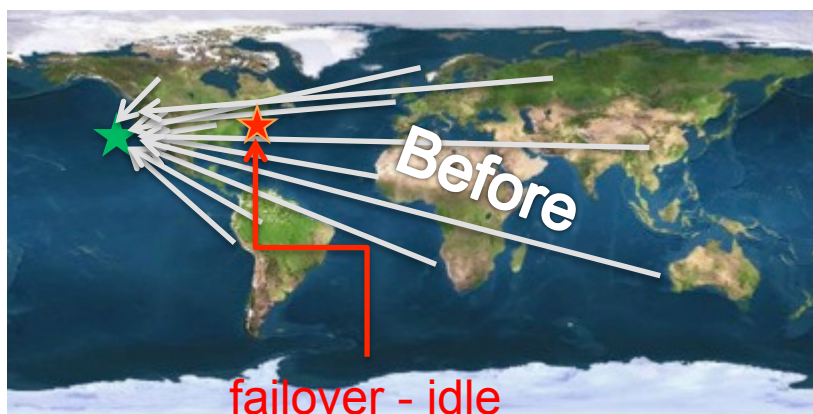
Virtual Machine (VM) Servers and Services (Green Circle)



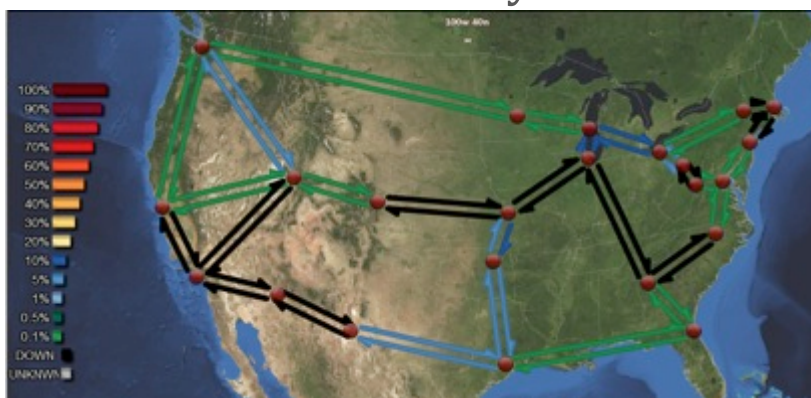
Benefits of Multiple Sites:

- Enterprise class datacenters
- Generator backed power and cooling
- Excellent internet connectivity
- Disaster Resiliency, Avoidance and Recovery
- Knowledgeable staff at both institutions

Global Load Balancing of Web Traffic



Example: During Internet 2 outage Sep 3, 2013, mitigated access to website successfully



- Increased reliability and faster access from East coast and Europe
- Balanced load on servers (no idle failover servers)

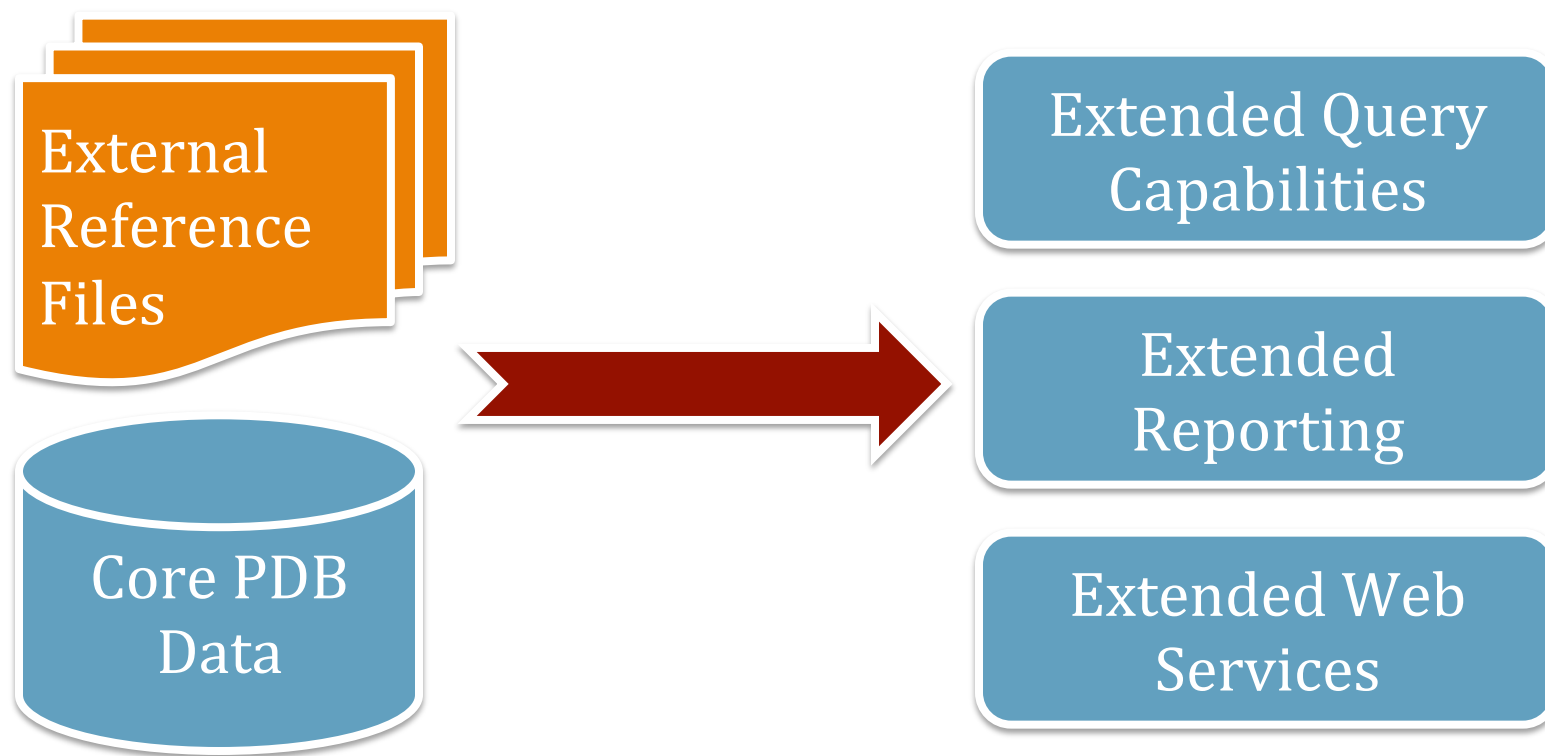
Plans for Next Grant Period

- Enhance searching based on external reference files
- Improve searching across diverse data types
- Improve reporting
- Provide pathway mapping
- Extend Web Services
- Improve structural analysis and visualization



External Reference Files (ERFs)

- A data file containing compilations of additional annotations on structure entries
- Can be extended as necessary without impacting archival data files



New ERF: Membrane Protein Annotation

- Annotation and classification of membrane proteins
- Query by membrane classification
- Visualization of membrane region
- Collaboration with SBKB and Stephen White (UC Irvine)

Prototype drilldown using *mpstruc* classification



Transmembrane Proteins

- ALPHA-HELICAL (1588)
- BETA-BARREL (308)
- MONOTOPIC MEMBRANE PROTEINS (247)



Transmembrane Proteins

- Channels: Potassium and Sodium ... (154)
- Bacterial and Algal Rhodopsins (121)
- Photosynthetic Reaction Centers (110)
- G Protein-Coupled Receptors (GPCRs) (93)
- P-type ATPase (87)
- Major Facilitator Superfamily (... (82)
- Multi-Drug Efflux Transporters (71)
- Other (877)



Transmembrane Proteins

- Rhodopsin (25)
- β_1 adrenergic r ... (15)
- A_{2A} adenosine receptor (12)
- β_2 adrenergic r ... (11)
- CXCR4 chemokine receptor comple ... (5)
- Rhodopsin, Squid (4)
- Apelin receptor (AR) TM helix 1 ... (4)
- Other (17)

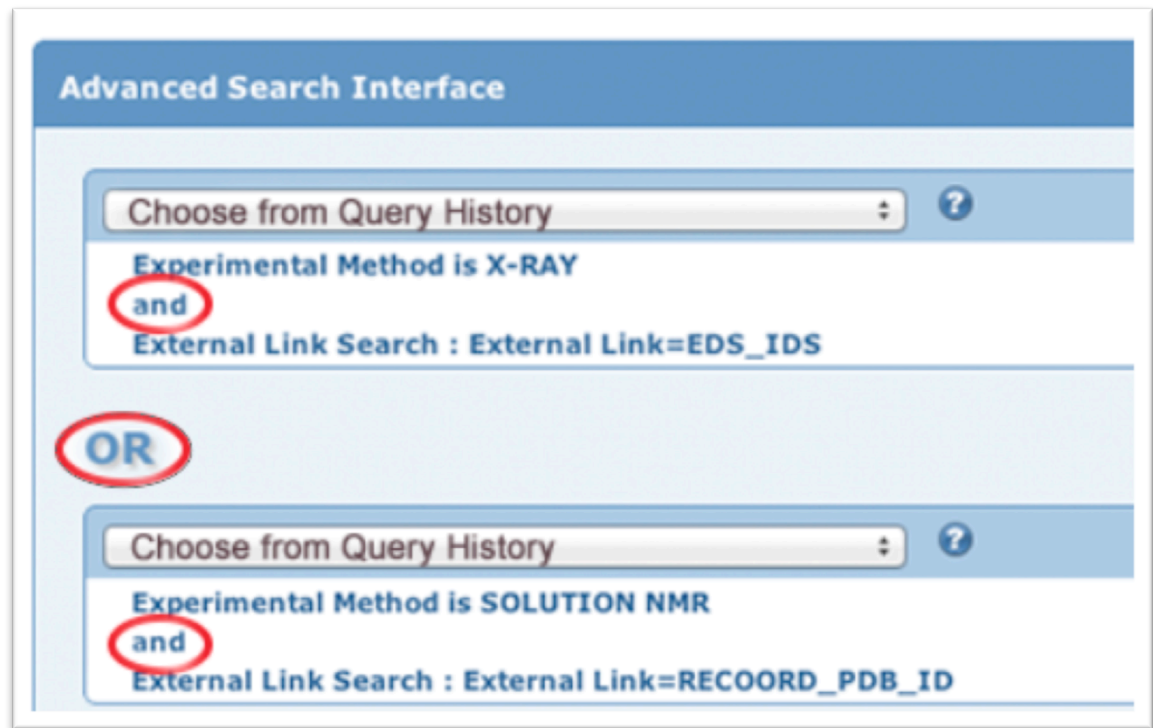
New ERF: Ligand Classification

- Find and export ligands relevant to drug design
- Classify ligands by
 - Modified residues vs. non-polymer ligand
 - Organic vs. inorganic
 - Buffer components, cryo-protectant, etc.
 - Approved drug or experimental compound
 - Action(s): inhibitor, agonist, antagonist, substrate, etc.
 - Ontologies (ChEBI, MeSH)



Improve Searching

- Search across diverse data types such as sequence, structure, ligands, function(s)
- Complex Boolean expressions combining sub-queries



The screenshot displays the 'Advanced Search Interface' with two search criteria blocks. The first block contains the text 'Experimental Method is X-RAY' and 'External Link Search : External Link=EDS_IDS', with the word 'and' circled in red. The second block contains the text 'Experimental Method is SOLUTION NMR' and 'External Link Search : External Link=RECOORD_PDB_ID', also with the word 'and' circled in red. A large 'OR' is circled in red between the two blocks, indicating a disjunctive search.



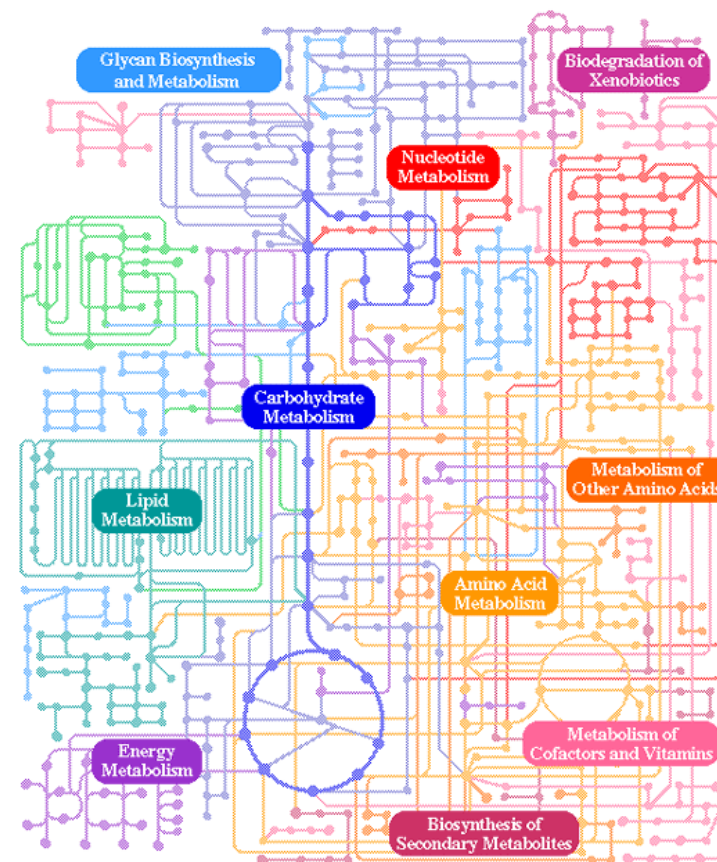
Improve Reporting

- Develop flexible framework to plot PDB data as timelines, networks, scatter plots, heat maps, histograms, and other display formats
- Reporting functionality will use annotations and summarized data from ERFs



Map Structures to Pathways: Why?

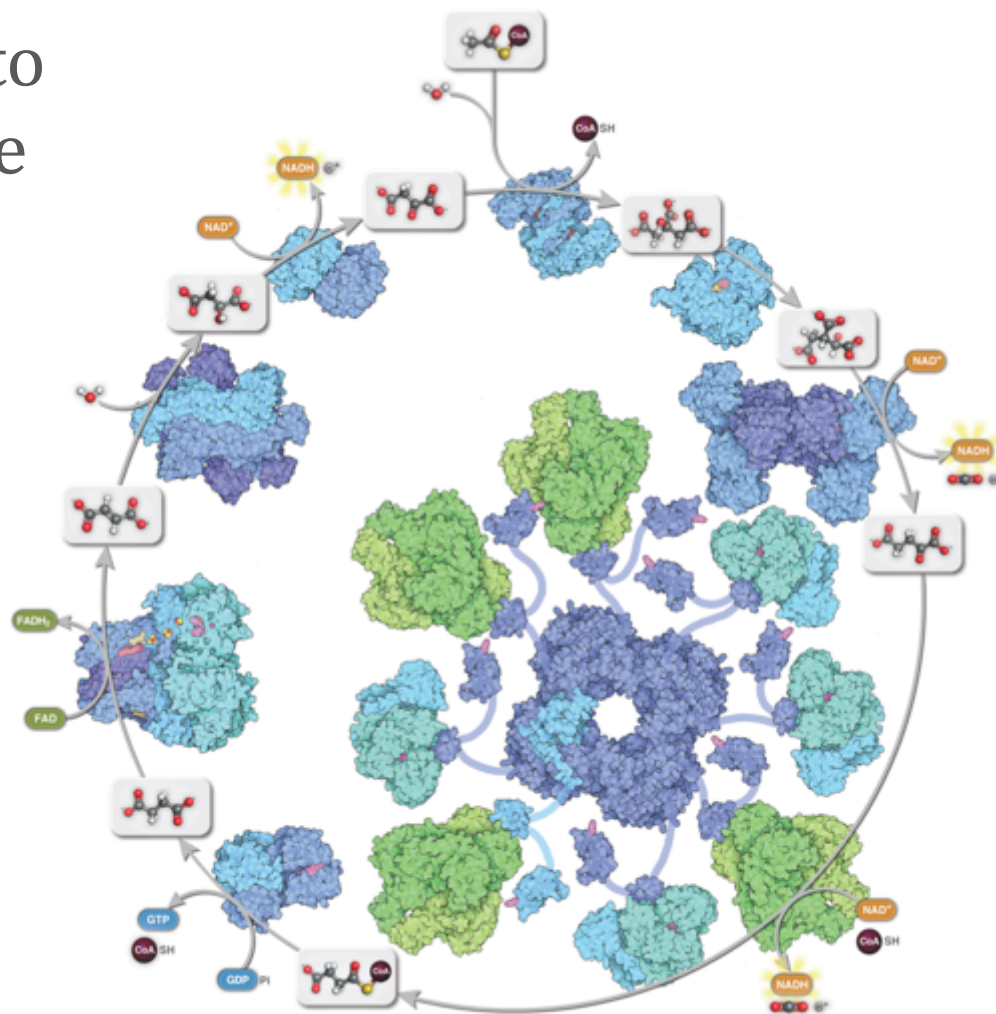
- Proteins do not act in isolation
- To provide a Structural View of Biology
- Repeated user requests for searching of metabolic pathways, enzymes, and metabolites



01100 5/31/04 Image source from KEGG

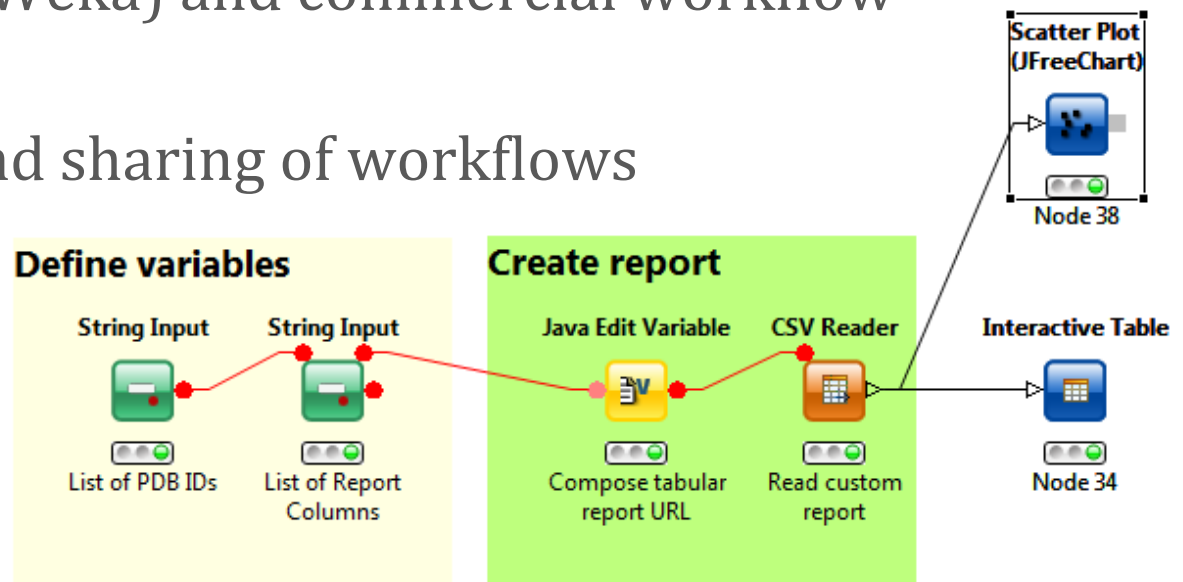
Map Structures to Pathways: Example

- Automated mapping to pathways in Reactome and other model organism databases stored in ERFs
- Semi-automatic in-depth mapping of selected pathways
- Queries and visualization of structural coverage



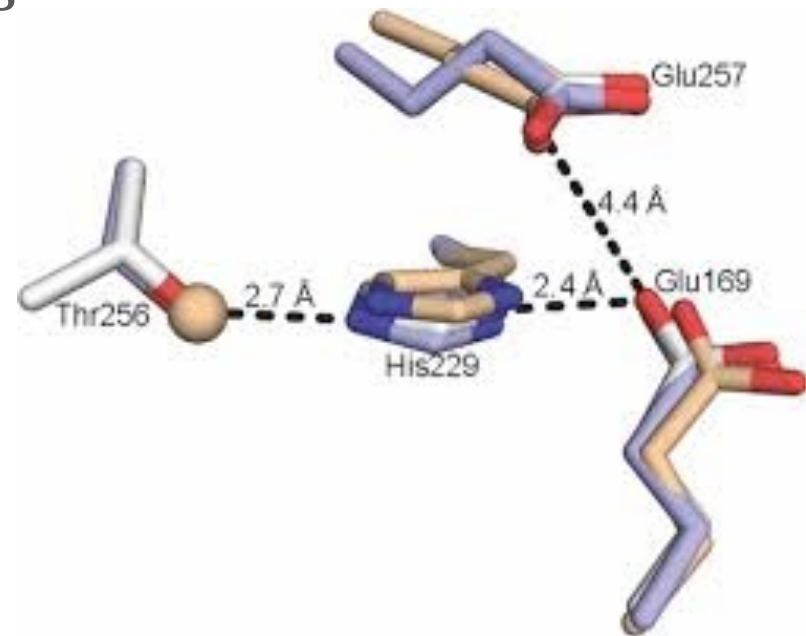
Extend Web Services

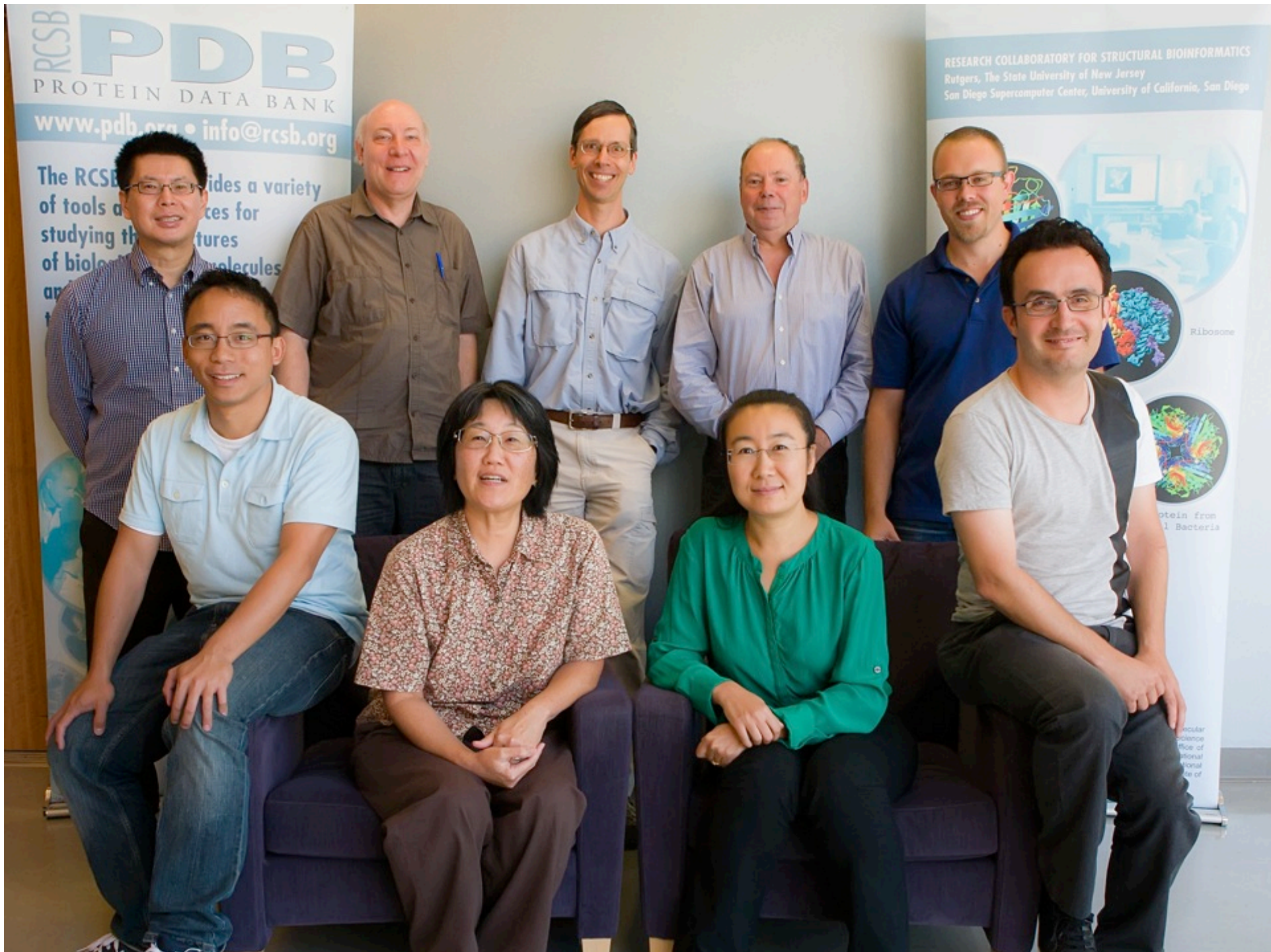
- Provide programmatic access to archival PDB data plus extra annotations from ERFs
- Provide workflow modules that can be used by common workflow tools such as KNIME or Galaxy
- Enable use of PDB data in conjunction with other open source (R, Weka) and commercial workflow modules
- Enable saving and sharing of workflows



Structural Analysis/Visualization

- Predefined queries for common structural motifs (metal coordination, cis-peptide bonds, *etc.*)
- Geometric queries for macromolecules and ligands
- Search for similar structural motifs and binding sites





RCSB **PDB**
PROTEIN DATA BANK

www.pdb.org • info@rcsb.org

The RCSB provides a variety of tools and resources for studying the structures of biological molecules and their interactions.

RESEARCH COLLABORATORY FOR STRUCTURAL BIOINFORMATICS
Rutgers, The State University of New Jersey
San Diego Supercomputer Center, University of California, San Diego

Ribosome

Protein from
Bacteria

Molecular
Science
Place of
National
Science
Foundation

Outreach & Education

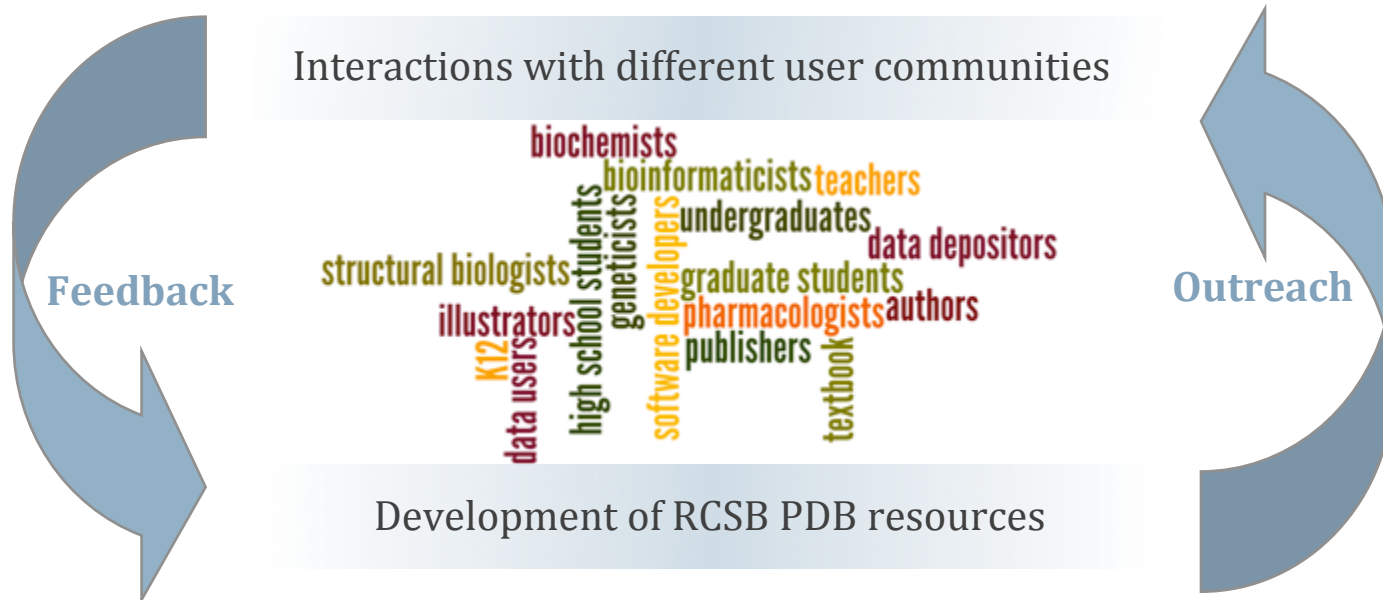
Christine Zardecki

Shuchismita Dutta



Goals

- To facilitate RCSB PDB's mission in the interests of science, medicine and education
- To promote a Structural View of Biology so that experts and non-experts can understand biological processes at a molecular level in 3D



International User Communities

Who are our users?	What are they using?	How do we know?
Biologists: structural biology, biophysics, biochemistry, genetics, Immunology, pharmacology, cell and molecular biology, ...	RCSB PDB website, deposition tools, data	Publication requests, website usage, annotator and Help Desk requests, community outreach, surveys
Other scientists: bioinformatics, software developers, ...	Web Services, search engines, data	Publication requests, website usage, Help Desk requests, community outreach, surveys
Students & teachers	PDB-101	Increase in web hits, email, meeting interactions, specialized workshops/events
Media: Writers, textbook authors, patient advocacy groups, ...	Images, data, information, outreach material, e.g., posters	Publications, image requests
General public: Curious/interested individuals, artists, sculptors, ...	Images, <i>Molecule of the Month</i> , external media	Concerts, media, Wikipedia



User Feedback

Online Survey (Nov 2012)



RCSB PDB User Survey

1. Gender

- Male
- Female

33%



67%

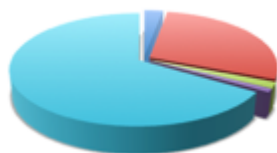
2. Ethnicity

- Hispanic or Latino
- Not Hispanic or Latino

(12% Hispanic or Latino)

3. Race

- American Indian or Alaska Native
- Asian
- Black or African American
- Native Hawaiian or Other Pacific Islander
- White



Examples of Impact and Assessment

- Outreach
 - Annual Biomedical Research Conference for Minority Students (ABRCMS; Nashville, TN, November 13-16, 2013)
 - Plans for additional surveys Y2 and Y4
- Data Out
 - Continued development and support for website features
- Data In
 - Continued depositor-directed surveying using deposition and annotation tools and interactions

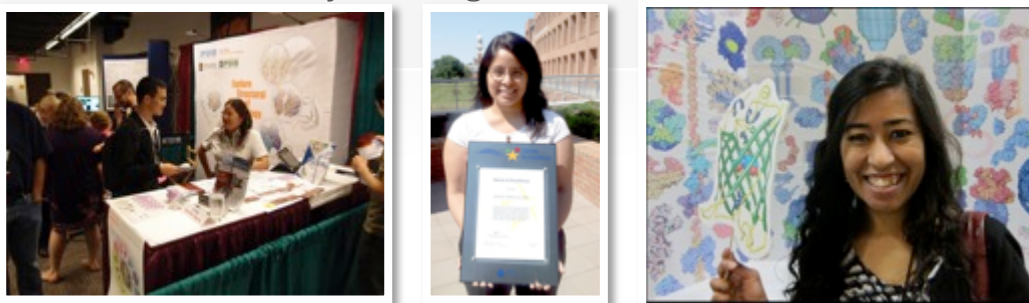


User Communication

Online News & Help Desk



Professional Society Meetings



Local Festivals



Seminars and Classes



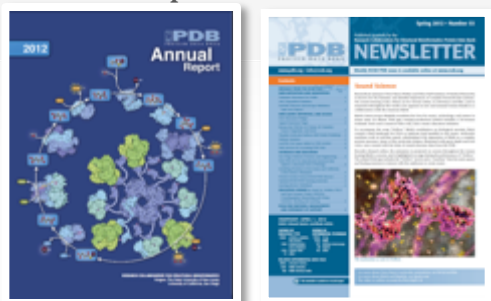
Facebook & Twitter



Staff Activities

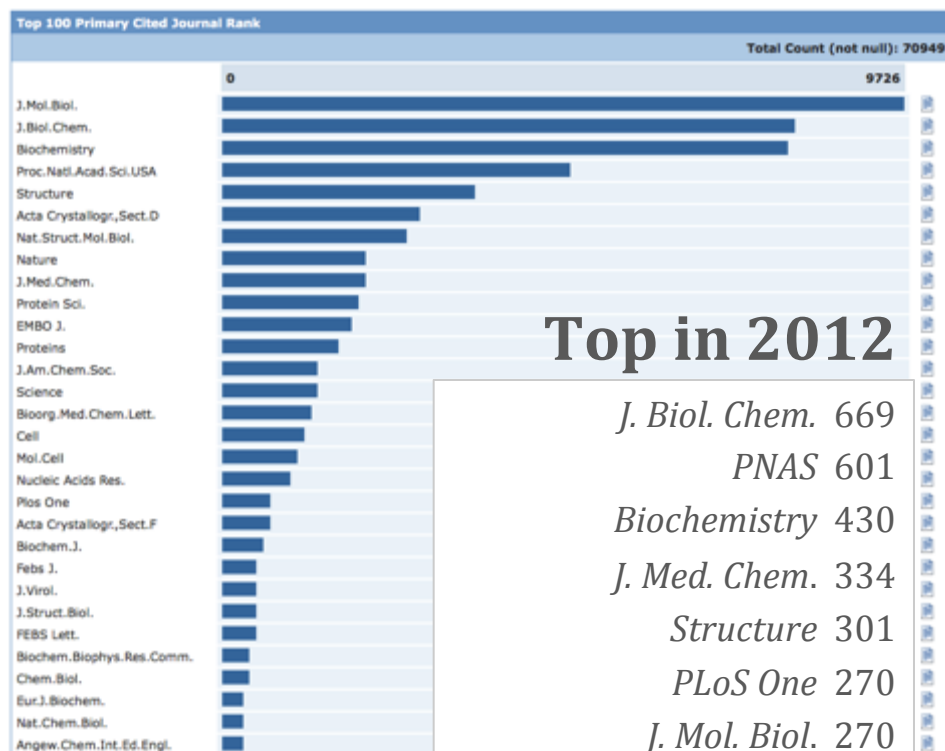


Reports & Newsletters



Journal Interactions

Top Overall



Top in 2012

<i>J. Biol. Chem.</i>	669
<i>PNAS</i>	601
<i>Biochemistry</i>	430
<i>J. Med. Chem.</i>	334
<i>Structure</i>	301
<i>PLoS One</i>	270
<i>J. Mol. Biol.</i>	270
<i>Acta Crystallogr. D</i>	269
<i>Nature</i>	217
<i>NSMB</i>	209

Structure Publication Notifications

Acta Crystallogr. D&F, FEBS J., J. Biol. Chem., J. Mol. Biol., Nature, Nature Comm., NSMB, Nat. Chem. Biol., Proteins, Nucleic Acids Research, PNAS, Science

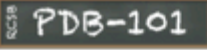
Validation Reports

IUCr Journals


jbc THE JOURNAL OF BIOLOGICAL CHEMISTRY

eLIFE




Online Resources



PDB-101



PDB
PROTEIN DATA BANK

A MEMBER OF THE   


An Educational Resource for Exploring a Structural View of Biology

Contact Us | Print
Jump to a Molecule:


Structural View of Biology | [Educational Resources](#) | [Molecule of the Month](#) | [Understanding PDB Data](#) | [Author Profiles](#)

Structural View of Biology

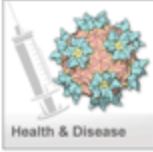
Select one of the key topics below to start exploring. Each subcategory leads to related *Molecule of the Month* articles and examples of proteins and nucleic acids.



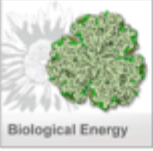
Protein Synthesis



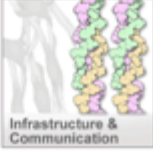
Enzymes




Health & Disease



Biological Energy



Infrastructure & Communication



Biotechnology & Nanotechnology

PDB-101 News

August 2013 Molecule of the Month: Serotonin Receptor

Follow us on Twitter
Get interesting structural biology news and more at @buildmodels

Like us on Facebook

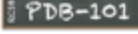
Video Tour of PDB-101
Learn about the different resources available at PDB-101 in this brief video.

About PDB-101


Life is three-dimensional. This extends to life's molecular building blocks—proteins, DNA, and RNA. PDB-101 offers tools to explore the molecules of biological processes that define life.

The **Structural View of Biology** interface starts with key topic categories and subcategories that drill down to individual molecules. It is built around the **Molecule of the Month** series, which regularly describes the structure and function of a molecule. **Educational Resources** provides activities and materials for learning, and **Understanding PDB Data** helps interpret the data archived in the PDB. **Author Profiles** are a new and unique historical and educational tool that offers a timeline display of all structures associated with a particular researcher.




The RCSB PDB develops these resources to support exploration of the structures found in the Protein Data Bank archive of experimentally-determined structures of proteins, nucleic acids, and complex assemblies.



PDB-101



PDB
PROTEIN DATA BANK

A MEMBER OF THE   

An Educational Resource for Exploring a Structural View of Biology

Contact Us | Print
Jump to a Molecule:


Structural View of Biology | [Educational Resources](#) | [Molecule of the Month](#) | [Understanding PDB Data](#) | [Author Profiles](#)

HIV Capsid


July 2013 Molecule of the Month by David Goodby
doi: 10.2210/pdb/mov/mon_2013_7 (PubMed Version)

Keywords: human immunodeficiency virus, capsid, virus assembly, HIV, viral restriction


Discussed Structures



HIV capsid and antibody Fab



HIV capsid hexamer




HIV capsid pentamer

Introduction

Viruses come in many shapes and sizes, ranging from simple protein shells filled with RNA or DNA to membrane-enveloped particles that rival cells in complexity. HIV is one of these complex viruses, surrounded by a membrane and filled with a diverse collection of viral and cellular molecules. The genome of HIV, which is composed of two strands of RNA, is packaged inside a distinctive cone-shaped capsid, which protects the RNA and delivers it to the cells that HIV infects.

Build a Model of the HIV Capsid



Click on the image above to download a paper model of the HIV capsid.

Staying Flexible

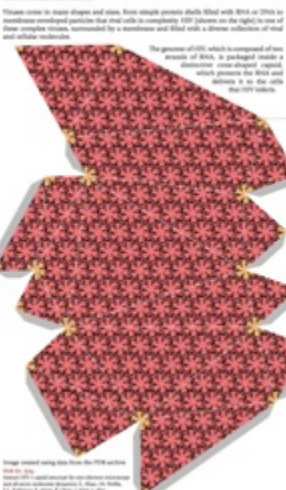
The HIV capsid is built from a single protein, called capsid protein, shown here on the left from HIV entry [1]. Capsid protein, also known as CA or p24, folds to form two domains connected by a flexible linker. This flexibility gives the protein a lot of options for assembly. The larger domain associates with other copies of the protein to form rings of six, and slightly less often, rings of five. The smaller domain then links these rings together to form the larger structure.

Breaking Symmetry

This flexibility allows the formation of structures that aren't as perfectly symmetrical as the capsids of viruses like poliovirus or rhinovirus. Instead, HIV capsid forms an unusual cone-shaped structure, with twelve of the pentamer rings (shown here in orange) and over a hundred hexamers (shown here in red). This model was constructed based on electron micrographs, using PDB entries 2H47 and 2H48 and refining the model with molecular dynamics. Two models were obtained that are consistent with the electron micrograph images: a model with 218 hexamers (shown here from PDB entry 3D41) and a slightly smaller

Build a Paper Model of HIV Capsid

Three views in three shapes and sizes, from simple protein shells filled with RNA or DNA to membrane-enveloped particles that rival cells in complexity. HIV (shown on the right) is one of these complex viruses, surrounded by a membrane and filled with a diverse collection of viral and cellular molecules.



The genome of HIV, which is composed of two strands of RNA, is packaged inside a distinctive cone-shaped capsid, which protects the RNA and delivers it to the cells that HIV infects.

red enzymes and accessory proteins and membrane

capsid

structural proteins and membrane

To build an HIV capsid at 1.8 nm resolution, cut out the model pieces, fold along the white lines, and tape or glue the glue flaps. Add two pieces of string (each with 3.0 cm long) to model the RNA strands.

For an extra challenge, try assembling the model without consulting the diagram or building a model that is more similar to the actual capsid.

For more information about the model, please visit the PDB-101 website at <http://www.rcsb.org/pdb/101>.


Image created using the Protein Data Bank archive and PDB-101.

© 2013 RCSB PDB. All rights reserved. For more information, please visit <http://www.rcsb.org/pdb/101>.

RCSB Protein Data Bank

July 10 10

Download our new PDF to build a 3D paper model of the HIV capsid: http://www.rcsb.org/pdb/101/static/101.do?year=education_discussion&educational_resource&index.html#Paper-Models



Like Comment Share

Genki Terashi, Amanda Qi, Mukesh Kumar and 3 others like this.

RCSB Protein Data Bank, HIV capsid is July's Molecule of the Month <http://www.rcsb.org/pdb/101> https://doi.org/10.2210/pdb/mov/mon_2013_7 July 10 at 11:40am

Christopher Weir My little brother will like this 😊 July 10 at 5:15pm

Resources (cont.)



How to fold PDB-101's DNA paper model



What is a Protein?

MOLÉCULA DEL MES

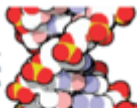
ADN: EL ACIDO DESOXIRRIBONUCLEICO

doi: 10.2210/rcsb_pdb/mom_2001_11

Memoria de sólo lectura

El ADN es una memoria molecular que sólo se puede leer, no es modificable. La información está almacenada de forma organizada en las cadenas del ADN, guardadas a buen recaudo dentro de las células. Cada molécula de ADN está compuesta de una larga cadena *Sinclair* de nucleótidos de nucleótidos, que se manipulan e taría formar la que cada cada pediatá aparados, e enroscan un 2.2.1.2.2...

manipulaciones presentes en los alimentos modificados genéticamente. Para otros en cambio, representa los avances en análisis e identificaciones forenses que han sido



Cómo construir un modelo de la estructura de ADN (Ácido desoxirribonucleico) en papel:

Utilice este folleto para construir un giro completo de una cadena de doble hélice de ADN. Escoja entre: un modelo esquemático para llenar los espacios con los nombres de las bases (a la derecha) o un modelo detallado que demuestra todos los átomos en cada nucleótido (otro lado del papel).



Corte el modelo.

Primero doble todos los pliegues marcados por una línea sólida gris.

Doble las líneas de puntos grises de manera que queden escondidas en el pliegue.

What is a Protein?

Proteins play countless roles throughout the biological world, from catalyzing chemical reactions to building the structures of all living things. Despite this wide range of functions all proteins are made out of the same twenty amino acids, but combined in different ways. The way these twenty amino acids are arranged dictates the folding of the protein into its unique final shape. Since protein function is based on the ability to recognize and bind to specific molecules, having the correct shape is critical for proteins to do their jobs correctly.

Primary Structure
Primary structure is the linear sequence of amino acids as encoded by the DNA. This sequence dictates how the protein will fold and therefore also defines how it will function. A single change in the amino acid sequence of hemoglobin can cause the protein to clump together, resulting in the disease sickle cell anemia.

Secondary Structure
Hydrogen bonds between amino acids form two particularly stable structural elements in proteins: alpha helices and beta sheets. Alpha helices (shown in blue) are the basic structural elements found in hemoglobin, but many other proteins also include beta sheets. The image highlights the pattern of hydrogen bonds (shown in green) that stabilizes alpha helices.

Tertiary Structure
Many functional proteins fold into a compact globular shape, with many carbon-rich amino acids clustered inside away from the surrounding water. The folded structure of hemoglobin includes a pocket to hold heme, which is the molecule that carries oxygen as it is transported throughout the body.

Quaternary Structure
Two or more polypeptide chains can come together to form one functional molecule with several subunits. The four subunits of hemoglobin cooperate so that the complex picks up and delivers more oxygen than is possible with single subunits.

PDB
PROTEIN DATA BANK
www.rcsb.org

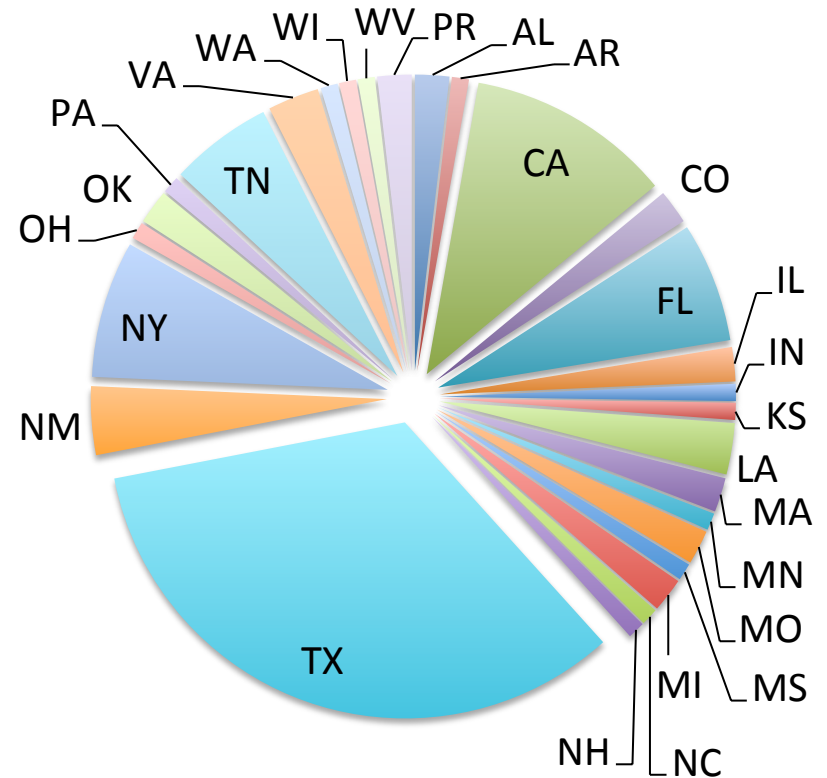


National Reach

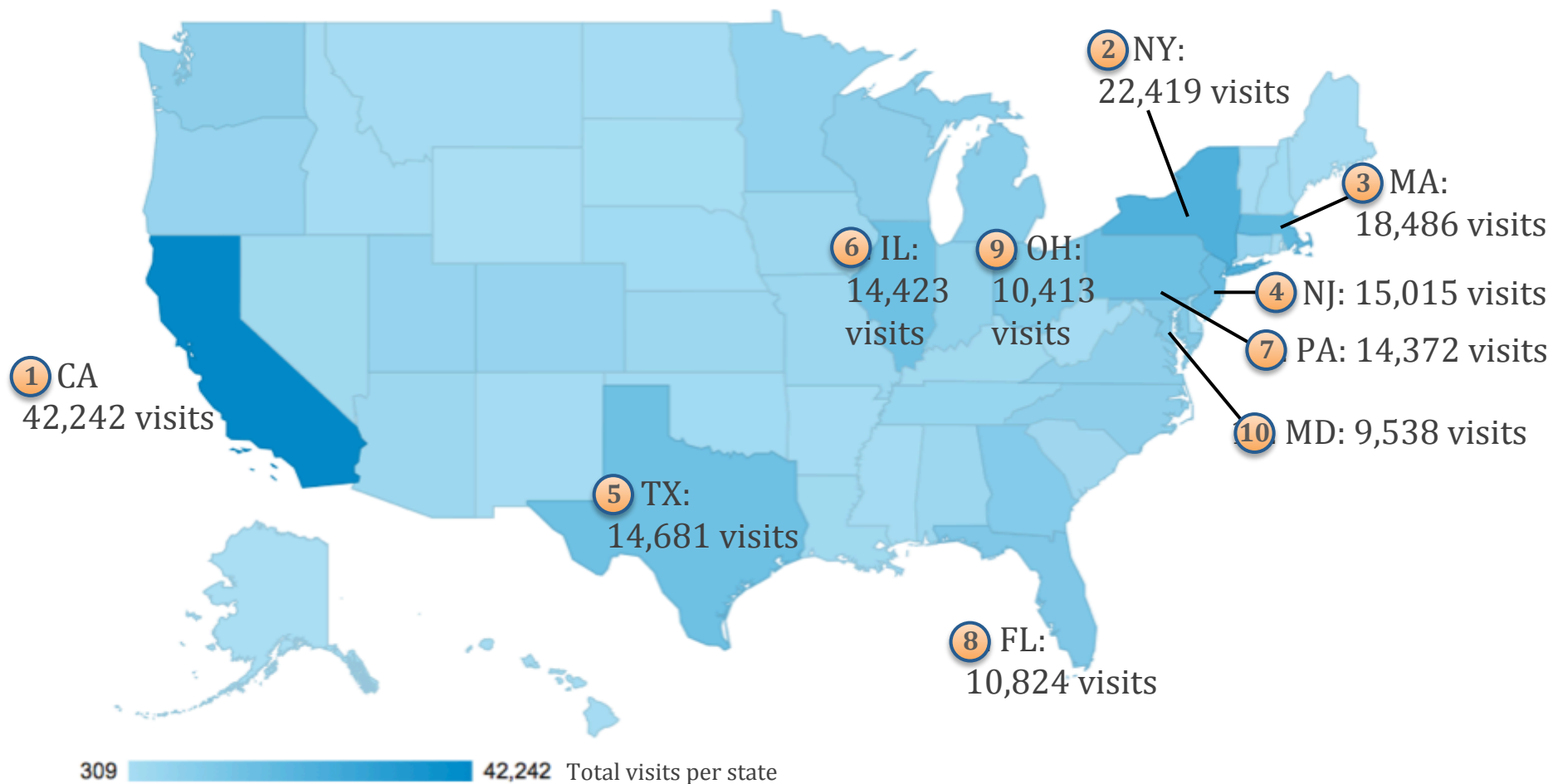
RCSB PDB used in national high school and undergraduate programs, *e.g.*,

- Protein Modeling event at Science Olympiad (MSOE/RCSB PDB)
- SMART Teams (**S**tudents **M**odeling **A** **R**esearch **T**opic; MSOE)
- WestEd: *In Touch With Molecules* project
- RCSB PDB's *Education Corner* describes high school and undergraduate programs at Brookhaven National Laboratory, Oregon State Univ., Stony Brook Univ., Univ. Kansas, and more

RCSB PDB Exhibit Visitors at NSTA 2013
San Antonio, TX



PDB-101 Usage in 2012: US



Educational Experiences

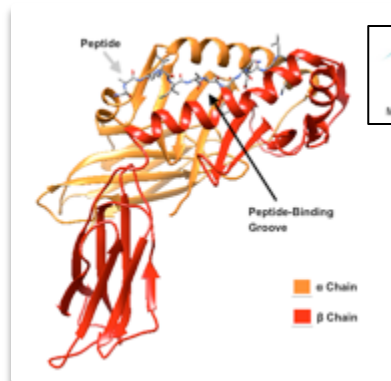


RCSB PDB: An Educational Incubator

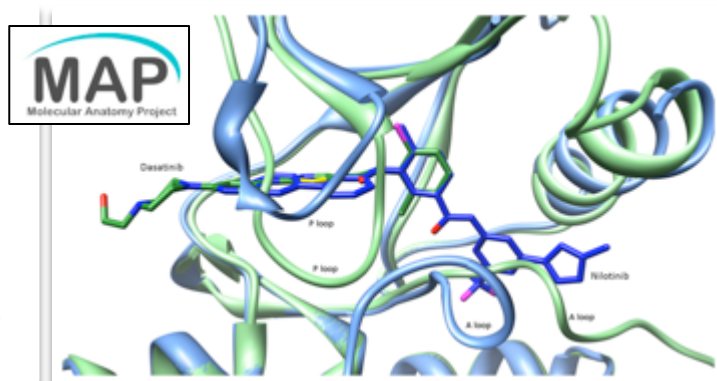
- Specialized workshops
- Courses/publications on a structural view of biology
- Research opportunities
- eLearning/Massive Open Online Courses (MOOCs)



International HS Teacher Workshop,
Princeton Univ., 2013



Student report,
Rutgers Univ., 2013

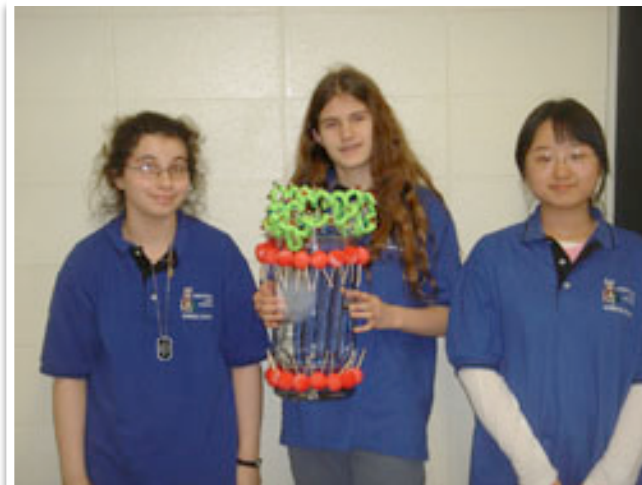


Reviewed and published report,
Rutgers Univ., 2013

PDB on Coursera
UC San Diego, 2013

Science Olympiad: Protein Modeling

- Students build 3D models and answer questions on structure and function
- Developed by Milwaukee School of Engineering (MSOE)
- RCSB PDB
 - Organized training workshops
 - Supervised and judged the event
- Event on hiatus until 2015
- Currently collaborating with MSOE to develop the next Protein Modeling competition (NIDA grant)



Princeton High School students, 2007



NJSO, 2008

Reflecting on Our Experiences



What have we observed?	What are we doing?
Scientists and students presume structural data is too complex	Designing courses to train students and scientists holistically
Specialist knowledge in biology is not required	Developing programs for high school teachers
Developing and disseminating educational materials takes much time and effort	Consolidating education and outreach efforts around key biomedical themes



Working Together to Visualize

	Jan	Feb	Mar	Apr	May	Jun	Jul	Aug	Sep	Oct	Nov	Dec
2012	Student workshop, PHS								Announce Program	Teacher workshop/webinar		HS student learning
2013			HS student learning			Submit reports	Select winners					HS interns

Working Together to Visualize
Molecular Visualization Program for Teachers and Students

Visualizing molecular structures plays a critical role in understanding its function and interactions. The RCSB Protein Data Bank (PDB) provides access to experimentally determined structures of biological macromolecules such as proteins, nucleic acids, and their complexes with various drugs or other small molecules.

The PDB faculty and staff at Rutgers University invite you to participate in a collaborative program designed to teach high school teachers and students about a structural view of chemistry and biology.

Why should you be interested in this program?

- It provides content that covers various NJ Core Curriculum Content Standards in Science, science practice, scientific explanations, reflecting on scientific knowledge
- Technology: impact of technology in learning about human health and disease
- It provides opportunities to interact with authentic data and access to large databases
- It facilitates interdisciplinary study in chemistry, biology, medicine and technology
- It provides teachers with professional development credits for attending hands-on workshops and selected students a summer internship opportunity at the PDB.

Program description:
This program engages high school teachers and students to cutting edge science by introducing them to knowledge, skills, resources and data that is used by scientists in academic and pharmaceutical research.

The program has two phases: (1) teacher training, (2) student training and exploration. In the first phase, high school teachers are invited to participate in a hands-on professional development workshop where they learn to use the PDB archive, visualize bio-molecular structures (like hemoglobin, insulin and DNA), and understand their functions in health and disease. In the second phase, these teachers will teach students about molecular structures and visualization using data archived in the PDB. With support from program developers and their peers, the teachers will find suitable examples of key biological macromolecules for the students to explore.

At the end of a guided exploration, students will create a report that describes a structural view of the molecule and its functions. The best reports from each school will be sent to the program developers for review. Two students with the best reports from all participating schools will be invited to participate in a summer internship with the PDB at Rutgers University.

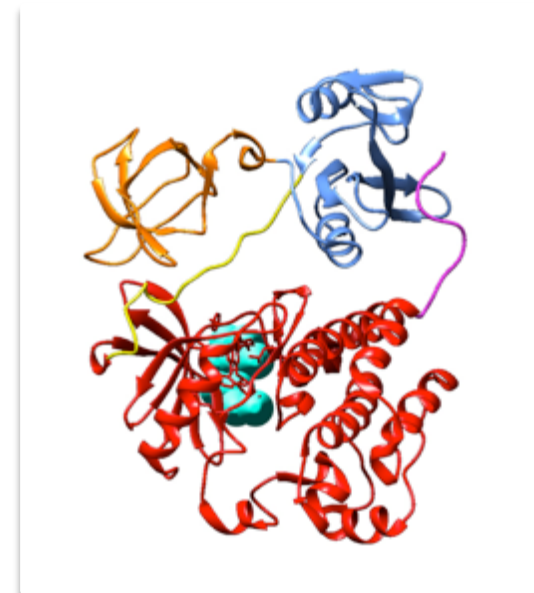
The knowledge, skills and tools that students learn through this program can be applied to various subjects throughout the school year and beyond. Towards the end of the school year, all participating teachers will meet to discuss their experience and summarize them in a brief report. Based on the reports and discussions, selected teachers will be featured in the RCSB PDB newsletter's Education Corner.

Teacher Training Details:
The teacher workshops will be provided on 2 consecutive Wednesdays (October 17 and 24, 2012), between 5:00 and 7:30pm at the Center for Integrative Proteomics Research at Rutgers University.
Tea/Coffee and light refreshments will be provided.
All participants are required to bring in their own laptops to participate in the workshop. Teachers will receive materials and suggestions for implementing the program within the framework of their current curriculum. It is recommended that at least 2 teachers from each participating school attend the workshop so that they can discuss school specific implementation issues. Teachers from each school may choose to report back their experiences individually or as a group.

Registration:
The program is free of charge. To register email what@csb.rutgers.edu. Please use "Working Together to Visualize" in the subject line. In the email include the school name and participant's name and email address.

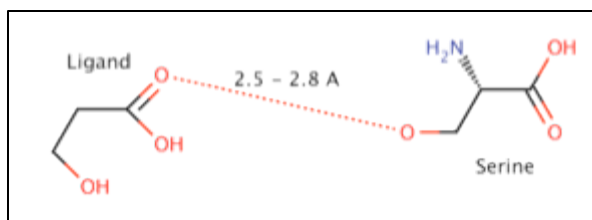
RCSB PDB PROTEIN DATA BANK
Center for Integrative Proteomics Research
Rutgers, the State University of New Jersey
174 Frelinghuysen Rd
Piscataway, NJ 08854
www.rcsb.org | info@pdb.org

- Teacher training workshops
- Teachers teach students
- Student report competition

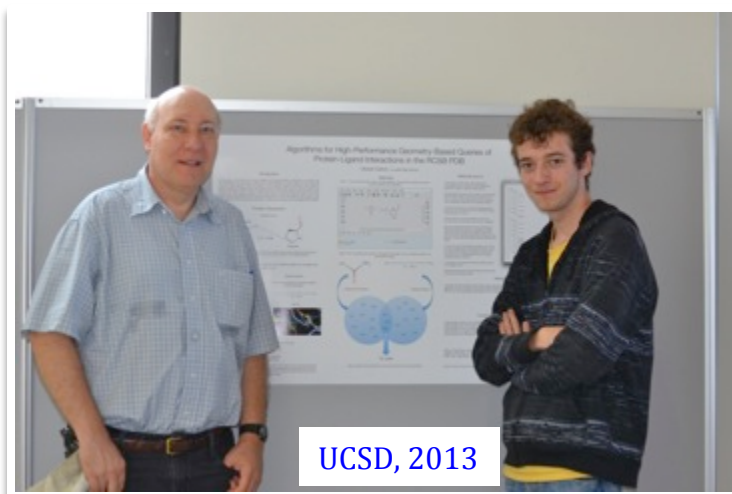
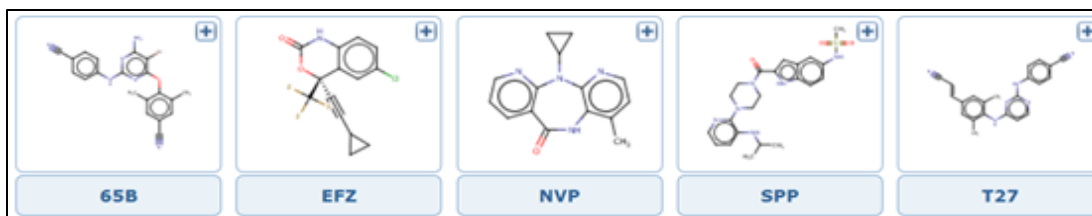
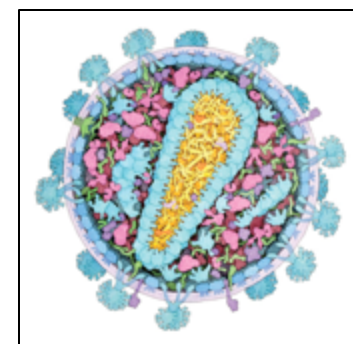


Engaging HS Students in Research

- Algorithms for High-Performance Geometry-Based Queries of Protein-Ligand Interactions in PDB



- Identification of HIV/AIDS related macromolecules in PDB
- Organization of anti-HIV drugs



Outreach Themes

2013/2014/2015: HIV/AIDS
2014/2015/2016: Diabetes
Onwards: *TBD* with educators

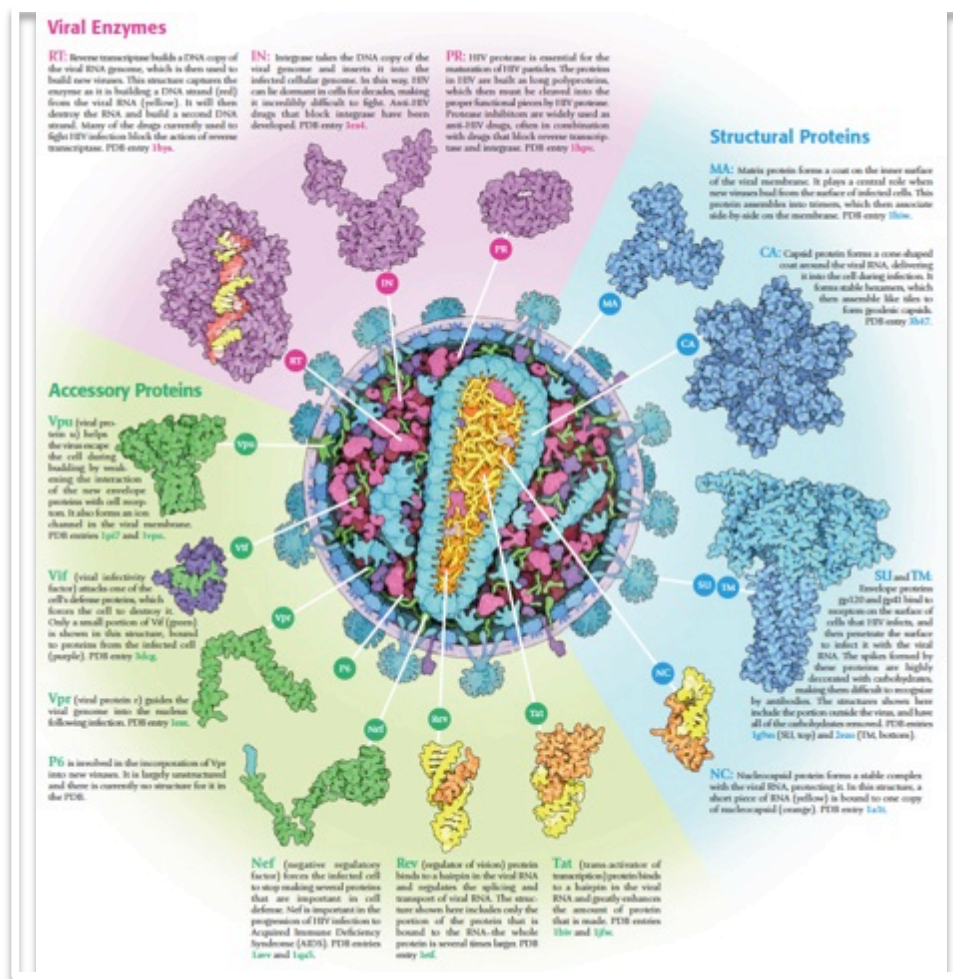
- Help consolidate our efforts
 - Facilitates collaboration with other programs (e.g., RWJMS AIDS program)
- Allow reuse of educational materials with distinct wrappers for different audiences
- Address key health concerns in the nation and provide new perspectives on etiology and treatment
 - Provides access to other funding resources (e.g., Global Health, GAIA centers-Rutgers Univ.)
- Lay the foundation for annual student video challenges
 - Community-created outreach materials



National High School Video Challenge

Year 1: The 3D of HIV/AIDS Video Contest

- RCSB PDB will challenge high school students to create short videos (<5') that tell stories about the 3D molecular processes involved with HIV/AIDS
- Videos will be judged on scientific accuracy, creativity, and educational value by an expert panel



Targeting HS Teachers and Students

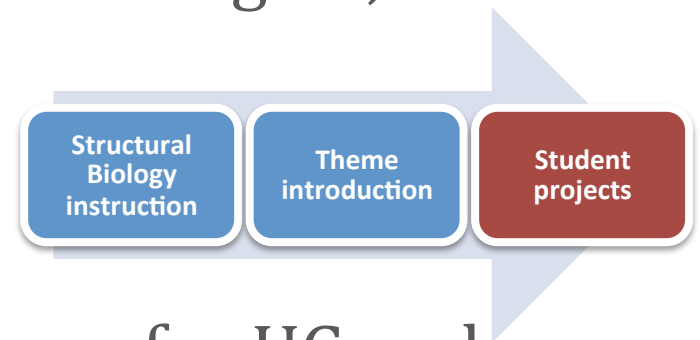
- NJ Pilot program “*Working Together to Visualize: A Molecular Structural View of HIV/AIDS*”, 2013-2014
- National program and Video Challenge, 2014-2015

	Jan	Feb	Mar	Apr	May	Jun	Jul	Aug	Sep	Oct	Nov	Dec
2013								Announce program		Teacher workshop/webinar, symposium		
2014	HS student learning				Submit videos	Select winners, post winning entries	Teacher workshop/webinar		Announce Challenge	Theme based symposium		
	Develop educational materials		NSTA workshop						HS student learning			
2015				Submit videos	Judge & announce winners		Post winning entries online					
	HS student learning											



Designing Courses for Students & Scientists

- Franchising courses first offered at Rutgers, then piloted at
 - Wellesley College
 - Georgetown University
- Developing bi-coastal short courses for UG and Grad students in collaboration with BioMaPS (NIH grant under review)
- Planning to offer these courses through e-learning platforms



Reaching Out Nationwide

- Participating in national education meetings
 - National Science Teachers' Association (NSTA)
- Offering summer workshops for teachers from around the nation at Rutgers and UC San Diego
- Offering webinars for interested teachers unable to attend the Rutgers/UC San Diego workshops
- Creating an online forum within PDB-101 for educators and students to share/discuss experiences and provide feedback



Future of Outreach and Education

- Designing courses for students and scientists in range of institutions, extensible through e-learning platforms*
- Facilitating a Structural View of Biology for students, educators, and scientists through theme based outreach and education
- Targeting HS teachers and students nationwide in analysis and visualization of molecular structures
- Developing online resources that capitalize on growth of mobile device usage in classroom

* Grant under review



Outreach Team

