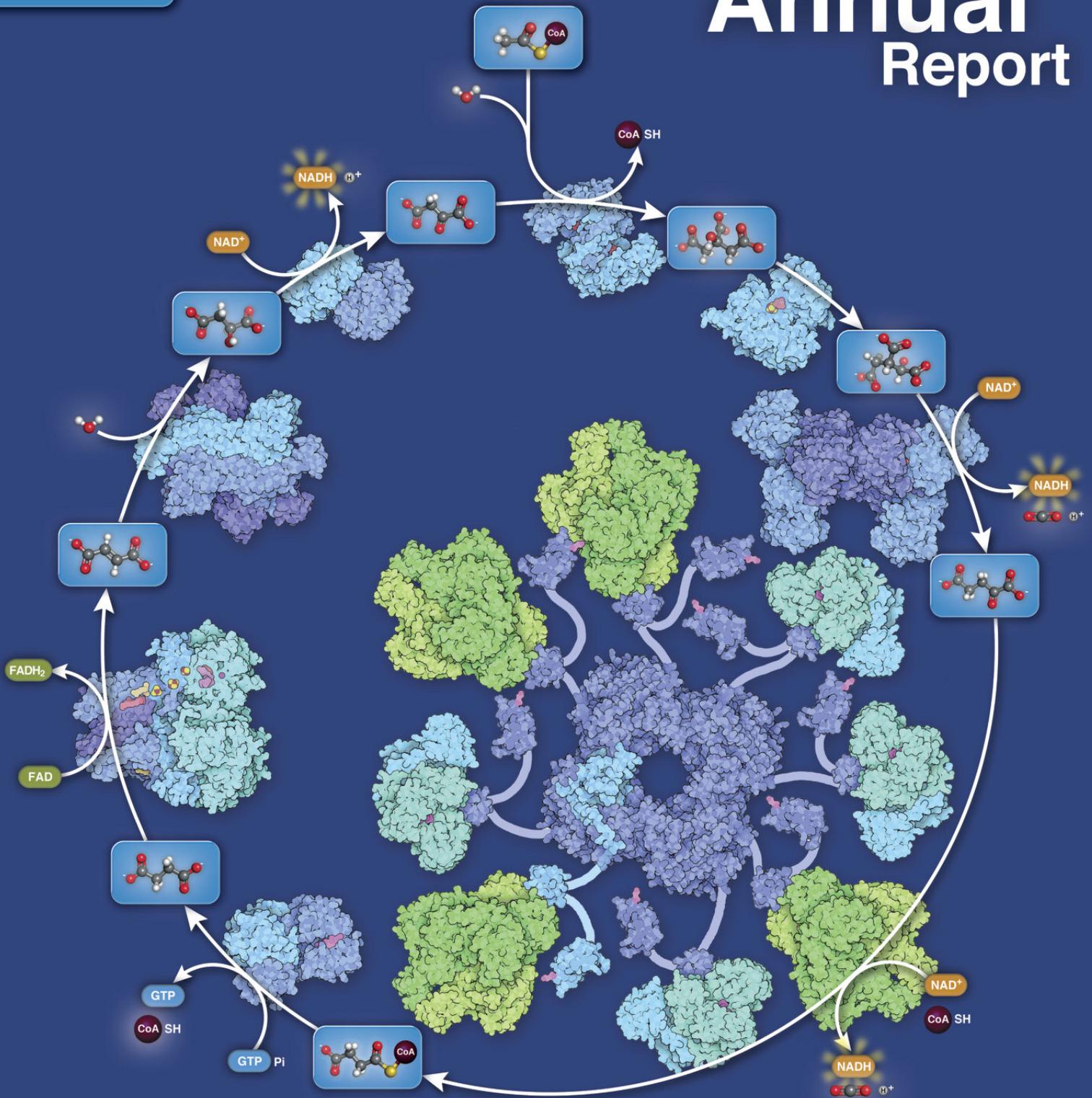


2012

RCSB **PDB**  
PROTEIN DATA BANK

# Annual Report



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

# Table of Contents

	Message from the Director.....	3
	The RCSB PDB and the PDB Archive .....	4
	PDB Structures .....	5
	Data Deposition, Processing, and Annotation .....	6
	Related wwPDB Initiatives .....	8
	Data Access, Query, and Reporting .....	10
	PDB-101, Outreach, and Education .....	14
	Related Resources .....	18
	References and Acknowledgements .....	18

## About the Cover

The front cover maps the proteins and chemical reactions involved with the metabolic pathway called the **citric acid cycle**.

Also known as the Krebs cycle or the tricarboxylic acid cycle, the citric acid cycle is at the center of cellular metabolism. It plays a starring role in both the process of energy production and biosynthesis. The cycle finishes the sugar-breaking job started in glycolysis and fuels the production of ATP in the process. It is also a central hub in biosynthetic reactions, providing intermediates that are used to build amino acids and other molecules. Citric acid cycle enzymes are found in all cells that use oxygen, and even in some cells that don't.

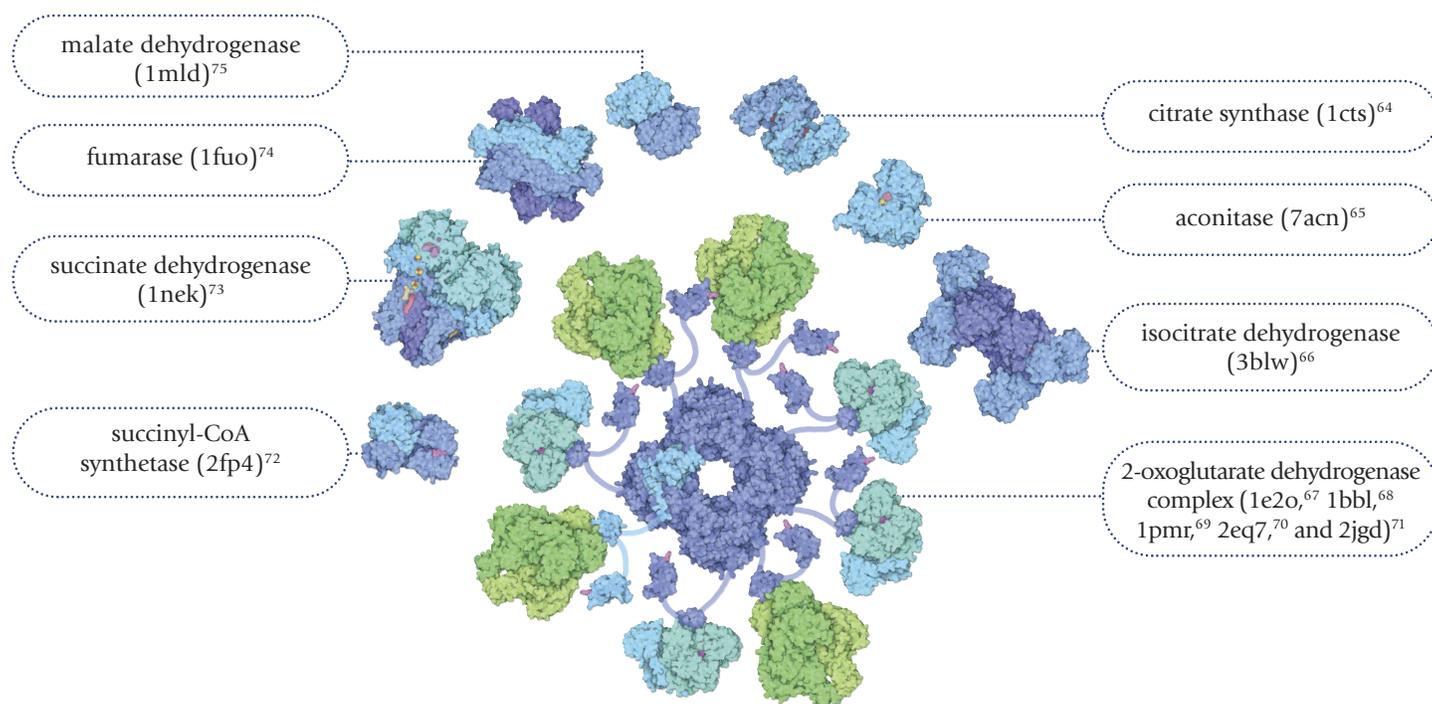
The eight reactions of the citric acid cycle use a small molecule—oxaloacetate—as a catalyst. The cycle starts by addition of an acetyl group to oxaloacetate, then, in eight steps, the acetyl group is completely broken apart, restoring the oxaloacetate molecule for another round. In a typically biological twist, it's not quite this simple. You might imagine that the enzymes could just pop off the two carbon atoms of the acetyl group, using the oxaloacetate as a convenient carrier. However, by carefully labeling particular carbon atoms in these molecules, scientists have found that things get shuffled around a bit, and two carbon atoms in the original oxaloacetate are the parts that are actually released as carbon dioxide. Then, at the end of the cycle, the original acetate atoms are shuffled around to recreate the oxaloacetate.

The structures shown here, taken from several different organisms, are explored further in the October 2012 *Molecule of the Month* ([doi:10.2210/rcsb\\_pdb/mom\\_2012\\_10](https://doi.org/10.2210/rcsb_pdb/mom_2012_10)). Illustration created with guidance from *Fundamentals of Biochemistry*.<sup>63</sup>

The metabolite atoms on the front cover are depicted as follows:

 Oxygen    Carbon    Hydrogen    Sulfur

## Citric Acid Cycle Enzymes





With the goal of enabling a structural view of biology, the RCSB PDB organization takes on many roles. At its core is the Protein Data Bank (PDB) archive, a key repository of information describing proteins, nucleic acids, and other important biological macromolecular machines that help inform our understanding of biology and medicine. Along with our Worldwide PDB collaborators, the RCSB PDB curates, annotates, and makes publicly available the PDB data deposited by scientists around the globe. The RCSB PDB then provides a window to these data through a rich online resource with powerful searching, reporting, and visualization tools for researchers. This information is then streamlined for students and teachers in the form of the educational site called PDB-101.

This annual report gives a glimpse of the RCSB PDB activities that help drive these efforts. During the reporting period of July 1, 2011 - June 30, 2012, the RCSB PDB has reached several milestones while simultaneously preparing for an exciting future.

In March 2012, 80,000 entries became available in the archive. The PDB reached the 70,000 mark in December 2010. Other PDB numbers: 9250 entries were deposited and annotated in the calendar year 2011, and 8076 were released into the archive (compared with 8878 depositions and 7897 released entries in the previous year). Data were downloaded 379 million times in 2011, compared with 209 million in 2010.

Other milestones have been reached near and far. Of local note, team members at Rutgers are happy to report that they have settled into the Center for Integrative Proteomics Research building. This new facility was built to join together structural biology, bioinformatics, and proteomics research groups on campus. These include the RCSB PDB, core facilities, research laboratories, and meeting space. This building has already hosted several meetings and workshops, including the Worldwide PDB's Small Angle Scattering Task Force.

Globally, the wwPDB continues to collaborate on several projects that will take the PDB into the next generation. As the amount of data deposited to the PDB increases, the need to develop more streamlined methods for processing and validation becomes greater. To this end, the wwPDB has made significant progress on the development of a Common Deposition and Annotation tool that will be used by all wwPDB sites to more effectively capture and process PDB depositions. External testing will begin in 2013.

The versatile, knowledgeable, and creative team of scientists, developers, and educators at the RCSB PDB look forward to reaching even more milestones in the coming year.



*Center for Integrative Proteomics Research in Piscataway, NJ*

A handwritten signature in black ink, which appears to read "H. Berman". The signature is fluid and cursive.

**Helen M. Berman**

Director, RCSB PDB  
Board of Governors Professor of Chemistry and Chemical Biology  
Rutgers, The State University of New Jersey

## The RCSB PDB:

### Enabling Science and Education Worldwide

The RCSB Protein Data Bank<sup>1</sup> provides a global resource for the advancement of research and education in biology and medicine by curating, integrating, and disseminating biological macromolecular structural information in the context of function, biological processes, evolution, pathways, and disease states.

Administered by the Research Collaboratory for Structural Bioinformatics (RCSB), the RCSB PDB builds upon the data available in the PDB archive by creating tools and educational materials for use by researchers, teachers, and students studying a variety of fields focused on biology: molecular biology, structural biology, computational biology, pharmacology, and others.

Through its online resources and interactions with its user communities, the RCSB PDB helps to answer broad biological questions where macromolecular structure is key.

### PDB: A Community Resource for Science

In the late 1950s, scientists began to decipher the 3D shapes of proteins at an atomic level. As structures were determined using X-ray crystallography, early computer graphics programs provided interactive views of these macromolecules. The possibilities for science and knowledge seen in these glimpses of myoglobin,<sup>2,3</sup> hemoglobin,<sup>4,5</sup> lysozyme,<sup>6,7</sup> and ribonuclease<sup>8,9</sup> inspired a new field of structural biology.

The potential research that could be enabled by archiving and sharing data from these experiments moved the scientific community to action. Beginning with only seven structures, the PDB archive was established in 1971 to provide both a home and access point for the data produced from these experiments.<sup>10</sup> Scientists would send their coordinate data to the PDB, who would then mail them to interested users.

Today, the PDB contains and supports online access to more than 80,000 biomacromolecules that help researchers understand all aspects of biomedicine and agriculture, from protein synthesis to health and disease to biological energy. Data are deposited by scientists from around the world, and used by a growing number of researchers in a variety of fields.



The PDB archive is managed by the Worldwide Protein Data Bank (wwPDB),<sup>11</sup> a consortium of groups that host deposition, annotation, and distribution centers for PDB data and collaborate on a variety of projects and outreach efforts.

Each wwPDB member develops separate websites, tools, and resources to access and analyze PDB data. As a member, the RCSB PDB collaborates with the wwPDB on matters relating to getting *data in* the archive, while independently developing tools and resources to get *data out*.

### Organization

The RCSB PDB is jointly managed at Rutgers, The State University of New Jersey and the University of California, San Diego (San Diego Supercomputer Center and the Skaggs School of Pharmacy and Pharmaceutical Sciences).

Helen M. Berman, Director of the RCSB PDB, is a Board of Governors professor of chemistry and chemical biology at Rutgers. Professor Berman

## Funding

The RCSB PDB is supported by funds from the:

- National Science Foundation (NSF DBI 0829586)
- National Institute of General Medical Sciences (NIGMS)
- Office of Science, Department of Energy (DOE)
- National Library of Medicine (NLM)
- National Cancer Institute (NCI)
- National Institute of Neurological Disorders and Stroke (NINDS)
- National Institute of Diabetes & Digestive & Kidney Diseases (NIDDK)



was part of the team that first envisioned the PDB archive. Dr. Martha Quesada, Deputy Director (Rutgers), and Professor Philip E. Bourne, Associate Director (UCSD), join her in RCSB PDB management.

The RCSB PDB Team is composed of experts in diverse fields of computer science, biology, chemistry, and education. In addition to working with PDB data, RCSB PDB members co-author scientific papers, exhibit at meetings, present posters and papers, and attend and organize workshops. Staff members also serve as tutors, teachers, and mentors to students of all ages.

The RCSB PDB receives input from an advisory board of experts in X-ray crystallography, nuclear magnetic resonance (NMR), 3D electron microscopy (3DEM), bioinformatics, and education.



2012 RCSB PDB Advisory Committee Meeting

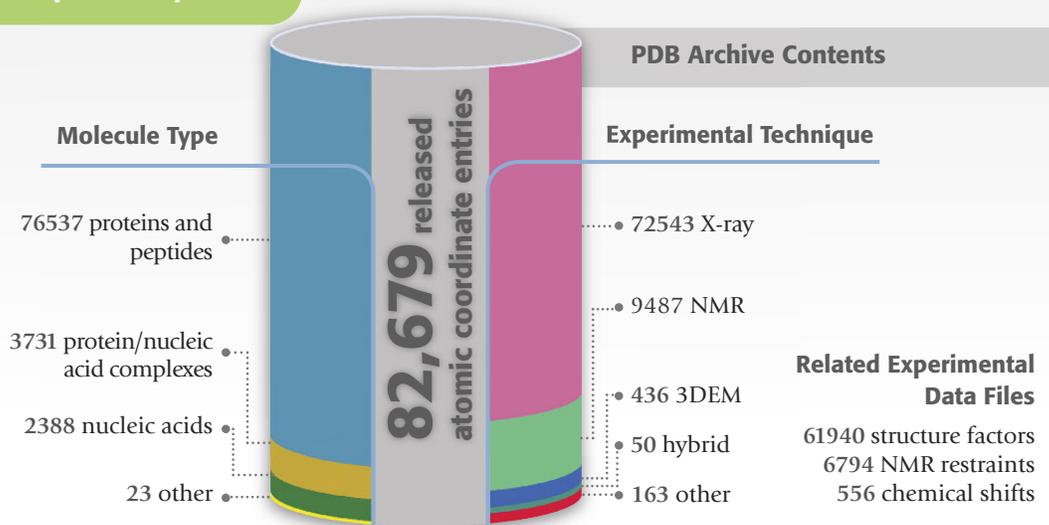


Snapshot: July 1, 2012

Scientists located around the world experimentally determine the 3D structures of biological macromolecules and submit them to be archived in the PDB.

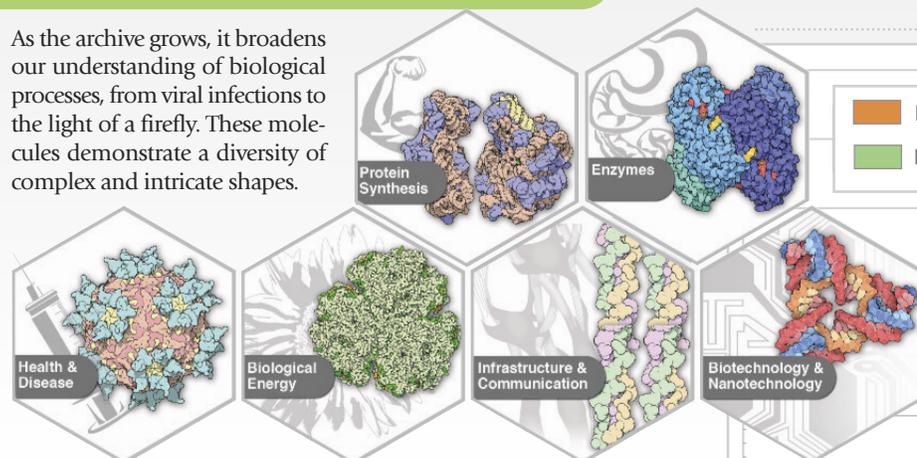
wwPDB annotators then carefully curate each entry. Validation checks are run and reported back to the depositor.

Data from the wwPDB sites are released into the public archive weekly.

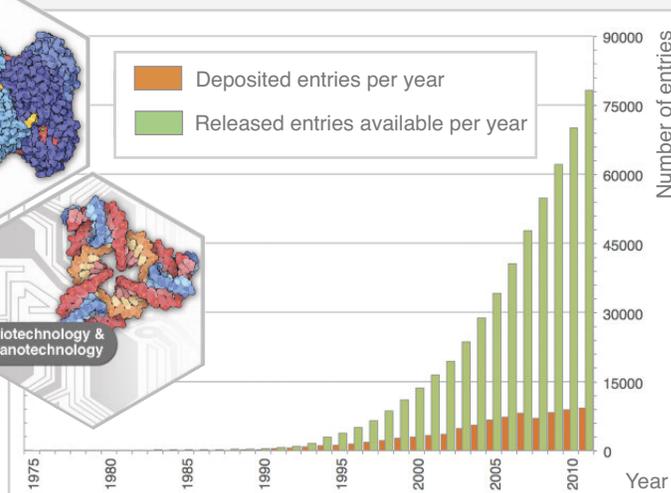


## Nucleic acids, proteins, and complex assemblies

As the archive grows, it broadens our understanding of biological processes, from viral infections to the light of a firefly. These molecules demonstrate a diversity of complex and intricate shapes.



## Growth of the PDB archive

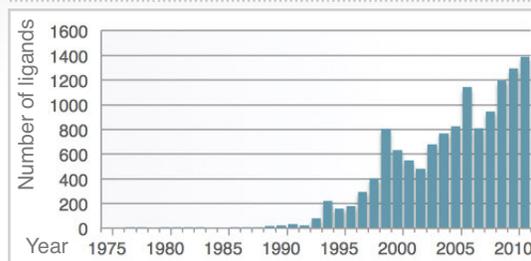


## Chemical components

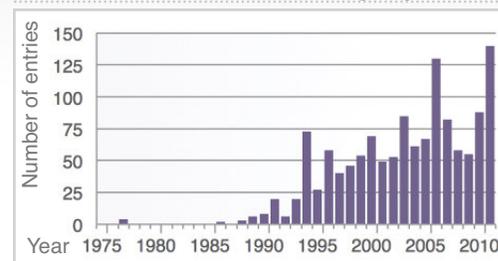
Many of the structures in the archive contain important chemical components—the ligands and other small molecules that are key to drug discovery and development.

Many pharmaceutical blockbusters can be found in the PDB.

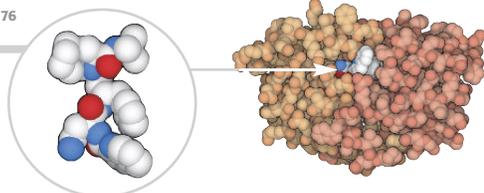
### Number of ligands released per year



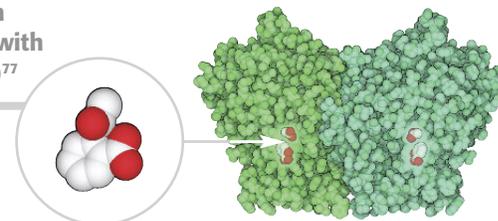
### Number of entries with peptide-like inhibitors/antibiotics released per year



### HIV protease with saquinavir (1hxb)<sup>76</sup>



### Prostaglandin H2 synthase with aspirin (1pth)<sup>77</sup>





## Overview

The goal for data annotation is that each PDB entry accurately represents the structure and experiment.

9682 entries were deposited to the PDB archive and prepared for release by the wwPDB during the period of this report. 7585 of these entries were deposited to the RCSB PDB. On average, structures are processed, reviewed by the author, and finalized for release in two weeks.

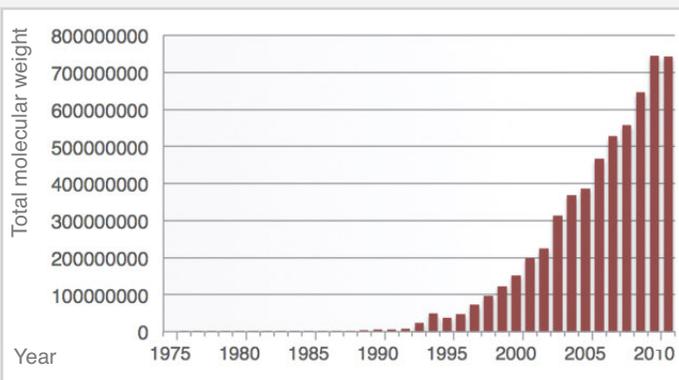
As part of data processing, the annotation team reviews protein sequence and chemistry of small molecule ligands, cross references to other databases, experimental details, correspondence of coordinates with primary data, protein conformation (Ramachandran plot), biological assemblies, and crystal packing. Annotators communicate with depositors to make sure the data are represented in the best way possible.

Sequences for 42% of these depositions were made publicly available prior to the release of the coordinate entry. This helps prevent duplication of structure determination efforts and promotes blind testing of structure prediction and modeling techniques.

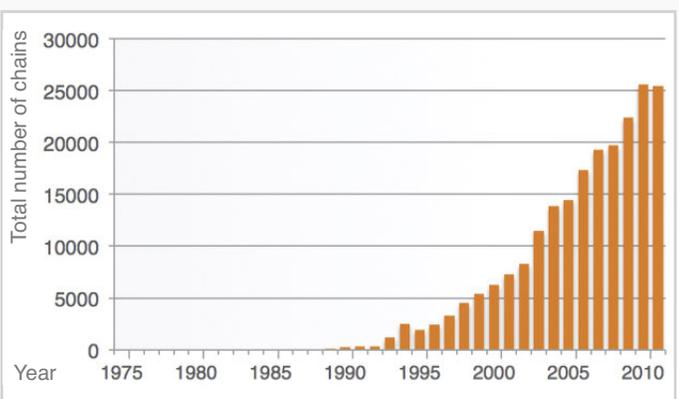
Data files describing each entry are released in different file formats (PDB, PDBx, PDBML-XML)<sup>12,13</sup> along with related experimental data.

## Growth of the complexity of PDB depositions

### Polymeric molecular weight released per year



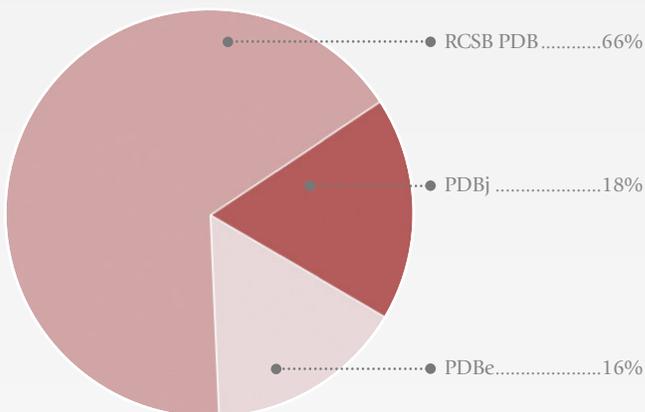
### Number of chains per asymmetric unit released per year



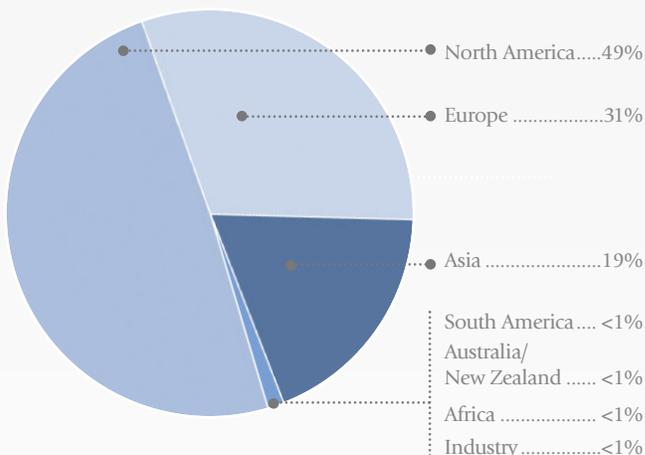
## Deposition statistics

July 1, 2011 - June 30, 2012

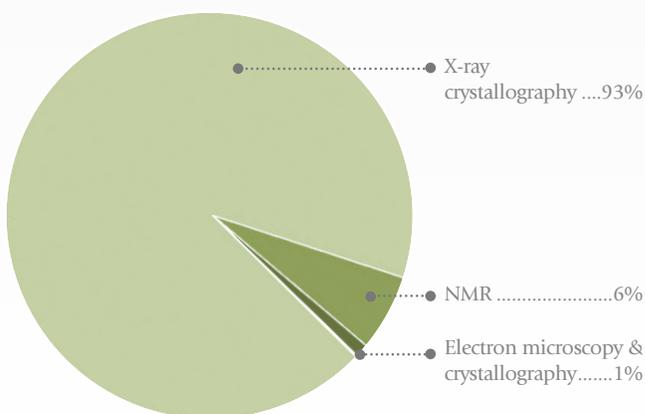
### Processing site for PDB depositions



### Depositor location



### Experimental source





## Tools for Deposition and Validation

The RCSB PDB develops tools that facilitate data validation and submission for depositors, even as structures are in the process of being determined. Software downloads, web servers, and documentation are available at [deposit.rcsb.org](http://deposit.rcsb.org).

### Validation Server

Creates a validation report outside of the annotation process to help depositors prepare depositions.

### SF-Tool

Validates model coordinates against structure factor data; translates structure factors among different formats; and checks for twinned or detwinned data. The latest release supports hybrid data and data from Phenix ([www.phenix-online.org](http://www.phenix-online.org)).

### pdb\_extract

Merges key details from output files produced by X-ray crystallographic and NMR programs into data files for validation and deposition.<sup>14</sup> The latest release supports data from hybrid method experiments, NCS and TLS ranges in BUSTER ([\[ing.com\]\(http://ing.com\)\) and REFMAC \(\[www.ccp4.ac.uk\]\(http://www.ccp4.ac.uk\)\), improved MTZ to mmCIF format conversion, and quality assessment of X-ray data.](http://www.globalphas-</a></p></div><div data-bbox=)

### Ligand Expo

Provides tools for accessing, visualizing, and viewing information in the Chemical Component Dictionary.<sup>15,16</sup> This tool is used internally, and by depositors preparing data submissions.

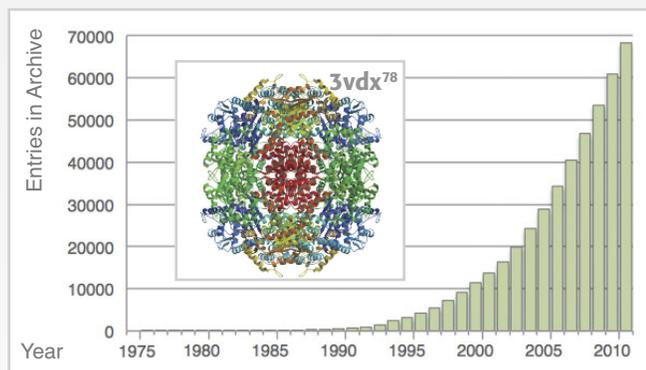
### Deposition Tools

Currently, experimentally-based programs are used by depositors to submit data to the PDB. ADIT<sup>1</sup> is used for most depositions (RCSB PDB and PDBj). ADIT-NMR is a single tool for NMR depositions (BMRB and PDBj). Coordinates, constraint data, and chemical shifts are processed and released by the RCSB PDB and PDBj, while other NMR spectral data are processed and archived by BMRB. Similarly, EMDep integrates the deposition of 3DEM maps to EMDB and model coordinates to PDB (RCSB PDB and PDBe).

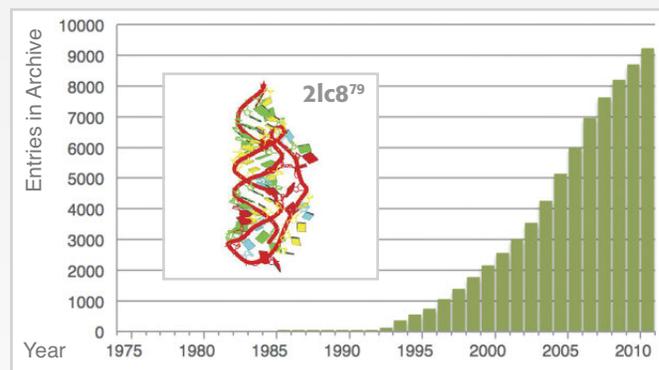
RCSB PDB annotators use an internal version of ADIT that incorporates modules from the wwPDB Common Deposition and Annotation tool (see next page) to prepare entries for release.

## Deposition growth by experimental method

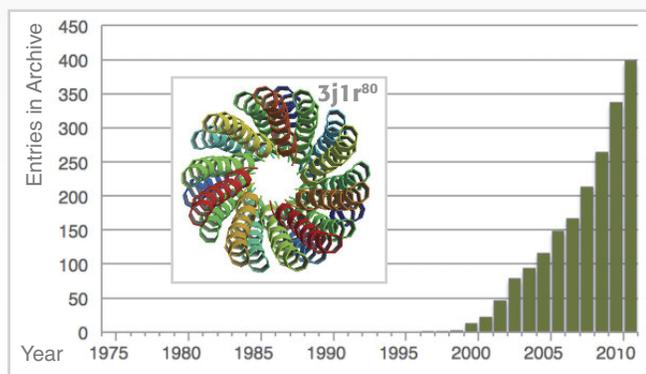
### Growth of X-ray entries



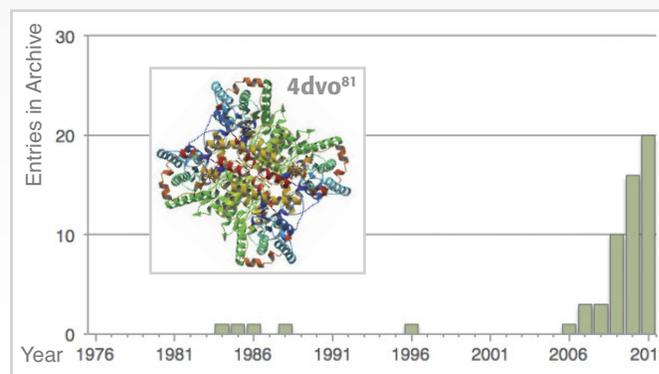
### Growth of NMR entries



### Growth of 3DEM entries



### Growth of Hybrid entries





## Overview

A major focus of the wwPDB is maintaining the consistency and accuracy of data across the archive. As the PDB grows, developments in structure determination methods and technologies can challenge how all structures are represented.

Several projects are underway that help the wwPDB support high quality and uniform data in the PDB archive. A key project is the development of the Common Deposition & Annotation Tool that will help wwPDB annotators address increases in the complexity and experimental variety of PDB depositions, maximize the efficiency and effectiveness of data processing and annotation, and provide for the higher quality and completeness of PDB entries.

Information about all of these projects can be found at [wwpdb.org](http://wwpdb.org).

## Data Reviews and Remediation

The wwPDB regularly reviews the archive to correct errors and inconsistencies. These remediation efforts involve the creation of new versions of the data files.<sup>16-18</sup>

Past remediation efforts have improved the representation of sequences, ligand chemistry and nomenclature, biological assemblies, residual B-factors, peptide inhibitors and antibiotics, and entries in nonstandard crystal frames, and other improvements. Since the July 2011 remediation effort, detailed descriptions of any changes made to data are recorded in the data entry.

During this report period, a multi-year review, analysis, and remediation effort began. Areas being addressed are the representation of carbohydrates, protein modifications, metal-containing ligands, translation of non-standard crystal frame, recalculation of B factors, translation of dissociated assemblies, and X-ray multiple models. It is anticipated that remediated data will be released in 2013 and 2014.

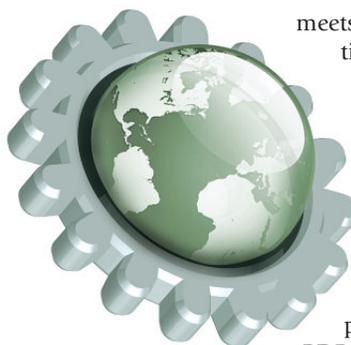
## Data Dictionaries

PDB entries are carefully curated and annotated using a standard data dictionary for macromolecular structure (mmCIF)<sup>19</sup> and for the small chemical components found in PDB entries (Chemical Component Dictionary).<sup>16</sup>

A new Biologically Interesting molecule Reference Dictionary (BIRD) has been created for peptide-like antibiotic and inhibitor molecules. BIRD contains chemical descriptions, sequence and linkage information, and functional and classification information as taken from the core structures and from external resources. Infrastructure has been developed and deployed for the identification and annotation of peptide-like antibiotics and inhibitors found in PDB entries.

## A Common Deposition & Annotation Tool for the Future

The wwPDB partners have been developing the next generation of processing software to ensure the value and effectiveness of the PDB in the future. The new system of tools and processes will maximize the efficiency and effectiveness of data handling and support for the scientific community going forward. In doing so, they will address anticipated increases in complexity and experimental variety of future submissions. Made up of experts from all partner sites, the wwPDB Common Deposition and Annotation (D&A) Tool project team



meets regularly and works in conversation with the wwPDB Directors and user communities.

The new D&A system uses modular architecture that will enable maintenance and extensibility over its lifetime. Anchor components, including the Ligand Processing Module, have already become part of the production pipeline at the RCSB PDB and PDBj. The Ligand Processing Module has been shown to deliver up to a 70% increase in processing efficiency. The Sequence Processing Module, Validation Module,<sup>26</sup> and functionality covering all added-value annotations, both calculated and manual, are in testing at all sites.

The overall system offers: automated batch data uploads, webform data entry, deposition restart and re-upload without loss of general information, new submissions easily built on previous depositions, display of percentage of deposition completeness, and detailed structure validation reports.

A representative group of depositors will begin testing in 2013.

## PDBx/mmCIF Working Group

A working group has been established to enable the direct use of PDBx/mmCIF format files in the major macromolecular crystallographic software tools.

At a September 2011 meeting held at PDBe, the wwPDB met with key software developers (including CCP4, Phenix, and Global Phasing Ltd., which account for ~%85 of recent depositions) to discuss the limitations of the PDB file format for supporting data for large structures, complex chemistry, and new and hybrid experimental methods.



*PDBx/mmCIF Working Group*

Rather than designing an entirely new format, the consensus among developers was to actively switch to the PDBx/mmCIF format, particularly for use in deposition. A meeting outcome was the formation of a working group for the adoption of the PDBx/mmCIF format for deposition and as an exchange format between programs for macromolecular crystallography. Part of this process will be to make it easier to capture more information in deposition and archiving.

## wwPDB Validation Reports and Task Forces

To help ensure the accuracy of entries in the PDB, submitted data is compared with community-accepted standards during the process of validation. As part of the annotation process, wwPDB members provide depositors with detailed reports that include the results of geometric and experimental data validation.<sup>20-23</sup> These reports, available as PDFs, provide an assessment of structure quality while keeping the coordinate data confidential.

Several journals use these validation reports as part of the manuscript review process, including the *Journal of Biological Chemistry* and the journals of the International Union of Crystallography. The wwPDB encourages all journal editors and referees to include these reports in the manuscript submission and review process.

To improve validation methods in the PDB, method-specific Validation Task Forces in X-ray Crystallography, NMR, 3DEM, and Small Angle Scattering have been convened to collect recommendations and develop consensus on additional validation that should be performed, and to identify software applications to perform validation tasks. Reports from the X-ray<sup>24</sup> and 3DEM<sup>25</sup> committees have been published, with other reports to follow.

These recommendations will be incorporated into the wwPDB data processing procedures and tools as part of the wwPDB Common Deposition & Annotation Tool development.<sup>26</sup>

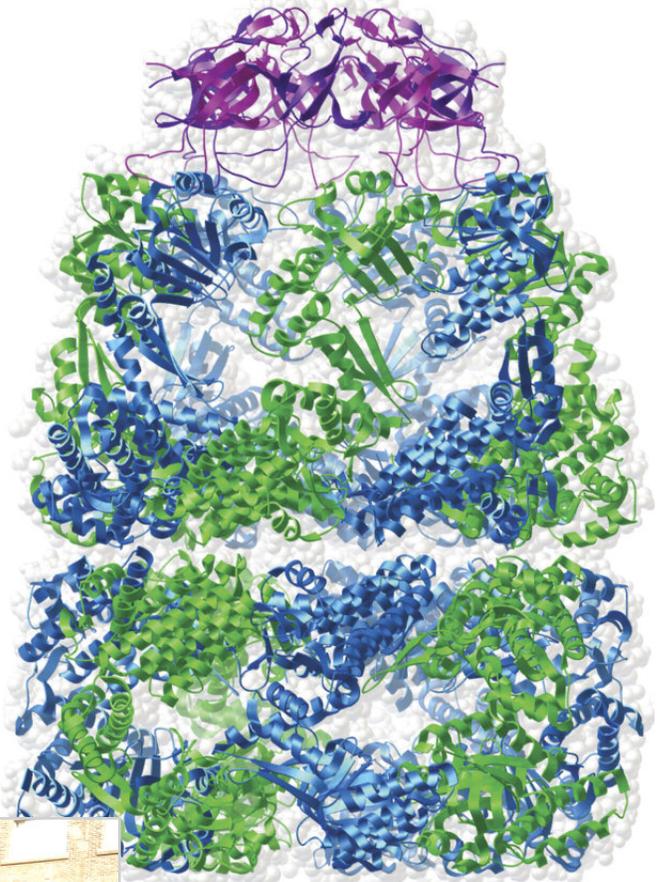
## Time-stamped Archival Copies

As part of a wwPDB initiative, time-stamped snapshots of the PDB archive are added each year to <ftp://snapshots.wwpdb.org> to provide readily identifiable data sets for research on the PDB archive. Scripts are available to help users create local copies of all or part of the PDB archive or snapshots.

## 3DEM Validation Task Force Report

Outcome of the first electron microscopy validation task force meeting

R. Henderson, A. Sali, M. L. Baker, B. Carragher, B. Devkota, K. H. Downing, E. H. Egelman, Z. Feng, J. Frank, N. Grigorieff, W. Jiang, S. J. Ludtke, O. Medalia, P. A. Penczek, P. B. Rosenthal, M. G. Rossmann, M. F. Schmid, G. F. Schroder, A. C. Steven, D. L. Stokes, J. D. Westbrook, W. Wriggers, H. Yang, J. Young, H. M. Berman, W. Chiu, G. J. Kleywegt, C. L. Lawson (2012) *Structure* 20: 205-214.



▲ **PDB ID: 1pcq** C. Chaudhry, G. W. Farr, M. J. Todd, H. S. Rye, A. T. Brunger, P. D. Adams, A. L. Horwich, P. B. Sigler (2003) Role of the gamma-phosphate of ATP in triggering protein folding by GroEL-GroES: function, structure and energetics. *EMBO J.* 22: 4877-4887. Image created using PyMOL ([www.pymol.org](http://www.pymol.org)).

▲ Participants at the Small Angle Scattering Task Force held at Rutgers in July 2012. From left to right: Martha Quesada (RCSB PDB), Torsten Schwede (University of Basel), Wayne Hendrickson (Columbia University), Helen Berman (RCSB PDB), Gerard Kleywegt (PDBe), John Tainer (The Scripps Research Institute), Mamoru Sato (Yokohama City University), Catherine Lawson (RCSB PDB), Jill Trehwella (Task Force Chair, University of Sydney), Jasmine Young, John Westbrook (RCSB PDB)



## Data archiving and distribution



### PDB Archive

Coordinated updates of the PDB archive and the wwPDB websites occur weekly. As the archive keeper for the wwPDB, the RCSB PDB maintains the PDB archive at <ftp://ftp.wwpdb.org>. Updates of the PDB archive and the RCSB PDB website are coordinated to occur at the same time as the other wwPDB sites.

A total of 8701 coordinate entries, 8034 structure factors, 559 constraints, and 465 chemical shifts were released during this period.

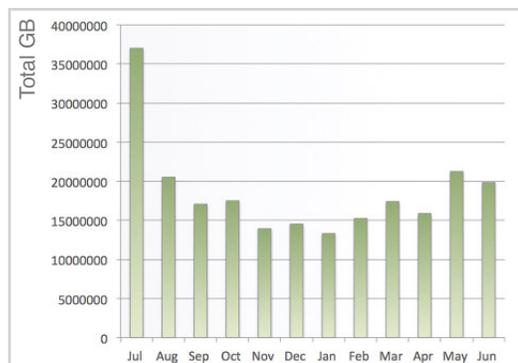
Monthly traffic from the FTP and rsync protocols typically amounts to about 18 million downloads (3.5 terabytes of data) by about 11,500 unique users for a total of approximately 2.5 terabytes of data.

PDB data are also downloaded from the wwPDB member FTP sites and individual web portals at PDBe and PDBj.



### Data downloads (FTP + rsync) per month (GB)

July 1, 2011 – June 30, 2012



### RCSB PDB

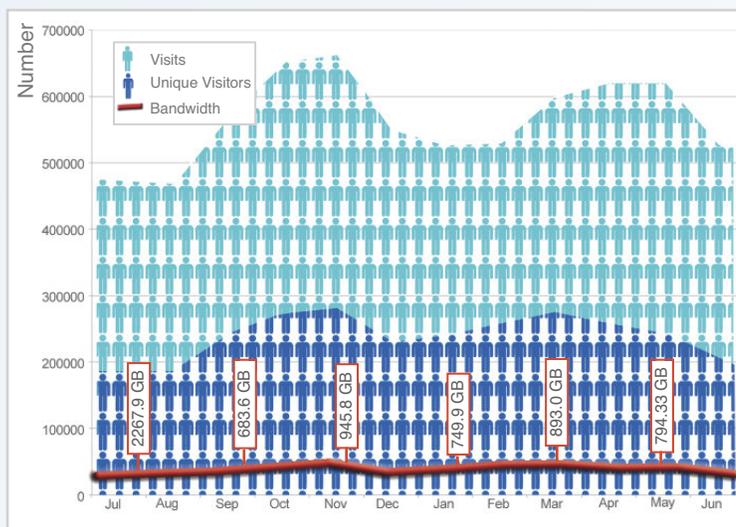


The RCSB PDB offers access to PDB data via the website, FTP server (supporting FTP and rsync access), mobile devices, and Web Services.<sup>27</sup>



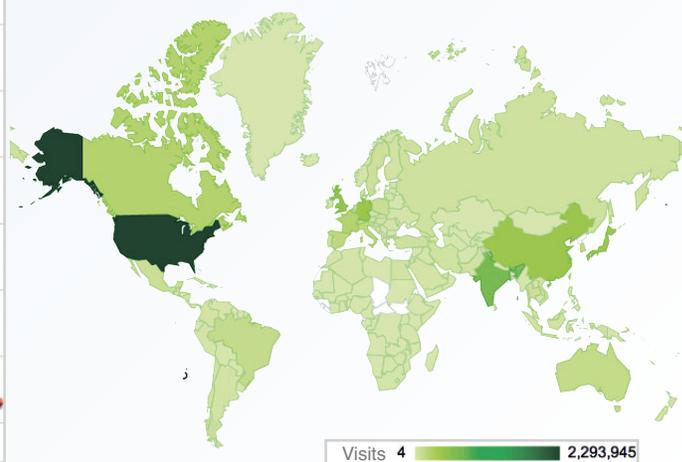
The website ([www.rcsb.org](http://www.rcsb.org)) is accessed by about 241,000 unique visitors per month (up 14% from last year's 211,000) from 140 countries. Around 750 Gigabytes of data are transferred each month from the website. Website statistics are tracked using Google Analytics.

### Website traffic (July 1, 2011 – June 30, 2012)



### RCSB PDB website usage by country

(July 1, 2011 – June 30, 2012)





## Tools for searching PDB data



### Quick Searching

The top search bar at [www.rcsb.org](http://www.rcsb.org) facilitates easy, intuitive, and precise queries.

Typing text in the top search bar launches an interactive pop-up box that suggests possible matches that are organized by categories ranging from author name to ontology terms. The top bar search can be also be limited to quick searches on author, macromolecule name, sequence, or ligand by selecting the related icon.

The top menu search recognizes particular types of syntax.



### Advanced Searching

The Advanced Search tool provides the capability to combine multiple searches for specific types of data using a logical AND or OR. The result is a list of structures that comply with ALL or ANY of the search criteria, respectively.

Individual data items are organized by category; contextual help and examples are available to guide users. New options



### Drill-down Pie Charts

Standard characteristics of PDB entries—resolution, release date, experimental method, polymer type, organism, taxonomy—are used to create searchable data distribution summaries. Users can tour the PDB archive by drilling down through combinations of these significant properties. These pie charts are available from the home page and for all query results. Frequently asked questions such as *What is the distribution of enzymes?* can be quickly answered for the entire PDB, a search results set, or the latest release by using these pie charts.



### Database Browsing

Database browsers offer another way to navigate the PDB by hierarchical classifications (EC number,<sup>31</sup> SCOP,<sup>32</sup> CATH,<sup>33</sup> Transporter Classification<sup>34</sup>) or ontologies (Gene Ontology/GO,<sup>35</sup> NCBI Taxonomy).<sup>36</sup>



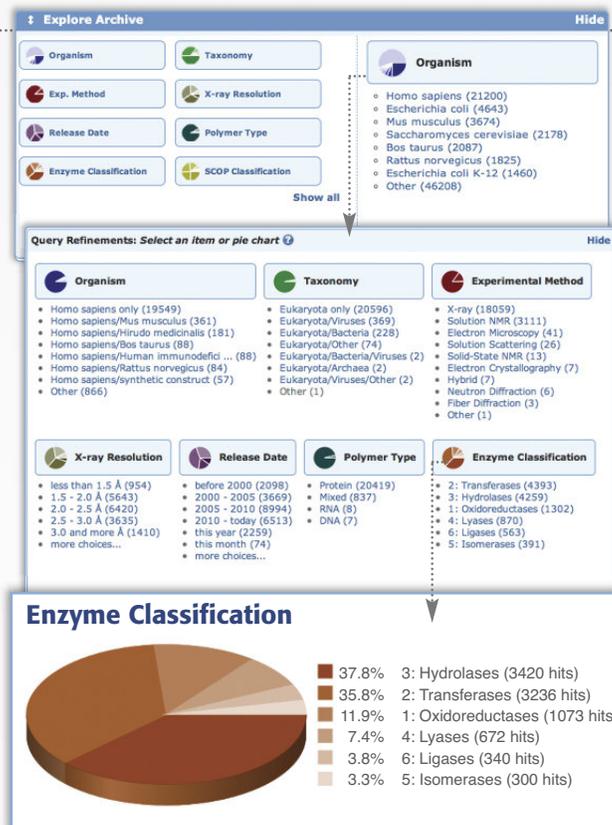
### Web Services

The RCSB PDB supports RESTful Web Services for programmatic access to PDB data. New services provide access to: pre-released sequences in FASTA format, which is useful for protein or nucleic acid structure blind predictions; custom reports in XML, CSV, and Excel format for sequence, structure, function, ligand, experimental details, and structure annotation data; Pfam annotations;<sup>30</sup> and precalculated domain-based structural alignments.

Entering SMILES strings<sup>28</sup> will suggest options to perform substructure, exact structure, or similar chemical structure searches, while typing in a sequence will offer different BLAST<sup>29</sup> search options.

Using these categories helps to quickly differentiate among possible search results. For example, autosuggestions for the input *bird* identify structures with authors whose names contain "bird" and structures from the organism bird. Users who want to perform simple, non-categorized text searches can click the magnifying glass icon or press return.

include quick searches by experimental and/or molecule type, searches based on structure determination method, the ability to find structures containing interresidue connectivity (LINK records) that cannot be inferred from the primary structure, and Pfam annotations, which are updated weekly by running Pfam's hidden Markov models.<sup>30</sup>



This example shows the path to the EC distribution of structures from humans. Clicking on any link returns the structures that match all selected parameters. This feature is available for all search results and for the entire PDB archive.



## Exploring Search Results

### Individual structures

For every entry in the PDB, an RCSB PDB Structure Summary page provides an overview of the structure; derived data from CATH, SCOP, Pfam, and GO; tools to examine the sequence, sequence domains, and sequence similarity; detailed information relating to the entry's citation, biology and chemistry, experiment, and geometry; and links to related resources. Several molecular viewers, including Jmol<sup>37</sup> and the RCSB PDB's Protein Workshop and Ligand Explorer,<sup>38</sup> help users view the molecule interactively. The Literature view provides the abstract, publication details, and a list of other articles referencing the entry.<sup>39</sup>

A new widget displays descriptions of the revision history of a PDB entry. Since July 2011, the details of revisions made to a PDB entry have been publicly recorded in the mmCIF/PDBx file (in the category PDBX\_VERSION). The Revision History widget lists these descriptions in the right hand column of the Structure Summary page.

### Structure result sets

The Query Results Browser lets users refine search results, access related *Molecule of the Month* features, and review individual entries.

The default view of query results (called the "detailed view") includes information about each entry from a variety of fields, such as authors, compound, citation, classification, and residue count. Users can toggle through this and other views, including a condensed view (title and macromolecule name), a gallery of images, and a timeline display of images. The views are synchronized; selecting or deselecting a structure in one view will have the same effect in the others.

Search results can be also be refined by editing the query using Advanced Search and drilling down through pie charts.

The final results set can be used to create a variety of tabular reports, download all sequences, or access all coordinate files.

Filter:  Check All View: Condensed Reports: Select one... Sort: Release Date

PDB ID	Title	Macromolecule Name
<input checked="" type="checkbox"/> 1BL8	POTASSIUM CHANNEL (KCSA) FROM STREPTOMYCES LIVIDANS	PROTEIN (POTASSIUM CHANNEL
<input checked="" type="checkbox"/> 1F6G	POTASSIUM CHANNEL (KCSA) FULL-LENGTH FOLD	VOLTAGE-GATED POTASSIUM CHANNEL
<input checked="" type="checkbox"/> 1J95	KCSA potassium channel with TBA (tetraethylammonium) and potassium	VOLTAGE-GATED POTASSIUM CHANNEL
<input checked="" type="checkbox"/> 1JQ1	POTASSIUM CHANNEL (KCSA) OPEN GATE MODEL	VOLTAGE-GATED POTASSIUM CHANNEL
<input checked="" type="checkbox"/> 1JQ2	POTASSIUM CHANNEL (KCSA) OPEN GATE MODEL	VOLTAGE-GATED POTASSIUM CHANNEL
<input checked="" type="checkbox"/> 1K4C	Potassium Channel KcsA-Fab complex in high concentration of K+	antibody Fab fragment heavy chain, ...

The results for a query for voltage-gated potassium channels can be browsed using "condensed" and "gallery" views.

### Ligands

For each chemical component in the PDB, a Ligand Summary page is available. Improved Jmol visualization and 2D images, information about subcomponents, and links to DrugBank<sup>40</sup> are some of the features recently added to these pages. Additionally, binding affinity data from PDBbind<sup>41</sup> is available from Structure Summary pages.

For search results, Ligand Summary Reports include information about selected ligands such as formula, molecular weight, name, SMILES string, which PDB entries are related to the ligand, and how they are related. For each ligand included in the report, a sub-table can be selected to show lists of all related PDB entries that contain the ligand, the entries that contain the ligand as a free ligand, and entries that contain the ligand as part of a larger, polymeric ligand. Image collages can also be created.

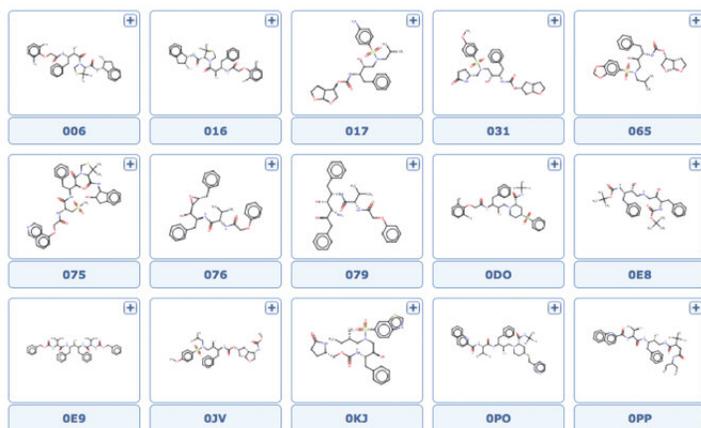


Image collage for the ligands associated with HIV protease structures in the PDB.



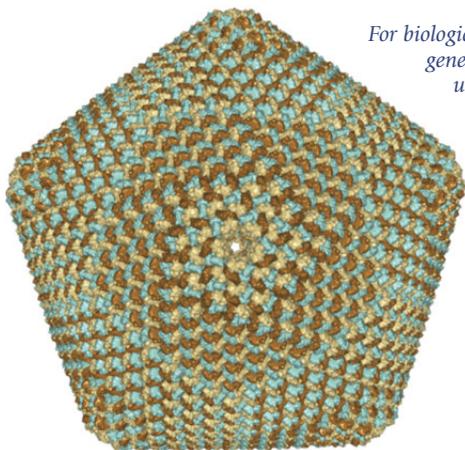
Ligand Explorer view of HIV protease with nelfinavir from entry 1ohr.<sup>56</sup>



## New Features

### Visualizing molecular surfaces

Protein Workshop supports molecular surfaces to aid in the display of quaternary structure, protein-protein interactions, and binding sites. Surfaces can be color-coded by chain, entity, and hydrophobicity using colorblind-safe color schemes.



For biological assemblies, surfaces are generated in Protein Workshop using space group symmetry operators. This allows the display of even the largest assemblies in the PDB on a standard laptop computer, such as the PBCV-1 virus capsid with 5040 chains shown (PDB ID: 1m4x).<sup>57</sup> Options to export custom high-resolution images are provided for the creation of posters and publications.

### Personalize the RCSB PDB with MyPDB



MyPDB provides personalized access to PDB data. It can store any type of RCSB PDB structure search, including a particular keyword, sequence, or Advanced Search composite query, to be run at any time with the click of a button. MyPDB can be customized to provide email alerts when new entries matching a saved search are released in the PDB

archive. Personal annotations and notes can be saved on any entry's Structure Summary page, along with a bookmark list of favorite structures. Accounts can be created and accessed from the home page.

### RCSB PDB Mobile

The new RCSB PDB Mobile app provides fast, on-the-go access. Search the entire PDB, view the latest weekly release of structures, access MyPDB, view the entire catalog of *Molecule of the Month* articles, and more using either a WiFi or cellular data connection and an iPad/iPhone device. RCSB PDB Mobile includes an extremely fast molecular viewer, developed by collaborator Takanori Nakane (Kyoto University), and enables the user to download and view any structure from the RCSB PDB website within seconds. Visit the Apple Store to access the free download of RCSB PDB Mobile.

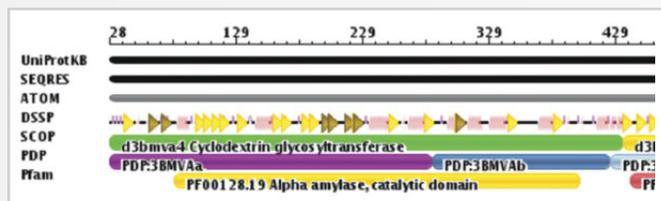
A version of the app for the Android platform is in development.



Available on the App Store

## Domain-based structural alignments

The Comparison Tool and Structure Summary's 3D Similarity option provide pre-calculated systematic structure comparisons for all PDB proteins. The new version of this feature provides domain-based protein structure alignments instead of chain-based alignments. A graphical display of the domain architecture has been added to the 3D Similarity tab, where a user can select a domain of interest and then retrieve structurally similar entries.



The sequence diagram for PDB ID 3BMV<sup>42</sup> shows the corresponding UniProtKB sequence, the SEQRES and ATOM records, and the various annotations available.

Results for domain d3bmva4

SCOP Domain ID d3bmva4 (chain 1) vs. representatives of other sequence clusters (chain 2)

Rank	Results	Domain 2	Title	P-value	Score	Rmsd	Len1	Len2	%ID	%Cov1	%Cov2
1	view	d2guya2	Alpha-amylase A2guy A:1	1.11E-16	788.51	2.61	406	381	22	91	97
2	view	PDP:4AEE	ALPHA AMYLASE, CATALY	1.33E-15	643.75	2.01	406	309	20	74	97
3	view	d1j0ha3	neopullulanase1j0h A:124	1.64E-14	766.83	3.21	406	382	22	89	94
4	view	PDP:4AEF	NEOPULLULANASE (ALPHA	1.05E-13	684.99	2.5	406	404	21	82	82
5	view	d1j1a3	ALPHA-AMYLASE 1j11 A:1	1.08E-12	705.92	3.17	406	432	21	90	85
6	view	d11wja2	4-ALPHA-GLUCANOTRANSI	2.32E-12	661.58	3.0	406	391	19	83	86
7	view	d1wzaa2	alpha-amylase A1wza A:2	2.9E-12	670.87	3.07	406	409	21	85	84
8	view	PDP:3DHU	Alpha-amylase	1.56E-11	594.12	3.05	406	338	15	79	95
9	view	d2bhua3	MALTOOLIGOSYLTREHALC	4.15E-11	592.41	2.75	406	411	14	81	80
10	view	d1jdca2	1,4-ALPHA MALTOTETRAH	1.66E-10	558.27	3.03	406	357	15	77	88
				91E-10	558.78	3.01	406	361	18	77	86
				4E-10	551.68	3.02	406	354	15	76	88
				39E-10	532.21	3.07	406	344	15	79	93
				08E-9	523.60	3.18	406	347	14	77	90
				1.7E-9	567.45	3.09	406	472	13	84	72

View 1 - 15 of 97

In the Structure Summary 3D Similarity tab for 3BMV,<sup>42</sup> selecting the "view" circled for the domain PDP:3DHUa launches the structure alignment view for the alpha amylase domains of 3BMV and 3DHU.<sup>58</sup>



From students learning about protein synthesis to animators creating a music video, the biological macromolecules of the PDB entertain and educate a diverse audience. To empower and inspire non-experts, the PDB-101 website was created

to package together resources that promote exploration in the world of proteins and nucleic acids for teachers, students, and the general public.

Major topic areas include Author Profiles, *Molecule of the Month*, Educational Resources, the Structural View of Biology browser, and the *Understanding PDB Data* resource.

## Author Profiles

Author Profiles are a new and unique historical and educational tool that offers a timeline display of all structures associated with a particular researcher. Structures are sorted by deposition date, with the first instance of a protein or protein complex highlighted. A search box is provided to search by author or structural genomics center name.

## Molecule of the Month

*Molecule of the Month* articles describe everything from AAA+ proteases to zinc fingers. Since 2000, the RCSB PDB has published articles that describe the structure and function of a molecule along with customized interactive views, discussion topics, and high resolution images. Created by David S. Goodsell (RCSB PDB, The Scripps Research Institute), this feature provides an easy introduction to macromolecular structures, shows how structures function, and highlights their importance in our lives.

Selected structures discussed in an article can be examined in individual Structure Focus pages, which describe how the entry is related to the *Molecule of the Month* article, display Jmol and sequence views, highlight the related ligands and ligand interactions, and link to the PubMed abstract.

The *Molecule of the Month* archive is organized by Title, Date, and Category, with links to individual PDF and ePub versions. A quick alphabetical pulldown menu available from the top of every PDB-101 page helps users jump from cholera toxin to influenza neuramidinase to anthrax toxin.

From the main RCSB PDB site, *Molecule of the Month* articles are available from the home page, top bar searches, structure search results, and corresponding Structure Summary pages.

2012

**4GBR**
**October 24**

Molecule: Beta-2 adrenergic receptor, Ly...

N-Terminal T4 Lysozyme Fusion Facilitates Crystallization of a G Protein Coupled...

**4DKL**
**March 21**

Lysozyme, Mu-type oploid receptor

Crystal structure of the mu-opioid receptor bound to a morphinan antagonist

*Truncated view of Author Profile for Brian Kobilka. Author Profiles display all entries associated with a particular author.*

## Molecule of the Month Milestones

Date	Molecule	Milestone #
January 2000	Myoglobin	#1
January 2002	Thrombin	#25
February 2004	Glycolytic Enzymes	#50
March 2006	Tissue Factor	#75
April 2008	Adrenergic Receptors	#100
May 2010	Parvoviruses	#125
June 2012	Sliding Clamps	#150

*The 150<sup>th</sup> article on Sliding Clamps was published during this report period.*



## Educational Resources

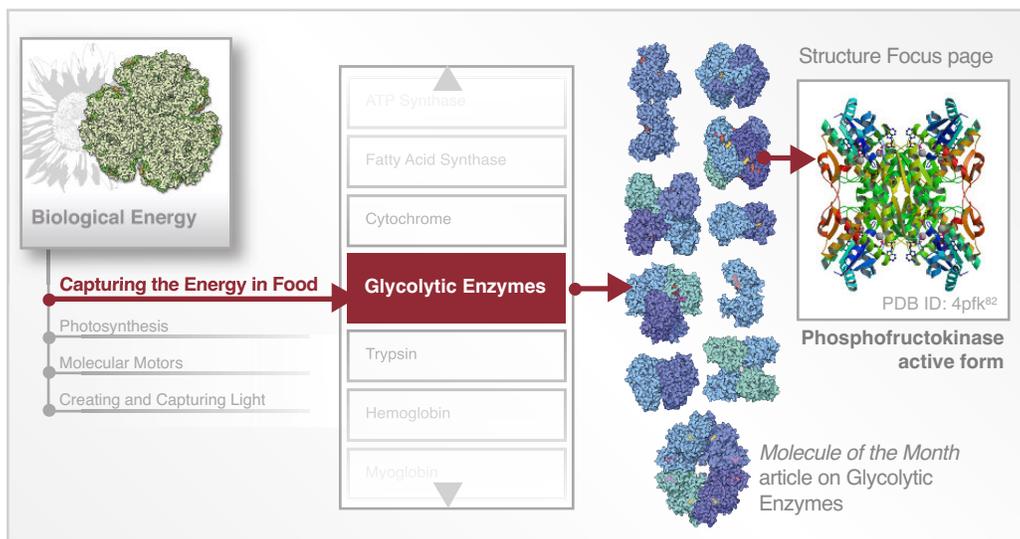
Educational Resources offer a variety of free materials, including posters describing the shape and size of different *Virus Structures* and *How Do Drugs Work*. Animated images illustrate concepts and structures, such as the *Structural Biology of HIV*. Downloadable activities and lesson plans can be used to build models of the dengue virus, DNA, and a new model of tRNA.

An archive of the RCSB PDB Newsletter's *Education Corner* shows how the PDB archive and RCSB PDB resources are used in classrooms all over the world. Past topics have included a crystallography program for middle school students, building protein models with high school students, and the educational experiences that led recent RCSB PDB Poster Prize awardees to study structural biology in graduate school.

## Structural View of Biology

Built around the *Molecule of the Month* series, the *Structural View of Biology* promotes a top-down exploration of the PDB in the context of biology. Readers can travel through high-level functional categories (such as Protein Synthesis and Health and Disease) and descriptive subcategories (such as Replication or Immune System) to access relevant articles that describe molecules in simple terms and provide specific examples.

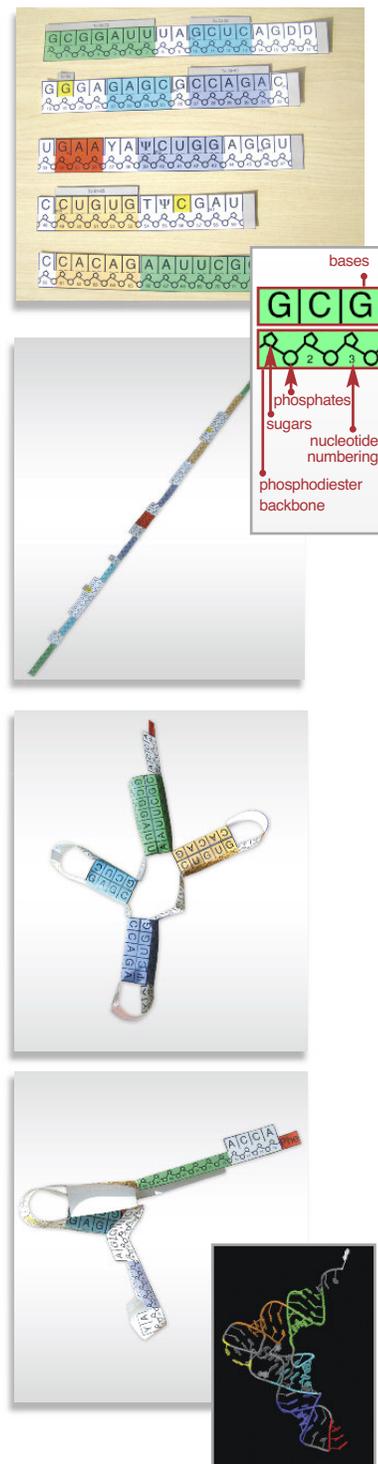
From the Biological Energy category, for example, users can first select the subcategory Capturing the Energy in Food to find an article about the Glycolytic Enzymes involved in breaking sugar for energy and then explore highlighted examples of these enzymes, such as the active and inactive forms of phosphofructokinase.



Use the Structural View of Biology to browse from biological function to PDB entry.

## Understanding PDB Data

Understanding PDB Data is a reference to help explore and interpret individual PDB entries. PDB data files can challenge experts and non-experts alike. An understanding of how a data file came to be—from the experiment to data release—is needed to analyze the data. Using text, images, and interactive views, this feature helps researchers and educators understand how to interpret PDB data, visualize structures, read coordinate files, and know about potential challenges.



Learn about the structure and function of tRNA by building a paper model in this new PDB-101 activity.



## Community Interactions

The RCSB PDB works with different user communities to develop a resource that supports a variety of audiences. Our outreach efforts aim to inform users about the RCSB PDB while collecting feedback to help develop a powerful resource for science, medicine, and education.

Users include biologists from a variety of specialties, scientists from other disciplines, students and educators at all levels, authors and illustrators, and the general public.

While [www.rcsb.org](http://www.rcsb.org) and PDB-101 serve as the primary tool for outreach, staff interact directly with users at international meetings, workshops, presentations, festivals, and more. Prizes are awarded for student posters at selected meetings.

Electronic help desks, newsletters, and flyers are also used to communicate with users and solicit detailed feedback about the resource.



Lead Annotator Jasmine Young at the American Crystallographic Association's 2012 Meeting.

## Publications

The RCSB PDB website is updated regularly with news about improvements to existing tools and new resources.

Published in print and online, the quarterly newsletter describes and highlights recent activities. Flyers, brochures, and tutorials are distributed to users at meetings and online to promote tools such as *pdb\_extract*, *MyPDB*, *Ligand Explorer*, *RCSB PDB Mobile*, and more.

Journal articles covering a diverse array of subjects are published regularly. Recent articles have described the *wwPDB*,<sup>43</sup> the history of the PDB<sup>44,45</sup> and the 2011 Symposium celebrating the 40<sup>th</sup> anniversary of the archive;<sup>46</sup> circular permutation in proteins;<sup>47</sup> CASP9;<sup>48</sup> and the foundations of structural molecular biology.<sup>49</sup> Reports from the X-ray crystallography<sup>24</sup> and 3DEM<sup>25</sup> Validation Task Forces were also published.

## Online help

Users of the RCSB PDB write daily to the electronic help desk ([info@rcsb.org](mailto:info@rcsb.org)) to receive guidance and offer suggestions on all aspects of the archive, website, and beyond. During this report period, users from around the world initiated more than 1000 conversations with the help desk. The majority of these questions related to data deposition and processing and to query functionality. Users also

## Popular Questions to [info@rcsb.org](mailto:info@rcsb.org)

**Q:** *Can I use this figure in my lecture/thesis/book?*

**A:** All images, including Molecule of the Month illustrations, can be reprinted. Citation and full usage information is at <http://bit.ly/RV2Vsh>

**Q:** *What is a PDB chain?*

**A:** The term "chain" refers to a protein or nucleic acid macromolecule. A PDB entry may have multiple copies of the same "chain." For example, hemoglobin entry 4hhb contains 4 protein chains: 2 copies of the alpha chain and 2 copies of the beta chain. See PDB-101's *Understanding PDB Data* for more.

**Q:** *How can structures on hold for publication get released?*

**A:** The wwPDB aims to release entries as close to the publication date as possible. With some entries, journals provide advance publication notice, and for others, citation information is sent by the structure authors or other users. Software is also regularly used to search for matches in PubMed. Have a citation? Please send it to the help desk.

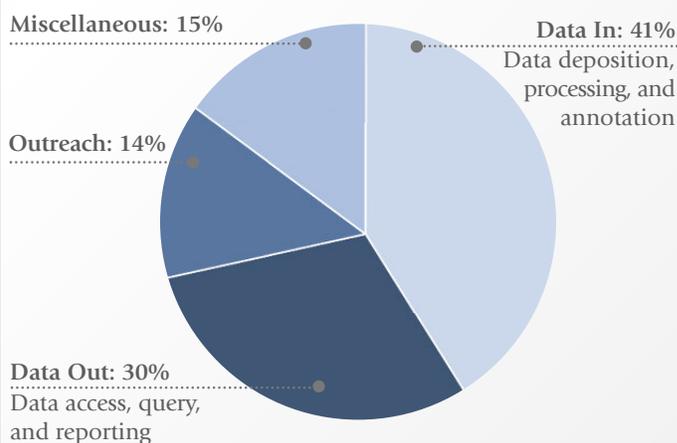
**Q:** *How can I find macromolecular complexes composed of multiple protein components?*

**A:** Advanced Search can find heteromultimers by specifying the number of entities in an entry, and can be used to limit or include nucleic acids.

**Q:** *Which structures have zinc finger motifs?*

**A:** Use the Sequence Features>Sequence Motif feature of Advanced Search to find a variety of sequence patterns, including zinc fingers.

## Help Desk Queries



wrote in with questions about downloading coordinate files, coordinate file format, citation information, or to obtain site-use or subject matter background assistance. Almost half of the query-related emails will be related to website enhancements.

The address [deposit@deposit.rcsb.org](mailto:deposit@deposit.rcsb.org) is also maintained for general data processing questions and annotator-depositor correspondence.



## Educational Programs and Activities

While most RCSB PDB programs are aimed at scientific researchers, our efforts also promote scientific literacy and structural biology to a broad audience.<sup>50,51</sup>

To interact with the broader community, the RCSB PDB participates in large-scale events. At the San Diego Festival of Science and Engineering, UCSD's Triton Day, and Rutgers Day, staff built virus and DNA models and discussed protein structures with thousands of event attendees. Many of the hands-on activities and materials used at these events are available from PDB-101.

On campus at Rutgers and UCSD, RCSB PDB leaders teach graduate and undergraduate students how to understand and visualize PDB data in the context of biology. Recent internships have focused on developing tools for the RCSB PDB and examining the representation of carbohydrates and lipids in the archive. Specialized undergraduate courses exploring molecular views of human anatomy are offered at Rutgers, with collaborations on similar courses at Georgetown University and Wellesley College.

Other programs focus on getting students and teachers at the middle and high school level interested and involved in the fields of structural biology and bioinformatics. A recent workshop introduced protein visualization to high school students.



▲ Programmer Chunxiao Bi at the San Diego Science & Engineering Festival's Expo Day.



▶ High school teachers built DNA models at the New Jersey Science Convention.

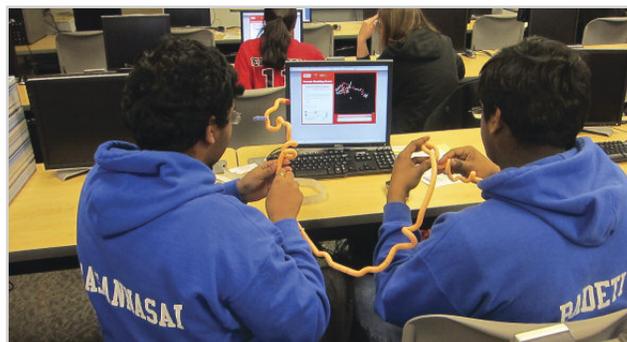
## Science Olympiad

In 2012, the RCSB PDB sponsored the protein modeling event at the Science Olympiad competitions held in New Jersey and, for the first time, in San Diego, CA. In this challenge, high school teams demonstrated their understanding of proteins involved in the regulation of apoptosis through their mastery of hand-built models and onsite exams. Information and resources were online at [education.pdb.org](http://education.pdb.org) and [twitter.com/buildmodels](https://twitter.com/buildmodels), and at workshops organized by the RCSB PDB at UCSD. The event is organized nationally by the MSOE Center for BioMolecular Modeling ([cbm.msoe.edu](http://cbm.msoe.edu)).

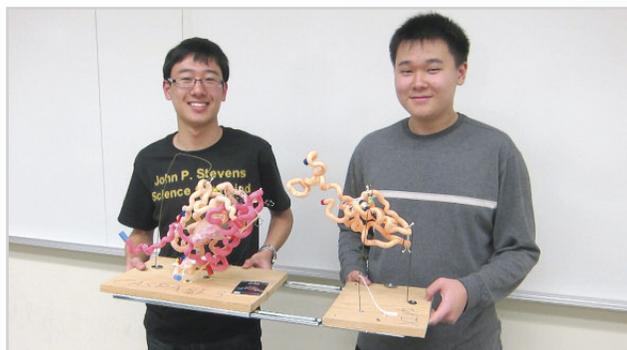
The protein modeling event will be on hiatus from the Olympiad for two years as other events are incorporated. The RCSB PDB's twitter feed [@buildmodels](https://twitter.com/buildmodels) will continue to post education and PDB-related news and links.



Annotators scored the models and provided feedback.



Teams build models using Toobers and a 3D viewer.



NJ state champions from JP Stevens High School.



## High-throughput Structural Studies and the PSI Structural Biology Knowledgebase

Worldwide Structural Genomics (SG) efforts have solved more than 11,500 structures to date, representing 15% of the PDB archive. RCSB PDB works closely with these centers in data deposition and annotation.

The Protein Structure Initiative (PSI) Structural Biology Knowledgebase ([sbkb.org](http://sbkb.org)) was created to highlight the impact of SG research to the community.<sup>53,54</sup> It is a researcher's resource, integrating structural data with other publicly available biological information to give a comprehensive view of a protein sequence. The SBKB can be searched by sequence, plain text, PDB ID, or UniProt accession code, and will return a list of all related PDB entries, selected structural targets with histories and associated experimental protocols, annotations from more than 100 biological resources, pre-calculated homology models, available DNA clones and materials from the PSI, as well as related technologies and publications. An SBKB search is integrated within every entry on the RCSB PDB website.

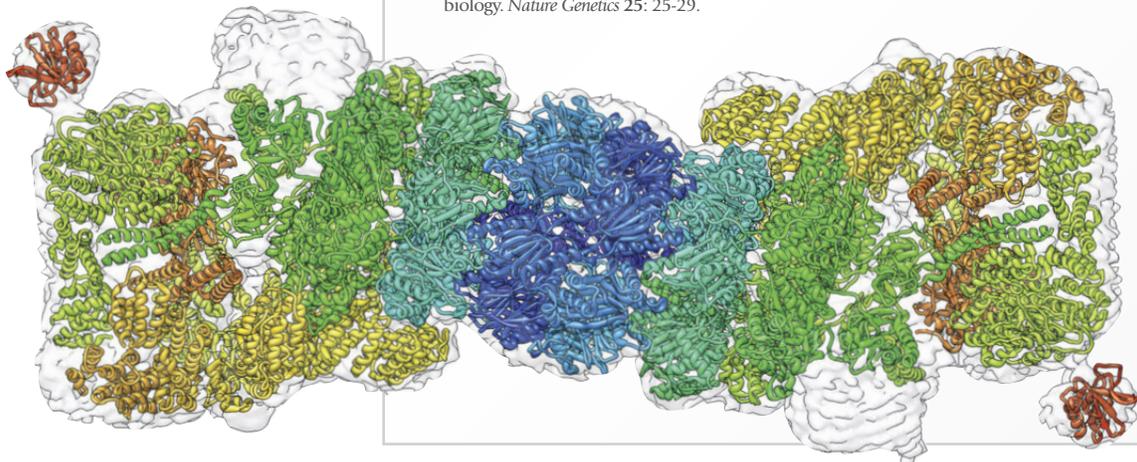
Tools and features developed by the PSI and SBKB are organized into a new "Hub" interface accessible from the SBKB homepage. They collect information related to Structural Targets, Methods, Membrane Proteins, Structure/Sequence/Function, and Homology modeling into a single place for interested users. Monthly editorial updates provided by Nature Publishing Group help to further broaden the view of structural biology.

## EMDataBank: One Stop Shop for 3DEM Deposition and Retrieval

The PDB archives large biological assemblies determined by 3DEM, a maturing methodology in structural biology that bridges the gap between cell biology and the experimental techniques of X-ray crystallography and NMR. 3DEM experiments produce 3D density maps, currently archived in the EM Data Bank, and often yield fitted coordinate models, which are archived in the PDB. [EMDataBank.org](http://EMDataBank.org) is a deposition and retrieval network for 3DEM map, model, and associated metadata.<sup>52</sup>

The EMDB map archive was recently merged with the PDB archive to enable access to EM maps and models from a single archive. This work has been carried out in collaboration with the RCSB PDB at Rutgers, PDBe, and the National Center for Macromolecular Imaging at Baylor College of Medicine.

26S proteasome structures<sup>59-62</sup> determined using 3DEM have been recently added to the PDB archive. The 26S proteasome is a key eukaryotic macromolecular machine that carries out the controlled degradation of a wide range of proteins tagged by ubiquitin. These new structures provide unprecedented detail for this complex assembly. The central 20S core particle that performs the proteolysis is capped at either end by two 19S regulatory particles that recognize the ubiquitin tags. Shown: PDB ID 4b4t and EMD-2165.<sup>59</sup> Image created using Chimera.<sup>55</sup>



## References

- H.M. Berman, J.D. Westbrook, Z. Feng, *et al.* (2000) The Protein Data Bank. *Nucleic Acids Res.* 28: 235-242.
- J.C. Kendrew, G. Bodo, H.M. Dintzis, *et al.* (1958) A three-dimensional model of the myoglobin molecule obtained by x-ray analysis. *Nature* 181: 662-666.
- J.C. Kendrew, R.E. Dickerson, B.E. Strandberg, *et al.* (1960) Structure of myoglobin: A three-dimensional Fourier synthesis at 2 Å resolution. *Nature* 185: 422-427.
- M.F. Perutz, M.G. Rossmann, A.F. Cullis, *et al.* (1960) Structure of haemoglobin: a three-dimensional Fourier synthesis at 5.5 Å resolution, obtained by X-ray analysis. *Nature* 185: 416-422.
- W. Bolton, M. F. Perutz. (1970) Three dimensional fourier synthesis of horse deoxyhaemoglobin at 2.8 Angstrom units resolution. *Nature* 228: 551-552.
- C.C.F. Blake, D.F. Koenig, G.A. Mair, *et al.* (1965) Structure of hen egg-white lysozyme. A three dimensional Fourier synthesis at 2 Å resolution. *Nature* 206: 757-761.
- C.C.F. Blake, L.N. Johnson, G.A. Mair, *et al.* (1967) Crystallographic studies of the activity of hen egg-white lysozyme. *Proc. R. Soc. London Ser. B* 167: 378-388.
- G. Kartha, J. Bello, D. Harker. (1967) Tertiary structure of ribonuclease. *Nature* 213: 862-865.
- H.W. Wyckoff, K.D. Hardman, N.M. Allewell, *et al.* (1967) The structure of ribonuclease-S at 6 Å resolution. *J. Biol. Chem.* 242: 3749-3753.
- F.C. Bernstein, T.F. Koetzle, G.J.B. Williams, *et al.* (1977) Protein Data Bank: a computer-based archival file for macromolecular structures. *J. Mol. Biol.* 112: 535-542.
- H.M. Berman, K. Henrick, H. Nakamura. (2003) Announcing the worldwide Protein Data Bank. *Nat Struct Biol* 10: 980.
- J.D. Westbrook, P.M.D. Fitzgerald. (2009). Chapter 10 The PDB format, mmCIF formats, and other data formats. In *Structural Bioinformatics, Second Edition* (P. E. Bourne & J. Gu, eds.), pp. 271-291. John Wiley & Sons, Inc., Hoboken, NJ.
- J.D. Westbrook, N. Ito, H. Nakamura, *et al.* (2005) PDBML: The representation of archival macromolecular structure data in XML. *Bioinformatics* 21: 988-992.
- H. Yang, V. Guranovic, S. Dutta, *et al.* (2004) Automated and accurate deposition of structures solved by X-ray diffraction to the Protein Data Bank. *Acta Cryst.* D60: 1833-1839.
- Z. Feng, L. Chen, H. Maddala, *et al.* (2004) Ligand Depot: A data warehouse for ligands bound to macromolecules. *Bioinformatics* 20: 2153-2155.
- K. Henrick, Z. Feng, W.F. Bluhm, *et al.* (2008) Remediation of the Protein Data Bank Archive. *Nucleic Acids Res* 36: D426-D433.
- H. M. Berman, K. Henrick, H. Nakamura, J. L. Markley. (2007) The Worldwide Protein Data Bank (wwPDB): Ensuring a single, uniform archive of PDB data. *Nucleic Acids Res* 35: D301-303.
- C.L. Lawson, S. Dutta, J.D. Westbrook, *et al.* (2008) Representation of viruses in the remediated PDB archive. *Acta Cryst.* D64: 874-882.
- P.M.D. Fitzgerald, J.D. Westbrook, P.E. Bourne, *et al.* (2005). 4.5 Macromolecular dictionary (mmCIF). In *International Tables for Crystallography* (S.R. Hall & B. McMahon, eds.), Vol. C. Definition and exchange of crystallographic data, pp. 295-443. Springer, Dordrecht, The Netherlands.
- I.W. Davis, L.W. Murray, J.S. Richardson, D.C. Richardson. (2004) MOLPROBITY: structure validation and all-atom contact analysis for nucleic acids and their complexes. *Nucleic Acids Res* 32: W615-619.
- R.A. Laskowski, M.W. McArthur, D.S. Moss, J.M. Thornton. (1993) PROCHECK: a program to check the stereochemical quality of protein structures. *J. Appl. Cryst.* 26: 283-291.
- A.A. Vaguine, J. Richelle, S.J. Wodak. (1999) SFCHECK: a unified set of procedures for evaluating the quality of macromolecular structure-factor data and their agreement with the atomic model. *Acta Cryst.* D55: 191-205.
- Z. Feng, J. Westbrook, H. M. Berman. Rutgers University, New Brunswick, NJ. (1998). NUCheck.NDB-407.
- R.J. Read, P.D. Adams, W.B. Arendall III, *et al.* (2011) A new generation of crystallographic validation tools for the Protein Data Bank. *Structure* 19: 1395-1412.
- R. Henderson, A. Sali, M.L. Baker, *et al.* (2012) Outcome of the first electron microscopy validation task force meeting. *Structure* 20: 205-214.
- S. Gore, S. Velankar, G. J. Kleywegt. (2012) Implementing an X-ray validation pipeline for the Protein Data Bank. *Acta Cryst* D68: 478-483.
- P.W. Rose, B. Beran, C. Bi, *et al.* (2011) The RCSB Protein Data Bank: redesigned web site and web services. *Nucleic Acids Res* 39: D392-D401
- D. Weininger. (1988) SMILES 1. Introduction and encoding rules. *J. Chem. Inf. Comput. Sci.* 28: 31.
- S.F. Altschul, W. Gish, W. Miller, *et al.* (1990) Basic local alignment search tool. *J. Mol. Biol.* 215: 403-410.
- M. Punta, P.C. Coggill, R.Y. Eberhardt, *et al.* (2012) The Pfam protein families database. *Nucleic Acids Res* 40: D290-301.
- Recommendations of the Nomenclature Committee of the International Union of Biochemistry and Molecular Biology on the Nomenclature and Classification of Enzyme-Catalysed Reactions. [www.chem.qmw.ac.uk/iubmb/enzyme/](http://www.chem.qmw.ac.uk/iubmb/enzyme/)
- A.G. Murzin, S.E. Brenner, T. Hubbard, C. Chothia. (1995) SCOP: a structural classification of proteins database for the investigation of sequences and structures. *J. Mol. Biol.* 247: 536-540.
- C.A. Orengo, A.D. Michie, S. Jones, *et al.* (1997) CATH—a hierarchical classification of protein domain structures. *Structure* 5: 1093-1108.
- M.H. Saier Jr., M.R. Yen, K. Noto, *et al.* (2009) The Transporter Classification Database: recent advances. *Nucleic Acids Res* 37: D274-278.
- The Gene Ontology Consortium. (2000) Gene Ontology: tool for the unification of biology. *Nature Genetics* 25: 25-29.

36. E.W. Sayers, T. Barrett, D.A. Benson, *et al.* (2012) Database resources of the National Center for Biotechnology Information. *Nucleic Acids Res* 40: D13-25.
37. Jmol: an open-source Java viewer for chemical structures in 3D. [www.jmol.org](http://www.jmol.org)
38. J.L. Moreland, A. Gramada, O.V. Buzko, *et al.* (2005) The Molecular Biology Toolkit (MBT): a modular platform for developing molecular visualization applications. *BMC Bioinformatics* 6: 21.
39. A. Prlić, M.A. Martínez, D. Dimitropoulos, *et al.* (2010) Integration of open access literature into the RCSB Protein Data Bank using BioLit. *BMC Bioinformatics* 11: 220.
40. C. Knox, V. Law, T. Jewison, *et al.* (2011) DrugBank 3.0: a comprehensive resource for 'omics' research on drugs. *Nucleic Acids Res.* 39: D1035-1041.
41. R. Wang, X. Fang, Y. Lu, *et al.* (2005) The PDBBind database: methodologies and updates. *J Med Chem* 48: 4111-4119.
42. R.M. Kelly, H. Leemhuis, H.J. Rozeboom, *et al.* (2008) Elimination of competing hydrolysis and coupling side reactions of a cyclodextrin glucanotransferase by directed evolution. *Biochem J* 413: 517-525.
43. H.M. Berman, K. Henrick, G. Kleywegt, *et al.* (2012). The Worldwide Protein Data Bank. In *International Tables for X-Ray Crystallography* (E. Arnold, D. M. Himmel & M. G. Rossmann, eds.), Vol. Volume F: Crystallography of biological macromolecules, pp. 827-832 Springer, Dordrecht, The Netherlands.
44. H.M. Berman. (2012). The Protein Data Bank. In *Leadership in Science and Technology: A Reference Handbook* (W. S. Bainbridge, ed.), Vol. 2, pp. 661-667. SAGE Publications, Inc. Thousand Oaks, CA.
45. T.L. Blundell, S.C. Harrison, R.M. Stroud, *et al.* (2011) Celebrating structural biology. *Nature Structural & Molecular Biology* 18: 1304-1316.
46. H.M. Berman, G.J. Kleywegt, H. Nakamura, J.L. Markley. (2012) The Protein Data Bank at 40: Reflecting on the Past to Prepare for the Future. *Structure* 20: 391-396.
47. S. Bliven, A. Prlić. (2012) Circular permutation in proteins. *PLoS Comput Biol.* 8: e1002445.
48. A. Kryshatovych, J. Moul, S.G. Bartual, *et al.* (2011) Target highlights in CASP9: Experimental target structures for the critical assessment of techniques for protein structure prediction. *Proteins* 10: 6-20.
49. D. Goodsell. (2011) Atomic evidence: the foundations of structural molecular biology. *Science Progress* 94: 414-430.
50. C. Zaredecki. (2008) Interesting structures: Education and outreach at the RCSB Protein Data Bank. *PLoS Biol* 6: e117.
51. S. Dutta, C. Zaredecki, D. Goodsell, H. M. Berman. (2010) Promoting a structural view of biology for varied audiences: an overview of RCSB PDB resources and experiences. *J Appl Cryst* 43: 1224-1229.
52. C.L. Lawson, M.L. Baker, C. Best, *et al.* (2011) EMDataBank.org: unified data resource for CryoEM. *Nucleic Acids Res* 39: D456-D464.
53. H.M. Berman, J.D. Westbrook, M.J. Gabanyi, *et al.* (2009) The protein structure initiative structural genomics knowledgebase. *Nucleic Acids Res* 37: D365-368.
54. M.J. Gabanyi, P.D. Adams, K. Arnold, *et al.* (2011) The Structural Biology Knowledgebase: a portal to protein structures, sequences, functions, and methods. *J Struct Funct Genomics* 12: 45-54.
55. E.F. Pettersen, T.D. Goddard, C.C. Huang, *et al.* (2004) UCSF Chimera—a visualization system for exploratory research and analysis. *J Comput Chem* 25: 1605-1612.
56. S.W. Kaldor, V.J. Kalish, J.F. Davies II, *et al.* (1997) Viracept (nelfinavir mesylate, AG1343): a potent, orally bioavailable inhibitor of HIV-1 protease. *J. Med. Chem.* 40: 3979-3985.
57. N. Nandhagopal, A. A. Simpson, J. R. Gurmon, *et al.* (2002) The structure and evolution of the major capsid protein of a large, lipid-containing DNA virus. *Proc Natl Acad Sci USA* 99: 14758-14763.
58. J.B. Bonanno, M. Dickey, K.T. Bain, *et al.* (2008) Crystal structure of an alpha-amylase from *Lactobacillus plantarum*. doi: 10.2210/pdb2213dhu/pdb.
59. F. Beck, P. Unverdorben, S. Bohn, *et al.* (2012) Near-atomic resolution structural model of the yeast 26S proteasome. *Proc Natl Acad Sci USA* 109: 14870-14875.
60. G.C. Lander, E. Estrin, M. E. Matyskiela, *et al.* (2012) Complete subunit architecture of the proteasome regulatory particle. *Nature* 482: 186-191.
61. K. Lasker, F. Forster, S. Bohn, *et al.* (2012) Molecular architecture of the 26S proteasome holocomplex determined by an integrative approach. *Proc Natl Acad Sci USA* 109: 1380-1387.
62. P. C. da Fonseca, J. He, E.P. Morris. (2012) Molecular model of the human 26S proteasome. *Mol Cell* 46: 54-66.
63. D. Voet, J. G. Voet, C. W. Pratt. (2002). *Fundamentals of Biochemistry*, John Wiley & Sons, Inc., New York, NY.
64. S. Remington, G. Wiegand, R. Huber. (1982) Crystallographic refinement and atomic models of two different forms of citrate synthase at 2.7 and 1.7 Å resolution. *J Mol Biol* 158: 111-152.
65. H. Lauble, M. C. Kennedy, H. Beinert, C. D. Stout. (1992) Crystal structures of aconitase with isocitrate and nitroisocitrate bound. *Biochemistry* 31: 2735-2748.
66. A.B. Taylor, G. Hu, P.J. Hart, L. McAllister-Henn. (2008) Allosteric motions in structures of yeast NAD<sup>+</sup>-specific isocitrate dehydrogenase. *J Biol Chem* 283: 10872-10880.
67. J.E. Knapp, D.T. Mitchell, M.A. Yazdi, *et al.* (1998) Crystal structure of the truncated cubic core component of the *Escherichia coli* 2-oxoglutarate dehydrogenase multienzyme complex. *J Mol Biol* 280: 655-668.
68. M.A. Robien, G.M. Clore, J.G. Omichinski, *et al.* (1992) Three-dimensional solution structure of the E3-binding domain of the dihydrolipoamide succinyltransferase core from the 2-oxoglutarate dehydrogenase multienzyme complex of *Escherichia coli*. *Biochemistry* 31: 3463-3471.
69. P.M. Ricaud, M.J. Howard, E.L. Roberts, *et al.* (1996) Three-dimensional structure of the lipoyl domain from the dihydrolipoamide succinyltransferase component of the 2-oxoglutarate dehydrogenase multienzyme complex of *Escherichia coli*. *J Mol Biol* 264: 179-190.
70. T. Nakai, N. Kamiya. (2008) Crystal structure of lipoamide dehydrogenase from *thermus thermophilus* HB8 with psbd. doi:10.2210/pdb2eq7/pdb
71. R.A. Frank, A.J. Price, F.D. Northrop, *et al.* (2007) Crystal structure of the E1 component of the *Escherichia coli* 2-oxoglutarate dehydrogenase multienzyme complex. *J Mol Biol* 368: 639-651.
72. M.E. Fraser, K. Hayakawa, M.S. Hume, *et al.* (2006) Interactions of GTP with the ATP-grasp domain of GTP-specific succinyl-CoA synthetase. *J Biol Chem* 281: 11058-11065.
73. V. Yankovskaya, R. Horsefield, S. Tomroth, *et al.* (2003) Architecture of succinate dehydrogenase and reactive oxygen species generation. *Science* 299: 700-704.
74. T. Weaver, L. Banaszak. (1996) Crystallographic studies of the catalytic and a second site in fumarate C from *Escherichia coli*. *Biochemistry* 35: 13955-13965
75. W.B. Gleason, Z. Fu, J. Birktoft, L. Banaszak. (1994) Refined crystal structure of mitochondrial malate dehydrogenase from porcine heart and the consensus structure for dicarboxylic acid oxidoreductases. *Biochemistry* 33: 2078-2088.
76. A. Krohn, S. Redshaw, J. C. Ritchie, *et al.* (1991) Novel binding mode of highly potent HIV-proteinase inhibitors incorporating the (R)-hydroxyethylamine isostere. *J. Med. Chem.* 34: 3340-3342.
77. P.J. Loll, D. Picot, R.M. Garavito (1995) The structural basis of aspirin activity inferred from the crystal structure of inactivated prostaglandin H2 synthase. *Nat. Struct. Biol.* 2: 637-643.
78. Y.T. Lai, D. Cascio, T.O. Yeates (2012) Structure of a 16-nm cage designed by using protein oligomers. *Science* 336: 1129-1129.
79. B. Houck-Loomis, M.A. Durney, C. Salguero, *et al.* (2011) An equilibrium-dependent retroviral mRNA switch regulates translational recoding. *Nature* 480: 561-564
80. X. Yu, C. Goforth, C. Meyer, *et al.* (2012) Filaments from *Ignicoccus hospitalis* show diversity of packing in proteins containing N-terminal type IV pilin helices. *J.Mol.Biol.* 422: 274-281.
81. A. Kovalevsky, B. L. Hanson, S. A. Mason, *et al.* (2012) Inhibition of D-xylose isomerase by polyols: atomic details by joint X-ray/neutron crystallography. *Acta Cryst.* D68: 1201-1206.
82. P.R. Evans, G. W. Farrants, P. J. Hudson. (1981) Phosphofructokinase: structure and control. *Philos. Trans. R. Soc. Lond., B, Biol. Sci.* 293: 53-62.

## wwPDB Members

RCSB PDB: [www.rcsb.org](http://www.rcsb.org)

Protein Data Bank Europe: [pdbe.org](http://pdbe.org)

Protein Data Bank Japan: [pdbj.org](http://pdbj.org)

BioMagResBank: [www.bmrwisc.edu](http://www.bmrwisc.edu)

## RCSB PDB Partners

Rutgers, The State University of New Jersey

Center for Integrative Proteomics Research

174 Frelinghuysen Road

Piscataway, NJ 08854-8087

San Diego Supercomputer Center and the Skaggs

School of Pharmacy and Pharmaceutical Sciences

University of California, San Diego

9500 Gilman Drive

La Jolla, CA 92093-0743

## RCSB PDB Management

Dr. Helen M. Berman

Director

Board of Governors Professor of

Chemistry & Chemical Biology

Rutgers, The State University

of New Jersey

[berman@rcsb.rutgers.edu](mailto:berman@rcsb.rutgers.edu)



Dr. Martha Quesada

Deputy Director

Rutgers, The State University

of New Jersey

[mquesada@rcsb.rutgers.edu](mailto:mquesada@rcsb.rutgers.edu)



Dr. Philip E. Bourne

Associate Director

San Diego Supercomputer

Center and the Skaggs School

of Pharmacy and Pharmaceutical

Sciences

University of California, San Diego

[bourne@sdsc.edu](mailto:bourne@sdsc.edu)



[www.rcsb.org](http://www.rcsb.org) • [info@rcsb.org](mailto:info@rcsb.org)