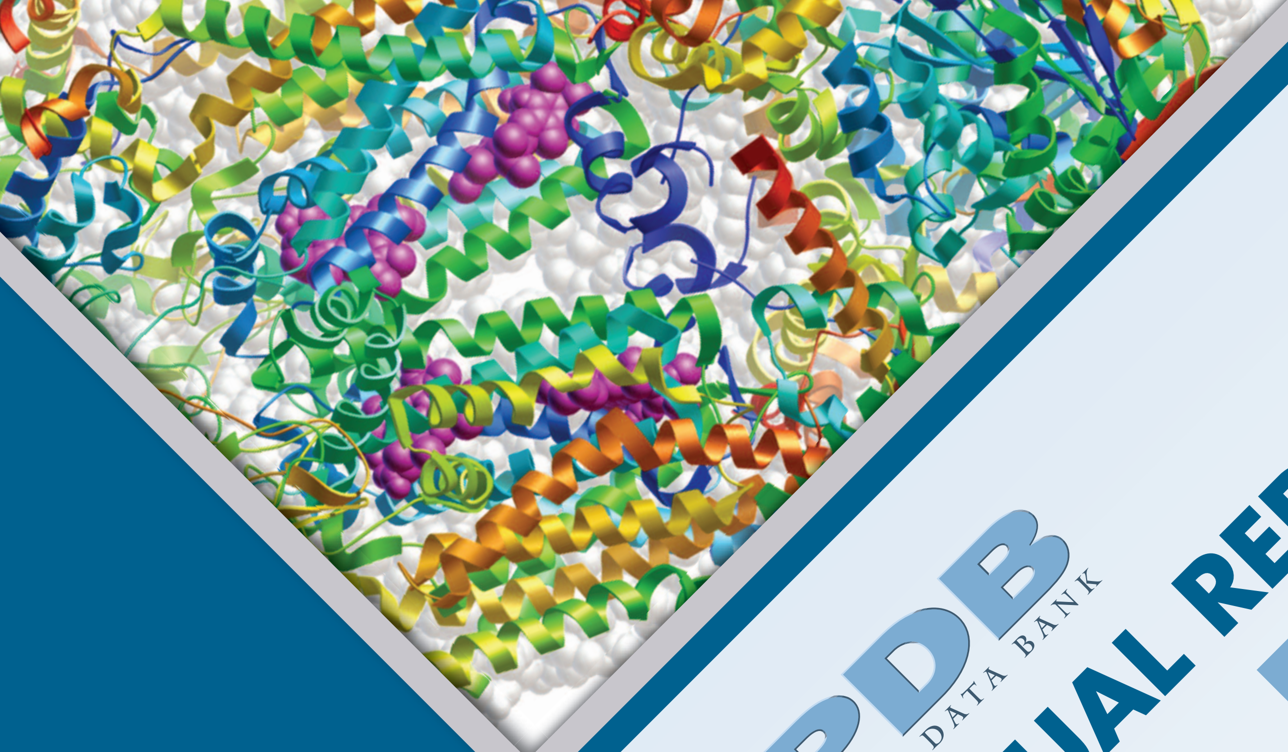
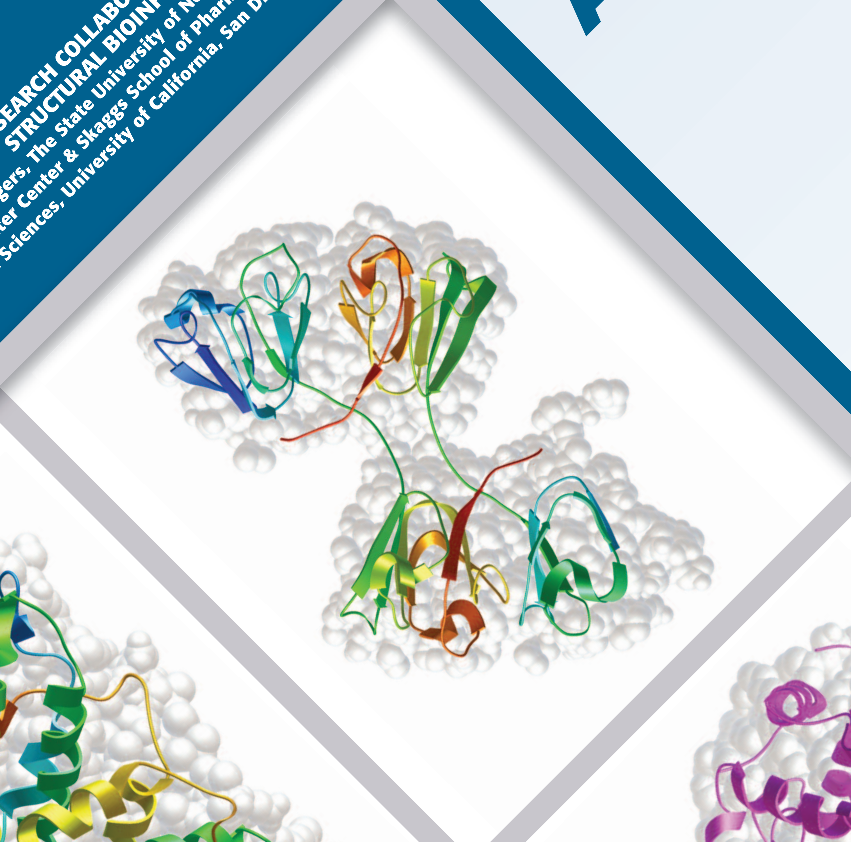
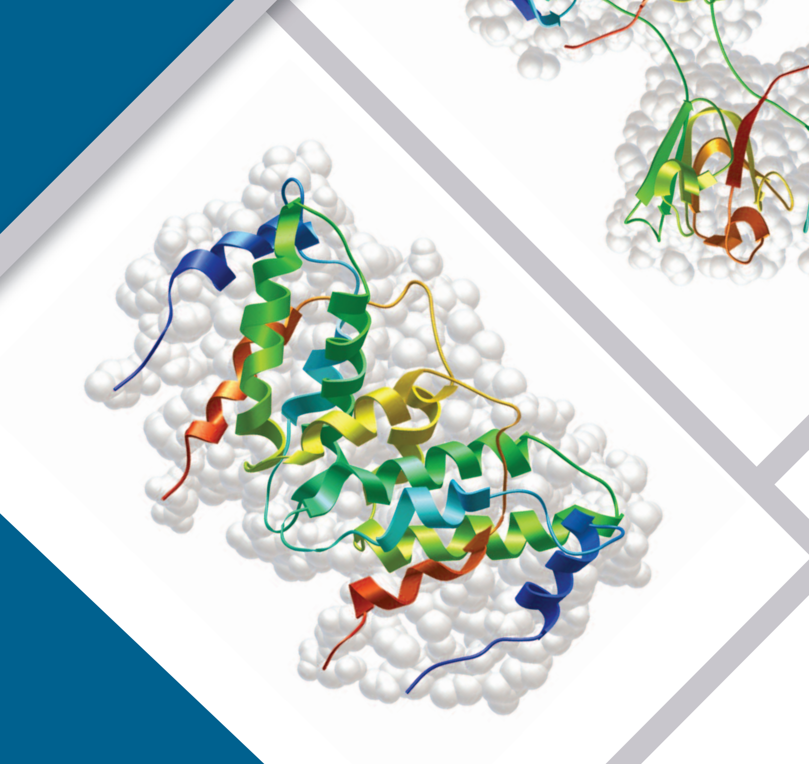


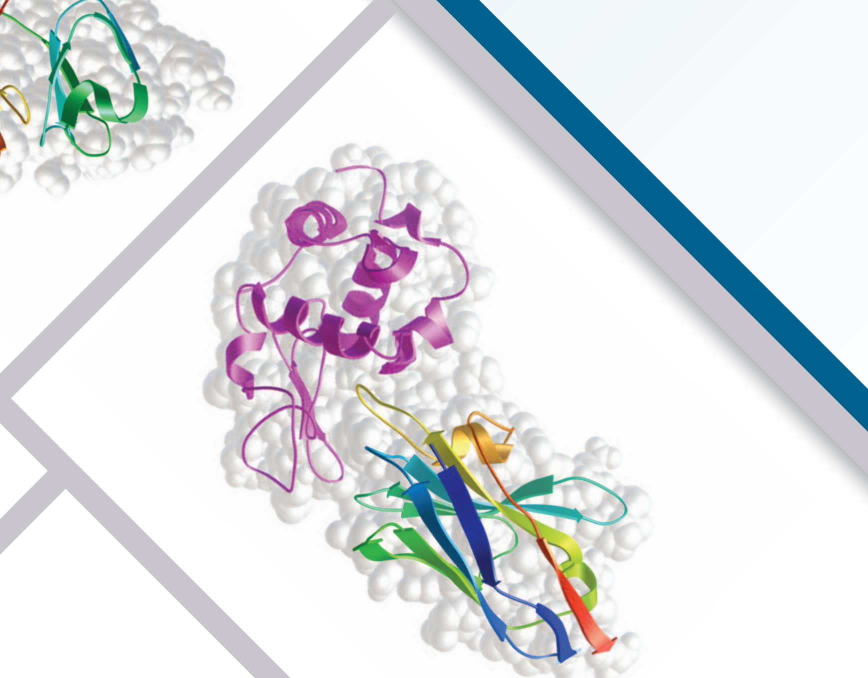
RCSB PDB PROTEIN DATA BANK

ANNUAL REPORT 2011

RESEARCH COLLABORATORY FOR STRUCTURAL BIOINFORMATICS
Rutgers, The State University of New Jersey
San Diego Supercomputer Center & Skaggs School of Pharmacy & Pharmaceutical Sciences, University of California, San Diego







RCSB PDB Annual Report

July 1, 2010 – June 30, 2011

Table of Contents

About the Cover

The RCSB PDB's *Molecule of the Month* feature is a gateway to learning about the structures in the PDB archive. Important biological molecules and how they function are illustrated through descriptive text, pictures, links to specific PDB entries, discussion questions, and 3D interactive molecular views. As the keystone of the PDB-101 educational resource, the *Molecule of the Month* supports teachers and students at all levels in their exploration of biology at a structural level.

The molecules featured on the cover and throughout this document were highlighted in columns published during the period of this report.

These images were created using Python Molecular Viewer (Michel F. Sanner (1999) Python. A Programming Language for Software Integration and Development. *J Mol Graphics Mod* **17**: 57-61. **mgltools.scripps.edu/packages/pmv**).

Front:

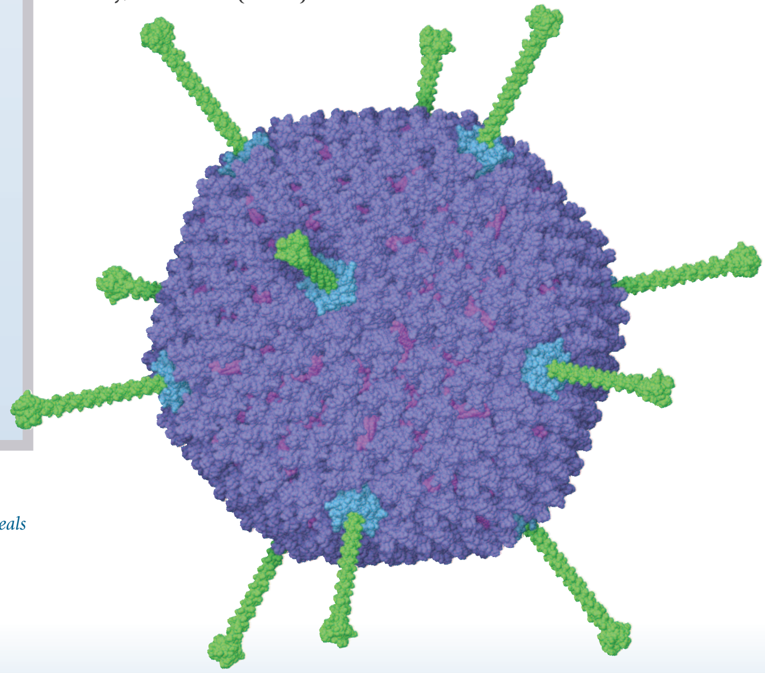
From Interferons, PDB ID: 1rfb. Structure of recombinant bovine interferon-gamma at 3.0 Å resolution. C.T. Samudzi, J.R. Rubin (1993) *Acta Cryst* **D49**: 513-521.

From Crystallins, PDB ID: 1blb. Close packing of an oligomeric eye lens beta-crystallin induces loss of symmetry and ordering of sequence extensions. V. Nalini, B. Bax, H. Driessen, D.S. Moss, P.F. Lindley, C. Slingsby (1994) *J Mol Biol* **236**: 1250-1258.

From Nanobodies, PDB ID: 1mel. Crystal structure of a camel single-domain VH antibody fragment in complex with lysozyme. A. Desmyter, T.R. Transue, M.A. Ghahroudi, M.H. Thi, F. Poortmans, R. Hamers, S. Muyldermans, L. Wyns (1996) *Nat Struct Biol* **3**: 803-811.

Back:

From Cytochrome bc1, PDB ID: 3h1j. Electron transfer by domain movement in cytochrome bc1. Z. Zhang, L.-S. Huang, V.M. Shulmeister, Y.I. Chi, K.K. Kim, L.W. Hung, A.R. Crofts, E.A. Berry, S.H. Kim (1998) *Nature* **392**: 677-684.



From Adenovirus, PDB ID: 1qiu. *A triple ß-spiral in the adenovirus fibre shaft reveals a new structural motif for a fibrous protein. M.J. Van Raaij, A. Mitraki, G. Lavigne, S. Cusack (1999)* Nature ***401****: 935-938.*

Message from the Director

With the goal of promoting a structural view of biology in research and education, the RCSB PDB has focused on developing new tools for searching, reporting, and visualizing PDB data during this most recent reporting period. At the same time, preparations were underway for a very special anniversary event.

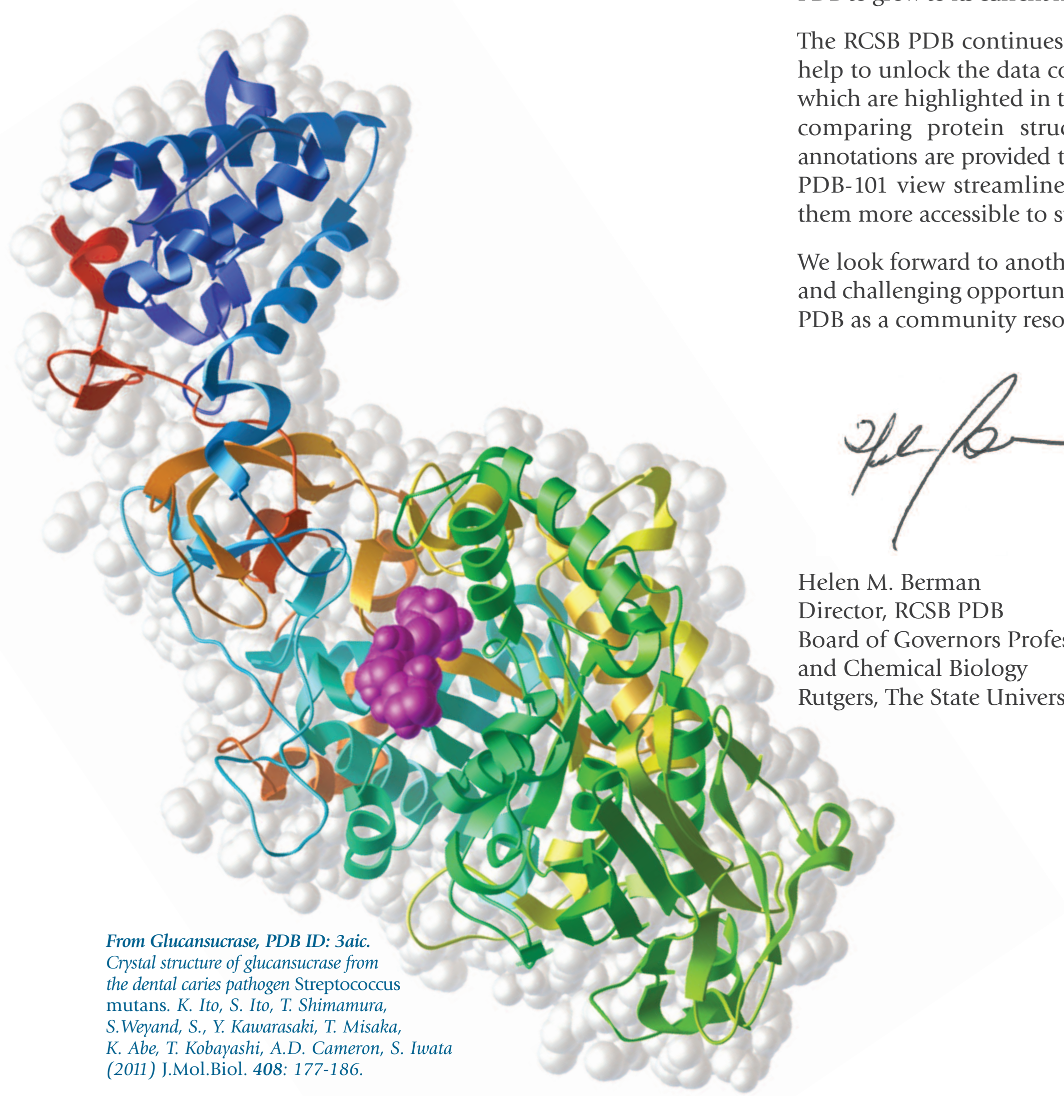
In the late 1960s, scientists, excited by the possibilities hinted at by the early structures available, realized that they could benefit from open access to the data describing them. Scientists were amenable to sharing the data in the hopes of enabling further research and progress. The PDB was established in 1971 as the first electronic, open access resource for biological sciences. As more structures were determined and added to the archive, the community helped to shape the resource by developing guidelines for deposition and tools for viewing and analyzing structure.

In Fall 2011, a symposium was organized by the Worldwide PDB (wwPDB) to commemorate the 40th anniversary of the Protein Data Bank (PDB) archive. The program highlighted the breadth and depth of the field of structural biology that has enabled the PDB to grow to its current holdings of more than 77,000 structures.

The RCSB PDB continues to develop features and services that help to unlock the data contained in the PDB archive, many of which are highlighted in this report. Resources such as tools for comparing protein structures and for exploring sequence annotations are provided to support researchers. The educational PDB-101 view streamlines the PDB archive and data, making them more accessible to students and the general public.

We look forward to another year of new and exciting structures and challenging opportunities in always working to improve the PDB as a community resource.

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



From Glucansucrase, PDB ID: 3aic.
Crystal structure of glucansucrase from the dental caries pathogen Streptococcus mutans. *K. Ito, S. Ito, T. Shimamura, S.Weyand, S., Y. Kawarasaki, T. Misaka, K. Abe, T. Kobayashi, A.D. Cameron, S. Iwata (2011)* J.Mol.Biol. ***408****: 177-186.*

The Protein Data Bank and the RCSB PDB

The RCSB PDB and the Power of Biological Macromolecules

The RCSB PDB enables vital research and study of biology at the biomolecular level.

The resource is powered by the data contained in the Protein Data Bank archive[1-3]–information about the 3D shapes of proteins, nucleic acids, and complex assemblies. These structures help researchers understand all aspects of biomedicine and agriculture, from protein synthesis to health and disease to biological energy.

As a member of the wwPDB, the RCSB PDB maintains the PDB archive of biomolecules determined from experiments in X-ray crystallography, nuclear magnetic resonance (NMR), and electron microscopy (EM). From around the world, scientists determine the atomic arrangements of these molecules, and submit data about their experiments to the PDB. This information is processed, validated, annotated, and made publicly and freely available by the wwPDB.

The RCSB PDB builds upon the data by creating tools and educational materials used by researchers, teachers, and students interested in molecular biology, structural biology, computational biology, and beyond. The online RCSB PDB provides access to curated and integrated biological macromolecular information in the context of function, biological processes, evolution, pathways, and disease states.

The annotation and dissemination of PDB data is enhanced by the RCSB PDB's targeted education and outreach activities.

Funding

The RCSB PDB is supported by funds from the:

- National Science Foundation (NSF)
- 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)

The Advisory Committee with RCSB PDB team members

The RCSB PDB Team

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

Helen M. Berman, Director of the RCSB PDB, is a Board of Governors professor of chemistry and chemical biology at Rutgers. Professor Berman 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 comprised of experts in diverse fields of computer science, biology, chemistry, and education.

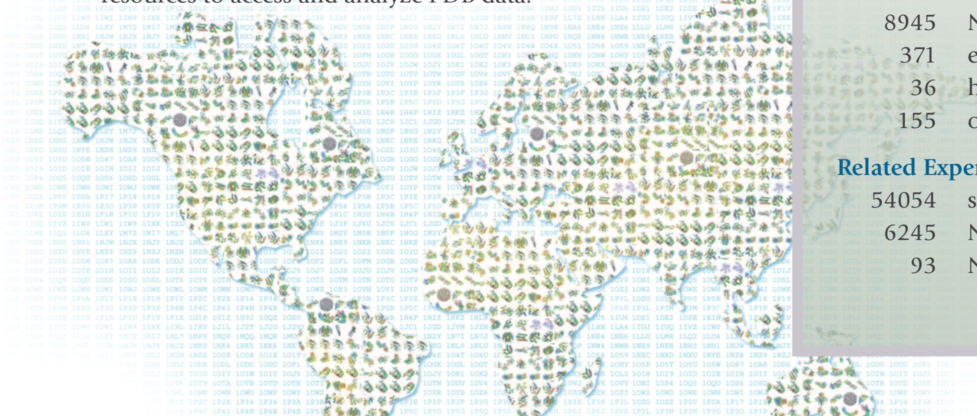
The RCSB PDB receives input from an Advisory Committee of experts in X-ray crystallography, NMR, electron microscopy, bioinformatics, and education.

The wwPDB

The wwPDB organization (wwpdb.org)[2] provides the authoritative archive of experimentally determined 3D structures of biological macromolecules. This unique scientific collaboration ensures that the PDB archive, now and in the future, is freely and publicly available to the global community.

wwPDB members (RCSB PDB, Protein Data Bank in Europe (PDBe),[4] Protein Data Bank Japan (PDBj),[5] and BioMagResBank (BMRB)[6]) host deposition, processing, 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.



PDB Archive Snapshot: July 1, 2011

Number of Entries

74140	released atomic coordinate entries

By Molecule Type

68651	proteins, peptides, and viruses
2265	nucleic acids
3185	protein/nucleic acid complexes
39	other

By Experimental Technique

64633	X-ray
8945	NMR
371	electron microscopy
36	hybrid
155	other

Related Experimental Data Files

54054	structure factors
6245	NMR restraints
93	NMR chemical shifts

Getting Data In and Out of the PDB

Structure Determination

Scientists located around the world experimentally determine the 3D structures of the proteins, nucleic acids, and molecular machines to be archived in the PDB.

Data Deposition

Data are deposited to wwPDB member sites using specially-developed software.

Approximately 170 new structures are deposited each week.

Data Processing

Annotators at each wwPDB site carefully curate and review each entry.

Important validation checks are run and reported back to the depositor.

On average, it takes two weeks to fully process, validate, and finalize an entry.

Data Archiving

Data from the wwPDB sites are combined and updated into the public PDB archive once a week. With sole write access to the FTP, the RCSB PDB is the *archive keeper*.

During this report period, 7918 entries were released into the archive.

Data Distribution & Query

Data can be accessed via FTP and online websites.

There were approximately 300 million data downloads from the wwPDB FTP sites during this report period.

WORLDWIDE PDB PROTEIN DATA BANK

RCSB PDB PROTEIN DATA BANK

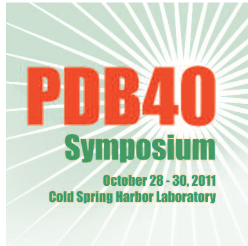
EMBL-EBI PDBe PROTEIN DATA BANK EUROPE

PDBj Protein Data Bank Japan

BMRB

40 Years of the Protein Data Bank

Photos by Constance Brukin.



Scenes from PDB40. Top row: discussions between presentations and at the poster session. Bottom row: PDB team members, past and present.

In June 1971, a symposium on Structure and Function of Proteins at the Three Dimensional Level was held at Cold Spring Harbor Laboratory (CSHL).

With its lively conversations, debates, and plans for the future, that meeting defined the beginning of the PDB as an archive for the experimentally determined 3D structures of biological macromolecules.

The wwPDB returned to CSHL October 28-20, 2011, to commemorate the 40th anniversary of the event, to celebrate the field of structural biology, and to inspire conversations and collaborations that will help guide the PDB for the next 40 years.

PDB40's distinguished speakers described structural biology's past as well as current research in related fields. More than 250 people attended the event, with close to 100 posters presented. Travel awards were made for 34 early career scientists from around the world.

PDB40 was organized by Helen M. Berman (RCSB PDB), Gerard Kleywegt (PDBe), Haruki Nakamura (PDBj), John Markley (BMRB), and Stephen K. Burley (wwPDB Advisory Committee and Eli Lilly & Company). Generous funding for the meeting and travel awards came from CSHL, the National Science Foundation, National Institutes of Health, Wellcome Trust, Japan Society for the Promotion of Science, and more than 20 industrial sponsors.

For presentations, pictures, and more, see **wwpdb.org**.

2011 — PDB40 Symposium. More than 77000 structures in the archive

2010 — Deposition of chemical shift data mandatory

2008 — Deposition of experimental data mandatory. 50000th structure is released

2007 — First release of remediated data

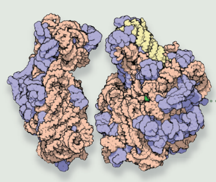
2006 — BMRB joins wwPDB

2005 — Last PDB archive distribution by postal mail (8 DVDs)

2003 — wwPDB established by RCSB PDB, PDBe, PDBj

2002 — EMDB established at MSD-EBI

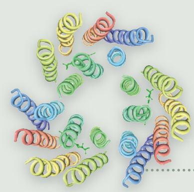
2000 — Osaka University opens a PDB data deposition center



First ribosome structures determined [37-39]

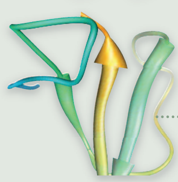
1999 — 10000th structure is released

1998 — PDB moves to RCSB PDB. MSD at the EBI becomes a deposition center for PDB data



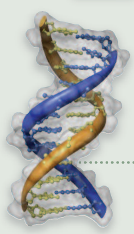
First EM entry released in PDB: bacteriorhodopsin[40]

1991

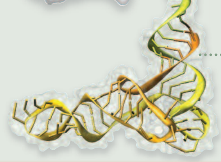


1989 — First NMR entry released in PDB: protein BDS-I[41]

IUCr policy on data deposition published

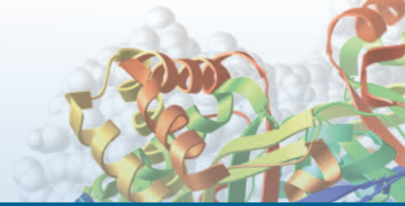


1981 — First B-DNA structure determined[42]



1973 — First tRNA structures determined[43, 44]

1971 — PDB established



Data Deposition and Processing

Overview

Using locally developed tools, the RCSB PDB carefully curates and annotates entries using an integrated system that features standard data dictionaries for macromolecular structure[7] and for small chemical components found in PDB entries. The goal is that each PDB entry accurately represents the structure and experiment.

8990 entries were deposited to the PDB archive and prepared for release by the wwPDB during the period of this report. 6872 of these entries were deposited to the RCSB PDB.

As part of data processing, the annotation team reviews protein sequence and chemistry of small molecule ligands, correspondence and 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.

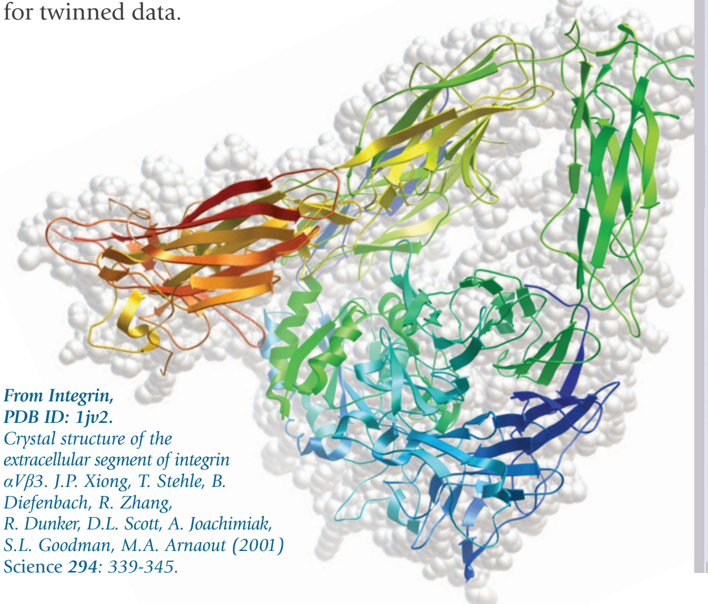
Sequences for 49% of these depositions were made publicly available prior to the release of the coordinate entry. This helps prevent duplication of structure determination efforts and promotes blind testing of structure prediction and modeling techniques.

Data files describing each entry are released in different file formats (PDB, PDBx, PDBML-XML)[8,9] along with related experimental data.

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**.

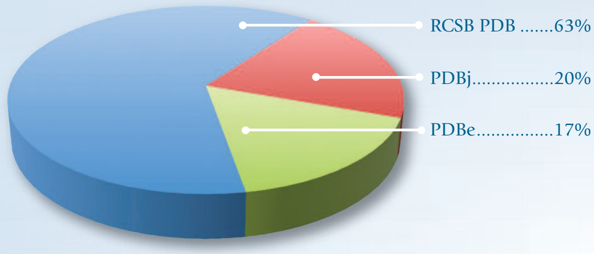
New highlights include a new validation report server, the updated pdb_extract data harvesting tool for automatically collecting key details from NMR and X-ray crystallographic applications for deposition,[10] and an updated version of SF-Tool for validating model coordinates against structure factor data, translating structure factors into different formats, and checking for twinned data.



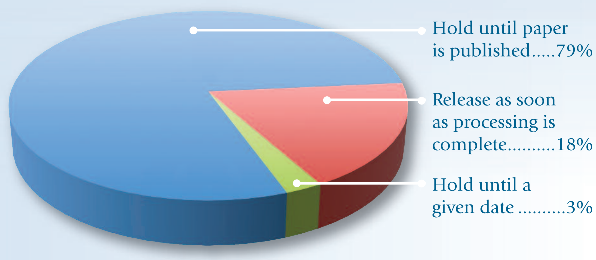
From Integrin, PDB ID: 1jv2.
Crystal structure of the extracellular segment of integrin αVβ3. J.P. Xiong, T. Stehle, B. Diefenbach, R. Zhang, R. Dunker, D.L. Scott, A. Joachimiak, S.L. Goodman, M.A. Arnaout (2001) Science ***294****: 339-345.*

Deposition Statistics
July 1, 2010 – June 30, 2011

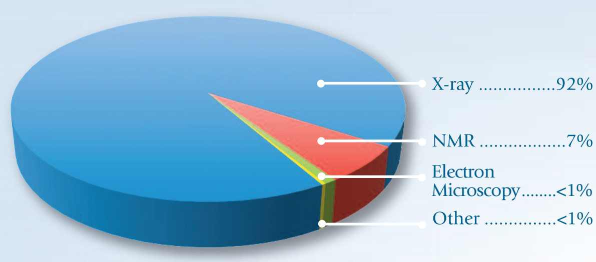
wwPDB Processing Site



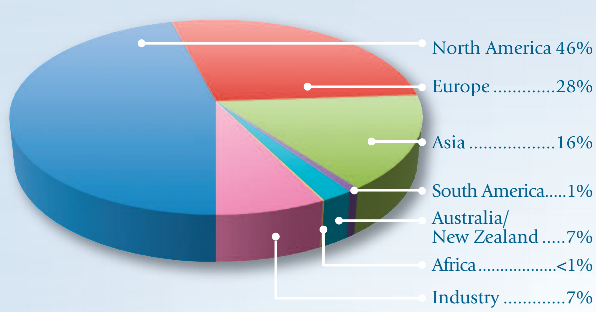
Author-defined Release Status



Experimental Source



Depositor Location



Ensuring a Uniform Archive

A major focus of the wwPDB is maintaining consistency and accuracy across the archive. As the PDB grows, developments in structure determination methods and technologies can challenge how all structures are represented. The wwPDB addresses these challenges with regular reviews of data processing procedures and coordinates remediation efforts to improve data representation.

A recent and ambitious review and remediation effort tackled complex problems, including the representation of biological assemblies, residual B factors, peptide inhibitors and antibiotics, and entries in the nonstandard crystal frame.

A Common Deposition & Annotation Tool for the Future

In order to ensure the value and effectiveness of the PDB in the future, the wwPDB partners have embarked on the development of the next generation of processing software. The new systems will maximize the efficiency and effectiveness of data handling and support for the scientific community going forward. They will also address anticipated increases in complexity and experimental variety of future submissions. Made up of experts from all partner sites, the Common Deposition and Annotation (D&A) Tool project team meets regularly and works in conversation with the wwPDB Directors and user communities.

The new system is based on a modular architecture that will enable maintenance and extensibility over its lifetime. Sequence and ligand processing modules have been completed and feature greatly enhanced visualization and integration of the process components. The ligand module is in production testing by annotators and has been shown to deliver up to a 70% increase in efficiency.

Development of the deposition interface is underway, with planned features to include automated batch data uploads, flexible manual 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.

We anticipate an initial release of the new system in the 4th quarter of 2012.

wwPDB Validation Reports and Task Forces

Validation is the process of checking submitted values against community-accepted standards. It helps to ensure that the data deposited and released in the PDB are accurate. The results of validation checking also provide a way of assessing the quality of a structure.

X-ray Validation Task Force Report:
A new generation of crystallographic validation tools for the Protein Data Bank.

R.J. Read, P.D. Adams, W.B. Arendall III, A.T. Brunger, P. Emsley, R.P. Joosten, G.J. Kleywegt, E.B. Krissinel, T. Lutteke, Z. Otwinowski, A. Perrakis, J.S. Richardson, W.H. Sheffler, J.L. Smith, I.J. Tickle, G. Vriend, P.H. Zwart (2011) *Structure* **19**: 1395-1412.
doi: 10.1016/j.str.2011.08.006

As part of the current structure annotation process, wwPDB members provide depositors with detailed validation reports that include the results of geometric and experimental data checking.[11-14] These reports, available as PDFs, provide an assessment of structure quality while keeping the coordinate data confidential. We encourage journal editors and referees to include these reports in the manuscript submission and review process. PDB validation reports are already required by the International Union of Crystallography (IUCr) journals.

To improve validation methods in the PDB, method-specific Validation Task Forces in X-ray Crystallography, NMR, EM, 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.

The first report of recommendations has recently been published by the X-ray Validation Task Force in *Structure*.[15] These recommendations will be incorporated into the wwPDB data processing procedures and tools as part of the D&A Common Tool development.

The EM Validation Task Force convened at Rutgers September 28-29, 2010.

Data Archiving and Distribution

As the archive keeper for the wwPDB, the RCSB PDB maintains the PDB archive at ftp://ftp.wwpdb.org. It has sole write access to the FTP archive, and controls updates and distribution to our global partners. Coordinated updates of the PDB archive and the wwPDB websites occur weekly.

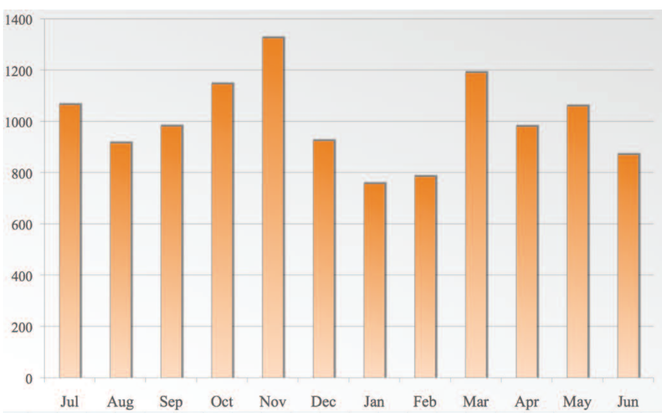
During this report period, 7918 PDB entries, 7343 structure factor files, and 511 restraint files were added to the archive. The wwPDB FTP sites combined supported approximately 300 million data downloads. Each wwPDB member also hosts a website to access PDB data.

The RCSB PDB website is a portal to these data files, with powerful searching and reporting tools. The website integrates these data with more than 30 external sources of related data (e.g., sequence annotations, structure annotations, taxonomy, functional descriptors), tools for visualization, and related editorial and educational content.

Each month, around 211,000 unique visitors access the website (up 17% from last year's 180,000) from 140 different countries. Approximately 1 Terabyte of data is transferred.

Web Services[16] are also used to remotely search the RCSB PDB database and retrieve data for any given structure. These services, including the RESTful Web Services, help software developers create tools that interact efficiently with PDB data.

Data downloads from www.rcsb.org
(Per month, in Gigabytes, July 1, 2010 - June 30, 2011)



PDB*Mobile*

PDB*Mobile* can be used to access basic structure and citation information, read the *Molecule of the Month*, and learn about the latest RCSB PDB news.

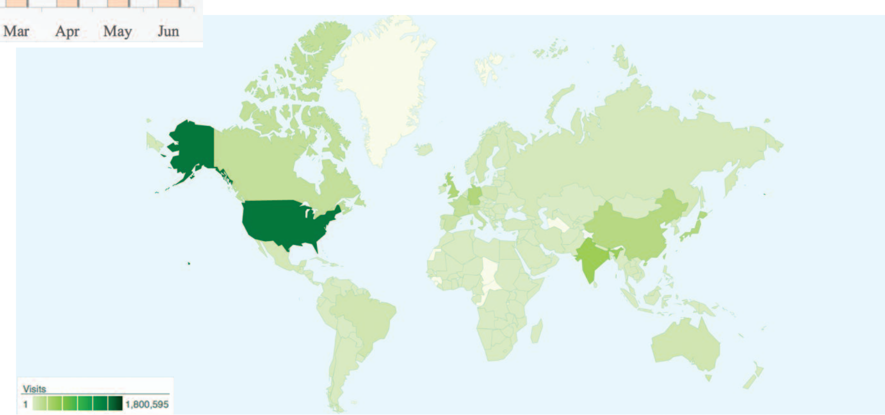
MyPDB: Personalized View of the RCSB PDB

MyPDB stores any type of search, such as keyword, sequence, ligand, and any composite query built with Advanced Search. These queries can be run at any time with the click of a button. Stored searches can also be set to run with each update. Automatic email alerts are sent when new entries matching the search are released.

Users can use MyPDB to save personal annotations and notes on the Structure Summary tab of any PDB entry, and add structures to a "favorites list." The Personal Annotations summary page provides easy access to all tagged structures and annotations.

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 these archival snapshots.



Website User Location
(July 1, 2010 - June 30, 2011)

Searching the PDB

Quick Searching: The top search bar facilitates easy, intuitive, and precise queries. After typing a few characters, an interactive pop-up box will appear with suggestions of common PDB search terms, organized in different categories.

These categories return more precise results than simple text searches. For example, entering the word *human* presents several options organized by category (such as Molecule Name, Author, Organism). Suggestions display the number of results and link to the set of matching structures. Results will be limited to a specific category. For example, results for organism *Homo sapiens* (human) will not include entries from author J. Human.

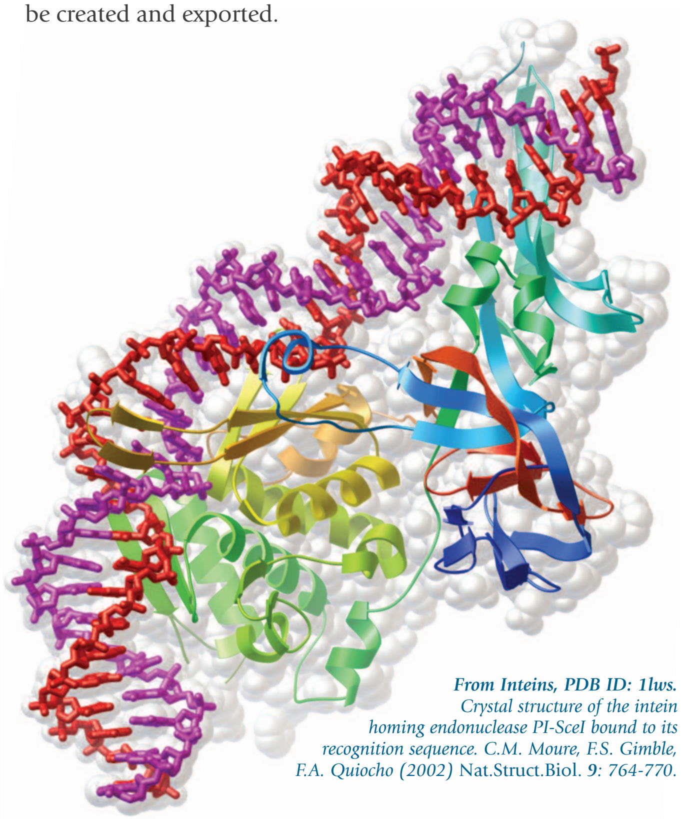
The top bar search can be be limited to quick searches on Author, Macromolecule name, Sequence, or Ligand by selecting the related icon.

The order of results of a PDB text search or a sequence search is based on the relevance of the term (for a text search) or the alignment score (for a sequence search). Search results can be further refined using a variety of options.

Drill-down Pie Charts: The entire archive, the latest update, and any search result set can be browsed by a combination of categories, including source organism, taxonomy, experimental method, resolution, release date, polymer type, Enzyme Classification (EC)[17] and SCOP[18] classification. An initial search result can be refined iteratively by selecting a subset of the results from the categories; for example a user may "drill down" to structures of human proteins, and then select another category to explore.

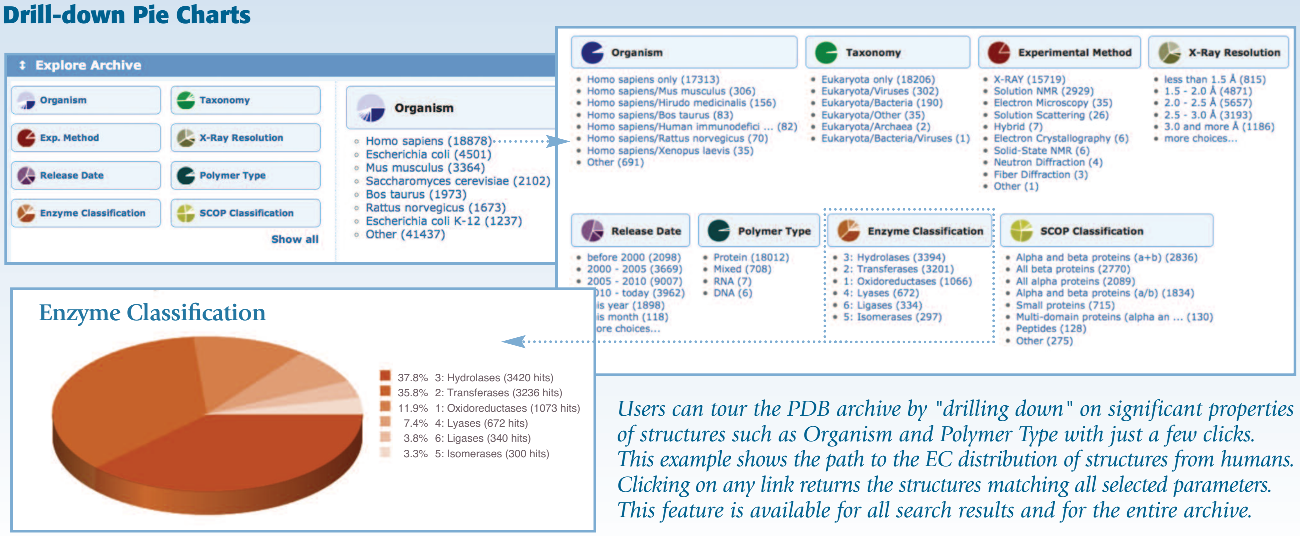
Database Browsing: Database browsers offer another way to navigate the PDB by hierarchical classifications (EC number, SCOP, CATH,[19] Transporter Classification[20]) or ontologies (Gene Ontology/GO,[21] NCBI Taxonomy[22]).

All search result sets can be explored or further refined. Individual structures can be examined at any point. Search results can be presented as a list of structures matching the query; a PubMed-like list of the primary citations for the structures; a list of ligands known to interact with the structures; a list of any RCSB PDB web pages that contain a particular keyword, including *Molecule of the Month* features; and a list of corresponding GO, SCOP, or CATH hits. Tabular reports highlighting a variety of features can be created and exported.



From Inteins, PDB ID: 1lws. *Crystal structure of the intein homing endonuclease PI-SceI bound to its recognition sequence. C.M. Moure, F.S. Gimble, F.A. Quiocho (2002)* Nat.Struct.Biol. ***9****: 764-770.*

Drill-down Pie Charts

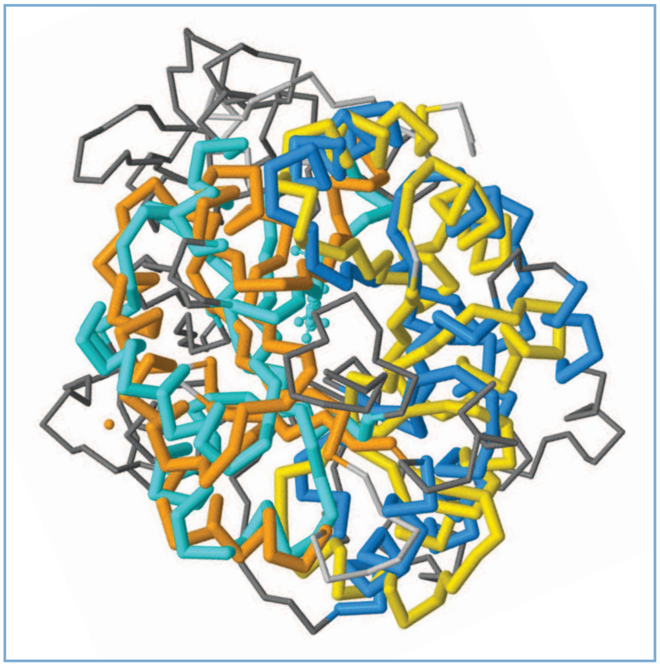


Users can tour the PDB archive by "drilling down" on significant properties of structures such as Organism and Polymer Type with just a few clicks. This example shows the path to the EC distribution of structures from humans. Clicking on any link returns the structures matching all selected parameters. This feature is available for all search results and for the entire archive.

Exploring PDB Structures

For every PDB entry, a Structure Summary page provides an overview of the structure; external domain annotations from CATH, SCOP, Pfam,[23] and GO; information and links from the Structural Biology Knowledgebase; tools to examine sequence, sequence domains, and sequence similarity; detailed information relating to the entry's published citation, biology and chemistry, experiment, and geometry; and links to related resources. Several molecular viewers, including Jmol[24] and the RCSB PDB's Protein Workshop and Ligand Explorer,[25] help users view the molecule interactively and generate images.

Comparison Tool for exploring sequence and 3D structure alignments



Proteins can have various degrees of similarity. If two proteins show high similarity in their amino acid sequence, it is generally assumed that they are closely related evolutionarily. If the sequence similarity is low, proteins can still show similar function and have an overall similar 3D structure. The detection of such remote similarities is important in studies of functional and evolutionary relationships between protein families.

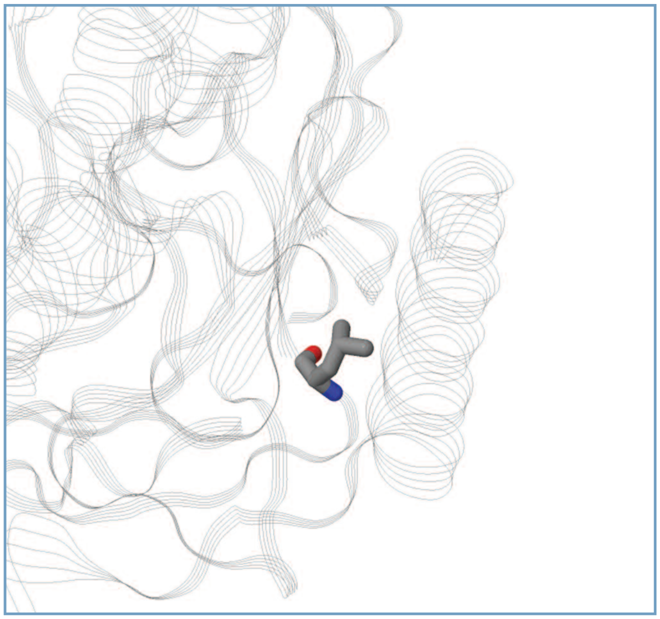
The RCSB PDB's Comparison Tool[26] can be used to quickly calculate pairwise sequence and structure alignments using a variety of algorithms.

Special features include support for both rigid-body and flexible alignments and detection of circular permutations, such as the example shown of a comparison of two proteins with a TIM barrel fold. Two alignment blocks are shown, one in turquoise and orange, the other in blue and yellow.

***PDB ID: 1bd0** (turquoise/blue/light grey). Reaction of alanine racemase with 1-aminoethylphosphonic acid forms a stable external aldimine. G.F. Stamper, A.A. Morollo, D. Ringe, C.G. Stamper (1998)* Biochemistry ***37**: 10438-10445.*

***PDB ID: 1cdg** (orange/yellow/dark grey). Nucleotide sequence and X-ray structure of cyclodextrin glycosyltransferase from* Bacillus circulans *strain 251 in a maltose-dependent crystal form. C.L. Lawson, R. van Montfort, B. Strokopytov, H.J. Rozeboom, K.H. Kalk, G.E. de Vries, D. Penninga, L. Dijkhuizen, B.W. Dijkstra (1994)* J Mol Biol ***236**: 590-600.*

Structure Summary: Sequence



A variety of annotations, including domain assignments, secondary structure, structural features, and ligand binding sites, can be mapped onto an entry's sequence diagram or 3D Jmol view.

In the example structure shown, arylsulfatase A contains a Pro to Leu mutation that is linked to the genetic disorder metachromatic leukodystrophy. This SNP (single nucleotide polymorphism) is highlighted in Jmol using the tools available from the Structure Summary's Sequence Tab.[27]

***PDB ID: 1e33**. Defective oligomerization of arylsulfatase A as a cause of its instability in lysosomes and metachromatic leukodystrophy. R. Von Bulow, B. Schmidt, T. Dierks, N. Schwabauer, K. Schilling, E. Weber, I. Uson, K. Von Figura (2002)* J Biol Chem *277: 9455-9461.*

Exploring Ligands

All residue and small molecule components found in PDB entries, including standard and modified amino acids/nucleotides, small molecule ligands, and solvent molecules, are described in the wwPDB's Chemical Component Dictionary. The information stored in the dictionary is easily searched from the RCSB PDB website.

The top bar search can quickly suggest and locate ligands based on chemical component name or 3-character dictionary ID. An advanced ligand search form can be used to search by a combination of chemical structure (SMILES/SMARTS string, MarvinSketch), molecule name (includes fuzzy searching and synonyms), formula, and weight.

These searches will return a Ligand Summary page that contains an overview of the chemical component, 2D and 3D images, links to other resources, and links to related ligands and PDB entries.

All general PDB structure searches return a Ligand Hits Tab containing the Ligand ID, image, chemical formula, molecular weight, molecular name, SMILES string, and a listing of PDB entries containing the ligand.

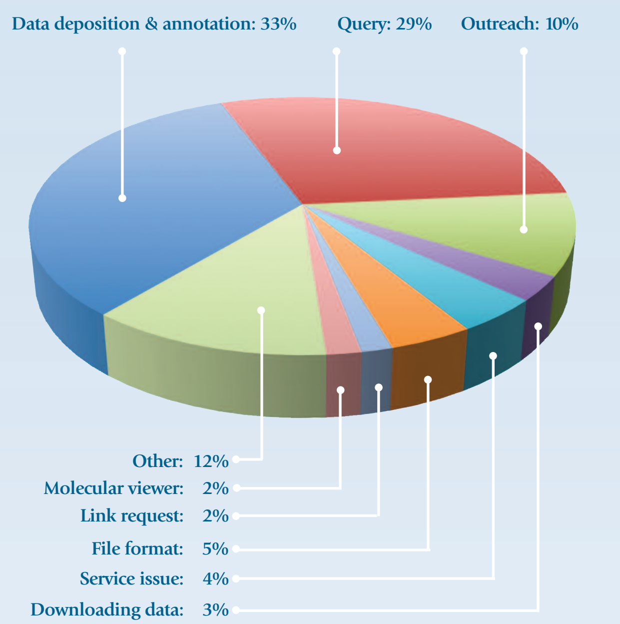
User Outreach and Support

The RCSB PDB supports a variety of audiences. Users include biologists from diverse fields, scientists from other disciplines, computational researchers, students and educators at all levels, authors and illustrators, and the general public.

The website provides an important vehicle for communication. Web pages can be customized to create views that reflect specific interests. News and announcements are updated regularly, along with detailed descriptions of new website releases.

The electronic help desk at **info@rcsb.org** provides around-the-clock support for website resources and beyond. The RCSB PDB hears daily from users from around the world. Questions range in scope from image reprint requests and entry release notifications, to requests for help with multi-parameter queries and detailed file format explanations. The help desk supports students new to studying biology to scientific specialists looking to expand their research. Many requests sent to the help desk lead to new website enhancements and features.

Types of questions sent to info@rcsb.org
July 1, 2010 – June 30, 2011



Other help desks assist with queries specific to data deposition and annotation. The list server at **pdb-l@rcsb.org** is also maintained for community announcements and discussion.

Related materials for learning about the RCSB PDB website include a suite of tools (narrated tutorial, slides, and exercises) created with OpenHelix; short online screencasts demonstrating how to use different RCSB PDB features; and contextual help and examples available where ? appears.

While the website serves as the primary tool for outreach, staff also interact directly with users at international meetings, workshops, presentations, festivals, and more.

Our outreach efforts inform users about new developments while collecting feedback from the PDB community. This helps the RCSB PDB resource to meet its mission in the interest of science, medicine, and education.

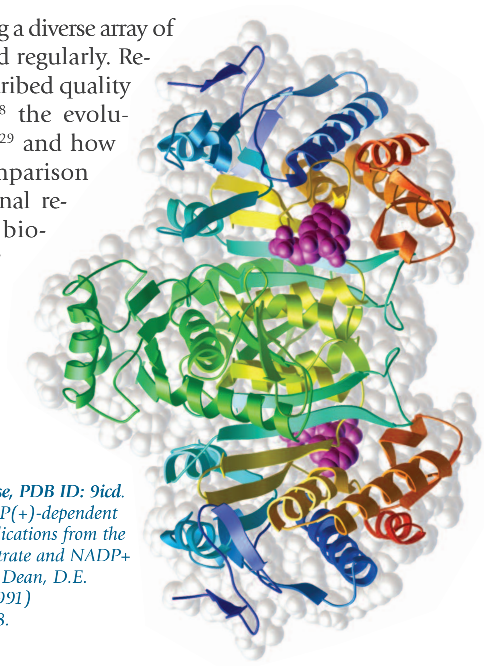
The RCSB PDB promoted a Structural View of Biology at the February 2011 meeting of the American Association for the Advancement of Science in Washington, DC.

Publications

The RCSB PDB website is updated regularly with news, recent developments, new resources, improvements to existing resources, and educational resources.

Published in print and online, the quarterly *Newsletter* describes and highlights recent activities. A variety of flyers, brochures, and tutorials are distributed to users at meetings and online.

Journal articles covering a diverse array of subjects are published regularly. Recent articles have described quality assurance practices,[28] the evolution of the website[16,29] and how to use it,[30] the Comparison Tool,[26] and educational resources for exploring biological structures.[31-33]



From Isocitrate Dehydrogenase, PDB ID: 9icd. *Catalytic mechanism of NADP(+)-dependent isocitrate dehydrogenase: implications from the structures of magnesium-isocitrate and NADP+ complexes. J.H. Hurley, A.M. Dean, D.E. Koshland Jr., R.M. Stroud (1991)* Biochemistry *30: 8671-8678.*

PDB-101: Introduction to a Structural View of Biology

PDB-101 is a new and unique view of the RCSB PDB website that places educational materials front and center. It packages together resources of interest to teachers, students, and the general public to promote exploration in the world of proteins and nucleic acids:

• ***Molecule of the Month*** series. Describing molecules from AAA+ proteases to zinc fingers, this feature provides an easy introduction to macromolecular structures, shows how the structures function, and highlights their importance in our lives.

Created by David S. Goodsell (RCSB PDB, The Scripps Research Institute), each illustrated entry links to high resolution images, asks discussion questions, and hosts customized interactive 3D molecular views. Biologically interesting examples of structures are shown in specific PDB-101 views that supplement summary information about the entry with functional context.

The *Molecule of the Month* resource is used in many classrooms to introduce structures to students, and is an integral part of the Science Olympiad protein modeling competitions.

• **Structural View of Biology.** Built around the *Molecule of the Month* series, this interface promotes a top-down exploration of the PDB. Beginning with high-level functional categories, readers can browse through descriptive subcategories to access relevant articles and related PDB entries. Mouseovers, pulldown menus, and carousels all offer easy navigational tools to promote learning.

• Available **Educational Resources** include posters, animations, and classroom lessons and activities. Recent additions include a downloadable PDF that can be used to build 3D models of DNA.

• **Understanding PDB Data** provides a reference to help explore and interpret individual PDB entries. Broad topics include how to understand PDB data, how to visualize structures, how to read coordinate files, and potential challenges to exploring the archive.

Educational Programs and Activities

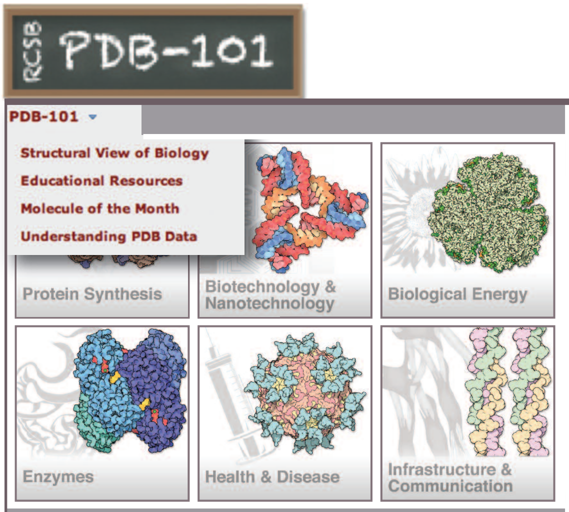
The RCSB PDB participates in large-scale events for local communities, such as the San Diego Science Festival and Rutgers Day. At these festivals, event attendees build hands-on virus and DNA models and learn about protein and nucleic acid structure.

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. Area high school classes visit and tour the RCSB PDB. These experiences often lead to new initiatives and collaborations.

At Rutgers and UCSD, the RCSB PDB leaders are involved in graduate and undergraduate courses and summer student research projects that depend heavily on the data in the PDB archive.

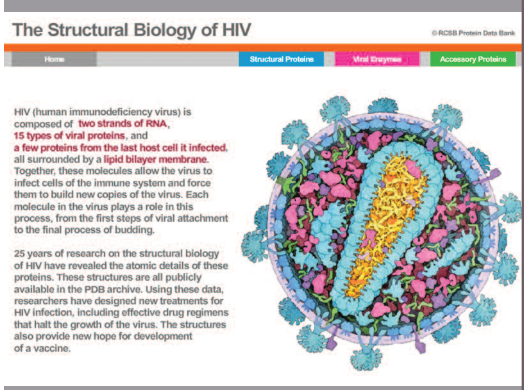
In 2012, the RCSB PDB will sponsor the protein modeling event at the Science Olympiad competitions held in New Jersey and in San Diego, CA. In this challenge, high school teams will demonstrate their understanding of proteins involved in the regulation of apoptosis through their mastery of hand-built models and onsite exams. Information and resources can be found at **education.pdb.org** and **twitter.com/buildmodels**. The event is organized nationally by the MSOE Center for BioMolecular Modeling (**cbm.msoe.edu**).

Visitors to RCSB PDB booth at the San Diego Science Festival learned about structures and built DNA models.



The Structural View of Biology promotes top-down exploration of the PDB. With a few mouse clicks, travel from descriptions of *Biotechnology & Nanotechnology* to *Industrial Applications* down to examples of *Antifreeze Proteins*.

Download, print, and fold this template to build a 3D DNA model.



The structural proteins, viral enzymes, and accessory proteins in the HIV virus are described in a poster and interactive animation.

The RCSB PDB also collaborates with resources that help to promote a broad view of structural biology.

PSI | nature
StructuralBiologyKnowledgebase

Structural Genomics and the PSI Structural Biology Knowledgebase

More than 10,000 structures–nearly 15% of the PDB archive–are the result of structural genomics efforts worldwide. The RCSB PDB works closely with these centers in data deposition and annotation, and collaborates with related resources.

The largest American structural genomics effort, the NIGMS's Protein Structure Initiative:Biology Network, recently launched TargetTrack to capture and disseminate the experimental progress of proteins studied by the PSI. Protein production and structure determination information is released to the public prior to publication in order to let others learn from the PSI efforts. TargetTrack is one of the databases supported by the PSI Structural Biology Knowledgebase (PSI SBKB, **sbkb.org**).

Through a web portal, the SBKB integrates the results of structural biology and genomics efforts with other publicly available biological information to facilitate further research. A search using a sequence, plain text, or PDB ID provides access to all PDB entries, target tracking information, available protocols, available clones and materials, annotations from more than 100 biological resources, homology models, publications, and technologies. Similar results can be automatically found with each PDB summary on the RCSB PDB website. Monthly editorial updates help to further broaden the view of structural biology.

BioSync

The Structural Biology Synchrotron Users Organization

BioSync is a guide to high energy data collection facilities for structural biologists.[34] Data collected from synchrotron beamlines currently account for 75% of all X-ray crystallographic entries deposited to the PDB. **BioSync.sbkb.org** provides detailed information and usage statistics for macromolecular beamlines at synchrotron facilities worldwide. Users can access information about specific capabilities, equipment, services and functions, and more.

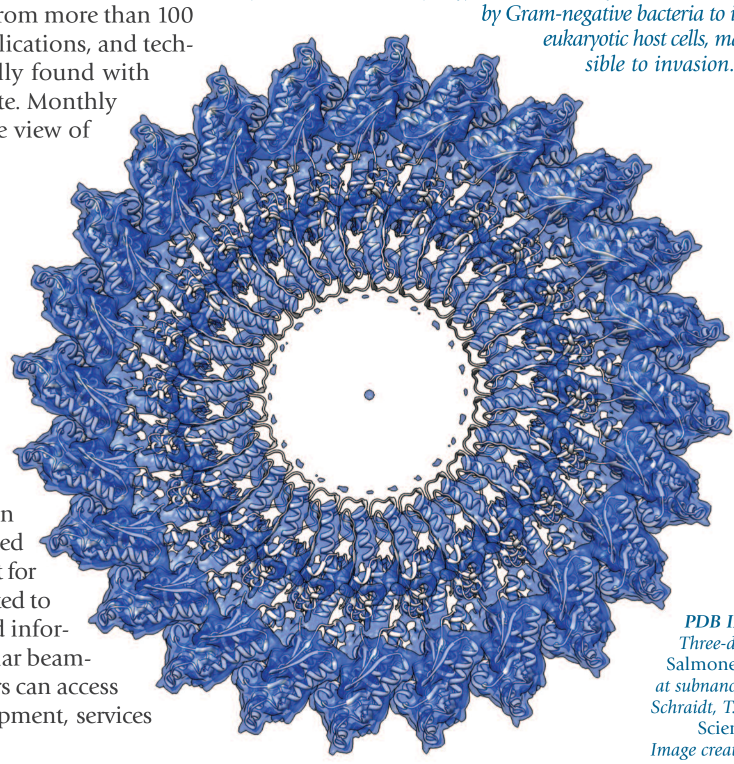
EMDataBank
Unified Data Resource for 3DEM

One Stop Shop for 3DEM Deposition and Retrieval

The PDB archives large biological assemblies determined by 3D electron microscopy (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**[36] is a deposition and retrieval network for 3DEM map, model, and associated metadata.

In 2012, the EMDB map archive will be merged with the PDB archive, enabling 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.

***Below:** 24-fold symmetric periplasmic domain of the inner membrane-associated ring IR1 of the type III secretion system needle complex from* Salmonella typhimurium. *Model coordinates are shown with the associated experimental EM map. Type III secretion systems are virulence factors used by Gram-negative bacteria to inject proteins into eukaryotic host cells, making them accessible to invasion.*



PDB ID 2y9j/EMD-1874. Three-dimensional model of Salmonella's *needle complex at subnanometer resolution. O. Schraidt, T.C. Marlovits (2011)* Science *331: 1192-1195. Image created using Chimera.*[35]

References

1. H.M. Berman, J.D. Westbrook, Z. Feng, G. Gilliland, T.N. Bhat, H. Weissig, I.N. Shindyalov, P.E. Bourne. (2000) The Protein Data Bank. *Nucleic Acids Res* **28**: 235-242.
2. H.M. Berman, K. Henrick, H. Nakamura. (2003) Announcing the worldwide Protein Data Bank. *Nat Struct Biol* **10**: 980.
3. F.C. Bernstein, T.F. Koetzle, G.J.B. Williams, E.F. Meyer Jr., M.D. Brice, J.R. Rodgers, O. Kennard, T. Shimanouchi, M. Tasumi. (1977) Protein Data Bank: a computer-based archival file for macromolecular structures. *J Mol Biol* **112**: 535-542.
4. S. Velankar *et al.* (2011) PDBe: Protein Data Bank in Europe. *Nucleic Acids Res* **39**: D402-D410
5. D.M. Standley, A.R. Kinjo, K. Kinoshita, H. Nakamura. (2008) Protein structure databases with new web services for structural biology and biomedical research. *Brief Bioinform* **9**: 276-285.
6. E.L. Ulrich *et al.* (2008) BioMagResBank. *Nucleic Acids Res* **36**: D402-408.
7. P.M.D. Fitzgerald, J.D. Westbrook, P.E. Bourne, B. McMahon, K.D. Watenpaugh, H.M. Berman. (2005). 4.5 Macromolecular dictionary (mmCIF). In *International Tables for Crystallography* (S. R. Hall & B. McMahon, eds.), Vol. G. Definition and exchange of crystallographic data, pp. 295-443. Springer, Dordrecht, The Netherlands.
8. 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.
9. J.D. Westbrook, N. Ito, H. Nakamura, K. Henrick, H.M. Berman. (2005) PDBML: The representation of archival macromolecular structure data in XML. *Bioinformatics* **21**: 988-992.
10. H. Yang, V. Guranovic, S. Dutta, Z. Feng, H. M. Berman, J. Westbrook. (2004) Automated and accurate deposition of structures solved by X-ray diffraction to the Protein Data Bank. *Acta Cryst* **D60**: 1833-1839.
11. 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.
12. 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.
13. 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.
14. Rutgers University, New Brunswick, NJ. (1998). NUCheck. Z. Feng, J. Westbrook, H. M. Berman. NDB 407.
15. R. J. Read *et al.* (2011) A new generation of crystallographic validation tools for the Protein Data Bank. *Structure* **19**: 1395-1412.
16. P.W. Rose *et al.* (2011) The RCSB Protein Data Bank: redesigned web site and web services. *Nucleic Acids Res* **39**: D392-D401.
17. Enzyme Nomenclature, Enzyme Classification **www.chem.qmw.ac.uk/iubmb/enzyme.**
18. 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.
19. C.A. Orengo, A.D. Michie, S. Jones, D.T. Jones, M.B. Swindells, J.M. Thornton. (1997) CATH–a hierarchic classification of protein domain structures. *Structure* **5**: 1093-1108.
20. M.H. Saier, Jr., M.R. Yen, K. Noto, D.G. Tamang, C. Elkan. (2009) The Transporter Classification Database: recent advances. *Nucleic Acids Res* **37**: D274-278.
21. M. Ashburner *et al.* (2000) Gene ontology: tool for the unification of biology. The Gene Ontology Consortium. *Nat Genet* **25**: 25-29.
22. E.W. Sayers *et al.* (2011) Database resources of the National Center for Biotechnology Information. *Nucleic Acids Res* **39**: D38-51.
23. R.D. Finn *et al.* (2010) The Pfam protein families database. *Nucleic Acids Res* **38**: D211-222.
24. Jmol: an open-source Java viewer for chemical structures in 3D. **www.jmol.org.**
25. J.L. Moreland, A. Gramada, O. V. Buzko, Q. Zhang, P.E. Bourne. (2005) The Molecular Biology Toolkit (MBT): a modular platform for developing molecular visualization applications. *BMC Bioinformatics* **6**: 21.
26. A. Prlić, S. Bliven, P.W. Rose, W. F. Bluhm, C. Bizon, A. Godzik, P.E. Bourne. (2010) Pre-calculated protein structure alignments at the RCSB PDB website. *Bioinformatics* **26**: 2983-2985.
27. M. Ryan, M. Diekhans, S. Lien, Y. Liu, R. Karchin. (2009) LS-SNP/PDB: annotated non-synonymous SNPs mapped to Protein Data Bank structures. *Bioinformatics* **25**: 1431-1432.
28. W.F. Bluhm, B. Beran, C. Bi, D. Dimitropoulos, A. Prlić, G.B. Quinn, P.W. Rose, C. Shah, B. Yukich, H.M. Berman, P.E. Bourne. (2011) Quality assurance for the query and distribution systems of the RCSB Protein Data Bank. *Database* **doi: 10.1093/database/bar003**.
29. P.E. Bourne *et al.* (2011) The evolution of the RCSB Protein Data Bank website. *WIREs Computational Molecular Science* **1**: 782-789.
30. P.W. Rose, W.F. Bluhm, B. Beran, C. Bi, D. Dimitropoulos, D.S. Goodsell, A. Prlić, G.B. Quinn, B. Yukich, H.M. Berman. (2011) The RCSB Protein Data Bank: site functionality and bioinformatics use cases. *NCI-Nature Pathway Interaction Database* Bioinformatics Primer **doi:10.1038/pid.2011.1.**
31. S. Dutta, C. Zardecki, 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.
32. D.S. Goodsell. (2010) The Protein Data Bank: Exploring Biomolecular Structure *Nature Education* **3**: 39.
33. D.S. Goodsell. (2011) Miniseries: Illustrating the machinery of life: Eukaryotic cell panorama. *Biochem Mol Biol Educ* **39**: 91-101.
34. A. Kuller, W. Fleri, W.F. Bluhm, J.L. Smith, J. Westbrook, P.E. Bourne. (2002) A biologist's guide to synchrotron facilities: the BioSync web resource. *TIBS* **27**: 213-215.
35. E.F. Pettersen, T.D. Goddard, C.C. Huang, G.S. Couch, D.M. Greenblatt, E.C. Meng, T.E. Ferrin. (2004) UCSF Chimera–a visualization system for exploratory research and analysis. *J Comput Chem* **25**: 1605-1612.
36. C.L. Lawson *et al.* (2011) EMDataBank.org: unified data resource for CryoEM. *Nucleic Acids Res* **39**: D456-D464.
37. N. Ban, P. Nissen, J. Hansen, P.B. Moore, T.A. Steitz. (2000) The complete atomic structure of the large ribosomal subunit at a 2.4 Å resolution. *Science* **289**: 905-920.
38. A.P. Carter, W.M. Clemons, D.E. Brodersen, R.J. Morgan-Warren, B.T. Wimberly, V. Ramakrishnan. (2000) Functional insights from the structure of the 30S ribosomal subunit and its interactions with antibiotics. *Nature* **407**: 340-348.
39. F. Schluenzen, A. Tocilj, R. Zarivach, J. Harms, M. Gluehmann, D. Janell, A. Bashan, H. Bartels, I. Agmon, F. Franceschi, A. Yonath. (2000) Structure of functionally activated small ribosomal subunit at 3.3 Å resolution. *Cell* **102**: 615-623.
40. R. Henderson, J.M. Baldwin, T.A. Ceska, F. Zemlin, E. Beckmann, K.H. Downing. (1990) Model for the structure of bacteriorhodopsin based on high-resolution electron cryo-microscopy. *J Mol Biol* **213**: 899-929.
41. P.C. Driscoll, A.M. Gronenborn, L. Beress, G.M. Clore. (1989) Determination of the three-dimensional solution structure of the antihypertensive and antiviral protein BDS-I from the sea anemone *Anemonia sulcata*: a study using nuclear magnetic resonance and hybrid distance geometry-dynamical simulated annealing. *Biochemistry* **28**: 2188-2198.
42. H.R. Drew, R.M. Wing, T. Takano, C. Broka, S. Tanaka, K. Itakura, R.E. Dickerson. (1981) Structure of a B-DNA dodecamer: conformation and dynamics. *Proc Natl Acad Sci USA* **78**: 2179-2183.
43. S.H. Kim, G.J. Quigley, F.L. Suddath, A. McPherson, D. Sneden, J. J. Kim, J. Weinzierl, A. Rich. (1973) Three-dimensional structure of yeast phenylalanine transfer RNA: folding of the polynucleotide chain. *Science* **179**: 285-288.
44. J.D. Robertus, J. E. Ladner, J. T. Finch, D. Rhodes, R. S. Brown, B.F.C. Clark, A. Klug. (1974) Structure of yeast phenylalanine tRNA at 3 Å resolution. *Nature* **250**: 546-551.

RCSB PDB

wwPDB Members

WORLDWIDE PDB PROTEIN DATA BANK

RCSB PDB
www.rcsb.org
Protein Data Bank Europe
pdbe.org
Protein Data Bank Japan
www.pdbj.org
BioMagResBank
www.bmrb.wisc.edu

RCSB PDB Partners

RUTGERS

Rutgers, The State University of New Jersey,
Department of Chemistry and Chemical Biology
610 Taylor Road
Piscataway, NJ 08854-8087

UCSD SDSC. SKAGGS SCHOOL of PHARMACY and PHARMACEUTICAL SCIENCES

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

Dr. Martha Quesada
Deputy Director
Rutgers, The State University of New Jersey
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

A list of current RCSB PDB Team Members is available at **www.pdb.org**.

www.rcsb.org • info@rcsb.org