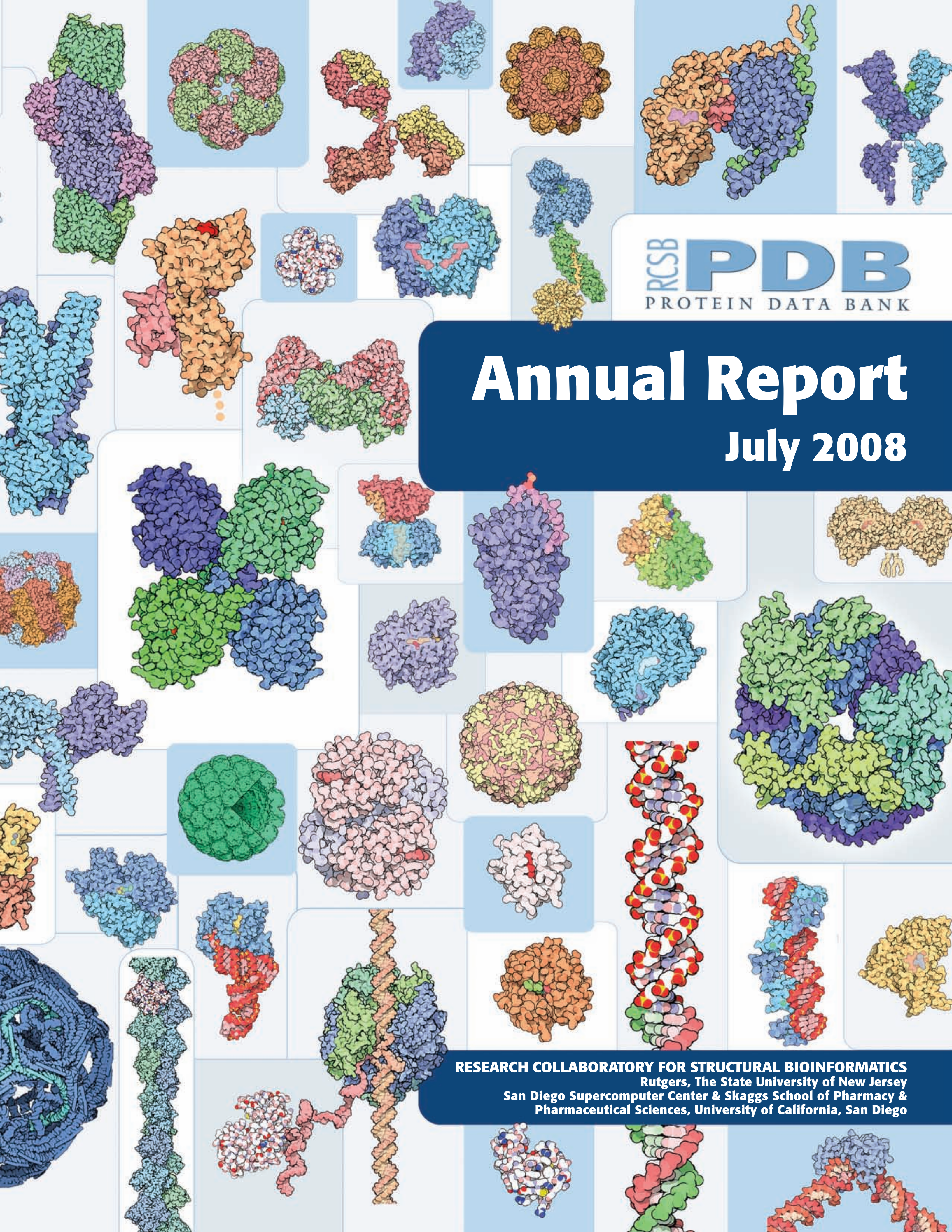


RCSB PDB
PROTEIN DATA BANK

Annual Report
July 2008

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

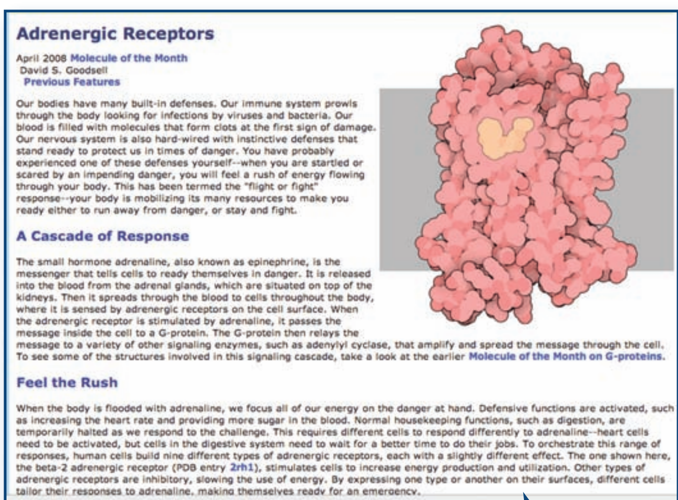
About the Cover

A milestone was reached in April 2008 when **the 100th installment of the *RCSB PDB Molecule of the Month*** was published on adrenergic receptors. This report cover highlights all of the structures published in this series since January 2000.

Created by David S. Goodsell (RCSB PDB, The Scripps Research Institute), the feature regularly provides an easy introduction to the RCSB PDB for teachers and students. It is used in many classrooms to introduce structures to students, and is an integral part of the protein modeling event at high school Science Olympiad competitions. Over time, the series has explored the structure and function of proteins and nucleic acids found in the PDB archive such as insulin, anthrax toxin, and multidrug resistance transporters.

The descriptions also provide links to example structures in the RCSB PDB and offer interactive Jmol[1] views for further exploration.

New *Molecule of the Month* features and the archive of previous editions are available from the RCSB PDB home page at **www.pdb.org**.



Adrenergic Receptors

April 2008 Molecule of the Month
David S. Goodsell
Previous Features

Our bodies have many built-in defenses. Our immune system prowls through the body looking for infections by viruses and bacteria. Our blood is filled with molecules that form clots at the first sign of damage. Our nervous system is also hard-wired with instinctive defenses that stand ready to protect us in times of danger. You have probably experienced one of these defenses yourself--when you are startled or scared by an impending danger, you will feel a rush of energy flowing through your body. This has been termed the "flight or fight" response--your body is mobilizing its many resources to make you ready either to run away from danger, or stay and fight.

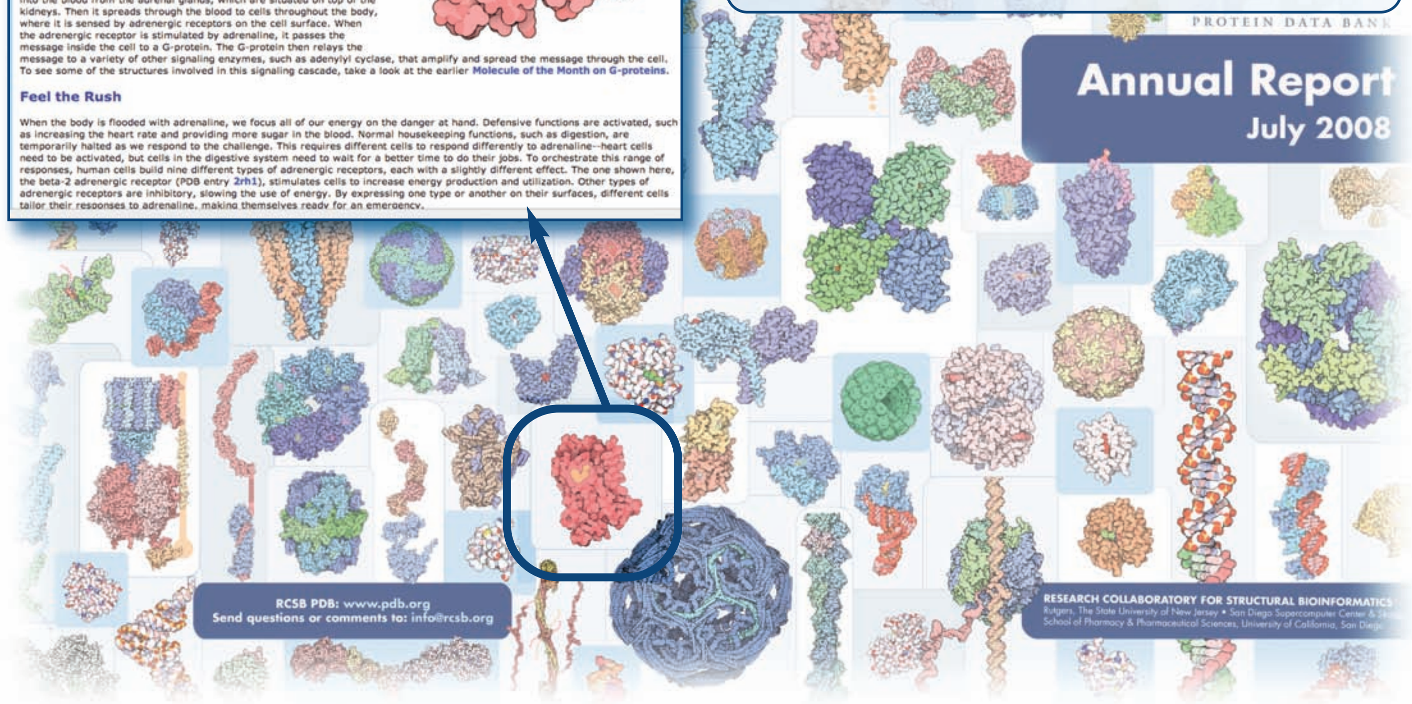
A Cascade of Response

The small hormone adrenaline, also known as epinephrine, is the messenger that tells cells to ready themselves in danger. It is released into the blood from the adrenal glands, which are situated on top of the kidneys. Then it spreads through the blood to cells throughout the body, where it is sensed by adrenergic receptors on the cell surface. When the adrenergic receptor is stimulated by adrenaline, it passes the message inside the cell to a G-protein. The G-protein then relays the message to a variety of other signaling enzymes, such as adenylyl cyclase, that amplify and spread the message through the cell. To see some of the structures involved in this signaling cascade, take a look at the earlier Molecule of the Month on G-proteins.

Feel the Rush

When the body is flooded with adrenaline, we focus all of our energy on the danger at hand. Defensive functions are activated, such as increasing the heart rate and providing more sugar in the blood. Normal housekeeping functions, such as digestion, are temporarily halted as we respond to the challenge. This requires different cells to respond differently to adrenaline--heart cells need to be activated, but cells in the digestive system need to wait for a better time to do their jobs. To orchestrate this range of responses, human cells build nine different types of adrenergic receptors, each with a slightly different effect. The one shown here, the beta-2 adrenergic receptor (PDB entry 2rh1), stimulates cells to increase energy production and utilization. Other types of adrenergic receptors are inhibitory, slowing the use of energy. By expressing one type or another on their surfaces, different cells tailor their responses to adrenaline, making themselves ready for an emergency.

Contents



References for the structures depicted on the cover are available from each Molecule of the Month *feature. Images are not drawn to scale.*

Message from the Director

When the Protein Data Bank (PDB) archive was founded with seven structures in 1971, it was difficult to predict the future of structural biology. With the release of the 50,000th structure in the archive in April 2008, the PDB holdings tell many stories about the progress of this science through the years.

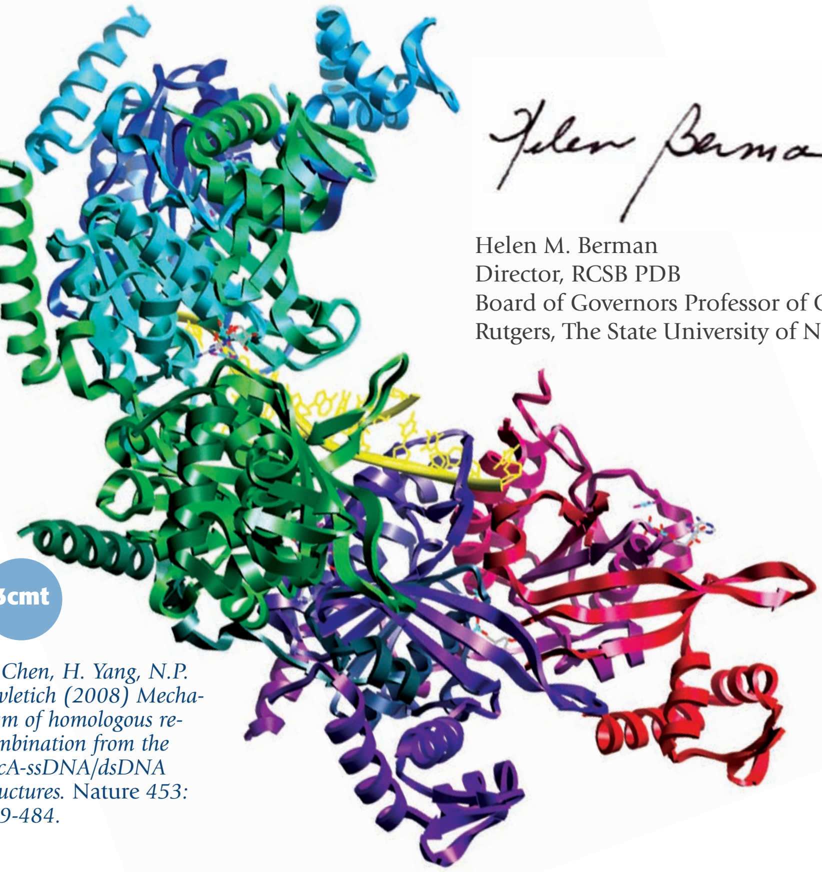
Improvements in the technologies and methods used in structure determination have enabled scientists to make exciting new discoveries. Structures that were once thought impossible to determine (such as ribosomes, large viruses, and enzyme complexes) are now commonplace. These large and complex molecular machines–along with other proteins and nucleic acids of all sizes–tell us about the different processes that happen in living cells.

But we're not done yet. Laboratories are producing structural data at a constantly increasing rate. Using the power of high throughput crystallography and NMR techniques, structural genomics projects worldwide have produced a large number of new and novel structures. The function of many of these structures is unknown, offering ongoing challenges to the scientific community.

More than 20 million files are accessed from the PDB archive every month through wwPDB websites and FTP archive downloads. This tells the story of how the PDB user community has expanded from 1970s, when a handful of structural biologists shared their data by postal mail. Today, PDB users are scientists working on basic and applied research in biology, biochemistry, genetics, pharmacology, biophysics, bioinformatics, and beyond. Computer scientists and software developers build and use tools for PDB data analysis and visualization. Students and educators at all levels of experience and interest utilize PDB data in their classes and projects.

Estimates indicate that the size of the PDB archive will triple to 150,000 structures by the year 2014. We look forward to a future in which biology and medicine are increasingly described in molecular terms.

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



3cmt

Z. Chen, H. Yang, N.P. Pavletich (2008) Mechanism of homologous recombination from the RecA-ssDNA/dsDNA structures. Nature *453: 489-484.*

About the RCSB PDB

The RCSB Protein Data Bank, administered by the nonprofit Research Collaboratory for Structural Bioinformatics (RCSB), provides an essential library of biomolecular structures to support scientific research and education worldwide.

The Protein Data Bank (PDB) archive is the single repository of information about the 3D structures of large biological molecules, primarily proteins and nucleic acids. These are the molecules of life that are found in all organisms, from bacteria and plants to animals and humans.

Among the more than **50,000** structures currently represented in the PDB are:

- **Transfer RNA (tRNA)**–The nucleic acid molecule that translates genomic information into the amino acid sequence of a protein
- **Dengue fever virus**–The fever-causing virus transmitted by tropical mosquitoes
- **Anabolic steroid receptors**–Hormones, including testosterone, that regulate protein synthesis
- **Prion proteins**–Misshapen proteins that are the cause of many diseases, including Mad Cow Disease
- **Circadian clock proteins**–Molecules that measure out a 24-hour day
- **p53 tumor suppressor**–"Guardian of the cell" that initiates programmed cell death in damaged or cancerous cells
- **Amyloid peptide**–A protein implicated in Alzheimer's disease
- **Photosynthetic reaction centers**–Proteins that capture light for the creation of sugar in plants

Snapshot: July 1, 2008

51491 released atomic coordinate entries

Molecule Type

47526	proteins, peptides, and viruses
1870	nucleic acids
2062	protein/nucleic acid complexes
33	other

Experimental Technique

43855	X-ray
7355	NMR
182	electron microscopy
99	other
33017	structure factor files
4054	NMR restraint files

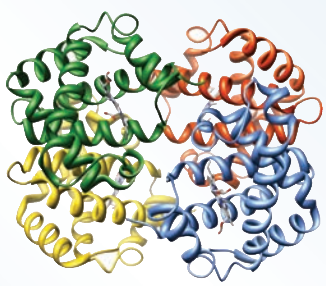
In addition to developing tools and systems for the deposition, annotation, and release of data in the PDB archive, the RCSB PDB supports a website where visitors can perform simple and complex queries on the data, visualize the structures, and analyze the results.

The RCSB PDB is a valuable resource used by researchers from a wide variety of scientific disciplines, students, teachers, and the general public.

Why Study Biomolecular Structure?

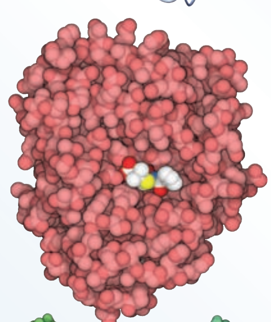
4hhb

G. Fermi, M.F. Perutz, B. Shaanan, R. Fourme (1984) The crystal structure of human deoxyhaemoglobin at 1.74 Å resolution. J.Mol.Biol. *175: 159-174.*



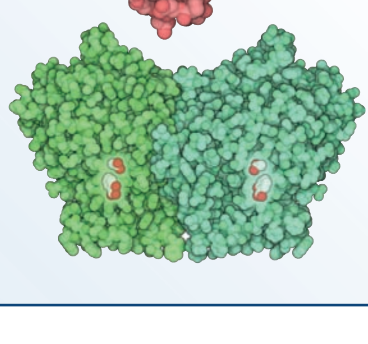
1pwc

N.R. Silvaggi, H.R. Josephine, A.P. Kuzin, R. Nagarajan, R.F. Pratt, J.A. Kelly (2005) Crystal structures of complexes between the R61 DD-peptidase and peptidoglycan-mimetic beta-lactams: a non-covalent complex with a "perfect penicillin." J.Mol.Biol. *345: 521-533. Image by David S. Goodsell.*



1pth

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. Image by David S. Goodsell.*



Understanding the shape of a molecule helps to understand how it works. Biomolecules are the main building blocks of living organisms. They come in a variety of shapes, ranging from tiny proteins and bits of DNA to complex and large molecular machines like the ribosome and ATP synthase. These different shapes enable the structures to do their jobs.

For example, the four chains of hemoglobin are arranged in a way that allows efficient pickup and delivery of oxygen and other gasses.

By studying these structures, scientists can deduce their functions in human health and disease.

This knowledge is also used to develop drugs, the small molecules that bind to a specific protein and modify its action. Some very powerful drugs, such as antibiotics or anticancer drugs, are used to completely disable a critical molecular machine. These drugs can kill a bacterial or cancer cell. Other drug molecules, such as aspirin, gently block less critical proteins for a few hours.

The structures in the PDB archive provide a vital resource for science and medicine.

About the RCSB PDB

History of the PDB Archive

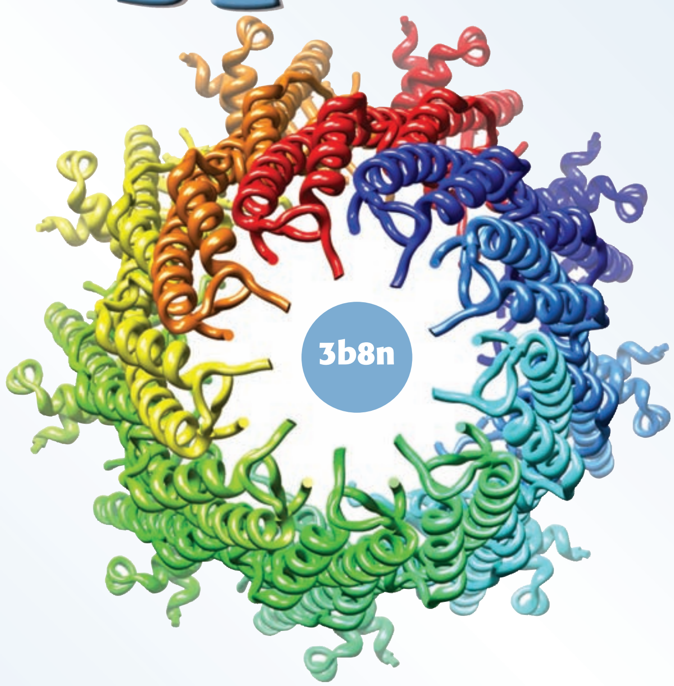
Almost as soon as it became possible to deduce the form of a protein, scientists realized that this important information must be shared. The PDB archive is the sole international repository for 3D structures of biological macromolecules. The resource includes proteins, nucleic acids (including DNA and RNA), and protein-nucleic acid complexes.

The PDB was established in 1971 at Brookhaven National Laboratory and originally contained 7 structures.[2] In 1998, the Research Collaboratory for Structural Bioinformatics (RCSB) became responsible for the management of the PDB.[3] In 2003, the Worldwide Protein Data Bank (wwPDB) was formed to ensure that the archive would be freely and publicly available to the global community.[4] The wwPDB consists of organizations that act as deposition, data processing and distribution centers for PDB data. As 'archive keeper' for the wwPDB, the RCSB PDB controls the central repository of the PDB archive.

Today, the PDB archive receives approximately 20 new structures daily. Curators then check each deposition for errors or omissions, ensuring a consistent format and accurate data for each entry.

The Life of a Structure

1 **Solving the puzzle of a protein's shape requires advanced techniques and careful analysis. Scientists use methods such as X-ray crystallography, nuclear magnetic resonance (NMR), and 3D electron microscopy (3D EM) to acquire structural data.**



2 **When a scientist determines a new structure, this information is deposited in the PDB archive. Annotators review and annotate the data before it is made freely available to the world.**

3 **Structures are then downloaded and viewed from online resources, such as the RCSB PDB website that lets users search databases, tabulate data, and analyze the results.**

Different visualization programs allow users to "see" a 3D picture of a protein structure.

Data are picked up by countless external resources to populate other databases for use in further studies.

4 **As these valuable data are shared around the globe, they contribute to further understanding of biomolecular structure and to human health and welfare.**

For example, a structure could be used in a study of how proteins interact with one another, explored by a student researching bioinformatics, or used in the design of candidates in drug-development trials.

3b8n: A. Tocilj, C. Munger, A. Proteau, R. Morona, L. Purins, E. Ajamian, J. Wagner, M. Papadopoulos, L. Van Den Bosch, J.L. Rubinstein, J. Fethiere, A. Matte, M. Cygler (2008) Bacterial polysaccharide co-polymerases share a common framework for control of polymer length. Nat.Struct.Mol.Biol. 15: 130-138.

About the RCSB PDB

The RCSB PDB and collaborators

The RCSB PDB Organization

The RCSB PDB member institutions jointly manage the project: 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 chemical biology at Rutgers. Professor Berman was part of the team that first envisioned the PDB archive. She is joined in RCSB PDB management by Dr. Martha Quesada, Deputy Director (Rutgers), and Professor Philip E. Bourne, Associate Director (UCSD).

The RCSB PDB Team is comprised 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 international advisory board (The RCSB PDB Advisory Committee), made up of experts in X-ray crystallography, NMR, 3D EM, bioinformatics, and education. wwPDB Task Forces focusing on NMR and X-ray Validation also guide the direction of the resource.

The wwPDB Organization

In addition to the RCSB PDB, the member institutions of the wwPDB include Protein Data Bank Europe (PDBe), Protein Data Bank Japan (PDBj), and the BioMagResBank (BMRB). In addition to maintaining the archive, the groups collaborate on a variety of projects and outreach efforts.

www.wwpdb.org provides links to member sites for deposition, access, and searching. This portal hosts extensive documentation describing data deposition and annotation. Statistics for the number of structures deposited and processed, and download statistics for each structure are updated regularly.

Funding

The RCSB PDB is supported by funds from the National Science Foundation (NSF), the National Institute of General Medical Sciences (NIGMS), the Office of Science, Department of Energy (DOE), the National Library of Medicine (NLM), the National Cancer Institute (NCI), the National Center for Research Resources (NCRR), the National Institute of Biomedical Imaging and Bioengineering (NIBIB), the National Institute of Neurological Disorders and Stroke (NINDS), and the National Institute of Diabetes & Digestive & Kidney Diseases (NIDDK).

NSF NIGMS Office of Science U.S. DEPARTMENT OF ENERGY NATIONAL LIBRARY OF MEDICINE NATIONAL CANCER INSTITUTE National Center for Research Resources NIBIB NATIONAL INSTITUTE OF NEUROLOGICAL DISORDERS AND STROKE NIDDK

Services: Data Input

Data Input: Deposition, Validation, and Annotation

A key function of the wwPDB is the efficient capture (deposition) and curation (validation and annotation) of experimental structural data.[5] Scientists contribute data produced from structure determination experiments using deposition tools available from the wwPDB partners. These data are then validated and annotated before being made publicly available. Data processed at the other wwPDB sites are forwarded to the RCSB PDB for inclusion in the archive.

When a structure is deposited online using ADIT or ADIT-NMR, it is immediately assigned its own unique PDB ID.

Data annotators then work to represent PDB data in the best possible way. Using tools developed by the RCSB PDB, data entries are carefully reviewed and processed using an integrated system that is based on the use of a standard mmCIF (macromolecular Crystallographic Information File) dictionary.[6]

Annotators compare the sequence and citation reported in the deposition to external databases.[7-8] The entry is assigned a title, names and synonyms for the protein or other polymer, the scientific name of the source of the protein(s) and biological assembly information. Any format errors are corrected. After checking the structure visually, annotators send validation reports[9] and the completed coordinate file in mmCIF and PDB formats to the depositor for review. After corresponding with the depositor to finalize the entry for release, the complete entry, including its status information and PDB ID, is loaded into a relational database.

Depending upon the hold status selected by the depositor, data release occurs when a depositor gives approval to the annotated entry (status: REL), the hold date has expired (HOLD), or the journal article has been published (HPUB). Structures can be on HOLD or HPUB for no more than one year.

Software for Deposition and Validation

The RCSB PDB develops tools that facilitate data validation and deposition for depositors, even as structures are in the process of being determined. Software downloads, webservers, and documentation are available at **deposit.pdb.org**.

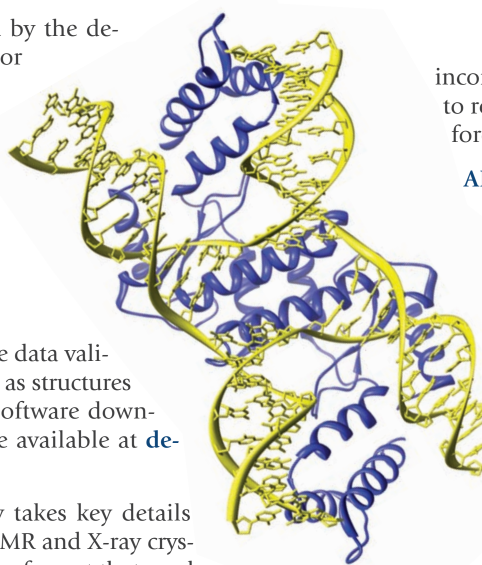
The program **pdb_extract**[10] automatically takes key details from the output files produced by many NMR and X-ray crystallographic applications and puts them in a format that can be used with ADIT and ADIT-NMR.

SF-Tool is a data validation and conversion tool that can validate model coordinates against structure factor data, translate structure factors into different formats, and check for twinned or detwinned data.

The **Validation Server** checks the format of the coordinate file and validates the overall structure before deposition.[9] It produces a validation report containing geometrical and experimental checks from several programs[11-13] and identifies any sequence and data inconsistencies. Researchers can use this program to review the quality of any released structure before using it in their own study.

ADIT (AutoDep Input Tool),[3] available from the RCSB PDB and PDBj, provides access to a collection of programs for data input, validation, annotation, and format exchange. **ADIT-NMR,** a similar program hosted at BMRB and PDBj, offers the same functionality, but also lets depositors submit NMR structural and experimental data using a single tool. Coordinates and constraint data are processed and released by the RCSB PDB and PDBj, while other NMR spectral data (such as chemical shifts, coupling constants, and relaxation parameters, *etc.*) are processed and archived by BMRB.

An enhanced version of ADIT that is currently in beta testing indicates possible remedies for file format problems; automatically validates the structure; enforces consistency between sequence and coordinates at deposition time; allows easier organization of sequence information and simplifies the way for entering author, title and citation information.



2qnc

C. Biertumpfel, W. Yang, D. Suck (2007) Crystal structure of T4 endonuclease VII resolving a Holliday junction. Nature *449: 616-662.*

Ligand Expo and the Chemical Component Dictionary

The Chemical Component Dictionary is as an external reference file describing all residue and small molecule components found in PDB entries. This dictionary contains detailed chemical descriptions for standard and modified amino acids/nucleotides, small molecule ligands, and solvent molecules. Each chemical definition includes descriptions of chemical properties such as stereochemical assignments, aromatic bond assignments, idealized coordinates, chemical descriptors (SMILES & InChI),[14,15] and systematic chemical names.

Users can search and browse the Chemical Component Dictionary using resources such as MSDchem[16] and Ligand Expo (**ligand-expo.rcsb.org**).

Developed by the RCSB PDB, Ligand Expo (formerly called Ligand Depot[17]), can be used to navigate the Chemical Component Dictionary. It integrates databases, services, tools and methods related to small molecules, and allows users to:

- **Search** for a chemical component
- **Browse** tables of components that contain
  – modified amino acids and nucleotides
  – trade and generic drugs names
  – common ring systems
- **Review** related information in chemical dictionaries and resource files (chemistry, geometry, atom nomenclature, and more)
- **Download** model and ideal chemical component coordinates
- **View** all instances of a component in released PDB entries

When depositing structures, users can search Ligand Expo to see if their ligand already exists in the PDB archive or use tools to build new components to be entered in the archive.

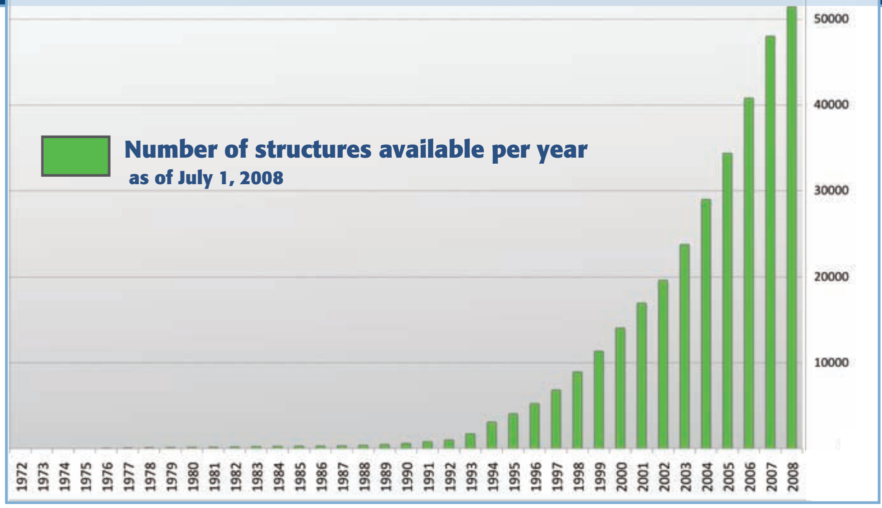
Services: Data Input

Statistics

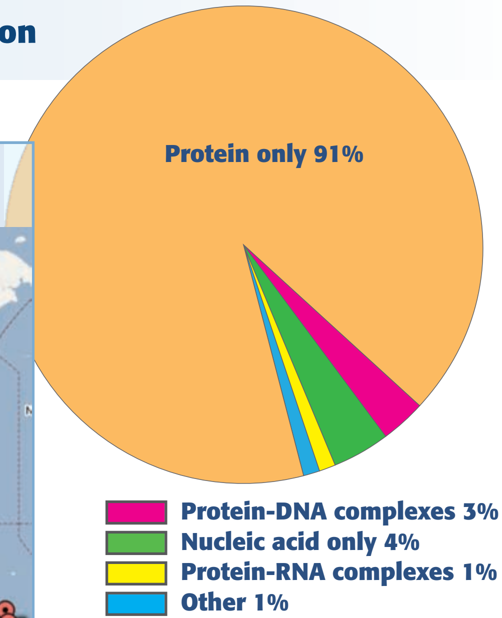
During the period covered by this report, 6978 files were deposited to the wwPDB from around the world. More than half were processed by the RCSB PDB; other structures were processed by wwPDB members PDBe and PDBj. Of these structures, approximately 96% were deposited with experimental data. Sequence data for about 55% of the depositions were released prior to the structure's release.

On average, structures are processed, reviewed by the author, and finalized for release in two weeks. RCSB PDB-developed software automatically tracks the status of data processing.

Growth of the Number of Structures Available in the PDB Archive



Structure Distribution



Distribution of proteins and nucleic acids in the archive in 2008

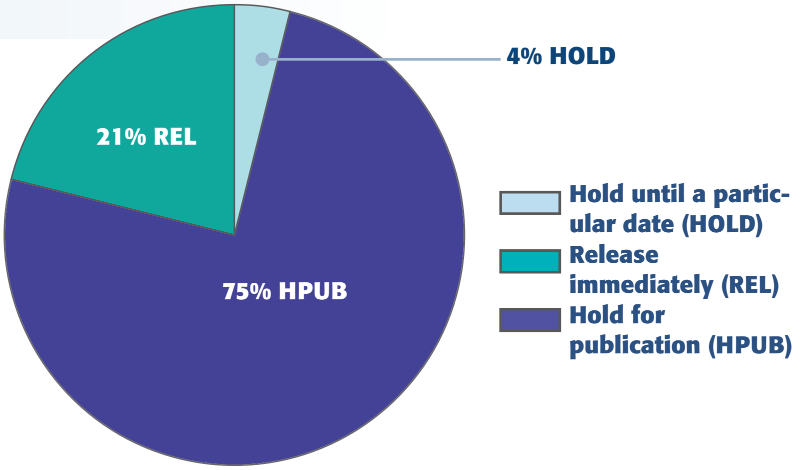
PDB Depositors by Location

Image created using Google Earth.[18]

Author-defined Release Status

After structures are fully processed and annotated, they are then released according to the status provided by the author at deposition. Of the 6978 entries deposited to the wwPDB during this report period, 75% were held until the corresponding journal publication was published.

The wwPDB receives publication dates and citation information directly from a few journals. Software is used to automatically scan for corresponding publications for most journals. Additionally, PDB depositors and users frequently send citation information to **deposit@deposit.rcsb.org**.



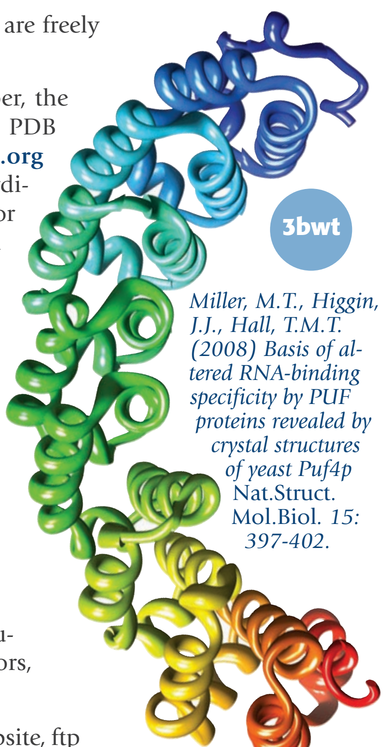
Services: Data Access, Query, and Reporting

FTP Downloads from the wwPDB in 2007. Image created using Google Earth.[18]

Data Distribution and Access

RCSB PDB services and data are freely available online.

As the wwPDB archive keeper, the RCSB PDB updates the PDB archive at **ftp://ftp.wwpdb.org** weekly. A total of 7319 coordinate files, 6087 structure factor files, and 554 constraint data files were released during this report period.



Miller, M.T., Higgin, J.J., Hall, T.M.T. (2008) Basis of altered RNA-binding specificity by PUF proteins revealed by crystal structures of yeast Puf4p Nat.Struct. Mol.Biol. *15:* 397-402.

Data are also loaded into a relational database,[19] integrated with more than 30 external sources of related data (*e.g.*, journal abstracts, functional descriptors, sequence annotations, structure annotations, taxonomy), and then made available for access by a large and diverse audience of scientists, educators, and students worldwide.

Access is available via the website, ftp server (supporting ftp and rsync access), and Web Services. The website is accessed by about 130,000 unique visitors per month from over 140 different countries. On a typical weekday, three to four pages from the website are viewed every second. Around 500 GigaBytes of data are transferred each month. At the same time, about 7,000 unique visitors download more than 10 million files from the FTP site at **ftp://ftp.wwpdb.org**, for a total of about 2 TB (TeraBytes) of data.

Data Query and Reporting

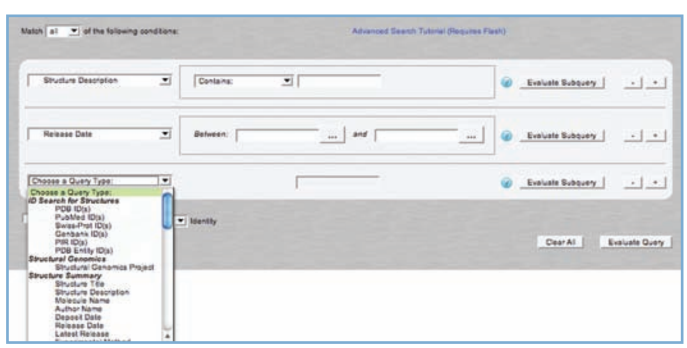
The website is designed to make every structure accessible and comprehensible to all of our data users–students and teachers, biologists, structural biologists, computational biologists, and the general public.

Users can search for structures using simple searches (PDB ID, keyword, sequence, or author) or by building more complex queries.

The Advanced Search interface can search across a large number of specific fields, such as external database IDs, keywords, structural features, biological and chemical information, sequences, experimental method, literature citations, and structural genomics center information. Subqueries can be combined *via* Boolean operations, and sequence-based redundancy can optionally be removed from search results.

Structures in the PDB may also be browsed using tree-like hierarchies based on Gene Ontology (GO) terms, Enzyme Classification, Medical Subject Headings, Source Organism, Genome Location, and SCOP and CATH classifications.[8, 20-23]

Structures in processing, on hold, or waiting for publication can be searched by author name and release status, and in some cases, structure name and sequence.

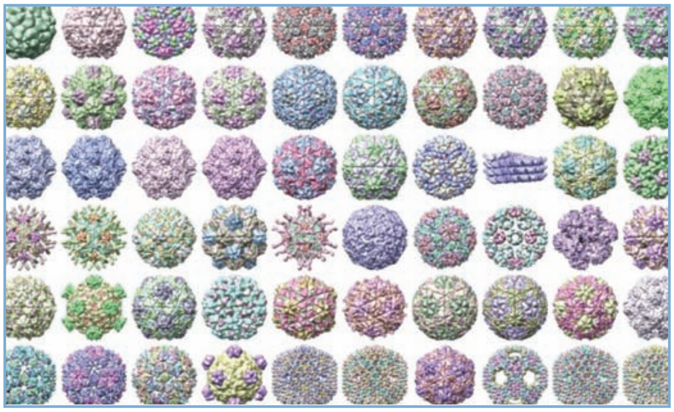


Advanced Search *can combine search criteria from a variety of categories.*

Services: Data Access, Query, and Reporting

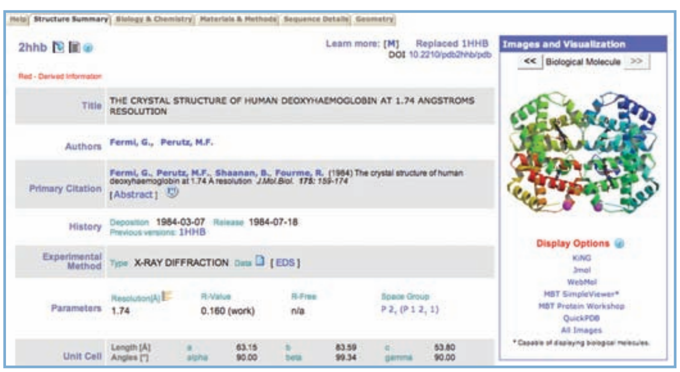
A variety of **Reports** can be created for any query results set. Tabs at the top of a search results set are available for: *citations*, for a PubMed-like list of the primary citations for the structures that match a query; *ligand hits*, for a list of ligands known to interact with the structures matching a query; *web page hits*, for a list of any RCSB PDB webpages that contain a particular keyword, including *Molecule of the Month* features; and tabs containing structures with corresponding *GO*, *SCOP*, or *CATH* hits. In the last three tabs, entries are returned in a tree browser that indicates where these structures reside in the respective hierarchies. The *SCOP* tab, for example, indicates which hits belong to which class of proteins.

Options to refine the query or create tabular reports from the search results are also available.



Search results, like the one shown for virus structures, can be viewed as a collage of structure images. Each image links to the Structure Summary page for the entry.

Whether they are looking at individual or multiple structures, users have a variety of options for learning about and visualizing the entries. Each individual entry has a **Structure Summary** page that provides summary information, static and interactive images of the molecule, and related links to other resources. Several molecular viewers, including the RCSB PDB's Protein Workshop,[24] are provided. A PDB, mmCIF or PDBML/XML format file for any structure can be **downloaded** as plain text or in compressed format. A sequence page, available for every structure, has been enhanced with new annotations such as new domain definitions, functional site mapping to sequence (if present in the entry), and calculated disulphide linkages.



An example Structure Summary page. Linked data items can be used to retrieve other structures with similar properties, such as author name, space group, or chemical component.

Statistical views of the archive are also presented as content distributions (experimental methods, resolution, organism, structural genomics center, *etc*.) and content growth (experimental methods, molecule type, and unique protein classifications).

All of these features are supported by help pages, Flash tutorials, and an active help desk.

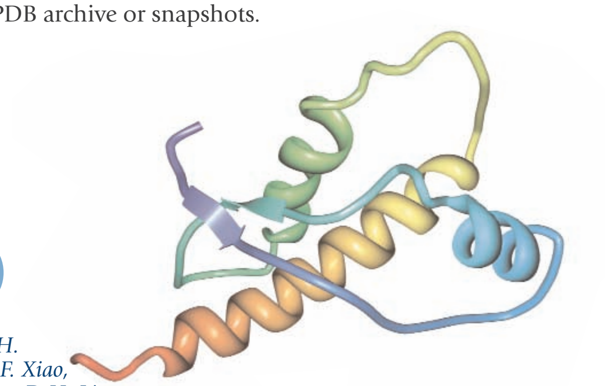
Additional enhancements and new resources are continually in development.

FTP: Downloading Tools and Time-stamped Copies

As part of a wwPDB initiative, time-stamped snapshots of the PDB archive are added each year to **ftp://snapshots.rcsb.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.

2jom



J. Li, F.H. Mei, G.F. Xiao, C.Y. Guo, D.H. Lin (2007) 1H, 13C and 15N resonance assignments of rabbit prion protein (91-228) J.Biomol.NMR *38: 181-181.*

This image was created using Protein Workshop, one of the many molecular viewer options available from the RCSB PDB website.

Web Services

Software tools can interact remotely with the RCSB PDB, instead of locally storing coordinate files and related data, through the use of Web Services. These services also allow developers to build platform-independent applications that use the advanced query functionality of the RCSB PDB.

Services offered include:

blastPDB: Runs a BLAST sequence search of the archive and returns a list of matching PDB IDs

runXMLQuery: Runs any query possible from the website. This includes advanced queries and combinations of queries with Boolean operations

The full RCSB PDB Web Services description language (WSDL) is at **www.pdb.org/pdb/services/pdbws?wsdl.**

Education and Outreach

Thousands of scientists visit **www.pdb.org** every day. However, advanced researchers are not the RCSB PDB's only users. The structures of proteins and nucleic acids are studied by teachers and students, and accessed by media writers, illustrators, and the general public. A number of resources, activities, and materials are available to engage our users at all levels to promote scientific literacy and a broader understanding of structural biology.

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

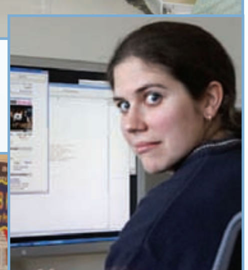
Other programs focus on getting students and teachers interested and involved in the fields of structural biology and bioinformatics. This includes supervising the protein modeling trial event (**education.pdb.org/olympiad**) at the New Jersey Science Olympiad competitions. At three events across the state, teams from all over the Garden State present hand-built 3D protein models using kits supplied by the RCSB PDB, along with an abstract, to be judged by our annotators. At the event, teams complete a written exam about the protein's structure and function and build another model.

Area high school classes frequently visit and tour the RCSB PDB. These experiences often lead to new initiatives and collaborations. For example, students studying protein structure-function relationships with the RCSB PDB, as part of a Howard Hughes Medical Institute program, created online presentations that can be seen at **www.scivee.tv.** This initiative will become part of a separate program to work with schools in-between San Diego and the Mexican border.

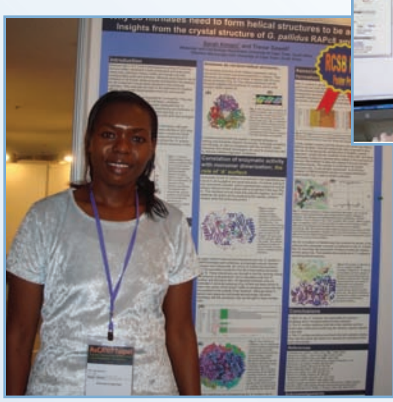
The RCSB PDB also exhibits at education-related meetings to talk to teachers and students about protein structure and function. At these meetings, the variety of resources available from the General Education section of **www.pdb.org**, in addition to the *RCSB PDB Molecule of the Month*, are discussed.

Education at All Levels

RCSB PDB's educational outreach to students of all ages and different interests requires a variety of activities.



Visits by high school classes and summer programs provide opportunities to learn about structural biology.



Student interns at both sites work on projects that become part of the online resource.

Serah Kimani (University of Cape Town) received the RCSB PDB Poster Prize at the 2007 Conference of the Asian Crystallographic Association. The prize is awarded to the best student poster at the affiliate meetings of the International Union of Crystallography and at the Intelligent Systems for Molecular Biology International Conference.

A lecture on data deposition was part of the 2008 Cold Spring Harbor Laboratory's X-ray Crystallography Course X-ray Methods in Structural Biology held at the Chinese Academy of Sciences in Beijing, China.

RCSB PDB members participate in a variety of outreach activities, including building virus models out of toothpicks and marshmallows with middle school students.

Number one in the state, Princeton High School students stand with their calmodulin model at the 2007 NJ Science Olympiad.

Education and Outreach

Community Interactions

Our outreach efforts are focused on informing users about the RCSB PDB and gaining input from the PDB community on how to improve and further develop resources.

Through a variety of channels, we collect feedback for further RCSB PDB development while providing materials that promote scientific literacy and a broader understanding of structural biology.

The electronic help desk at **info@rcsb.org** provides around-the-clock support for using RCSB PDB resources and beyond. Many requests made to this address become enhancements to the website. Other help desks assist in depositor-specific queries.

The RCSB PDB exhibit booths provide opportunities to meet with targeted audiences, like teachers and school district policy makers.

Team members actively participate in professional meetings through presenting lectures, giving demonstrations and tutorials, and collecting feedback through personal interactions.

Workshops are held to focus on particular topics. For example, the wwPDB sponsored a small meeting of software developers called *Future Challenges to the PDB* in the summer of 2007 to discuss data representation issues related to TLS refinement, alternate conformations and multiple models, structures larger then a typical PDB-formatted file, twinning, and new and hybrid experimental methods. Ideas generated in this meeting included forming an advisory council for input on such issues and meeting regularly with developers.

Later, the X-ray Validation Task Force was formed at the *Workshop on Next Generation Validation Tools* for the wwPDB held April 14-18, 2008 in Cambridge, UK to provide recommendations on additional validation that should be performed on PDB entries, and identify software applications to perform validation tasks.

Information about the Workshop on Next Generation Validation Tools *is available from* www.wwpdb.org.

Publications

www.pdb.org is updated weekly with news, recent developments, new resources, and improvements to existing documents. Educational features, such as new Molecule of the Month installments, are posted regularly.

Published quarterly, the RCSB PDB Newsletter describes and highlights recent RCSB PDB activities. Features include an Education Corner that describes how the PDB archive and RCSB PDB resources are used in the classroom, and a Community Focus interview with many of the scientific luminaries in the PDB user community.

A variety of flyers, brochures, and tutorials are distributed to users and published online.

The RCSB PDB regularly contributes articles to many peer-reviewed journals that cover a diverse array of subjects. Recent publications, among others,[25-26] have described:

- Depositing structures via the wwPDB[5], and particularly NMR depositions[27]
- The wwPDB Remediation Project[28]
- The history of the PDB archive[29]
- Outreach and education activities[30]

Education Resources

ImmersivePDB

At UCSD, students and scientists can engage with the ImmersivePDB virtual reality software for protein and nucleic acid structures at the Calit2 visualization lab. This program lets users see these structures in a very unique way.

The traveling *Art of Science* exhibit showcases the beauty and variety of protein shapes found in the PDB archive, bringing structural biology to art enthusiasts from New York to Austria. In the past year, this show included a display of photographs and sculptures of protein structures by artist Julian Voss-Andreae.

Interactive displays have been created to showcase structures at museums and at kiosks at RCSB PDB member sites.

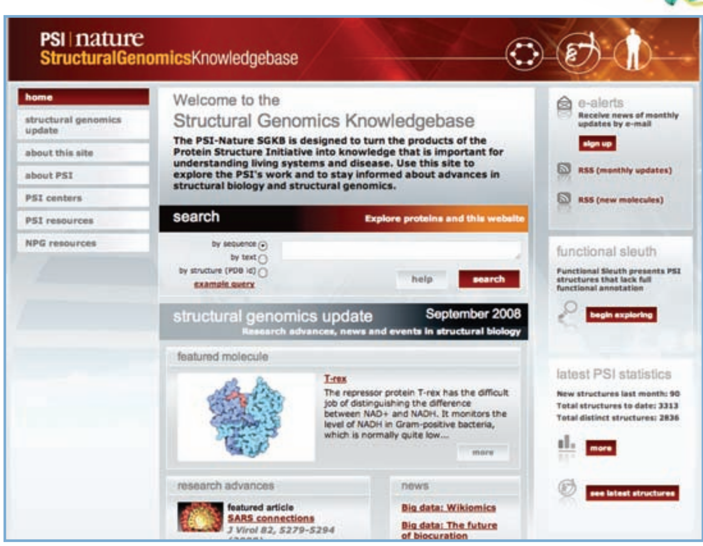
Related Resources and Collaborations

Structural Genomics

Structural genomics efforts are quickly determining a large number of novel structures in a high throughput mode. Since the PDB is the repository for these protein structures, the RCSB PDB works and collaborates with structural genomics resources and centers worldwide.

Current target progress and information is tracked by specific databases. The Target Registration Database (TargetDB; **targetdb.pdb.org**)[31] provides status information for targets selected for structure determination by various structural genomics centers. The Protein Expression Purification and Crystallization Database (PepcDB; **pepcdb.pdb.org**)[32] extends the content of TargetDB with status history, stop conditions, reusable text protocols and contact information collected from the Protein Structure Initiative (PSI) efforts and other structural genomics centers. These resources facilitate coordination among the different centers and provide valuable information about experimental design for use by biologists worldwide.

Structural genomics-related information from the PDB archive, along with the data from TargetDB and PepcDB, are integrated with key annotation, modeling, literature, and technology resources in the PSI SG Knowledgebase (PSI SGKB; **kb.psi-structuralgenomics.org**) funded by the NIGMS.[33] The PSI SGKB publicizes research advances in structural genomics and structural biology catalyzed by the PSI so that these structural, functional, and methodological advances are readily available for all biomedical scientists.

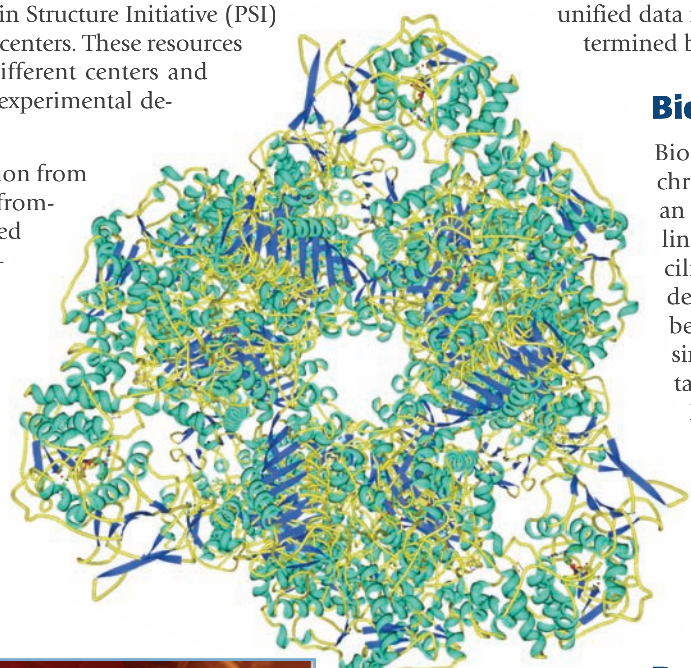


PSI SGKB was expanded recently when it joined forces with the Nature Publishing Group to create a PSI SGKB Gateway. The Gateway enhances the PSI SGKB's powerful query feature with editorial articles and features.

CryoEM

Cryo-electron microscopy (cryoEM) is a maturing methodology in structural biology that enables the determination of 3D structures of macromolecular complexes and cells from 2 to 100 Ångstrom resolution with information that bridges the gap between cell biology and crystallographic and NMR methods. A one-stop deposition and retrieval site for cryoEM density maps, atomic models and associated metadata is now available **emdatabank.org**.

The EM Data Bank, a joint effort among the Macromolecular Structure Database Group at the European Bioinformatics Institute, the RCSB PDB at Rutgers, and the National Center for Macromolecular Imaging at Baylor College of Medicine, aims to integrate the expertise and infrastructure of these centers to create a unified data resource for large complexes determined by cryoEM.



2vdc

M.Cottevieille, E. Larquet, S. Jonic, M.V. Petoukhov, G. Caprini, S. Paravisi, D.I. Svergun, M.A. Vanoni, N. Boisset (2008) The subnanometer resolution structure of the glutamate synthase 1.2-MDa hexamer by cryoelectron microscopy and its oligomerization behavior in solution: Functional implications. J.Biol.Chem. *283: 8237-8249.*

BioSync

BioSync (Structural Biology Synchrotron Users Organization)[34] is an online clearinghouse for beamline information at synchrotron facilities. The website contains descriptions of all synchrotron beamlines currently being used for single crystal macromolecular crystallography. The site also includes PDB deposition statistics with galleries of structures, cross-linked to the RCSB PDB, that are grouped by site and beamline. Separate tables of statistics and galleries are updated weekly for structural genomics depositions.

Data Uniformity

In August 2007, the wwPDB released remediated files for all PDB entries to ensure the uniformity of the data across the archive.[28] As part of the Remediation Project, many entries in the archive were updated and made consistent, with a focus on sequences (references to databases and taxonomies, plus differences between chemical and macromolecular sequences); primary citations; and assembly and virus information. In the Chemical Component Dictionary, the chemistry and nomenclature in monomers and ligands were standardized. This release of remediated data greatly improved searching and reporting capabilities across the PDB archive.

The wwPDB continues to actively collaborate in the improvement of data processing and annotation practices.

RCSB PDB

Citing PDB Structures and the RCSB PDB

Data files contained in the PDB archive (**ftp://ftp.wwpdb.org**) are free of all copyright restrictions and made fully and freely available for both non-commercial and commercial use. Users of the data should attribute the original authors of that structural data.

Structures should be cited with the PDB ID and the primary reference. For example, structure 102l should be referenced as: PDB ID: 102l. D.W. Heinz, W.A. Baase, F.W. Dahlquist, B.W. Matthews (1993) How amino-acid insertions are allowed in an alpha-helix of T4 lysozyme. *Nature* 361: 561-564.

Structures without a published reference can be cited with the PDB ID, author names, and title: PDB ID: 1ci0. W. Shi, D.A. Ostrov, S.E. Gerchman, V. Graziano, H. Kycia, B. Studier, S.C. Almo, S.K. Burley, New York Structural GenomiX Research Consortium (NYSGXRC) The structure of PNP oxidase from *S. cerevisiae.*

Structures may also be referenced using their Digital Object Identifier (DOI). The DOIs for PDB structures all have the same format:10.2210/pdbXXXX/pdb, where XXXX is the PDB ID. For example, the DOI for entry 4hhb is "10.2210/pdb4hhb/pdb". This links directly to the entry in the PDB file format on the FTP server.

The journal reference for the RCSB PDB is H.M. Berman, J. Westbrook, Z. Feng, G. Gilliland, T.N. Bhat, H. Weissig, I.N. Shindyalov, P.E. Bourne (2000) The Protein Data Bank. *Nucleic Acids Research* 28: 235-242. The RCSB PDB should also be referenced with the URL **www.pdb.org**.

The RCSB PDB is a member of the worldwide PDB (wwPDB): H.M. Berman, K. Henrick, H. Nakamura (2003) Announcing the worldwide Protein Data Bank. *Nature Structural Biology* 10: 980.

The *RCSB PDB Molecule of the Month* may be referenced using its DOI and the author/s of the article. The DOIs have the format: 10.2210/rcsb_pdb/mom_YYYY_MM, where YYYY is the year and MM the month (one or two digits). For example, the DOI for the May 2003 feature on hemoglobin by Shuchismita Dutta and David S. Goodsell is "10.2210/rcsb_pdb/mom_2003_5".

Joint Center for Structural Genomics (JCSG). Crystal structure of protein of unknown function (DUF1696) with pleckstrin-homology domain (ZP_00539245.1) from Exiguobacterium sibiricum *255-15 at 2.42 Å resolution. DOI: 10.2210/pdb3b77/pdb*

References

Except where noted, images were made using Chimera (**www.cgl.ucsf.edu/chimera**).[35]

[1]*Jmol: an open-source Java viewer for chemical structures in 3D.* www.jmol.org.

[2]*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.*

[3]*H.M. Berman, J. 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.*

[4]*H.M. Berman, K. Henrick, H. Nakamura (2003) Announcing the worldwide Protein Data Bank.* Nat Struct Biol. *10(12): 980.*

[5]*S. Dutta, K. Burkhardt, G.J. Swaminathan, T. Kosada, K. Henrick, H. Nakamura, H.M. Berman. Data deposition and annotation at the Worldwide Protein Data Bank in* Methods in Molecular Biology, *426: Structural Proteomics: High-Throughput Methods, B.G. Kobe, M. Guss, T. Huber, Editors. 2008, Humana Press: Totowa, NJ.*

[6]*P.M.D. Fitzgerald, J.D. Westbrook, P.E. Bourne, B. McMahon, K.D. Watenpaugh, H.M. Berman, 4.5 Macromolecular dictionary (mmCIF), in* International Tables for Crystallography, G. Definition and exchange of crystallographic data, *S.R. Hall, B. McMahon, Editors. 2005, Springer: Dordrecht, The Netherlands. p. 295-443.*

[7]*S.F. Altschul, W. Gish, W. Miller, E.W. Myers, and D.J. Lipman (1990) Basic local alignment search tool.* J. Mol. Biol. *215: 403-410.*

[8]*D.L. Wheeler, T. Barrett, D.A. Benson, S.H. Bryant, K. Canese, V. Chetvernin, D.M. Church, M. Dicuccio, R. Edgar, S. Federhen, M. Feolo, L.Y. Geer, W. Helmberg, Y. Kapustin, O. Khovayko, D. Landsman, D.J. Lipman, T.L. Madden, D.R. Maglott, V. Miller, J. Ostell, K.D. Pruitt, G.D. Schuler, M. Shumway, E. Sequeira, S.T. Sherry, K. Sirotkin, A. Souvorov, G. Starchenko, R.L. Tatusov, T.A. Tatusova, L. Wagner, E. Yaschenko (2008) Database resources of the National Center for Biotechnology Information.* Nucleic Acids Res. *36: D13-21.*

[9]*J. Westbrook, Z. Feng, K. Burkhardt, H.M. Berman (2003) Validation of protein structures for the PDB.* Meth Enz. *374: 370-385.*

[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 PDB.* Acta Crystallogr *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: W615-9.*

[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 Crystallogr *D55: 191-205.*

[14]*D. Weininger (1988) SMILES 1. Introduction and encoding rules.* J. Chem. Inf. Comput. Sci. *28: 31.*

[15]*© The International Union of Pure and Applied Chemistry, IUPAC International Chemical Identifier (InChI) 2005 (secretariat@iupac.org).*

[16]*D. Dimitropoulos, J. Ionides, K. Henrick (2007) Using MSDchem to search the PDB ligand dictionary.* Current Protocols in Bioinformatics. *2: Unit 14.3.*

[17]*Z. Feng, L. Chen, H. Maddula, O. Akcan, R. Oughtred, H.M. Berman, J. Westbrook (2004) Ligand Depot: A data warehouse for ligands bound to macromolecules.* Bioinformatics. *20: 2153-2155.*

[18] *Google Earth.* earth.google.com.

[19] *N. Deshpande, K.J. Addess, W.F. Bluhm, J.C. Merino-Ott, W. Townsend-Merino, Q. Zhang, C. Knezevich, L. Xie, L. Chen, Z. Feng, R. Kramer Green, J.L. Flippen-Anderson, J. Westbrook, H.M. Berman, P.E. Bourne (2005) The RCSB Protein Data Bank: a redesigned query system and relational database based on the mmCIF schema.* Nucleic Acids Res. *33: D233-D237.*

[20] *The Gene Ontology Consortium (2000) Gene Ontology: tool for the unification of biology.* Nature Genetics. *25: 25-29.*

[21] *Enzyme Classification. www.chem.qmw.ac.uk/iubmb/enzyme.*

[22] *L. Conte, A. Bart, T. Hubbard, S. Brenner, A. Murzin, C. Chothia (2000) SCOP: a structural classification of proteins database.* Nucleic Acids Res. *28: 257-259.*

[23] *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.*

[24] *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.*

[25] *S. Dutta, H.M. Berman, W.F. Bluhm (2007) Using the tools and resources of the RCSB Protein Data Bank.* Current Protocols in Bioinformatics. *20: 1.9.1-1.9.24.*

[26] *M. Andrec, D.A. Snyder, Z. Zhou, J. Young, G.T. Montelione, R.M. Levy (2007) A large data set comparison of protein structures determined by crystallography and NMR: statistical test for structural differences and the effect of crystal packing.* Proteins. *69(3): 449-65.*

[27] *J.L. Markley, E.L. Ulrich, H. Berman, K. Henrick, H. Nakamura, H. Akutsu (2008) BioMagResBank (BMRB) as a partner in the Worldwide Protein Data Bank (wwPDB): New policies affecting biomolecular NMR depositions.* J Biol NMR. *40: 153-155.*

[28] *K. Henrick, Z. Feng, W. Bluhm, D. Dimitropoulos, J.F. Doreleijers, S. Dutta, J.L. Flippen-Anderson, J. Ionides, C. Kamada, E. Krissinel, C.L. Lawson, J.L. Markley, H. Nakamura, R. Newman, Y. Shimizu, J. Swaminathan, S. Velankar, J. Ory, E.L. Ulrich, W. Vranken, J. Westbrook, R. Yamashita, H. Yang, J. Young, M. Yousufuddin, H. Berman (2008) Remediation of the PDB Archive.* Nucleic Acids Res. *36: D426-D433.*

[29] *H. Berman (2008) The Protein Data Bank: a historical perspective.* Acta Crystallogr *A: Foundations of Crystallography 64: 88-95.*

[30] *C. Zardecki (2008) Interesting structures: Education and outreach at the RCSB Protein Data Bank.* PLoS Biol. *6: e117.*

[31] *L. Chen, R. Oughtred, H.M. Berman, J. Westbrook (2004) TargetDB: A target registration database for structural genomics projects.* Bioinformatics. *20: 2860-2862.*

[32] *A. Kouranov, L. Xie, J. de la Cruz, L. Chen, J. Westbrook, P.E. Bourne, H.M. Berman (2006) The RCSB PDB information portal for structural genomics.* Nucleic Acids Res. *34: D302-D305.*

[33] *H. Berman (2008) Harnessing Knowledge from Structural Genomics.* Structure. *16: 16-18.*

[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: 1605-12.*

wwPDB Members

RCSB PDB
www.pdb.org

Protein Data Bank Europe
www.ebi.ac.uk/msd

Protein Data Bank Japan
www.pdbj.org

BioMagResBank
www.bmrb.wisc.edu

RCSB PDB Partners

Rutgers, The State University of New Jersey,
Department of Chemistry and Chemical Biology
610 Taylor 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 92093-0537

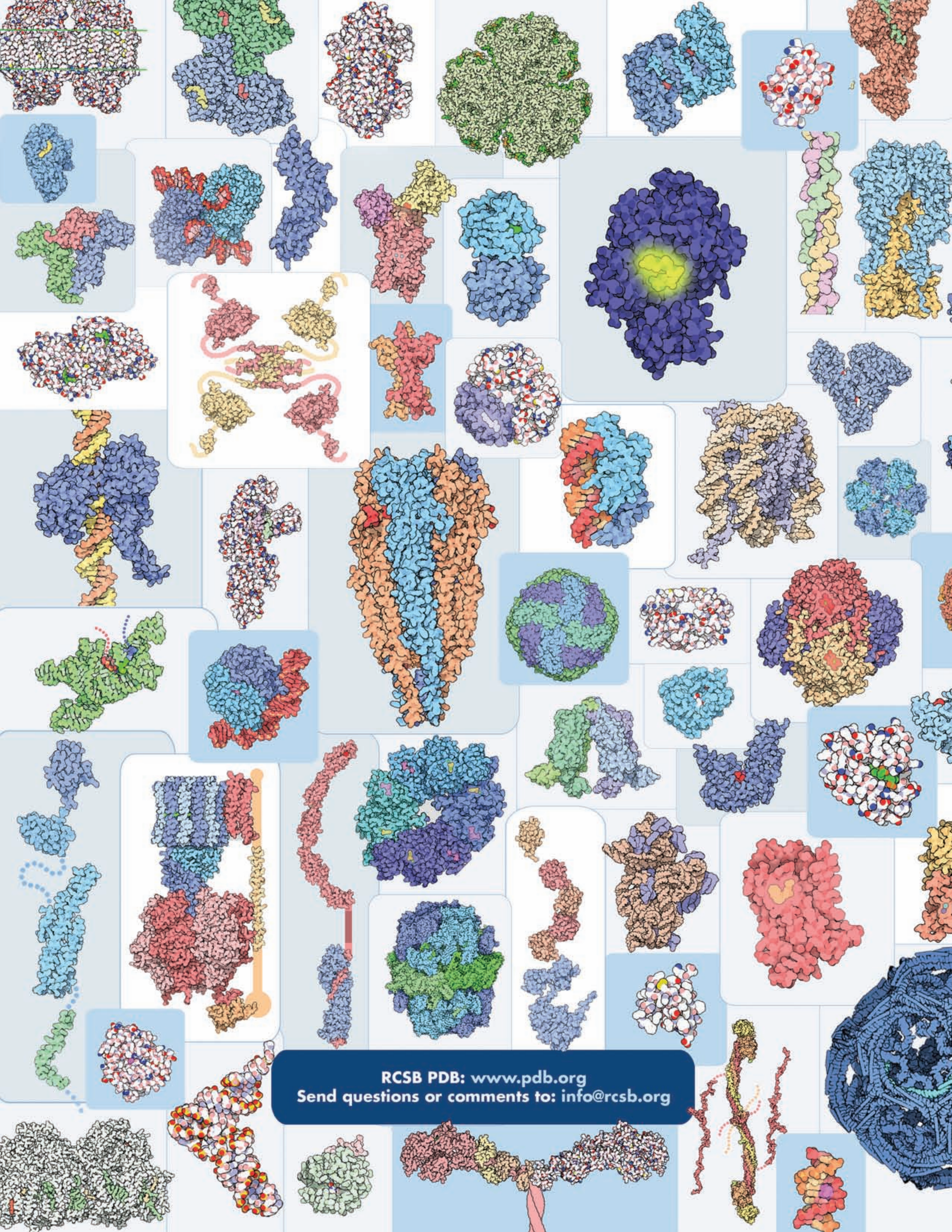
RCSB PDB Management

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**.



RCSB PDB: www.pdb.org
Send questions or comments to: info@rcsb.org