RCSB PDB
PROTEIN DATA BANK

ANNUAL REPORT
July 2009

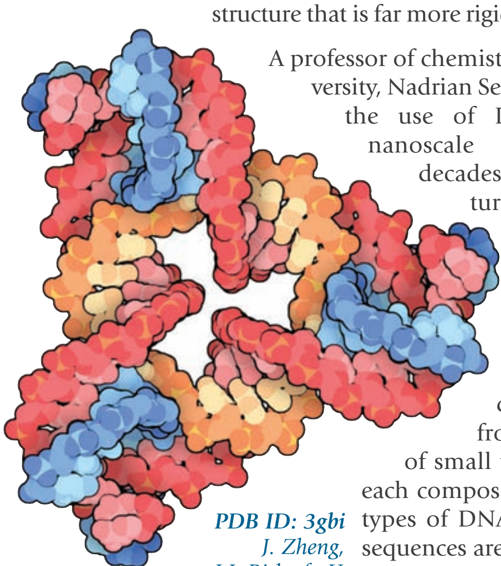
RESEARCH COLLABORATORY FOR STRUCTURAL BIOINFORMATICS
Rutgers, The State University of New Jersey
San Diego Supercomputer Center & Skaggs School of Pharmacy & Pharmaceutical Sciences, University of California, San Diego

About the Cover

Image and description from the Molecule of the Month *on* ***Designed DNA Crystal*** *doi: 10.2210/rcsb_pdb/mom_2009_11*.

DNA is a perfect raw material for constructing nanoscale structures. Since base-pairing has been selected by evolution to be highly specific, it is easy to design sequences that will link up with their proper mates. In this way, we can treat small pieces of DNA like Tinkertoys, designing individual components and then allowing them to assemble when we put them together. In addition, the chemistry of DNA synthesis has been completely automated, so custom pieces of DNA can be easily constructed, or even ordered from commercial biotechnology companies. This puts DNA nanotechnology in the hands of any modest laboratory, and many laboratories have taken advantage and created nanoscale scaffolds, tweezers, polyhedra, computers, and even tiny illustrations composed entirely of DNA.

DNA has the characteristic mix of flexibility and rigidity that is the hallmark of biological molecules. If the sequence of bases is correct, it zips up into a double helix that, at least in short lengths, is a sturdy cylinder. Longer stretches, however, start to show flexibility, and the DNA helix curves and bends. The trick in designing a DNA infrastructure is to develop ways to rigidify the overall structure. In most cases, this has been done by having the DNA strand weave back and forth between many parallel double helices. In this way, the bundle of helices forms a structure that is far more rigid than a single helix.



PDB ID: 3gbi *J. Zheng, J.J. Birktoft, Y. Chen, T. Wang, R. Sha, P.E. Constantinou, S.L. Ginell, C. Mao & N.C. Seeman. (2009) From molecular to macroscopic via the rational design of a self-assembled 3D DNA crystal.* Nature ***461****: 74-77.*

A professor of chemistry at New York University, Nadrian Seeman has pioneered the use of DNA for building nanoscale structures. After decades of work, the structure deposited to the PDB by Seeman's laboratory and highlighted on the cover is the first crystal structure of a DNA lattice completely designed from scratch. It is built of small triangular subunits, each composed of three separate types of DNA strands. The base sequences are carefully chosen so that they assemble into this one particular structure, and not any others. At the corners of the triangle, there are sticky ends that link to other triangles, stacking up in a predicable way into a three-dimensional scaffold.

These scaffolds are the first step towards the larger goals of nanotechnology. Two major applications of the scaffold have been proposed. First, the spaces in the lattice may be used to host other molecules, like proteins, providing a general way to orient many copies of the particular molecule. This would allow the determination of the structure by X-ray crystallography without the need to crystallize the protein by itself, which is always a hit-or-miss proposition. The second application is for building nanoscale electrical devices. In these speculative structures, individual components, such as electron transfer proteins, would be placed at the nodes in the scaffold. Since the base sequences of the DNA lattice are known, special connectors could be designed to attach the components in exactly the right place.

Table of Contents

Message from the Director

Welcome to a milestone edition of the RCSB Protein Data Bank's Annual Report. Not only is this our tenth edition, but the RCSB PDB's management of this resource was renewed for the second time during the period covered by this publication.

These bulletins were established in 2000 to provide a yearly snapshot of RCSB PDB activities and the state of the PDB archive. The span of this document–July through June–mirrors the period covered in reports to funding agencies. The annual report, one of several RCSB PDB publications, is distributed worldwide to researchers in biology, biochemistry, genetics, pharmacology, biophysics, and bioinformatics; computer scientists and software developers; and students and educators of all levels.

Our activities, as described in this report, follow and support the structure determination pipeline–from the deposition of data by scientists to the use of that data by researchers and students. These annual reports also offer historical and background information to help introduce new users to **www.pdb.org** and the world of biological macromolecules that the RCSB PDB hosts.

Just as the number of structures available in the PDB archive has increased over the past ten years, so has the number of services available from the RCSB PDB. Features such as MyPDB, which emails users when structures of interest are released, and our expansive outreach and education program are aimed to keep the RCSB PDB in the hands of its users.

We hope that you enjoy our 2009 report, and look forward to continued development of the resource.

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

About the RCSB PDB

Snapshot: July 1, 2009

58,588 released atomic coordinate entries

Molecule Type

54,141	proteins, peptides, and viruses
2,033	nucleic acids
2,381	protein/nucleic acid complexes
33	other

Experimental Technique

50,284	X-ray
7,914	NMR
243	electron microscopy
17	hybrid
130	other
39,459	structure factor files
5,200	NMR restraint files

The RCSB Protein Data Bank

The RCSB Protein Data Bank, administered by the nonprofit Research Collaboratory for Structural Bioinformatics (RCSB), supports scientific research and education worldwide by providing an essential resource of information about biomolecular structures. These molecules of life are found in all organisms, from bacteria and plants to animals and humans.

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. Scientists from around the world determine the shapes of these structures, and pass on data about their experiments to the PDB. This information is then made freely available in the PDB archive.

The RCSB PDB manages the PDB archive as a member of the Worldwide PDB (wwPDB).

The Worldwide Protein Data Bank

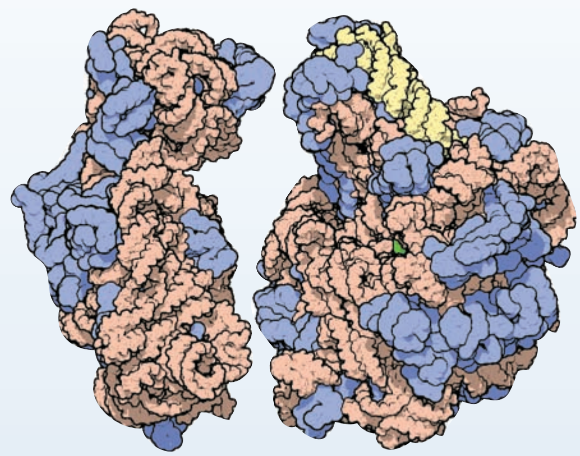
The wwPDB organization (**www.wwpdb.org**) was formed to ensure that the PDB archive is and will be freely and publicly available to the global community. The wwPDB members (RCSB PDB, Protein Data Bank in Europe (PDBe), Protein Data Bank Japan (PDBj), and the BioMagResBank (BMRB)) host deposition, processing, and distribution centers for PDB data and collaborate on a variety of projects and outreach efforts.

As 'archive keeper' for the wwPDB, the RCSB PDB maintains the central repository of the PDB archive.

The protein and nucleic acid structures in the PDB archive are vital resources for the study of science and medicine worldwide.

Biomolecules, the main building blocks of living organisms, come in a variety of shapes. They can be tiny globular proteins like myoglobin, or resemble a twisted ladder, like DNA. The look of collagen echoes that of a rope, while icosahedral viruses can look like soccer balls. Large molecular machines, such as the protein-building ribosome, have very complex architecture.

Each of these shapes enables biological molecules to perform specific tasks that are essential to life. By studying these structures, researchers can understand the functions of biomolecules in human health and disease. Scientists can also use this information to develop small molecule drugs that can assist or stop certain protein functions from happening.



*Three structural biologists won the 2009 Nobel Prize in Chemistry for studies of the structure and function of the ribosome–Venkatraman Ramakrishnan, Thomas A. Steitz, and Ada E. Yonath. The PDB depositions of their first complete ribosome subunit structures (**1fjg**, **1ffk**, and **1fka**) almost a decade ago ushered structural biology into a new era. Since that time, hundreds of ribosome structures consisting of 50S, 30S subunits and complete 70S ribosomes have been contributed by these Nobel scientists and others. These structures, complexed with and without antibiotics, tRNAs, mRNAs, initiation factors, and release factors, provide a basis for understanding how the ribosome works and are useful tools for drug development.*
Image from the **Molecule of the Month**.

Milestones

1970s

Community discusses how to establish a protein structure archive

PDB archive is established at Brookhaven National Laboratory in October 1971 with 7 structures

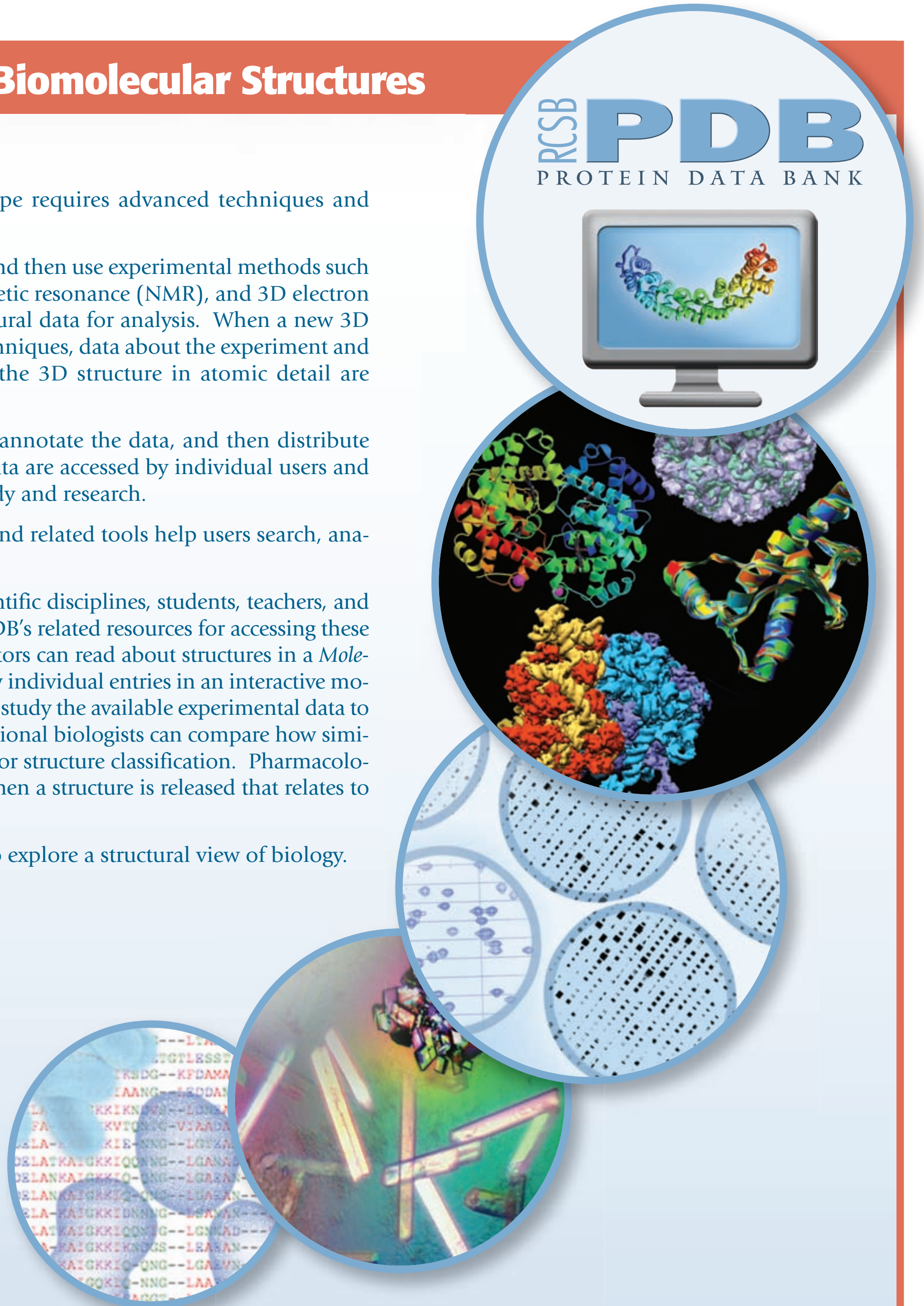
1980s

Number of structures increases as technology improves

Community discusses requiring data deposition to PDB archive

Deposition guidelines are established by the International Union of Crystallography

The PDB Archive of Biomolecular Structures



Solving the puzzle of a protein's shape requires advanced techniques and careful analysis.

Scientists prepare a protein for study, and then use experimental methods such as X-ray crystallography, nuclear magnetic resonance (NMR), and 3D electron microscopy (3D EM) to acquire structural data for analysis. When a new 3D structure is determined using these techniques, data about the experiment and the *x, y, z* coordinates that describe the 3D structure in atomic detail are deposited in the PDB archive.

wwPDB members review, curate, and annotate the data, and then distribute the data online to the world. These data are accessed by individual users and by countless external resources for study and research.

The RCSB PDB's searchable database and related tools help users search, analyze and visualize PDB structural data.

Researchers from a wide variety of scientific disciplines, students, teachers, and the general public also use the RCSB PDB's related resources for accessing these data. For example, students and educators can read about structures in a *Molecule of the Month* column, and then view individual entries in an interactive molecular viewer. Structural biologists can study the available experimental data to recreate similar experiments. Computational biologists can compare how similar different structures are by sequence or structure classification. Pharmacologists can receive MyPDB email alerts when a structure is released that relates to drugs under development.

All of these tools and resources help to explore a structural view of biology.

Many steps are involved when researchers experimentally determine the 3D shape of a protein or nucleic acid. After the data are made available in the PDB archive, they are used worldwide in research and education.

1990s

Structural genomics begins

PDB management moves to the RCSB PDB

2000s

wwPDB is formed

50,000th structure is released

2nd renewal of **RCSB PDB management**

About the RCSB PDB

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 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.

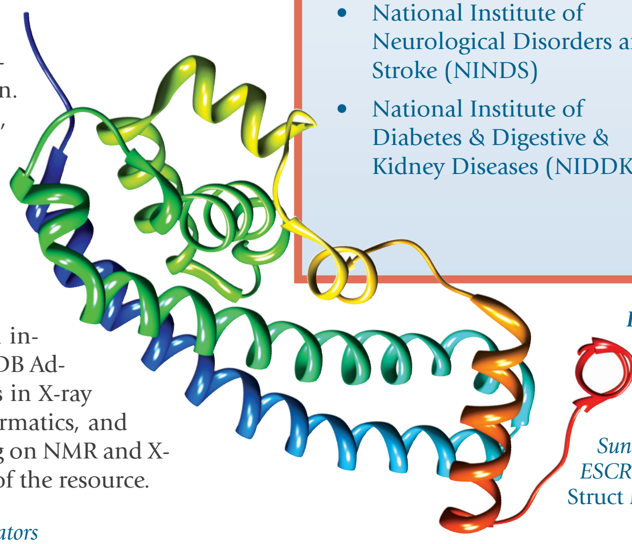
The RCSB PDB Team is comprised of experts in diverse fields of computer science, biology, chemistry, and education. In addition to working with PDB data, RCSB PDB members co-author scientific papers, exhibit at meetings, present posters and papers, and attend and organize workshops. Staff members also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

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).



PDB ID: 3frr
M. Bajorek, H.L. Schubert, J. McCullough, C. Langelier, D.M. Eckert, W.M. Stubblefield, N.T. Uter, D.G. Myszka, C.P. Hill & W.I. Sundquist. (2009) Structural basis for ESCRT-III protein autoinhibition. Nat Struct Mol Biol ***16****: 754-762.*

The RCSB PDB with some wwPDB Collaborators

Data Input

PDB depositors by location (1999–2009). Image created by Google Earth.[1]

Before Structures are Released: Data Deposition, Validation, and Annotation

A key function of the wwPDB is the efficient capture (deposition) and curation (validation and annotation) of experimental structural data.[2] 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, it is immediately assigned its own unique PDB ID. Using locally developed tools, the RCSB PDB carefully curates and annotates entries using an integrated system that is based on the use of a standard data dictionary.[3] The goal is to ensure the PDB entry accurately represents the structure and experiment.

Today, the wwPDB receives approximately 20 new structures daily. Annotation-curators then check each deposition for errors or omissions, ensuring a consistent format and accurate data for each entry. Annotators compare the sequence and citation reported in the deposition to external databases.[4; 5] 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[6] and the completed coordinate file 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. On average, structures are processed, reviewed by the author, and finalized for release in two weeks.

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 describing the structure has been published (HPUB). Structures can be on HOLD or HPUB for no longer than one year past the date of deposition.

The Chemical Component Dictionary

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. Each chemical definition describes properties such as stereochemical assignments, aromatic bond assignments, idealized coordinates, chemical descriptors (SMILES & InChI),[7; 8] and systematic chemical names. This dictionary is used and maintained as part of the data annotation process. Special care is taken to standardize the chemistry and nomenclature in this dictionary.

Users can search and browse the Chemical Component Dictionary using resources such as MSDchem[9] and Ligand Expo.[10]

Developed by the RCSB PDB, Ligand Expo integrates databases, services, tools, and methods related to small molecules. It allows users to search or browse for components; review the chemistry, geometry, atom nomenclature, and more; download model and ideal chemical component coordinates; and view all instances of a component in released PDB entries.

Large Assemblies and the EMDataBank

The PDB archives many large biological assemblies determined by X-ray crystallographic and EM methods. Examples include viruses (capsid and helical), ribosomes, and chaperones. Due to their size, curating these structures can involve complex symmetry operations and experimental data.

Cryo-electron microscopy (cryoEM) is a maturing methodology in structural biology that enables the determination of these large macromolecular complexes. These experiments provide information that bridges the gap between cell biology and crystallography/NMR. In addition to density maps and associated metadata, cryoEM experiments produce 3D coordinates. Through an NIH/NIGMS-funded collaboration, we have created a joint EM map and model deposition tool in which metadata provided in the map deposition step is automatically copied to the coordinate deposition session. This global deposition and retrieval network for cryoEM map, model, and associated metadata, is available alongside a portal for software tools for standardized map format conversion, map segmentation and model assessment, visualization, and data integration at **EMDataBank.org**. This work has been done in collaboration with PDBe at the European Bioinformatics Institute and the National Center for Macromolecular Imaging at Baylor College of Medicine.

Validation

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. Software used by the RCSB PDB (and made available to depositors as described below) checks the sequence and file format consistency, compares geometrical and chemical interactions to various standards, and reports any errors found. The software includes this information in a report that also compiles output from the established programs Molprobity,[11] Nucheck,[6] Procheck,[12] and SFCheck.[13] These reports also provide a way of assessing the "quality" of a structure.

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, web-servers, and documentation are available at **deposit.pdb.org**.

- **pdb_extract**[14] automatically takes key details from the output files produced by many NMR and X-ray crystallographic applications and puts them in a deposition-ready format.

- **SF-Tool** validates model coordinates against structure factor data, translates structure factors into different formats, and checks for twinned data.

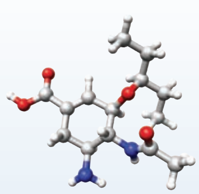
- The **Validation Server** checks the format of the coordinate file and validates the overall structure before deposition.[6] It produces a validation report containing geometrical and experimental checks from several programs[6; 11-13] and identifies any sequence and data inconsistencies. Researchers are encouraged to use this program to review the quality of any released structure before using it in their own study.

- **ADIT**[15] is used by the RCSB PDB and PDBj for data deposition and annotation. Depositors upload their data files and use the ADIT editor to add information and check the completeness of an entry. Annotators use a special view of ADIT to run checks and annotate the entry in preparation for release. **ADIT-NMR**, a similar program hosted at BMRB and PDBj, is a single tool for NMR structural and experimental data deposition. 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.

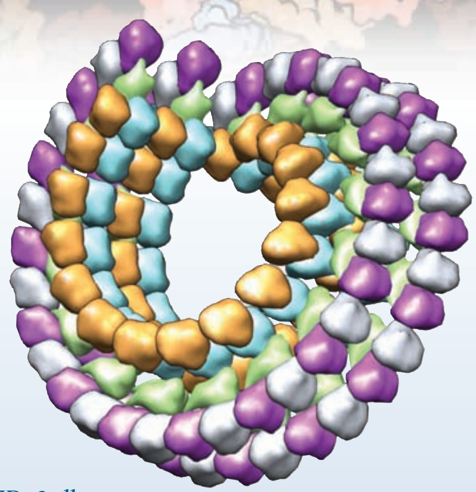
- Depositors can search **Ligand Expo** for matches to ligands already in the archive. Ligand Expo can also be used to build components new to the PDB.

Representing Molecules Small and Large

Representing different types of structures presents unique challenges. As new experimental techniques emerge, data validation methods and accurate descriptions are needed to provide a consistent and reliable archive. Similarly, special efforts are required to ensure that information about all biomolecules, from small molecules (such as drugs that bind to proteins) to large assemblies (such as viruses), is available to the PDB user community.



Ligand Expo* (ligand-expo.rcsb.org)** *is a tool for searching the Chemical Component Dictionary for small molecule components found in the PDB such as* ***oseltamivir *(Tamiflu; shown), finding the PDB structures that contain particular small molecules, building new chemical definitions, and more.*



PDB ID: 3edl
D. Tan, W.J. Rice & H. Sosa. (2008) Structure of the kinesin13-microtubule ring complex. Structure ***16****: 1732-1739.*

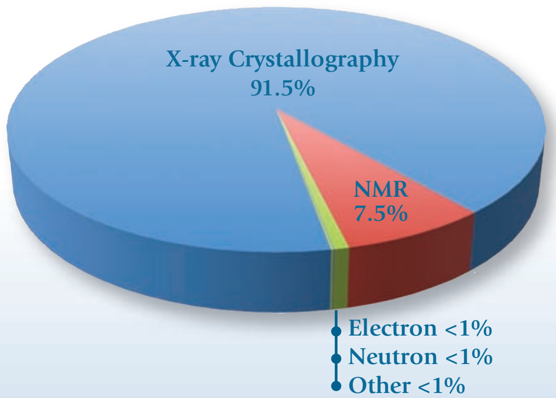
Statistics (July 1, 2008 - June 30, 2009)

During the period of this report, 7848 structures were deposited to the PDB archive and prepared for release by the wwPDB.

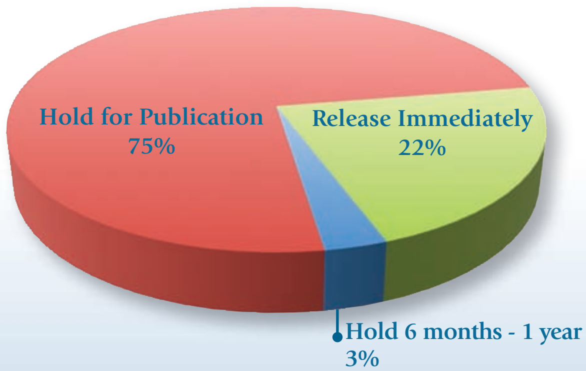
Sequences for 55% of these depositions were released before the coordinate files to help prevent unintended duplication of effort in structure determination and to promote blind testing of structure prediction and modeling techniques.

A look at the demographics of all of these structures is below.

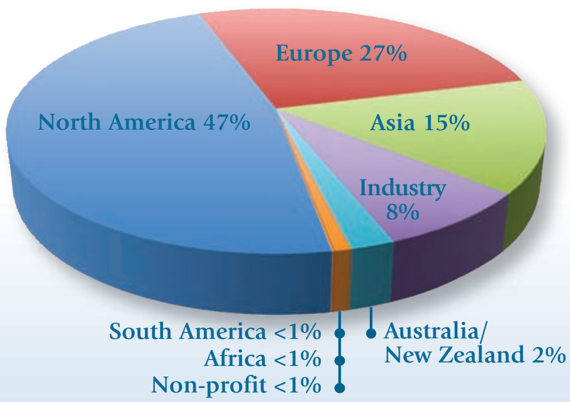
Experiment Type



Author-Defined Release Status



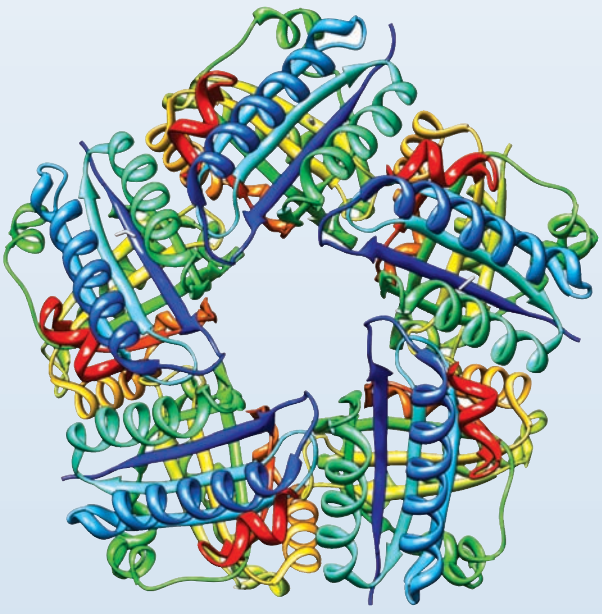
Deposition Demographics



wwPDB: Collaboration for a Consistent Archive

A major focus of the wwPDB is maintaining consistency and accuracy across the archive. As the PDB grows, new structures and new technologies can challenge how all structures are represented. To help face these challenges, the wwPDB initiated a project to carefully review how data were represented in the **PDB file format**, which is the default data view for many users. The definition of every data item found in a PDB formatted file was checked for scientific accuracy. The resulting PDB Format Version 3.2 Guide (**www.wwpdb.org/docs.html**) was released in September 2008 with new remarks and records that standardized experimental methods and structure model descriptions in an entry. The corresponding data dictionaries (for mmCIF and XML file formats) were also updated.

Following this evaluation, the entire archive was revised to restore and update all relevant records and remarks in the datafiles. For example, the records describing biological assembly information and the potentially important sites in an entry were either computed by current software or captured from relevant resources. Various identifiers were added as cross-references to several databases, such as taxonomy and PubMed identifiers. Miscellaneous issues were checked and corrected–such as microheterogeneity, chromophores, or inclusion of amino and/or acetyl groups in the polymer sequence. Data collection and refinement details were checked and revised where appropriate. The tools and resources created to standardize these various issues have been incorporated in ongoing data annotation efforts.



PDB ID: 3dtz
Midwest Center for Structural Genomics (MCSG)
Crystal structure of putative chlorite dismutase TA0507.
DOI: 10.2210/pdb3dtz/pdb.

Data Access, Query, and Reporting

Data Distribution and Access

RCSB PDB services and data are freely available online.

As the wwPDB archive keeper, the RCSB PDB is responsible for maintaining the PDB archive at **ftp://ftp.wwpdb.org**. This site is a primary source of data for all other sites worldwide. During this report period, 7237 coordinate files (in PDB, mmCIF, and PDBML/XML file formats), 6566 structure factor files, and 1153 constraint data files were added to the archive.

The RCSB PDB website at **www.pdb.org** also provides access to these data files. It is accessed by about 150,000 unique visitors per month from more than 130 different countries. On a typical weekday, three to four pages from the website are viewed every second. Around 800 gigabytes of data are transferred each month. At the same time, about 6,000 unique visitors download more than 10 million files from the FTP site at **ftp://ftp.wwpdb.org**, for a total of about 2 terabytes of data.

Web Services have been implemented to help the software developer community build tools that interact more effectively with PDB data. Instead of storing coordinate files and related data locally, these Web Services provide a way for software tools to remotely search the RCSB PDB database and retrieve data for any given structure.

Data Query and Reporting

The website provides access to a relational database[22; 23] that integrates PDB data with related information from external sources (such as journal abstracts, functional descriptors, sequence annotations, structure annotations, and taxonomy). Users can search for structures using simple searches (PDB ID, keyword, sequence, or author) or by building more complex queries with the Advanced Search feature.

PDB structures can also be browsed using tree-like hierarchies based on GO terms, Enzyme Classification,[24] Medical Subject Headings, Source Organism, Genome Location,[25] SCOP, and CATH classifications.

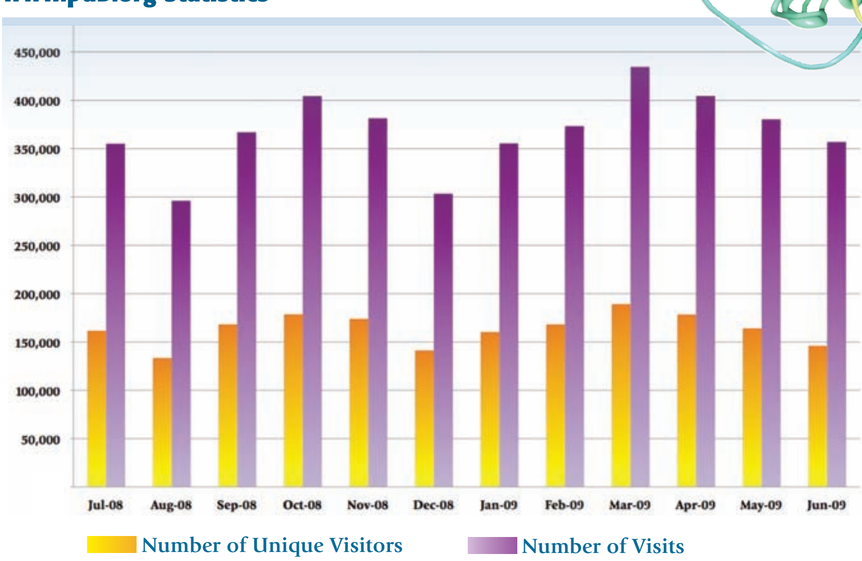
All search result sets can be further refined or explored. Individual structures can be examined at any point. Search results can be presented as a list of structures; a PubMed-like list of the primary citations for the structures; a list of ligands known to interact with the structures; a list of any 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.

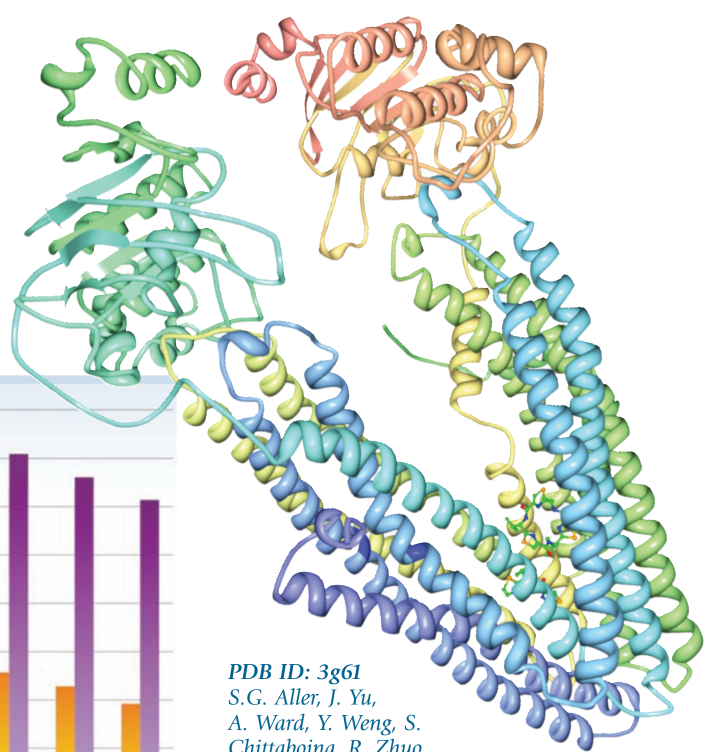
RCSB PDB website features are supported by help pages, tutorials, and an active help desk. Additional enhancements and new resources are continually in development.

Exploring Structures

The RCSB PDB website hosts a Structure Summary page for each individual entry that provides an overview of the structure; derived data from CATH, SCOP, Pfam, and Gene Ontology (GO);[16-19] tools to examine the sequence, sequence domains, and sequence similarity; detailed information relating to the entry's citation, biology and chemistry, experiment, and geometry; and links to related resources. Several molecular viewers, including Jmol[20] and the RCSB PDB's Protein Workshop,[21] are available to interactively view the molecule.

www.pdb.org Statistics





PDB ID: 3g61
S.G. Aller, J. Yu, A. Ward, Y. Weng, S. Chittaboina, R. Zhuo, P.M. Harrell, Y. T. Trinh, Q. Zhang, I.L. Urbatsch & G. Chang. (2009) Structure of P-glycoprotein reveals a molecular basis for poly-specific drug binding. Science *323: 1718-1722.*
Image created using Protein Workshop.

New Features

www.pdb.org, which also provides a variety of services and features, was recently redesigned to improve the accessibility and navigation of the site, and to allow users to customize the site. Several new features have been added, including:

MyPDB:

Keep up-to-date with new structures... Automatically!

The MyPDB service notifies users via email when the PDB releases structures that match customized queries.

With MyPDB, users can:

- Combine and save keyword, sequence, ligand, and other searches
- Run searches stored in MyPDB at any time
- Receive email alerts weekly or monthly
- Access structures directly from the email alerts for further exploration
- Utilize the RCSB PDB web interface to refine saved searches

To set up a MyPDB search, click on **MyPDB Login** from the top of **www.pdb.org**.

Literature View:

Looking at structures in PubMedCentral

The Literature View for any structure in the PDB illustrates how a given structure has been analyzed and presented in open access publications.

Developed in collaboration with the BioLit project (**biolit.ucsd.edu**), the Literature View lets users:

- Read the primary citation's abstract
- Search for structures with PubMed abstracts containing the same keywords
- Learn about other publications included in the PDB entry
- Explore the articles found in PubMedCentral that contain the entry's PDB ID, even those without a reference to the entry's primary citation. Links to the abstract and copyright information, along with figures and related legends from these open access articles, are displayed
- View any additional PDB entries cited in the related articles

To access these features, select the **Literature** tab for any structure.

Sequence Similarity View

Interested in finding homologous protein structures, or finding a non-redundant set of proteins? The new Sequence Similarity View is an easy way to get this information based upon a protein chain in a PDB entry.

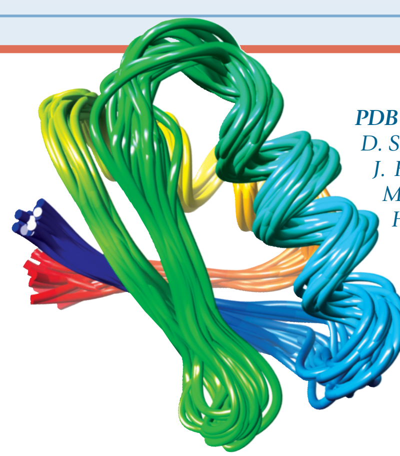
This view:

- Provides clusters of structures at different levels of sequence similarity based on a weekly BLAST[3] analysis of all proteins with more than 20 amino acids in the PDB
- Offers details about the structures of all homologous proteins found within a cluster
- Defines sequence similarity on a chain-by-chain basis, with the results returned for the entire structure

To see proteins that are from 30% to 100% identical with the alpha and beta chains of hemoglobin, explore the **Sequence Similarity View** for PDB entry 4hhb.

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.wwpdb.org** to provide readily identifiable data sets for research on the PDB archive. Scripts are available to help users create local copies of all or part of the PDB archive or snapshots.



PDB ID: 2rog
D. Sakakibara, A. Sasaki, T. Ikeya, J. Hamatsu, T. Hanashima, M. Mishima, M. Yoshimasu, N. Hayashi, T. Mikawa, M. Walchli, B.O. Smith, M. Shirakawa, P. Guntert & Y. Ito. (2009) Protein structure determination in living cells by in-cell NMR spectroscopy. Nature *458: 102-105.*

Educational Initiatives

One of the goals of the RCSB PDB is to make macromolecular structures accessible and comprehensible to all of our users–students and teachers, biologists, structural biologists, computational biologists, and the general public. Educational programs and resources are aimed at helping teachers and students at all levels understand PDB data and structures, while outreach efforts work towards enabling our users to take advantage of all RCSB PDB services. Outreach also involves getting feedback from users in order to develop the resource further.

While most RCSB PDB programs are aimed at scientific researchers, our efforts also promote scientific literacy and a broader understanding of structural biology.

Educational Resources and Programs

At the San Diego Science Festival, building virus structures provided a hands-on introduction to symmetry, molecular structures, and the PDB.

The RCSB PDB participated in two recent large-scale events aimed at general audiences in local communities. At the San Diego Science Festival and Rutgers Day, staff built virus models and discussed protein structures with thousands of event attendees.

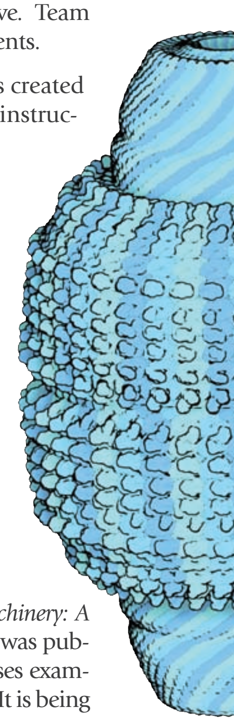
Other programs have focused on getting students and teachers at the middle and high school level interested and involved in the fields of structural biology and bioinformatics. Area high school classes also visit and tour the RCSB PDB. These experiences often lead to new initiatives and collaborations. The RCSB PDB also exhibits at education-related meetings to talk to teachers about incorporating protein structure and function in the classroom.

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. Team members also participate with many on-campus organizations and events.

The **Educational Resources** web page archives many of the materials created by the RCSB PDB for education at all levels, including animations, instructional handouts, and much more.

Online Series: *Molecule of the Month*

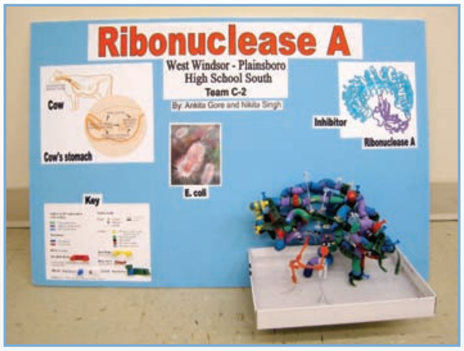
Created by David S. Goodsell (RCSB PDB, The Scripps Research Institute), the *Molecule of the Month* provides an easy introduction to macromolecular structures. This feature is used in many classrooms to introduce structures to students, and is an integral part of the Science Olympiad protein modeling 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. Accessible from the RCSB PDB homepage, *Molecule of the Month* descriptions also provide links to example structures in the RCSB PDB and offer interactive Jmol[20] views for further exploration.



How Do Drugs Work?

Inspired by the success and enduring popularity of 2002's *Molecular Machinery: A Tour of the Protein Data Bank*, a new poster entitled *How Do Drugs Work?* was published. In a style similar to the *Molecule of the Month* series, the poster uses examples from the PDB archive to describe different drug-protein interactions. It is being distributed at scientific and educational meetings.

The Protein Modeling Trial Event at New Jersey Science Olympiad



Before the competition, students research and build a structure by using the Molecule of the Month *and other resources.*

At the event, teams must quickly build an additional protein model.

RCSB PDB annotators then score the models with close attention given to the functionally important parts of the protein.

In 2009, teams from all over the Garden State became experts on ribonuclease for Science Olympiad tournaments held at the regional and state level. In 2010, teams will model proteins related to H1N1 influenza. This event is managed across the country by the Center for BioMolecular Modeling (**cbm.msoe.edu**). For information about the New Jersey event, see **education.pdb.org**.

User 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. From assembling formal task forces to speaking with users at professional society meetings, feedback is always collected to further RCSB PDB development.

A variety of online resources and outreach programs are developed to support many different RCSB PDB user communities.

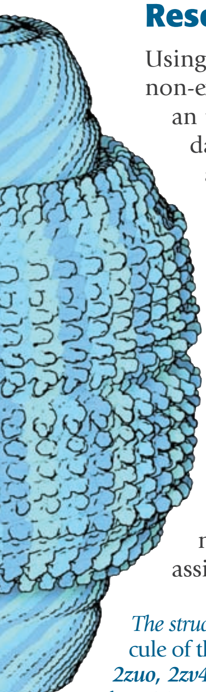
Exhibits and workshops at professional society meetings provide opportunities to meet with users, present recent work, and distribute related materials and educational resources. In addition, at the Congress and General Assembly of the International Union of Crystallography (shown) and other selected meetings, the RCSB PDB poster prize is awarded for the best student poster presentation.

Looking at Structures: An Online Resource for Understanding PDB Data

Using the data in the PDB archive can be challenging to the non-expert. To be able to analyze the data, users need to have an understanding of the experiments used to generate the data, and how these data are validated. Using text, images, and interactive Jmol views, ***Looking at Structures*** helps researchers and educators get the most out of the PDB archive. Broad topics include how to read PDB data files, how to visualize structures, how to read coordinate files, and potential challenges in exploring the archive. ***Looking at Structures*** is available from the left-hand menu of **www.pdb.org.**

Online Help

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 lead to new website enhancements and features. Other help desks assist in depositor-specific queries.



The structure of rat liver vault as illustrated in the June 2009 edition of Molecule of the Month *(dx.doi.org/10.2210/rcsb_pdb/mom_2009_06).* ***PDB IDs 2zuo, 2zv4, and 2zv5:*** *H. Tanaka, K. Kato, E. Yamashita, T. Sumizawa, Y. Zhou, M. Yao, K. Iwasaki, M. Yoshimura & T. Tsukihara. (2009) The structure of rat liver vault at 3.5 Ångstrom resolution.* Science ***323****: 384-388.*

Different Ways of Looking at Structure

At UCSD, students and scientists engage with the **ImmersivePDB** virtual reality software for protein and nucleic acid structures at the Calit2 visualization laboratory. This program lets users see these structures in a very unique way. Wearing stereo glasses, viewers can move through and around any PDB structure that is projected in the laboratory's special environment.

Interactive displays have been created to showcase structures at museums and at kiosks at RCSB PDB member sites. The same software can be used to highlight structures from a lab, institution, or class on your own computer by using the *Molecules in Motion Kiosk Viewer*. Using a list of PDB IDs, this full-screen animation program will display any PDB structure from different angles and perspectives. It also focuses on any chemical components within the structure. The Java viewer can be downloaded or launched from the **Educational Resources** page.

Short Course: Crystallography for Modelers

A short course called *Crystallography for Modelers* was held May 7-8 at Rutgers to help students learn how to understand PDB data in order to get the most out of it and avoid mistakes in research. The course was aimed at the practicing pharmaceutical/biophysical modeler to provide a deeper understanding of crystal structures and PDB files, their derivation, reliability, interpretation, and proper application and use as well as the practical aspects (with some theory) of electron density calculation, interpretation, use, and refinement approaches. 29 students from academia and industry attended the two-day course to hear lectures from RCSB PDB staff and participate in software demonstrations from industrial participants.

Publications

www.pdb.org is updated weekly with news, recent developments, new resources, and improvements to existing resources. Educational features, such as new *Molecule of the Month* installments, and help-related pages and tutorials are posted regularly. Statistical views of the archive are presented as content distributions and content growth.

Published quarterly, the *RCSB PDB Newsletter* describes and highlights recent 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 covering a diverse array of subjects. Recent publications have described the wwPDB collaboration,[26; 27] data deposition and annotation,[2; 27] data file formats,[28] and representing viruses in the data archive.[29]

Related Resources

Structural Genomics and the PSI Structural Genomics Knowledgebase

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.

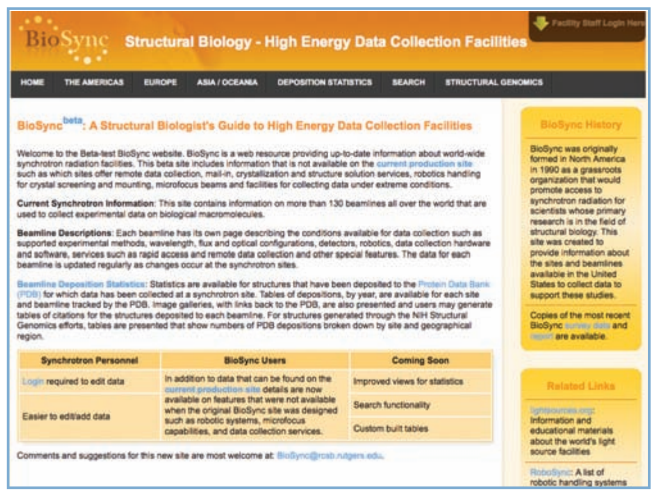
Current target progress and information is tracked by specific databases. The **Target Registration Database** (TargetDB; **targetdb.pdb.org**)[30] provides status information for targets selected for structure determination by various structural genomics centers. The **Protein Expression Purification and Crystallization Database** (PepcDB; **pepcdb.pdb.org**)[31] extends the content of TargetDB with status history, stop conditions, reusable text protocols, and contact information collected from the Protein Structure Initiative (PSI) efforts. 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 Structural Genomics Knowledgebase (PSI SGKB; **kb.psi-structuralgenomics.org**) funded by the NIGMS.[32] The PSI SGKB publicizes research advances in structural genomics and structural biology catalyzed by the PSI and beyond so that these structural, functional, and methodological advances are readily available for all biomedical scientists.

BioSync

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

BioSync has been redesigned to incorporate the dramatic changes in data collection capabilities at synchrotron beamlines (including remote data collection, mail-in, crystallization and structure solution services, robotics handling for crystal screening and mounting, microfocus beams and facilities for collecting data under extreme conditions). With its new look and feel, users can more easily find information about particular beamlines and to search for capabilities, services and equipment. These new capabilities are in beta testing at **biosync-beta.rutgers.edu**.



The redesigned BioSync resource

The structure of the TM1585 gene product of the bacterium Thermotoga maritima, *solved by the Joint Center for Structural Genomics, gives the first look at a novel enzyme in sugar metabolism and provides a new protein fold as a bonus. For a full description, see the Featured PSI Structures issue describing* **Glycerate Kinase** *at the PSI Structural Genomics Knowledgebase (doi:10.3942/psi_sgkb/fm_2008_3).*

PDB ID: 2b8n

R. Schwarzenbacher, D. McMullan, S.S. Krishna, Q. Xu, M.D. Miller, J.M. Canaves, M.A. Elsliger, R. Floyd, S.K. Grzechnik, L. Jaroszewski, H.E. Klock, E. Koesema, J.S. Kovarik, A. Kreusch, P. Kuhn, T.M. McPhillips, A.T. Morse, K. Quijano, G. Spraggon, R.C. Stevens, H. van den Bedem, G. Wolf, K. O. Hodgson, J. Wooley, A. M. Deacon, A. Godzik, S.A. Lesley & I.A. Wilson. (2006) Crystal structure of a glycerate kinase (TM1585) from Thermotoga maritima *at 2.70 Å resolution reveals a new fold.* Proteins **65**: *243-248. Image created using the Python Molecular Viewer (**mgltools.scripps.edu**).*[33]

References

Except where noted, molecular graphics images were produced using the UCSF Chimera[35] package from the Resource for Biocomputing, Visualization, and Informatics at the University of California, San Francisco (supported by NIH P41 RR-01081).

1. Google Earth. **earth.google.com**

2. S. Dutta, K. Burkhardt, J. Young, G.J. Swaminathan, T. Matsuura, K. Henrick, H. Nakamura & H.M. Berman. (2009) Data deposition and annotation at the worldwide Protein Data Bank. *Mol Biotechnol* **42**: 1-13.

3. 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.

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

5. 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.

6. J. Westbrook, Z. Feng, K. Burkhardt & H.M. Berman. (2003) Validation of protein structures for the Protein Data Bank. *Meth Enz* **374**: 370-385.

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

8. ©The International Union of Pure and Applied Chemistry. (2005) IUPAC International Chemical Identifier (InChI; contact: **secretariat@iupac.org**).

9. D. Dimitropoulos, J. Ionides & K. Henrick. (2007) Using MSDchem to search the PDB ligand dictionary. *Curr Prot Bioinformatics* **2**: Unit 14.13.

10. 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.

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. 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.

15. 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.

16. 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.

17. 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.

18. R.D. Finn, J. Tate, J. Mistry, P.C. Coggill, S.J. Sammut, H.R. Hotz, G. Ceric, K. Forslund, S.R. Eddy, E.L. Sonnhammer & A. Bateman. (2008) The Pfam protein families database. *Nucleic Acids Res* **36**: D281-288.

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

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

21. 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.

22. 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.

23. S. Dutta, H.M. Berman & W.F. Bluhm. (2007) Using the tools and resources of the RCSB Protein Data Bank. *Curr Prot Bioinformatics* **20**: 1.9.1-1.9.24.

24. Enzyme Nomenclature, Enzyme Classification. **www.chem.qmw.ac.uk/iubmb/enzyme**.

25. E.W. Sayers, 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, D. Landsman, D.J. Lipman, T.L. Madden, D.R. Maglott, V. Miller, I. Mizrachi, J. Ostell, K.D. Pruitt, G.D. Schuler, E. Sequeira, S.T. Sherry, M. Shumway, K. Sirotkin, A. Souvorov, G. Starchenko, T.A. Tatusova, L. Wagner, E. Yaschenko & J. Ye. (2009) Database resources of the National Center for Biotechnology Information. *Nucleic Acids Res* **37**: D5-15.

26. H.M. Berman, K. Henrick, H. Nakamura & J. L. Markley. (2009) Chapter 11 The Worldwide Protein Data Bank. In *Structural Bioinformatics*, Second Edition (J. Gu & P. E. Bourne, eds.), pp. 293-303. John Wiley & Sons, Inc., Hoboken, NJ.

27. S. Dutta, K. Burkhardt, G.J. Swaminathan, T. Kosada, K. Henrick, H. Nakamura & H.M. Berman. (2008) Data deposition and annotation at the Worldwide Protein Data Bank. In *Methods in Molecular Biology* (B. Kobe, M. Guss, T. Huber, eds.), Vol. 426: Structural Proteomics: High-Throughput Methods, pp. 81-101. Humana Press, Totowa, NJ.

28. 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.

29. C.L. Lawson, S. Dutta, J.D. Westbrook, K. Henrick & H.M. Berman. (2008) Representation of viruses in the remediated PDB archive. *Acta Cryst* **D64**: 874-882.

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

31. 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.

32. H.M. Berman, J.D. Westbrook, M.J. Gabanyi, W. Tao, R. Shah, A. Kouranov, T. Schwede, K. Arnold, F. Kiefer, L. Bordoli, J. Kopp, M. Podvinec, P. D. Adams, L.G. Carter, W. Minor, R. Nair & J. La Baer. (2009) The protein structure initiative structural genomics knowledgebase. *Nucleic Acids Res* **37**: D365-368.

33. M. Sanner. (1999) Python: A Programming Language for Software Integration and Development. *J Mol Graphics* **17**: 57-61.

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.

RCSB PDB

wwPDB Members

RCSB PDB
www.pdb.org

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

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 92093-0743

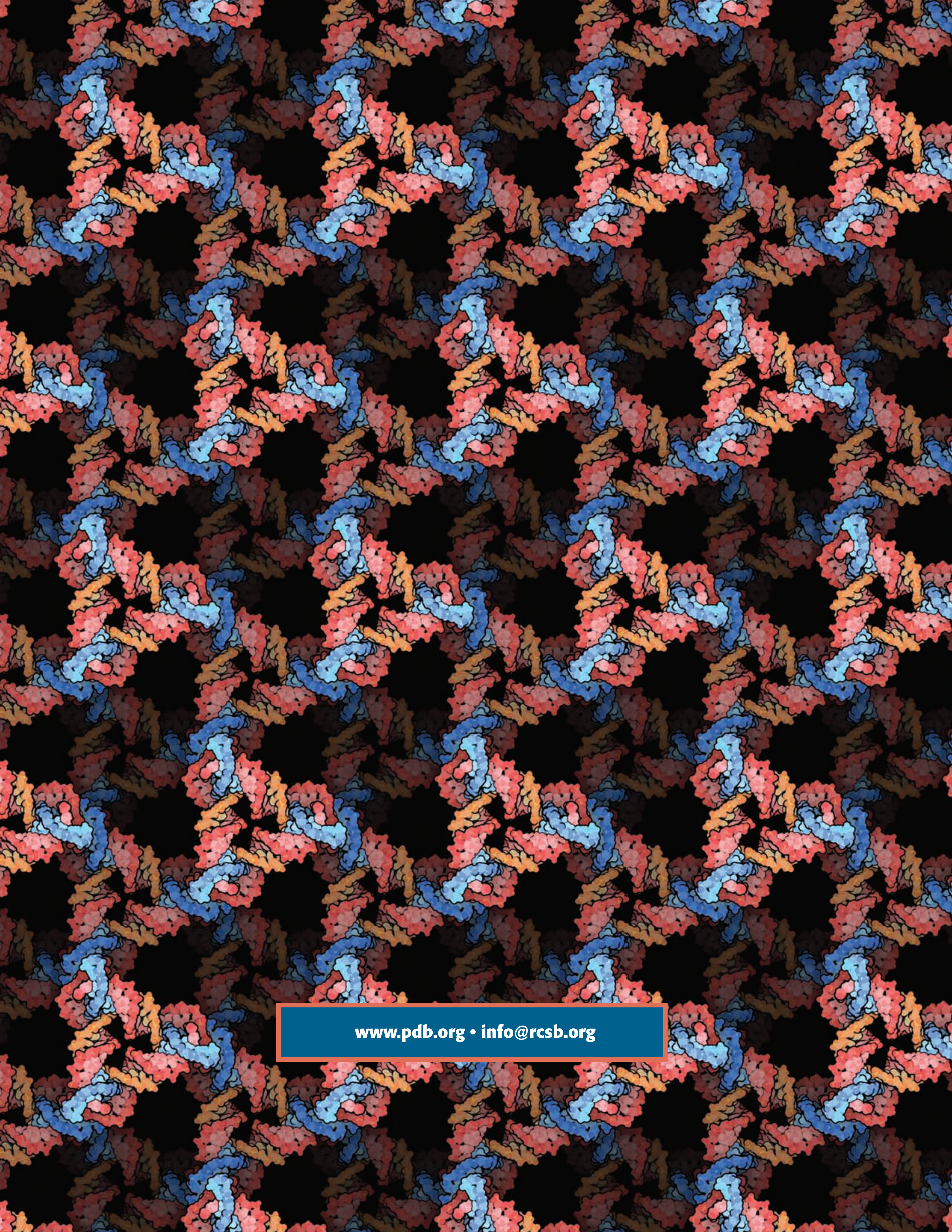
RCSB PDB Management

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.pdb.org • info@rcsb.org