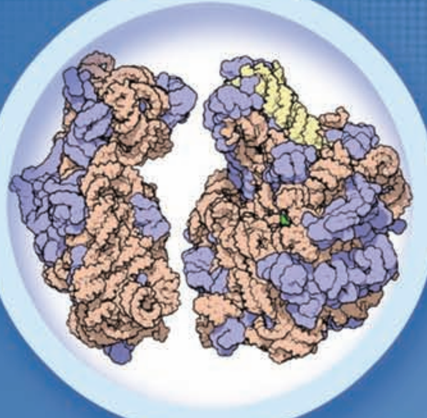
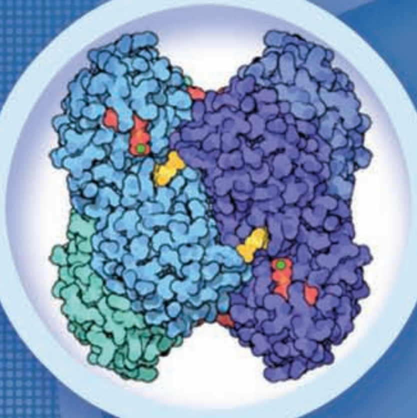
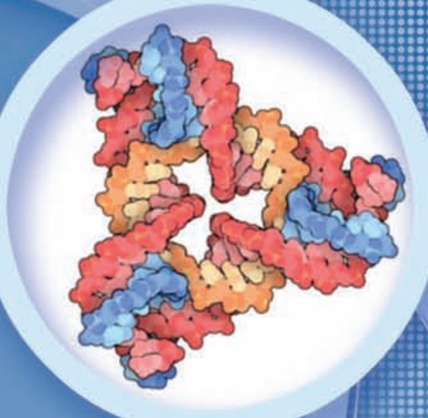
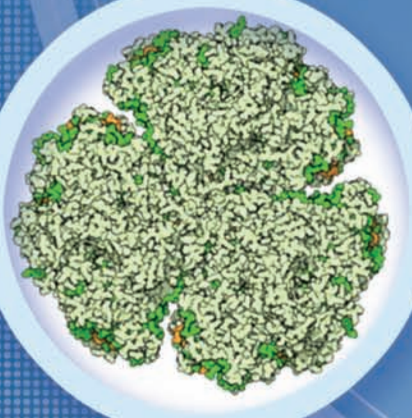


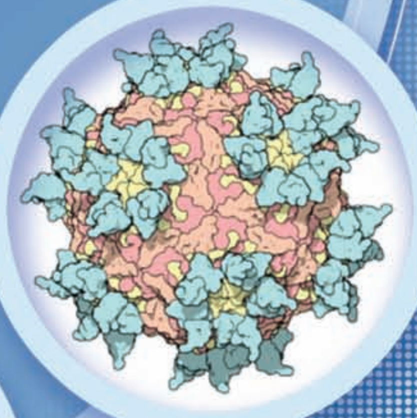


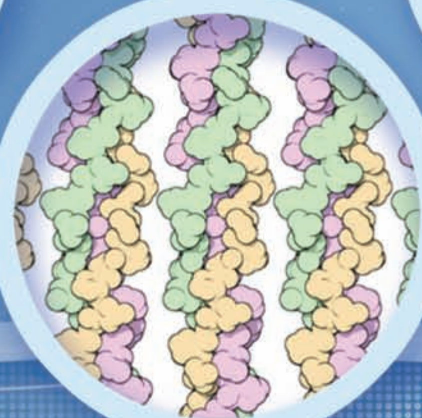


RCSB PDB
PROTEIN DATA BANK

# ANNUAL REPORT 2010







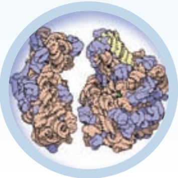

**RESEARCH COLLABORATORY FOR STRUCTURAL BIOINFORMATICS**
**Rutgers, The State University of New Jersey**
**University of California, San Diego**

# About the Cover

Since January 2000, the RCSB PDB's *Molecule of the Month* has been introducing readers to the structure and function of the many nucleic acids, proteins, and complex assemblies found in the PDB.

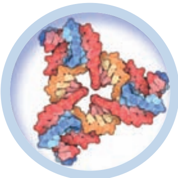
The cover of this report represents a new way of navigating through the archive of past *Molecule of the Month* columns. Each article has been sorted into a category or categories based on biological context.

Each structure shown represents a major, top-level functional category:



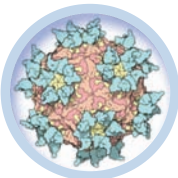
**PROTEIN SYNTHESIS** includes the major molecules of protein synthesis, from DNA to RNA to ribosomes to folded proteins.

Example: **ribosome**
**doi: 10.2210/rcsb_pdb/mom_2000_10**



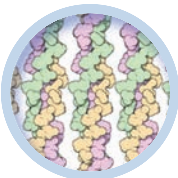
**BIOTECHNOLOGY AND NANOTECHNOLOGY** explores the ways scientists look to nature for inspiration, and harness biological machinery for use in science and technology.

Example: **designed DNA crystal**
**doi: 10.2210/rcsb_pdb/mom_2009_11**



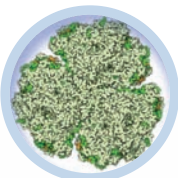
**HEALTH AND DISEASE** covers subjects such as drug action, viruses, and the molecular basis of disease.

Example: **rhinovirus**
**doi: 10.2210/rcsb_pdb/mom_2001_8**



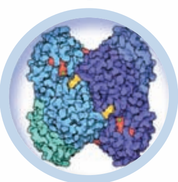
**INFRASTRUCTURE AND COMMUNICATION** includes topics such as structural proteins and cell signaling.

Example: **collagen**
**doi: 10.2210/rcsb_pdb/mom_2000_4**



**BIOLOGICAL ENERGY** presents the processes for capturing and converting energy in cells.

Example: **photosystem I**
**doi: 10.2210/rcsb_pdb/mom_2001_10**



The **ENZYMES** category organizes the many catalysts for all of the chemical transformations needed for life.

Example: **phosphofructokinase**
**doi: 10.2210/rcsb_pdb/mom_2004_2**

From these major topics, readers can browse descriptive subcategories that lead to the individual *Molecule of the Month* features. The Protein Synthesis category, for example, contains subcategories about the Structure of Nucleic Acids, DNA Maintenance, Replication, Transcription, Translation, and Protein Folding/Modification/Degradation. Each subcategory then reveals the relevant *Molecule of the Month* features. All articles can still be searched by date and feature title, but the category view lets users approach the *Molecule of the Month* archive with a biological interest, rather than a specific molecule, in mind.

We hope that this new interface will promote new ways of exploring these articles and the PDB archive.

*Molecule of the Month* articles can be referenced and accessed using their Digital Object Identifiers (DOI). The DOI can be used as part of a URL (**http://10.2210/rcsb_pdb/mom_2000_10**), or entered in a DOI resolver (such as **http://www.crossref.org/**) to automatically link to the feature at the RCSB PDB website.

# Table of Contents

# Message from the Director

The Protein Data Bank (PDB) is the central archive of the 3D structures of biological molecules. Originally developed as a repository for the structural biology community, it has grown in the past 39 years to serve a wide range of communities. Scientists from around the world depend upon PDB data for their research and discoveries, while students and teachers access the PDB to study biological processes and medicine at the molecular level.

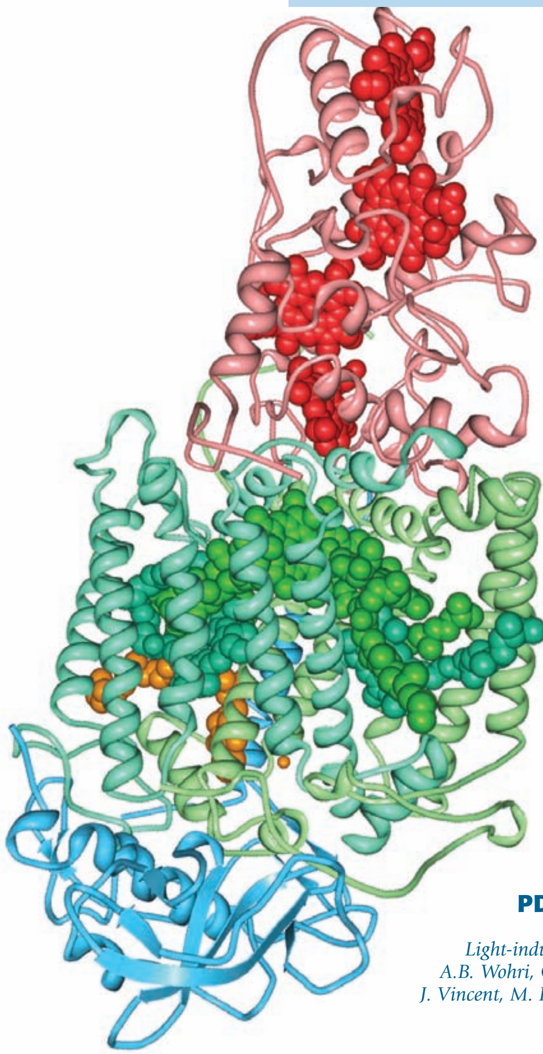
The PDB is by nature community-based. The archive itself is comprised of experimental data contributed by structural biologists. The free and open sharing and exchange of this data then drives science and education worldwide.

In partnership with the Worldwide Protein Data Bank, the RCSB PDB curates, annotates, and distributes PDB data. This collaboration ensures that PDB data are provided in an accurate and timely manner, and in a single, uniform, and global PDB archive.

The RCSB PDB website provides access to the PDB, along with related tools and resources to examine these structures in the context of function, biological processes, evolution, pathways and disease states. Conversations and feedback from RCSB PDB users guide the development of comprehensive, integrated, and unique views of the data that enable scientific innovation and education.

This report highlights the key components of the RCSB PDB– "data in", "data out", and outreach and education–along with the major highlights of the report period of July 1, 2009 - June 30, 2010. These activities promote access to and understanding of the PDB archive, with the goal of serving both the scientific and educational communities.

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



**PDB ID: 2x5u**

*Light-induced structural changes in a photosynthetic reaction center caught by Laue diffraction.*
*A.B. Wohri, G. Katona, L.C. Johansson, E. Fritz, E. Malmerberg, M. Andersson, J. Vincent, M. Eklund, M. Cammarata, M. Wulff, J. Davidsson, G. Groenhof, R. Neutze (2010)* Science *328: 630-633*

## Snapshot: July 1, 2010

### Number of Entries

| | |
|---|---|
| 66212 | released atomic coordinate entries |

### Entries by Molecule Type

| | |
|---|---|
| 61280 | proteins, peptides, and viruses |
| 2746 | protein/nucleic acid complexes |
| 2148 | nucleic acids |
| 38 | other |

### Entries by Experimental Technique

| | |
|---|---|
| 57298 | X-ray |
| 8449 | NMR |
| 295 | electron microscopy |
| 24 | hybrid methods |
| 146 | other |

### Related Experimental Data

| | |
|---|---|
| 46690 | structure factor files |
| 5742 | NMR restraint files |

# About the RCSB PDB

## The PDB Archive of Biomolecular Structures

As the single repository of information about the 3D structures of proteins, nucleic acids, and complex assemblies, the Protein Data Bank (PDB) archive[1-3] is a vital resource for worldwide research and study in the biomedical and agricultural sciences.

The archive contains biomolecules determined from experiments in X-ray crystallography, nuclear magnetic resonance (NMR), and electron microscopy. From around the world, scientists determine the atomic arrangements of these molecules, and submit data about their experiments to the PDB. This information is processed, validated, annotated, and made publicly available, at no cost, to the global community.

Countless individual users and external resources require PDB data for their study and research. By comparing the many structures available in the PDB, structural biologists have made basic discoveries in molecular evolution, enzyme action, molecular recognition, protein folding and assembly, and countless other central biological concepts. The structures in the PDB have been widely used for the development of new pharmaceutical agents; for instance, drugs developed from structures in the PDB have changed HIV infection from a uniformly deadly disease into a treatable condition. The broader community of biologists uses the structures in the PDB, augmented by sequence information and other data, to lend structural insight to their work, for instance, exploring the structural consequences of mutations. Finally, the PDB is widely used in the education and outreach community to create dynamic imagery that brings the results of basic and applied biological research to the general public.

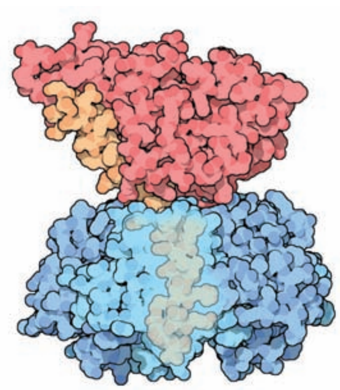
### As of July 2010, the PDB archive contained

**Public archive (109 GB)**
- More than 66,000 entries
- More than 500,000 files
- Data dictionaries and documentation
- Derived data files

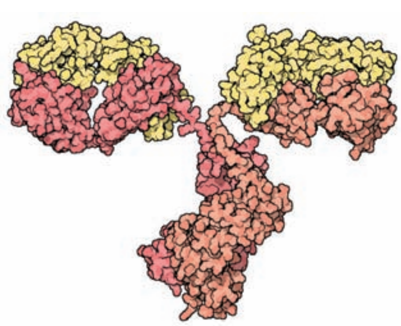
**For each PDB entry**
- Atomic coordinates
- Sequence information
- Description of structure
- Experimental data

**Internal archive**
- Depositor correspondence
- Depositor contact information
- Paper records
- Documentation
- Historical records

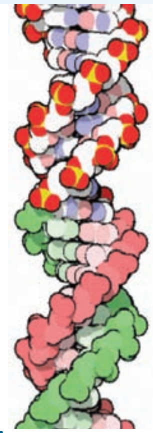
**The structures archived in the PDB are vital to the study of science and medicine worldwide. The examples shown below have been highlighted in past *Molecule of the Month* features.**



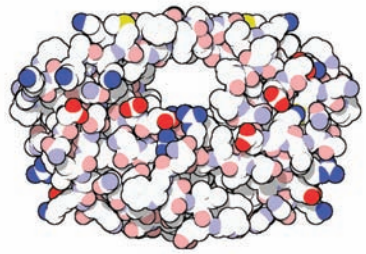
**Cholera Toxin**
doi: 10.2210/rcsb_pdb/mom_2005_9



**Antibodies**
doi: 10.2210/rcsb_pdb/mom_2001_9



**DNA**
10.2210/rcsb_pdb/mom_2001_11

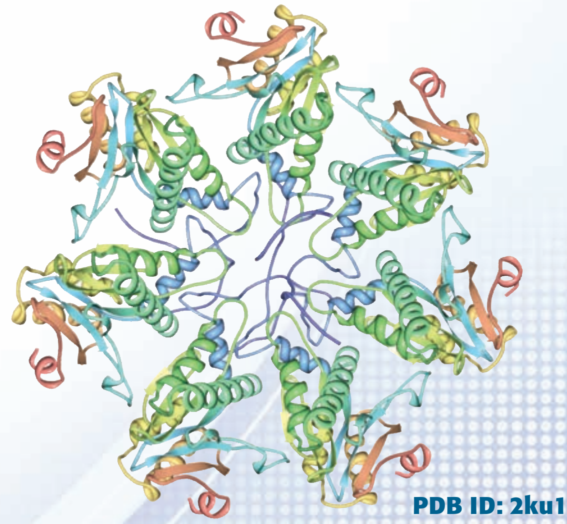


**HIV-1 Protease**
doi: 10.2210/rcsb_pdb/mom_2000_6

## What Does the PDB Enable?

Through the safe storage and free access to the scientific data in the PDB archive, scientists and students can

- Decipher the functions of molecules with known sequence and structure, but unknown biochemical function
- Study evolution over long-time scales where the sequence signal indicating a relationship is too weak to measure
- Better understand regulation, signaling and other pathways
- Study protein motion and allostery
- Explore the underlying mechanisms of organ systems in the human body, from circulatory to lymphatic to urinary systems
- Design drugs that will respond to emerging infectious diseases
- Understand biological principles in detail
- Predict and develop models for proteins with unknown 3D structure



**PDB ID: 2ku1**

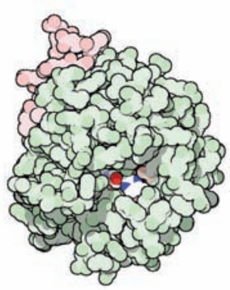
*Dynamic regulation of archaeal proteasome gate opening as studied by TROSY NMR. T. L. Religa, R. Sprangers, L. E. Kay (2010)* Science *328: 98-102*

## The RCSB Protein Data Bank

The RCSB Protein Data Bank, administered by the Research Collaboratory for Structural Bioinformatics (RCSB),[2] provides a global resource for the advancement of research and education in biology and medicine by curating, integrating, and disseminating biological macromolecular structural information in the context of function, biological processes, evolution, pathways, and disease states.

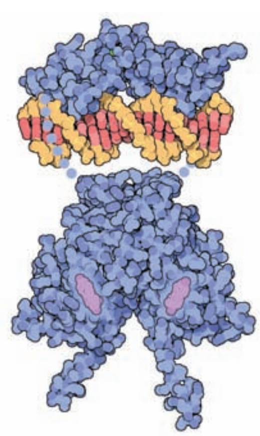
As a member of the Worldwide PDB (wwPDB), the RCSB PDB supports the deposition, processing, and annotation of experimental structure data. As the archive keeper for the wwPDB, the RCSB PDB maintains the PDB central repository.

The RCSB PDB's website offers a searchable database, tools, and related services for users to easily search, analyze, and visualize the data in the PDB archive. These efforts are enhanced by targeted education and outreach activities.



**Thrombin**

doi: 10.2210/rcsb_pdb/mom_2002_1



**Estrogen Receptor**

doi: 10.2210/rcsb_pdb/mom_2003_9

## The Worldwide Protein Data Bank

WORLDWIDE PDB PROTEIN DATA BANK

The wwPDB organization (**www.wwpdb.org**)[1] is a unique scientific collaboration that provides the authoritative resource for experimentally-determined 3D structures of biological macromolecules. It was formed to ensure that the PDB archive, now and in the future, is freely and publicly available to the global community. wwPDB members (RCSB PDB, Protein Data Bank in Europe (PDBe),[4] Protein Data Bank Japan (PDBj),[5] and the BioMagResBank (BMRB)[6]) host deposition, processing, and distribution centers for PDB data and collaborate on a variety of projects and outreach efforts.

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 situations, the wwPDB regularly reviews data processing procedures, coordinates "remediation" efforts to improve data representation in the archive, and implements policies and standards to accommodate new developments in structural biology. Policies are developed in concert with the wwPDB Advisory Committee and with Validation Task Forces comprised of experts in the experimental method communities. These Task Forces have been convened to collect recommendations and develop consensus on additional validation that should be performed, and to identify software applications to perform validation tasks.

# About the RCSB PDB

## Funding

The RCSB PDB is supported by funds from:

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

## The RCSB PDB Organization

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

Helen M. Berman, Director of the RCSB PDB, is a Board of Governors professor of chemistry and chemical biology at Rutgers. Professor Berman was part of the team that first envisioned the PDB archive. Dr. Martha Quesada, Deputy Director (Rutgers), and Professor Philip E. Bourne, Associate Director (UCSD), join her in RCSB PDB management.

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

The RCSB PDB receives input from an advisory board of experts in X-ray crystallography, NMR, 3D EM, bioinformatics, and education (the RCSB PDB Advisory Committee).

*The RCSB PDB team and collaborators*

# Data Input

## Before Structures Enter the Archive: Data Deposition, Validation, and Annotation

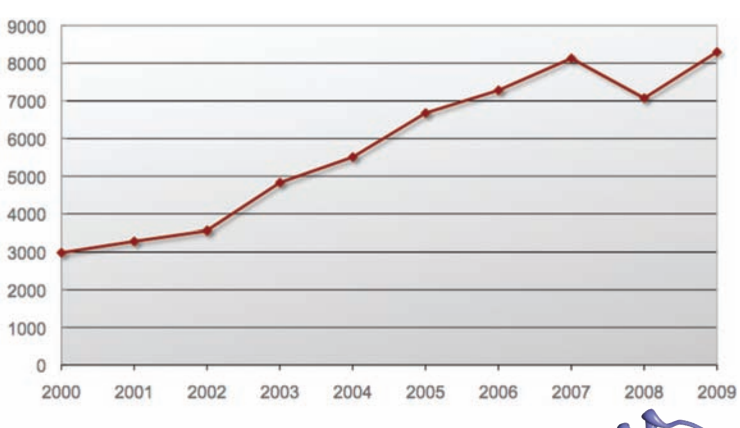
The data in the PDB is contributed from scientists located around the world, and then validated and curated by the wwPDB before being made publicly available in the archive.[7]

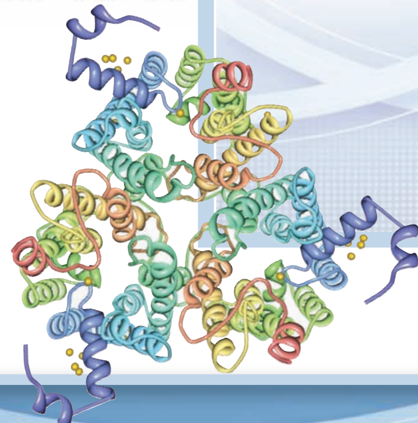
When a structure is deposited online, it is immediately assigned its own unique PDB ID. Approximately 25 new entries are deposited to the PDB each day. Using locally-developed tools, the RCSB PDB carefully curates and annotates entries using an integrated system that uses a standard data dictionary for macromolecular structure[8] and for the small chemical components found in PDB entries. The goal is that PDB entry accurately represents the structure and experiment.

Annotation-curators check each deposition for errors or omissions, and compare the reported sequence, source, and other information with external databases.[9-10] The entry is assigned a title, names and synonyms for the protein or other polymer, scientific name of macromolecule's source, and biological assembly information. After reviewing the structure visually, the completed entry is sent to the depositor, along with a summary validation report. Annotators then work with the depositor to address any issues and finalize the entry for public release. On average, this process takes two weeks.

Depending upon the hold status selected by the depositor, data release occurs when a depositor gives approval to the finalized 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.

### Growth in Data Depositions (2000-2009)



**PDB ID: 3k3f**

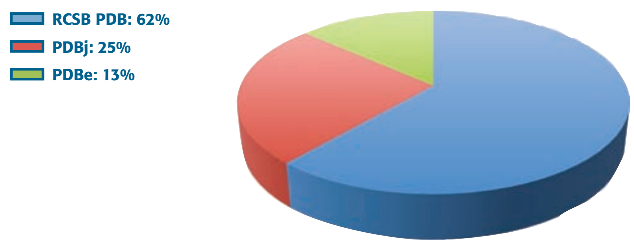
*Crystal structure of a bacterial homologue of the kidney urea transporter. E.J. Levin, M. Quick, M. Zhou (2009)* Nature *462: 757-761*

## Deposition Statistics July 1, 2009– June 30, 2010

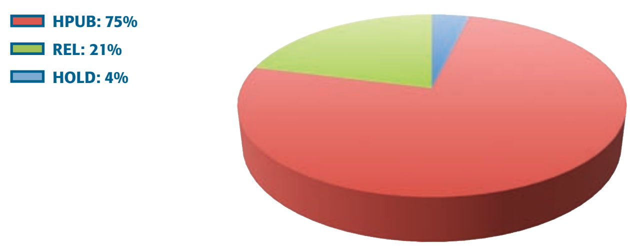
7848 structures were deposited to the PDB archive and prepared for release by the wwPDB during the period of this report. 6860 of these entries were deposited to the RCSB PDB.

Sequences for 55% of these depositions were made publicly available before the coordinate files. This helps to prevent duplication of structure determination effort and promotes blind testing of structure prediction and modeling techniques.

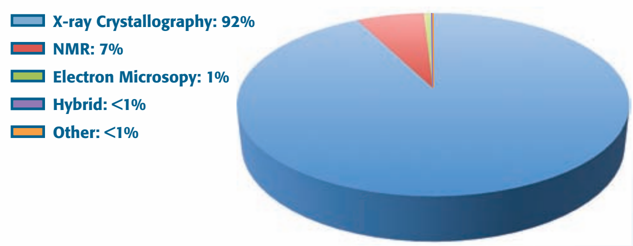
### wwPDB Processing Site



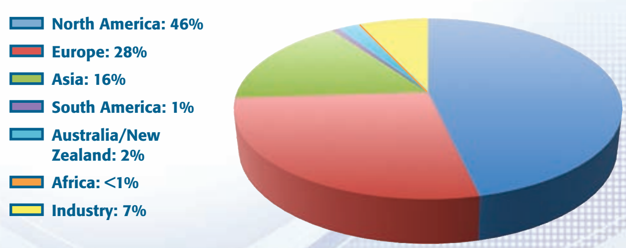

### Author-defined Release Status




### Experimental Source




### Depositor Location



# Data Input

## wwPDB Validation Reports

**Review**
output from validation programs

**Respond**
to any issues

**Share**
with journals and coauthors

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

As part of the structure annotation process, wwPDB members provide depositors with detailed validation reports that include the results of geometric and experimental data checking.[11-14] These reports, available as PDFs, can be easily reviewed and shared by depositors.

As these validation reports provide an assessment of structure quality while keeping the coordinate data confidential, we are encouraging journal editors and referees to request them from depositors as part of the manuscript submission and review process. PDB validation reports are already required by the International Union of Crystallography (IUCr) journals as part of their submission process. The reports will continue to be developed and improved as we receive recommendations from our Validation Task Forces and as we further develop our data deposition and processing procedures.

## Software for Deposition and Validation

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

• **pdb_extract**[15] automatically puts key details from the output files produced by many NMR and X-ray crystallographic applications 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. Researchers are encouraged to use this program to review the quality of any released structure before using it in their own study.

• The RCSB PDB and PDBj use **ADIT**[2] 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. A new and improved version of ADIT was released during this report period. With this version, depositors are required to run validation checks before the entry is deposited.

**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 are processed and archived by BMRB. Similarly, **EMDep** integrates map deposition to EMDB and model coordinate deposition to PDB.

Annotators use a special version of ADIT to prepare entries for release.

• **Ligand Expo** provides tools for accessing, visualizing, and viewing reports about the information in the wwPDB's Chemical Component Dictionary. It can also be used to identify structure entries containing particular small molecules; browse tables of components that contain modified amino acids, nucleotides, popular drugs (trade and generic names) and common ring systems; review related information in chemical dictionaries and resource files; and download the 3D structure of small molecule components. A sketch tool is also provided for building new chemical definitions from reported PDB chemical components. This tool is used by annotators and by depositors preparing submissions.

## Deposition in the Future: The wwPDB Common Deposition & Annotation Tool

While the wwPDB partners follow established procedures and formats for data processing, different tools are used for the annotation of data, depending upon what data were deposited and where. The wwPDB partners have started a project to produce a set of common deposition and annotation processes and tools that will enable the wwPDB to deliver a resource of increasingly high quality and dependability over the next 10 years. Specifically, the new processes and supporting systems will address the anticipated increase in complexity and experimental variety of submissions over a 10-year life span along with the increase in deposition throughput. The processes and tools will maximize the efficiency and effectiveness of data handling and support for the scientific community.

The tool is being developed module by module. Work during the period of this report has focused on the sequence and the ligand processing modules, and on the overall workflow manager. These modules are integrated with external resources and visualization tools.

This single, common tool is being developed by the wwPDB following standard project management procedures. Made up of experts from all partner sites, the project team meets regularly and works in conversation with the wwPDB Directors and the depositing community.

# Data Access, Query, and Reporting

## Data Distribution and Access

RCSB PDB services and data are freely available online.

As the archive keeper for the wwPDB, the RCSB PDB maintains the PDB archive at **ftp://ftp.wwpdb.org**. It has sole write access to the FTP archive, and controls updates and distribution to our global partners. Updates of the PDB archive and the wwPDB websites are coordinated to occur at the target time of Wednesday 00:00 UTC (Tuesday 5:00 p.m. Pacific).

During this report period, **7835 PDB entries** (in PDB, mmCIF, and PDBML/XML file formats), **7422 structure factor files**, and **551 constraint data files** were added to the archive.

The RCSB PDB website at **www.pdb.org** is a portal to these data files. The website also provides access to a relational database that integrates these data with more than 30 external sources of related data (e.g., sequence annotations, structure annotations, taxonomy, functional descriptors), tools for visualization, and related editorial and educational content. The website is accessed by about 180,000 unique visitors per month from more than 150 different countries. Around 1 terabyte of data is transferred each month this way. At the same time, about 11,000 unique visitors download more than 14 million files from the FTP site, for a total of about 2 terabytes of data.

Data are also accessed through Web Services that allow remote software to search the RCSB PDB database and retrieve data for any given structure. These services, including the new RESTful Web Services, help developers create tools that interact efficiently with PDB data.

## Website Overview

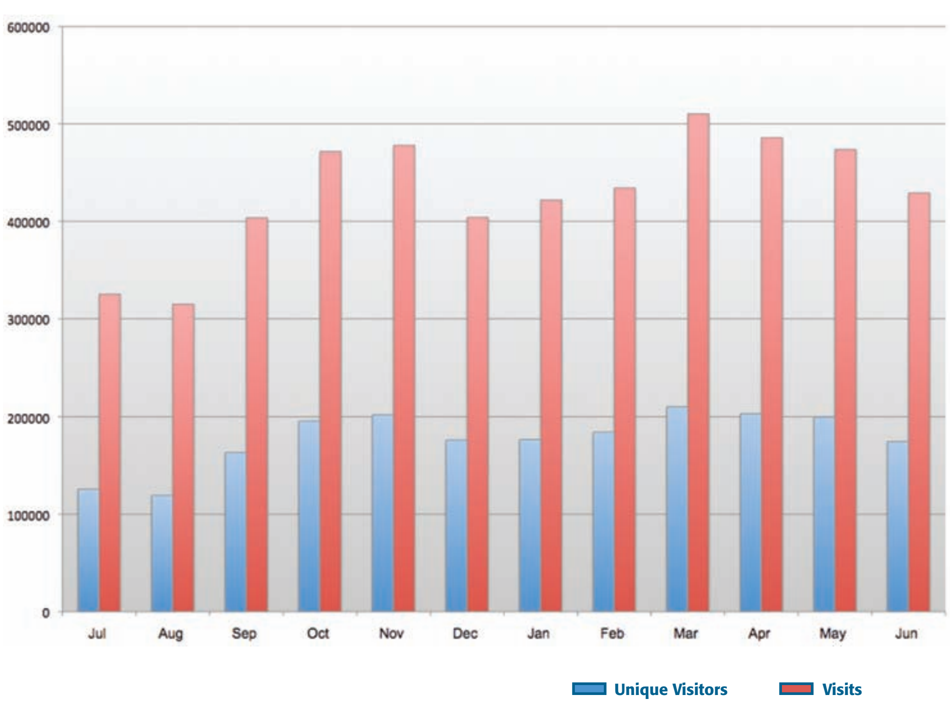
Recently, RCSB PDB web pages were reorganized in response to user feedback. Navigating the website and search results is easier and more intuitive.

Much of the website, including the home page, uses customizable web widgets. Boxes with a dark blue bar on top are widgets that can be moved on the page by dragging the arrow buttons, hidden by selecting "Hide," or included in a customized view. The left-hand menu now groups frequently-used web pages into sections–Deposition, Home, Search, Tools, My PDB, Help, and Education–that can be moved up and down to order by user interest. New widgets on the home page display and link to all entries released in the most recent update in the *Latest Structures* widget, while the *New Features* widget tiles through the enhancements recently added to the website, with links to detailed information.

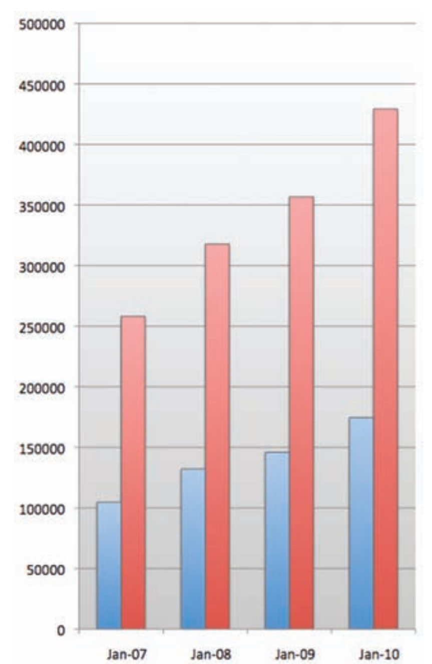
## Exploring Individual Structures

For every entry in the PDB, an RCSB PDB Structure Summary page provides an overview of the structure; derived data from CATH, SCOP, Pfam, and 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 and Ligand Explorer,[21] help users view the molecule interactively. A Literature view option provides the abstract and related details about the reference describing the structure, and includes information about other articles referencing the entry.[22]

**Unique Visitors and Visits to www.pdb.org (July 1, 2009 - June 30, 2010)**




**Growth in Website Access (2007-2010)**



# Data Access, Query, and Reporting

## Data Query and Reporting

The website provides access to a relational database[23-24] 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 tool *Advanced Search*.

PDB structures can be browsed using tree-like hierarchies based on GO terms, Enzyme Classification,[25] Medical Subject Headings, Source Organism, Genome Location,[10] SCOP, and CATH classifications.

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

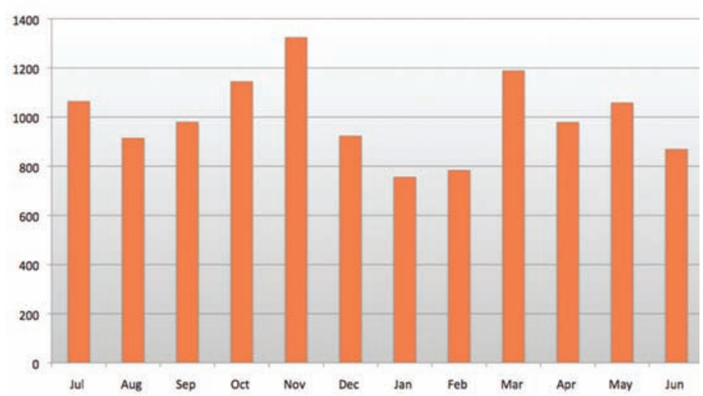
RCSB PDB website features are supported by help pages, tutorials, and an active help desk. Additional enhancements and new resources are continually in development.

### Data downloads from ftp.wwpdb.org

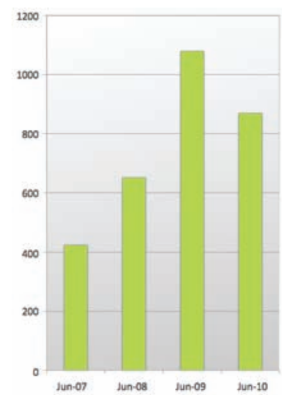
During this report period, ftp://ftp.wwpdb.org hosted

- 174,571,033 total downloads (14138 GB)
- An average of 129,165 unique users each month

### Data downloads per month from www.pdb.org in GB (July 1, 2009 - June 30, 2010)

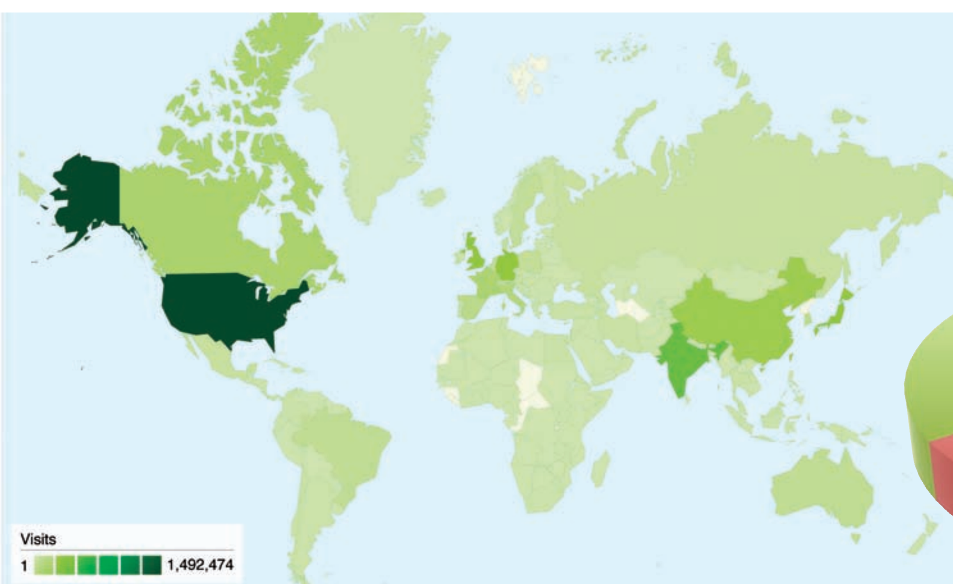



### Yearly growth in GB of web data downloads (2007-2010)

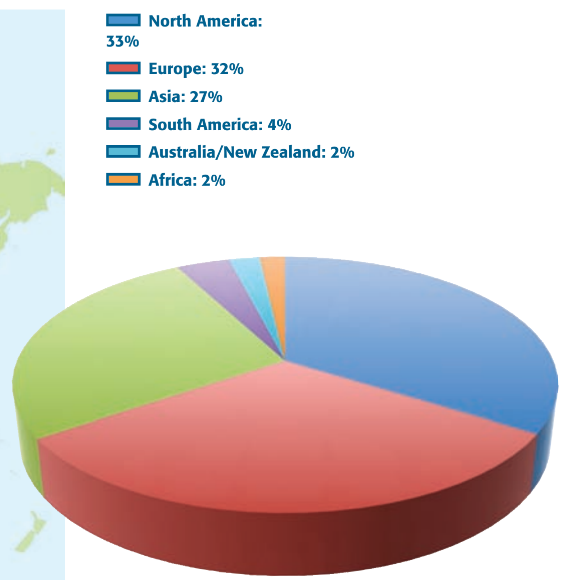



*The number of downloads in 2009 reflects the release of a standardized version of the entire archive.*

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



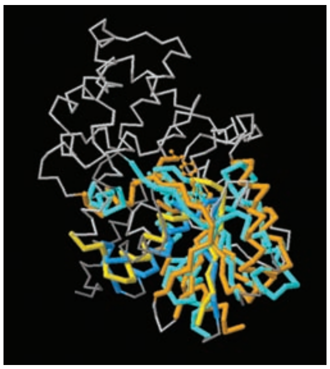

*Web traffic by country. Image from Google Analytics.*




*Web traffic by continent.*

## New Features

The **Comparison Tool** calculates pairwise sequence (blast2seq, Needleman-Wunsch, and Smith-Waterman)[26-28] and structure alignments (FATCAT, CE, Mammoth, TM-Align, TopMatch).[29-34]



*Structure alignment of two nucleotide-binding proteins.*[36-37] *The Comparison Tool's jCE option can detect the circular permutations that are difficult for many alignment algorithms to detect.*

Comparisons can be made for any protein in the archive and for customized or local files not in the PDB. Special features include support for both rigid-body and flexible alignments (via jFATCAT) and detection of circular permutations (via jCE).[35] jFATCAT and jCE alignments are performed on RCSB PDB servers, while Mammoth, TMAlign, and TopMatch comparisons are calculated by through links to the corresponding sites.

The Comparison Tool is available from the RCSB PDB as a web page, web widget, and stand-alone Java Web Start application. It is also integrated with the Sequence Clusters offered from each entry's Sequence Similarity Tab.

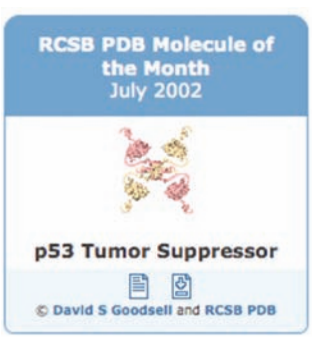
**Ligand searching and reporting** for ligands using the *Advanced Search*, the *Chemical Structure Search*, or the top-bar pulldown has been greatly improved.

• The Chemical Name search from the **top-bar pulldown** on every page returns matches to names of small molecules in the wwPDB Chemical Component Dictionary and any synonyms. Searches by Chemical Name or ID will return structures with matching chemical components that are free ligands or are part of a protein or nucleic acid chain.

• In ***Advanced Search***, searches can be specified to look for free and/or polymeric chemical components. A "sounds like" option searches for misspelled or incomplete names. Advanced Searches using SMILES strings use a similarity threshold while specifying polymeric type.

• The ***Chemical Structure Search*** utilizes the latest version of the ChemAxon MarvinSketch applet (5.3.0.1).

Small bits of code, called **Web Widgets**, embed RCSB PDB functionality in external websites.[38] The widgets currently available include a *Molecule of the Month* widget; a Tag Library that links to the corresponding Structure Summary page, Jmol view, and PDB file; a widget that embeds an image of given PDB ID; and a Comparison Tool widget.




Molecule of the Month *widget*

## MyPDB

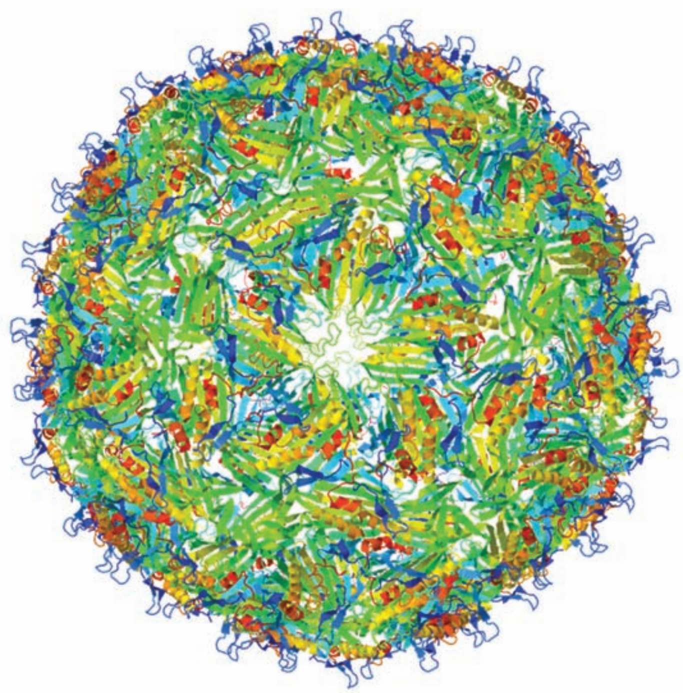
The MyPDB service automatically emails alerts when the PDB releases entries that match user-stored queries.

With MyPDB, users can

- combine and save keyword, sequence, ligand and other searches
- refine and/or run stored searches at any time
- receive email alerts weekly or monthly

To subscribe, select the MyPDB registration option from the left-hand menu at **www.pdb.org**.

**Display of Large Assemblies**: Providing a simple interactive view of extremely complex structures has been challenging. The RCSB PDB's Jmol view now supports all entries, including extremely large ribosome complexes split across multiple entries, and vault structures that contain a very large number of atoms and chains. Composite views of split entries can also be displayed for asymmetric units and the biological assemblies.



*Large structures, such as the bacteriophage structure 1gav,*[39] *can be viewed on the RCSB PDB site using Jmol. The Jmol window can be resized for better viewing, and a menu console is provided for easy use of the program.*

## Time-stamped Archival Copies

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

# User Outreach

## Community Interactions


The RCSB PDB works with different user communities to develop a resource that supports a variety of audiences.

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

*RCSB PDB members actively participate at several meetings throughout the year.*

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

While **www.pdb.org** serves as the primary tool for outreach, staff also interact directly with users at international meetings, workshops, presentations, festivals, and more. The RCSB PDB exhibits at a variety of conferences, including those of the *American Society for Biochemistry and Molecular Biology*, the *American Crystallographic Association*, and the international conference on *Intelligent Systems for Molecular Biology*. Formal and informal surveys at meetings are used to solicit detailed feedback about the resource.

## Publications

**www.pdb.org** is updated regularly with news, recent developments, new resources, and improvements to existing resources, and educational resources.

Published quarterly in print and online, the *RCSB PDB Newsletter* describes and highlights recent activities.

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

## Online Help

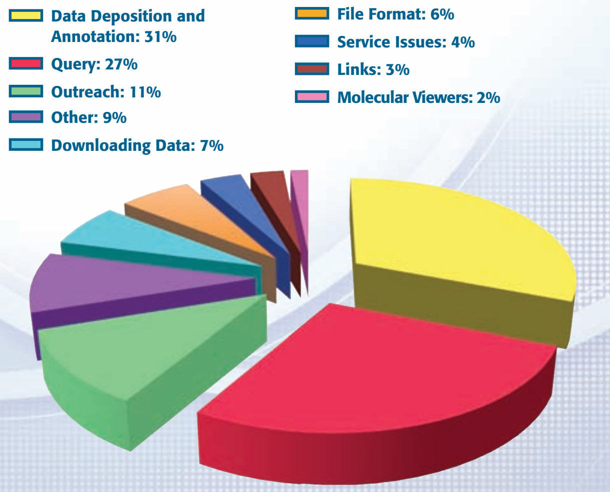
Many online resources and outreach programs are available to assist RCSB PDB users in taking advantage of all RCSB PDB services.

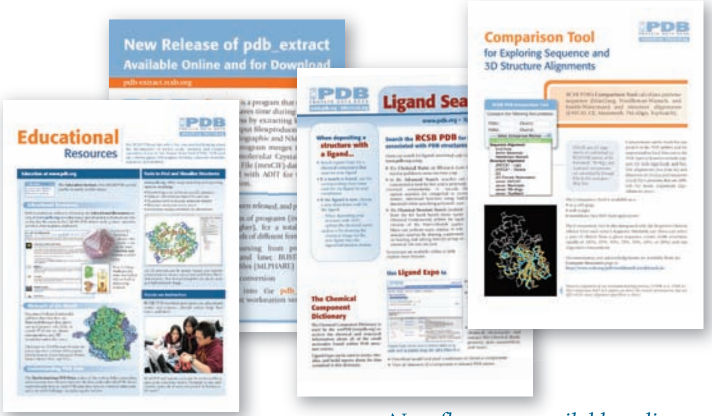
The electronic help desk at **info@rcsb.org** provides around-the-clock support for using RCSB PDB resources and beyond. Many requests sent to this help desk lead to new website enhancements and features. Other help desks are available to assist in queries specific to the data deposition and annotation process.

Brief, online video tutorials demonstrate how to use website features such as ligand searching and website customization tools. These screencasts are created and narrated by RCSB PDB staff.

A glossary and site map guide users through the website. Context-sensitive help pages are also available throughout the site where a ? appears.

### Help Desk Overview: Types of questions sent to info@rcsb.org




*New flyers are available online or in print by request to* ***info@rcsb.org****.*

The RCSB PDB regularly publishes articles covering a diverse array of subjects. Recent articles have described the new widget features,[38] the integration of open access literature with the database,[22] a tool for displaying domain assignments,[40] and a guide to using PDB data for illustrators.[41]

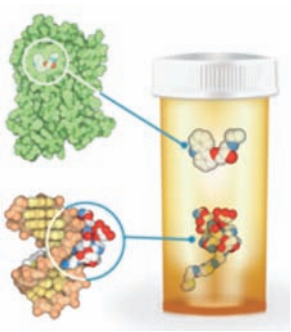
# Educational Initiatives

While most RCSB PDB programs are aimed at scientific researchers, our efforts also promote scientific literacy and structural biology to a much broader community.[42-43] Educational programs and resources are available to help teachers and students at all levels understand PDB data and structures.

## Online Features

### Educational Resources

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.



*The* How Do Drugs Work? *poster is one of the downloadable materials available from* ***www.pdb.org****. In a style similar to the* Molecule of the Month *series, the poster uses PDB entries to illustrate different drug-protein interactions.*

### *Molecule of the Month*

Created by David S. Goodsell (RCSB PDB, The Scripps Research Institute), this feature provides an easy introduction to macromolecular structures, shows how the structures function, and highlights their importance in our lives. Each illustrated entry links to specific PDB entries, asks discussion questions, and hosts customized interactive 3D molecular views. *Molecule of the Month* articles describe everything from AAA+ proteases to zinc fingers.

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

### Understanding PDB Data: Looking at Structures

Using the data in the PDB archive can be challenging to the non-expert. An understanding of how a data file came to be–from the experiment to data release–is needed to analyze the data. Using text, images, and interactive views, *Looking at Structures* helps researchers and educators understand how to read PDB data files, visualize structures, read coordinate files, and know about potential challenges.

### Education Corner

In addition to publicizing developments and activities, the quarterly RCSB PDB Newsletter hosts an *Education Corner* feature that describes how the PDB archive and RCSB PDB resources are used in classrooms all over the world. Past topics have included a rubric for biomacromolecular 3D literacy, the iPhone program Molecules, and physical 3D molecular models.

### Suite of Online Training Tools

For a detailed introduction to the features and functionality of the RCSB PDB, we have worked with OpenHelix to provide a comprehensive set of training materials. The tools include an narrated tutorial that demonstrates basic and advanced searching, report generation, exploration of individual structures, and many other research and education tools. The full tutorial runs for about an hour, and can be navigated by specific chapters.

The animated PowerPoint slides used as a basis for the tutorial can be downloaded, along with slide handouts and exercises. These materials are freely available to create classroom content.

## Educational Programs and Activities

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

Other programs focus on getting students and teachers at the middle and high school level interested and involved in the fields of structural biology and bioinformatics. Area high school classes visit and tour the RCSB PDB. These experiences often lead to new initiatives and collaborations.

At Rutgers and UCSD, the RCSB PDB leaders are involved in graduate and undergraduate courses that depend heavily on the data in the PDB archive. Team members also participate in many on-campus organizations and events.

### Science Olympiad

The RCSB PDB also sponsors a protein modeling event at the Science Olympiad held in New Jersey. In the 2010 trial challenge, high school teams demonstrated their understanding of hemagglutinin, neuraminidase, and how protein structure relates to the H1N1 influenza virus through hand-built models and onsite exams. Information and resources for the NJ event can be found at **education.pdb.org** and **twitter.com/buildmodels**. The event is organized nationally by the MSOE Center for BioMolecular Modeling (**cbm.msoe.edu**).

*As part of Alumni Weekend, the* Art of Science *exhibit was installed at Rutgers.*

*Children of all ages create viral icosahedron structures at a variety of events.*

*The 2010 NJ State Champions in protein modeling from West Windsor Plainsboro-North High School.*

*At the protein modeling event in NJ, annotators "validate" and score the students' models.*

# Related Resources

## Structural Genomics and the PSI Structural Biology Knowledgebase

Structural genomics efforts worldwide have determined more than 9,000 structures–nearly 15% of the PDB archive. The RCSB PDB works closely with these centers to make the data available and collaborates with related resources.

TargetDB (**targetdb.sbkb.org**)[44] tracks the experimental progress made on sequence targets by structural genomics projects worldwide. The Protein Expression Purification and Crystallization Database (PepcDB, **pepcdb.sbkb.org**)[45] collects protocol information and experimental progress for proteins targeted by the NIGMS Protein Structure Initiative. These databases are two of the many resources that are accessible through the Structural Biology Knowledgebase (SBKB, **sbkb.org**).[46]

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

*Each month, the SBKB's* Featured PSI Molecule *gives a detailed portrait solved by the PSI structural genomics efforts. Shown is the structure of Alg13 solved by the Northeast Structural Genomics Consortium. Image created using the Python Molecular Viewer.*[47]

**PDB ID: 2jzc**

*Solution structure of Alg13: the sugar donor subunit of a yeast N-acetylglucosamine transferase. X. Wang, T. Weldeghiorghis, G. Zhang, B. Imperiali, J.H. Prestegard (2008)* Structure ***16****: 965-975*

## BioSync

BioSync (Structural Biology Synchrotron Users Organization) is an online guide to high energy data collection facilities for structural biologists.[48]

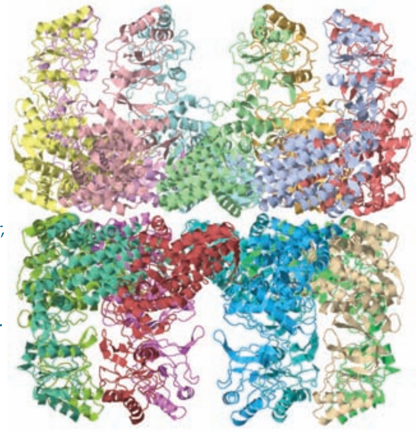
Through **biosync.sbkb.org**, it provides detailed information and usage statistics for macromolecular beamlines at synchrotron facilities worldwide. As data collected from synchrotron beamlines currently account for more than 80% of all X-ray crystallographic entries deposited to the PDB, the BioSync website has been recently upgraded with a new layout and features through funding from NIGMS.

BioSync offers information uploaded by beamline personnel about each synchrotron. Users can access information about specific capabilities, equipment, services and functions (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) in a report or through a search form.

PDB deposition statistics are grouped by geographical region and by synchrotron, with new graphs comparing beamlines within a site. Galleries of structures and tables containing citations and other general information are also available, with a separate set of statistics, images, and information provided for structures produced by structural genomics efforts.

## EMDataBank.org

The PDB archives large biological assemblies determined by cryo-electron microscopy (cryoEM), a maturing methodology in structural biology that bridges the gap between cell biology and the experimental techniques of X-ray crystallography and NMR. In addition to 3D density maps, cryoEM experiments often yield fitted coordinate models. Through an NIH/NIGMS-funded collaboration, a joint EM map and model deposition tool has been developed. The unified resource for deposition and retrieval network for cryoEM map, model, and associated metadata is available at **EMDataBank.org**.[49] This work has been carried out in collaboration with PDBe at the European Bioinformatics Institute and the National Center for Macromolecular Imaging at Baylor College of Medicine.



**PDB ID: 3iyf**

*Mechanism of folding chamber closure in a group II chaperonin. J. Zhang, M. L. Baker, G.F. Schroder, N. R. Douglas, S. Reissmann, J. Jakana, M. Dougherty, C.J. Fu, M. Levitt, S.J. Ludtke, J. Frydman, W. Chiu. (2010)* Nature *463: 379-383.*

*Image created using Jmol. The coordinates for this complex were jointly deposited along with the EM Data Bank map entry EMD-5140.*

# References

Except where noted, molecular graphics images were produced using the RCSB PDB's Protein Workshop.[21]


1. H. M. Berman, K. Henrick, H. Nakamura. (2003) Announcing the worldwide Protein Data Bank. *Nat. Struct. Biol.* **10**: 980.
2. H. M. Berman, J. D. Westbrook, Z. Feng, G. Gilliland, T. N. Bhat, H. Weissig, I. N. Shindyalov, P. E. Bourne. (2000) The Protein Data Bank. *Nucleic Acids Res.* **28**: 235-242.
3. F. C. Bernstein, T. F. Koetzle, G. J. B. Williams, E. F. Meyer Jr., M. D. Brice, J. R. Rodgers, O. Kennard, T. Shimanouchi, M. Tasumi. (1977) Protein Data Bank: a computer-based archival file for macromolecular structures. *J. Mol. Biol.* **112**: 535-542.
4. S. Velankar, Y. Alhroub, A. Alili, C. Best, H. C. Boutselakis, S. Caboche, M. J. Conroy, J. M. Dana, G. van Ginkel, A. Golovin, S. P. Gore, A. Gutmanas, P. Haslam, M. Hirshberg, M. John, I. Lagerstedt, S. Mir, L. E. Newman, T. J. Oldfield, C. J. Penkett, J. Pineda-Castillo, L. Rinaldi, G. Sahni, G. Sawka, S. Sen, R. Slowley, A. W. Sousa da Silva, A. Suarez-Uruena, G. J. Swaminathan, M. F. Symmons, W. F. Vranken, M. Wainwright, G. J. Kleywegt. (2011) PDBe: Protein Data Bank in Europe. *Nucleic Acids Res.*
5. Protein Data Bank Japan. http://www.pdbj.org/.
6. E. L. Ulrich, H. Akutsu, J. F. Doreleijers, Y. Harano, Y. E. Ioannidis, J. Lin, M. Livny, S. Mading, D. Maziuk, Z. Miller, E. Nakatani, C. F. Schulte, D. E. Tolmie, R. Kent Wenger, H. Yao, J. L. Markley. (2008) BioMagResBank. *Nucleic Acids Res.* **36**: D402-408.
7. 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.
8. 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.
9. 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.
10. E. W. Sayers, T. Barrett, D. A. Benson, E. Bolton, S. H. Bryant, K. Canese, V. Chetvernin, D. M. Church, M. Dicuccio, S. Federhen, M. Feolo, L. Y. Geer, W. Helmberg, Y. Kapustin, D. Landsman, D. J. Lipman, Z. Lu, T. L. Madden, T. Madej, D. R. Maglott, A. Marchler-Bauer, V. Miller, I. Mizrachi, J. Ostell, A. Panchenko, K. D. Pruitt, G. D. Schuler, E. Sequeira, S. T. Sherry, M. Shumway, K. Sirotkin, D. Slotta, A. Souvorov, G. Starchenko, T. A. Tatusova, L. Wagner, Y. Wang, W. John Wilbur, E. Yaschenko, J. Ye. (2010) Database resources of the National Center for Biotechnology Information. *Nucleic Acids Res.* **38**: D5-16.
11. J. Westbrook, Z. Feng, K. Burkhardt, H. M. Berman. (2003) Validation of protein structures for the Protein Data Bank. *Meth. Enz.* **374**: 370-385.
12. 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.
13. 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.
14. 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. D* **55**: 191-205.
15. 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 Crystallogr. D* **60**: 1833-1839.
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. 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.
18. E. L. Sonnhammer, S. R. Eddy, E. Birney, A. Bateman, R. Durbin. (1998) Pfam: multiple sequence alignments and HMM-profiles of protein domains. *Nucleic Acids Res.* **26**: 320-322.
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. http://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. A. Prlic, M. A. Martinez, D. Dimitropoulos, B. Beran, B. T. Yukich, P. W. Rose, P. E. Bourne, J. L. Fink. (2010) Integration of open access literature into the RCSB Protein Data Bank using BioLit. *BMC Bioinformatics* **11**: 220.
23. 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.
24. P. W. Rose, B. Beran, C. Bi, W. F. Bluhm, D. Dimitropoulos, D. S. Goodsell, A. Prlic, M. Quesada, G. B. Quinn, J. D. Westbrook, J. Young, B. Yukich, C. Zardecki, H. M. Berman, P. E. Bourne. (2011) The RCSB Protein Data Bank: redesigned website and web services. *Nucleic Acids Res.*
25. Enzyme Nomenclature, Enzyme Classification. http://www.chem.qmw.ac.uk/iubmb/enzyme.
26. T. A. Tatusova, T. L. Madden. (1999) BLAST 2 Sequences, a new tool for comparing protein and nucleotide sequences. *FEMS Microbiol. Lett.* **174**: 247-250.
27. S. B. Needleman, C. D. Wunsch. (1970) A general method applicable to the search for similarities in the amino acid sequence of two proteins. *J. Mol. Biol.* **48**: 443-453.
28. T. F. Smith, M. S. Waterman. (1981) Identification of common molecular subsequences. *J. Mol. Biol.* **147**: 195-197.
29. Y. Ye, A. Godzik. (2003) Flexible structure alignment by chaining aligned fragment pairs allowing twists. *Bioinformatics* **19**: ii246-255.
30. I. N. Shindyalov, P. E. Bourne. (1998) Protein structure alignment by incremental combinatory extension of the optimum path. *Protein Eng.* **11**: 739-747.
31. A. R. Ortiz, C. E. Strauss, O. Olmea. (2002) MAMMOTH (matching molecular models obtained from theory): an automated method for model comparison. *Protein Sci.* **11**: 2606-2621.
32. Y. Zhang, J. Skolnick. (2005) TM-align: a protein structure alignment algorithm based on the TM-score. *Nucleic Acids Res.* **33**: 2302-2309.
33. M. J. Sippl, M. Wiederstein. (2008) A note on difficult structure alignment problems. *Bioinformatics* **24**: 426-427.
34. M. J. Sippl. (2008) On distance and similarity in fold space. *Bioinformatics* **24**: 872-873.
35. R. C. Holland, T. A. Down, M. Pocock, A. Prlic, D. Huen, K. James, S. Foisy, A. Drager, A. Yates, M. Heuer, M. J. Schreiber. (2008) BioJava: an open-source framework for bioinformatics. *Bioinformatics* **24**: 2096-2097.
36. J. Yuvaniyama, J. M. Denu, J. E. Dixon, M. A. Saper. (1996) Crystal structure of the dual specificity protein phosphatase VHR. *Science* **272**: 1328-1331.
37. M. Soundararajan, F. S. Willard, A. J. Kimple, A. P. Turnbull, L. J. Ball, G. A. Schoch, C. Gileadi, O. Y. Fedorov, E. F. Dowler, V. A. Higman, S. Q. Hutsell, M. Sundstrom, D. A. Doyle, D. P. Siderovski. (2008) Structural diversity in the RGS domain and its interaction with heterotrimeric G protein alpha-subunits. *Proc. Natl. Acad. Sci. USA* **105**: 6457-6462.
38. P. E. Bourne, B. Beran, C. Bi, W. Bluhm, R. Dunbrack, A. Prlic, G. Quinn, P. Rose, R. Shah, W. Tao, B. Weitzner, B. Yukich. (2010) Will Widgets and Semantic Tagging Change Computational Biology? *PLoS Comput. Biol.* 6: e1000673.
39. K. Tars, M. Bundule, K. Fridborg, L. Liljas. (1997) The crystal structure of bacteriophage GA and a comparison of bacteriophages belonging to the major groups of Escherichia coli leviviruses. *J. Mol. Biol.* **271**: 759-773.
40. K. Alden, S. Veretnik, P. E. Bourne. (2010) dConsensus: a tool for displaying domain assignments by multiple structure-based algorithms and for construction of a consensus assignment. *BMC Bioinformatics* **11**: 310.
41. D. S. Goodsell. (2010) Getting the most out of the Protein Data Bank. *The Journal of Biocommunication* **35**: e52-e57.
42. C. Zardecki. (2008) Interesting structures: Education and outreach at the RCSB Protein Data Bank. *PLoS Biol.* **6**: e117.
43. S. Dutta, C. Zardecki, D. Goodsell, H. M. Berman. (2010) Promoting a structural view of biology for varied audiences: an overview of RCSB PDB resources and experiences. *J. Appl. Cryst.* **43**: 1224-1229.
44. L. Chen, R. Oughtred, H. M. Berman, J. Westbrook. (2004) TargetDB: A target registration database for structural genomics projects. *Bioinformatics* **20**: 2860-2862.
45. 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.
46. 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.
47. PMV: Python Molecular Viewer. http://mgltools.scripps.edu.
48. 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.
49. C. L. Lawson, M. L. Baker, C. Best, C. Bi, M. Dougherty, P. Feng, G. van Ginkel, B. Devkota, I. Lagerstedt, S. J. Ludtke, R. H. Newman, T. J. Oldfield, I. Rees, G. Sahni, R. Sala, S. Velankar, J. Warren, J. D. Westbrook, K. Henrick, G. J. Kleywegt, H. M. Berman, W. Chiu. (2011) EMDataBank.org: unified data resource for cryoEM. *Nucleic Acids Res.*


# RCSB PDB

## wwPDB Members

RCSB PDB
www.pdb.org

Protein Data Bank Europe
pdbe.org

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.

# MOLECULE OF THE MONTH ILLUSTRATIONS

Images of the structures below are from past issues of the *Molecule of the Month*. They were created using the PDB entries listed.

**antibodies:** 1igt, L. J. Harris, S. B. Larson, K. W. Hasel, A. McPherson. (1997) Refined structure of an intact IgG2a monoclonal antibody. *Biochemistry* **36**: 1581-1597.

**cholera toxin**: 1xtc, R. G. Zhang, D. L. Scott, M. L. Westbrook, S. Nance, B. D. Spangler, G. G. Shipley, E. M. Westbrook. (1995) The three-dimensional crystal structure of cholera toxin. *J. Mol. Biol.* **251**: 563-573.

**collagen**: 1cag, J. Bella, M. Eaton, B. Brodsky, H. M. Berman. (1994) Crystal and molecular structure of a collagen-like peptide at 1.9 Å resolution. *Science* **266**: 75-81.

**designed DNA crystal**: 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.

**DNA**: H. R. Drew, R. M. Wing, T. Takano, C. Broka, S. Tanaka, K. Itakura, R. E. Dickerson. (1981) Structure of a B-DNA dodecamer: conformation and dynamics. *Proc. Natl. Acad. Sci. USA* **78**: 2179-2183.

**estrogen receptor**: top, 1hcq, J. W. Schwabe, L. Chapman, J. T. Finch, D. Rhodes. (1993) The crystal structure of the estrogen receptor DNA-binding domain bound to DNA: how receptors discriminate between their response elements. *Cell* **75**: 567-578; bottom, 1a52, D. M. Tanenbaum, Y. Wang, S. P. Williams, P. B. Sigler. (1998) Crystallographic comparison of the estrogen and progesterone receptor's ligand binding domains. *Proc. Natl. Acad. Sci. USA* **95**: 5998-6003.

**phosphofructokinase**: 4pfk, P.R. Evans, G.W. Farrants, P.J. Hudson (1981) Phosphofructokinase: structure and control. *Philos.Trans.R.Soc.London, Ser.B* 293: 53-62.

**photosystem I**: 1jb0, P. Jordan, P. Fromme, H. T. Witt, O. Klukas, W. Saenger, N. Krauss. (2001) Three- dimensional structure of cyanobacterial photosystem I at 2.5 Å resolution. *Nature* **411**: 909-917.

**HIV-1 Protease**: 3hvp, A. Wlodawer, M. Miller, M. Jaskolski, B. K. Sathyanarayana, E. Baldwin, I. T. Weber, L. M. Selk, L. Clawson, J. Schneider, S. B. Kent. (1989) Conserved folding in retroviral proteases: crystal structure of a synthetic HIV-1 protease. *Science* **245**: 616-621.

**rhinovirus**: 1rvf, T. J. Smith, E. S. Chase, T. J. Schmidt, N. H. Olson, T. S. Baker. (1996) Neutralizing antibody to human rhinovirus 14 penetrates the receptor-binding canyon. *Nature* **383**: 350-354.

**ribosome**: 1ffk, N. Ban, P. Nissen, J. Hansen, P. B. Moore, T. A. Steitz. (2000) The complete atomic structure of the large ribosomal subunit at a 2.4 Å resolution. *Science* **289**: 905-920; 1fka, F. Schluenzen, A. Tocilj, R. Zarivach, J. Harms, M. Gluehmann, D. Janell, A. Bashan, H. Bartels, I. Agmon, F. Franceschi, A. Yonath. (2000) Structure of functionally activated small ribosomal subunit at 3.3 Å resolution. *Cell* **102**: 615-623.

**thrombin**: 1ppb, W. Bode, I. Mayr, U. Baumann, R. Huber, S. R. Stone, J. Hofsteenge. (1989) The refined 1.9 Å crystal structure of human alpha-thrombin: interaction with D-Phe-Pro-Arg chloromethylketone and significance of the Tyr-Pro-Pro-Trp insertion segment. *EMBO J.* **8**: 3467-3475.