PDB NEWSLETTER

NUMBER 11 ◆ FALL 2001 ◆ RCSB

PUBLISHED QUARTERLY BY THE RESEARCH COLLABORATORY FOR STRUCTURAL BIOINFORMATICS

CONTENTS

SNAPSHOT: SEPT. 30, 2001

16,121 released atomic coordinate entries

MOLECULE TYPE

14,427	proteins, peptides, and viruses
1,006	nucleic acids
670	protein/nucleic acid complexes
18	carbohydrates

EXPERIMENTAL TECHNIQUE

13,302	diffraction and other
6,031	structure factor files
2,477	NMR
1,142	NMR restraint files
342	theoretical modeling

RCSB

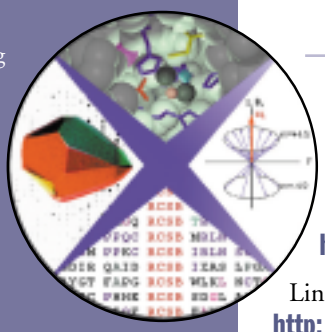
SDSC: www.rcsb.org

RUTGERS: rutgers.rcsb.org

NIST: nist.rcsb.org

E-MAIL: info@rcsb.org

FTP: ftp.rcsb.org



MESSAGE FROM THE PDB

One of the goals of the PDB is to make the archive as consistent and error-free as possible. The PDB's Data Uniformity Project enhances the consistency of existing (legacy) entries and maintains a consistent method of annotating current depositions. Recently, the PDB archive has been standardized and released in mmCIF format.

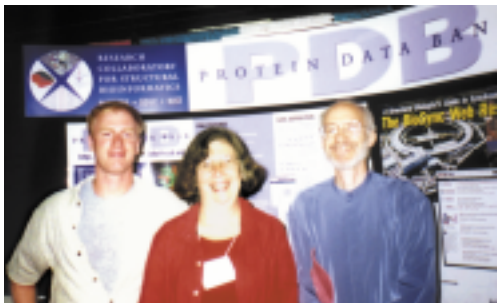
One focus of this work has been to resolve any inconsistencies between the specification of the chemical sequence and the sequence that is inferred from the deposited coordinate data. Another focus of this work was to include in the mmCIF data files the results of prior uniformity processing of individual PDB records. The standardized data for records such as compound name, citation, and source organism were previously accessible from the PDB database, but this information was not available in all of the data files. The mmCIF data files include the integration of all of this information, as well as additional macromolecular names and synonyms from related SwissProt sequence database entries.

All PDB entries are available in mmCIF format from the PDB beta FTP site at **ftp://beta.rcsb.org/pub/pdb/uniformity/data/mmCIF/**. The files follow the latest version of the mmCIF dictionary supplemented by an exchange dictionary developed by the PDB and the EBI. This exchange dictionary can be obtained from **http://deposit.pdb.org/mmcif/**.

An application program called CIFTr was made available for translating files in mmCIF format into files in PDB format. Further information on this program is available in this newsletter.

Comments on this, and all aspects of the PDB, are welcome at **info@rcsb.org**.

The PDB ✦



Director Helen M. Berman visits with former student Luke Rooney and UCSD Associate Vice Chancellor for Research, John Wooley at the ACA meeting in Los Angeles. See story on page 4.

The Research Collaboratory for Structural Bioinformatics (RCSB) is a non-profit consortium dedicated to improving our understanding of biological macromolecular structure.

Weekly PDB news is available on the Web at
http://www.rcsb.org/pdb/latest_news.html

Links to this and previous PDB newsletters are available at
http://www.rcsb.org/pdb/newsletter.html

Data Deposition and Processing

PDB Deposition Statistics

As we begin the final quarter of the year 2001, 2,422 structures have been deposited to the PDB this year—with 873 deposited in the last quarter.

Approximately 70% were deposited with a HPUB release status; 12% were deposited with a HOLD status; and 18% were deposited and released as soon as the annotation of the entry was completed. The release of HPUB structures is described in this newsletter.

Of these depositions, 80% were determined from X-ray diffraction experiments. 16% were determined from NMR studies.

CIFTr and Data Uniformity Files Released

An application program that can translate files between mmCIF format and PDB formats has been released. Called CIFTr, this program also provides the option of producing a file with a blank chain ID field for structures with a single chain, and the option of producing files with standard IUPAC hydrogen nomenclature for standard L-amino acids.

CIFTr was released this summer along with the files from the Data Uniformity Project (see Message from the PDB for more information). These files (in mmCIF format) can be accessed from the PDB beta FTP site at **ftp://beta.rcsb.org/pub/pdb/uniformity/data/mmCIF/**. Further information about the Data Uniformity project is at **http://www.rcsb.org/pdb/uniformity/**.

CIFTr works on UNIX platforms and can be downloaded at **http://pdb.rutgers.edu/software/**. Questions about this program may be sent to **info@rcsb.org**.

PDB Focus: How Are HPUB Structures Released?

The release status of structures is determined at the time of deposition by the author. The status HPUB is used to indicate that a structure will be released when the corresponding journal article is published. Publication is considered to be when the article is distributed by the publisher, either in print or electronically. Structures are released when the PDB can confirm that the article corresponds with the entry.

The PDB receives publication dates and citation information from some journals. For other journals, the PDB scans the literature for publication information. We also greatly appreciate the citation information that is sent to us at deposit@rcsb.rutgers.edu from the community.

ADIT Reaches One Year Mark at Osaka University

The ADIT deposition and annotation site established at the Institute for Protein Research at Osaka University in Osaka, Japan has been operational for a year. Entries deposited at this site have been processed by staff at the Laboratory of Protein Informatics (Head, Professor Haruki Nakamura) at the Institute for Protein Research at Osaka University. Under the direction of Drs. Masami Kusunoki and Genji Kurisu, these entries are processed by Takashi Kosada, Reiko Igarashi and Yumiko Kengaku using ADIT, and are incorporated into the PDB archive.

Picture of Yumiko Kengaku, Takashi Kosada, and Reiko Igarashi (Institute for Protein Research) taken by Kyle Burkhardt on a recent visit to Osaka University. Takashi, wearing his souvenir New Jersey shirt, has also visited the RCSB-Rutgers site.

Part of the success of this cooperative agreement is due to the productive visits that the RCSB and Osaka group members have made to both sites. We look forward to continuing collaborations with this group.

In addition to the ADIT Osaka site at **http://pdbdep.protein.osaka-u.ac.jp/adit/**, depositions to the PDB can also be made at the RCSB-Rutgers site (ADIT; **http://pdb.rutgers.edu/adit/**) and at the European Bioinformatics Institute (AutoDep; **http://autodep.ebi.ac.uk/**).

Data Query, Reporting, and Access

Redundancy Reduction Capability Now Available on the PDB Web Site

The Protein Data Bank holdings contain considerable redundancy in sequence and structure. An option that allows users to select a subset of structures from which homologous sequences have been largely removed is now available from the primary PDB Web site and its mirrors. This option, which is available from all search interfaces, filters subsets of structures that match a particular query.

Removing sequence homologues from queries via the home page and SearchLite returns representatives of protein structures with less than 90% sequence similarity. SearchFields provides the option of selecting either 50, 70, or 90% similarity as cut-off values. The user can then toggle between the complete set of results and the reduced subset by using the options menu at the top of the Query Result Browser.

While sequence homology is defined on a per chain basis, results are returned on a structure basis. Results may differ from other

non-redundant sets outside the PDB. The CD-HIT algorithm (Cluster Database at High Identity with Tolerance) is used to remove redundant sequences and leave only the representatives (Li, W., Jaroszewski, L. and Godzik, A.; *Bioinformatics*, (2001) 17:282-283). CD-HIT can be found at **http://bioinformatics.ljcrf.edu/cd-hi/**.

Further information about this new feature is available at **http://www.rcsb.org/pdb/redundancy.html**. Questions or comments on this feature may be sent to **info@rcsb.org**.

Enhanced Access to Primary PDB Web Site

PDB users can now enjoy further enhanced access to the primary PDB Web site at **http://www.rcsb.org/pdb/** thanks to enhanced connectivity at the SDSC-PDB site. SDSC now has 40 Mbits of exclusive Internet bandwidth, almost doubling its Internet capacity, with connectivity to two independent sources. This will provide even more reliable access to this PDB site.

Hardware Upgrades to PDB System

In order to provide the best access possible to the user community, the systems serving the primary distribution site have been upgraded. Two pairs of load-balanced Enterprise class Sun servers now administer the main Web server (**http://www.rcsb.org/pdb**) and the FTP server (**ftp://ftp.rcsb.org/**).

These redundant systems have been installed to ensure access—even in the case of a hardware failure. These enhancements, in conjunction with the two independent network paths that now provide Web access to the sites, will allow PDB users to enjoy even more robust connectivity to PDB's resources.

Database of NIGMS Protein Structure Initiative Target Sequences

Under the sponsorship of the NIGMS, the PDB has created a centralized registration database for target sequences from the NIH P50 structural genomics projects at **http://targetdb.pdb.org/**.

Target sequences and status information are collected weekly from the originally funded NIH structural genomics centers: the Berkeley Structural Genomics Center, the Joint Center for Structural Genomics, the Midwest Center for Structural Genomics, the Northeast Structural Genomics Consortium, the New York Structural Genomics Research Consortium, the Southeast Collaboratory for Structural Genomics, and the Tuberculosis Structural Genomics Consortium. The new P50 centers (the Center for Eukaryotic Structural Genomics and the Structural Genomics of Pathogenic Protozoa Consortium) will be added in the near future.

The target database can be searched by sequence using FASTA (Pearson, W.R. and Lipman, D.J. (1988) "Improved tools for biological sequence comparison" *PNAS* 85:2444-2448). Sequence searches may include only the P50 target sequences or the P50 and PDB sequences. Target sequences may also be searched by contributing P50 site, protein name, project tracking identifier, date of last modification, and the current status of the target (e.g. cloned, expressed, crystallized, ...). Search results may be viewed as HTML reports, FASTA data files, or in XML.

Target data for all of the NIH projects can be downloaded as an XML document. The XML document is organized following the recommendations of the International Task Forces on Target Tracking (see **http://www.nigms.nih.gov/news/meetings/airlie.html** for more information). This document type definition for the target data file can be retrieved from **http://targetdb.pdb.org/apps/target.dtd**.

PDB Web Site Statistics

A glance at the access statistics for the primary PDB Web site at **http://www.rcsb.org/pdb/** shows that the Web site hits received and files downloaded have resumed their frequency with the start of the new academic year.

The **http://www.rcsb.org/pdb/** address continues to receive the most traffic, though use of the mirror sites and beta test site is ever increasing. As always, PDB users are encouraged to access their most proximate RCSB mirror site at Rutgers (**http://rutgers.rcsb.org/**), NIST (**http://nist.rcsb.org/**), the Cambridge Crystallographic Data Centre in the United Kingdom (**http://pdb.ccdc.cam.ac.uk/**), the National University of Singapore (**http://pdb.bic.nus.edu.sg/**), Osaka University in Japan (**http://pdb.protein.osaka-u.ac.jp/**), or the Universidade Federal de Minas Gerais in Brazil (**http://www.pdb.ufmg.br/**). These sites are directly accessible from the PDB home page.

Users are also invited to preview new features at the PDB beta test site, accessible at **http://beta.rcsb.org/pdb/**. We appreciate your feedback!

Access Statistics for **www.rcsb.org**

	DAILY AVERAGE			MONTHLY TOTALS		
MONTH	HITS	FILES	SITES	KBYTES	FILES	HITS
Sept 01	135,709	108,309	59,052	113,485,643	3,249,277	4,071,270
Aug 01	93,011	73,385	52,087	68,849,896	2,274,960	2,883,358
July 01	85,948	66,336	27,070	60,512,717	2,056,420	2,664,390

PDB Outreach

PDB Director Speaks at Research Subcommittee Hearing

Helen M. Berman, Director of the PDB, spoke at a hearing held by the House Science Subcommittee on Research that examined the impact federal investment has had on promoting innovation in information technology.

At this session, "Innovation in Information Technology: Beyond Faster Computers and Higher Bandwidth", Prof. Berman described how the developments in technology have influenced the growth of the PDB. She also described how the way the PDB archives and distributes data has changed as computer and information technologies have advanced.

Details of the hearing are available at **http://www.house.gov/science/research/reshearings.htm**.

PDB Goes to ECM-20, ISMB, Protein Society, and ACA

Thanks to all the conference participants who visited the PDB exhibit booths and posters at various meetings this summer:

- **American Crystallographic Association's Annual Meeting** (July 21-26, Los Angeles, CA) The PDB's exhibit booth and users lunch were great successes.
- **Intelligent Systems for Molecular Biology** 9th International Conference (July 21-25, Copenhagen, Denmark) The PDB participated in the exhibition at this event.
- **15th Symposium of the Protein Society** (July 28-August 1, Philadelphia, PA) The PDB met with several users during the PDB's poster sessions.
- **20th European Crystallographic Meeting** (August 25-31, Krakow, Poland) A talk and PDB poster were presented.

We look forward to seeing you at future meetings.

Archival PDB Newsletters Available

While the PDB Newsletter was started in 1974, only issues dating back to 1993 have been accessible on-line until recently. Early paper copies have been scanned and are now available at **ftp://ftp.rcsb.org/pub/pdb/doc/newsletters/old_bnl/** in PDF format (see **http://www.adobe.com/products/acrobat/alternate.html** for a free reader).

Sixty-three PDB newsletters, ranging from September 1974 to January 1993, are included in this set. Changes in technology become very evident as you read through these newsletters. The earliest were prepared on a typewriter and some by pasting sections together. The later newsletters resemble the printed versions of today.

The history of the Protein Data Bank—the growth of the resource, the means of delivery of the data, and the evolution of standard formats—can be traced through these newsletters. We hope you enjoy them.

PDB Chapter Published in International Tables

A chapter describing the PDB's systems for the data resource has been published in the *International Tables for Crystallography*—"The Protein Data Bank, 1999–". H.M. Berman, J. Westbrook, Z. Feng, G.L. Gilliland, T.N. Bhat, H. Weissig, I.N. Shindyalov, and P.E. Bourne. *International Tables for Crystallography*. Volume F: Crystallography of Biological Macromolecules. M.G. Rossmann and E. Arnold, Editors. Dordrecht: Kluwer Academic Publishers. The Netherlands.

Also included are chapters on the Nucleic Acid Database, The Cambridge Structural Database, the Biological Macromolecule Crystallization Database, and the history of the PDB at Brookhaven.

Further information about the *International Tables* is available from the IUCr at **http://www.iucr.ac.uk/iucr-top/it/index.html**.

Structural Genomics Page Updated

The PDB frequently updates its Structural Genomics page at **http://www.rcsb.org/pdb/strucgen.html.** The purpose of this page, which is a compilation of structural genomics links, is to provide an entry point to additional information on structural genomics relevant to PDB users. A recent addition is the Target Registration Database, which is also described in this newsletter.

We appreciate the links and other structural genomics-related developments that have been sent to us by PDB users. Additional information for this page may be sent to the PDB by sending e-mail to **info@rcsb.org**.

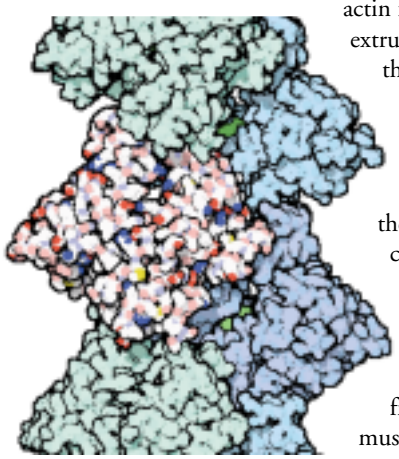
PDB Focus: Molecule of the Month

The Molecule of the Month series is a wonderful collection of short columns featuring a new PDB structure of interest each month. They describe the functions and significance of the selected biological macromolecules for a general audience, providing a basic understanding of structural interactions. Written and illustrated by Dr. David S. Goodsell of the Scripps Research Institute, this feature adds a unique aesthetic quality and informative educational resource to the PDB Web site. You can access the Molecule of the Month installations at **http://www.rcsb.org/pdb/molecules/molecule_list.html.**

Below is a sample of the information that was presented in this feature during the past quarter:

Actin: Dynamic Molecular Infrastructure

July, 2001—The complex ultrastructure of cells—their shape and internal structure—and the many motions of cells are largely supported by filaments of actin. A tangle of cross-linked actin filaments fills the cytoplasm of animal, plant and fungal cells, forming a "cytoskeleton" that gives the cell shape and form and provides a scaffold for organization. Tightly bundled actin filaments provide a sturdy backbone to extrude structures from the cell surface, such as the pseudopods used by amoebas for crawling and the finger-like microvilli of intestinal cells, which extend into the digestive tract and absorb nutrients. Actin also forms the ladder on which myosin climbs, providing the infrastructure for muscle contraction and creating the motion that we experience in our daily lives. Actin is plentiful throughout the body as it performs these basic structural tasks: it may comprise 5 percent of the protein in a typical cell, or up to one fifth of the protein in special cases, such as muscle cells.



PDB ID: **1atn**

W. Kabsch, H.G. Mannherz, D. Suck, E.F. Pai, K.C. Holmes (1990): Atomic structure of the actin: DNase I complex. Nature *347, p. 37.*

Actin has a rare combination of strength and sensitivity. Actin filaments are used in many of the most strenuous structural tasks, but at the same time, actin filaments are easily and continually disassembled. One of the great hall-

marks of actin is its dynamic character. Actin filaments are continually built and broken down as the needs of the cell change from moment to moment. In special cases, such as muscle actin or the actin bundles in microvilli, a collection of specialized actin-binding proteins stabilize the filament, forming a more permanent structure. But the bulk of actin in typical cells is in constant flux, constantly forming filaments and breaking down for each new task.

The dynamic character of actin is controlled by a molecule of ATP bound to each actin monomer. The state of this ATP determines the stability of the actin filament. Free actin typically holds an ATP molecule and binds tightly to growing filaments. After attaching, the ATP is broken and the actin subtly changes shape. This new form, with ADP bound, is not as stable in the filament and dissociates more easily. One of the unusual consequences of this behavior is "treadmilling." An actin filament will be continually building at one end, where new actin-ATP complexes are forming strong new connections, and at the same time slowly falling apart at the opposite end, where the actin-ADP form has weakened connections. Imagine the filament growing at one end and dissolving at the other, so that the whole structure slowly steps through the cell but never gets any longer or shorter.

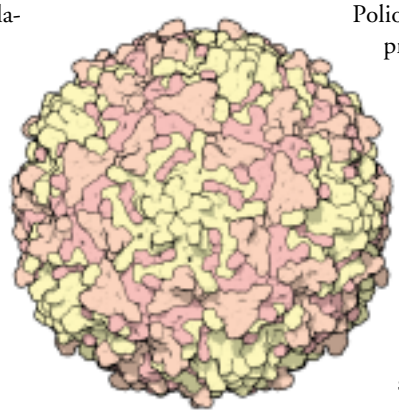
Of course, cells cannot have actin filaments growing uncontrollably all over the cytoplasm. The poison phalloidin, from the death cap mushroom, demonstrates what would happen. It promotes the growth of actin and ultimately clogs the cell with rigid actin filaments, causing fatal liver and kidney damage in unwary mushroom lovers. In cells, a variety of actin-severing proteins control the growth of actin, ensuring that the filaments grow only when needed. Two of these actin watchdogs are gelsolin and profilin. Gelsolin, from PDB entry **1yvn**, breaks actin filaments into short lengths when the level of calcium rises. Then, it remains bound to the end, blocking additional growth. Profilin, from PDB entry **1hlu**, binds to free actin and keeps it from adding to filaments, also inhibiting growth. Both bind to the actin monomer at a similar location, blocking part of the site that binds to neighboring actin molecules in a filament.

Large helical protein assemblies, such as actin filaments, are notoriously difficult to study by crystallography, because the filaments do not form perfect crystals. The structures of actin in the PDB all have something bound to them, blocking formation of a filament, so the structures contain only a single actin molecule, not an entire actin filament. PDB entry **1atn**, contains a DNA-cutting enzyme that just happens to bind to actin. Actin is a U-shaped molecule with ATP bound deep in the groove between the two arms. PDB entry **1alm**, presents a model of one myosin motor bound to a short actin filament formed of five molecules, based on data from electron microscopy. The file contains only alpha carbon positions for the proteins, so you'll need to use backbone diagrams when you look at it.

A list of all actin structures in the PDB as of July, 2001, is available at **http://www.rcsb.org/pdb/molecules/pdb19_report.html**. For more information about actin, see **http://www.rcsb.org/pdb/molecules/pdb19_4.html**.

Poliovirus and Rhinovirus: Little RNA Viruses

August, 2001—Viruses are biological hijackers. They attack a living cell and force it to make many new viruses, often destroying the cell in the process. Picornaviruses, or "little RNA viruses," are among the most simple viruses. They are composed of a modular protein shell, which seeks out and binds to a target cell surface, surrounding a short piece of RNA, which contains all of the information needed to co-opt the cell's machinery and direct the construction of new viruses. In spite of their simplicity, or perhaps because of it, the picornaviruses are also among the most important viruses for human health and welfare. Three familiar examples are: poliovirus (PDB entry **2plv**), rhinovirus (PDB entry **4rhv**), and the virus that causes foot and mouth disease in livestock (PDB entry **1bbt**).



PDB ID: ***4rhv***

E. Arnold, M.G. Rossmann (1988): The use of molecular-replacement phases for the refinement of the human rhinovirus 14 structure. Acta Crystallogr A *44, p. 270.*

Poliovirus and rhinovirus have specialized to attack primarily human beings, but they use two different approaches. Poliovirus, which is found in three similar forms, is designed to attack a given person only once. It makes its offspring and then is off to the next person. In most cases, poliovirus causes a simple flu-like disease as it attacks cells in the digestive system. This infection is rapidly cleared up by the immune system. But in about 1 in 100 cases, the virus spreads to the nerve cells that control muscle motion, causing paralysis—polio myelitis—as the nerve cells are infected.

Rhinovirus, on the other hand, is found in many different forms that attack a given person many times during their life. Each time you get a cold, a different form of rhinovirus (or occasionally, a different type of virus) is attacking. Your body learns how to fight it off, but you are still susceptible to the next form. On average, a person will have a new cold once every two years, so most of us are quite familiar with the symptoms of rhinovirus infection in our nose and respiratory tract. Because they are so simple, picornaviruses can be very stable. Rhinovirus can last for days on your hands and still be infectious. And because the virus is shed from infected people all through the period with symptoms and even for days after, it spreads effectively through contact from person to person.

Antibodies are our major defense against these small, efficient viruses. Vaccines prime the immune system with antibodies, making it ready to fight an infection. In the case of poliovirus, there are two types of vaccines. One is a killed version of the virus, which is slowly killed with formaldehyde over the course of several days so that it is inactivated, but still keeps its proper shape. The second is a weakened, but still live, strain of the virus that has been artificially bred to stimulate the immune system without causing disease. The immune system responds by making anti-

bodies to fight these weakened viruses, leaving it ready to fight the real thing when it comes along.

The polio vaccines are one of the triumphs of modern medicine, but many people would say that the lack of a cure for the common cold is one of the great failings. The difficulty of creating a vaccine for the common cold lies in the diversity of rhinovirus. Over one hundred types of rhinovirus have been discovered as they strike people around the world, and new strains appear continually. Rhinovirus is a moving target that is not effectively combated with a single vaccine. Antiviral drugs, however, are a possible solution.

Many viruses, including the picornaviruses and bacteriophage **phiX174** (discussed in an earlier Molecule of the Month), are icosahedral in shape. They are composed of 60 identical pieces that form a perfectly symmetrical shell, termed a capsid, around the viral genome. In the case of poliovirus and rhinovirus, the shell is composed of 60 copies of four different proteins for a total of 240 protein chains in all. These proteins are carefully designed to be stable, but not too stable.

They must be fairly sturdy to allow the virus to pass from host to host through the hostile environment. But at the same time, they must be able to fall apart when they enter the cell, releasing the RNA inside. A carefully orchestrated set of structural changes occur as the virus attaches to the surface of the cell and is drawn inside, allowing the virus to deliver its RNA into the unwitting host.

The RNA protected inside the capsid is seen only as a blurry tangle in these crystallographic structures. It is not as perfectly symmetrical as the many proteins in the shell. The rhinovirus genome, when analyzed by sequencing techniques, contains just enough information to direct the construction of eleven proteins. These include the four separate proteins for its capsid, another four proteins that replicate its RNA, two proteins to clip each of these proteins into the proper shape, and one additional protein with as-yet obscure function.

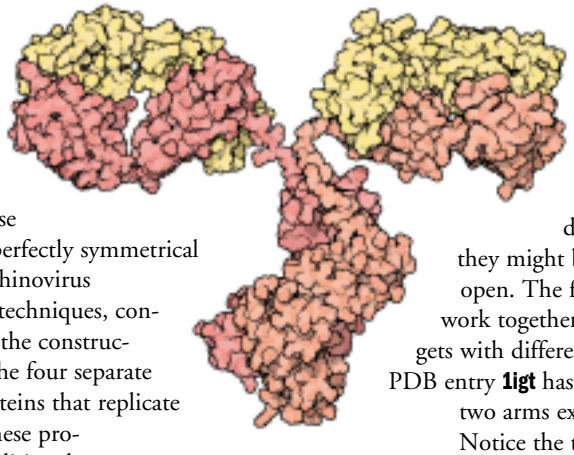
Antibodies bind to the surface of picornaviruses and stop them from attacking cells. Rhinovirus binds to a receptor protein on the cell surface. The receptor protein is gripped within a groove that encircles the five-fold symmetrical arrangement of proteins (known as the canyon in the picornavirus literature). Antibodies bind to the surface of rhinovirus and poliovirus in this same position and block their attachment to the surfaces of cells. PDB entry **1rvf** shows fragments of antibodies bound to rhinovirus. The intact antibodies are much larger than the small fragments seen here, so seven to ten antibodies are enough to form a bulky barrier on each virus to block attachment and infection.

Many structures of rhinovirus with antiviral drugs are available at the PDB, including the drug pleconaril, currently in clinical testing, (PDB entry **1c8m**). Most drugs act by blocking protein binding sites or destabilizing a key interaction. These drugs, on the other hand, may act differently. They actually stabilize the virus structure so that it cannot release its cargo of RNA. The drugs bind in a little pocket under the deep groove that grabs onto the cellular receptor. Normally, the binding of virus to receptor shifts the structure of the virus, ultimately allowing the virus to release RNA. The drug, however, glues the virus shut.

A list of all picornaviruses in the PDB as of August, 2001 is available at **http://www.rcsb.org/pdb/molecules/pdb20_report.html**. For more information on picornaviruses, see **http://www.rcsb.org/pdb/molecules/pdb20_5.html**.

Antibodies: Molecular Watchdogs

September, 2001—Antibodies are our molecular watchdogs, waiting and watching for viruses, bacteria and other unwelcome visitors. Antibodies circulate in the blood, scrutinizing every object that they touch. When they find an unfamiliar, foreign object, they bind tightly to its surface. In the case of viruses, like rhinovirus or poliovirus presented last month in the Molecule of the Month, a coating of bound antibodies may be enough to block infection. Antibodies alone, however, are no match for bacteria. When antibodies bind to a bacterial surface, they act as markers alerting the other powerful defensive mechanisms available in the immune system.



PDB ID: ***1igt***

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

Antibodies, and many of the other molecules used in the immune system, have a distinctive shape. Typically, they are composed of several flexible arms with binding sites at the end of each one. This makes perfect sense: since antibodies do not know in advance what attackers they might be fighting, they keep their options open. The flexible arms allow the binding sites to work together, grabbing with both arms onto targets with different overall shapes. The antibody in PDB entry **1igt** has two binding sites, at the tips of the two arms extending right and left at the top. Notice the thin, flexible chains that connect these arms to the central domain at the bottom. Some antibodies have longer flexible linkers connecting the arms together, allowing them even more latitude when finding purchase on a surface. Other antibodies have four or ten binding sites, so each contact can be weaker and still allow the whole antibody to bind firmly.

Your blood contains upwards of 100,000,000 different types of antibodies. Each type binds to a different target molecule. Remarkably, all of these antibodies are created before they ever see a virus or bacterium. You don't make a special antibody when a virus or bacterium infects your body. Instead, all of your antibodies are pre-fabricated, lying in wait until a virus or bacterium attacks. There are so many different kinds of antibodies that one or two are bound to be the right ones to fight the infection.

This amazingly huge collection of antibodies is created by the recombination of genes in lymphocytes, the blood cells that make antibodies. Each lymphocyte creates a different type of antibody,

based on how it has recombined its antibody genes. When an antibody encounters a virus or bacterium, the appropriate lymphocytes will multiply, flooding the blood with the particular antibodies needed to battle the invader. These lymphocytes may also make small adjustments on the antibodies they produce, tailoring their antibodies to bind more tightly and more specifically.

Antibodies are composed of four chains, two long heavy chains and two shorter light chains. The specific binding site is found at the tips of the two arms, in a pocket formed between the light and heavy chain. The binding site is composed of several loops in the protein chain that have very different lengths and amino acid composition. Differences in these "hypervariable loops" form the many types of pockets in different antibodies, each of which bind specifically to a different target. The rest of the antibody--the rest of the arms and the large constant domain that ties the two arms together—is relatively uniform in structure, providing a convenient handle when antibodies interact with the rest of the immune system.

When a foreign molecule is found in the blood, many different antibodies may bind to it, attacking at different angles. Three different antibodies that bind to the protein lysozyme can be found in PDB entries **1fdl**, **3hfl**, and **3hfm**. These entries each include only one arm of the antibody (termed "Fab" for "antigen-binding fragment"), which has been separated from the antibody for ease in study. The antibodies pick entirely different binding sites on the small lysozyme molecule.

Researchers have used the incredible functional diversity of the immune system in a clever way: to design new enzymes. Enzymes work by easing molecules through a difficult chemical change. For instance, take the Diels-Alder reaction. Two molecules come together, forming an unstable intermediate. Then, the intermediate falls apart, releasing sulfur dioxide and forming the desired product. Enzymes act by stabilizing the intermediate, smoothing the path from start to finish.

To make an antibody into an enzyme, we need to find an antibody that stabilizes this intermediate transition state in a similar way. Researchers have done this by finding antibodies that bind to a molecule that mimics the transition state. These antibody-enzymes are termed catalytic antibodies. The catalytic antibody shown in PDB entry **1c1e** performs the Diels-Alder condensation reaction. This is significant because this type of reaction is not performed by any natural enzymes. Antibodies that perform a number of other cleavage and condensation reactions, including reactions that are impossible any other way, may be found in the PDB.

Antibodies are very flexible, making it difficult to study an intact antibody. Most of the hundreds of antibody structures available at the PDB are fragments of antibodies, typically of just the Fab arm with the specific binding pocket. Three examples of intact antibodies are found in PDB entries **1igt**, **1igy**, and **1hzh**. All are nice examples for exploration. These entries show how antibodies are able to twist into different shapes, forced by packing into the different crystal lattices. This will give you some idea of the range of motion that these molecules are capable of as they bind to their targets.

A list of all antibodies in the PDB as of September, 2001 is available at **http://www.rcsb.org/pdb/molecules/pdb21_report.html**. For suggestions for further reading about antibodies, see **http://www.rcsb.org/pdb/molecules/pdb21_5.html**.

PDB Job Listings

PDB career opportunities are posted at **http://www.rcsb.org/pdb/jobs.html**. The current available openings are:

Systems and Applications Programmer

The Protein Data Bank at Rutgers University has a position open for an applications programmer to support and develop software for data processing operations at the Protein Data Bank. Programming areas include: macromolecular structure analysis and validation, molecular graphics, web application development, distributed object and relational database applications, and general scientific programming. Experience developing and maintaining object oriented software on UNIX platforms is required. Experience in the following is highly desirable: C/C++, JAVA, and CORBA. Please send resume to Dr. Helen Berman at **pdbjobs@rcsb.rutgers.edu**.

Biochemical Information Specialist

The Protein Data Bank at Rutgers University has a position open for a Biochemical Information Specialist to curate and standardize macromolecular structures for the Protein Data Bank. A background in biological chemistry, as well as some experience with UNIX-based computer systems, is required. Experience in crystallography and/or NMR spectroscopy is a strong advantage. The successful candidate should be well-motivated, able to pay close attention to detail, and meet deadlines. This position offers the opportunity to participate in an exciting project with significant impact on the scientific community. Please send resume to Dr. Helen Berman at **pdbjobs@rcsb.rutgers.edu.**

RCSB PDB Project Team Leaders

The overall operation of the RCSB PDB is managed by the RCSB Project Team Leaders. Technical and scientific support is provided by the RCSB PDB Members.

Dr. Helen M. Berman, Director
Rutgers University
berman@rcsb.rutgers.edu

Dr. Philip E. Bourne, Co-director
San Diego Supercomputer Center
bourne@sdsc.edu

Dr. Gary Gilliland, Co-director
National Institute of Standards and Technology
gary.gilliland@nist.gov

Dr. John Westbrook, Co-director
Rutgers University
jwest@rcsb.rutgers.edu

Protein Data Bank
http://www.pdb.org/

RCSB PDB Members

Rutgers University

Department of Chemistry and Chemical Biology
610 Taylor Road
Piscataway, NJ 08854-8087

Kyle Burkhardt
kburkhar@rcsb.rutgers.edu

Li Chen
lchen@rcsb.rutgers.edu

Victoria Colflesh
victoria@rcsb.rutgers.edu

Sharon Cousin
sharon@rcsb.rutgers.edu

Dr. Zukang Feng
zfeng@rcsb.rutgers.edu

Dr. Lisa Iype
lisa@rcsb.rutgers.edu

Dr. Shri Jain
sjain@rcsb.rutgers.edu

Jessica Marvin
jessica@rcsb.rutgers.edu

Gnanesh Patel
gnanesh@rcsb.rutgers.edu

Tania Rose Posa
tania@rcsb.rutgers.edu

Dr. Bohdan Schneider
bohdan@rcsb.rutgers.edu

Rosalina Valera
rosalina@rcsb.rutgers.edu

Christine Zardecki
zardecki@rcsb.rutgers.edu

SDSC

San Diego Supercomputer Center
UC San Diego
9500 Gilman Drive
La Jolla, CA 92093-0537

Dr. Peter Arzberger
parzberg@sdsc.edu

Bryan Banister
bryan@sdsc.edu

Tammy Battistuz
tammyb@sdsc.edu

Dr. Wolfgang F. Bluhm
wbluhm@sdsc.edu

Eliot Clingman
eliotc@sdsc.edu

Dr. Nita Deshpande
nita@sdsc.edu

Dr. Ward Fleri
ward@sdsc.edu

Dr. Douglas S. Greer
dsg@sdsc.edu

Dr. Michael Gribskov
gribskov@sdsc.edu

Dorothy Kegler
dkegler@sdsc.edu

David Padilla
dpadilla@sdsc.edu

Thomas Solomon
tsolomon@sdsc.edu

David Stoner
dstoner@sdsc.edu

Dr. Lynn F. Ten Eyck
teneyckl@sdsc.edu

Dr. Helge Weissig
helgew@sdsc.edu

NIST

Biotechnology Division
National Institute of Standards and Technology
100 Bureau Drive
Gaithersburg, MD 20899-8314

Dr. T.N. Bhat
bhat@nist.gov

Phoebe Fagan
phoebe.fagan@nist.gov

Dr. Diane Hancock
diane.hancock@nist.gov

Dr. Veerasamy Ravichandran
vravi@nist.gov

Dr. Narmada Thanki
narmada.thanki@nist.gov

Michael Tung
michael.tung@nist.gov

Dr. Greg Vasquez
gregory.vasquez@nist.gov

Statement of Support

The PDB is supported by funds from the National Science Foundation, the Office of Biological and Environmental Research at the Department of Energy, and two units of the National Institutes of Health: The National Institute of General Medical Sciences and the National Library of Medicine, in addition to resources and staff made available by the host institutions.