PDB NEWSLETTER

NUMBER 12 ◆ WINTER 2002 ◆ PUBLISHED QUARTERLY BY THE PROTEIN DATA BANK

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**

Contents

Snapshot: Jan. 1, 2002

16,972 released atomic coordinate entries

Molecule Type

15,200	proteins, peptides, and viruses
1,041	nucleic acids
713	protein/nucleic acid complexes
18	carbohydrates

Experimental Technique

14,019	diffraction and other
6,534	structure factor files
2,593	NMR
1,206	NMR restraint files
360	theoretical modeling

Participating RCSB Members

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

The year is off to a busy start for the PDB! In addition to handling the ever increasing data flow and continuing to provide new PDB features, PDB members have attended meetings in structural biology, structural genomics, computational biology and proteomics.

Talks were presented in the beginning of January all around the world: the Pacific Symposium on Biocomputing (January 3-7 in Mauna Lani, Hawaii), the Keystone Meeting "Structural Genomics: From Gene Sequence to Function" (January 5-11 in Breckenridge, CO) and the Cambridge Healthtech Institute's "Second Human Proteome Project Meeting" (January 7-10 in San Diego, CA). The PDB also participated in CCP4's "High-throughput Structure Determination Study Weekend (January 4-5 in York, UK). Some future events include:

A poster describing the PDB's ongoing efforts to make the data archive uniform will be presented at the Genome 9 meeting in Oakland, CA (January 27-31, 2002). These efforts are also described in the recent Nucleic Acids Research Database Issue ("The Protein Data Bank: unifying the archive" *Nucleic Acids Research*, 2002, Vol. 30, No. 1 245-248).

We will be exhibiting at the Biophysical Society's 46th Annual Meeting in San Francisco, CA (February 23-27, 2002)—we hope you will stop by booth 204 and say hello.

The PDB will also be exhibiting in an art gallery to introduce the beauty of protein structure to a general audience. Various representations of proteins found in the PDB will be on display in "The Art of Science" at the Gallery, a space dedicated to art exhibits at Rutgers University. The Art of Science will be on display from January 21 - February 9, 2002. For more information, please send email to **info@rcsb.org**.

Best wishes for 2002!

The PDB ✦

The Protein Data Bank (PDB) is the single worldwide repository for the processing and distribution of 3-D biological macromolecular structure data. The PDB is operated by Rutgers, the State University of New Jersey; the San Diego Supercomputer Center (SDSC) at the University of California, San Diego; and the National Institute of Standards and Technology (NIST) — three members of the Research Collaboratory for Structural Bioinformatics, a non-profit consortium dedicated to improving our understanding of biological systems.

MIRROR SITES

Cambridge Crystallographic Data Centre (UK): **http://pdb.ccdc.cam.ac.uk/**

National University of Singapore: **http://pdb.bic.nus.edu.sg/**

Osaka University (Japan): **http://pdb.protein.osaka-u.ac.jp/**

Universidade Federal de Minas Gerais (Brazil): **http://www.pdb.ufmg.br/**

Data Deposition and Processing

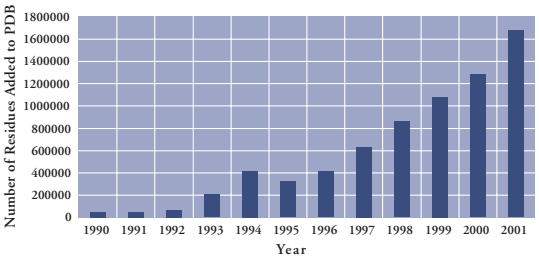
PDB Deposition and Growth Statistics for 2001

In 2001, 3,298 structures were deposited to the PDB, and were processed by teams at RCSB-Rutgers (75%), Osaka University (10%), and the European Bioinformatics Institute (14%).

Of the structures deposited, 72% were deposited with a release status of "hold until publication"; 16% were released as soon as annotation of the entry was complete; and 12% were held until a particular date. 82% of these entries were determined by X-ray crystallographic methods; 15% were determined by NMR methods.

The growth of the PDB can be seen in the increase in the number of residues added into the archive each year. 1,680,053 residues were released into the archive in 2001, as compared to 1,314,912 in 2000 and 1,068,340 in 1999.

GROWTH OF RESIDUES IN THE PDB



Alpha Version of ADIT Workstation Software Released

ADIT (AutoDep Input Tool) is the integrated software system used by PDB annotation staff for checking and editing PDB structure data entries.

A version of this software for workstation use has been released for alpha testing at **http://pdb.rutgers.edu/software/**. The system includes tools to help users check and prepare structure depositions. A file in mmCIF format can be created and then deposited to the PDB at **http://deposit.rcsb.org/adit/**.

The functionality of the workstation version is similar to that provided by the Web-based PDB deposition system (**http://deposit.rcsb.org/adit/**). The alpha version of the software is currently available in binary form for Linux platforms. Questions about this software may be sent to **help@rcsb.rutgers.edu**.

Improved Deposition for Cryo-electron Microscopy

PDB deposition systems have been extended to accept data items specifically describing cryo-electron microscopy (cryo-EM) methods, for those EM methods that generate fitted coordinates. These items were developed in collaboration with members of the cryo-EM community, the EBI-MSD group, and the PDB.

New versions of ADIT (**http://pdb.rutgers.edu/adit/** and **http://pdbdep.protein.osaka-u.ac.jp/adit/**) and AutoDep (**http://autodep.ebi.ac.uk**) that support these new data items for cryo-EM depositions are available.

Further information about AutoDep is available at **http://autodep.ebi.ac.uk/autodep-doc/html/release-notes.html**.

ADIT users can now select from the X-ray, NMR, or Cryo-EM deposition views.

Further information about ADIT is available at **http://pdb.rutgers.edu/**.

Cryo-EM data item definitions are available from the following sources:

http://deposit.rcsb.org/mmcif/dictionaries/mmcif_iims.dic
http://deposit.rcsb.org/mmcif/dictionaries/mmcif_iims.dic/Index/
http://iims.ebi.ac.uk/iims_dictionary/iims.dic
http://iims.ebi.ac.uk/cif_iims.dic/Groups/em_group.html

A PDB file format template for REMARKs is available at:

http://iims.ebi.ac.uk/3dem_pdb.html
http://deposit.rcsb.org/cryo-em.html

PDB Annotation Manual Online

The manual used as a guide by the PDB ADIT annotators for PDB Data Processing and Annotation is available online.

This document, a reference for the annotation staff, describes how the PDB data processing software system is used to produce the files that are released into the PDB archive. It is available in PostScript format from **http://www.rcsb.org/pdb/info.html#File_Formats_and_Standards** and can be viewed using readers such as Ghostview.

PDB Focus: CIFTr

CIFTr is an application program that translates files in mmCIF format into files in PDB format. CIFTr works on UNIX platforms, and can be downloaded at **http://pdb.rutgers.edu/software/**. CIFTr 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. These files, available 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 accessible at **http://www.rcsb.org/pdb/uniformity/**.

Data Query, Reporting, and Access

Alpha Version of OpenMMS Toolkit Released

The PDB has released an alpha version of the OpenMMS Toolkit, a suite of software tools that implement the Corba standard (OMG specification dtc/2001-04-06). The Corba specification provides a standard application programming interface (API) that allows remote programs direct access to the detailed experimental and macromolecular data available in the PDB.

More information about the PDB and Corba is available at **http://www.rcsb.org/pdb/newsletter/2001q1/corba.html**.

The OpenMMS Toolkit was developed to facilitate the use of macromolecular structure data by various scientific applications.

In addition to a reference implementation that demonstrates the functionality of the OMG Corba specifications, the toolkit also contains software for parsing data files and for creating and loading a relational database.

Compiled and source-only distributions of OpenMMS are available at **http://openmms.sdsc.edu/**. Questions may be sent to **info@rcsb.org**.

STING Millenium Suite Released on PDB Web Site

After a period of testing at the PDB beta test site, select components of STING Millennium Suite (SMS) are now available from the Structure Explorer pages of the primary PDB Web site and its mirrors. SMS, a set of Java-based tools for the simultaneous display of information about macromolecular structure and sequence, was developed by Dr. Goran Neshich of Embrapa-CNPTIA (Campinas, Brazil) and colleagues, in collaboration with Dr. Barry Honig's laboratory at Columbia University in New York City, NY. The SMS links from the PDB site are served by an SMS mirror that is being maintained at the San Diego Supercomputer Center.

The "Sequence Details" and "View Structure" sections of the Structure Explorer now link to two interactive structure and sequence SMS views for any PDB structure, which include options to access features such as a graphical display of amino acid contacts; these views require Chime and a Java-enabled Web browser. A simpler "Protein Dossier" view is also available from the "Sequence Details" section, offering a static graphical summary of sequence- and structure-based properties, such as relative entropy and temperature factors.

The "Geometry" section of the Structure Explorer now links to a Ramachandran plot for each PDB entry, also served from SMS, with options including the inter-connection of data in a dihedral angle plot with the 3-D structure of the molecule. This view also requires Java and Chime. SMS is also accessible from the "Other Sources" section of the Structure Explorer for each PDB entry, under the category of Visualization resources.

Further information about this suite of tools is available from the SMS home page at **http://mirrors.rcsb.org/SMS/**, and at **http://www.rcsb.org/pdb/help-results.html**. Comments may be sent to **info@rcsb.org**.

Phase Out of BNL FTP Archive

Since 1999, the Research Collaboratory for Structural Bioinformatics (RCSB) has maintained two distinct FTP sites: the RCSB Protein Data Bank (PDB) site at **ftp://ftp.rcsb.org/** (and its mirrors; see **http://www.rcsb.org/pdb/mirrors.html**), and the Brookhaven National Laboratory (BNL) PDB site at **ftp://bnlarchive.rcsb.org**.

In order to conserve resources and avoid confusions arising through the existence of two distinct PDB FTP sites, the RCSB will phase out the BNL PDB archive as of March 1, 2002, as announced on October 19, 2001 (**http://www.rcsb.org/pdb/lists/pdb-l/200110/msg00024.html**). This decision was made after consultation with the PDB Advisory Committee and review by members of the PDB user community.

The files available at **ftp://bnlarchive.rcsb.org/pub/resources/index/** are now available at **ftp://ftp.rcsb.org/pub/pdb/derived_data/index/**.

Current users of the BNL PDB archive are encouraged to consider the option of mirroring the RCSB FTP archive. The RCSB FTP archive can be found at **ftp://ftp.rcsb.org** and instructions for mirroring it can be found at **http://www.rcsb.org/pdb/ftpproc.final.html**.

A Perl script is provided to assist with conversion of existing BNL FTP directory structure to the RCSB FTP directory structure. Further information about the script is available at **ftp://ftp.rcsb.org/pub/pdb/software/bnl2rcsb.pl**.

Please send your questions or concerns, or requests for assistance regarding this change, to **info@rcsb.org**.

PDB Focus: Enzyme Queries Using SearchFields

Enzymes in the PDB have been classified in a hierarchical tree structure using the standards of the Enzyme Commission (EC). This classification permits users to search for enzymes by EC number or EC class/name through the SearchFields interface. It also allows users to "browse" through all enzymes using an Enzyme Browser interface, also available through SearchFields.

The PDB Enzyme Browser, available through the SearchFields interface

EC searches can be accomplished by selecting the "EC Number and Classification" option at the bottom of the SearchFields form, and then pressing the "New Form" button. New fields for "EC Number" and "Enzyme Class/Name" will then be available for searches based on these parameters. For example, searches for HIV-1 Protease in the PDB can be performed by entering "Retropepsin" in the "Enzyme Class/Name" field or by entering "3.4.23.16" in the "EC Number" field.

Selecting the "Browse and Select from Enzyme Classification" link under the "Enzyme Class/Name" box will launch the Enzyme Browser in a separate window. This interface allows users to navigate through different echelons of enzyme classification to arrive at a subset of particular interest. In the Enzyme Browser table, clicking on the "EC Number" column allows you to move up and down the EC tree. Clicking on a number in the "# of Structures in the PDB" column item gives you access to all those structures in the "Query Result Browser". Selecting the "use" link for any row will bring that EC Number and Enzyme Class/Name into the query form so that you can search the PDB database for that specific item.

The enzyme name/classification can be used only with the nomenclature or subclass exactly as given in the Enzyme Commission Nomenclature or substrings thereof.

Enzyme Nomenclature (1992) Recommendations of the Nomenclature Committee of the International Union of Biochemistry and Molecular Biology. Academic Press, New York.

Related information is available at the following Web sites:

http://www.chem.qmw.ac.uk/iubmb/
http://www.genome.ad.jp/
http://www.expasy.ch/enzyme/
http://www.rrz.uni-hamburg.de/biologie/b_online/d18_1/ec.htm

Further details on using this feature can be accessed at **http://www.rcsb.org/pdb/help-searchfields.html**. A basic query tutorial that offers guidance on using other SearchFields features is available at **http://www.rcsb.org/pdb/query_tut.html**.

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 continued their frequency with the start of the new 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
Dec 01	108,134	82,501	59,181	93,434,566	2,557,560	3,352,175
Nov 01	152,121	118,446	74,439	115,609,565	3,553,387	4,563,639
Oct 01	152,048	116,443	79,370	113,280,685	3,609,759	4,713,496

PDB Outreach

PDB Annual Report 2001 Now Available Online

The PDB Annual Report 2001 is now available from the PDB Web site in PDF format at **http://www.rcsb.org/pdb/annual_report01.pdf**. This document features a detailed look at the second full year of the RCSB's operation of the PDB from July 1, 2000 through June 30, 2001. It highlights PDB functions, accomplishments during this period, and plans for the coming year. Printed copies can also be obtained by sending your request and postal address to **AnnualReport@rcsb.org**.

PDB CD-ROM Issue 99 Released

The latest PDB CD-ROM (release #99) set is currently being distributed. This release contains the macromolecular structure entries for the 16,972 structures available as of January 1, 2002. The CD-ROMs are produced quarterly as of the last update of the PDB Web site for March, June, September and December. The experimental data (X-ray structure factors and NMR constraints) are also included, if available. Further information is available at **http://www.rcsb.org/pdb/cdrom.html**.

PDB Molecules of the Quarter

Photosystem I, DNA, and Glycogen Phosphorylase

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:

Photosystem I: Capturing Light

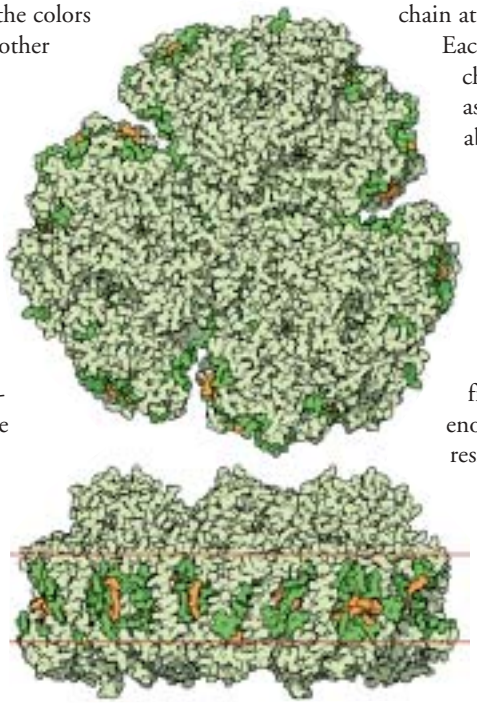
October, 2001—Look around. Just about everywhere that you go, you will see something green. Plants cover the Earth, and their smaller cousins, algae and photosynthetic bacteria, can be found in nearly every corner. Everywhere, they are busy converting carbon dioxide into sugar, creating living organic molecules out of air using the energy of sunlight as power. This process, termed photosynthesis, provides the material foundation on which all life rests.

At the center of photosynthesis is a class of proteins termed photosynthetic reaction centers. These proteins capture individual light photons and use them to provide power for building sugar. One example is photosystem I (PDB entry **1jb0**), one of the two large reaction centers used in cyanobacteria, algae and plants. Photosystem I is a trimeric complex that forms a large disk. In cells, the complex floats in a membrane with the large flat faces exposed above and below the membrane.

Each of the three subunits of photosystem I is a complex of a dozen proteins, which together support and position over a hundred cofactors. Some of these cofactors are exposed around the edge of complex and many others are buried inside. Cofactors are small organic molecules that are used to perform chemical tasks that are beyond the capabilities of pure protein molecules. The cofactors in photosystem I include many small, brightly-colored

molecules such as chlorophyll, which is bright green, and carotenoids, which are orange. The colors are, in fact, the reason that these molecules are useful: the colors are an indication that the cofactors absorb other colors strongly. For instance, chlorophyll absorbs blue and red light, leaving the beautiful greens for us to see. The energy from these absorbed colors is then captured to perform photosynthesis.

The heart of photosystem I is an electron transfer chain, a chain of chlorophyll, phylloquinone, and three iron-sulfur clusters. These cofactors convert the energy from light into energy that the cell can use. The two chlorophyll molecules at the bottom of the chain capture the light first. When they do, an electron is excited into a higher energy state. Normally this electron would quickly decay, releasing heat or releasing a new photon of slightly lower energy. But before this has a chance to happen, photosystem I passes this electron on, up the chain of cofactors. At the top, the electron is transferred to a small ferredoxin protein, which then ferries it on to the other steps of photosynthesis. At the bottom, the hole left by this wandering electron is filled by an electron from another protein, plastocyanin.



PDB ID ***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, p. 909.***

This may seem rather mundane until you see the trick that the photosystem is performing. The proteins at both ends of this process, ferredoxin and plastocyanin, are carefully chosen. Because of the special design of their own cofactors, it is more difficult to add an electron to ferredoxin than it is to plastocyanin--normally, the flow would be in the opposite direction. But photosystem I uses the energy from light to energize the electron, moving it in a difficult direction. Then, since the electron is placed in such an energetic position, it can be used to perform unfavorable duties such as the production of sugar from carbon dioxide.

Different photosystems are used by different photosynthetic organisms. Higher plants, algae, and some bacteria have photosystem I, and a second one termed photosystem II. A low resolution structure of photosystem II is available in PDB entry 1fe1. Photosystem II uses water instead of plastocyanin as the donor of electrons to fill the hole left when the energized electron is passed up the chain. When it grabs electrons from a water molecule, photosystem II splits the water and releases oxygen gas. This reaction is the source of all of the oxygen that we breathe. Some photosynthetic bacteria contain a smaller photosynthetic reaction center, such as the one in PDB entry **1prc**. As in photosystem I, a stack of chlorophyll and other cofactors transfer a light-energized electron up to an energetic electron carrier.

Of course, plants do not rely on the slim chance of a photon running into one tiny chlorophyll molecule in the middle of the reaction center. As with all things in life, cells have found an even better way. Photosystem I contains an electron transfer chain at the center of each of the three subunits. Each one is surrounded by a dense ring of chlorophyll and carotenoid molecules that act as antennas. These antenna molecules each absorb light and transfer energy to their neighbors. Rapidly, all of the energy funnels into the three reaction centers, where is captured to create activated electrons.

You can view the many photosystem I cofactors of the electron transfer chain and the antenna in PDB entry **1jb0**. Only one of the three subunits is included in the file, but you will find that this is complicated enough. Two special chlorophyll molecules, residues 1140 and 1239, act as a bridge between the reaction center in the middle and the many molecules in the surrounding antenna. Magnesium ions lie at the center of each chlorophyll. The residue numbers for the electron transfer chain are 1011-1013 and 1021-1023 for the chlorophylls, 2001-2002 for the phylloquinones, and 3001-3003 for the iron sulfur clusters.

A list of all photosystem I structures in the PDB as of October, 2001, is available at **http://www.rcsb.org/pdb/molecules/pdb22_report.html**. For more information about photosystem I, see **http://www.rcsb.org/pdb/molecules/pdb22_6.html**.

DNA: Your Inheritance

November, 2001—Each of the cells in your body carries about 1.5 gigabytes of genetic information, an amount of information that would fill two CD-ROMs or a small hard disk drive. Surprisingly, when placed in an appropriate egg cell, this amount of information is enough to build an entire living, breathing, thinking human being. Through the efforts of the international human genome sequencing projects, you can now read this information. Along with most of the biological research community, you can marvel at the complexity of this information and try to understand what it means. At the same time, you can wonder at the simplicity of this information when compared to the intricacy of the human body.

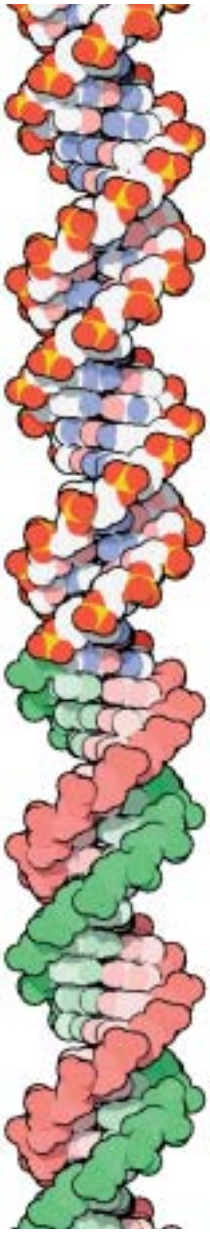
DNA is read-only memory, archived safely inside cells. Genetic information is stored in an orderly manner in strands of DNA. DNA is composed of a long linear strand of millions of nucleotides, and is most often found paired with a partner strand. These strands wrap around one another in the familiar double helix. The code is quite easy to read: you simply step down the strand of DNA one nucleotide at a time and read off the bases: A, T, C or G. This is exactly what your cells do: they scan down a messenger RNA (copied from the DNA), and use ribosomes to build proteins based on the code that is read. This is also how researchers determine the sequence of a DNA strand: they clip off one nucleotide at a time to see what it is.

Your genetic information, inherited from your parents, is your most precious possession. It guided the construction of your body in the first nine months of your life and it continues to control all of the basic functions of living. Each of your cells is constantly using this information, asking questions about how to control blood sugar levels and body temperature, how to digest different foods and how to deal with new environmental challenges, and thousands of other important questions. The answers are held in the DNA. Hundreds of different proteins are built to interact with this information: to read it and use it to build new proteins, to copy it when the cell divides, to store and protect it when it is not actively being used, and to repair the information when it becomes corrupted by chemicals or radiation.

DNA is arguably one of the most beautiful molecules in living cells. Its graceful helix is pleasing to the eye. DNA is also one of the most familiar molecules, the central icon of molecular biology, easily recognized by everyone. To some, it may carry a negative connotation, being a pervasive symbol for activists against genetically engineered produce. To others, it may bring to mind advances in forensics such as the DNA fingerprinting used in many recent high-profile trials. Some may have seen it in science fiction, modified to build dinosaurs or store cryptic messages from aliens. To all it is a pervasive symbol of our growing understanding of the human body and our close kinship with the rest of the biosphere, and the moral and ethical issues that must be addressed in the face of that knowledge.

DNA is perfect for the storage and readout of information. It is laden with information. Every surface and edge of the molecule carries information. The basic mechanism by which DNA stores and transmits genetic information was discovered in the 1950's by Watson and Crick. This basic information is stored in the way that the bases match one another on opposite sides of the double helix--adenine with thymine, guanine with cytosine—forming a set of complementary hydrogen bonds.

Additional 'extragenetic' information is read from the surfaces that are left exposed in the double helix. In the major groove (the wider of the two grooves in the structure), the different base pairs have a characteristic pattern of chemical groups that carry information. These include hydrogen bond donors and acceptors, as well as a site with a large, bulky group in adenine-thymine base pairs or a small group in guanine-cytosine base pairs. In the minor groove, there is a different arrangement of chemical groups that carry additional information. As revealed in hundreds of structures in the PDB, this extragenetic information is used by proteins to read the genetic code in DNA without unwinding the double helix. It is also targeted by a number of toxins and drugs that attack DNA.



PDB ID ***1bna***

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, p. 2179.*

DNA adopts the familiar smooth double helix, termed a B-helix, under the typical conditions found in living cells. An example is shown in the PDB entry **1bna**. Under other conditions, however, DNA can form other structures, as revealed in two early crystal structures: PDB entries **1ana** and **2dcg**. The structure found in PDB entry **1ana**, with tipped bases and a deep major groove, is termed A-DNA. It is formed under dehydrating conditions. Also, RNA most often shows this form, because its extra hydroxyl group on the sugar gets in the way, making the B-form unstable (look, for instance, at the A-helical structure of transfer RNA shown in a previous Molecule of the Month). The form found in PDB entry **2dcg**, which winds in the opposite direction from A-DNA and B- DNA, is termed Z-DNA. It is found under high salt conditions and requires a special type of base sequence, with many alternating cytosine-guanine and guanine-cytosine base pairs.

We often think of DNA as a perfect, smooth double helix. In reality, DNA has a lot of local structure. The small piece of DNA shown in PDB entry **1bna**, shows some of the common variations. At the top, the helix is bent to one side, distorted by the way that the helices are packed into the crystal. At the bottom, two of the bases are strongly propeller twisted—they are not in one perfect plane. This improves the way that the bases stack on top of one another along each strand, stabilizing the whole double helix. As more and more structures of DNA are studied, it is becoming clear that DNA is a dynamic molecule, quite flexible on its own, which is bent, kinked, knotted and unknotted, unwound and rewound by the proteins that interact with it.

To locate DNA structures in the PDB using the SearchFields interface, in the "Contains Chain Type" section select DNA-YES and all others NO. A list of all DNA structures in the PDB as of November, 2001 is available at **http://www.rcsb.org/pdb/molecules/pdb23_report.html**. For more information on DNA, see **http://www.rcsb.org/pdb/molecules/pdb23_5.html**.

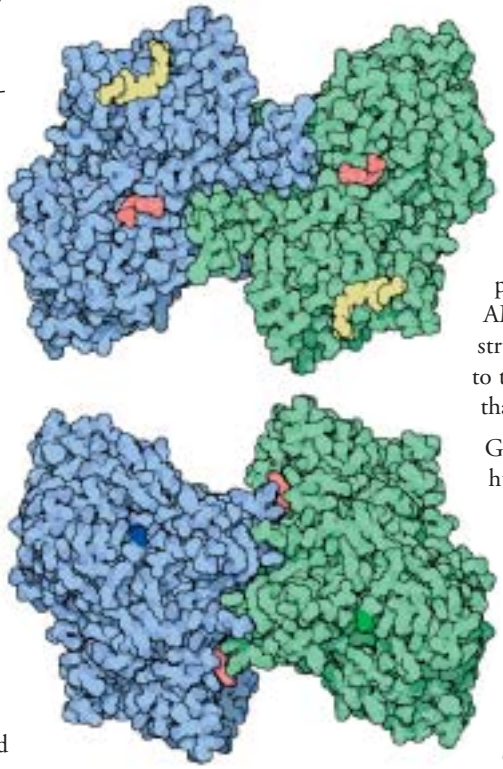
Glycogen Phosphorylase: Moderation for the Sweet Tooth

December, 2001—Although it may not seem so during the holiday season, we do not have to eat continually throughout the day. Our cells do require a constant supply of sugars and other nourishment, but fortunately our bodies contain a mechanism for storing sugar during meals and then metering it out for the rest of the day. The sugars are stored in glycogen, a large molecule that contains up to 10,000 glucose molecules connected in a dense ball of branching chains. Your muscles store enough glycogen to power your daily activities, and your liver stores enough to feed your nervous system and other tissues all through the day and on through the night.

Sugar is released from glycogen by the enzyme glycogen phosphorylase. It clips glucose from the chains on the surface of a glycogen granule. The enzyme is a dimer of two identical subunits, which can be seen in PDB entry **6gpb**. Two nucleotides, are bound at the entrance to the active site, which is found in a deep cleft. Short chains of sugars similar to the ends of glycogen chains bind into another cleft that the enzyme uses to grip the glycogen granule. In its cleavage reaction, glycogen phosphorylase uses a phosphate molecule, connecting it to the sugar as it is released. A second enzyme, phosphoglucomutase, then shifts the position of the phosphate to a neighboring carbon atom in the sugar, making the sugar ready for breakdown by glycolysis.

As you might imagine, this process is highly regulated. Traffic of sugar into and out of storage in glycogen is used to control the level of glucose in the blood, so glycogen phosphorylase must be activated when sugar is needed and quickly shut down when sugar is plentiful. It is controlled in several ways. First, the enzyme is activated by adding a phosphate molecule to a serine amino acid (serine 14) on the back side of the enzyme. The phosphate causes a large shift in the shape of the enzyme, shifting it into the active conformation. Two special enzymes control the addition and removal of this phosphate, based on levels of the sugar-monitoring hormones insulin and glucagon, and other hormones such as epinephrine (adrenaline).

Also, binding of other molecules can modify the activity of the molecule. For instance, AMP (adenine monophosphate) binds to a different site on the back side of the molecule, causing the same shift to the active conformation. This is useful, because AMP is a product of ATP breakdown and will be more plentiful when energy levels are low and more sugar is needed.



PDB ID ***6gpb***

L.N. Johnson, K.R. Acharya, M.D. Jordan, P.J. McLaughlin (1990): Refined crystal structure of the phosphorylase-heptulose 2-phosphate-oligosaccharide-AMP complex. J. Mol. Biol. ***211****, p. 645.*

Glycogen phosphorylase is activated by a change of shape. The structure in PDB entry **8gpb** is in the inactive T state, and the structure in PDB entry **1gpa** is in the active R state. (T stands for tense and R for relaxed, a notation developed when the first allosteric enzymes were being studied, although structures such as these have shown that the idea of tension does not really apply at the molecular level). The shift between the two shapes is controlled by phosphorylation of serine 14 or binding of AMP to the regulatory site. The R-state structure shown here has phosphates attached to the serines and a sulfate group in the site that binds to AMP.

Glycogen is used in many organisms, from humans to yeast. Much of the scientific work on the enzyme has been done with rabbit glycogen phosphorylase, such as those found in PDB entries **8gpb** and **1gpa**. You can look at the slightly different enzyme from yeast in PDB entry **1ygp**. This file contains the two protein chains and several small molecules. One such small molecule is the cofactor pyridoxal phosphate, a reactive molecule which binds tightly in the active site and is used to assist in the reaction. A phosphate is bound in each subunit next to the key threonine amino acid that is used for regulation, controlling an allosteric change similarly to serine 14 in the rabbit form. As you are looking at this enzyme, notice how the two protein chains wrap arms around one another. This allows the subunits to work together when responding to the small changes in shape that are used for control.

A list of all glycogen phosphorylases in the PDB as of December, 2001 is available at **http://www.rcsb.org/pdb/molecules/pdb24_report.html**. For suggestions for further reading about antibodies, see **http://www.rcsb.org/pdb/molecules/pdb24_4.html**.

PDB Job Listings

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

PDB Project Team Leaders

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

PDB Members

Rutgers

Rutgers, The State University of New Jersey
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

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

Olivera Tosic
olivera@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

David Archbell
dave@sdsc.edu

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

Dorothy Kegler
dkegler@sdsc.edu

David Padilla
dpadilla@sdsc.edu

Thomas Solomon
tsolomon@sdsc.edu

David Stoner
dstoner@sdsc.edu

Dr. Helge Weissig
helgew@sdsc.edu

NIST

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

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

Phoebe Fagan
phoebe.fagan@nist.gov

Dr. Veerasamy Ravichandran
vravi@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.