RCSB PDB NEWSLETTER

NUMBER 24 ◆ WINTER 2005 ◆ *Published quarterly by the Research Collaboratory for Structural Bioinformatics Protein Data Bank*

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

Links to RCSB PDB newsletters are available at www.rcsb.org/pdb/newsletter.html

Contents

Snapshot: January 1, 2005

28,992 released atomic coordinate entries

Molecule Type		Experimental Technique	
26,363	proteins, peptides, and viruses	24,706	diffraction and other
1,396	nucleic acids	4,286	NMR
1,220	protein/nucleic acid complexes	15,244	structure factor files
13	carbohydrates	2,349	NMR restraint files

Participating RCSB Members

Rutgers: rutgers.rcsb.org

SDSC/UCSD: www.pdb.org

CARB/UMBI/NIST: nist.rcsb.org

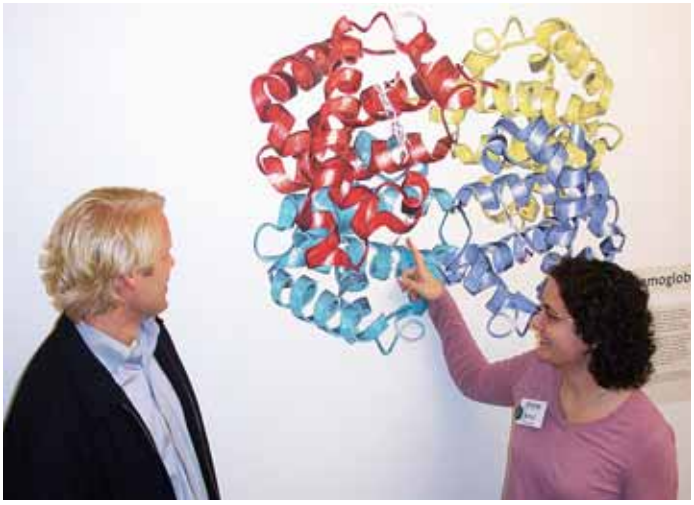
E-mail: info@rcsb.org

FTP: ftp.rcsb.org

The RCSB PDB is a member of the wwPDB (www.wwpdb.org)

Message from the RCSB PDB

At the end of 2004, an open house was held at the RCSB-Rutgers site. The December 9th event included the opening of the Molecular Art Mural that is described in this quarter's Education Corner. Demonstrations of annotation, the beta site, educational resources, and structural genomics databases were offered to students, teachers, and members of the local community.



Annotator Leader Kyle Burkhardt discusses hemoglobin with Edmund Scheer at the RCSB-Rutgers Open House

In 2005, the RCSB plans to have an exhibit stand at the meetings listed below.

Biophysical Society: February 11–15, 2005 *Long Beach, CA*

American Society for Biochemistry & Molecular Biology (ASBMB): April 2–6, 2005; San Diego, CA. There will also be a special session entitled "PDB in the Classroom and Beyond."

American Chemical Society Mid-Atlantic Regional Meeting (MARM): May 22–25, 2005 *Piscataway, NJ*

American Crystallographic Association (ACA): May 28–June 2, 2005 *Orlando, FL*

Intelligent Systems for Molecular Biology (ISMB): June 25–29, 2005 *Detroit, Michigan*

Congress of the International Union of Crystallography (IUCr): August 23–31, 2005 *Florence, Italy*

We hope to see you at these meetings, and at other RCSB PDB presentations to be made throughout the year.

Best wishes for 2005!

The RCSB PDB ✦

Data Deposition and Processing

Enhanced Design and Content of RCSB PDB Data Deposition Services

The RCSB PDB has further developed the web pages that describe data deposition. The main page at **deposit.pdb.org** links to the software and resources that can be used for preparing, checking, validating and depositing structural data to the PDB. It also provides individual guidelines for preparing structures determined by X-ray crystallography, NMR, or electron microscopy.

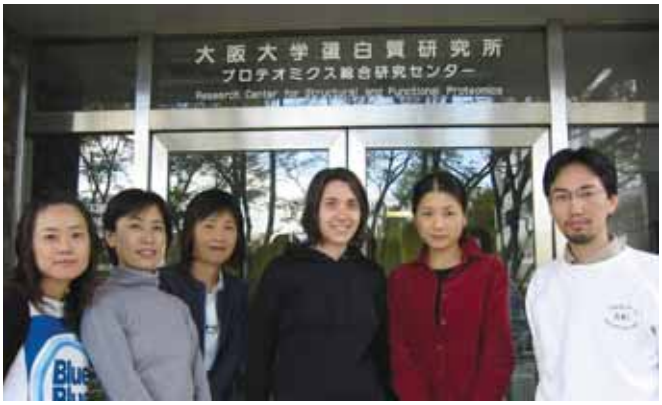
Tutorials are available for ADIT, pdb_extract, the Validation Suite, and Ligand Depot. A guide for converting a structure factor file into mmCIF format is also available.

A FAQ is provided to answer questions on a variety of subjects—deposition, update, and release policies (such as *How do I replace coordinates? How do I deposit a new chemical component/ligand?*), ADIT (*How can I base a new deposition on an older one?*), Validation (*I keep getting an error message – what could be wrong?*), and pdb_extract (*What does pdb_extract do?*).

It is recommended that all depositors review this page and its associated links in order to make their data deposition process quicker, easier, and more complete and accurate. We welcome your comments on these pages at **deposit@rcsb.rutgers.edu**.

PDB Focus: Worldwide Data Annotation

PDB data are processed by an international effort involving members of the wwPDB—RCSB, MSD-EBI, and PDBj. Structures deposited using ADIT are annotated by staff from the RCSB (at Rutgers University in New Jersey and remotely at the Academy of Sciences of the Czech Republic) and PDBj in Osaka, Japan. AutoDep depositions are processed by the MSD-EBI team in the United Kingdom.



RCSB annotator Suzanne Richman on a visit to PDBj, where she processed data for two months. From left to right: Shinobu Saeki, Yumiko Kengaku, Reiko Igarashi, Suzanne Richman, Yasuyo Morita, and Takashi Kosada.

PDB Deposition Statistics

In 2004, 5356 experimentally-determined structures were deposited to the PDB archive—a 14½% increase over 2003's 4677 depositions. The entries were processed by teams at RCSB-Rutgers, Osaka University, and the European Bioinformatics Institute. Of the structures deposited in 2004, 76% were deposited with a release status of "hold until publication;" 14% were released as soon as annotation of the entry was complete; and 10% were held until a particular date.

83% of these entries were determined by X-ray crystallographic methods; 14% were determined by NMR methods. 78% of these depositions were deposited with experimental data.

58% of these depositions released the sequence in advance of the structure's release.

PDB Focus: Ligand Depot—A small molecule information resource

Ligand Depot (**ligand-depot.rutgers.edu**) is a data warehouse that integrates databases, services, tools, and methods related to small molecules bound to macromolecules.

An important tool to use when depositing PDB structures, this resource can be used to find codes for existing ligands, to link to other entries with a particular ligand, and to search for substructures.

If a ligand related to a deposition is not in Ligand Depot, please email the chemical diagram, name, and formula to **deposit@rcsb.rutgers.edu**.

Zukang Feng, Li Chen, Himabindu Maddula, Ozgur Akcan, Rose Oughtred, Helen M. Berman, and John Westbrook. Ligand Depot: a data warehouse for ligands bound to macromolecules. Bioinformatics *(2004) 20: 2153-2155.*

Data Query, Reporting, and Access

RCSB Beta Site Features

In early July, the RCSB PDB released its newly reengineered beta site (**pdbbeta.rcsb.org**) for public testing. Some of the features of this site are described below.

- **Using PubMed Abstracts**

Under an agreement with the National Library of Medicine (NLM; **www.nlm.nih.gov**), PubMed abstracts for the primary citation for structures are available on the beta site. For any given structure, the PubMed abstract can be accessed from the Structure Explorer Summary Page by clicking on the PubMed link under the Primary Citation. A web page is returned with the article title, abstract, keywords, authors, organizational affiliation, journal, and PubMed identifier. Clicking on the icon shown will launch a browser window containing the corresponding entry at PubMed (**www.ncbi.nlm.nih.gov/entrez/query.fcgi?db=PubMed**).

The text box at the bottom of this page can be used to search for structures in the PDB using any word in the abstract or keyword fields. Terms can be entered into the text box either by typing the word manually or by clicking the mouse over any word in the abstract or the keyword fields. This provides an alternative means to finding structures of interest.

- **Searching the PDB based on NMR Experimental Details**

A new SearchFields feature is the ability to search for structures

based upon specific NMR experimental parameters. These include NMR Experiment Type, Refinement Method, Selection Criteria, Spectrometer Manufacturer, Spectrometer Model, and Sample Conditions (for pH and Degrees Kelvin). As with the current production site, searches may also be performed based upon experimental technique and the availability of constraint files.

• Database Browsers

A major feature of the beta site is the ability to easily browse through database content based on the categories Biological Process, Cellular Component, Molecular Function, Enzyme Classification, Source Organism, Disease, Genome, SCOP Classification, CATH Classification, and Mutant Resource.

Each browser provides a hierarchical view of all the structures that are associated with the particular category or classification. Users can explore the category's hierarchy, view the number of associated PDB structures, and search for specific associated structures. The browsers are available from the left hand navigation bar and from the Search Fields menu items on the beta site. Detailed help and browsing examples are available via the RoboHelp built into the reengineered site (in the left hand navigation bar, click on *Browse Database* and then on *How to Browse*).

Any comments and suggestions about the browsers or any other beta site features may be sent to **betafeedback@rcsb.org**.

Website Statistics

The PDB is available from several Web and FTP sites located around the world. Users are also invited to preview new features at the RCSB PDB beta test site, accessible at **beta.rcsb.org/pdb**.

The access statistics are given below for the primary RCSB PDB website at **www.pdb.org**.

Access Statistics for **www.pdb.org**

	DAILY AVERAGE		MONTHLY TOTALS			
MONTH	HITS	FILES	SITES	KBYTES	FILES	HITS
Oct 04	300,332	216,139	132,048	260,835,564	6,484,171	9,009,988
Nov 04	311,939	225,940	130,967	258,547,191	6,552,271	9,046,247
Dec 04	242,497	176,578	113,468	218,396,096	5,297,346	7,274,937

Data CD Distribution

The last 2004 release of PDB data on CD-ROM has been distributed. For 2005, the RCSB will distribute data to subscribers on DVD or CD-ROM each quarter.

The two-DVD quarterly sets will contain all structural and experimental data available in the archive at that time. The same data set on CD-ROM would currently require 20 disks.

The CD-ROM quarterly sets will contain structures released or modified during the previous three months.

Subscription ordering information is available at **www.rcsb.org/pdb/data_cd.html**.

Orders received after production has begun (January 7, April 8, July 8, or October 7, 2005) will be added to the next quarterly production cycle. Questions should be directed to **pdbcd@rcsb.org**. ✦

Outreach and Education

Recent Publications

• RCSB PDB 2004 Annual Report

This snapshot of the RCSB PDB, which covers the period of July 1, 2003–June 30, 2004, is intended to provide background information about the resource and describe recent progress and accomplishments. Available online as a PDF, this report is one of the many RCSB publications designed to keep our user community informed and involved (**www.rcsb.org/pdb/info.html#Press_Releases**).

• pdb_extract and data deposition tools described in Acta D

The options, procedures, and tools for accurate and automated PDB deposition are discussed in a recently published paper. Highlighted are pdb_extract, the PDB Validation Suite, and ADIT.

H. Yang, V. Guranovic, S. Dutta, Z. Feng, H. M. Berman and J. D. Westbrook. Automated and accurate deposition of structures solved by X-ray diffraction to the Protein Data Bank. Acta Cryst. *(2004). D60: 1833-1839.*

• XML data representation–PDBML

A paper discussing the PDB exchange dictionary and the PDB archive files (collectively named "PDBML") has been published online. PDBML files are available from **ftp://beta.rcsb.org/pub/pdb/uniformity/data/XML/**.

John Westbrook, Nobutoshi Ito, Haruki Nakamura, Kim Henrick, and Helen M. Berman. PDBML: the representation of archival macromolecular structure data in XML. Bioinformatics *(online October 2004)* **bioinformatics.oupjournals.org/cgi/content/abstract/bti082**

• PDB Oral History

RCSB PDB Director Helen M. Berman was interviewed by David Berol (currently a writer for the Eastern Research Group) to discuss the rich history of the Protein Data Bank. Since her beginning days as a crystallographer, Berman believed that protein structure data should be available for archiving and sharing. In this interview, she discusses how the PDB was established at Brookhaven National Laboratory, the impact of community and technology on the archive, and the PDB's transition to the RCSB. Access to this oral history can be arranged through the Chemical Heritage Foundation (**www.chemheritage.org**).

Meetings, Exhibits, & Workshops

• Cryo-EM Workshop held at RCSB-Rutgers

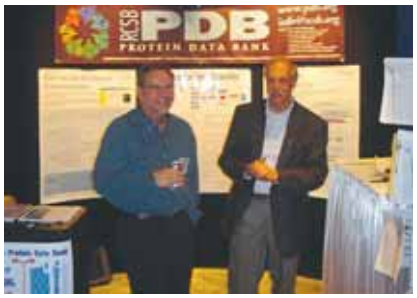
A workshop was held to develop community consensus on the data items needed for deposition of 3D density maps and atomic models derived from cryo-electron microscopy studies. Organized by Helen M. Berman (Rutgers, The State University of New Jersey), Wah Chiu (Baylor College of Medicine), and Michael Rossmann (Purdue University), the Cryo-Microscopy Structure Deposition Workshop was held at RCSB-Rutgers in Piscataway, NJ (October 23-24, 2004). The workshop examined the data items currently collected by EMDep and ADIT for such depositions and discussed desirable additions. Workshop presentations are available from the meeting website at **rcsb-dev.rutgers.edu:5015**.

Attendees at the Cryo-EM Workshop

The workshop was sponsored by the RCSB PDB and the Computational Center for Biomolecular Complexes (**ncmi.bcm.tmc.edu/ccbc**).

• International Conference on Structural Genomics

The RCSB PDB exhibited at the 2004 International Conference on Structural Genomics in Washington, DC (ISGO; November 17-21) and met with the structural genomics community.



Eldon Ulrich visits Gary Gilliland at the RCSB PDB booth at the ISGO Meeting

• RCSB PDB Art of Science Exhibit

In celebration of National Chemistry Week, the RCSB PDB Art of Science exhibit was at Fairleigh Dickinson University (FDU) in Teaneck, NJ. Sponsored by the School of Natural Sciences and the Weiner Library of FDU and the Hudson Bergen Chemical Society, an opening reception was held November 19. The exhibit ran until December 19.

• Symposium: Database Challenges in Biology

This report by Gary L. Gilliland on the September 10, 2004 symposium held at RCSB-CARB also appears in the Winter 2004 ACA Newsletter.

This meeting highlighted many of the challenges facing biological database resources that include the increasing rate of data acquisition and complexity, database integration, data validation and data mining. In attendance were nearly 100 scientists from academic and research institutions, and government agencies.

After a welcome by Gary Gilliland (RCSB PDB and CARB), Mark Ellisman (University of California, San Diego) gave his lecture *Multi-scale Imaging and Databasing of the Nervous System with Advanced Cyberinfrastructure*. He focused on data acquisition and analysis issues associated with nervous system data, and provided an overview of several aspects of the Biomedical Informatics Research Network (BIRN, **www.nbirn.net**). Next, Wah Chiu (Baylor College of Medicine) lectured on the *Database for Cryo-Electron Microscopy*. He described his activities that center around the use of electron crystallography and cryo-electron microscopy to determine the three-dimensional structure of macromolecular assemblies. His presentation included a description of The National Center for Macromolecular Imaging (**ncmi.bcm.tmc.edu**), an extensive network of collaborative projects, many of which are focused on structural investigation targets that may be critical for use in developing drugs for healthcare. The next presentation by John Johnson (The Scripps Research Institute), *VIrus Particle ExploreR (VIPER): a Database of Standardized Atom Coordinates for Icosahedral Viruses and Derived Description of Subunit Interactions*, highlighted the resource at **mmtsb.scripps.edu/viper**. The structural data for these structures has been put into a uniform format that allows viewing and analyzing the complete capsid structure.

In the afternoon, John Markley (University of Wisconsin) presented *Data Management in the Laboratory: User Facilities and Research on Small and Large Scales*. The facilities involved include the National Magnetic Resonance Facility (**www.nmrfam.wisc.edu**) and BioMagResBank (**www.bmrb.wisc.edu**). His lecture described the issues associated with the complex data associated with NMR structure determinations, from sample preparation to data analysis, and an approach for capturing these data. Stephen Bryant (National Center for Biotechnology Information; NCBI, **www.ncbi.nlm.nih.gov**) then discussed the *Conserved Domain Database: A Protein Family Database*. His presentation included a description of novel tools for visualizing and analyzing structural similarities, and illustrated how the structural data is integrated with functional annotation data and reference information generated and maintained by NCBI. Next, Cathy Wu (Georgetown University Medical Center) presented a talk entitled *PIR Integrated Bioinformatics for Functional Genomics and Proteomics*. She described the current state of the Protein Information Resource (PIR; **pir.georgetown.edu**) and how this resource is now involved in UniProt (Universal Protein Resource, **www.pir.uniprot.org**), a joint effort by PIR, the European Bioinformatics Institute (EBI) and the Swiss Institute of Bioinformatics to consolidate protein sequence information from diverse sources. The final presentation of the day, entitled *The Protein Data Bank: An Integrated Resource for Structural Biology*, was given by Helen Berman (Rutgers University and the RCSB PDB). Her talk gave a historical perspective of the PDB, its current status, and future challenges. She capped the day by highlighting many of the issues and advances associated with international efforts to insure a single archive for the structural data. The diversity and challenges of the database activities described by the speakers made this a memorable event.

Molecules of the Quarter: G Proteins, Photosystem II, & Ubiquitin

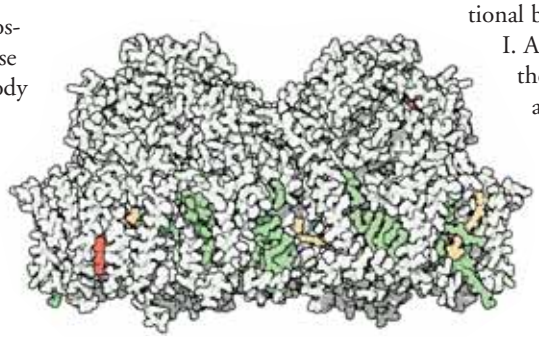
The *Molecule of the Month* series by David S. Goodsell explores the functions and significance of selected biological macromolecules for a general audience.

October 2004—*G Proteins*

Cells communicate by passing small, disposable messages to one another. Some of these messengers travel to distant parts of the body through the blood; others simply diffuse over to a neighboring cell. Then, another cell picks the message up and reads it. In many cases, these molecular messengers never get inside cells. Instead, the message is picked up by a receptor on the cell surface and the signal is then passed from outside to inside through a chain of signaling molecules. G proteins form the central link in this chain of communication. The G protein system is the most common method of signaling in our cells. Thousands of G-protein-coupled receptors have been found on our cells, each waiting for its own particular messenger. Some recognize hormones and make changes in the level of metabolism. Others are used in the nervous system to transmit nerve signals. Our sense of sight also relies on a G protein system that is sensitive to light, and a thousand different forms of these receptors, each recognizing the odor of a different molecule, control our sense of smell. They all share the combination of a receptor that receives a message and a G protein that delivers it inside the cell.

November 2004—*Photosystem II*

Photosystem II is the first link in the chain of photosynthesis. It captures photons and uses the energy to extract electrons from water molecules. These electrons are used in several ways. First, when the electrons are removed, the water molecule is broken into oxygen gas, which bubbles away, and hydrogen ions, which are used to power ATP synthesis. This is the source of all of the oxygen that we breathe. Second, the electrons are passed down a chain of electron-carrying proteins, getting an additional boost along the way from photosystem I. As these electrons flow down the chain, they are used to pump hydrogen ions across the membrane, providing even more power for ATP synthesis. Finally, the electrons are placed on a carrier molecule, NADPH, which delivers them to enzymes that build sugar from water and carbon dioxide.



Photosystem II. PDB ID: 1s5l

K.N. Ferreira, T.M. Iverson, K. Maghlaoui, J. Barber, S. Iwata. Architecture of the Photosynthetic Oxygen-Evolving Center Science *(2004) 303:1831-1838*

December 2004—*Ubiquitin*

As its name implies, ubiquitin is found in all eukaryotic cells and in cells throughout your body. The Nobel Prize in Chemistry was awarded this year to the three researchers who discovered its essential function in 1980. In the subsequent years, it has become apparent that apart from its role in protein disposal, ubiquitin is also used for other tasks, such as directing the transport of proteins in and out of the cell. By connecting ubiquitin together in short or long chains, or using different types of linkages between the molecules, many different signals may be encoded. Because of the important roles it plays, ubiquitin has changed very little over the evolution of life, so you can find a similar form in yeast cells, plant cells, and in our own cells.

The full Molecule of the Month features are available from **www.rcsb.org/pdb/molecules/molecule_list.html** ✦

PDB Community Focus: John L. Markley, BMRB & Center for Eukaryotic Structural Genomics

John L. Markley received a Ph.D. in Biophysics from Harvard University in 1969 where he worked with Oleg Jardetzky and Elkan R. Blout. His graduate research included NMR studies of helix coil transitions in polyamino acids and the preparation and NMR investigation of selectively deuterated proteins. The latter project was made possible through funding and access to excellent research facilities at the Merck Research Laboratories in Rahway, New Jersey, where he spent 30 months. As a National Institutes of Health Postdoctoral Fellow with Melvin P. Klein at the University of California, Berkeley, he made the transition from continuous-wave to pulse Fourier transform NMR spectroscopy and investigated NMR relaxation mechanisms. He joined the faculty of the Chemistry Department at Purdue University as an Assistant Professor in 1972, and by 1981 had moved up the ranks to Professor. He relocated to the Biochemistry Department at the University of Wisconsin-Madison in 1983, where he founded the National Magnetic Resonance Facility at Madison (1985), the BioMagResBank (BMRB; 1990), and the Center for Eukaryotic Structural Genomics (2000). He currently is Steenbock Professor of Biomolecular Structure and chairs the Graduate Program in Biophysics at the University of Wisconsin-Madison.

Q What is the history behind the BMRB?

A NMR is unique among biophysical approaches in its ability to provide a broad range of atomic-level information relevant to the structural, dynamic, and chemical properties of biological macromolecules. Since my days as a graduate student, I have been deeply impressed by the value of chemical information available from NMR data assigned to specific sites in proteins. In 1984, Eldon Ulrich and I wrote a comprehensive review of all assigned chemical shifts in proteins, which required digging information from individual publications in the literature. In 1985, during a mini-sabbatical with the late Professor Yoshimasa Kyogoku at the Protein Research Institute in Osaka, Japan, I formulated the idea for organizing a publicly available data bank for assigned protein

NMR parameters. In the meantime, the first NMR structures of proteins began appearing. Ulrich, Kyogoku, and I refined the idea for a data repository for NMR information about protein structure and dynamics and published a proposal in 1989. We secured funding for a pilot study from the National Library of Medicine, which enabled us to begin developing data models. These initially took the form of flat files with a rigid format. Later, BMRB adopted a relational database format and following discussions with Helen Berman and other developers of mmCIF (which is a variant of the STAR format devised by Sydney Hall and Nick Spadaccini) eventually developed the current NMR-STAR format. The data exchange format created at BMRB has gained rapid and widespread acceptance in the community. NMR-STAR is extensible and thus can accommodate the addition of new NMR parameters of interest to the biomolecular NMR community. Its tag-value nature and the tabular organization of the data model make it easy to interconvert NMR-STAR with relational or XML formats. BMRB's holdings have grown to include chemical shifts, *J*-couplings, relaxation rates, residual dipolar couplings, and chemical information derived from NMR investigations (such as hydrogen exchange rates, pK_a values, and structural restraints). New types of data collected are raw (time domain) data for structure determinations (largely from structural genomics centers) and solid-state NMR data. Our early idea was that BMRB annotators would gather and enter data from the literature, but the enormous growth in the biomolecular NMR field and the reluctance of journals to publish this information soon made this impractical. As with the PDB, BMRB relies on depositions from scientists. We are grateful that the NLM has funded BMRB continuously since its founding. BMRB now has mirror sites in Florence, Italy, and Osaka, Japan. The Osaka site is beginning to take responsibility for data depositions from that part of the world.

Q What are the interactions with the BMRB and the RCSB PDB?

A BMRB is a member of the RCSB and has close ties with the PDB. Our interactions are warm and collegial and span common interests in standards for data representation, software development, and data interchange. Recent collaboration with PDB has centered on unifying the nomenclature used for X-ray and NMR structures and the underlying data. BMRB and PDB have been pursuing the common goal of providing the tools for harvesting the full range of information in digital form that constitutes the normal "Methods" section of an article in a journal such as *J. Biol. Chem.* BMRB has adapted the PDB ADIT deposition software for data entry as a way of making data entry more uniform across the two data banks. Currently, data relevant to BMRB associated with NMR structures deposited at PDB are transferred automatically to BMRB for processing. BMRB and PDB are close to releasing jointly developed software that will provide one-stop data entry for NMR structures. This will simplify the task of depositors as well as streamline the work of annotators at BMRB and PDB.

Q Does your work in structural genomics influence your work with the BMRB, and vice versa?

A The goals of structural genomics are to enlarge knowledge of sequence-structure-function interrelationships and to lower the costs of solving structures. At the same time, its technology and products are to be made available to the community. Structural genomics both reinforces the kind of work that has gone on at the BMRB and the PDB, and offers challenges to these data banks. In some ways, the data dictionary work at BMRB and PDB anticipated many of the demands of structural genomics. Both data banks were well prepared to handle increasing numbers of structures and the increasing level of detail about the experiments demanded by structural genomics. The challenges have been to provide streamlined data deposition and pre-validation tools. BMRB has worked closely with all structural genomics centers that utilize NMR spectroscopy to determine their needs and to seek suggestions for improving the operations of the data bank. These interactions have led to new developments at BMRB, such as the use of validation software developed at the Northeast Structural Genomics Consortium, the extension of NMR-STAR format for chemical shifts to include probabilistic assignments as developed at the National Magnetic Resonance Facility at Madison, and the repository of collections of time-domain data sets used in structure determinations. Also, in response to the structural genomics community, BMRB has been increasing its links to other sites on the Web.

Q The amount of experimental data to be archived is growing exponentially. What is the future of data management?

A We are confident of our ability at BMRB, given the current level of approved funding, to keep up with the growth in the field. Our strategy at BMRB is to develop procedures and software that enable us to manage data more efficiently. The new data deposition system mentioned above, which jointly handles coordinates as well as underlying data and information about the biological system and experiments conducted, allows for more automated data harvesting and validation. I anticipate that BMRB, like PDB, will become increasingly internationalized. The new data deposition site at Osaka represents a positive step in this direction. At first, data collected at Osaka will be transferred to Madison for annotation, but the plan is to develop this capability in Osaka.

Q The size and complexity of structures being solved by X-ray and EM methods continues to increase over time. Is there a limit to the size of structures that can be investigated with NMR methods?

A As recent publications from the laboratories of Kurt Wüthrich and others have shown, it is difficult to place an absolute molecular weight limit on NMR structures. With conventional methods of uniform ^{13}C +^{15}N labeling, the practical limit for high-throughput NMR structure determinations is about 20 kDa. However, by expending additional effort and by utilizing ^{2}H labeling, the bar can be raised to 30 kDa and above. Recent results from Masatsune Kainosho's laboratory with stereo array isotope labeling suggest that this approach could increase the practical limit to 40 kDa. Selective labeling approaches also are increasing the sizes of RNA structures that can be solved by NMR. It is important to recognize that NMR can provide useful information about structure-function relationships even in the absence of a three-dimensional structure. BMRB captures this kind of information, which can be obtained from systems 150 kDa and larger.

PDB Education Corner: Early Protein Structures

Part of the RCSB-Rutgers Open House (described in this newsletter's Message from the RCSB PDB) was a celebration marking the opening of the Molecular Art Mural. Painted by local artist Jessica Milazzo onto the walls of RCSB-Rutgers, this mural depicts some of the earliest structures solved by X-ray crystallography. The mural's accompanying text notes, reprinted here, are from the RCSB PDB's Molecule of the Month *series and RCSB PDB staff. These images from the event show the individual pieces of the mural, and some of the visitors—Jessica Milazzo, Maggie England, and Dania Jorgensen.*

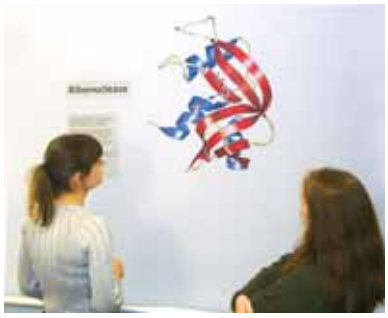
Hemoglobin

The science of protein structure began with the structure of hemoglobin. After years of arduous work, Max Perutz and his coworkers determined its atomic structure. Perutz's pioneering work in X-ray crystallography of proteins—including his study of hemoglobin won him the Nobel Prize in 1962.

Hemoglobin is the protein that makes blood red. It is composed of four protein chains—two alpha chains and two beta chains, each with a ring-like heme group containing an iron atom. Oxygen binds reversibly to these iron atoms and is transported through blood. Each of the protein chains is similar in structure to myoglobin, the protein used to store oxygen in muscles and other tissues.

PDB ID: 2dhb. *W. Bolton and M.F. Perutz. Three-dimensional Fourier synthesis of horse deoxyhaemoglobin at 2.8 Ångstrom units resolution.* Nature *(1970) 228:551-552. M.F. Perutz, M.G. Rossmann, A.F. Cullis, G. Muirhead, and G Will. Structure of haemoglobin: a three-dimensional Fourier synthesis at 5.5 Ångstrom resolution.* Nature *(1960) 185:416-422.*

Ribonuclease



The structure of ribonuclease was the third protein—after myoglobin and lysozyme—that was determined by X-ray crystallography. Two independent ribonuclease structures were reported in 1967.

Ribonucleases are small enzymes that catalyze the breakdown of single-stranded ribonucleic acid (RNA) by cleaving a phosphodiester bond. Ribonucleases have many biological functions, such as cutting harmful RNA into smaller components in order to remove them from the cell. Ribonuclease's structure contains a cleft in which the RNA is held during cleavage.

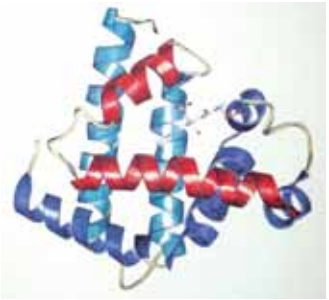
G. Kartha, J. Bello, J and D. Harker. Tertiary structure of ribonuclease. Nature, *(1967) 213:862-865. H.W. Wyckoff, K.D. Hardman, N.M. Allewell, T. Inagami, D. Tsernoglou, L.N. Johnson, and F.M. Richards. The structure of ribonuclease-S at 6 Ångstrom resolution.* J. Biol. Chem. *(1967) 242:3749-3753.*

Lysozyme

Lysozyme protects biological organisms from the ever-present danger of bacterial infection. This small enzyme attacks the protective cell walls of bacteria. Bacteria build a tough skin of carbohydrate chains to brace their delicate cell membranes against changes in osmotic pressure. Lysozyme breaks these carbohydrate chains, which destroys the structural integrity of the bacterial cell wall. The bacteria then burst under their own internal pressure. Lysozyme is present in many places that are rich with potential food for bacterial growth. The lysozyme pictured here is from hen egg white, where lysozyme serves to protect the proteins and fats that will nourish the developing chick.

PDB ID: 2lyz. *C.C.F. Blake, D.F. Koenig, G.A. Mair, A.C.T North, D.C. Phillips, and V.R. Sarma. Structure of hen egg-white lysozyme. A three dimensional Fourier synthesis at 2 Ångstrom resolution.* Nature *(1965) 206:757-761. C.C.F. Blake, L.N. Johnson, G.A. Mair, A.C.T. North, D.C. Phillips, and V.R. Sarma, Crystallographic studies of the activity of hen egg-white lysozyme.* Proc. R. Soc. London Ser. B *(1967) 167:378-388.*

Myoglobin



Myoglobin was the first reported protein structure. It represented a milestone in structural biology for which John Kendrew shared the Nobel Prize in Chemistry in 1962. This structure, along with the work on hemoglobin being carried out by Max Perutz, set the stage for developing our emerging understanding of biology at the atomic level.

Myoglobin is a small, bright red protein. It is very common in muscle cells, and gives meat much of its red color. Its biological function is to store oxygen obtained from hemoglobin that is carried in the blood for use when muscles are hard at work. The myoglobin used in the structure shown was taken from sperm whale muscles. Marine whales and dolphins have a great need for myoglobin, so that they can store extra oxygen for use in their deep dives undersea.

PDB ID: 1mbn. *J.C. Kendrew, G. Bodo, H.M. Dintzis, R.G. Parrish, and H. Wyckoff. A three-dimensional model of the myoglobin molecule obtained by X-ray analysis.* Nature *(1958) 181:662-666. H.C. Watson, The stereochemistry of the protein myoglobin.* Prog. Stereochem. *(1969) 4:299.*

RCSB PDB Partners

The RCSB PDB is managed by three partner sites of the Research Collaboratory for Structural Bioinformatics:

Rutgers, The State University of New Jersey
Department of Chemistry and Chemical Biology
610 Taylor Road
Piscataway, NJ 08854-8087

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

Center for Advanced Research in Biotechnology/UMBI/NIST
9600 Gudelsky Drive
Rockville, MD 20850-3479

The RCSB PDB is a member of the Worldwide PDB (www.wwpdb.org)

Mirror Sites

Cambridge Crystallographic Data Centre (UK): pdb.ccdc.cam.ac.uk
National University of Singapore: pdb.bic.nus.edu.sg
Osaka University (Japan): pdb.protein.osaka-u.ac.jp
Max Delbrück Center for Molecular Medicine (Germany): www.pdb.mdc-berlin.de

Statement of Support

The RCSB PDB is supported by funds from the National Science Foundation, the National Institute of General Medical Sciences, the Office of Science, Department of Energy, the National Library of Medicine, the National Cancer Institute, the National Center for Research Resources, the National Institute of Biomedical Imaging and Bioengineering, and the National Institute of Neurological Disorders and Stroke.

RCSB PDB Leadership Team

The overall operation of the PDB is managed by the RCSB PDB Leadership Team.

Dr. Helen M. Berman, *Director*
Rutgers, The State University of New Jersey
berman@rcsb.rutgers.edu

Dr. Philip E. Bourne, *Co-Director*
San Diego Supercomputer Center
University of California, San Diego
bourne@sdsc.edu

Judith L. Flippen-Anderson, *Production and Outreach Leader*
Rutgers, The State University of New Jersey
flippen@rcsb.rutgers.edu

Dr. Gary L. Gilliland, *Co-Director*
Center for Advanced Research in Biotechnology/UMBI/NIST
gary.gilliland@nist.gov

Dr. John Westbrook, *Co-Director*
Rutgers, The State University of New Jersey
jwest@rcsb.rutgers.edu

A list of current RCSB PDB Team Members is available at
www.rcsb.org/pdb/rcsb-group.html

Job openings available at the RCSB PDB are listed at
www.rcsb.org/pdb/jobs.html

RCSB PROTEIN DATA BANK
www.pdb.org

San Diego Supercomputer Center
University of California, San Diego
9500 Gilman Drive
La Jolla, CA 92093-0537

Return Service Requested