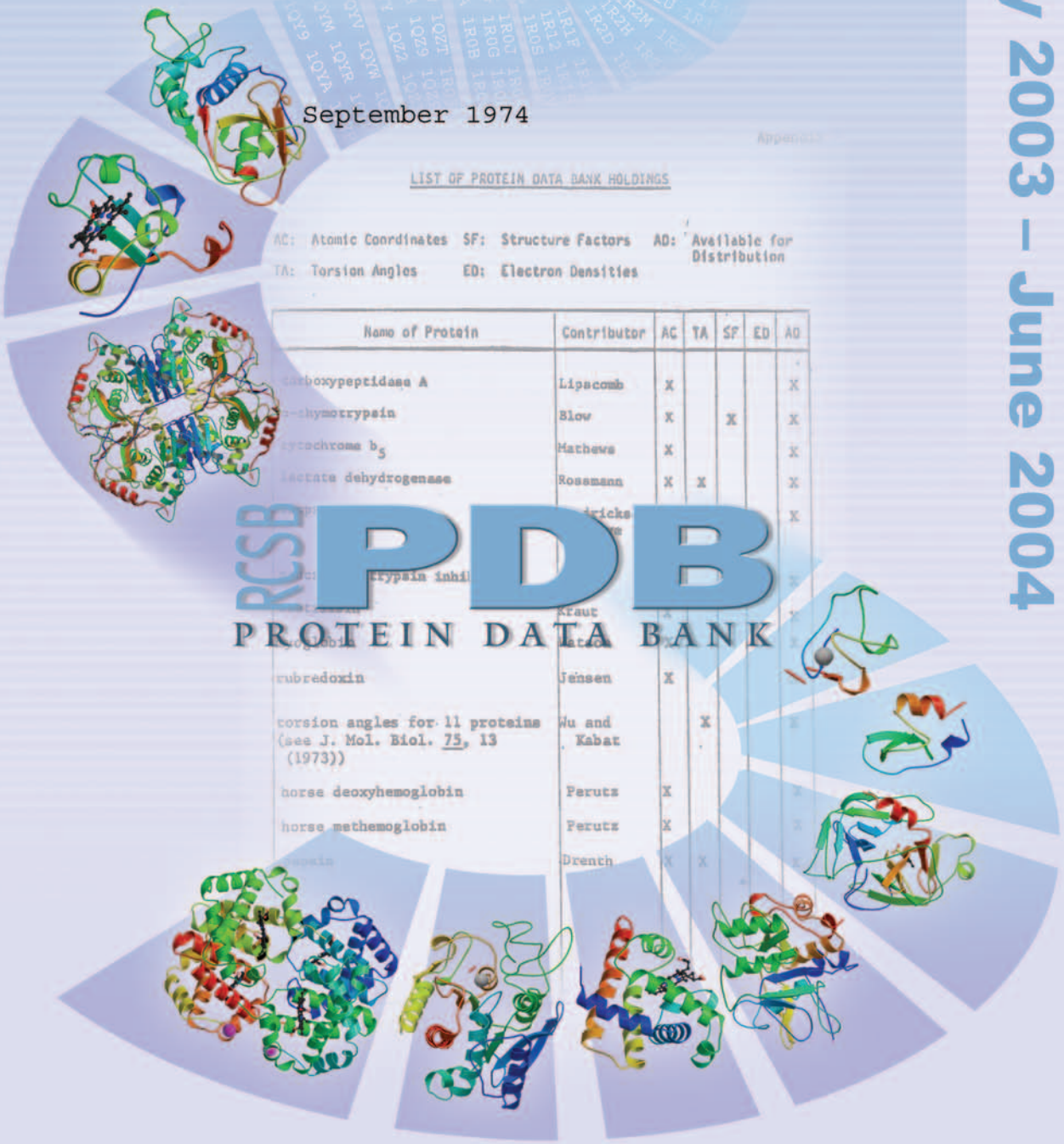
ANNUAL REPORT

for July 2003 – June 2004



RCSB PDB

PROTEIN DATA BANK

Research Collaboratory for Structural Bioinformatics

Rutgers, The State University of New Jersey
San Diego Supercomputer Center, UC San Diego
Center for Advanced Research in Biotechnology/UMBI/NIST

CONTENTS

ABOUT THE COVER

2004 marks the 30th anniversary of the Protein Data Bank's first newsletter, which was published in September 1974. It contained forms for depositing data, provided detailed directions for retrieving data from the PDB, and listed the original 12 coordinate sets available at that time. This cover shows the beginning of the PDB's journey through structural biology, as it started with the newsletter and an archive of a few structures—hemoglobin,[1-3] carboxypeptidase A,[4] myoglobin,[5] subtilisin,[6] alpha-chymotrypsin,[7] pancreatic trypsin inhibitor,[8] rubredoxin,[9] lactate dehydrogenase,[10] cytochrome b_5,[11] and papain.[12]

COVER REFERENCES

[1] *Bolton, W. and Perutz, M.F. (1970) Three dimensional fourier synthesis of horse deoxyhaemoglobin at 2.8 Ångstrom units resolution.* Nature, *228, 551-552.*

[2] *Perutz, M.F., Muirhead, H., Cox, J.M. and Goaman, L.C. (1968) Three-dimensional Fourier synthesis of horse oxyhaemoglobin at 2.8 Å resolution: the atomic model.* Nature, *219, 131-139.*

[3] *Hendrickson, W.A., Love, W.E. and Karle, J. (1973) Crystal structure analysis of sea lamprey hemoglobin at 2 Å resolution.* J Mol Biol, *74, 331-361.*

[4] *Quiocho, F.A. and Lipscomb, W.N. (1971) Carboxypeptidase A: a protein and an enzyme.* Adv Protein Chem, *25, 1-78.*

[5] *Watson, H.C. (1969) The stereochemistry of the protein myoglobin.* Prog. Stereochem., *4, 299.*

[6] *Alden, R.A., Birktoft, J.J., Kraut, J., Robertus, J.D. and Wright, C.S. (1971) Atomic coordinates for subtilisin BPN' (or Novo).* Biochem Biophys Res Commun, *45, 337-344.*

[7] *Birktoft, J.J. and Blow, D.M. (1972) Structure of crystalline alpha-chymotrypsin. V. The atomic structure of tosyl-alpha-chymotrypsin at 2 Å resolution.* J Mol Biol, *68, 187-240.*

[8] *Huber, R., Kukla, D., Ruhlmann, A., Epp, O. and Formanek, H. (1970) The basic trypsin inhibitor of bovine pancreas. I. Structure analysis and conformation of the polypeptide chain.* Naturwissenschaften, *57, 389-392.*

[9] *Watenpaugh, K.D., Sieker, L.C., Herriott, J.R. and Jensen, L.H. (1973) Refinement of the model of a protein: rubredoxin at 1.5 Å resolution.* Acta Crystallogr B, *29, 943-956.*

[10] *White, J.L., Hackert, M.L., Buehner, M., Adams, M.J., Ford, G.C., Lentz Jr., P.J., Smilely, I.E., Steindel, S.J. and Rossman, M.G. (1976) A comparison of the structures of apo dogfish M4 lactate dehydrogenase and its ternary complexes.* J Mol Biol, *102, 759-779.*

[11] *Mathews, F.S., Argos, P. and Levine, M. (1972) The structure of cytochrome* b_5 *at 2.0 Ångstrom resolution.* Cold Spring Harb Symp Quant Biol, *36, 387-395.*

[12] *Drenth, J., Jansonius, J.N., Koekoek, R., Swen, H.M. and Wolthers, B.G. (1968) Structure of papain.* Nature, *218, 929-932.*

MESSAGE FROM THE RCSB PDB

The RCSB Protein Data Bank is a publicly accessible information portal for researchers and students interested in structural biology. At its center is the PDB archive—the sole international repository for the 3-dimensional structure data of biological macromolecules. These structures hold significant promise for the pharmaceutical and biotechnology industries in the search for new drugs and in efforts to understand the mysteries of human disease.

The primary mission of the RCSB PDB is to provide accurate, well-annotated data in the most timely and efficient way possible to facilitate new discoveries and scientific advances. The RCSB processes, stores, and disseminates these important data, and develops the software tools needed to assist users in depositing and accessing structural information.

Staff are located at three member institutions of the Research Collaboratory for Structural Bioinformatics (RCSB): Rutgers, The State University of New Jersey; the San Diego Supercomputer Center (SDSC) at the University of California, San Diego (UCSD); and the Center for Advanced Research in Biotechnology (CARB/UMBI/NIST).

This report details the RCSB PDB's accomplishments from July 1, 2003, through June 30, 2004, and looks at the history, contents, and use of the RCSB PDB resource. As a popular and publicly available biological database, the RCSB PDB server is extremely active: in an average month during this report period, approximately 429 structures were deposited to the PDB archive and processed, 411 structures were released, and 5 million files of individual structures were downloaded.

The RCSB is dedicated to providing resources that fully utilize the data contained in the archive. To this end, the RCSB PDB website and database have been reengineered using feedback derived from correspondence with our help desks, conference attendance, focus groups and other personal interactions between the users of the PDB and RCSB staff. This effort has improved the usability and navigation of the website, and has introduced new capabilities for browsing and searching the PDB archive that provide more reliable query results. This resource, currently undergoing public beta testing at **pdbbeta.rcsb.org**, is described in greater detail on pages 6-7.

The RCSB PDB Team

Other highlights during this period include:

- Enhanced distribution features, including a new QuickSearch feature, and the availability of bulk non-redundant data (page 6)
- Growth and promotion of educational resources for all varieties of PDB users (page 8)
- Formation of the Worldwide Protein Data Bank (wwPDB), a collaboration with the RCSB, PDBj, and the MSD-EBI which ensures a single, uniform and enduring archive of experimentally determined 3-dimensional biological structures (page 4)
- Availability of Ligand Depot—a small molecule information web resource (page 5)
- Expansion of resources for structural genomics, including the development of the Protein Expression Purification and Crystallization Database (PepcDB; pages 8-9)
- Release of PDB_EXTRACT, a program used to capture information from various steps of X-ray crystallographic applications for use in preparing PDB depositions (page 5)

The continued growth of our varied services in the areas of data deposition, annotation, distribution, community outreach, and education are detailed in this report.

We welcome and appreciate your feedback, which helps us to further improve and refine this important resource.

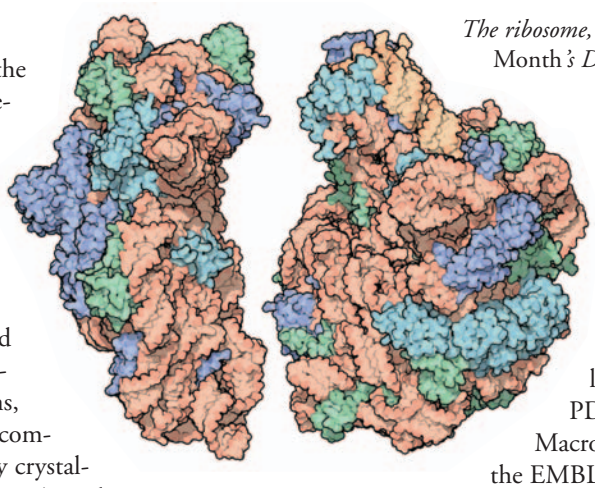
Helen M. Berman (Director), Philip E. Bourne, Gary L. Gilliland, John Westbrook (Co-Directors), Judith L. Flippen-Anderson (Production and Outreach Leader)

for the RCSB Protein Data Bank

The RCSB PDB Leadership Team: (left to right) Philip E. Bourne, Judith L. Flippen-Anderson, Helen M. Berman, John Westbrook, Gary L. Gilliland

WHAT IS THE PDB?

The PDB[1, 2] was founded in 1971 at Brookhaven National Laboratory as the sole international repository for three-dimensional structure data of biological macromolecules. Since July 1, 1999, the PDB archives have been managed by three member institutions of the RCSB.



The ribosome, as illustrated by the Molecule of the Month*'s David S. Goodsell*

The growing PDB archive contains the three-dimensional coordinates and related information about biological macromolecules. These structures, including proteins, nucleic acids, and large macromolecular complexes, have been determined using X-ray crystallography, nuclear magnetic resonance (NMR), and cryo-electron microscopy. PDB data are used by pharmaceutical and biotechnology industries and academic researchers in their search for new drugs and in efforts to understand the underlying causes of human disease. Understanding the shape of these structures and how they bind to various molecules may well provide the key to how they function.

Since the information about these structures is contained in a database, the RCSB PDB provides several tools for searching the database and creating tabular and graphical reports of the search results. Other resources for related aspects of structural biology, including structural genomics, data representation formats, downloadable software, and educational materials, are also made available.

In 2003, the RCSB formed a collaboration called the Worldwide PDB (wwPDB)[3] with the Macromolecular Structure Database at the EMBL's European Bioinformatics Institute (MSD-EBI), and Protein Data Bank Japan (PDBj). All wwPDB sites share responsibilities for data deposition, data processing, and distribution, and agree to support a single, standardized archive of structural data. (see also page 4)

PDB HOLDINGS

as of July 1, 2004

26144 Total Structures

Molecule Type

23676
proteins, peptides, and viruses

1338
nucleic acids

1112
protein/nucleic acid complexes

18
carbohydrates

Experimental Technique

22306
X-ray diffraction and other

3838
NMR

13216
structure factor files

1964
NMR restraint files

BIOLOGICAL MACROMOLECULES IN THE PDB

Each molecule has a role in a biological process, and many have several biological functions. The structures range from small pieces of protein or DNA to complex machines, such as the ribosome (pictured).

Among the most intricate and remarkable of nature's machines, the ribosome's structure includes more than 100,000 atoms, reflecting its complex and crucial function as the cell's protein-assembly factory.

Also represented in the PDB's archive are enzymes, viruses and molecular assemblies (such as the nucleosome). Their structures provide insight into these molecules' roles in fundamental biological processes and, in some cases, how they function in disease or drug interactions.

Educators and students
RCSB PDB member Kyle Burkhardt discusses protein structure with a high school class (from The Pingry School in Martinsville, NJ) at the RCSB PDB site at Rutgers.

RCSB PDB USERS

As the sole repository for three-dimensional structure data of biological macromolecules, the PDB is a critical resource for academic, pharmaceutical, and biotechnology research. Depositors from these groups provide these data, which are used by these groups to perform further research. Students and educators depend upon the RCSB PDB's collection of resources to help elucidate various

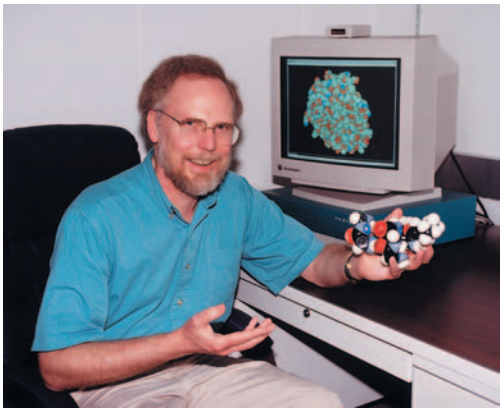
subjects related to biological structure. Some of these diverse user groups are highlighted here.

Educators and students use the resources made available from the website for their own research in structural biology. The RCSB prepares materials for use by all audiences, including a Molecule of the Month feature that highlights a different molecule in the PDB (see also page 8).

Structural biologists in industry and academia who focus on pharmaceutical and biotechnology research contribute PDB data as part of their work.

Computational biologists use data from the PDB to discover how macromolecules fold and function.

Scientists working in structural genomics use tools produced by the RCSB PDB. Structural genomics is a worldwide initiative aimed at determining a large number of protein structures in a high throughput mode. The RCSB PDB has developed tools to facilitate the rapid deposition of data produced by the structural genomics initiatives and has created databases to track the progress of their work.



Computational Biologists

J. Andrew McCammon, an HHMI Investigator at UCSD, studies biochemical reactions.

Structural Biologists Using X-ray Crystallography

Robert Sweet and others review a diffraction pattern at the National Synchrotron Light Source at Brookhaven National Laboratory. Collecting X-ray diffraction data at a synchrotron beamline is part of the process of determining a protein structure.

RCSB PDB MANAGEMENT

This resource is managed by three member institutions of the RCSB—Rutgers, The State University of New Jersey; the San Diego Supercomputer Center (SDSC), an

Structural Biologists using NMR

Ed S. Mooberry and Gregory Zornetzer stand in front of a bare Fleckvieh Varian Unity Inova 900 MHz spectrometer at the National Magnetic Resonance Facility at the University of Wisconsin-Madison. NMRFAM provides state of the art NMR spectrometer facilities used in structure determination to users locally and nationwide.

ADVISORY COMMITTEE

The RCSB PDB solicits the advice of several international committees. The members of the Advisory Committee are experts in a variety of areas, including X-ray crystallography, NMR, cryo-electron microscopy, bioinformatics, and education.

Stephen K. Burley (Chair)
Chief Scientific Officer & Senior Vice President, Research, Structural GenomiX

Frank Allen
Executive Director, Cambridge Crystallographic Data Centre

Edward N. Baker
Professor, University of Auckland

Wah Chiu
Professor, Baylor College of Medicine

Juli Feigon
Professor, University of California, Los Angeles

Nobuhiro Go
Professor, Japan Atomic Energy Research Institute

Barry Honig
Professor & HHMI Investigator, Columbia University

Robert Kaptein
Professor, Utrecht University

Sung-Hou Kim
Professor, University of California, Berkeley

Seth Pinsky
Senior Vice President, Global Research and Development & Chief Technical Officer, MDL

David Searls
Senior Vice-President of Bioinformatics, GlaxoSmithKline

Judith Voet
Professor, Swarthmore College

organized research unit of the University of California, San Diego (UCSD); and the Center for Advanced Research in Biotechnology (CARB), part of a joint venture between the University of Maryland Biotechnology Institute (UMBI) and the National Institute of Standards and Technology (NIST). The RCSB PDB is supported by funds from the National Science Foundation (NSF), the National Institute of General Medical Sciences (NIGMS), the Office of Science, Department of Energy (DOE), the National Library of Medicine (NLM), the National Cancer Institute (NCI), the National Center for Research Resources (NCRR), the National Institute of Biomedical Imaging and Bioengineering (NIBIB), and the National Institute of Neurological Disorders and Stroke (NINDS).

Five project leaders manage the operations of the RCSB PDB.

- Helen M. Berman, a Board of Governors Professor of Chemistry and Chemical Biology at Rutgers, is the overall Director of the RCSB PDB. She was part of the original team that developed the PDB at Brookhaven National Laboratory, and is a co-founder of the Nucleic Acid Database.

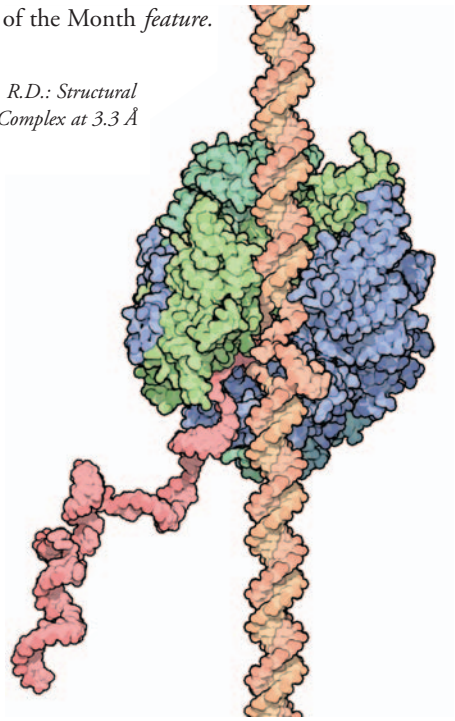
Three Co-Directors oversee activities at their respective sites:

- John Westbrook, Research Associate Professor of Chemistry at Rutgers University
- Philip E. Bourne, Professor of Pharmacology at UCSD and Adjunct Professor at the Burnham Institute
- Gary Gilliland, CARB Fellow, Adjunct Professor at UMBI, and NIST Research Chemist.

Judith L. Flippen-Anderson, formerly with the Naval Research Laboratory, is the Production and Outreach Leader across the three sites.

RNA Polymerase, from the April 2003 Molecule of the Month *feature.*
PDB ID: **1i6h**

Gnatt, A.L., Cramer, P., Fu, J., Bushnell, D.A., Kornberg, R.D.: Structural Basis of Transcription: An RNA Polymerase II Elongation Complex at 3.3 Å Resolution (2001) Science, ***292****, 1876.*



THE MISSION OF THE RCSB PDB TEAM

The RCSB seeks to enable science worldwide by offering a variety of resources to improve the understanding of structure-function relationships in biological systems. The RCSB believes that the availability of consistent, well-annotated three-dimensional data will facilitate new scientific advances. For the data to be truly useful, we must deliver it in a timely and efficient manner. To fulfill this mission, the capabilities of the RCSB PDB are continually being upgraded and significantly extended.

RCSB COLLABORATIONS: WWPDB

The RCSB, the Macromolecular Structure Database at the European Bioinformatics Institute (MSD-EBI), and the Protein Data Bank Japan (PDBj; at the Institute for Protein Research at Osaka University) have formed the Worldwide PDB (wwPDB).[3] The mission of the wwPDB is to maintain a single PDB archive of macromolecular structural data that is freely and publicly available to the global community. wwPDB formalizes the international character of the PDB and ensures that the archive will remain single and uniform. All three organizations serve as deposition, data processing and distribution sites for the PDB Archive. Each wwPDB site provides its own view of the primary data, thus providing a variety of tools and resources. Additionally, wwPDB members collaborate on key projects essential to the maintenance of the PDB archive.

MSD-EBI (**www.ebi.ac.uk/msd**) continues to provide weekly updates of the structures deposited and processed at their AutoDep site. A key accomplishment of this collaboration is the agreement on an exchange dictionary that is used by wwPDB members for data exchange.

Close collaborations have been maintained with PDBj (**www.pdbj.org**), where ADIT is used for structure deposition and annotation.

The RCSB, MSD-EBI, and PDBj are also collaborating on an PDBML/XML representation of PDB data based on the PDB exchange dictionary (page 9).

RCSB COLLABORATIONS: BMRB

The collaboration on NMR data deposition continues with the BioMagResBank (BMRB; **www.bmrb.wisc.edu**)[4], stressing the development of a data dictionary and an integrated deposition system based on ADIT (page 5).

RCSB PDB SERVICES: DATA DEPOSITION

DATA INPUT: DEPOSITION, VALIDATION, AND ANNOTATION

A key component of RCSB PDB operations is the efficient capture (deposition) and curation (validation and processing/annotation) of experimental structural data. Data from experiments using X-ray crystallography, NMR, cryo-electron microscopy and other methods are deposited in the PDB. Scientists can contribute their data using tools available at RCSB-Rutgers, PDBj, and MSD-EBI. Data are also accepted via FTP and e-mail. Data annotated at PDBj and MSD-EBI are sent to the RCSB for release.

The deposition tool ADIT (AutoDep Input Tool) is available online from the RCSB and PDBj sites and as a software download for standalone desktop use. ADIT provides a user interface to a collection of programs for data input, validation, annotation, and format exchange. The ADIT system uses the PDB exchange format that is based on the macromolecular Crystallographic Information File (mmCIF) dictionary. mmCIF is an ontology of more than 2,500 terms defining macromolecular structure and related experiments.[5]

After the annotation staff performs checks on the data, validation reports and a completed PDB file are returned to the depositor for review. Depositors also have the option of independently performing many of these checks using validation software available from the RCSB. When finalized, the complete entry, including its status information and PDB ID, is loaded into the core relational database. This entire process is completed with an average turnaround of less than two weeks. Depending upon the hold status selected by the depositor, data release occurs when a depositor gives approval (REL), the hold date has expired (HOLD), or the journal article has been published (HPUB). As of May 2003, there is a one-year limit on the length of a hold period, including HPUBs. If the citation for a structure is not published within the one-year period, depositors are given the option to either release or withdraw the deposition.

SOFTWARE FOR DEPOSITION AND VALIDATION

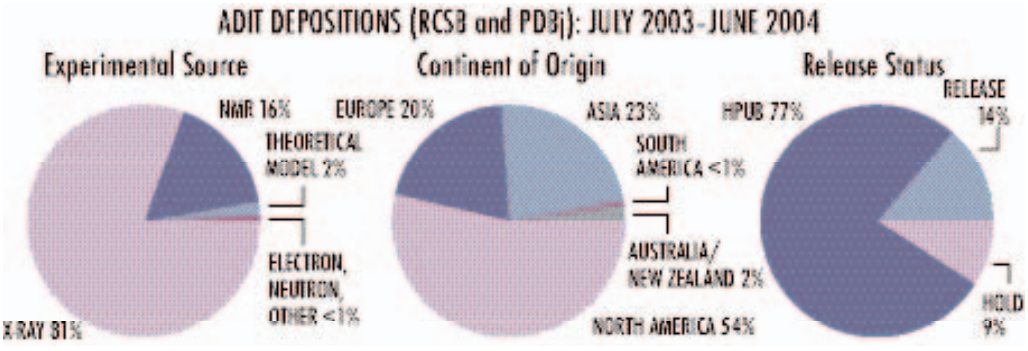
The RCSB PDB has developed a variety of tools to facilitate the validation and deposition of structural data by depositors. These tools include PDB_EXTRACT,[6] the PDB Validation Suite,[1] Ligand Depot,[7] and ADIT.[1]

These programs can be accessed from the RCSB PDB website. Some are also available for download and desktop installation from **deposit.pdb.org/software**. These programs are updated regularly and are available for a variety of platforms.

The RCSB and BMRB have collaborated on a version of ADIT that collects NMR experimental data. ADIT-NMR is available from the BMRB site, **www.bmrb.wisc.edu**.

STATISTICS

During the period covered by this report, 5153 files were deposited to the PDB archive through an international effort. Of the structures deposited during the period of this report, approximately 74% were deposited with experimental data. Around 56% of the depositions released their sequence data prior to the structure's release.



PDB_EXTRACT takes information about data collection, phasing, density modification, and the final structure refinement from the files produced by many crystal structure determination applications. The collected information is organized into mmCIF files that are ready for validation and deposition. In addition to the web-based and standalone versions available from the RCSB PDB, PDB_EXTRACT has been integrated with the CCP4i interface of the CCP4 program suite[8] for protein crystallography (version 5.0).

The **PDB Validation Suite** can be used to check the format of the coordinates and validate the overall structure before deposition. Sequence/coordinate alignment, missing and extra atoms or residues, and data inconsistencies are reported. The validation report also contains links to geometrical and experimental checks from the programs NUCHECK,[9] PROCHECK,[10] and SFCHECK.[11]

Ligand Depot integrates databases, services, tools and methods related to small molecules bound to macromolecules into an online data warehouse. It can be used to find codes for existing ligands, to link to PDB entries with a particular ligand, and to search for ligand substructures.

ADIT can be used to check file formats, run the PDB Validation Suite, and add or edit information for a deposition. An mmCIF file can be written from the desktop version of ADIT and uploaded to the web version or sent to the PDB for deposition.

RCSB PDB SERVICES: DATA DISTRIBUTION AND ACCESS

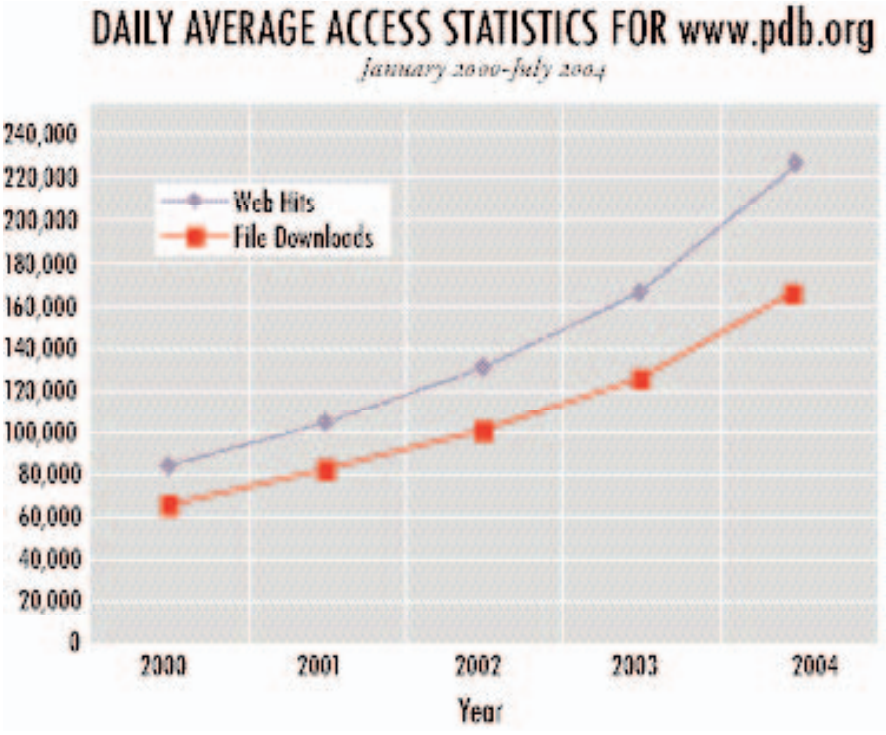
DATA DISTRIBUTION AND ACCESS

PDB data and RCSB services are available without cost through the Internet. The main RCSB PDB website at SDSC-UCSD receives an average of more than 220,000 hits per day from all over the world. On average, more than one file is downloaded every second, 24 hours a day, seven days a week. Additionally, there are six RCSB PDB mirror sites around the world at RCSB-Rutgers (US), RCSB-CARB (US), Osaka University (Japan), the National University of Singapore (Singapore), the Cambridge Crystallographic Data Centre (United Kingdom), and the Max Delbrück Center for Molecular Medicine (Germany). All RCSB sites are available 24 hours a day, seven days a week. New structures are added to the PDB holdings by 1:00 A.M. Pacific Daylight Time each Wednesday, 52 weeks per year.

There are a number of different ways to query the database. Entering the PDB ID of the target macromolecule in the search box on the home page performs the simplest search; the ID is assigned at the time of deposition, and is referenced in the journal article describing its structure. This search returns a Structure Explorer page that provides summary information about the entry as well as links to the atomic coordinates, derived geometric data, and experimental data (X-ray structure factors and NMR constraint data, where available). Links to structurally similar "neighbor" entries are also provided, along with options to study other aspects of the molecule, such as the secondary structure or primary amino acid sequence. Dynamic links to the structure's entry in other databases are provided by the Molecular Information Agent (MIA), and are accessible under the Other Sources section of the Structure Explorer page. Views of the structure's asymmetric and biologically active units are provided as static images generated by MolScript[12] and Raster3D,[13] and interactive views are presented in VRML, RasMol,[14,15] MICE,[16,17] Chime,[18] Swiss-Pdb Viewer,[19,20] STING,[21] and QuickPDB (Java).[22]

Multiple structures can be retrieved using the keyword search functionality on the RCSB PDB home page. SearchLite and the customizable SearchFields interfaces are provided to perform queries using parameters selected by the user. SearchFields can be used to limit the results to a subset of structures based upon the percentage of sequence identity. Files that list these clusters and their rankings at 50%, 70% and 90% sequence identity are available from the FTP server. QuickSearch is a new keyword search feature that searches across the archive and/or the static web pages.

The Query Result Browser lists all molecules that meet the user's query specifications, and allows exploration of one or more of the structures returned. Options to refine the query or create tabular reports from the search results are also available. A PDB or mmCIF format file for any structure can be downloaded as plain text or in one of several compressed formats. Files may also be downloaded from the FTP server.



REENGINEERED DATABASE AND WEBSITE

The RCSB PDB is dedicated to providing resources that fully utilize the data contained in the PDB archive.

To this end, we invite our users to participate in the beta test of the reengineered RCSB PDB site and database. The site was developed using feedback derived from the RCSB PDB help desk, conference attendees, focus groups and other personal interactions between our users and RCSB staff. The RCSB PDB has worked with both users and usability experts to improve navigation and ease of access. Experts with backgrounds in molecular and structural biology are analyzing the system for errors at each new phase of development.

This new system utilizes data improvements resulting from our ongoing efforts in uniformity and standardization (page 9). It offers improved accessibility for our diverse user community and will allow for rapid addition of new features based on input from users and in

response to the underlying evolution in science and technology.

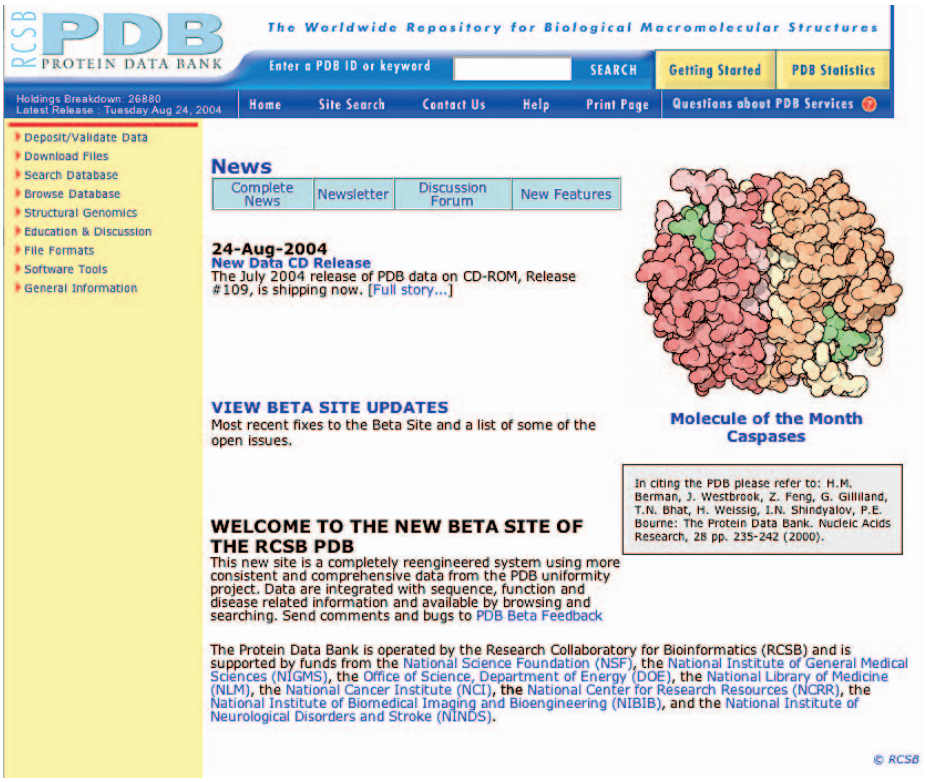
The new system implements an industry-standard three-tier system composed of very distinct database, application logic, and presentation layers. The partitioning of functionality between these layers and the use of Java Enterprise Edition (J2EE) creates a modular architecture that greatly simplifies maintenance and extensibility of this new system. It also enables us to incorporate readily available tools of high quality to improve existing components of the site, such as the keyword search engine and the help documentation, which has been converted to one dynamic, content-specific system supported by RoboHelp.[23]

We have loaded all of the data from our uniformity processing efforts and the calculated derived geometric structural features into a relational database that provides a single source of data for the reengineered site. This is a marked improvement over the current query system, which stores data in four separate databases and flat files that are integrated through a CGI layer. The new relational database implements the PDB exchange data dictionary schema (**pdbschema.sdsc.edu**).

The reengineering development process included an internal requirements analysis and active testing. The strengths, weaknesses, and major usage patterns of our current system were examined, documented, and used in planning the new site.

The beta test site is available at **pdbbeta.rcsb.org** alongside the current production site. Both sites are updated once a week.

The RCSB is very excited about this resource, and encourages the PDB community to test the new site. We look forward to your feedback at **betafeedback@rcsb.org**.



The RCSB PDB beta test home page at **pdbbeta.rcsb.org**

WHAT CAN YOU DO WITH THE BETA RCSB PDB WEBSITE?

- Browse for structures by exploring various features associated with sequence, structure, and function.
- Search for ligands and review protein-ligand interactions.
- Query and browse data incorporated from external sources for extensive cross-referencing within the database. Examples include browsing by genome location, taxonomy, Enzyme Commission number, SCOP,[25] CATH,[26] MEDLINE,[24] and Gene Ontology[27] terms and locating structures by their relationship to known diseases.
- Utilize high quality visualization features based on the molecular biology toolkit (**mbt.sdsc.edu**), KiNG (Kinemage Next Generation; **kinemage.biochem.duke.edu/**), Jmol (open source Java Viewer), and WebMol.[28]
- Search the database using improved tools and produce reports organized like a scientific paper, with sections materials and methods, structure details, biology and chemistry, references.
- Navigate the site easily using a new help system, site search browser, alphabetical site index, improved navigation tools, and recall of previous queries.
- Review more comprehensive structure summaries on screen and in hand using the printer-friendly option.
- Create more detailed and sortable reports for search results containing multiple structures.
- Search database using queries based upon sequence, 3D structure of ligands, and proteins.

OTHER RCSB PDB SERVICES

OUTREACH

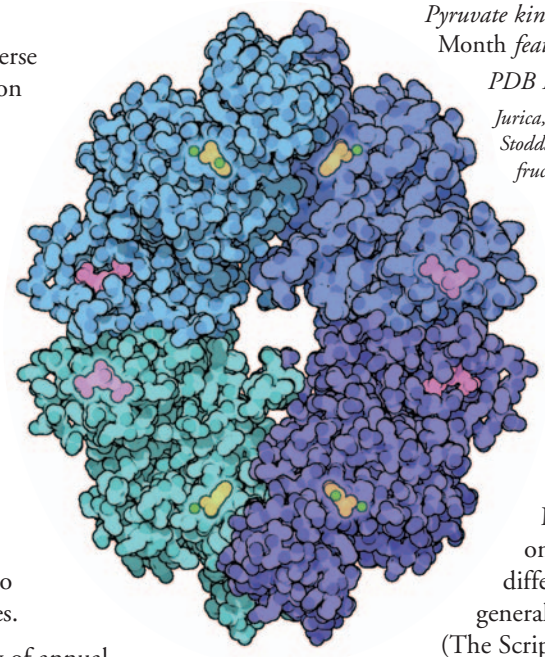
The RCSB PDB interacts with its diverse user community to provide information about the resource, gain feedback for development, and provide materials that promote a broader understanding of structural biology. The RCSB staff responds to general and specific inquiries sent to help desks, and offers a listserv forum for exchange among members of the PDB community.

Recent developments and activities are announced weekly on the RCSB PDB website. The RCSB PDB distributes a quarterly newsletter and a variety of online and print materials about the resource. Our efforts are also described in journal papers and articles.

The RCSB PDB exhibited at a variety of annual meetings, including the American Crystallographic Association (ACA), the Biophysical Society, the Experimental Nuclear Magnetic Resonance Conference, and the International Conference on Intelligent Systems for Molecular Biology. Staff members presented talks, demonstrations, and posters at more than thirty meetings around the world, including the National Science Teachers Association's annual meeting.

The 2004 RCSB PDB Poster Prize recognized student poster presentations involving macromolecular crystallography at each of the meetings of the International Union of Crystallography's Regional Associates–the ACA, the Asian Crystallographic Association, and the European Crystallographic Association. The prize was also awarded at the Annual International Conference on Research in Computational Molecular Biology (RECOMB).

The RCSB PDB's "Art of Science" art exhibit highlights various representations of proteins found in the PDB, including large-scale depictions of the images available from Structure Explorer pages and pictures from the Molecule of the Month series. It traveled to the University of Wisconsin-Madison and California State University, Fullerton during the period of this report.



Pyruvate kinase, from the February 2004 Molecule of the Month *feature on glycolytic enzymes.*

PDB ID: ***1a3w***

Jurica, M.S. Mesecar, A. Heath, P.J. Shi, W. Nowak, T. Stoddard, B.L. The allosteric regulation of pyruvate kinase by fructose-1,6-bisphosphate. (1998) Structure, ***6****, 195-210.*

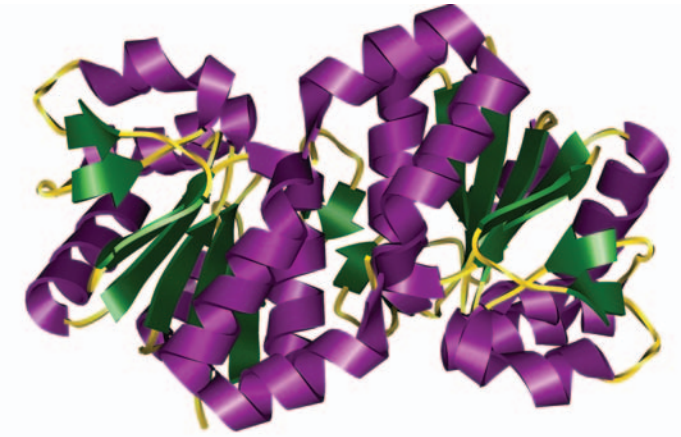
MOLECULE OF THE MONTH AND EDUCATIONAL RESOURCES

Emphasis is placed on educating new users, students, educators, and the general public. The RCSB maintains an extensive portal to educational resources for different levels of expertise. The Molecule of the Month column highlighted on the RCSB PDB home page focuses on a different biological molecule each month for a general audience. Produced by David S. Goodsell (The Scripps Research Institute), each installment includes an introduction to the structure and function of the molecule, a discussion of the relevance of the molecule to human health and welfare, and suggestions for how visitors might view these structures and access further details.

Curricular materials for educators are solicited and posted on the RCSB PDB website and highlighted in our newsletters. Tutorials and help documentation are also available on the website, with enhanced user help features available on the RCSB PDB beta test site.

STRUCTURAL GENOMICS

Structural genomics is a worldwide initiative aimed at determining a large number of protein structures in a high throughput mode. The RCSB PDB continues to be actively involved in developing the informatics of structural genomics projects. This



Crystal Structure Of Human DJ-1.

PDB ID: ***1p5f***

Wilson, M. A., Collins, J. L., Hod, Y., Ringe, D., Petsko, G. A.: The 1.1 Å Resolution Crystal Structure of DJ-1, the Protein Mutated in Autosomal Recessive Early Onset Parkinson's Disease. (2003) Proc.Nat.Acad.Sci.USA ***100****, 9256-9261.*

effort includes participation in task forces, meetings, and individual interactions with each of the structural genomics centers. The RCSB PDB is responsible for maintaining the proposed data deposition specifications and the data dictionary supporting these specifications (**deposit.pdb.org/mmcif**). In addition, the RCSB PDB has created software to help integrate data from standard structure determination packages, and has created the resources Target Registration Database (TargetDB)[29] and PepcDB.

Efficient structure solution on a genomic scale requires a centralized coordination effort, to which the timely availability of status information on the progress of protein production and structure solution is key. TargetDB collects target sequences from the nine P50 NIH structural genomics centers and from other genomics centers worldwide. As of this writing, more than 65,400 sequences have been entered in TargetDB. PepcDB, the Protein Expression Purification and Crystallization Database, extends the content of TargetDB with status history, stop conditions, reusable text protocols and contact information collected from the NIH Protein Structure Initiative (PSI) Centers. A demonstration site has been created (**pepcdb.rutgers.edu**).

Solution Structure Of A Dimeric Lactose DNA-Binding Domain Complexed To A Nonspecific DNA Sequence.

PDB ID: ***1osl***

Kalodimos, C.G., Biris, N., Bonvin, A.M.J.J., Levandoski, M.M., Guennuegues, M., Boelens, R., Kaptein, R. Structure and Flexibility Adaptation in Nonspecific and Specific Protein-DNA Complexes. (2004) Science ***305****, 386-389.*

DATA UNIFORMITY: RELEASE OF A STANDARDIZED ARCHIVE

The RCSB PDB's focus on making the archive as consistent and error-free as possible is ongoing.[30-31] This work involves continuous archive-wide standardization of nomenclature and usage within existing data items. Standardization is required to support reliable comparisons of data across the archive as well as for integration with other database resources.

All PDB entries have been standardized and released in mmCIF format. These remediated files can be accessed from the Download/Display File section of the Structure Explorer page for any entry, or for a set of query results. The remediated mmCIF files are also available from the FTP site. The files follow the latest version of the PDB exchange data dictionary (**deposit.pdb.org/mmcif**) that was developed by the RCSB PDB and the MSD-EBI. Software for translating these data into PDB-formatted files can be downloaded from **deposit.pdb.org/software**.

All remediated data files are also available in PDBML/XML format from the beta FTP site and through queries on the production site. The PDBML/XML data files have been created by software-facilitated translation of the remediated mmCIF data files. As a result, the element and attribute names in the PDBML/XML data files directly correspond to the item names defined in the PDB Exchange Dictionary. An PDBML/XML schema documenting the translated format is available at **deposit.pdb.org/mmcif**. The delivery of PDB data in PDBML/XML format is the product of a collaboration among PDBj, MSD-EBI, and the RCSB.

DATA CD DISTRIBUTION

The RCSB PDB currently distributes a quarterly CD-ROM of PDB holdings upon request. All PDB entries and related experimental data are distributed in the January release followed by incremental updates in the remaining periods. Provided at no cost to users, the CD-ROMs are designed to help researchers who have limited Internet access or need subsets or a complete set of the structure files for their research.

PHYSICAL ARCHIVES

The PDB Physical Archives contain the paper documents, magnetic tapes, and other materials generated by BNL and the RCSB PDB over the course of its history. The archives continue to be called on periodically to resolve issues involving older structures. Older data have been restored and incorporated into a file structure that was designed for easy maintenance and expansion. The overall goal of the PDB Physical Archives is to preserve not only the data submitted by the depositors, but also the records associated with the transactions and activities that are part of the evolution of the resource. Access and availability of this information to the RCSB PDB staff provides a resource for resolving issues concerning specific entries, aides in uniformity and value-added annotation, facilitates disaster recovery, and makes the information available for research.

RECENT PUBLICATIONS

Berman, H.M. (2003) Structural databases of biological macromolecules. In *Encyclopedia of the Human Genome*. Nature Publishing Group Macmillian Publishers Ltd., London pp. 406-412.

Berman, H.M., Bourne, P.E., and Westbrook, J. (2004) The Protein Data Bank: A case study in management of community data. *Current Proteomics* **1**, 49-57.

Berman, H.M., Bourne, P.E., Westbrook, J., and Zardecki, C. (2003) The Protein Data Bank. In *Protein Structure. Determination, analysis, and applications for drug discovery.* Chasman, D. (ed.), Marcel Dekker, Inc., New York, NY pp. 389-405.

Berman, H.M., Henrick, K., and Nakamura, H. (2003) Announcing the worldwide Protein Data Bank. *Nat Struct Biol* **10**, 980.

Berman, H.M., and Westbrook, J. (2003) The need for dictionaries, ontologies, and controlled vocabularies. OMICS. *A Journal of Integrative Biology* 7, 9-10.

Berman, H.M., and Westbrook, J. (2004) The impact of structural genomics on the Protein Data Bank. *Am J Pharmacogenomics* **4**, 257-252.

Bourne, P.E., Addess, K.J., Bluhm, W.F., Chen, L., Deshpande, N., Feng, Z., Fleri, W., Green, R., Merino-Ott, J.C., Townsend-Merino, W., Weissig, H., Westbrook, J., and Berman, H.M. (2004) The distribution and query systems of the RCSB Protein Data Bank. *Nucleic Acids Res* **32**, D223-D225.

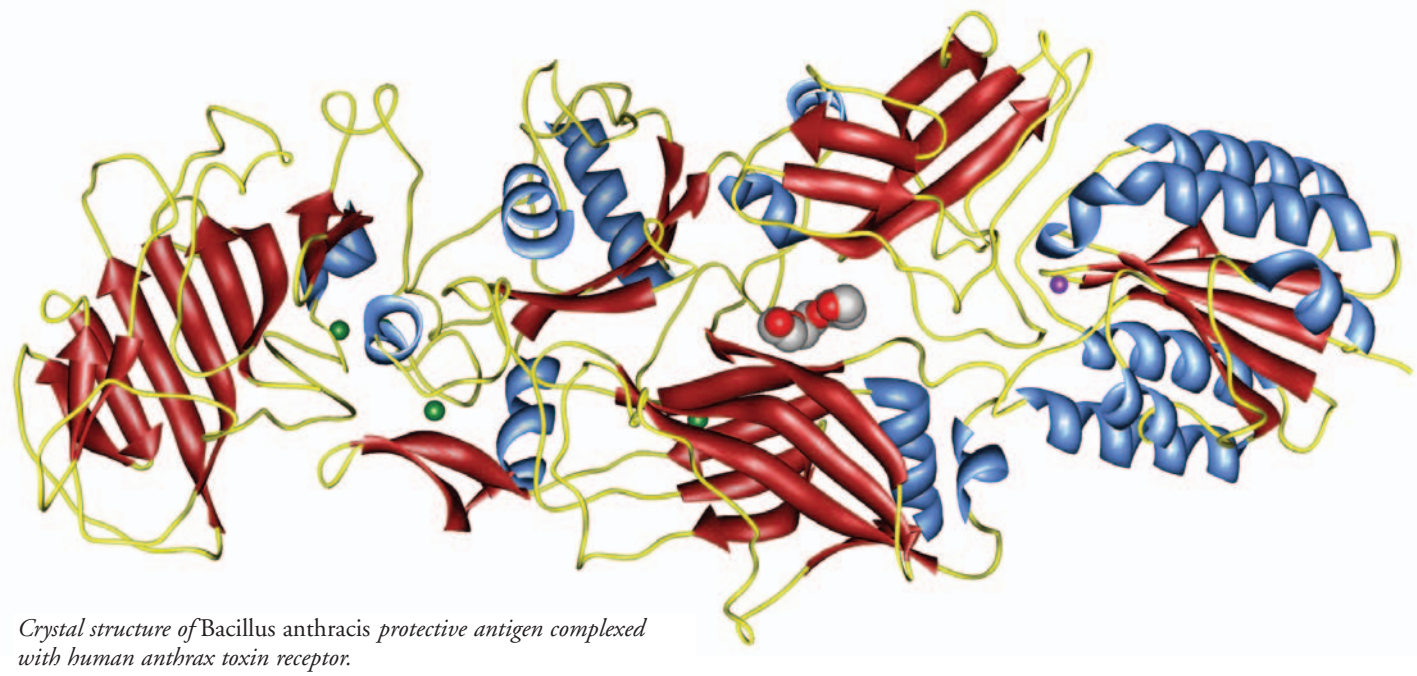
Bourne, P.E., Westbrook, J., and Berman, H.M. (2004) The Protein Data Bank and lessons in data management. *Brief Bioinform* **5**, 23-30.

Chen, L., Oughtred, R., Berman, H.M., and Westbrook, J. (2004) TargetDB: A target registration database for structural genomics projects. *Bioinformatics* **bioinformatics.oupjournals.org/cgi/content/abstract/bth300v1**.

Feng, Z., Chen, L., Maddula, H., Akcan, O., Oughtred, R., Berman, H.M., and Westbrook, J. (2004) Ligand Depot: A Data Warehouse for Ligands Bound to Macromolecules. Bioinformatics **20**, 2153-2155.

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

Westbrook, J., Feng, Z., Chen, L., Yang, H., and Berman, H.M. (2003) The Protein Data Bank and structural genomics. *Nucleic Acids Res* **31**, 489-491.



Crystal structure of Bacillus anthracis *protective antigen complexed with human anthrax toxin receptor.*

PDB ID: ***1t6b***

Santelli, E., Bankston, L. A., Leppla, S. H., Liddington, R. C. (2004) Crystal Structure of a Complex Between Anthrax Toxin and its Host Cell Receptor. Nature, ***430****, 905-908.*

IMPORTANT WEB ADDRESSES

BETA TEST SITE
pdbbeta.rcsb.org

DATA UNIFORMITY
www.rcsb.org/pdb/uniformity

DEPOSITION SERVICES INFORMATION
deposit.pdb.org

EDUCATION RESOURCES
www.rcsb.org/pdb/education.html

FTP
ftp://ftp.rcsb.org

MMCIF RESOURCES
deposit.pdb.org/mmcif

MOLECULE OF THE MONTH
www.rcsb.org/pdb/molecules/molecule_list.html

PEPCDB
pepcdb.rutgers.edu

RCSB PROTEIN DATA BANK
www.pdb.org

RCSB PDB NEWSLETTER
www.rcsb.org/pdb/newsletter.html

STRUCTURAL GENOMICS
www.rcsb.org/pdb/strucgen.html

SOFTWARE RESOURCES
www.rcsb.org/pdb/software-list.html

TARGETDB
targetdb.pdb.org

PDBML/XML BETA DATA FILES
ftp://beta.rcsb.org/pub/pdb/uniformity/data/XML

WWPDB
www.wwpdb.org

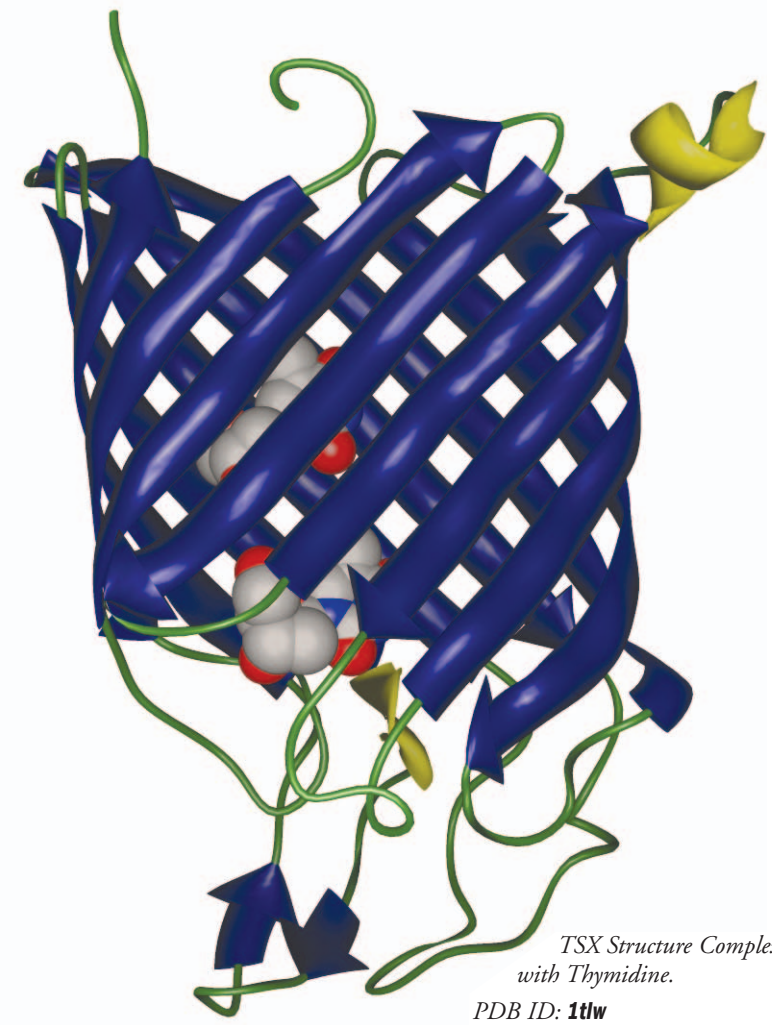
HELP DESKS

BETA SITE FEEDBACK
betafeedback@rcsb.org

GENERAL PDB INQUIRIES
info@rcsb.org

GENERAL DEPOSITION QUESTIONS
deposit@rcsb.rutgers.edu

ADIT INFORMATION
help@rcsb.rutgers.edu



TSX Structure Complexed with Thymidine.

PDB ID: ***1tlw***

Ye, J., van den Berg, B.: Crystal Structure of the Bacterial Nucleoside Transporter TSX. (2004) EMBO J. ***23****, 3187-3195.*

SELECTED REFERENCES

1. Berman, H.M., Westbrook, J., Feng, Z., Gilliland, G., Bhat, T.N., Weissig, H., Shindyalov, I.N. and Bourne, P.E. (2000) The Protein Data Bank. *Nucleic Acids Res.*, **28**, 235-242.
2. Bernstein, F.C., Koetzle, T.F., Williams, G.J.B., Meyer Jr., E.F., Brice, M.D., Rodgers, J.R., Kennard, O., Shimanouchi, T. and Tasumi, M. (1977) Protein Data Bank: a computer-based archival file for macromolecular structures. *J. Mol. Biol.*, **112**, 535-542.
3. Berman, H.M., Henrick, K. and Nakamura, H. (2003) Announcing the worldwide Protein Data Bank. *Nat. Struct. Biol.*, **10**, 980.
4. Doreleijers, J.F., Mading, S., Maziuk, D., Sojourner, K., Yin, L., Zhu, J., Markley, J.L. and Ulrich, E.L. (2003) BioMagResBank database with sets of experimental NMR constraints corresponding to the structures of over 1400 biomolecules deposited in the Protein Data Bank. *J. Biomol. NMR*, **26**, 139-146.
5. Bourne, P.E., Berman, H.M., Watenpaugh, K., Westbrook, J.D. and Fitzgerald, P.M.D. (1997) The macromolecular Crystallographic Information File (mmCIF). *Meth. Enzymol.*, **277**, 571-590.
6. Yang, H., Guranovic, V., Dutta, S., Feng, Z., Berman, H.M. and Westbrook, J. (2004) Automated and accurate deposition of structures solved by X-ray diffraction to the Protein Data Bank. *Acta Crystallogr. D Biol. Crystallogr.,* **60**, 1833-1839.
7. Feng, Z., Chen, L., Maddula, H., Akcan, O., Oughtred, R., Berman, H.M. and Westbrook, J. (2004) Ligand Depot: A Data Warehouse for Ligands Bound to Macromolecules. *Bioinformatics*, 20, 2153-2155.
8. CCP4. (1994) The Collaborative Computational Project Number 4 suite: programs for protein crystallography. *Acta Crystallogr. D Biol. Crystallogr.,* **50**, 760-763.
9. Feng, Z., Westbrook, J. and Berman, H.M. (1998) NUCHECK. Rutgers University, New Brunswick, NJ.
10. Laskowski, R.A., Rullmann, J.A., MacArthur, M.W., Kaptein, R. and Thornton, J.M. (1996) AQUA and PROCHECK-NMR: programs for checking the quality of protein structures solved by NMR. *J. Biomol. NMR*, **8**, 477-486.
11. Vaguine, A.A., Richelle, J. and Wodak, S.J. (1999) SFCHECK: a unified set of procedures for evaluating the quality of macromolecular structure-factor data and their agreement with the atomic model. *Acta Crystallogr D Biol Crystallogr.*, **55**, 191-205.
12. Kraulis, P. (1991) MOLSCRIPT: a program to produce both detailed and schematic plots of protein structures. *J. Appl. Cryst.*, **24**, 946-950.
13. Merrit, E.A. and Bacon, D.J. (1997) Raster3D: Photorealistic molecular graphics. *Meth. Enzymol.*, **277**, 505-524.
14. Sayle, R. and Milner-White, E.J. (1995) RasMol: biomolecular graphics for all. *Trends Biochem. Sci.*, **20**, 374.
15. Bernstein, H.J. (2000) Recent changes to RasMol, recombining the variants. *Trends Biochem. Sci.*, **25**, 453-455.
16. Bourne, P.E., Gribskov, M., Johnson, G., Moreland, J. and Weissig, H. (1998) A prototype Molecular Interactive Collaborative Environment (MICE). In Altman, R., Dunker, K., Hunter, L. and Klein, T. (eds.), *Pacific Symposium on Biocomputing*, pp. 118-129.
17. Tate, J.G., Moreland, J. and Bourne, P.E. (1999) MSG (Molecular Scene Generator): a Web-based application for the visualization of macromolecular structures. *Journal of Applied Crystallography*, **32**, 1027-1028.
18. MDL Information Systems Inc. Chime. United States.
19. Kaplan, W. and Littlejohn, T. (2001) Swiss-PDB Viewer (Deep View). *Brief Bioinform*, **2**, 195-197.
20. Guex, N. and Peitsch, M.C. (1997) SWISS-MODEL and the Swiss-PdbViewer: An environment for comparative protein modeling. *Electrophoresis*, **18**, 2714-2723.
21. Neshich, G., Togawa, R., Vilella, W. and Honig, B. (1998) STING (Sequence To and withIN Graphics) PDBViewer. *Protein Data Bank Quarterly Newsletter*, **85**, 6-7.
22. Shindyalov, I. and Bourne, P.E. Applet QuickPDB. ©1996-1998 San Diego Supercomputer Center.
23. eHelp Corporation. Robohelp. ©1992-2003. All Rights Reserved.
24. National Library of Medicine. (1989-) MEDLINE [database online]. Bethesda (MD). Updated weekly. Available from: National Library of Medicine; OVID, Murray, UT; The Dialog Corporation, Palo Alto, CA.x.
25. Conte, L., Bart, A., Hubbard, T., Brenner, S., Murzin, A. and Chothia, C. (2000) SCOP: a structural classification of proteins database. *Nucleic Acids Res.*, **28**, 257-259.
26. Orengo, C.A., Michie, A.D., Jones, S., Jones, D.T., Swindells, M.B. and Thornton, J.M. (1997) CATH–a hierarchic classification of protein domain structures. *Structure*, **5**, 1093-1108.
27. The Gene Ontology Consortium. (2000) Gene Ontology: tool for the unification of biology. *Nature Genetics*, **25**, 25-29.
28. Walther, D. (1997) WebMol--a Java-based PDB viewer. *Trends Biochem Sci*, **22**, 274-275.
29. Chen, L., Oughtred, R., Berman, H.M. and Westbrook, J. (2004) TargetDB: A target registration database for structural genomics projects. *Bioinformatics*, **bioinformatics.oupjournals.org/cgi/content/abstract/bth300v1**.
30. Westbrook, J., Feng, Z., Jain, S., Bhat, T.N., Thanki, N., Ravichandran, V., Gilliland, G.L., Bluhm, W., Weissig, H., Greer, D.S. et al. (2002) The Protein Data Bank: Unifying the archive. *Nucleic Acids Res.*, **30**, 245-248.
31. Bhat, T.N., Bourne, P., Feng, Z., Gilliland, G., Jain, S., Ravichandran, V., Schneider, B., Schneider, K., Thanki, N., Weissig, H. et al. (2001) The PDB data uniformity project. *Nucleic Acids Res.*, **29**, 214-218.

MIRRORS

San Diego Supercomputer Center, UCSD

www.pdb.org
ftp://ftp.rcsb.org

Rutgers, The State University of New Jersey

rutgers.rcsb.org

Center for Advanced Research in Biotechnology/UMBI/NIST

nist.rcsb.org

Cambridge Crystallographic Data Centre, UK

pdb.ccdc.cam.ac.uk
ftp://pdb.ccdc.cam.ac.uk/rcsb

National University of Singapore

pdb.bic.nus.edu.sg
ftp://pdb.bic.nus.edu.sg/pub/pdb

Osaka University, Japan

pdb.protein.osaka-u.ac.jp
ftp://ftp.protein.osaka-u.ac.jp/pub/pdb

Max Delbrück Center for Molecular Medicine, Germany

www.pdb.mdc-berlin.de
ftp://ftp.pub.mdc-berlin.de/pub/pdb

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
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

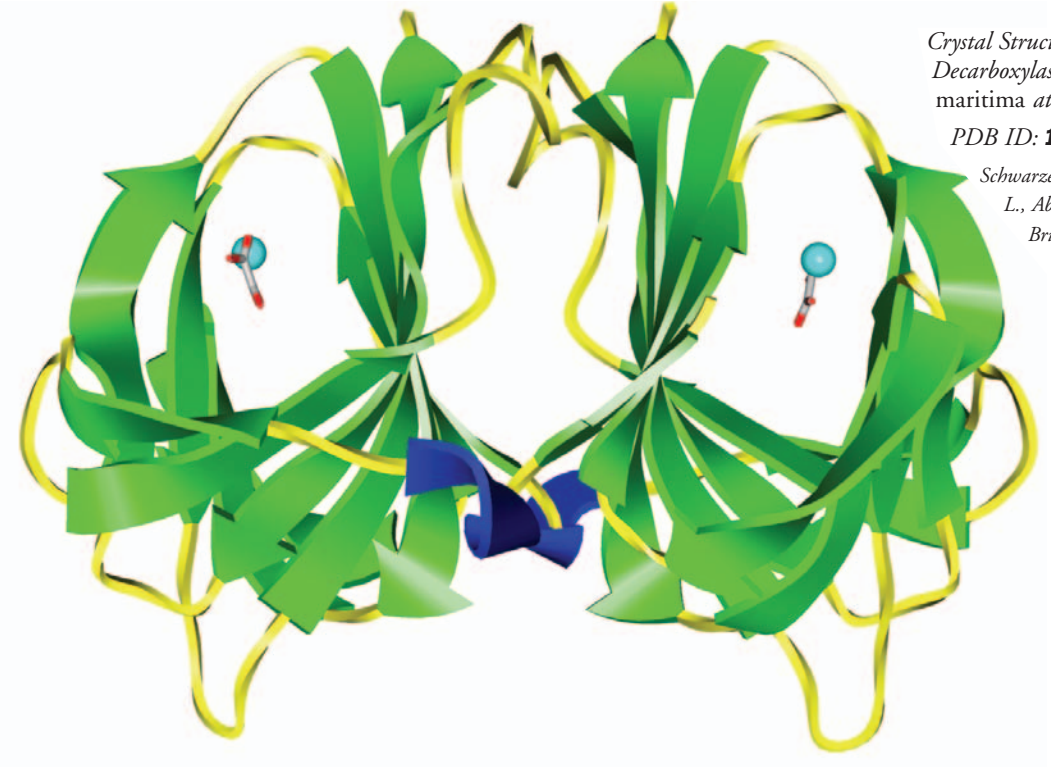
A list of current RCSB PDB Team Members is available at
www.rcsb.org/pdb/rcsb-group.html

RCSB PDB PARTNERS

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

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

Center for Advanced Research in Biotechnology, part of a joint venture between the University of Maryland Biotechnology Institute and the National Institute of Standards and Technology
9600 Gudelsky Drive
Rockville, MD 20850-3479



Crystal Structure Of Putative Oxalate Decarboxylase (TM1287) from Thermotoga maritima *at 1.95 Å Resolution.*

PDB ID: **1o4t**.

Schwarzenbacher, R., Von Delft, F., Jaroszewski, L., Abdubek, P., Ambing, E., Biorac, T., Brinen, L. S., Canaves, J. M., Cambell, J., Chiu, H. J., Dai, X., Deacon, A. M., Didonato, M., Elsliger, M. A., Eshagi, S., Floyd, R., Godzik, A., Grittini, C., Grzechnik, S. K., Hampton, E., Karlak, C., Klock, H. E., Koesema, E., K, J. S.: Crystal Structure of a Putative Oxalate Decarboxylase (TM1287) from Thermotoga maritima *at 1.95 Å Resolution. (2004)* Proteins **56**, *392.*

www.pdb.org

Send questions or comments to:

info@rcsb.org