RCSB PDB
PROTEIN DATA BANK

ANNUAL REPORT

July 2002 - June 2003

RESEARCH COLLABORATORY FOR STRUCTURAL BIOINFORMATICS

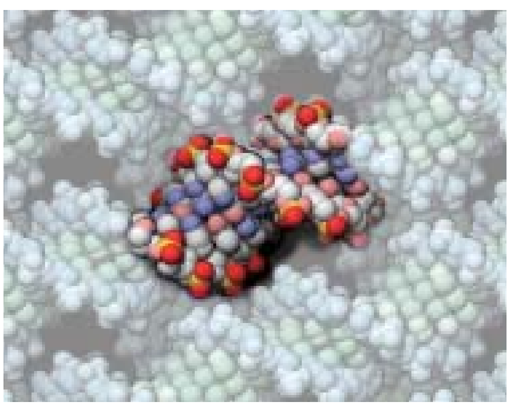
Rutgers, The State University of New Jersey

San Diego Supercomputer Center at the University of California, San Diego

Center for Advanced Research in Biotechnology of the National Institute of Standards and Technology

CONTENTS

ABOUT THE COVER



April 25, 2003, marked the 50th anniversary of the publication of the description of the structure of the double helix. Since that time, the DNA double helix has become an icon of modern life and its atomic structure has been further elucidated through extensive study with a variety of techniques, including fiber diffraction, NMR spectroscopy, and X-ray crystallography. These studies have been extended to include the interaction of DNA with various drugs and proteins. Many such structures can be found in the PDB. The cover image depicts the crystal-packing of a DNA decamer from entry ***196d****. The end-to-end packing of multiple copies of the small oligonucleotide creates the appearance of long, continuous double helices that mimic natural DNA.*

PDB ID: 196d

Goodsell, D. S., Grzeskowiak, K., Dickerson, R. E. (1995): Crystal structure of C-T-C-T-C-G-A-G-A-G. Implications for the structure of the Holliday junction. Biochemistry *34, pp. 1022-9.*

RCSB PDB PARTNER SITES

San Diego Supercomputer Center at the University of California San Diego, La Jolla, CA, USA

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

Rutgers, The State University of New Jersey, Piscataway, NJ, USA

rutgers.rcsb.org

Center for Advanced Research in Biotechnology of the National Institute of Standards and Technology, Rockville, MD, USA

nist.rcsb.org

OTHER RCSB PDB MIRRORS

Cambridge Crystallographic Data Centre, United Kingdom

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

National University of Singapore, 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

Universidade Federal de Minas Gerais, Brazil

www.pdb.ufmg.br
ftp://vega.cenapad.ufmg.br/pub/pdb

Max Delbrück Center for Molecular Medicine, Germany

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

DEPOSITION SITES

ADIT

RCSB
deposit.pdb.org/adit

Osaka University
pdbdep.protein.osaka-u.ac.jp/adit

AutoDep

EBI
autodep.ebi.ac.uk

MESSAGE FROM THE RCSB PDB

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

As the sole repository for three-dimensional structure data of biological macromolecules, the Protein Data Bank (PDB) is a critical resource for academic, pharmaceutical, and biotechnology research. Students and educators also depend upon the PDB's diverse collection of resources to help elucidate various subjects related to biological structure.

The PDB's primary mission is to provide accurate, well-annotated data in the most timely and efficient way possible to facilitate new discoveries and advances in science.

The PDB 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) of the National Institute of Standards and Technology (NIST).

This report highlights the PDB's accomplishments from July 1, 2002, through June 30, 2003. It is also designed to serve as a useful guide to the history, contents, and use of the PDB. As a popular and publicly available biological database, the PDB server is extremely active: in an average month, approximately 351 structures are deposited and processed, 303 structures are released, and 5 million files of individual structure entries are downloaded from the archive.

Working with colleagues and collaborators worldwide, the PDB team has accomplished a great deal in the past year, including:

- Released standalone software for the preparation, validation, and deposition of macromolecular data.
- Developed a tool for joint deposition of NMR data with the BioMagResBank (BMRB).
- Released the PDB archive in XML format, a product of a collaboration with the Protein Data Bank Japan (PDBj), and the Macromolecular Structure Database (MSD) group at the European Bioinformatics Institute (EBI).
- Released files and images of the biologically active units for all structures in the archive.
- Established a new mirror site in Berlin, Germany.
- Continued to improve the efficiency of the production data processing and query systems.

Our project to re-engineer and redesign the PDB database, Web site, and query system has progressed to the alpha stage. This effort will introduce new capabilities for browsing and searching the PDB archive, improved usability and navigation of the Web site, and more reliable query results.

Some members of the RCSB PDB Team: (front row, left to right) Nita Deshpande, Elizabeth Walker, Rose Oughtred, Sharon Cousin, Jasmin Yang, Irina Persikova, Lew-Christiane Fernandez, Suzanne Richman, Zukang Feng, Shuchismita Dutta, Tammy Battistuz, (middle row, left to right) Li Chen, Kyle Burkhardt, Prentice Bisbal, Melcoir Rosas, Gary L. Gilliland, Judith L. Flippen-Anderson, Christine Zardecki, Helen M. Berman, (back row, left to right) Wayne Townsend-Merino, Philip E. Bourne, John Westbrook, Veerasamy Ravichandran, Wolfgang Bluhm, Al Carlson, Vladimir Guranovic, Shri Jain, Huanwang Yang, Jeff Merino-Ott

Ongoing services in the areas of data deposition, annotation, and distribution, as well as community outreach and education, continue to grow and to be successful.

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

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

for the RCSB Protein Data Bank

WHAT IS THE PDB?

EXPERIMENTAL TECHNIQUE	PROTEINS, PEPTIDES & VIRUSES	PROTEIN-NUCLEIC ACID COMPLEXES	NUCLEIC ACIDS	CARBOHYDRATES	TOTAL
X-RAY DIFFRACTION & OTHER	16,784	809	675	14	18,282
NMR	2,667	93	526	4	3,290
TOTAL	19,451	902	1,201	18	21,572

PDB HOLDINGS AS OF JULY 1, 2003

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 has been managed by three member institutions of the RCSB.

The PDB processes, stores, and disseminates structural coordinates and related information about proteins, nucleic acids, and protein-nucleic acid complexes. Some examples of the types of structures that can be found in the PDB archive are DNA, RNA, viruses, and ribosomes. The PDB also provides resources for related aspects of structural biology, including structural genomics, data representation formats, downloadable software, and educational materials.

Why is it important?

Understanding the shape of a macromolecule aids in understanding how it functions. The data contained in the PDB assist the pharmaceutical and biotechnology industries in understanding diseases and developing drugs. Medical researchers use the three-dimensional structures of proteins and other biological macromolecules to unlock their therapeutic potential. Scientists also use PDB structural information in research directed at understanding the chemistry and biochemistry of natural processes. These efforts require the most consistent, well-annotated information available about the atomic structure of complex biological molecules.

Research in the area of structural genomics, where scientists attempt to elucidate as many structures as possible using high throughput techniques, is expected to lead to a greater understanding of the relationship between structure and function. The PDB is adapting the most recent technologies available to facilitate the collection, validation, annotation, organization, and distribution of the tremendous amount of data that will be generated by these newly evolving techniques.

WHAT CAN YOU DO WITH THE PDB?

- Access pictures and coordinates for the biologically active unit of a structure.
- Find all structures with a particular ligand, enzyme classification (EC) number, or by a specific citation author, directly from an entry's Structure Explorer page.
- Perform a sophisticated keyword query that allows use of a wildcard and checks spelling.
- Search for bound ligands.
- Study a *Molecule of the Month*–such as Green Fluorescent Protein–and find examples of this structure in the PDB.
- Locate information about a protein that was described in a journal article in the PDB and in related resources and databases.
- Download software to validate data and prepare depositions on your own computer.
- Search for sequences of unreleased structures.
- Explore the sequence and secondary structure of a single molecule or a group of molecules.
- Compare a molecule with its structurally similar neighbors.

How is the PDB managed?

The PDB is managed by Rutgers, The State University of New Jersey; the San Diego Supercomputer Center (SDSC), an organized research unit of the University of California, San Diego (UCSD); and the Center for Advanced Research in Biotechnology (CARB) of the National Institute of Standards and Technology (NIST)—three member institutions of the RCSB. The RCSB PDB is funded by the National Science Foundation (NSF), the Department of Energy (DOE), and the National Institutes of Health (NIH).

The RCSB PDB project leaders manage the overall operation of the PDB. Helen M. Berman, a Board of Governors Professor of Chemistry and Chemical Biology at Rutgers, is the Director of the PDB. She was part of the original team that developed the PDB at Brookhaven National Laboratory, and is the 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 Director of Integrative Biosciences at SDSC; and Gary L. Gilliland, Associate Director of CARB. Judith L. Flippen-Anderson, formerly with the Naval Research Laboratory, is the Production and Outreach Leader across the three sites.

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 PDB are continually being upgraded and significantly extended.

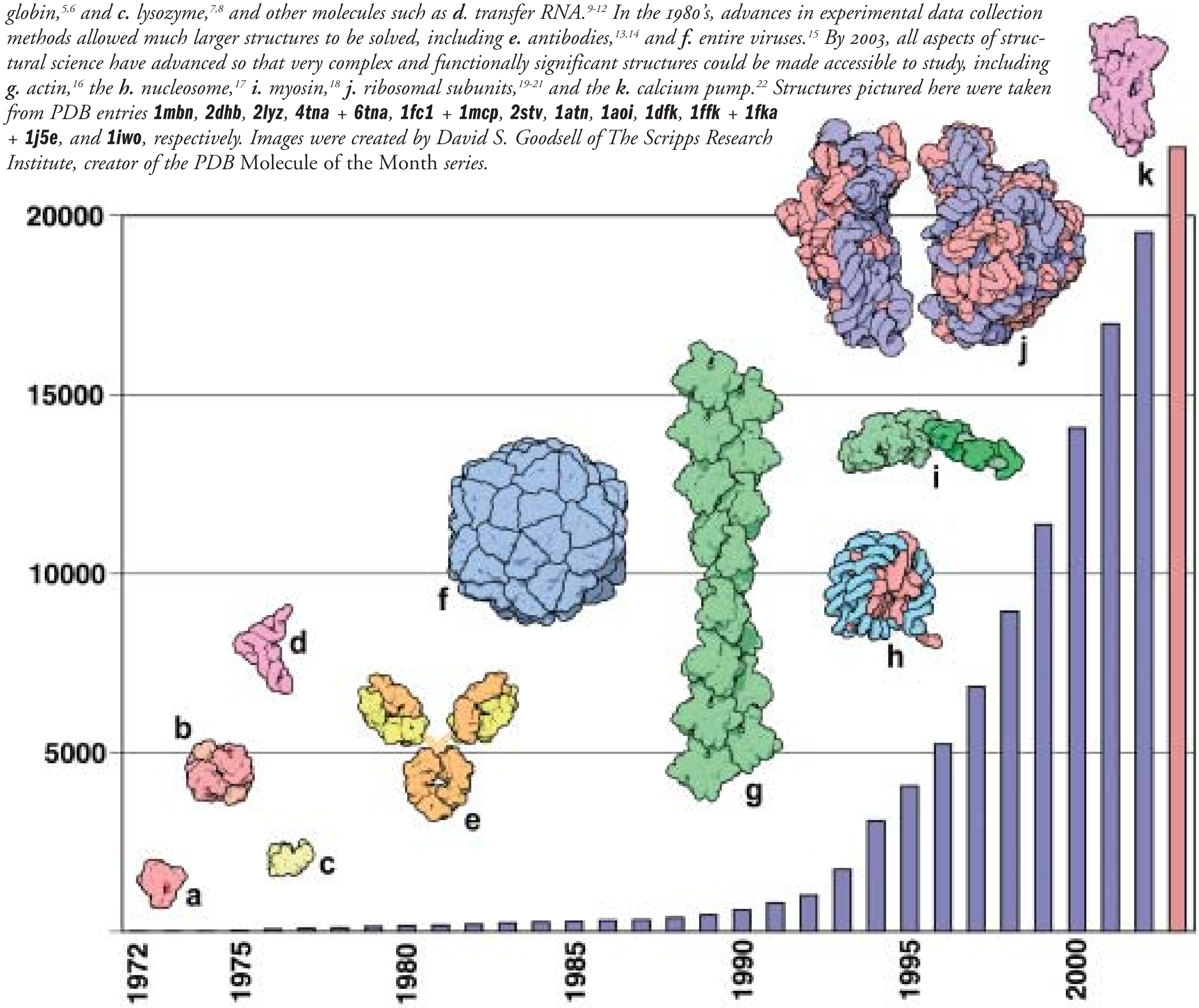
HOW DOES IT WORK?

Data Input: Deposition, Validation, and Annotation

A key component of the PDB is efficient capture (deposition) and curation (validation and annotation) of experimental structural data. Data from experiments using X-ray crystallography, nuclear magnetic resonance (NMR), cryo-electron microscopy (cryo-EM) and other methods are deposited in the PDB. Scientists can contribute their data using tools available at the RCSB-Rutgers (US), the Institute for Protein Research at Osaka University (Japan), and the European Bioinformatics Institute (EBI/UK). Data are also accepted via FTP and e-mail. Data processed at Osaka University and the EBI are forwarded to the RCSB for release.

The main deposition tool, ADIT (AutoDep Input Tool), is available online from sites at RCSB-Rutgers (US) and the Institute for Protein Research (Japan), and is also available 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 (**deposit.pdb.org/mmcif**). mmCIF is an ontology of more than 2,500

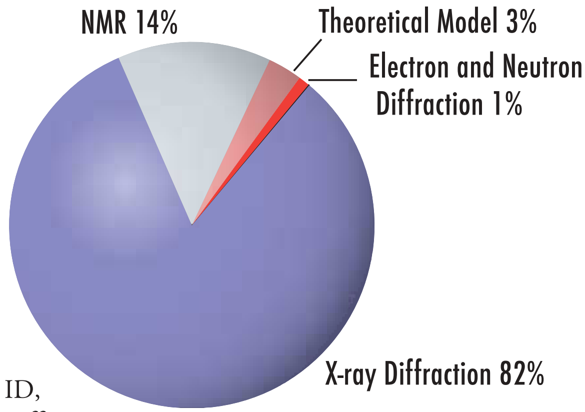
This chart shows the increase in the total number of structures in the PDB per year, through July 1, 2003, as well as examples of the increasing complexity of these structures. In the 1970's, the first structures available to the scientific community included proteins such as ***a.*** *myoglobin,*[3,4] ***b.*** *hemoglobin,*[5,6] *and* ***c.*** *lysozyme,*[7,8] *and other molecules such as* ***d.*** *transfer RNA.*[9-12] *In the 1980's, advances in experimental data collection methods allowed much larger structures to be solved, including* ***e.*** *antibodies,*[13,14] *and* ***f.*** *entire viruses.*[15] *By 2003, all aspects of structural science have advanced so that very complex and functionally significant structures could be made accessible to study, including* ***g.*** *actin,*[16] *the* ***h.*** *nucleosome,*[17] ***i.*** *myosin,*[18] ***j.*** *ribosomal subunits,*[19-21] *and the* ***k.*** *calcium pump.*[22] *Structures pictured here were taken from PDB entries* **1mbn**, **2dhb**, **2lyz**, **4tna** + **6tna**, **1fc1** + **1mcp**, **2stv**, **1atn**, **1aoi**, **1dfk**, **1ffk** + **1fka** + **1j5e**, *and* **1iwo**, *respectively. Images were created by David S. Goodsell of The Scripps Research Institute, creator of the PDB* Molecule of the Month *series.*



terms defining macromolecular structure and related experiments.[23] After checks are performed by PDB staff, 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 PDB. When finalized, the complete entry, including its status information and PDB ID, is loaded into the core relational database. PDB staff complete this entire process with an average turnaround of less than two weeks.

EXPERIMENTAL SOURCE FOR ADIT DEPOSITIONS

July 2002–June 2003

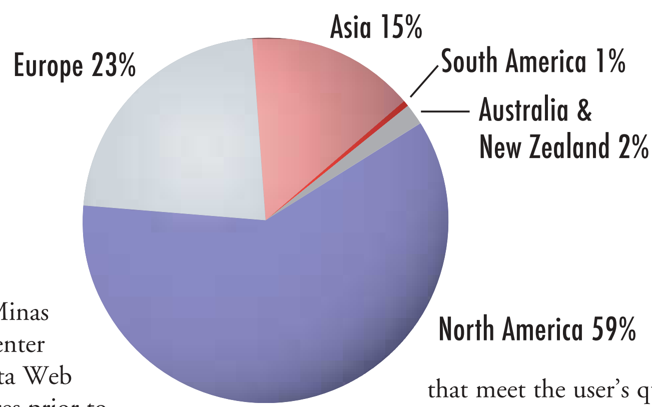


Data Distribution and Access

The PDB is a free service available through the internet. The main PDB Web site at SDSC/UCSD receives an average of more than 160,000 hits per day from all over the world. On average, more than one file is downloaded every second, 24 hours per day, seven days per week. Additionally, there are seven RCSB PDB mirror sites around the world at Rutgers (US), CARB/NIST (US), Osaka University (Japan), the National University of Singapore (Singapore), the Cambridge Crystallographic Data Centre (United Kingdom), the Universidade Federal de Minas Gerais (Brazil), and the Max Delbrück Center for Molecular Medicine (Germany). A beta Web site is available for users to test new features prior to their incorporation into the main Web site and its mirrors. All RCSB sites are maintained 24 hours per day, seven days per week. New structures are added to the PDB holdings by 1:00 a.m. Pacific Time each Wednesday, 52 weeks per year.

ADIT DEPOSITION DEMOGRAPHICS

July 2002–June 2003

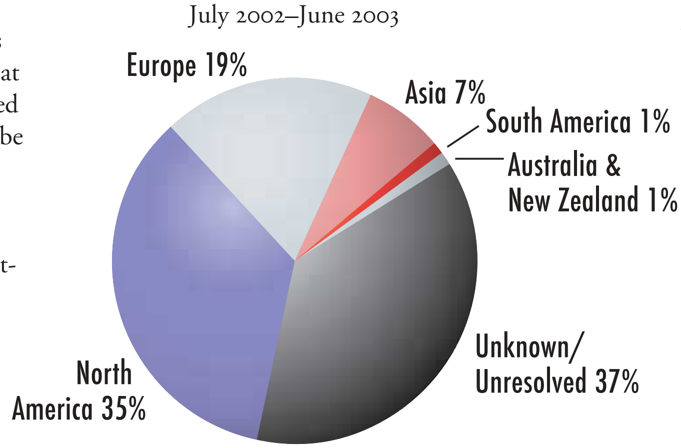


There are a number of different ways to query the database. Entering the PDB ID of the target macromolecule in the search box on the home page performs the simplest search; the ID is assigned at the time of deposition, and is referenced in the paper that is published to describe a structure. This search returns a Structure Explorer page that provides summary information about the entry, 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 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 (unique) and biologically active units are provided as static images generated by MolScript[24] and Raster3D[25], and interactive views are presented in VRML, RasMol[26], MICE[27,28], Chime[29], Swiss-Pdb Viewer[30,31], STING[32], and QuickPDB[33].

Multiple structures can be retrieved using the keyword search functionality on the PDB home page, the SearchLite interface, or the customizable SearchFields interface that searches on parameters selected by the user. The resulting Query Result Browser lists all molecules that meet the user's query specifications, and allows exploration of one or more of the resulting structures. 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 from the PDB Web site. Files may also be downloaded from the PDB FTP server.

ACCESS BY REGION TO www.pdb.org

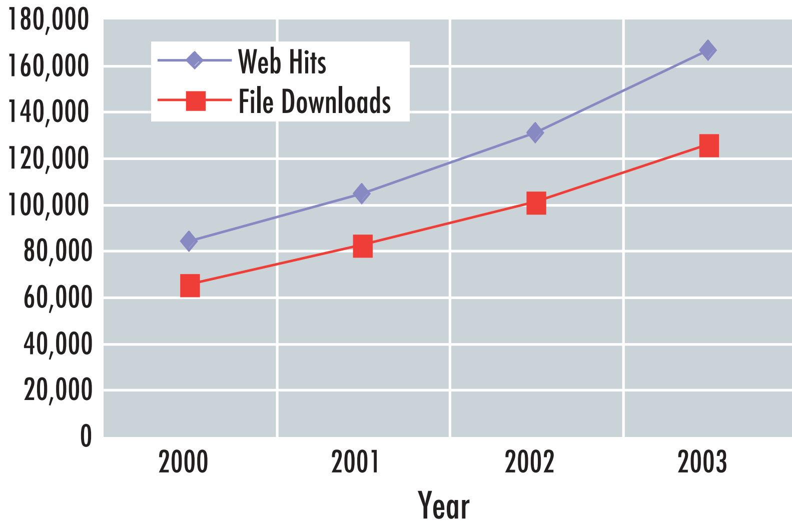
July 2002–June 2003



Direct data access is also supported. The OpenMMS Toolkit, based on the CORBA standard (OMG specification dtc/2001-04-06), offers an applications programming interface along with tools for loading or parsing data. Scripts for mirroring the PDB FTP site, or for downloading files from a specific update, are also available.

DAILY AVERAGE ACCESS STATISTICS FOR www.pdb.org

January 2000-July 2003



Physical Archive

The PDB Physical Archive contains the files and documents associated with the history of each entry in addition to backup copies of current data files. This resource, containing paper, magnetic, and electronic records, is maintained at the CARB/NIST site. A backup copy of the complete query and distribution production system is produced each month by SDSC/UCSD and sent to CARB/NIST for long-term archiving. The overall goal of the PDB Physical Archive 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 and maintenance of the resource. The access and availability of this information to the PDB staff provides a resource for resolving issues concerning specific entries, aiding in uniformity and value-added annotation, facilitating disaster recovery, and making the information available for research.

Outreach and Education

The PDB interacts with its diverse user community to provide information about the resource, to gain feedback for development considerations, and to provide materials that promote a broader understanding of structural biology. This is achieved through accessibility—the PDB maintains an active help desk and has a strong presence at meetings through presentations, workshops, user groups, and exhibit booths. The PDB staff responds to general and specific inquiries sent to the PDB help desks. In addition, the **pdb-l@rcsb.org** listserv offers a forum for exchange among members of the PDB community.

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

Emphasis is placed on educating new users, students, educators, and the general public. To this end, the PDB maintains an extensive portal to educational resources for different levels of expertise. David S. Goodsell's *Molecule of the Month* column, a feature focusing on a different biological molecule each month for a general audience, is highlighted on the PDB home page. This collaboration has also resulted in the widely distributed *Molecular Machinery* poster that depicts 75 PDB structures at a relative scale. PDB staff members attend teacher conferences and educate users at meetings, through our help desks, and through the display of images and exhibits such as the popular *Art of Science* collection. Curricular materials used by educators are solicited and posted on the PDB Web site and highlighted in our newsletters. Tutorials and help documentation are also available on the Web site.

CD-ROM

The PDB distributes a quarterly CD-ROM of its current holdings upon request. 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.

ADVISORY COMMITTEE

The PDB continues to solicit the advice of its PDB Advisory Committee, an international team of experts in X-ray crystallography, NMR, modeling, bioinformatics, and education. The eleven members of this group are:

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

Frank Allen
Executive Director, Cambridge Crystallographic Data Centre

Edward N. Baker
Professor of Structural Biology, University of Auckland

Juli Feigon
Professor of Biochemistry, University of California, Los Angeles

Nobuhiro Go
Professor of Bioinformatics, Japan Atomic Energy Research Institute

Barry Honig
Investigator, Howard Hughes Medical Institute, Department of Biochemistry and Molecular Biophysics, Columbia University

Robert Kaptein
Professor of Chemistry, Utrecht University

Sung-Hou Kim
Professor of Chemistry, University of California, Berkeley

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

David B. Searls
Senior Vice President, Bioinformatics, GlaxoSmithKline Pharmaceuticals

Judith Voet
Professor of Biochemistry, Swarthmore College

PROGRESS & ACHIEVEMENTS

A Collection of Highlights

Structural Genomics

The PDB continues to be actively involved in developing the informatics of structural genomics projects. This effort includes participation in task forces, meetings, and individual interactions with each of the structural genomics centers. The PDB is responsible for maintaining the proposed data deposition specifications and the data dictionary supporting these specifications (**deposit.pdb.org/mmcif**). In addition, the PDB has created software to help integrate data from standard structure determination packages, and established and maintained the Target Registration Database (TargetDB; **targetdb.pdb.org**).

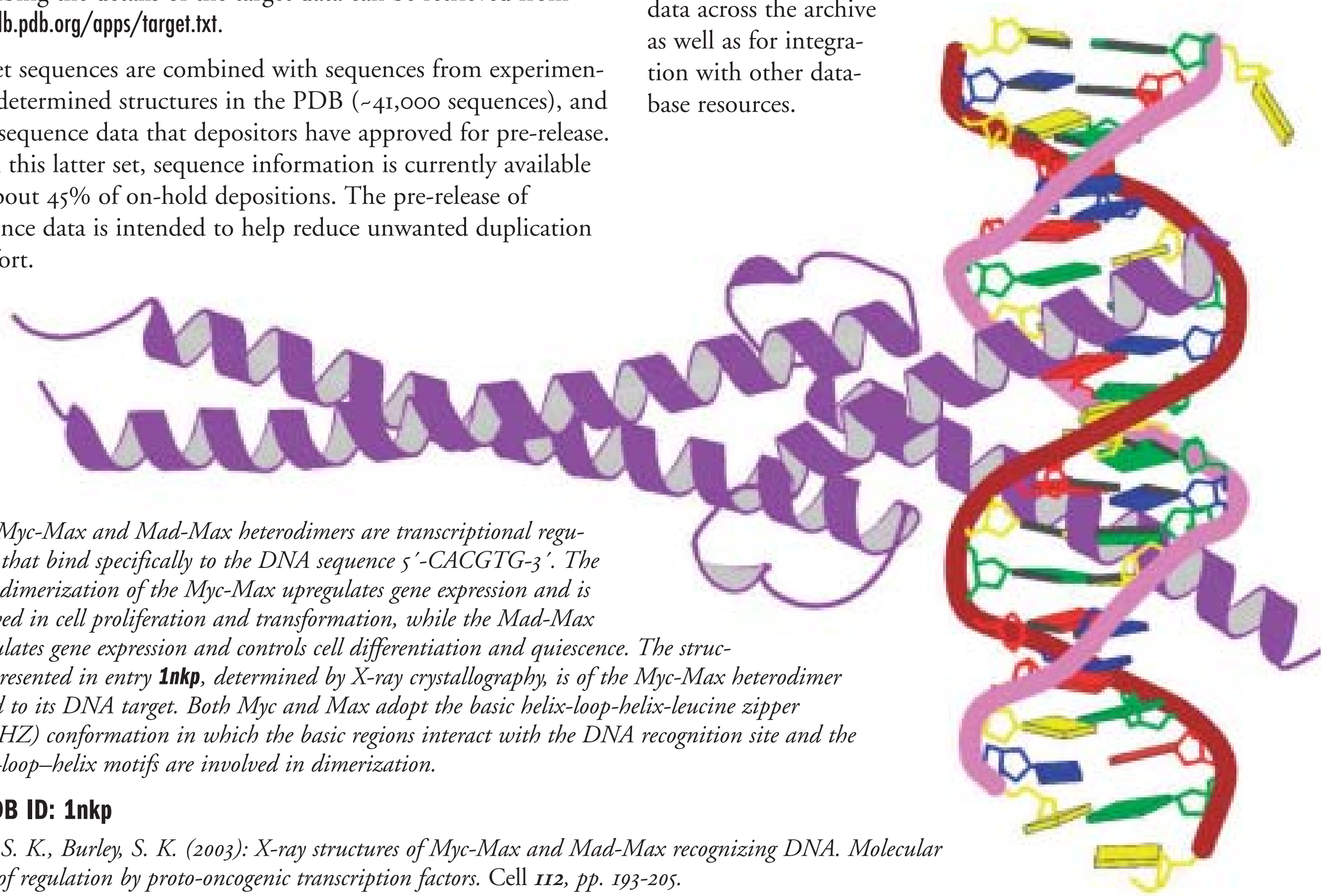
Efficient structure solution on a genomic scale requires a centralized coordination effort, to which the timely availability of status information on the progress of protein production and structure solution is key. TargetDB currently includes over 37,000 target sequences from 18 centers worldwide, including all the NIH-P50 structural genomics centers. Target data are organized following the recommendations of the International Task Force on Target Tracking, which include definitions for the set of states used to track target progress. These states span the details of protein production, structure solution and the ultimate deposition of experimental and structural data. The document type definition file describing the details of the target data can be retrieved from **targetdb.pdb.org/apps/target.txt**.

Target sequences are combined with sequences from experimentally determined structures in the PDB (~41,000 sequences), and with sequence data that depositors have approved for pre-release. From this latter set, sequence information is currently available for about 45% of on-hold depositions. The pre-release of sequence data is intended to help reduce unwanted duplication of effort.

In addition to providing the consolidated target data file, target data from the structural genomics centers are loaded into a relational database. All, or subsets, of these sequence data may be searched using the FASTA sequence comparison method. A simple search form is provided to permit queries of each target data element, including contributing site, protein name, sequence, project tracking identifier, date of last modification, current status of the target, and source organism. Search results may be viewed as HTML reports, FASTA data files, or XML documents. The HTML reports include active links to related information at the contributing project site and in other databases such as the PDB, GenBank[34], and SWISS-PROT[35]. TargetDB also tracks target status temporally, so it is possible to create reports on the evolution of target progress over a selected time interval. TargetDB is updated weekly.

Data Uniformity: Release of a Standardized Archive

The PDB's focus on making the archive as consistent and error-free as possible is ongoing. 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.

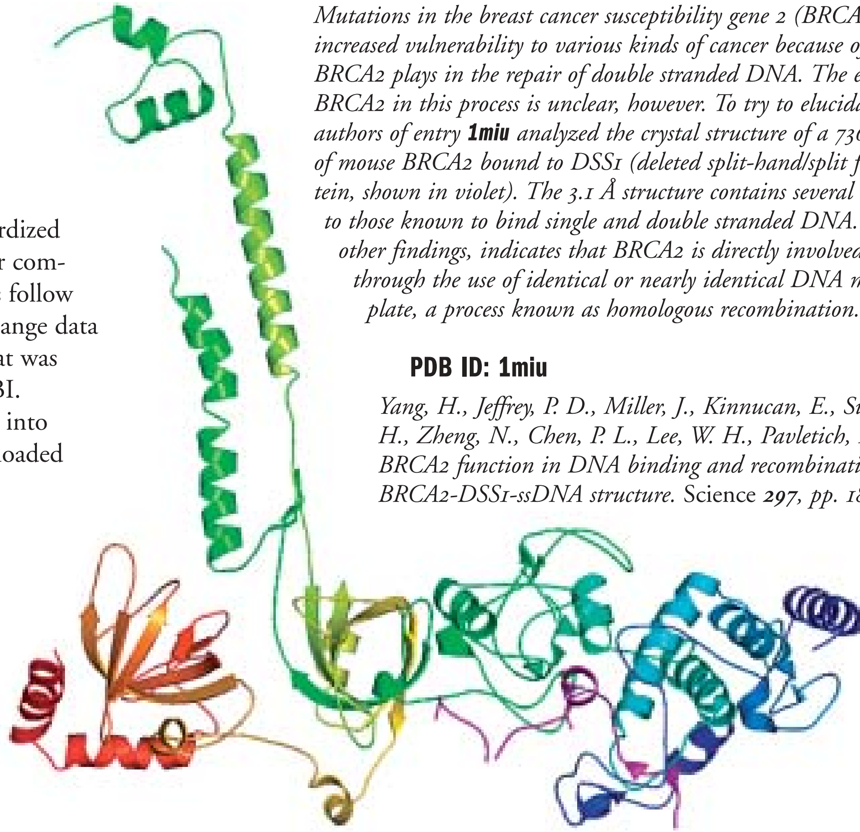


Both Myc-Max and Mad-Max heterodimers are transcriptional regulators that bind specifically to the DNA sequence 5′-CACGTG-3′. The heterodimerization of the Myc-Max upregulates gene expression and is involved in cell proliferation and transformation, while the Mad-Max deregulates gene expression and controls cell differentiation and quiescence. The structure presented in entry ***1nkp****, determined by X-ray crystallography, is of the Myc-Max heterodimer bound to its DNA target. Both Myc and Max adopt the basic helix-loop-helix-leucine zipper (bHLHZ) conformation in which the basic regions interact with the DNA recognition site and the helix–loop–helix motifs are involved in dimerization.*

PDB ID: 1nkp

Nair, S. K., Burley, S. K. (2003): X-ray structures of Myc-Max and Mad-Max recognizing DNA. Molecular bases of regulation by proto-oncogenic transcription factors. Cell *112, pp. 193-205.*

The PDB archive has been standardized and released in mmCIF format for community review. These curated files follow the latest version of the PDB exchange data dictionary (deposit.pdb.org/mmcif) that was developed by the PDB and the EBI. Software for translating these data into PDB-formatted files can be downloaded from deposit.pdb.org/software.



Mutations in the breast cancer susceptibility gene 2 (BRCA2) result in increased vulnerability to various kinds of cancer because of the role that BRCA2 plays in the repair of double stranded DNA. The exact role of BRCA2 in this process is unclear, however. To try to elucidate this role, the authors of entry **1miu** *analyzed the crystal structure of a 736-residue domain of mouse BRCA2 bound to DSS1 (deleted split-hand/split foot syndrome protein, shown in violet). The 3.1 Å structure contains several domains similar to those known to bind single and double stranded DNA. This, along with other findings, indicates that BRCA2 is directly involved in DNA repair through the use of identical or nearly identical DNA molecules as a template, a process known as homologous recombination.*

PDB ID: 1miu

Yang, H., Jeffrey, P. D., Miller, J., Kinnucan, E., Sun, Y., Thomä, N. H., Zheng, N., Chen, P. L., Lee, W. H., Pavletich, N. P. (2002): BRCA2 function in DNA binding and recombination from a BRCA2-DSS1-ssDNA structure. Science ***297****, pp. 1837-48.*

Release of PDB Data in XML Format

All released PDB entries are now available in XML format from the PDB beta FTP site at ftp://beta.rcsb.org/pub/pdb/uniformity/data/XML. The 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 XML data files directly correspond to the item names defined in the PDB Exchange Dictionary. An XML schema documenting the translated format is available at deposit.pdb.org/mmcif. The delivery of PDB data in XML format is the product of a collaboration between the Protein Data Bank Japan (PDBj), the Macromolecular Structure Database (MSD) group at European Bioinformatics Institute (EBI), and the RCSB.

Deposition and Processing

During the period covered by this report, 4,213 files were deposited to the PDB through an international effort. These data were most often processed and returned to the depositor in less than two weeks. Of the structures deposited during the period of this report, approximately 67% were deposited with experimental data. Around 50% of the depositions released their sequence ahead of the structure's release.

The programs used to prepare, validate, and deposit these data are available for download and desktop installation from deposit.pdb.org/software.

These programs are updated regularly and are available for a variety of platforms.

DOWNLOADABLE DATA DEPOSITION AND PROCESSING SOFTWARE

(deposit.pdb.org/software)

- PDB_EXTRACT—Extracts mmCIF data from structure determination applications
- ADIT—Prepares files that can be e-mailed to the PDB for deposition
- PDB Validation Suite—Processes and checks structure data
- CIFTr—Translates files in mmCIF format into files in PDB format
- mmCIF Loader—Loads mmCIF data into relational databases and XML
- MAXIT—Processes and curates macromolecular structure data
- CIFOBJ—A class library of mmCIF dictionary access tools
- OpenMMS Toolkit—A suite of Java source code that includes an mmCIF parser, RDBMS loader, XML translator, and CORBA server

Enhancements to Query, Reporting, and Access

The capabilities of the PDB resource continue to be extended. During the past year, new content and features were added in response to user requests. The developments described here have undergone significant testing on the beta Web site prior to their addition to the primary PDB Web site and its mirrors.

The Summary Information section of the Structure Explorer page for any PDB entry can be used to search the PDB for all structures that match any author listed in the primary citation, that include a particular ligand listed, or that have the same EC number.

Links to view or download the curated mmCIF files are available from the Structure Explorer and the Query Result Browser page.

The availability of experimental data for a particu-

IMPORTANT WEB ADDRESSES

Beta Test Site
beta.rcsb.org

Data Uniformity
www.rcsb.org/pdb/uniformity

Deposition FAQ
deposit.pdb.org

Educational Resources
www.rcsb.org/pdb/education.html

FTP Site
ftp://ftp.rcsb.org

mmCIF Resources
deposit.pdb.org/mmcif

Molecule of the Month
www.rcsb.org/pdb/molecules/molecule_list.html

New PDB Features
www.rcsb.org/pdb/new_features.html

OpenMMS
openmms.sdsc.edu

PDB Home Page
www.pdb.org

PDB Newsletter
www.rcsb.org/pdb/newsletter.html

Software Resources
www.rcsb.org/pdb/software-list.html

Structural Genomics
www.rcsb.org/pdb/strucgen.html

XML Beta Data Files
ftp://beta.rcsb.org/pub/pdb/uniformity/data/XML

HELP DESKS

General PDB Inquiries
info@rcsb.org

General Deposition Questions
deposit@rcsb.rutgers.edu

ADIT Information
help@rcsb.rutgers.edu

lar structure (structure factor or constraint file) can be included in reports generated for a set of structures resulting from a PDB search. This feature is included in the Structure Summary report and is offered as an option for the customizable reports. Tabular reports of results can be saved in HTML format or plain text format; the latter can also be imported into a spreadsheet program such as Excel.

The results of the weekly clustering of protein chains in the PDB are included in the FTP archive at **ftp://ftp.rcsb.org/pub/pdb/derived_data/NR**. These clusters are used in the "remove sequence homologs" feature on the PDB Web sites. Files that list the clusters and their rankings at 50%, 70%, and 90% sequence identity are available. Smaller rank numbers indicate higher (better) ranking; chains with rank number 1 are considered to be the best representative of their cluster.

The View Structure section of the Structure Explorer pages offers static ribbon images of the assumed biological unit(s) for structures, where relevant, in addition to the static images of the asymmetric unit. Links to the coordinate files that are used to generate the biological unit images are also accessible here, as well as from the Download/Display File section of the Structure Explorer.

BioEditor[36], a program for creating and viewing structure presentations, is accessible from the PDB Web site. BioEditor (**bioeditor.sdsc.edu**) is a tool to bridge the gap between printed literature and current Web-based presentation formats for macromolecular structures. This standalone Windows application can be used to prepare and present structure annotations containing formatted text, graphics, sequence data, and interactive molecular views—all in a single document or set of documents. BioEditor facilitates the communication of structural data to a diverse audience by allowing users to create and view dynamic content in a uniform format that can be widely distributed through the internet.

Swiss-Pdb Viewer is linked from the View Structure section of Structure Explorer pages. The Structure Explorer page for any NMR-solved entry links to that structure's entry in the BioMagResBank (BMRB).

Several new features are available for testing on the beta Web site. Lucene[37], a new keyword search engine that uses a ranking system and indices of the curated mmCIF files, facilitates searches on the home page, SearchLite, and the Text Search field of the SearchFields interface. A simple search interface for theoretical model structures is in development and testing on the PDB beta Web site at **beta.rcsb.org/pdb/cgi/models.cgi**.

Mirroring

During this period, a new mirror site was added at Max-Delbrück-Center for Molecular Medicine in Berlin, Germany.

Tools are available to aid the mirroring process for PDB users. An rsync script assists users in setting up their own local mirrors of the PDB FTP site. A simple Perl program downloads all files from the PDB associated with a particular update date, including coordinate files, structure factors, and NMR restraints. These scripts are available from **ftp://ftp.rcsb.org/pub/pdb/software**.

CD-ROM Distribution

Copies of the coordinate and experimental data files are produced quarterly and released on CD-ROM. With the October 2002 release, theoretical models are placed in a *models* directory separate from the experimental coordinate data. NMR constraints and X-ray structure factors are available as separate products. A full release of the PDB currently requires five disks for coordinate files, and six disks for experimental data files. As of 2003, the entire PDB archive is distributed once in January; incremental updates are provided each quarter. This greatly reduces the number of disks needed to have a recent copy of the PDB archive.

Developments in NMR

The PDB is working to improve structure deposition and annotation services for data acquired from NMR experiments. We are collaborating with the BMRB (a member of the RCSB) to provide a single integrated deposition system for NMR data that will accept both experimental and structure data. A beta version of an ADIT server for the data items collected by the BMRB is at **tuna.bmrb.wisc.edu:8000/bmrb-adit**. It uses an mmCIF-like dictionary (NMRIF) that was derived from the NMR STAR deposition form used by the BMRB. The ADIT system has been extended to display deposition information that closely models the current BMRB NMR STAR data presentation, and is now being used by BMRB to collect NMR experimental data.

The BioMagResBank team: (top row, left to right) Jurgen Doreleijers, David Tolmie, Steve Mading, Dimitri Maziuk, John Markley, Dan Sinclair, Eldon Ulrich, (bottom row, left to right) Jundong Lin, R. Kent Wenger, and Hongyang Yao.

Outreach and Education

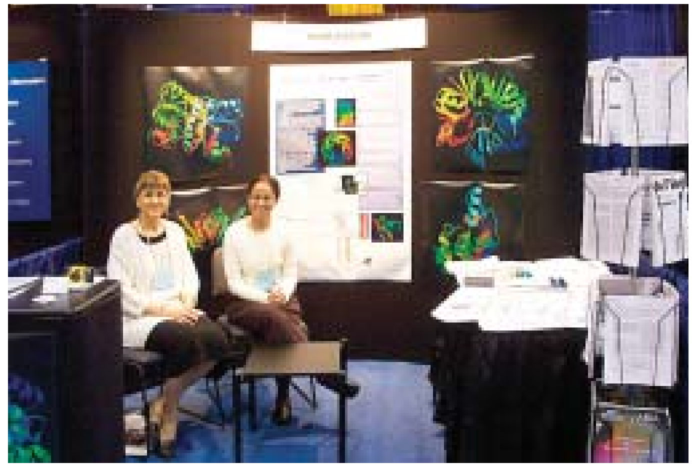
The PDB participated in exhibitions at the Intelligent Systems for Molecular Biology (ISMB) conference in Edmonton, Canada; the Protein Society's Annual Symposium in San Diego, CA; the XIX Congress and General Assembly of the International Union of Crystallography (IUCr) in Geneva, Switzerland; and the Biophysical Society Meeting in San Antonio, TX. PDB staff members presented talks, demonstrations, and posters at more than thirty meetings around the world, including the International Structural Genomics Organisation (ISGO) International Conference on Structural Genomics in Berlin, Germany. User group meetings were held at the IUCr meeting and locally at RCSB sites.

The quarterly PDB Newsletter has been redesigned so that it is printed in full color, and includes new features such as interviews with community members and an "Education Corner" where instructors describe ways that they use PDB in the classroom.

Images from the PDB's Art of Science *exhibit were on display at Purdue University from April 11 through May 17, 2003. The PDB would like to see the* Art of Science *travel to other places. For information on sponsoring this exhibit at your institution, contact* **info@rcsb.org**.

The PDB's *Art of Science* travelling art exhibit highlights various representations of proteins found in the PDB, including large-scale depictions of the images available from PDB Structure Explorer pages and pictures from the PDB's *Molecule of the Month* series. It was part of the Centre for Cellular and Molecular Biology's (CCMB) Silver Jubilee Celebrations and Symposium on "The Current Excitement in Biology" in Hyderabad, India (November 24-29, 2002) and was on display at the CCMB campus's main building from November 15, 2002, through February 28, 2003. Additionally, the PDB exhibit, along with molecular images from Purdue University's Structural Biology Center, were featured in "Watson's Crick," a commons area in the Department of Biological Sciences where local artists present their work. The exhibit was open from April 11, 2003, until May 17, 2003.

We announced the initiation of the PDB Poster Prize, which will recognize student poster presentations involving macromolecular crystallography. The prize will be awarded to the best poster by an undergraduate or graduate student at each of the meetings of the IUCr Regional Associates—the American Crystallographic Association (ACA), the Asian Crystallographic Association (AsCA), and the European Crystallographic Association (ECA)—as well as at the IUCr Congress itself. Each award will consist of two educational books; this year's prize will be signed copies of *Biochemistry*—Vol. I, by Donald and Judith G. Voet, and *Introduction to Macromolecular Crystallography,* by Alexander McPherson. Winners will be announced on the PDB Web site and in the PDB, ACA, and IUCr newsletters.



RCSB PDB members Judith L. Flippen-Anderson (Production and Outreach Leader) and Shuchismita Dutta at the PDB booth at the Biophysical Society Annual Meeting.

The PDB met with science reporter Patrick Regan for a segment that appeared on the New Jersey Network (NJN) News in December 2002, and was highlighted in the article "Banking on Structures" that appeared in *BioIT World* (October 2002).

Collaborations with Other Organizations

The collaboration on NMR data deposition continues with the BioMagResBank (BMRB), stressing the development of a data dictionary and an integrated deposition system based on ADIT.

The PDB continues to work with the Cambridge Crystallographic Data Centre (CCDC) on methods for ligand validation, and mirrors ReLiBase+, a CCDC ligand resource.

STING Millennium, a Web-based suite of programs that provides simultaneous analysis and display of structure and sequence, is being mirrored through collaboration with the Brazilian Agricultural Research Corporation (Embrapa).

We are working with the deCODE Biostructures Group (Emerald Biosystems) on crystallization data representation and exchange, and on special handling of structures with licensing restrictions.

The European Bioinformatics Institute (MSD-EBI) continues to provide weekly updates of the structures deposited and processed at their AutoDep site. A key accomplishment of the EBI collaboration is the agreement on an exchange dictionary that is currently in the final stages of testing. This permits all data associated with each deposition to be exchanged. The release of PDB entries in XML format is the product of a collaboration with the MSD-EBI, PDBj, and the RCSB.

PDB staff members have worked with Alexander Wlodawer (National Cancer Institute) and Jiri Vondresek (Institute of Organic Chemistry and Biochemistry, Prague) to move the HIV Protease Database from NCI to NIST, incorporating uniformity-compliant PDB file data. The database is available at srdata.nist.gov/hivdb.

RCSB PDB members Philip E. Bourne (Co-Director), David Padilla, and Gregory Vasquez respond to questions from PDB user Michael Braxenthaller at PDB's exhibit at the ISMB 2002 Conference.

The PDB is currently working with IBM to evaluate prototype technology for fault-tolerant storage and high-performance database technology and the use of Websphere and DB2 as part of the re-engineered PDB.

Close collaborations have been maintained with the Institute for Protein Research at Osaka University (PDBj), where ADIT is used for the deposition and processing of structures. The RCSB, MSD-EBI, and PDBj are also collaborating on an XML representation of PDB data based on the PDB exchange dictionary.

Colleagues at the National Center for Biotechnology Information (NCBI) at the NIH continue to work with the PDB on ways to ensure that PDB files can be used by NCBI-developed databases.

David S. Goodsell of The Scripps Research Institute continues to contribute the *Molecule of the Month* feature on the PDB site, which provides introductory level descriptions of the structure and function of key molecules found in the PDB. This collaboration has also resulted in the popular *Molecular Machinery* poster that continues to be distributed at conferences and classrooms, and in response to user requests.

Other collaborators include Paul Adams (Lawrence Berkeley National Laboratory), Wladek Minor (University of Virginia) and Zbyszek Otwinowski (University of Texas) on developing software for structural genomics, Alexei Adzhubei (University of Oslo) on a models and mmCIF database project, as well as Peter Karp (Metacyc), Ernest Laue (CCPN), Eric Martz (University of Massachusetts), Cherri Pancake (Oregon State University), Dietmar Schomburg (BRENDA), Wolf-D. Ihlenfeldt (CACTVS), Anthony Williams (ACDLABS), and Jane Richardson (Duke University).

THE FUTURE OF THE PDB

The PDB has already exceeded the goals set at the outset of this project, and we plan to continue to improve our services by adding new functionality and by assessing and anticipating the needs of our ever-increasing user base.

Data Deposition and Processing

The PDB will continue to streamline deposition and data processing procedures. We will continue to strengthen the data processing software that has been released both as source code and as executables.

Structural Genomics

Community collaborations to develop the required data items for structural genomics will continue. The PDB will also work with the software developers and beamline operators to integrate the output of applications programs with PDB deposition software. Efforts to produce tools to facilitate the extraction and integration of the data from existing structure determination software will be expanded. The PDB will further encourage the structural genomics centers to use PDB software tools in their respective data processing operations.

NMR

The PDB will continue to maintain an active dialog with the NMR community through contacts with individual depositors and structural genomics centers, and as an active participant in the CCPN software initiative. The development of a common interface for the coordinates, constraints and chemical shift data with the BMRB is currently in alpha testing (tuna.bmrb.wisc.edu:8000/bmrb-adit). We anticipate a wider beta test and possible deployment within the next project year.

DATA DISTRIBUTION AND QUERY

The PDB has mounted a significant effort to redesign and re-engineer our distribution and query system. This new system utilizes data improvements resulting from our ongoing efforts in uniformity and standardization. It will offer improved accessibility for the 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 re-engineered system will provide new functionality of value to all users of the PDB. Examples include browsing by (i) relatedness to disease, (ii) molecular function, (iii) biochemical process, and (iv) cellular location, as well as a better description of ligands bound to macromolecules. Published Medline[38] abstracts will be available for searching directly from each entry's Structure Explorer page. Of major significance is the introduction of Web services that will permit a remote application program or database to easily access PDB data at various levels of detail from atoms to complete structures.

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 high quality, readily available tools to improve existing components of the site, such as the keyword search engine—now facilitated by Lucene, and the help documentation—converted to one dynamic, content-specific system supported by Robohelp[39].

We have loaded all of the data from our uniformity processing efforts and the calculated derived geometric structural features into a relational database engine that provides a single source of data for the re-engineered 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.pdb.org), which is being reviewed by community members and experts. This effort demonstrates the viability of using the PDB exchange dictionary as the conceptual model for the production PDB database.

The re-engineering development process includes an internal requirements analysis and active testing. The strengths, weaknesses, and major usage patterns of our current system have been examined and documented, and have provided a good indication of future need. An iterative process of loading, testing, and correcting new data will be used to further refine the system as these data become available.

The PDB has worked with both users and usability experts to improve navigation and ease of access to the new PDB site. Experts with backgrounds in molecular and structural biology are analyzing the system for errors at each new phase of development. Focus group sessions to test navigation and query features of the new system are being conducted with representatives from different sects of the community, including researchers, educators, students, and the general public. Feedback from these efforts, as well as from other interactions with users, is documented, reviewed, and incorporated into our developments on an ongoing basis. The system is currently in an alpha testing phase, with a beta release expected during the first part of 2004.

RECENT RCSB PDB PUBLICATIONS

Data Uniformity

Efforts on the PDB Data Uniformity Project will continue. The PDB intends to ensure that legacy and current files are annotated consistently. Some of these activities will help define and refine the controlled vocabulary used in annotation.

Outreach and Education

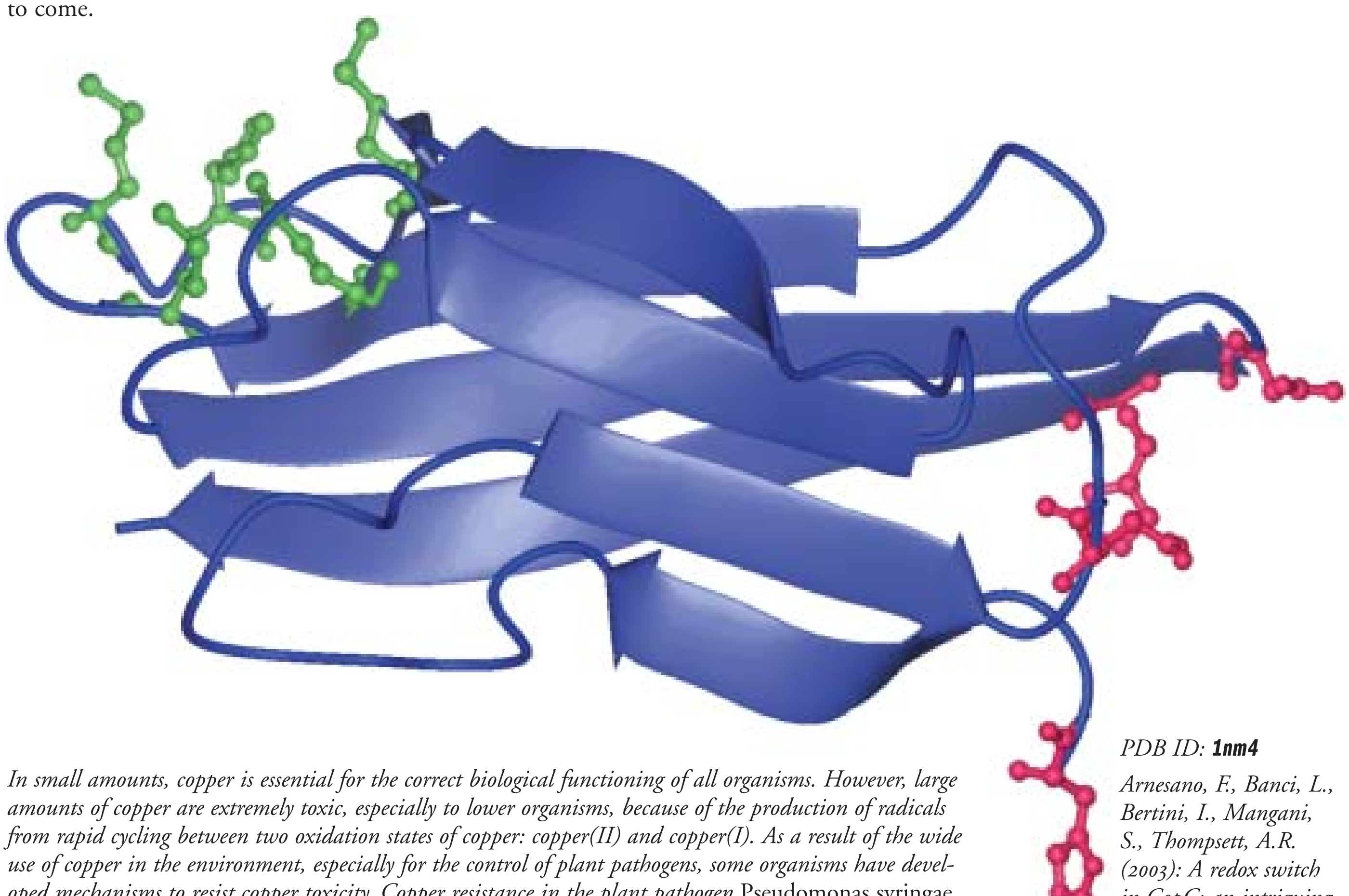
The PDB will continue to solicit the input of its diverse user community of researchers, teachers, and students through interactions at future meetings and through help desk services. Workshops to educate teachers and students will be held, and publications and online information resources will be further developed. New ways to educate and convey the PDB's mission to new audiences will be explored. We will remain engaged with users to ensure that the PDB continues to serve as a prominent resource to further scientific advances worldwide for many years to come.

Greer, D. S., Westbrook, J. D., Bourne, P. E. (2002): An ontology driven architecture for derived representations of macromolecular structure. *Bioinformatics,* 18(9), pp. 1280-1.

Westbrook, J., Feng, Z., Chen, L., Yang, H., and Berman, H. M. (2003): The Protein Data Bank and structural genomics. *Nucleic Acids Res.* **31**, pp. 489-91.

The PDB Team (2003): The Protein Data Bank. *Structural Bioinformatics*. P. E. Bourne and H. Weissig. Hoboken, NJ, John Wiley & Sons, Inc. pp. 181-98.

Westbrook, J., and Fitzgerald, P. M. (2003): The PDB format, mmCIF formats and other data formats. *Structural Bioinformatics.* P. E. Bourne and H. Weissig. Hoboken, NJ, John Wiley & Sons, Inc. pp. 161-79.



In small amounts, copper is essential for the correct biological functioning of all organisms. However, large amounts of copper are extremely toxic, especially to lower organisms, because of the production of radicals from rapid cycling between two oxidation states of copper: copper(II) and copper(I). As a result of the wide use of copper in the environment, especially for the control of plant pathogens, some organisms have developed mechanisms to resist copper toxicity. Copper resistance in the plant pathogen Pseudomonas syringae *arises from four proteins, CopABCD, that sequester excess copper in the cell's periplasm and outer membrane. One of these proteins, CopC (shown here from the unbound NMR solution structure in entry* **1nm4***), contains two copper-binding sites. Site A (shown in red) selectively binds copper(II) using residues His-1, Glu-27, Asp-89, and His-91. Site B (shown in green), located about 30Å away, selectively binds copper(I) using His-48, Met-40, Met-43, Met-46, and Met-51. Oxidation or reduction causes copper to be shuttled from one site to the other indicating a mechanism for copper trafficking that is activated by a redox switch.*

PDB ID: **1nm4**

Arnesano, F., Banci, L., Bertini, I., Mangani, S., Thompsett, A.R. (2003): A redox switch in CopC: an intriguing copper trafficking protein that binds copper(I) and copper(II) at different sites. Proc. Nat. Acad. Sci. USA ***100****, pp. 3814-9.*

SELECTED REFERENCES

1. Berman, H. M., Westbrook, J., Feng, Z., Gilliland, G., Bhat, T. N., Weissig, H., Shindyalov, I. N. & Bourne, P. E. (2000): The Protein Data Bank. *Nucleic Acids Res.* **28**, pp. 235-42.
2. Bernstein, F. C., Koetzle, T. F., Williams, G. J., Meyer, E. E., Brice, M. D., Rodgers, J. R., Kennard, O., Shimanouchi, T. & Tasumi, M. (1977): Protein Data Bank: a computer-based archival file for macromolecular structures. *J. Mol. Biol.* **112**, pp. 535-42.
3. Watson, H. C. (1969): The stereochemistry of the protein myoglobin. *Prog. Stereochem.* **4**, p. 299-333.
4. Kendrew, J. C., Bodo, G., Dintzis, H. M., Parrish, R. G., Wyckoff, H. (1958): A three-dimensional model of the myoglobin molecule obtained by X-ray analysis. *Nature* **181**, pp. 662-6.
5. Bolton, W., Perutz, M. F. (1970): Three dimensional Fourier synthesis of horse deoxy-haemoglobin at 2.8 Å units resolution. *Nature* **228**, pp. 551-2.
6. Perutz, M. F., Rossmann, M. G., Cullis, A. F., Muirhead, G., Will, G. (1960): Structure of haemoglobin: a three-dimensional Fourier synthesis at 5.5 Å resolution. *Nature* **185**, pp. 416-22.
7. Blake, C. C. F., Koenig, D. F., Mair, G. A., North, A. C. T., Phillips, D. C., Sarma, V. R. (1965): Structure of hen egg-white lysozyme. A three dimensional Fourier synthesis at 2 Å resolution. *Nature* **206**, pp. 757-61.
8. Blake, C. C. F., Johnson, L. N., Mair, G. A., North, A. C. T., Phillips, D. C., Sarma, V. R. (1967): Crystallographic studies of the activity of hen egg-white lysozyme. *Proc. R. Soc. London Ser. B* **167**, pp. 378-88.
9. Hingerty, B. E., Brown, R. S., Jack, A. (1978): Further refinement of the structure of yeast tRNAPhe. *J. Mol. Biol.* **124**, pp. 523-4.
10. Sussman, J. L., Holbrook, S. R., Warrant, R. W., Church, G. M., Kim, S.-H. (1978): Crystal structure of yeast phenylalanine transfer RNA. I. Crystallographic refinement. *J. Mol. Biol.* **123**, pp. 607-30.
11. Suddath, F., Quigley, G., McPherson, A., Sneden, D., Kim, J., Kim, S., Rich, A. (1974): Three-dimensional structure of yeast phenylalanine transfer RNA at 3.0 Ångstroms resolution. *Nature* **248**, pp. 20-4.
12. Robertus, J. D., Ladner, J. E., Finch, J. T., Rhodes, D., Brown, R. S., Clark, B. F. C., Klug, A. (1974): Structure of yeast phenylalanine tRNA at 3 Å resolution. *Nature* **250**, pp. 546-51.
13. Deisenhofer, J. (1981): Crystallographic refinement and atomic models of a human Fc fragment and its complex with fragment B of protein A from *Staphylococcus aureus* at 2.9 and 2.8 Å resolution. *Biochemistry* **20**, pp. 2361-70.
14. Satow, Y., Cohen, G. H., Padlan, E. A., Davies, D. R. (1986): Phosphocholine binding immunoglobulin Fab McPC603. An X-ray diffraction study at 2.7 Å. *J. Mol. Biol.* **190**, pp. 593-604.
15. Jones, T. A., Liljas, L. (1984): Structure of satellite tobacco necrosis virus after crystallographic refinement at 2.5 Å resolution. *J. Mol. Biol.* **177**, pp. 735-67.
16. Kabsch, W., Mannherz, H. G., Suck, D., Pai, E. F., Holmes, K. C. (1990): Atomic structure of the actin:DNase I complex. *Nature* **347**, pp. 37-44.
17. Luger, K., Mader, A. W., Richmond, R. K., Sargent, D. F., Richmond, T. J. (1997): Crystal structure of the nucleosome core particle at 2.8Å resolution. *Nature* **389**, pp. 251-60.
18. Houdusse, A., Szent-Gyorgyi, A. G., Cohen, C. (2000): Three conformational states of scallop myosin S1. *Proc. Nat. Acad. Sci. USA* **97**, pp. 11238-43.
19. Ban, N., Nissen, P., Hansen, J., Moore, P. B., Steitz, T. A. (2000): The complete atomic structure of the large ribosomal subunit at a 2.4 Å resolution. *Science* **289**, pp. 905-20.
20. Schluenzen, F., Tocilj, A., Zarivach, R., Harms, J., Gluehmann, M., Janell, D., Bashan, A., Bartels, H., Agmon, I., Franceschi, F., Yonath, A. (2000): Structure of functionally activated small ribosomal subunit at 3.3 Å resolution. *Cell* **102**, pp. 615-23.
21. Wimberly, B. T., Brodersen, D. E., Clemons Jr., W. M., Morgan-Warren, R., Carter, A. P., Vonrhein, C., Hartsch, T., Ramakrishnan, V. (2000): Structure of the 30S ribosomal subunit. *Nature* **407**, pp. 327-39.
22. Toyoshima, C., Nomura, H. (2002): Structural changes in the calcium pump accompanying the dissociation of calcium. *Nature* **418**, pp. 605-11.
23. Bourne, P. E., Berman, H. M., Watenpaugh, K., Westbrook, J. D. & Fitzgerald, P. M. D. (1997): The macromolecular Crystallographic Information File (mmCIF). *Meth. Enzymol.* **277**, pp. 571-90.
24. Kraulis, P. (1991): MOLSCRIPT: a program to produce both detailed and schematic plots of protein structures. *J. Appl. Cryst.* **24**, pp. 946-50.
25. Merritt, E. A. & Bacon, D. J. (1997): Raster3D: Photorealistic molecular graphics. *Meth. Enzymol.* **277**, pp. 505-24.
26. Sayle, R. & Milner-White, E. J. (1995): RasMol: biomolecular graphics for all. *Trends Biochem. Sci.* **20**, p. 374.
27. Bourne, P. E., Gribskov, M., Johnson, G., Moreland, J. & Weissig, H. (1998): *Pacific Symposium on Biocomputing.*
28. Tate, J. G., Moreland, J. & Bourne, P. E. (1999): MSG (Molecular Scene Generator): a Web-based application for the visualization of macromolecular structures. *J. Appl. Cryst.* **32**, pp. 1027-8.
29. MDL Information Systems, Inc. Chime. United States.
30. Guex, N. & Peitsch, M. C. (1997): SWISS-MODEL and the Swiss-PdbViewer: An environment for comparative protein modeling. *Electrophoresis* **18**, pp. 2714-23.
31. Kaplan, W. & Littlejohn, T. (2001): Swiss-PDB Viewer (Deep View). *Brief Bioinform 2*, pp. 195-7.
32. Neshich, G., Togawa, R., Vilella, W. & Honig, B. (1998): STING (Sequence To and withIN Graphics) PDB_Viewer. *Protein Data Bank Quarterly Newsletter* **85**, pp. 6-7.
33. Shindyalov, I. & Bourne, P. E. Applet QuickPDB. Copyright © 1996–1998 San Diego Supercomputer Center.
34. Benson, D. A., Karsch-Mizrachi, I., Lipman, D. J., Ostell, J., Rapp, B. A. & Wheeler, D. L. (2000): GenBank. *Nucleic Acids Res.* **28**, pp. 15-8.
35. Bairoch, A. & Apweiler, R. (2000): The SWISS-PROT protein sequence database and its supplement TrEMBL in 2000. *Nucleic Acids Res.* **28**(1), pp. 45-8.
36. Yang, P., Craig, P. A., Goodsell, D. S., Bourne, P. E. (2003): BioEditor—simplifying macromolecular structure annotation. *Bioinformatics* **19**(7), pp. 897-8.
37. Apache Software Foundation. Lucene. Copyright © 1999–2003.
38. 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.
39. eHelp Corporation. Robohelp. Copyright © 1992–2003, All Rights Reserved.

In the nucleus, an RNA-protein assembly called the splicesosome excises non-coding regions of precursor mRNA and splices together the remaining exons to form mature mRNA. The splicesosome is composed of five uridine-rich small nuclear ribonucleoproteins (U snRNPs) containing small nuclear RNAs (snRNAs) and at least 80 associated proteins. Each U snRNP complex contains a 110-180 nucleotide snRNA, snRNP-specific proteins, and seven Sm or Sm-like proteins that are common to each snRNP. The seven Sm proteins contain a common sequence motif and assemble with snRNA to form a "doughnut-shaped" heteroheptameric core. In addition to the Sm and Sm-like proteins found in eukaryotes, a family of Sm-like proteins has been identified in archae (SmAP). While an atomic resolution structure of a eukaryotic snRNP core has not yet been determined, smAP monomer and dimer structures are nearly identical to those found in humans and thus can be used to model the eukaryotic snRNP core. The structure presented in PDB entry **1lnx** *contains the crystallographic structure determination of an SmAP1 homoheptamer from* Pyrobaculum aerophilum. *Each monomer is comprised of a five stranded beta sheet (shown in yellow) and a short N-terminal alpha helix (shown in red). The seven monomers are arranged to form a central pore, approximately 8 to 9 Å in diameter that the authors propose as the oligouridine-binding site.*

PDB ID: 1lnx

Mura, C., Kozhukhovsky, A., Gingery, M., Phillips, M., Eisenberg, D. (2003): The oligomerization and ligand-binding properties of Sm-like archaeal proteins (SmAPs). Protein Sci. ***12****, pp. 832-47.*

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 of the National Institute of Standards and Technology
9600 Gudelsky Drive
Rockville, MD 20850

www.pdb.org

Send questions or comments to:

info@rcsb.org

RCSB PDB TEAM MEMBERS

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

Dr. Helen M. Berman, *Director*
Department of Chemistry and Chemical Biology, Rutgers University
610 Taylor Road, Piscataway, NJ 08854-8087
732-445-4667
Fax: 732-445-4320
berman@rcsb.rutgers.edu

Dr. Philip E. Bourne, *Co-Director*
San Diego Supercomputer Center, University of California, San Diego
9500 Gilman Drive
La Jolla, CA 92093-0537
858-534-8301
Fax: 858-822-0873
bourne@sdsc.edu

Judith L. Flippen-Anderson, *Production and Outreach Leader*
Department of Chemistry and Chemical Biology, Rutgers University
610 Taylor Road
Piscataway, NJ 08854-8087
732-445-0103
Fax: 732-445-4320
flippen@rcsb.rutgers.edu

Dr. Gary L. Gilliland, *Co-Director*
Center for Advanced Research in Biotechnology, National Institute of Standards and Technology
9600 Gudelsky Drive
Rockville, MD 20850
301-738-6262
Fax: 301-738-6255
gary.gilliland@nist.gov

Dr. John Westbrook, *Co-Director*
Department of Chemistry and Chemical Biology, Rutgers University
610 Taylor Road
Piscataway, NJ 08854-8087
732-445-4290
Fax: 732-445-4320
jwest@rcsb.rutgers.edu