PDB NEWSLETTER

NUMBER 19 ◆ FALL 2003 ◆ PUBLISHED QUARTERLY BY THE PROTEIN DATA BANK

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

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

Contents

Snapshot: October 1, 2003

22,700 released atomic coordinate entries

Molecule Type

20,501	proteins, peptides, and viruses
1,233	nucleic acids
948	protein/nucleic acid complexes
18	carbohydrates

Experimental Technique

19,285	diffraction and other
10,652	structure factor files
3,415	NMR
1,697	NMR restraint files

Participating RCSB Members

SDSC/UCSD: www.pdb.org

Rutgers: rutgers.rcsb.org

CARB/NIST: nist.rcsb.org

E-mail: info@rcsb.org

FTP: ftp.rcsb.org

Message from the PDB

This past summer, the PDB attended several meetings and discussed with users several of the new features available. One of these developments is the release of files and images of the biologically active units for all structures in the archive. Another is the release of standalone software for the preparation, validation, and deposition of macromolecular data.

- The PDB has provided separate coordinate and image files containing the biological molecule. These can be accessed from the Structure Explorer "View Structure" and "Download/Display File" pages.
- Software is available for download and installation on your desktop from **deposit.pdb.org/software**. Software binaries are provided as compressed files; README files are included with installation instructions. Some of the programs available include: PDB_EXTRACT, which extracts mmCIF data from structure determination applications; ADIT, which prepares files that can be emailed to the PDB for deposition; and the PDB Validation Suite, which processes and checks structure data.

We welcome your comments and questions about these features at info@rcsb.org.

The PDB ✦

PDB members, including Huanwang Yang (center) and Kyle Burkhardt, (right), demonstrated new deposition software to Rongguang Zhang (Argonne National Laboratory) and other attendees at the ACA 2003 meeting this past summer.

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

MIRROR SITES

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

National University of Singapore: pdb.bic.nus.edu.sg

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

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

Max Delbrück Center for Molecular Medicine (Germany): www.pdb.mdc-berlin.de

Data Deposition and Processing

PDB Deposition Statistics

Since the beginning of the year, approximately 3,500 structures have been deposited to the PDB. Eighty percent are deposited with a "hold until publication status," 12% are deposited with a "release immediately" status, and 8% are held until a specified date.

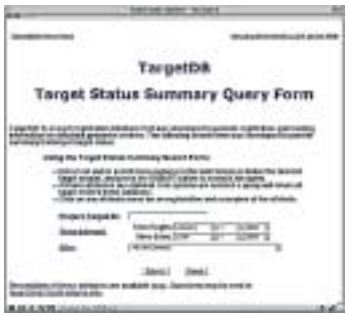
Eighty-three percent are from X-ray crystallographic studies and 12% are from NMR studies.

Fifty-five percent of these depositions release the sequence in advance of the structure's release. Seventy-one percent of these depositions were deposited with experimental data.

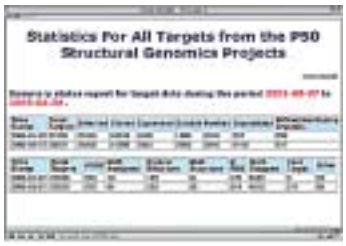
New Summary Report Available from TargetDB

TargetDB (**targetdb.pdb.org**) is a database of registration and tracking information for structural genomics centers worldwide. TargetDB provides timely status and tracking information on the progress of the production and solution of structures.

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



*The new Target Status Summary Query Form (**targetdb.pdb.org/nih**)*



Summary report for all NIH centers (January 1–June 1, 2003) from TargetDB's new Target Status Summary Query Form

A new feature at the TargetDB site is a search form that can provide summary tracking of target status. This form is now available from TargetDB's main page. Users can search by Target ID, date, or site(s). The summary report includes the number of targets at any given stage in the pipeline. With this new option it is possible to track the progress of a target or of an entire center over a user defined time interval. For instance, reports can be created for each NIH center describing the number of targets in each status category by project year.

Further information about TargetDB and links to structural genomics resources are available at **www.rcsb.org/pdb/strucgen.html**. ✦

Data Query, Reporting, and Access

New Features on the PDB Web Site and FTP Archives

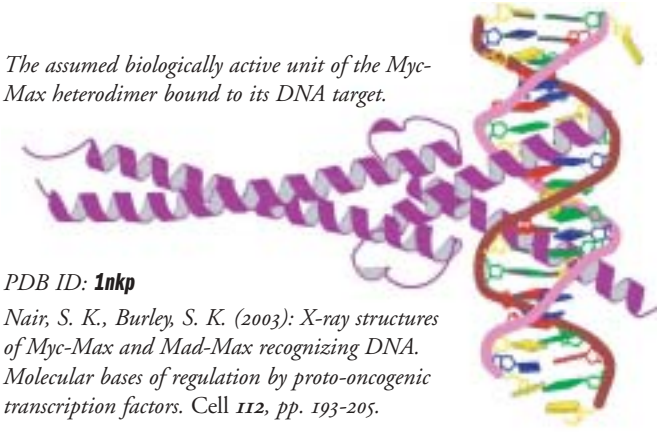
During this past quarter, several new developments that were previously in beta test have been incorporated into the primary PDB Web site, its mirrors, and on the PDB FTP site:

- **Remediated mmCIF Files**
  The primary PDB Web site and its mirrors now offer the remediated mmCIF files, previously referred to as the "beta" mmCIF files. These files have replaced the set of automatically translated mmCIF files that were created with pdb2cif.pl. The 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 PDB FTP site at **ftp://ftp.rcsb.org/pub/pdb/data/structures/all/mmCIF/**. The translated files have been removed from the FTP site, and will be made available upon individual request.

  For every experimentally-solved structure, both current and obsolete, there is now a remediated mmCIF file in the corresponding directory of the FTP archive in Unix compressed (.Z) format. mmCIF files are not provided for theoretical models. However, the translation software (pdb2cif.pl) is provided at **www.bernstein-plus-sons.com/software/pdb2cif** for users who wish to generate the translated mmCIF files.

The assumed biologically active unit of the Myc-Max heterodimer bound to its DNA target.



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.*

- **Biological Unit Images and Coordinate Files**
  The View Structure section of the Structure Explorer now offers still ribbon images of the assumed biological unit(s) for structures, where relevant, in addition to 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.

  Biological unit coordinate files are also available from the FTP archive at **ftp://ftp.rcsb.org/pub/pdb/data/biounit**. These files are accessible in gzipped (.gz) format. Links to a variety of free uncompression tools can be found at **www.rcsb.org/pdb/help-general.html#format_structure_compressed**.

- **Enzyme Names and EC Number Query** The Summary Information section of the Structure Explorer page for enzymes in the PDB now displays the enzyme name for entries with complete EC numbers. This section also supports queries for all other PDB entries with the same EC number by clicking on the number shown for that entry.

Questions about these new features may be sent to **info@rcsb.org**.

New Version of OpenMMS Toolkit Released

Version 1.5.1 of the OpenMMS software toolkit is now available on the OpenMMS Web site at **openmms.sdsc.edu**. In addition to the fast database loader, this version contains all of the pdbx fields from the PDB Exchange Dictionary (**deposit.pdb.org/mmcif/**). This includes attributes such as the model number in the atom_site record which is needed for NMR structures. Also available in this release is the "xconv" program which converts mmCIF files to XML files with the standard XML/PDB format.

PDB Focus: Using rsync to Mirror the PDB FTP Site

One freely available method for establishing and maintaining a local copy of the PDB FTP Site is rsync. The RCSB-created script, rsyncPDB.sh, is a template for using rsync to mirror the FTP archive from an anonymous rsync server. This script can be found at **ftp://ftp.rcsb.org/pub/pdb/software/**, and the comments in the script explain its usage. Prior to running it, users will need to set three variables in rsyncPDB.sh to suit their local setup. This script is used by the PDB to maintain its FTP mirrors.

General rsync documentation can be found at **www.samba.org/rsync**. An overview of PDB FTP mirroring procedures is offered at **www.rcsb.org/pdb/ftpproc.final.html**, and the layout of the PDB FTP archive is accessible at **www.rcsb.org/pdb/ftp_plan.html**.

For more information about rsync or mirroring the PDB FTP Site, please send e-mail to **info@rcsb.org**.

PDB Web Site Statistics

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

The access statistics are given below for the main PDB Web site at **www.pdb.org**.

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

	DAILY AVERAGE			MONTHLY TOTALS		
MONTH	HITS	FILES	SITES	KBYTES	FILES	HITS
Sep 03	202,285	153,680	96,603	198,614,136	4,456,720	5,866,265
Aug 03	131,014	100,838	65,563	132,975,741	3,025,151	3,930,448
Jul 03	146,399	111,723	75,051	131,039,156	3,351,718	4,391,985

PDB Co-Director, Philip E. Bourne, presents "The Re-engineered PDB" to ISMB Conference attendees.

PDB Outreach

PDB at the ISMB, ACA, and Protein Society Meetings

PDB staff members participated in several meetings this summer. We received excellent feedback from those in attendance at these events, which will help to guide PDB's future developments:

11th International Conference on Intelligent Systems for Molecular Biology *(ISMB; June 29-July 3, Brisbane, Queensland, Australia)*
New query and reporting features of the re-engineered PDB site were demonstrated at PDB's exhibit and as part of a presentation on "The Re-engineered PDB" at this conference.

American Crystallographic Association's Annual Meeting *(ACA; July 26-31, Northern Kentucky Convention Center)*
Desktop versions of various software programs were demonstrated at the PDB exhibit, and a standalone version of ADIT software was distributed at this meeting. Posters were presented on "TargetDB: A Target Registration Database for Structural Genomics" and on the Ligand Depot. A presentation on "PDB Data Assembly and Validation Tools" was part of the Computational Methods session. The PDB Poster Prize, which is described further in this newsletter, made its debut at the ACA meeting.

17th Symposium of the Protein Society *(July 27-29, Boston, MA)*
A poster on the "New Features of the Protein Data Bank" was presented at this symposium.

PDB Poster Prize—2003 Winners Announced

The PDB Poster Prize was awarded to the best student poster presentations at this year's meetings of the IUCr Regional Associates—the American Crystallographic Association (ACA), the Asian Crystallographic Association (AsCA), and the European Crystallographic Association (ECA/ECM).

The winners are:

- **ACA**—The first-ever PDB Poster Prize was awarded to Ty Gould for the poster "Quorum Sensing Signal Generation by the AHL Synthsase LasI in *Pseudomonas aeruginosa* Pathogenesis" (T.A. Gould[1], R.C. Murphy[1], H.P. Schweizer[2], M.E.A. Churchill[1]; [1]Dept. of Pharmacology, Univ. of Colorado Health Sciences Center, Denver, CO; [2]Dept. of Microbiology, Colorado State Univ., Fort Collins, CO). A runner-up award was made to Paul

Ty Gould (left), PDB Poster Prize winner, and Paul Hubbard (right), PDB Poster Prize runner up

Hubbard for the poster "Structure and Catalytic Mechanism of Bacterial 2, 4–Dienoyl CoA Reductase." (Xiquan Liang[3], Horst Schulz[3], Jung-Ja Kim, Department of Biochemistry, Medical College of Wisconsin; [3]Department of Chemistry, The City University of New York). Special thanks to the ACA PDB Poster Prize Committee members—Vivien Yee (Chair), Victor Young, Tom Koetzle, Sylvie Doublie, Marvin L. Hackert,—and the committee's organizer, Jeanette Krause Bauer.

- **AsCA**—The prize was awarded to Janet Deane for the poster "Crystal structure of a complex of FLINC4, an intramolecular LMO4:LDB1 complex" (Janet E. Deane, Megan Maher, J. Mitchell Guss, and Jacqueline M. Matthews, School of Molecular and Microbial Biosciences, University of Sydney). Special thanks to the judges of all of the student posters at AsCA—Ted Baker (Chair), Peter Colman, Janet Smith, Mark Spackman, Colin Raston, and Yu Wang.
- **ECM—**The prize was awarded to Carina Lobley for the poster "Structural Studies of the Enzymes of Pantothenate Synthesis" (Carina M.C. Lobley[1], Mairi L. Kilkenny[1], Florian Schmitzberger[1], Michael E. Webb[2], Chris Abell[2], Alison G. Smith[3,1], Tom L. Blundell[1]; [1]Department of Biochemistry, Cambridge; [2]University Chemical Laboratory, Cambridge; [3]Department of Plant Sciences, Cambridge). Special thanks to the judges of all of the student posters at ECM—G. Davies, C. Kenyon, E.F. Garman, and A. Roodt.

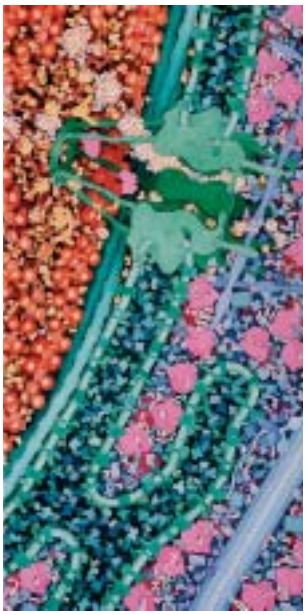
PDB Art Part of Molecular Gallery Show at Cal State Fullerton

The PDB's traveling art exhibit is part of the *Art of Science* show at California State University, Fullerton. Featured are images from the PDB and 3-D models by David Goodsell and Arthur Olson (The Scripps Research Institute). Also included are objects from the Cal State Fullerton Keck Center for Molecular Structure, such as an older generation X-ray camera. The *Art of Science* will run through December 19 at the Atrium Gallery in the Paulina June & George Pollak Library.

This 3-D model of the digestive enzyme chymotrypsin is part of the Art of Science *exhibit at Cal State Fullerton.*

The PDB's *Art of Science* exhibit includes large-scale depictions of proteins and images and text from the *Molecule of the Month* series. The PDB would like to see the *Art of Science* travel to other places. If you would be interested in sponsoring this exhibit at your institution, please let us know at info@rcsb.org.

*Illustrations of "Macrophage and Bacterium" by Goodsell Win Award in NSF/*Science *Visualization Contest*



Part of "Macrophage and Bacterium 2,000,000X" by David S. Goodsell, The Scripps Research Institute

PDB contributor and author of the *Molecule of the Month* series, David S. Goodsell, has been awarded second prize in the 2003 Science and Engineering Visualization Challenge for his illustration "Macrophage and Bacterium 2,000,000X". The challenge is a joint project of the National Science Foundation and *Science* Magazine that encourages and promotes the visual and conceptual beauty of science and engineering.

Dr. Goodsell is an Associate Professor of Molecular Biology at The Scripps Research Institute in La Jolla, California. His research involves computational chemistry and biomolecular computer graphics. Goodsell's artistic talents are multi-faceted, and include many renderings of molecules and cells that are drawn, painted, or generated with the aid of graphics programs. Previously he was awarded the Association of Medical Illustrators Literary Award, among other distinctions he has received. He has contributed many wonderful installments for the PDB *Molecule of the Month* feature and has authored the popular PDB *Molecular Machinery* poster—a recent interview with Dr. Goodsell that further describes these efforts can be found in the PDB Newsletter's Spring 2003 issue.

Dr. Goodsell's winning series of three paintings shows a macrophage engulfing a bacterium, including all of the macromolecules in the two cells and in the surrounding blood serum. Goodsell used hundreds of PDB structures for the paintings to get the sizes and shapes of the molecules right: "The *Molecular Machinery* poster is a good example of the type of information that I start with when I approach a new painting—lots of structures all drawn at a consistent size."

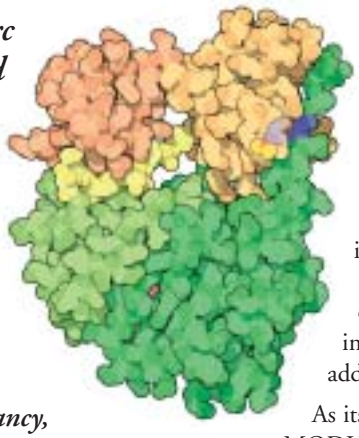
The full winning entry can be viewed at www.scripps.edu/pub/goodsell/gallery/macrophagebacterium.html. The original paintings are currently on display in the Center for Integrative Molecular Biosciences at The Scripps Research Institute in La Jolla.

PDB Education Listserv

A listserv for educators who use the PDB has been established by Dr. Judith Voet, Professor of Biochemistry at Swarthmore College and member of the PDB Advisory Committee. The purpose of this forum is to provide the PDB with feedback on its present usability by students and educators, and ideas for future directions in support of education. If you would like to participate in the discussion, send an e-mail message to jvoet1@swarthmore.edu with the request to subscribe.

PDB Molecules of the Quarter: Src Tyrosine Kinase, Calmodulin, and Estrogen Receptor

The *Molecule of the Month* series, by David S. Goodsell, explores the functions and significance of selected biological macromolecules for a general audience. These installments are available at www.rcsb.org/pdb/molecules/molecule_list.html. A sample of the molecules featured during this past quarter are included below:



Src Tyrosine Kinase
PDB ID: **2src**
W. Xu, A. Doshi, M. Lei, M.J. Eck, S.C. Harrison (1999): Crystal structures of c-Src reveal new features of its autoinhibitory mechanism. Molecular Cell ***3****, pp. 629-38.*

Src Tyrosine Kinase: Signaling, Redundancy, and Cancer

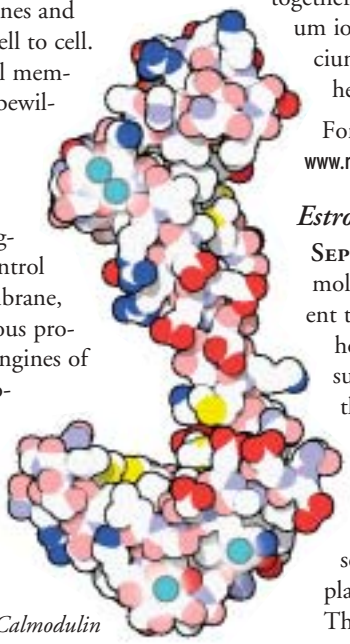
July, 2003—Your body is a democratic nation of cells. Each cell is an individual with its own needs, but all of your cells work together to keep you alive. As you might imagine, this requires an incredible amount of cooperation. Cells are in constant communication to inform their neighbors of their needs and future plans. They send messages to each other, passing hormones and chemokines and other molecular messages from cell to cell. These messages are received by proteins in the cell membrane, which transmit the signal inside. There, a bewilderingly complex collection of proteins relays the message to all of the appropriate places inside the cell.

The Src protein, shown in PDB entry **2src**, is a signaling protein that specializes in messages that control the growth of cells. It sits just inside the cell membrane, where it assists in the passing of signals from various protein receptors to the proteins that turn "on" the engines of protein synthesis and cellular growth. Src is a tyrosine kinase, so it relays its messages by adding phosphate groups to special tyrosine amino acids in protein chains. It adds phosphate groups to a wide variety of proteins that control cellular structure, cell communication, and cellular growth, turning them "on" and releasing them to perform their individual tasks.

For more information about Src tyrosine kinase, see www.rcsb.org/pdb/molecules/pdb43_2.html.

Calmodulin: Sensing Calcium

August, 2003—Calcium is the most plentiful mineral element found in your body, with phosphorous coming in second. This probably doesn't come as a surprise, since your bones are strengthened and supported by about two kilograms of calcium and phosphorous. Your body also uses a small amount of calcium, in the form of calcium ions, to perform more active duties. Calcium ions play essential roles in cell signaling, helping to control processes such as muscle contraction, nerve signaling, fertilization and cell division. Through the action of calcium pumps and several kinds of calcium binding proteins, cells keep their internal calcium levels 1,000-10,000 times lower than the calcium levels in the blood. Thus when calcium is released into cells, it can interact with calcium sensing proteins and trigger different biological effects, causing a muscle to contract, releasing insulin from the pancreas, or blocking the entry of additional sperm cells once an egg has been fertilized.



Calmodulin
PDB ID: **3cln**
Y.S. Babu, C.E. Bugg, W.J. Cook (1988): Structure of calmodulin refined at 2.2 Å resolution. J. Mol. Biol. ***204****, pp. 191-204.*

As its name suggests, calmodulin is a CALcium MODULated proteIN. It is abundant in the cytoplasm of all higher cells and has been highly conserved through evolution. Calmodulin acts as an intermediary protein that senses calcium levels and relays signals to various calcium-sensitive enzymes, ion channels and other proteins. Calmodulin is a small dumbbell-shaped protein composed of two globular domains connected together by a flexible linker. Each end binds to two calcium ions. PDB entry **3cln** has all four sites filled with calcium ions and the linker has formed a long alpha helix separating the two calcium-binding domains.

For more information on calmodulin, see www.rcsb.org/pdb/molecules/pdb44_2.html.

Estrogen Receptor: a Large Family

September, 2003—Estrogens are small, carbon-rich molecules built from cholesterol. This is quite different than larger hormones, such as insulin and growth hormone, which are sensed by receptors on the cell surface. Estrogens pass directly into cells throughout the body, so the cell can use receptors that are in the nucleus, right at the site of action on the DNA. When estrogen enters the nucleus, it binds to the estrogen receptor, causing it to pair up and form a dimer. This dimer then binds to several dozen specific sites in the DNA, strategically placed next to the genes that need to be activated. Then, the DNA-bound receptor activates the DNA-reading machinery and starts the production of messenger RNA.

When researchers looked into the human genome, they found over 150 proteins that are similar to the estrogen receptor. This is a large family of nuclear receptors that sense the levels of small hormones and other signaling molecules, such as steroid and thyroid hormones, vitamin D, and retinoic acid. Like estrogen, these are all small molecules that pass directly into cells and find their way to the nucleus. These receptors each bind to a specific signaling molecule and then activate or repress their own set of 50-100 genes. For more information on nuclear receptors from a genomic perspective, take a look at the *Protein of the Month* feature at the European Bioinformatics Institute. (www.ebi.ac.uk/interpro/potm/archive.html)

For more information about the estrogen receptor, see www.rcsb.org/pdb/molecules/pdb45_2.html. ✦

PDB Community Focus: Brian W. Matthews

Brian W. Matthews is a long-time PDB depositor. From 1982 through the present time, he has contributed approximately 500 sets of coordinates to the archive—more than any other single author. A Professor of Physics and a member of the Institute of Molecular Biology at the University of Oregon, Matthews' numerous distinctions and achievements also include an appointment as a Howard Hughes Medical Institute investigator and membership in the U.S. National Academy of Sciences. He is involved in studying some of the fundamental problems in biology, and his X-ray studies have always played a critical role in his research. The PDB recently interviewed Prof. Matthews on his experiences in this field:

PDB: Over the years you have been an integral part of the tremendous technical revolution that has taken place in protein crystallography. How long did it take you to solve your first structure and what methods did you use? How long would it take to do the same structure today and why?

Prof. Matthews: The first structure that I solved was of a "small molecule" of 14 atoms including a sulfur that was used as a "heavy atom". It took me over a year in part because all of the calculations had to be done by hand and also because the Patterson function gave misleading information regarding the location of the sulfur. Today such a structure would be solved in hours, if not minutes.

The first protein structure determination with which I was involved was that of alpha-chymotrypsin. David Blow had been working on the project at the MRC Lab in Cambridge and he and Michael Rossmann had used the structure as a test case during their early development of the molecular replacement technique. Their calculations suggested that the two molecules in the asymmetric unit were probably related by a local two-fold axis. Shortly after I arrived in Cambridge, Barbara Jeffries, a technician working on the project, found that the inclusion of 2% dioxane in the crystallizing medium eliminated persistent twinning.

This made it possible to grow large, untwinned crystals and to make a serious attempt at high-resolution data collection. Data collection was by precession photography. A single data set necessitated 20-30 film packs, took perhaps three months to collect and a similar amount of time to process. Paul Sigler joined the project about six months after I had arrived and Richard Henderson joined the group later as a starting graduate student. From the time that I joined the project until the structure was solved took over three years, but as I mentioned, the project had already been underway for several years. We used the method of isomorphous replacement following the pattern set with myoglobin and lysozyme. In the case of alpha-chymotrypsin we could average the density for the two molecules in the asymmetric unit and this substantially improved the quality of the electron density map. We also included information from anomalous scattering measurements incorporating ideas that were being developed around that time. Given high quality crystals the structure could probably be solved today within a month or so. As I mentioned, however, the initial crystals were subject to frequent twinning. Without the introduction of dioxane by our technician, Barbara Jeffries, the structure might still remain unsolved even today.

PDB: How has your research program evolved over the years? What role has crystallography played in that evolution?

Prof. Matthews: Because of my involvement with the alpha-chymotrypsin project, I had an early interest in proteolytic enzymes. Therefore it was natural for me, when I started my own laboratory at the University of Oregon, to work on the thermostable protease thermolysin. Because it was a zinc peptidase we were curious as to whether the structure might be related to that of carboxypeptidase which by then had been determined by the Lipscomb group at Harvard. Also I had begun to develop an interest in protein stability and folding.

Crystallography has been central to all of the subsequent work that we have done. The structure of the Cro protein immediately suggested how it might bind to DNA. The introduction of site-directed mutagenesis, coupled with crystallography, led to our exploitation of the T4 lysozyme system as a way to try to understand the structural basis of protein stability. It also contributed to what has been described as our "pollution" of the Protein Data Bank with structures of mutant lysozymes.

PDB: You deposited your first structure approximately 20 years ago - what do you think crystallography will be like for someone depositing their first structure 20 years from now?

Prof. Matthews: I was extremely excited when we were able to determine our first structure in my own lab. This was thanks to the outstanding contributions of Peter Colman, my first postdoc, together with Hans Jansonius who spent 1971 in Eugene as a sabbatical visitor. I am pleased to see that students and postdocs who determine their first structure in my group still share that same excitement. I certainly hope that this will remain 20 years in the future. At the time that I started my career in structural biology it was sufficient to be just a crystallographer. Now knowledge of crystallography is just one tool in the repertoire of skills that are necessary. I like to think that crystallography remains as the most powerful tool that we have and hope that it will remain so in the future.

PDB: The PDB plans to be there and to be ready—do you have any advice for us based on your experience?

Prof. Matthews: Based on my own experience I have nothing but good things to say about the PDB. On a technical level the PDB has adapted extremely well to handle the ever-increasing number of depositions. I have also found the PDB willing to consider suggestions for improvement. Also the PDB has played an essential role in helping to ensure that key structural information are preserved and made available in a timely fashion to the community at large. Keep up the good work!

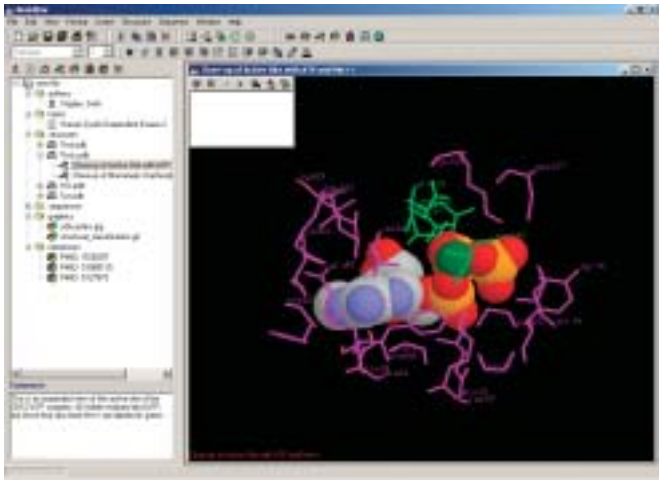
PDB Education Corner

PDB's Education Corner features a different teacher each quarter, offering an account of how he or she uses the PDB to educate students. This quarter's column is by Prof. Paul A. Craig, Associate Professor of Chemistry at the Rochester Institute of Technology:

The Protein Data Bank plays a critical role in teaching, learning and research in Biochemistry education at the Rochester Institute of Technology (RIT), a career-oriented comprehensive university with about 17,000 students. In the College of Science, students majoring in Biology, Chemistry, Biochemistry, Biotechnology and Clinical Chemistry all take courses in Biochemistry. The complete Biochemistry sequence consists of three courses which last a full academic year under our quarter system. The first course covers amino acids, proteins and membranes, the second is metabolism and the third is about nucleic acids and molecular genetics.

Teaching and Learning

We use *Biochemistry* by Berg, Tymoczko and Stryer (Berg JM, Tymoczko JL, Stryer L, 5th edition, W.H. Freeman, New York, 2002) in our courses. Tim Driscoll of molvisions (**molvisions.com**) has developed an excellent set of *Living Figures* for the publisher which we use in lecture to demonstrate the structural features of proteins and nucleic acids throughout the course sequence. We believe that the students learn much more, however, when they work on developing their own structure documentaries based on PDB files. To emphasize the importance of understanding structure and function in these courses, these documentaries account for 20% of their course grades. For these projects, students are asked to select a particular type of protein (e.g., a kinase or a protease) from the PDB which includes either an inhibitor (for the first course in our sequence) or a complex with a nucleic acid (for the third course, which focuses on molecular biology). They then use the PDB Structure Explorer page to identify the primary reference for the structure. From this starting point, they go to PubMed (**www.ncbi.nlm.nih.gov**) to identify at least two additional referred resources. From these sources the students prepare a structure annotation that includes a brief introduction, a statement of the physiological role of the structure, a detailed description of the structure of the protein which includes general features (e.g., secondary structures) and specific features of the active site and the interaction between the protein and inhibitor (or nucleic acid), and a summary of the family relationships of the structure.



Screenshot of a BioEditor documentary on human cyclin dependent kinase 2, created by Seth Staples, RIT.

Students are also required to prepare a series of molecular visualizations for the structure part of the documentary. This requires that students spend a few hours learning how to read PDB files and compare information in their files with the related publications. After earlier attempts with Kinemage and Rasmol, we spent several years helping students prepare Chime pages, with varying degrees of success. For the last two years the students have been preparing their structure documentaries using a program called BioEditor (**bioeditor.sdsc.edu**) that was developed at the San Diego Supercomputer Center (more on that below).

A more detailed description of the project requirements can be found at **www.rit.edu/~pac8612/Biochemistry/502(702)/protein_vis.htm**. A number of students have produced stellar documentaries using BioEditor and these can be viewed on the BioEditor Web site at **bioeditor.sdsc.edu/documentaries.shtml**.

Software Development

My professional interests are at the interface between computers and biology. This has included simulations of several protein separation processes. The most successful of these is an electrophoresis simulation (**www.rit.edu/~pac8612/electro/Electro_Sim.html**). In communicating with the programmers, I have always found that PDB files are an effective starting point to explain sequence and structure information.

I was fortunate to spend the 2001-2002 academic year on sabbatical at the San Diego Supercomputer Center where I worked Peng Yang (a programmer), David Goodsell (a molecular biologist and a wonderful artist—see the PDB *Molecule of the Month* pages) and Phil Bourne on developing a software tool called BioEditor (**bioeditor.sdsc.edu**), which is designed to facilitate the annotation of macromolecular structure (Yang P, Craig PA, Goodsell D, Bourne PE. BioEditor—Simplifying Macromolecular Structure Annotation, *Bioinformatics* (2003) **19**: 897-898). This tool enables users to assemble information and images they need to create documentaries about a particular structure or family of structures. It includes a number of powerful features that enable users to prepare structure documentaries: a built-in browser based on Internet Explorer which enables users to directly download and incorporate PDB structure files, a molecular visualization interface based on the Chime plug-in, a tool for cataloguing images that are created by the user or collected from other sites, and the ability to assemble references based only on their PubMed ID numbers (PMIDs).

Students in Computer Science and Bioinformatics at RIT continue to work on improving the ease of use and power of BioEditor as part of their undergraduate and M.S. degree programs. ✦

PDB Project Team Leaders

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

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

Dr. Philip E. Bourne
Co-Director
SDSC/UCSD
bourne@sdsc.edu

Judith Flippen-Anderson
Production and Outreach Leader
Rutgers University
flippen@rcsb.rutgers.edu

Dr. Gary Gilliland
Co-Director
CARB/NIST
gary.gilliland@nist.gov

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

PDB Members

Rutgers

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

Prentice Bisbal
prentice@rcsb.rutgers.edu

Kyle Burkhardt
kburkhar@rcsb.rutgers.edu

Li Chen
lchen@rcsb.rutgers.edu

Sharon Cousin
sharon@rcsb.rutgers.edu

Dr. Shuchismita Dutta
sdutta@rcsb.rutgers.edu

Dr. Zukang Feng
zfeng@rcsb.rutgers.edu

Lew-Christiane Fernandez
fernandz@rcsb.rutgers.edu

Dr. Rachel Kramer Green
kramer@rcsb.rutgers.edu

Vladimir Guranovic
vladimir@rcsb.rutgers.edu

Dr. Shri Jain
sjain@rcsb.rutgers.edu

Dr. Rose Oughtred
rose@rcsb.rutgers.edu

Dr. Irina Persikova
persikova@rcsb.rutgers.edu

Suzanne Richman
richman@rcsb.rutgers.edu

Melcoir Rosas
melcoir@rcsb.rutgers.edu

Dr. Bohdan Schneider
bohdan@rcsb.rutgers.edu

Dr. Huanwang Yang
hyang@rcsb.rutgers.edu

Dr. Jasmin Yang
jasmin@rcsb.rutgers.edu

Christine Zardecki
zardecki@rcsb.rutgers.edu

SDSC/UCSD

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

Dr. Ken Addess
addess@sdsc.edu

David Archbell
dave@sdsc.edu

Tammy Battistuz
tammyb@sdsc.edu

Dr. Wolfgang F. Bluhm
wbluhm@sdsc.edu

Dr. Nita Deshpande
nita@sdsc.edu

Jeff Merino-Ott
jott@sdsc.edu

Wayne Townsend-Merino
wayne@sdsc.edu

CARB/NIST

Center for Advanced Research in Biotechnology
National Institute of Standards and Technology
9600 Gudelsky Drive
Rockville, MD 20850

Al Carlson
carlson@umbi.umd.edu

Dr. Veerasamy Ravichandran
vravi@nist.gov

Padma Priya Paragi Vedanthi
padmap@nist.gov

Elizabeth Walker
walkere@umbi.umd.edu

Statement of Support

The PDB is supported by funds from the National Science Foundation, the Department of Energy, and the National Institutes of Health, in addition to resources and staff made available by the host institutions.

PROTEIN DATA BANK
www.pdb.org

U.S. Department of Commerce
National Institute of Standards and Technology
100 Bureau Drive, Stop 8314
Gaithersburg, MD 20899-8314

OFFICIAL BUSINESS
PENALTY FOR PRIVATE USE, $300

Return Service Requested

FIRST-CLASS MAIL
POSTAGE & FEES PAID
NIST
PERMIT No. G195