RCSB PDB NEWSLETTER

NUMBER 25 ◆ SPRING 2005 ◆ *Published quarterly by the Research Collaboratory for Structural Bioinformatics Protein Data Bank*

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

Contents

Snapshot: April 1, 2005

30263 released atomic coordinate entries

Molecule Type	Experimental Technique
27579 proteins, peptides, and viruses	25803 diffraction and other
1247 nucleic acids	4460 NMR
1424 protein/nucleic acid complexes	16188 structure factor files
13 carbohydrates	2469 NMR restraint files

Participating RCSB Members

Rutgers: rutgers.rcsb.org

SDSC/UCSD: www.pdb.org

E-mail: info@rcsb.org

FTP: ftp.rcsb.org

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

Message from the RCSB PDB

This summer, the RCSB PDB will be demonstrating the beta site and other tools and resources at a variety of meetings:

• The RCSB PDB exhibit booth will be part of the ***American Chemical Society Mid-Atlantic Regional Meeting*** (MARM; May 22-25, 2005; Rutgers, The State University of New Jersey). *The Art of Science* show will also be on view at The Gallery, a space dedicated to art exhibits at Rutgers University.

• Kyle Burkhardt will be presenting "Using RCSB PDB tools to validate macromolecular structures" as part of the macromolecular structure validation workshop being held on Saturday May 28, in addition to the exhibit booth at the ***American Crystallographic Association Meeting*** (ACA; May 28 - June 2, 2005; Orlando, FL).

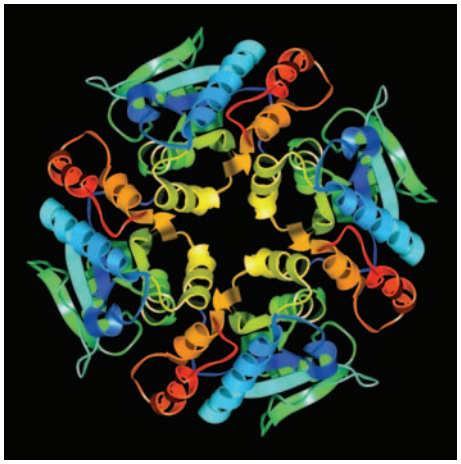
• At the ***Intelligent Systems for Molecular Biology*** (ISMB) conference (June 25 - 29, 2005; Detroit, MI), RCSB PDB beta site demonstrations will be the focus of the exhibit booth.

• The RCSB PDB will be exhibiting alongside MSD-EBI and PDBj in the wwPDB stand at the ***XX Congress of the International Union of Crystallography*** (IUCr; August 23-31 2005; Florence, Italy). Presentations about validation and mmCIF applications, as well as a round table discussion about Data Mining from the PDB, will also be part of the program.

The *RCSB PDB Poster Prize* will also be awarded at the ACA and IUCr meetings, as well as at the Conference on Research in Computational Molecular Biology (RECOMB).

Further details about all of these events will be made available on the website. We look forward to seeing the diverse PDB community this summer!

The RCSB PDB ◊



The Protein Structure Initiative (PSI), a national program aimed at determining the three-dimensional shapes of a wide range of proteins, has now produced more than 1,000 different structures. These structures will shed light on how proteins function in many life processes and could lead to targets for the development of new medicines.

The full story from the NIGMS is available from www.nigms.nih.gov/news/releases/021005.html. Shown: **PDB ID: 1XN4**

Data Deposition and Processing

Weekly Deadlines for Entry Release and Modification Requests

PDB entries are processed by the three members of the wwPDB (RCSB, MSD-EBI and PDBj), and are released immediately (REL), when the corresponding paper is published (HPUB), or on a particular date (HOLD).

Each week, all files scheduled for release or modification are checked and validated one final time. Authors may be contacted to resolve any issues that may arise while preparing the entries for release.

When the release of HPUB structures is requested, the PDB staff routinely confirms the primary citation. If this is not accomplished within that release cycle, the entry may be scheduled to be released in a later update.

For entry release/modifications to be included in the next update, any required author correspondence should be sent to the appropriate wwPDB member by the following times:

- **RCSB (deposit@rcsb.rutgers.edu): 15:00 EST Friday**
- **MSD-EBI (pdbhelp@ebi.ac.uk): 15:00 GMT Thursday (10:00 EST Thursday)**
- **PDBj (adit@adit.protein.osaka-u.ac.jp): 13:00 JST Thursday (23:00 EST Wednesday)**

All entries due for release are transferred to the RCSB for final packaging into the master PDB ftp archive. These files are then released by 4:00 EST each Wednesday.

Requests received after these cutoff times will be processed during the next update cycle.

Changes and Corrections to Entries in the PDB Archive

Corrections to entries originally processed by the RCSB, EBI-MSD, or PDBj are handled by the same annotation staff and subsequently reviewed by the author(s) depositing the structure. Any changes in released PDB entries are described in the PDB REVDAT records and in the mmCIF/XML category DATABASE_PDB_REV_RECORD.

When replacement coordinates for a released entry are provided by the depositing author, the original entry is made obsolete and the replacement coordinates are released in a new superseding PDB entry. The relationship between obsolete and superseding entries is stored in OBSLTE/SPRSDE PDB records and in the mmCIF/XML category PDBX_DATABASE_PDB_OBS_SPR. Queries of obsoleted entries on the RCSB PDB website always produce the most recent superseding entry. Obsoleted entries remain available in a separate area of the PDB ftp site, **ftp://ftp.rcsb.org/pub/pdb/data/structures/obsolete/**.

For the entries deposited prior to 1998, a variety of consistency checks have been performed. This has been done as part of an ongoing wwPDB project to maintain uniformity within the PDB archive. Examples of uniformity corrections include corrections related to atomic nomenclature for both macromolecule(s) and ligand(s), sequence-coordinate consistency, and the addition of missing records (*e.g.* citations, synonyms, and sequence database references). Corrections in the pre-1998 entries have been made only in the mmCIF and XML data files. The mmCIF and XML data files are offered as download options on the RCSB PDB website and are also available via ftp. Data uniformity efforts by MSD-EBI and PDBj will be incorporated into these data in the near future.

The XML data files were produced as part of a joint project by all wwPDB members, and these files are in the final stage of beta testing. Both mmCIF and XML data files conform to the PDB Exchange data dictionary. This dictionary is available in both mmCIF and XML schema form at **deposit.pdb.org/mmcif/**.

PDB Deposition Statistics

In the first quarter of 2005, 1398 experimentally-determined structures were deposited to the PDB archive.

The entries were processed by wwPDB team members at RCSB-Rutgers, MSD-EBI, and PDBj. Of the structures deposited, 71% were deposited with a release status of HPUB; 18% with REL; and 11% with HOLD.

84% of these entries were determined by X-ray crystallography; 14% were determined by NMR. 82% were deposited with experimental data. 57% released the sequence in advance of the structure's release.

Data Query, Reporting, and Access

RCSB PDB Beta Site Features

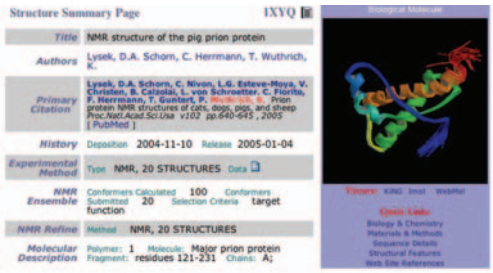
In July 2004, the RCSB PDB released a reengineered beta site (**pdbbeta.rcsb.org**) for public testing. Some of the features of this site are described below.

Comments and suggestions about the beta site are welcomed at **betafeedback@rcsb.org**.

- **Query-by-Example in Structure Explorer Pages**

The Structure Explorer page for any PDB entry can perform several "query-by-example" searches. Using the information available in a particular entry, structures containing the same data can be quickly found by clicking on that data item.

For example, all other entries by a specific primary citation author



Query-by-Example: Selecting the author name "Wüthrich, K."

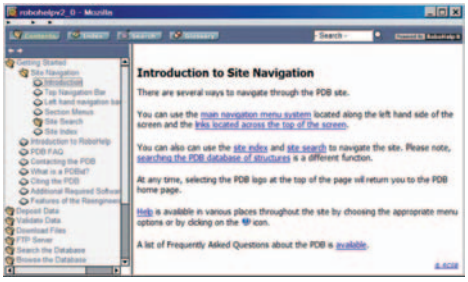
Returns all PDB entries with that author in the primary citation.

can be found by clicking on the author's name. A search of the database for all structures in which the selected author appears in the citation will be presented in a Query Results Browser.

Authors, Primary Citation Authors, Chemical Components, and GO, CATH, and SCOP classifications can all be used in "query-by-example" searches.

- **Detailed Help System**

The beta site uses Macromedia® RoboHelp® software to offer detailed guidance for navigating the site and understanding the information available from the RCSB PDB. The integrated system organizes all the help pages on the site and offers a table of contents, index, and glossary on every page. These tools help users to access the rich content offered by the help pages.



The beta site help system

The help system launches into a separate browser window to allow users to access the help information and the beta site at the same time. Once users are in the RoboHelp system, they can access help on specific topics by browsing through the table of contents listed in the left hand menu. These topics include Getting Started, Deposit Data, Validate Data, Download Files, FTP Server, Search/Browse the Database, Query Results Browser, and Viewing Molecules.

The Help link located at the top of the home page launches the system. Links to specific help features and topics (such as "How to Validate Data", "How to Search", and "Molecular Viewer Help") are located in the left hand navigation bars, menus and pages throughout the website. An introduction to using RoboHelp is also availableon the beta site.

The Glossary offered by the help system provides definitions for terms related to structural biology, and is an excellent source of information.

- **New Customizable Tabular and Structure Reports**

Customizable reports are now available as a new feature on the beta site. Users can select specific data items to display for individual structures or for a group of structures. These reports can be saved in a CSV format that can be opened in spreadsheet application programs.

Individual customizable reports can consolidate information from the four major areas of the "Summarize" pulldown menu found on the 'Structure Explorer' page (Biology and Chemistry, Materials and Methods, Sequence Details and Structural Features). A report also can display information specific for the type of the experimental method (X-Ray and NMR). In addition, a report can also provide ligand details, SCOP and CATH classifications, and the primary citation. An option to generate a custom structure report can be found in the pulldown menu on the 'Structure Explorer' page for a single structure. This type of reports includes images of the asymmetric and biological units. A PDF can be viewed and saved by clicking on the 'Print Page' link in the navigation bar at the top of the page.

Reports for a set of structures can include information from a variety of areas, including Sequence, Ligands, Biological Details, and Experimental Details. After selecting the set of structures, these reports can be generated by using the "Customize Report" option from the "Report" pulldown menu on the 'Query Results Browser' page. The report generated can be sorted in ascending/descending order of any column by clicking on the column headers.

Time-Stamped Copies of PDB Archive Available via FTP

Starting with January 2005, time-stamped yearly snapshots of the PDB Archive will be available from **ftp://snapshots.rcsb.org/**. It is hoped that these snapshots will provide readily identifiable data sets for research on the PDB archive.

Currently, the directory 20050106 is available. This directory contains the exact and complete contents of the FTP archive as it appeared on January 6, 2005. This includes the 29040 experimentally-determined coordinate files that were current (*i.e.*, not obsolete) in PDB and mmCIF formats. Data in XML (PDBML) format are not included in this first snapshot, but will be made available on DVD in the near future. Subsequent snapshots on this FTP server will include data in PDB, mmCIF, and PDBML formats.

This snapshot follows the historical directory structure -- coordinate files are contained in subdirectories named after the two middle characters of the PDB ID, for example, 100d is found in the directory '00'.

The date and time stamp of each file indicates the last time the file was modified. Entries in the PDB archive have been processed by the three members of the wwPDB (RCSB, MSD-EBI, and PDBj).

Website Statistics

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

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

	DAILY AVERAGE		MONTHLY TOTALS			
MONTH	HITS	FILES	SITES	KBYTES	FILES	HITS
Mar 05	290,914	212,084	144,602	273,109,948	6,362,527	8,727,441
Feb 05	286,463	209,274	127,832	215,481,454	5,650,399	7,734,506
Jan 05	267,916	198,327	122,062	212,312,171	6,148,162	8,305,402

Outreach and Education

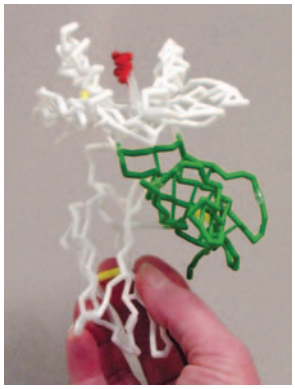
Status Report on "Large Macromolecular Complexes in the Protein Data Bank" Published

An overview of the large complexes in the PDB and some of the challenges experienced in their representation, visualization, analysis, and archiving has been published.

Large Macromolecular Complexes in the Protein Data Bank: A Status Report
Shuchismita Dutta and Helen M. Berman
Structure, Vol 13, 381-388, March 2005

www.structure.org/cgi/content/abstract/13/3/381/

Physical Model of February's Molecule of the Month - Major Histocompatibility Complex - Available on Loan to Educators



As part of a project by Molecule of the Month author David S. Goodsell (The Scripps Institute) and Tim Herman (Center for BioMolecular Modeling), a physical model of Class I MHC is available on loan to educators. This study will explore how physical models can be used to enhance the value of the online Molecule of the Month teaching resource. The models may be borrowed for a period of two weeks. The only cost involved is for return shipping. Further information about this opportunity is available at **www.rpc.msoe.edu/cbm/borrow_mhc.php**.

Biophysical Society's Annual Meeting

Lei Xie prepares to demonstrate the beta site at the Biophysical Society Meeting

RCSB PDB exhibited at the 49th Annual Meeting of the Biophysical Society (February 12-16, 2005) in Long Beach, California. Demonstrations of the beta site were presented.

Our Best Wishes to Gary

After working for almost 20 years at CARB and NIST, and serving as a co-director of the RCSB PDB since 1998, Gary L. Gilliland has taken a position with Centocor (a subsidiary of Johnson and Johnson) in Pennsylvania.

Under Gary's leadership, the RCSB PDB team at CARB/NIST was responsible for cataloging, archiving, and preserving the legacy PDB materials that have been accumulating since the PDB was established in 1971 at Brookhaven National Laboratory. This team also distributed data CDs of PDB files worldwide and maintained an exact copy of the RCSB PDB production site which served as both a mirror site and a fail-over system. The functions formerly performed at the CARB/NIST site will now be assumed by the groups at Rutgers and UCSD/SDSC.

The transition has now been successfully completed. The entire RCSB PDB team thanks Gary and wishes him the best of luck in his new adventures in the world of macromolecular structural biology.

PDB Education Corner: A short history of visualizing structures in the PDB by Judith Voet, J. H. Hammons Professor, Swarthmore College

The Protein Data Bank is now 34 years old. The last **PDB Education Corner** article highlighted a mural depicting some of the first protein structures to be determined (Winter 2005, Issue 24). The study of protein structure and function is filled with such images, and the use of computer graphics software has become vital for their visualization and analysis. Before the advent of computer graphics, building a 3D model of a protein was such an intensive effort that it would often take a class the whole semester to complete. The use of computers sophisticated enough for such visualization and analysis by the general educational community is a relatively recent phenomenon, although it is now pervasive. PDB Education Corner articles have described their use in courses ranging from high school to graduate school. But it wasn't always that way. To get an idea of just how recent this phenomenon is, I thought I'd reminisce a bit on my own introduction to the world of molecular graphics and the PDB.

I have been surrounded by computers for my whole professional life, but only succeeded in feeling comfortable with a computer when the Macintosh, with its miraculous mouse, arrived at Swarthmore College (and on my desk at home) in 1983. I was never able to remember the keyboard commands required for almost any action before the advent of the menu from which to select my choice. The first draft of the first edition of my textbook *Biochemistry*, co-authored with Donald Voet, was hand-written and then typed into a mainframe computer by an assistant using a dumb terminal (word processing programs had yet to be invented). When I got my first Macintosh, the whole world opened up for me – well, almost. Writing by hand with its difficult revision process was now a thing of the past, but I was still unable to visualize macromolecules on my computer. The molecular graphics software and color monitors necessary for the visualization of a PDB file still only operated on specialized and expensive graphics computers. At this time, there were around 200 structures available. This small number of structures, combined with the difficulty and expense of obtaining a molecular graphics computer, kept

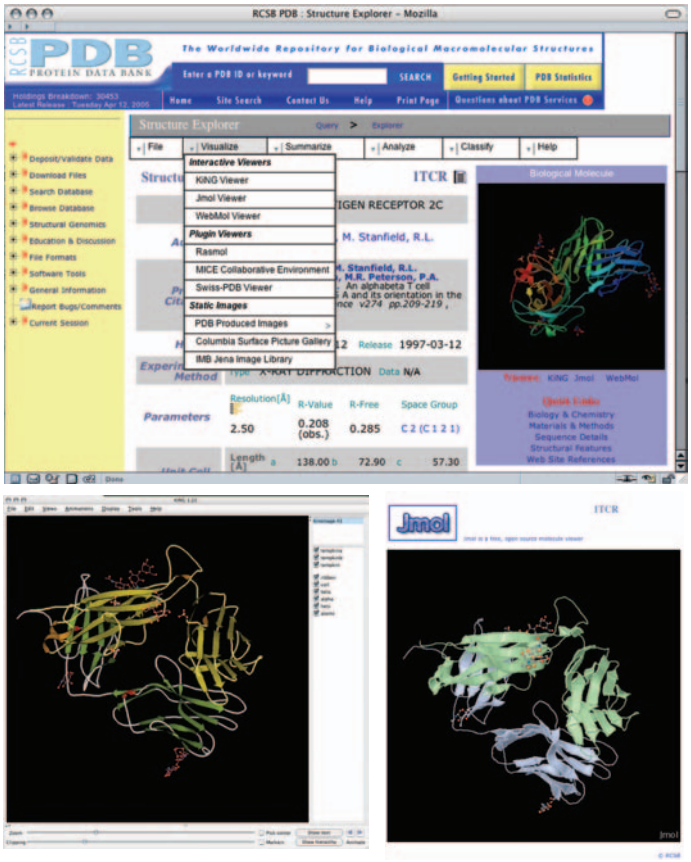
the PDB out of the hands of the general education community. By 1990, when the first edition of *Biochemistry* was published, structure visualization still required more sophisticated and expensive computers than were available for purely educational purposes. If we wanted to use a figure of a molecular structure that we had seen in a journal article in the textbook, we most often obtained it from the author. I remember our excitement when we had one structure generated for us by a colleague with a molecular graphics computer.

In 1993, while working with Joel Sussman at the Weizmann Institute in Rehovot, Israel, I had access to a molecular graphics computer for the first time and began to see the power of the PDB, not just as a storage facility for 3D coordinates, but as an interactive tool for search and discovery. At about the same time, Michael Levitt created a piece of software, MacImdad®, that allowed PDB files to be displayed and manipulated on a Macintosh computer. As it happened, at that time Levitt had a joint appointment at Stanford University and at the Weizmann Institute. For most of the time I was in Israel, he was in the United States, and I was assigned his desk and his Mac. The connection was too much to be ignored. I became a disciple of MacImdad. In 1994, on my return from Israel, I obtained the program for Swarthmore College, installed it on several Macs in the biochemistry laboratory and my students began to use computers to visualize protein structures and reaction mechanisms for the first time. Still, it was not easy to obtain the original 3D coordinates from the PDB. Believe it or not, there was not yet an easily accessible internet! Transfer of PDB files required access to an ftp site, still the province only of mainframe computers. The PDB would provide compact disks containing their data on request, but there was an energy barrier to obtaining this data. If you had a specific protein you wanted to study, most likely you wrote to the researcher who had determined the structure and asked him/her to send you the coordinates. MacImdad was wonderful because it came with all the data contained in the PDB at that time, in a compressed form accessible by the program.

Also in the same timeframe, Jane and David Richardson developed another Macintosh-friendly format for visualizing molecular structures–the Kinemage (short for Kinetic Image). Kinemages are displayed using the freely available program MAGE that is usable on most computer platforms. The Richardsons created many Kinemages that were extremely valuable for educational use in visualizing and understanding macromolecular structure and function. I availed myself of several kinemages and of the MAGE program for use in the biochemistry laboratory.

Due to the generosity of the Richardsons, John Wiley & Sons produced and distributed a supplementary disk to accompany the second edition of *Biochemistry* (1995) that contained many Kinemages that we generated to help visualize and interpret the molecular structures contained in the text. The creation of each Kinemage required the use of PDB files of the 3D coordinates of the macromolecule under study.

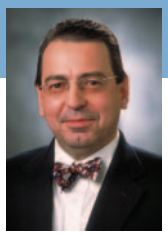
A third addition to my accessible molecular visualization software list was RasMol®, developed by Roger Sayle at MDL® and donated to the scientific community.



*Visualization 2005: Options available at the PDB Beta Site include KiNG (**kinemage.biochem.duke.edu**) and Jmol (**jmol.sourceforge.net**). Structure shown is from March's Molecule of the Month: T-Cell Receptor*

By 1998, I was back in Joel Sussman's laboratory and working on *Fundamentals of Biochemistry* (D. Voet, J.G. Voet and C. Pratt), the internet had made its phenomenal appearance, and we found a great web-based tool for macromolecular visualization: MDL Chime (**www.mdl.com**). This plug-in was based on RasMol, but ran directly on a web browser. Eric Martz created a website and many support documents to guide educators in its use for the development of visualization exercises for students, and also developed the very powerful molecular visualization tool Protein Explorer (**www.umass.edu/microbio/rasmol/**). At the time of this writing, MDL is no longer supporting Chime, but a Java-based, platform-independent successor, Jmol, promises to be even better. Kinemages can now be manipulated directly on the Web using KiNG.

The ability to visualize and manipulate macromolecules using relatively low level computers is now pervasive and included in most biochemistry classes as a matter of course. Most textbooks come with ancillary materials that allow students to examine and manipulate molecules using resources available at the publishers' websites. This growth in visualization capability would not be nearly as much use without the parallel growth in the ease of use and availability of the PDB. Hand in hand these tools and resources have developed a whole new way of teaching and learning about molecular structure that was only a dream 25 years ago.



PDB Community Focus: Stephen K. Burley, Structual GenomiX

Stephen K. Burley (M.D., D.Phil., F.R.S.C.) is the chief scientific officer of Structural GenomiX, Inc. (SGX; www.stromix.com), located in San Diego, California. SGX is an oncology focused drug discovery and development company, with Troxatyl® in clinical trials and multiple protein kinase inhibitors in preclinical development. Prior to joining SGX, he was the Richard M. and Isabel P. Furlaud professor and chief academic officer at The Rockefeller University, and a full investigator in the Howard Hughes Medical Institute. Burley received an M.D. degree from Harvard Medical School in the joint Harvard-MIT Health Sciences and Technology program and, as a Rhodes Scholar, received a D.Phil. in Molecular Biophysics from Oxford University. He trained in internal medicine at the Brigham and Women's Hospital, and did post-doctoral work with Gregory A. Petsko at the Massachusetts Institute of Technology and William N. Lipscomb at Harvard University. With William J. Rutter and others at the University of California, San Francisco and The Rockefeller University, Burley co-founded Prospect Genomics, Inc., which was subsequently acquired by SGX. He is a fellow of the Royal Society of Canada and of the New York Academy of Sciences.

Q What made you decide to leave academia for a position in a biotechology firm?

A Although my much publicized departure from The Rockefeller University and the Howard Hughes Medical Institute was initially greeted with surprise, one of my closest colleagues told me that on reflection he realized it was inevitable. He was referring to my training in both structural biology and medicine, followed by internship and residency in internal medicine. He was also reflecting on my long-standing commitment to exploring new areas of science and technology. In short, I value the breadth of my training and am continually looking for new opportunities to learn.

The advent of high-throughput X-ray crystallography and improvements in computational chemistry convinced me that a leadership role in the right industrial environment would allow me to help change the way in which new drugs are discovered. For much of the late 1990s, I wondered which company and when. My involvement in Structural GenomiX, Inc. (SGX) as a member of the scientific advisory board and my role as a co-founder of Prospect Genomics, Inc. (an in silico drug discovery company) made it clear that the time was ripe for such a move. Shortly after SGX acquired Prospect Genomics, I volunteered myself for the post of Chief Scientific Officer.

Three plus years later there have been many changes, both personal and professional. I left the excitement of New York City for southern California, where my family and I live in idyllic surroundings complete with palm trees, the Pacific ocean, and as much fly fishing as I can fit in to an extremely busy professional schedule. At SGX, I have helped guide our transition from a gene-to-structure platform company to an oncology drug discovery and development company. We have combined high-throughput X-ray crystallography with combinatorial synthesis and state-of-the-art computational chemistry tools to create a best-in-class lead discovery engine that is delivering drug development candidates targeted at protein kinases that cause cancer. On the clinical front, SGX now has a compound in clinical trials for acute myeloid leukemia and various solid tumor malignancies.

Q Chairing the RCSB PDB Advisory Committee requires a lot of time and energy – what is your interest in this board?

A My involvement with the PDB coincided with its transition from Brookhaven National Laboratory to the RCSB, when Helen Berman asked me to chair the advisory committee. Unlike some invitations to do committee work, this one was easy to accept. The PDB is a mission critical global archive on which all of biomedical research relies. As an industrial scientist, I also appreciate the enormous economic benefit that the PDB provides to pharmaceutical and biotechnology companies. Without facile access to this comprehensive, single archive of experimentally-derived three-dimensional structures of biological macromolecules, our drug discovery activities would be considerably handicapped. Finally, the PDB represents a vitally important intellectual and educational resource. During the next decade, I foresee that the wealth of structures in the PDB will play a decisive role in integrating our understanding of distinct chemical and signaling steps as we build up "systems biology" views of how cells and organ systems work at the molecular level.

I would add that the quality of the RCSB leadership and the committee makes my job as chair of the RCSB PDB Advisory Committee both interesting and fun. My fellow committee members were recruited from among the top ranks of industrial and academic scientists. Working closely with Helen, John Westbrook, and Phil Bourne and the advisory committee, we have forged a strong partnership that provides the best possible advice to the RCSB and serves as a vocal advocate for PDB user community.

Q You are also a member of the advisory committee of the wwPDB – how do you see the future of this organization?

A In my answer to your previous question, I described the PDB as a global archive. I view the PDB as a resource that must grow and develop for the collective benefit of all humanity. I share Helen's vision for the wwPDB, and feel honored to be playing an advisory role now that the RCSB, the MSD, and PDBj are working closely together to make it happen. As the keeper of the single, global archive, the RCSB and the US government funding agencies together have a central role that brings with it a special responsibility to ensure the security and stability of the archive. I believe that the wwPDB advisory committee provides an important venue to help ensure that the wwPDB membership effectively coordinates effort and funding from different parts of the globe.

Q There are now ~30,000 structures in the PDB. How many more structures do you think there would be if the drug companies deposited all the structures they have been working on? Should the drug companies be encouraged to deposit these structures?

A Extrapolating from structure determination efforts at SGX over the past five years, I suspect that there are well in excess of ten

thousand protein structures in the proprietary databases of biotech and pharma companies world wide. Even before my move to industry, I was a strong advocate of including depositions from drug companies in the PDB. These proprietary databases contain a wealth of structures illuminating how protein targets recognize small molecule ligands. If present in the PDB, I believe that this enormous body of information could be brought together and used to synthesize a more predictive, quantitative understanding of protein-drug interactions that would serve to accelerate drug discovery activities in both academe and industry. Companies are being encouraged to deposit their structures, and many do so to a limited extent. I am aware of three impediments to increasing the number of such depositions.

First, there are understandable intellectual property concerns relating to proprietary small molecules that can only be resolved by deferring depositions until the necessary composition-of-matter patents pertaining to the small molecules have been issued. Once full disclosure has occurred, the structures can and should be included in the PDB.

Second, there are practical limits to resources that company scientists can devote to depositing structures in the PDB. The business of these companies is drug discovery and development, not publications and PDB depositions. The RCSB is committed to helping companies get their data into the PDB with as little effort as possible. Working together, SGX and the RCSB have developed a facile method for porting large numbers of structures from our company database to the PDB. I am optimistic that this advance will be used by other companies.

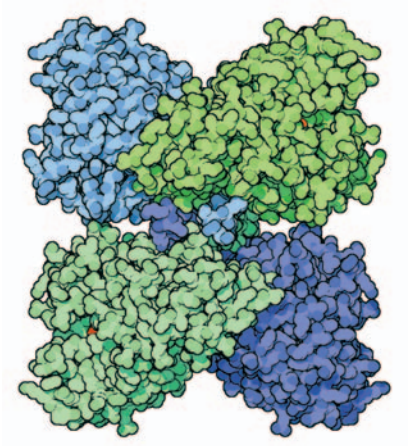
Finally, some companies have taken the position that de novo protein structures should be kept confidential to avoid enabling competitors. From my experience at SGX, I can tell you that this strategy rarely yields a sustainable competitive advantage. We routinely determine structures that others companies are holding in confidence, thereby leveling the playing field. I would like to see biotechnology and pharmaceutical companies come to a common recognition that de novo protein structures represent pre-competitive information that should be shared, just as with single nucleotide polymorphism data. Structural genomics programs in the US, Europe, and Japan should help the industry come to this view sooner rather than later. With the enormous flood of new structures coming into the PDB over the next decade, pre-competitive structural information will soon be freely available for many drug discovery targets. Once the deluge starts, companies will have nothing to gain by keeping their de novo structures confidential and much publicity value to be had by depositing them to the PDB.

Molecules of the Quarter: Phenylalanine Hydroxylase, Major Histocompatibility Complex, T-Cell Receptor

The *Molecule of the Month* series by David S. Goodsell explores the functions and significance of selected biological macromolecules for a general audience. The full Molecule of the Month features are available from **www.rcsb.org/pdb/molecules/molecule_list.html**

January 2005 — *Phenylalanine Hydroxylase*

Four molecules of phenylalanine hydroxylase interact to form a tetramer, which is the functional unit for this enzyme. Each molecule in the tetramer is organized into three domains: a regulatory domain, a catalytic domain where the enzyme activity resides and a tetramerization domain that assembles four chains into the tetramer. At the heart of each catalytic domain is an iron ion that plays an important role in the enzyme action. A model structure of the complete enzyme tetramer is shown here. This is composed of two PDB files: **2pah**, which includes the structure of the catalytic and tetramerization domains of the enzyme, and **1phz**, which includes the regulatory domain flexibly attached to the catalytic domain.



From January's Molecule of the Month. **2pah** *(Fusetti, F., Erlandsen, H., Flatmark, T., Stevens, R. C.: Structure of Tetrameric Human Phenylalanine Hydroxylase and its Implications for Phenylketonuria* J.Biol.Chem. *273: 16962 (1998) &* **1phz** *Kobe, B., Jennings, I. G., House, C. M., Michell, B. J., Goodwill, K. E., Santarsiero, B. D., Stevens, R. C., Cotton, R. G., Kemp, B. E.: Structural basis of autoregulation of phenylalanine hydroxylase.* Nat Struct Biol *6: 442 (1999)*

February 2005 — *Major Histocompatibility Complex*

Viruses are insidious enemies, so we must have numerous defenses against them. Antibodies are our first line of defense. Antibodies bind to viruses, mobilizing blood cells to destroy them. But what happens if viruses slip past this defense and get inside a cell? Then, antibodies have no way of finding them and the viruses are safe...but not quite.

Each cell has a second line of defense that it uses to signal the immune system when something goes wrong inside. Cells continually break apart a few of their old, obsolete proteins and display the pieces on their surfaces. The small peptides are held in the Major Histocompatibility Complex (MHC), which grips the peptides and allows the immune system to examine them. In this way, the immune system can monitor what is going on inside the cell. If all the peptides displayed on the cell surface are normal, the immune system leaves the cell alone. But if there is a virus multiplying inside the cell, many of the MHC molecules carry unusual peptides from viral proteins, and the immune system kills the cell.

March 2005 — *T-Cell Receptor*

T-cell receptors are similar to one arm of an antibody. Like antibodies, they are composed of two chains. The binding site is at the tip of the molecule, and is formed by several loops of the protein chains. The amino acids in these loops are very different in different T-cell receptors, so they are often called hypervariable loops. Each chain also includes a segment at one end that crosses through the membrane, connecting the receptor to the cell surface.

RCSB PDB Partners

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

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

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

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

Mirror Sites

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

Statement of Support

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

RCSB PDB Leadership Team

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

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

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

Allison Clarke, Operations Coordinator
Rutgers, The State University of New Jersey
aclarke@rcsb.rutgers.edu

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

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

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

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

RCSB PROTEIN DATA BANK
www.pdb.org

9650 Rockville Pike
Bethesda, MD 20814

Return Service Requested