

# RCSB PDB NEWSLETTER

NUMBER 22 ♦ SUMMER 2004 ♦ 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](http://www.rcsb.org/pdb/latest_news.html)

Links to RCSB PDB newsletters are available at [www.rcsb.org/pdb/newsletter.html](http://www.rcsb.org/pdb/newsletter.html)

## CONTENTS

MESSAGE FROM THE RCSB PDB.....	1
DATA DEPOSITION AND PROCESSING	
PDB_EXTRACT Integrated into	
CCP4 Program Suite .....	2
PDB Deposition Statistics.....	2
DATA QUERY, REPORTING, AND ACCESS	
FASTA Sequence Files on RCSB PDB FTP Archives .....	2
Batch File Download Script Now Available .....	2
Website Statistics .....	2
RCSB PDB OUTREACH AND EDUCATION	
Three RCSB PDB Papers Published: TargetDB,	
Ligand Depot, and Lessons in Data Management .....	3
RCSB Poster Prize Awarded at RECOMB .....	3
RCSB Exhibits: ENC, RECOMB, and NSTA .....	3
CD-ROM Subscriber Questionnaire Results Released .....	3
New CD-ROM Update Release.....	3
Molecules of the Quarter:	
Growth Hormone, Serpins, Acetylcholinesterase .....	3
PDB COMMUNITY FOCUS: KIM HENRICK, E-MSD .....	5
PDB EDUCATION CORNER:	
USING VISUALIZATION TOOLS, BY J. RICKY COX (MSU)...	6
RELATED LINKS: NMR RESOURCES .....	7
RCSB PDB LEADERSHIP TEAM, RCSB PDB MEMBERS,	
AND STATEMENT OF SUPPORT .....	8

## SNAPSHOT: JULY 1, 2004

26,144 released atomic coordinate entries

MOLECULE TYPE	EXPERIMENTAL TECHNIQUE
23,676 proteins, peptides, and viruses	22,306 diffraction and other
1,338 nucleic acids	13,216 structure factor files
1,112 protein/nucleic acid complexes	3,838 NMR
18 carbohydrates	1,964 NMR restraint files

## PARTICIPATING RCSB MEMBERS

RUTGERS: [rutgers.rcsb.org](http://rutgers.rcsb.org)

SDSC/UCSD: [www.pdb.org](http://www.pdb.org)

CARB/NIST: [nist.rcsb.org](http://nist.rcsb.org)

E-MAIL: [info@rcsb.org](mailto:info@rcsb.org)

FTP: [ftp.rcsb.org](ftp://ftp.rcsb.org)

*The RCSB PDB is a member of the wwPDB  
([www.wwpdb.org](http://www.wwpdb.org))*

## MESSAGE FROM THE RCSB PDB

The RCSB PDB is dedicated to providing resources that utilize the data contained in the PDB archive. To this end, we invite our users to participate in the beta test of the re-engineered RCSB PDB site and database. The site was developed using feedback derived from the PDB help desk, conference attendance, focus groups and other personal interactions between the users of the PDB and RCSB staff.

The new system has been designed using an Enterprise Java framework and is based on a three-tier model—an underlying database, a pres-

The screenshot shows the RCSB PDB website interface. At the top, it says "Protein Data Bank - Mozilla". Below that is the RCSB PDB logo and the tagline "The Worldwide Repository for Biological Macromolecular Structures". There is a search bar with the text "Enter a PDB ID or keyword" and buttons for "SEARCH", "Getting Started", and "PDB Statistics". A navigation menu includes "Home", "Site Search", "Contact Us", "Help", "Print Page", and "Questions about PDB Services". A "Quick Links" section lists various services like "Deposit/Validate Data", "Download Files", "Search Database", "Browse Database", "Structural Genomics", "Education & Discussion", "File Formats", "Software Tools", and "General Information". The "News" section features a headline: "06-July-2004 Come visit us in 'Data Alley' -- RCSB PDB Demonstrations, Poster Prizes, Workshop, and More at the ACA Annual Meeting July 17-22 in Chicago, IL." Below the news is a molecular structure image of DNA Ligase. There are also sections for "VIEW BETA SITE UPDATES" and "WELCOME TO THE NEW BETA SITE OF THE RCSB PDB".

The beta RCSB PDB site at [pdbbeta.rcsb.org](http://pdbbeta.rcsb.org). Comments about the website should be sent to [betafeedback@rcsb.org](mailto:betafeedback@rcsb.org).

entation layer, and a middle tier connecting them. The underlying database consists of curated mmCIF files resulting from the data uniformity project, which will allow improved query access to the unified data.

The beta test site will be available at [pdbbeta.rcsb.org](http://pdbbeta.rcsb.org) alongside the current production site. Both sites will be updated regularly.

The RCSB is very excited about this resource, and encourage the PDB community to test the new site. We look forward to your feedback.

**The RCSB PDB ♦**

## MIRROR SITES

Cambridge Crystallographic Data Centre (UK): [pdb.ccdc.cam.ac.uk](http://pdb.ccdc.cam.ac.uk)

National University of Singapore: [pdb.bic.nus.edu.sg](http://pdb.bic.nus.edu.sg)

Osaka University (Japan): [pdb.protein.osaka-u.ac.jp](http://pdb.protein.osaka-u.ac.jp)

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

RCSB PDB Beta test site: [pdbbeta.rcsb.org](http://pdbbeta.rcsb.org)

## DATA DEPOSITION AND PROCESSING

### *PDB\_EXTRACT Integrated into CCP4 Program Suite*

**P**DB\_EXTRACT is a program suite developed by the RCSB that contains tools and examples for extracting mmCIF data from structure determination applications. It has been integrated with the CCP4i interface of the CCP4 program suite for protein crystallography (version 5.0).



PDB\_EXTRACT is also available for download (both source and executable) from [deposit.pdb.org/mmcif/PDB\\_EXTRACT/index.html](http://deposit.pdb.org/mmcif/PDB_EXTRACT/index.html).

PDB\_EXTRACT extracts information about data processing, heavy atom phasing, molecular replacement, density modification, and final structure refinement from the output files produced by many X-ray crystallographic applications. The program merges these data into macromolecular Crystallographic Information File (mmCIF) data files that can be used with ADIT to perform validation and to add any additional information for PDB deposition.

PDB\_EXTRACT can also be used online at [pdb-extract.rutgers.edu](http://pdb-extract.rutgers.edu).

### *PDB Deposition Statistics*

**I**n the first half of 2004, approximately 2726 structures were deposited in the PDB archive. Of the structures received, 79% were deposited with a “hold until publication” release status; 13% with a “release immediately” status; and 8% with a specific release date.

83% of these entries were the result of X-ray crystallographic experiments; 14% were determined by NMR methods.

## DATA QUERY, REPORTING, AND ACCESS

### *FASTA Sequence Files on RCSB PDB FTP Archives*

**T**he RCSB PDB maintains several FASTA formatted sequence files on the FTP archives. The sequences for all currently released experimental structures are contained in `pdb_seqres.txt`, available in uncompressed form at [ftp://ftp.rcsb.org/pub/pdb/derived\\_data/pdb\\_seqres.txt](http://ftp.rcsb.org/pub/pdb/derived_data/pdb_seqres.txt) and in Unix compressed (“Z”) format at [ftp://ftp.rcsb.org/pub/pdb/derived\\_data/pdb\\_seqres.txt.Z](http://ftp.rcsb.org/pub/pdb/derived_data/pdb_seqres.txt.Z). These two files contain all sequences for structures queried on the Home Page, QuickSearch, SearchLite, and SearchFields.

PDB depositors are given the opportunity to prerelease the sequences of their structures before releasing the coordinate data.

Prereleased sequences for unreleased structures are contained in the separate file `pre-released.seq`, available in uncompressed form at [ftp://ftp.rcsb.org/pub/pdb/derived\\_data/index/pre-released.seq](http://ftp.rcsb.org/pub/pdb/derived_data/index/pre-released.seq). Unreleased structures can be queried on the Status Query page.

### *Batch File Download Script Now Available*

**A** script to download large numbers of files from the PDB FTP site is now available at [ftp://ftp.rcsb.org/pub/pdb/software/getPdbStructures.pl](http://ftp.rcsb.org/pub/pdb/software/getPdbStructures.pl). This simple Perl script can be run locally to download files from a user’s list of PDB IDs. Options are available to download coordinate files in either PDB or mmCIF format, as well as experimental data files. The script creates a directory structure for the downloaded files. Further details regarding usage of this script can be found at [ftp://ftp.rcsb.org/pub/pdb/software/getPdbStructures.html](http://ftp.rcsb.org/pub/pdb/software/getPdbStructures.html).

### *Website Statistics*

The RCSB PDB is available from several Web and FTP sites located around the world. Users are also invited to preview the newly reengineered RCSB website at [pdbbeta.rcsb.org](http://pdbbeta.rcsb.org).

The access statistics are given below for the primary RCSB PDB website at [www.pdb.org](http://www.pdb.org).

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

MONTH	DAILY AVERAGE			MONTHLY TOTALS		
	HITS	FILES	SITES	KBYTES	FILES	HITS
Apr 04	270,777	199,230	115,792	246,788,624	5,777,687	7,852,545
May 04	226,369	168,320	101,883	212,372,632	5,049,625	6,791,070
Jun 04	218,377	161,922	88,025	228,966,379	4,695,739	6,332,951

## RCSB PDB OUTREACH AND EDUCATION

### *Three RCSB PDB Papers Published: TargetDB, Ligand Depot, and Lessons in Data Management*

**T**wo papers have been published online that describe the TargetDB and Ligand Depot resources. TargetDB is a centralized target registration database that includes protein target data from the NIH structural genomics centers and a number of international sites ([targetdb.pdb.org](http://targetdb.pdb.org)). Ligand Depot is an integrated data resource for finding information about small molecules bound to proteins and nucleic acids ([ligand-depot.rutgers.edu](http://ligand-depot.rutgers.edu)).

**TARGETDB:** a target registration database for structural genomics projects (*Li Chen, Rose Oughtred, Helen M. Berman, and John Westbrook*) **BIOINFORMATICS:** [bioinformatics.oupjournals.org/cgi/content/abstract/bth300v1](http://bioinformatics.oupjournals.org/cgi/content/abstract/bth300v1)

**LIGAND DEPOT:** a data warehouse for ligands bound to macromolecules (*Zukang Feng, Li Chen, Himabindu Maddula, Ozgur Akcan, Rose Oughtred, Helen M. Berman, and John Westbrook*) **BIOINFORMATICS:** [bioinformatics.oupjournals.org/cgi/content/abstract/bth214v1](http://bioinformatics.oupjournals.org/cgi/content/abstract/bth214v1)

The history of the PDB is used to highlight practices important to developers of current biological databases in an article published in the journal *Briefings in Bioinformatics*. The role of the

“human factor” in the form of users, collaborators, scientific societies, and *ad hoc* committees is also included.

Philip E. Bourne, John Westbrook, Helen M. Berman The Protein Data Bank and lessons in data management (2004) *Briefings in Bioinformatics* ([www.henrystewart.com/journals/bib/](http://www.henrystewart.com/journals/bib/)) 5, pp 23-30.

### RCSB Poster Prize Awarded at RECOMB

**T**hanks to the students and judges who participated in the competition for the best student poster presentation in the category of “Protein Structure” at the Eighth Annual International Conference on Research in Computational Molecular Biology (RECOMB 2004; March 27-31, San Diego, CA).

The RCSB PDB Poster Prize was awarded to Boris E. Shakhnovich for the poster “Protein Structure and Evolutionary History Determine Sequence Space Topology” (Boris E. Shakhnovich<sup>1</sup>, Eric Deeds<sup>2</sup>, Charles Delisi<sup>1</sup>, and Eugene I. Shakhnovich<sup>3</sup>)

Information about the 2004 RCSB PDB Poster Prize is at [www.rcsb.org/pdb/poster\\_prize.html](http://www.rcsb.org/pdb/poster_prize.html).

### RCSB Exhibits: Experimental Nuclear Magnetic Resonance Conference, RECOMB, and NSTA

**T**he RCSB PDB exhibited with a tabletop display at the 45th Experimental Nuclear Magnetic Resonance Conference (ENC) that was held April 18-23 at the Asilomar Conference Center (Pacific Grove, CA).

At RECOMB, the RCSB PDB demonstrated the reengineered website in addition to awarding the RCSB PDB Poster Prize award.

The RCSB PDB participated at the National Science Teachers Association (NSTA) National Convention April 1-4, 2004 in Atlanta, GA. Molecule of the Month materials were made available at the MSOE Center for BioMolecular Modeling’s exhibit booth. SMART teams (see Education Corner from Spring 2004) presented their models built from PDB coordinates at the meeting.



The SMART team from Kettle Moraine High School (Wales, WI) presented the models they created using data from the PDB at the NSTA meeting.

### CD-ROM Subscriber Questionnaire Results Released

**A** short online questionnaire was emailed to 1450 RCSB PDB CD-ROM subscribers in February 2004. The purpose of this instrument was to get to know subscribers, ask their opinion on how well we

are doing, and gauge interest in a DVD data product.

The questionnaire had a 23% response rate with an approval rating of 84.5%. Many respondents provided additional information in the comments section and took additional time to write us separately.

Problems are being addressed immediately. Suggestions for improvements are being reviewed and considered by RCSB PDB staff. Suggested improvements already in development include a top level index and a DVD data product. The results of the questionnaire have been compiled and are detailed in the CD-ROM Subscriber Questionnaire Results report ([www.rcsb.org/pdb/2004CDUserReport.html](http://www.rcsb.org/pdb/2004CDUserReport.html)).

### New CD-ROM Update Release

Two products were distributed for the April 2004 CD-ROM. Release 108U contained the incremental set of experimentally determined structures (1,490) and models (84) deposited between January 1, 2004 and April 1, 2004, on a single CD-ROM disk. Release 108U-EXP contains the experimental data, both X-ray structure factors (865) and NMR constraints (83), deposited between January 1, 2004 and April 1, 2004, on a single CD-ROM disk. Questions should be directed to [pdbrcd@rcsb.org](mailto:pdbrcd@rcsb.org). Ordering information is available at [www.rcsb.org/pdb/cdrom.html](http://www.rcsb.org/pdb/cdrom.html).

### PDB Molecules of the Quarter: Growth Hormone, Serpins, Acetylcholinesterase

**T**he *Molecule of the Month* series, by David S. Goodsell, explores the functions and significance of selected biological macromolecules ([www.rcsb.org/pdb/molecules/molecule\\_list.html](http://www.rcsb.org/pdb/molecules/molecule_list.html)). Structures highlighted during this past quarter were:

#### Growth Hormone

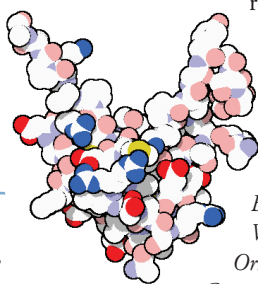
**APRIL 2004**—As children grow, their height, weight and strength increase. Numerous factors influence this growth, including the genetic makeup of the child, nutrition and environmental factors. Specific messengers released by the body also stimulate and regulate growth. Growth hormone is one key growth signal released from the pituitary, a pea-sized gland located at the base of the brain. Lack of this hormone in children can cause them to remain shorter than average, while in its excess they may grow taller than most. Growth hormone continues its work in adults, playing an important role in repair and maintenance of different tissues in the body.

The pituitary releases several hormones including growth hormone, prolactin and placental lactogen. These small protein hormones are similar in their sequence and structure and play crucial roles in growth, development and milk production.

Growth hormone travels through the blood and stimulates the liver to produce a protein called insulin-like growth factor (**IGF-1**), as in PDB entry

PDB ID **1h02**.

Brzozowski, A., Dodson, E., Dodson, G., Murshudov, G., Verma, C., Turkenburg, J., De Bree, F., Dauter, Z.: *Structural Origins of the Functional Divergence of Human Insulin-Like Growth Factor-I and Insulin* *Biochemistry* 41 pp. 9389 (2002)



<sup>1</sup> Bioinformatics Program, Boston University

<sup>2</sup> Department of Molecular and Cellular Biology, Harvard University

<sup>3</sup> Department of Chemistry and Chemical Biology, Harvard University



**1h02. IGF-1** helps the cartilage cells located at the ends of long bones to multiply. In children, this leads to growth in the length of the bones and increases the child's height. By puberty, however, the cartilage at the ends of most long bones is converted to bone and subsequent action of growth hormone or **IGF-1** usually cannot increase their length. **IGF-1** also acts on immature muscle cells to increase muscle mass. Aside from these growth stimulating functions, growth hormone participates in regulating the body's metabolism. It acts on fat cells to reduce the amount of stored fats, promotes protein synthesis in cells and plays a role in regulating the sugar levels in the blood. Thus growth hormone has multiple effects on the overall form and function of a growing body.

For more information on growth hormone, by Shuchismita Dutta and David S. Goodsell, see [www.rcsb.org/pdb/molecules/pdb52\\_1.html](http://www.rcsb.org/pdb/molecules/pdb52_1.html).

### Serpins

**MAY 2004**—Our cells are often forced to work with dangerous machinery. For instance, cells build many machines for demolition, such as nucleases that break down DNA and RNA, amylases and related enzymes that break down carbohydrates, lipases that chew up lipids, and proteases that disassemble proteins. These destructive enzymes are needed in many capacities. They are used in digestion to break food molecules into workable pieces. They are used in defense to attack invading viruses and bacteria. They are used to break down defective or obsolete molecules inside cells. They are also used in signaling cascades to activate signaling molecules instantly when a message is received. These enzymes are essential when used at the proper place and time, but can spell disaster if they get loose.

To control these destructive machines, our cells also build a host of proteins that block their action and neutralize the danger. The serpins are one class of these molecules, designed to seek out and destroy specific serine proteases. The name serpin, although sounding like something from Greek mythology, is taken from their function: SERine Protease INhibitors. An example is alpha-antitrypsin, from PDB entry **1psi**. It is found in the bloodstream, where it protects the surrounding tissues from the protein-cutting enzyme elastase. Neutrophils (a type of

white blood cell) secrete elastase in sites of inflammation, where it breaks down connective tissue and allows blood cells to enter and do their jobs in defense and repair. The serpin protects the neighboring areas and ensures that the elastase doesn't spread throughout the body.

Over thirty different human serpins (a number of which are available in the PDB) have been studied, each with a different essential job. Many are found in the blood. Several control the process of blood clotting: antithrombin limits the action of thrombin when a clot is forming, and antiplasmin limits the action of plasmin when blood clots are being disassembled. Other serpins control the action of proteases used in the complement system, which protects us from bacterial infection.

For more information on serpins, see [www.rcsb.org/pdb/molecules/pdb53\\_1.html](http://www.rcsb.org/pdb/molecules/pdb53_1.html).

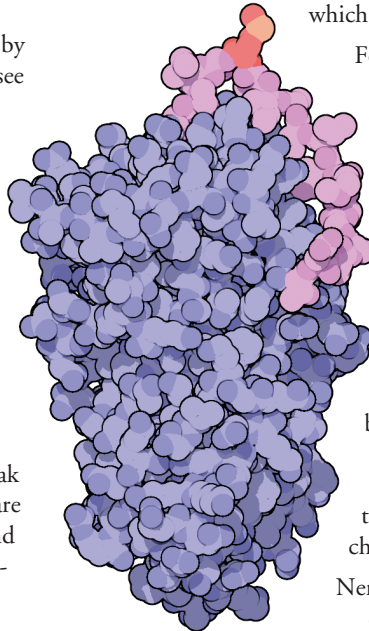
### Acetylcholinesterase

**JUNE 2004**—Every time you move a muscle and every time you think a thought, your nerve cells are hard at work. They are processing information: receiving signals, deciding what to do with them, and dispatching new messages off to their neighbors. Some nerve cells communicate directly with muscle cells, sending them the signal to contract. Other nerve cells are involved solely in the bureaucracy of information, spending their lives communicating only with other nerve cells. But unlike our human bureaucracies, this processing of information must be fast in order to keep up with the ever-changing demands of life.

Nerves communicate with one another and with muscle cells by using neurotransmitters. These are small molecules that are released from the nerve cell and rapidly diffuse to neighboring cells, stimulating a response once they arrive. Many different neurotransmitters are used for different jobs: glutamate excites nerves into action; GABA inhibits the passing of information; dopamine and serotonin are involved in the subtle messages of thought and cognition. The main job of the neurotransmitter acetylcholine is to carry the signal from nerve cells to muscle cells. When a motor nerve cell gets the proper signal from the nervous system, it releases acetylcholine into its synapses with muscle cells. There, acetylcholine opens receptors on the muscle cells, triggering the process of contraction. Of course, once the message

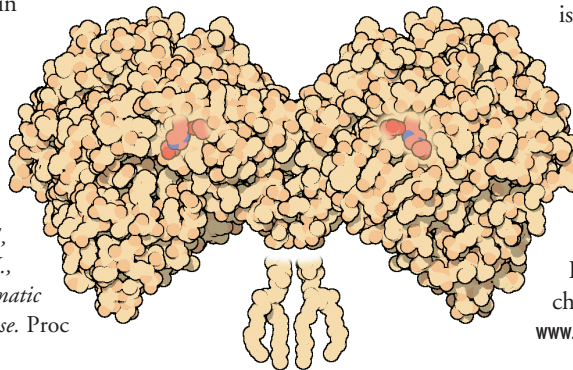
is passed, the neurotransmitter must be destroyed, otherwise later signals would get mixed up in a jumble of obsolete neurotransmitter molecules. The cleanup of old acetylcholine is the job of acetylcholinesterase.

For more information on acetylcholinesterase, see [www.rcsb.org/pdb/molecules/pdb54\\_1.html](http://www.rcsb.org/pdb/molecules/pdb54_1.html).



PDB ID **1psi**.

Elliott, P. R., Lomas, D. A., Carrell, R. W., Abrahams, J. P.: *Inhibitory conformation of the reactive loop of alpha 1-antitrypsin*. Nat Struct Biol 3 pp. 676 (1996)



### ACETYLCHOLINESTERASE

PDB ID **1acj**.

Harel, M., Schalk, I., Ehret-Sabatier, L., Bouet, F., Goeldner, M., Hirth, C., Axelsen, P. H., Silman, I., Sussman, J. L.: *Quaternary ligand binding to aromatic residues in the active-site gorge of acetylcholinesterase*. Proc Natl Acad Sci U S A 90 pp. 9031 (1993)



*Kim Henrick is currently the group leader of the European Bioinformatics Institute (EBI) Macromolecular Structure Database (E-MSD). He was born in Australia and received his PhD at the University of Western Australia in Crawley, Western Australia. He was a post-doctoral fellow at the Imperial College of Science and Technology in London, The Cambridge Centre for Protein Engineering in Cambridge, and the National Institute for Medical Research in London. Kim also spent some time working on the CCP4 project and as the senior scientific officer at the UK Science and Engineering Research Council (SERC) Laboratory in Daresbury. His interest in things structural has spanned the gamut from solving the structure of small organic molecules (reacting thionyl chloride with tricyclo*

*[3,2,1,0<sup>o</sup>,4]oct-6-en-ols) to developing Web-based systems for the deposition and validation of high-resolution electron microscopy macromolecular structure information. He has been with the European Bioinformatics Institute (EBI) since 1996, and became the head of the Macromolecular Structure Database (E-MSD) in 2001. Kim is also actively involved with the newly formed World Wide PDB (wwPDB).*

**Q** You started out doing small molecule crystallography and are now involved in the protein world. How and why did this transition take place for you?

**A** The institute where I was working in the 1980's was in financial difficulties so I took voluntary redundancy and moved into protein crystallography. This was initially in bewilderment coming from 2-theta as a measure of resolution to Ångstroms and to the realization that a B-factor in the protein world had no physical reality.

**Q** As macromolecular crystallography moves into the age of structural genomics do you see parallels with changes that have taken place in small molecule crystallography over the years?

**A** In 'small molecule crystallography' (what a dreadful expression) we had George Sheldrick. The release of SHELX76 changed the world in crystallography from struggling with programs with card image formats (SHELX76 was testing in 1975 – the same year Microsoft was founded by Bill Gates and Paul Allen). This was followed closely by Digital Equipment Corporation introducing the VAX 11/780 (1978). These two events made crystallography relatively easy. With George's SHELXTL, he put into a single package data reduction, phasing, refinement, graphics, and report generation. Protein crystallographers at present have a status that has been lost by most chemical crystallographers although both are collecting data with the same machines (cryo-cooling and image plates). Currently data collection and structure solution in chemistry is usually carried out by departmental service crystallographers. The speed of data collection and structure solution for protein ligand complexes or mutant analysis is now much the same as for chemical crystallography. If the various high throughput projects being undertaken, such as BIOXHIT in Europe, achieve their aims it should soon be possible for new protein structures to be solved at similar rates, (and now 'we' have George Sheldrick as well, especially with SHELXD and SHELXE).

**Q** What is the scope of the MSD project?

**A** The aim of the European Macromolecular Structure Database, MSD, is to support deposition of new structures determined by X-ray, NMR and cryo-3D-electron microscopy techniques, to provide storage and organization of the structural data, and to support search and analysis tools to query the data, i.e. exactly the same as the RCSB and PDBj.

**Q** What is the nature of your interactions with the RCSB PDB and PDBj?

**A** The members of the wwPDB have identical aims in managing, processing and providing publicly accessible structural data. Not just the best we can do but really in the right way. The groups consist of people dedicated to this end of providing a global service for 3D structure data and we each know our individual processes are the best approach - which leads to resolvable niggling tensions at times. However these separate developments give an invaluable means of cross checking PDB entries, which would otherwise be missed. The members work under pressure from scientists who want the data in a form they require now, while in part being unsympathetic to the standardization of their structures.

**Q** What do you think wwPDB should accomplish?

**A** The wwPDB should reach the situation where the PDB is recognized as a global resource and its role defined by the partners and their scientific advisory boards. This will take considerable effort from the partners who are each restrained by the specific demands of their funding agencies. However, the function of the PDB should be seen as not the sole prerogative of the USA funding agencies.

**Q** Where do you see structural biology evolving over the next decade and how will the PDB need to change to keep pace?

**A** Genome projects and sequence-oriented bioinformaticians see their work as dealing with molecular biology. However, on its own, the DNA sequence tells us little about what the genome does or how it works. DNA sequences code for proteins but proteins are only functional once they have folded up into a unique 3D structure. Similarities and differences found between gene sequences can only really be understood once the 3D structure of a member of a homologous family is known. The PDB currently handles 3D structure data for proteins from the experimental techniques of X-ray crystallography and NMR spectroscopy. Significant 3D structural information is given from cryo-3D-electron microscopy and tomographic reconstructions. The PDB perhaps should expand to include all 3D experimental data and concentrate on data integration with other biological databases.



*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 J. Ricky Cox, an associate professor of chemistry at Murray State University in Murray, KY.*

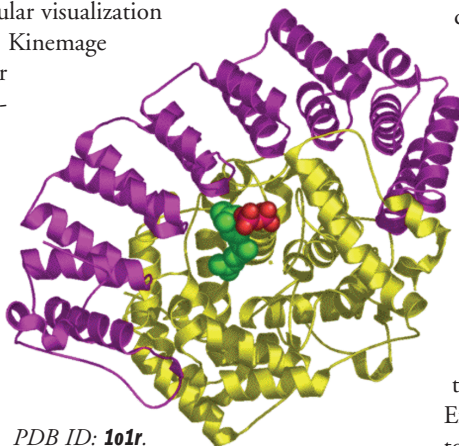
While a graduate student, I developed my love and passion for macromolecular structure, especially protein structure. It was also during this time that I fully realized the importance of noncovalent interactions in chemistry and biology. One of my favorite quotes while in graduate school was a paper from the lab of Professor Alan Fersht<sup>1</sup>: "Biology is dominated by the chemistry of the noncovalent bond." This quote is on the wall of my office and my students would certainly agree that I emphasize noncovalent interactions in molecular systems. In the case of inter- and intramolecular interactions in biological molecules, it is very difficult for textbooks to present the diverse array of possible noncovalent interactions that can exist between functional groups and the role they play in the structure and function of macromolecular systems. This is why computer-based molecular visualization programs such as Protein Explorer, RasMol, Kinemage and Deep View have opened many doors for instructors and students to explore the structural nature of small and large systems. These programs allow the user to gain access to the wealth of information contained in the PDB. Although the structures in the PDB were not strictly obtained for educational purposes, chemistry and biology instructors can tap into this tremendous resource for pedagogical purposes. Currently, my students and I use Protein Explorer<sup>2</sup> to access the PDB, although one can visit the World Index of Biomolecular Visualization Resources ([www.molvisindex.org](http://www.molvisindex.org)) and choose a suitable program.

### *Molecular Visualization in the Classroom*

I have been thoroughly impressed with how my colleagues have integrated the use of molecular visualization software and the PDB into the biochemistry and biology classroom.<sup>3</sup> Like many instructors that teach macromolecular structure, I utilize the PDB to teach the structural nature of these systems and to support material covered in the textbook. I also use Protein Explorer and the PDB to emphasize the nature and importance of noncovalent interactions that exist in proteins and protein complexes. Along with the normal classifications given to amino-acid side chains, I also categorize amino acids by their ability to form the various types of noncovalent interactions, especially the interactions involving aromatic systems ( $\pi$ - $\pi$  interactions).  $\pi$ -Type interactions in protein systems are not widely discussed in the biochemistry textbooks, even though they are an area of interest among the research community.

Over the last several years, through my own mining of the PDB and student exploration projects, I have cataloged several examples of the different types of noncovalent interactions that exist between protein side chains and in protein-ligand complexes. I show these interactions during class when teaching protein structure and when I discuss the different types of complexes proteins can make with other biomolecules such as nucleic acids, carbohydrates and lipids. Even in the metabolism chapters, I try to show the structure of metabolic enzymes to support the discussion of the mechanistic features of these enzymes and the structure-function relationship that is so important in these chapters.

In one biochemistry course I teach, the focus is on medicinal biochemistry, membrane structure and signal transduction. The students in this course have already completed a previous biochemistry course and are familiar with macromolecular structure. One might expect that the use of Protein Explorer and the PDB might be limited; however, this is certainly not the case. When teaching this course, I often find myself before lecture searching through the PDB looking for structures that can contribute to that day's lecture. I am usually successful at finding at least one structure in the PDB that can be shown and dissected via Protein Explorer that has direct application to the topic of the lecture. This type of approach leads to a balanced lecture period where there is an appropriate mix of lecture, student discussion and computer-based visualization. For example, a few days ago I was discussing the role of lipid tags (anchors) in the structure and function of peripheral membrane proteins. One method by which these proteins are associated with membranes is prenylation of C-



**PDB ID: 1o1r.**

*Turek-Etienne, T. C., Strickland, C. L., Distefano, M. D.: Biochemical and Structural Studies with Prenyl Diphosphate Analogues Provide Insights Into Isoprenoid Recognition by Protein Farnesyl Transferase Biochemistry 42 pp. 3716 (2003)*

terminal cysteine residues. This involves covalent modification with a farnesyl or geranylgeranyl group to produce a thioether linkage to the side chain of the cysteine residue. While looking through the PDB for a structure to discuss in class, I came across the PDB record (PDB ID: 1o1r) of a Protein Farnesyltransferase with bound geranylgeranyl diphosphate.<sup>4</sup> This  $\alpha$ -helical dimer is a truly interesting and beautiful protein structure that caught my eye right away. I used Protein Explorer to visualize this structure in class as it was quite complementary to the discussion of lipid tags and it sparked a discussion on the structure and function of Ras proteins. I may have been able to show an image of this enzyme in class; however, the ability to manipulate this structure in Protein Explorer allowed me to show subtle aspects of the structure (three small  $\beta$ -sheet regions) and to highlight some of



the residues (Lys-164, Lys-294 and Arg-291 and Trp-303) involved in ligand binding. The basic residues interact with the phosphate groups of the geranylgeranyl diphosphate ligand via ionic interactions and the Trp side chain seems to be interacting with the hydrocarbon moiety of the ligand through CH- $\pi$  interactions. The involvement of specific amino-acid residues in the structure and function of protein systems is always an important aspect of my biochemistry courses even though macromolecular structure might not be the major focus.

### *Student Use of Visualization and PDB Tools*

As a biochemistry instructor, I believe that utilizing computer-based visualization software, and data housed in the PDB, has been a tremendous asset in the teaching-learning process. I am very excited about the ability of students to use the software and the PDB data to explore macromolecular structure and search for noncovalent interactions. Over the past few years, I have had two different approaches to getting students involved in protein exploration projects. The first approach involves assigning individual students or student groups a particular macromolecule or giving them the option of finding their own structure in the PDB in which they have an interest. The students use visualization software to dissect the structure of the macromolecule and find several examples of noncovalent interactions important in stabilizing the structure of the protein or protein-ligand complex.<sup>5</sup> The students present their findings to the class in oral presentations and field questions from me and other students in the class. Another approach I use, especially in larger classes, is to have students complete worksheets that contain questions and tasks concerning a particular structure from the PDB. These assignments are highly structured and require students to use Protein Explorer to investigate the structure of a particular protein or protein complex.<sup>3</sup> The tasks and question for a particular assignment are provided below.

PDB ID: **1lou**

**TASK #1:** List and describe four ionic interactions.

**TASK #2:** List and describe three H-bonds involving the side chain of Tyr or Gln.

**TASK #3:** Find a Ser residue forming a H-bond with a water molecule.

**INTERACTION QUESTION #1:** Is there a cation- $\pi$  interaction between Phe97 and Lys39? If yes, describe the nature of the interaction. If not, provide details on why the two side chains are not participating in a stabilizing interaction.

These assignments are designed such that students have to acquire or recall information concerning the structures of amino acids and their ability to form noncovalent interactions. The students also have to demonstrate that they understand the distance and geometrical requirements for the formation of stabilizing interactions, which can be difficult when studying a macromolecule. The description required to complete the tasks and questions in the assignments involves measuring distances and manipulating the structures in Protein Explorer. Students quickly realize that proximity of amino-acid side chains does not guarantee that they are participating in a stabilizing (or repulsive) interaction.

In my opinion, there is no way that traditional tests or quizzes

can measure the ability of students to recognize and justify the existence of noncovalent interactions in protein systems. Other assignments I have used require students to comment on structural motifs that exist in proteins and to identify interactions involved in protein-ligand complexes. In some cases, students are asked to write a short paper on the biological function of protein system that is the focus of the assignment.

In some of my biochemistry courses, I ask students to give oral presentations on the molecular basis of disease. To complement the information given on the various aspects of the diseases, students have searched the PDB to find relevant structures that can be visualized in Protein Explorer during the presentation. For example, a student recently gave a presentation on Phenylketonuria, a genetic disease linked to mutations in the enzyme phenylalanine hydroxylase (PDB IDs: **1pah** and **2pah**). Protein Explorer was used to show the multimeric nature of the enzyme, the catalytic and tetramerization domains, the catalytic iron ion and its coordinating side chains and the side chain of amino acids that are mutated in the disease-causing form of the enzyme. The use of molecular visualization in the student presentations was complementary to the other material presented and enabled the students to better explain the molecular basis of disease.

### *References:*

- <sup>1</sup>Fersht, A. R.; Matouschek, A.; Serrano, L. *The Folding of an Enzyme: Theory of Protein Engineering Analysis of Stability and Pathway of Protein Folding*. J. Mol. Biol. 1992, 224, 771-782.
- <sup>2</sup>Martz, E. *Protein Explorer: Easy yet Powerful Macromolecular Visualization*. Trends Biochem. Sci. 2002, 27, 107-109.
- <sup>3</sup>Honey, D. W.; Cox, J. R. *Lesson Plan for Protein Exploration in a Large Biochemistry Class*. Biochem. Mol. Biol. Educ. 2003, 31, 356-362.
- <sup>4</sup>Turek-Etienne, T. C.; Strickland, C. L.; Distefano, M. D. *Biochemical and Structural Studies with Prenyl Diphosphate Analogues Provide Insights into Isoprenoid Recognition by Protein Farnesyl Transferase*. Biochemistry 2003, 42, 3716-3724.
- <sup>5</sup>Cox, J. R. *Teaching Noncovalent Interactions in the Biochemistry Curriculum Through Molecular Visualization: The Search for  $\pi$ -Interactions*. J. Chem. Educ. 2000, 77, 1424-1428.

## RELATED LINKS: NMR RESOURCES

[manatee.bmrb.wisc.edu:8000/bmrb-adit](http://manatee.bmrb.wisc.edu:8000/bmrb-adit)

The ADIT-NMR Deposition Tool for submitting experimental results to the BMRB is similar to the ADIT interface for deposition structural data to the PDB, and is the result of a collaboration between the BMRB and the RCSB-PDB.

[www.ccpn.ac.uk](http://www.ccpn.ac.uk)

Collaborative Computing Project for the NMR Community

[www.rcsb.org/pdb/software-list.html#NMR](http://www.rcsb.org/pdb/software-list.html#NMR)

Refinement programs for NMR

[www.usm.maine.edu/%7erhodes/ModQual](http://www.usm.maine.edu/%7erhodes/ModQual)

"Judging the Quality of Macromolecular Models, A Glossary of Terms from Crystallography, NMR, and Homology Modeling," by Gale Rhodes

[www.cis.rit.edu/htbooks/nmr/inside.htm](http://www.cis.rit.edu/htbooks/nmr/inside.htm)

"The Basics of NMR," by Joseph P. Hornak

## RCSB PDB LEADERSHIP TEAM

*The overall operation of the PDB is managed by the RCSB PDB Leadership Team.  
Technical and scientific support are provided by the RCSB PDB Members.*

DR. HELEN M. BERMAN, *Director*  
Rutgers University  
[berman@rcsb.rutgers.edu](mailto:berman@rcsb.rutgers.edu)

DR. PHILIP E. BOURNE  
*Co-Director*  
SDSC/UCSD  
[bourne@sdsc.edu](mailto:bourne@sdsc.edu)

JUDITH L. FLIPPEN-ANDERSON  
*Production and Outreach Leader*  
Rutgers University  
[flippen@rcsb.rutgers.edu](mailto:flippen@rcsb.rutgers.edu)

DR. GARY L. GILLILAND  
*Co-Director*  
CARB/UMBI/NIST  
[gary.gilliland@nist.gov](mailto:gary.gilliland@nist.gov)

DR. JOHN WESTBROOK  
*Co-Director*  
Rutgers University  
[jwest@rcsb.rutgers.edu](mailto:jwest@rcsb.rutgers.edu)

## RCSB PDB MEMBERS

*The RCSB PDB is a member of the Worldwide PDB ([www.wwpdb.org](http://www.wwpdb.org))*

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

### Rutgers

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

KYLE BURKHARDT  
[kburkhar@rcsb.rutgers.edu](mailto:kburkhar@rcsb.rutgers.edu)

LI CHEN  
[lchen@rcsb.rutgers.edu](mailto:lchen@rcsb.rutgers.edu)

SHARON COUSIN  
[sharon@rcsb.rutgers.edu](mailto:sharon@rcsb.rutgers.edu)

JIM CROKER  
[jimc@rcsb.rutgers.edu](mailto:jimc@rcsb.rutgers.edu)

DR. SHUCHISMITA DUTTA  
[sdutta@rcsb.rutgers.edu](mailto:sdutta@rcsb.rutgers.edu)

DR. ZUKANG FENG  
[zfeng@rcsb.rutgers.edu](mailto:zfeng@rcsb.rutgers.edu)

LEW-CHRISTIANE  
FERNANDEZ  
[fernandz@rcsb.rutgers.edu](mailto:fernandz@rcsb.rutgers.edu)

DR. RACHEL KRAMER  
GREEN  
[kramer@rcsb.rutgers.edu](mailto:kramer@rcsb.rutgers.edu)

VLADIMIR GURANOVIC  
[vladimir@rcsb.rutgers.edu](mailto:vladimir@rcsb.rutgers.edu)

DR. SHRI JAIN  
[sjain@rcsb.rutgers.edu](mailto:sjain@rcsb.rutgers.edu)

DR. IRINA PERSIKOVA  
[persikova@rcsb.rutgers.edu](mailto:persikova@rcsb.rutgers.edu)

SUZANNE RICHMAN  
[richman@rcsb.rutgers.edu](mailto:richman@rcsb.rutgers.edu)

MELCOIR ROSAS  
[melcoir@rcsb.rutgers.edu](mailto:melcoir@rcsb.rutgers.edu)

DR. BOHDAN SCHNEIDER  
[bohndan@rcsb.rutgers.edu](mailto:bohndan@rcsb.rutgers.edu)

SUSAN VAN ARNUM  
[susan@rcsb.rutgers.edu](mailto:susan@rcsb.rutgers.edu)

ALICE XENACHIS  
[alicex@rcsb.rutgers.edu](mailto:alicex@rcsb.rutgers.edu)

DR. HUANWANG YANG  
[hyang@rcsb.rutgers.edu](mailto:hyang@rcsb.rutgers.edu)

DR. JASMIN YANG  
[jasmin@rcsb.rutgers.edu](mailto:jasmin@rcsb.rutgers.edu)

CHRISTINE ZARDECKI  
[zardecki@rcsb.rutgers.edu](mailto: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  
[address@sdsc.edu](mailto:address@sdsc.edu)

DR. WOLFGANG F. BLUHM  
[wbluhm@sdsc.edu](mailto:wbluhm@sdsc.edu)

DR. NITA DESHPANDE  
[nita@sdsc.edu](mailto:nita@sdsc.edu)

ANN KAGEHIRO  
[kagehiro@sdsc.edu](mailto:kagehiro@sdsc.edu)

CHARLIE KNEZEVICH  
[charlie@sdsc.edu](mailto:charlie@sdsc.edu)

JEFF MERINO-OTT  
[jott@sdsc.edu](mailto:jott@sdsc.edu)

WAYNE TOWNSEND-  
MERINO  
[wayne@sdsc.edu](mailto:wayne@sdsc.edu)

### CARB/UMBI/NIST

Center for Advanced Research in  
Biotechnology  
9600 Gudelsky Drive  
Rockville, MD 20850

AL CARLSON  
[carlson@umbi.umd.edu](mailto:carlson@umbi.umd.edu)

KEVIN LWINMOE  
[lwinmoe@umbi.umd.edu](mailto:lwinmoe@umbi.umd.edu)

DR. VEERASAMY RAVICHANDRAN  
[vravi@nist.gov](mailto:vravi@nist.gov)

KATHRYN ROSECRANS  
[rosecran@umbi.umd.edu](mailto:rosecran@umbi.umd.edu)

ELIZABETH WALKER  
[walkere@umbi.umd.edu](mailto:walkere@umbi.umd.edu)

### 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 PROTEIN DATA BANK  
[www.pdb.org](http://www.pdb.org)

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

*Return Service Requested*

Non-Profit Org  
US Postage  
**PAID**  
Permit No 1889  
San Diego, CA