Summer 2005 • Number 26

Published quarterly by the Research Collaboratory for Structural Bioinformatics Protein Data Bank

RCSB PDB PROTEIN DATA BANK NEWSLETTER

Contents

SNAPSHOT: JULY 1, 2005
31535 released atomic coordinate entries

MOLECULE TYPE		EXPERIMENTAL TECHNIQUE	
28739	proteins, peptides, and viruses	26932	diffraction and other
1481	nucleic acids	4603	NMR
1302	protein/nucleic acid complexes	17182	structure factor files
13	carbohydrates	2552	NMR restraint files

PARTICIPATING RCSB MEMBERS:
- **RUTGERS:** rutgers.rcsb.org
- **SDSC/UCSD:** www.pdb.org

E-mail: info@rcsb.org
FTP: ftp.rcsb.org
RCSB PDB Beta: pdbbeta.rcsb.org
The RCSB PDB is a member of the wwPDB (www.wwpdb.org)

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

Message from the RCSB PDB

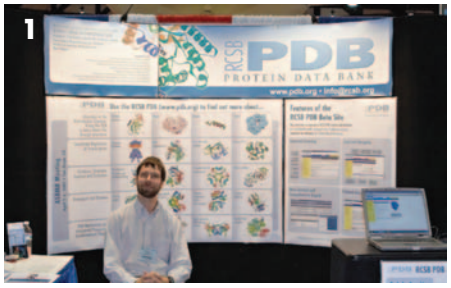
In July 2004, the RCSB PDB released a reengineered website and database (**pdbbeta.rcsb.org**) for public beta testing. The underlying database consists of curated mmCIF files resulting from the data uniformity project, which allows for improved query access to the unified data.

Thanks to feedback sent from users, this website is constantly being refined and enhanced. Some of the features of this site are described in this newsletter, and in previous newsletters.

This RCSB PDB Newsletter has also been redesigned – the layout has been revised to make it easier to read and to find articles of interest.

The beta RCSB PDB website, along with tools for deposition, will be demonstrated at the wwPDB exhibit stand (Booth number 12) at this year's XX Congress & General Assembly of the International Union of Crystallography (IUCr; August 23 - 31 in Florence, Italy). Our wwPDB partners MSD-EBI and PDBj will also demonstrate their websites and we will all be available throughout the exhibition to meet with PDB users. A roundtable discussion about data mining from the PDB will be held on August 27 with presentations from Helen M. Berman, Kim Henrick, Haruki Nakamura, and Philip E. Bourne. Other wwPDB-related activities at this meeting will be announced on the RCSB PDB website. We look forward to seeing you there. ❖ *The RCSB PDB*

RCSB PDB at Recent Meetings



1. Wolfgang F. Bluhm at the American Society for Biochemistry and Molecular Biology's (ASBMB) Annual Meeting (April 2-6, San Diego, CA) 2. The RCSB PDB exhibit booth at the American Chemical Society's Mid-Atlantic Regional Meeting (MARM; May 22-25, 2005; Rutgers, University) 3. Kyle Burkhardt meets with Darrell Hurt at the American Crystallographic Association's Annual Meeting (ACA; May 28-June 2, 2005; Orlando, FL) 4. wwPDB collaborators Kim Henrick (Head, MSD-EBI) and Helen M. Berman (Director, RCSB PDB) will be alongside other wwPDB members at the upcoming IUCr meeting

Data Deposition and Processing

PDB Focus: pdb_extract Makes Deposition Easier

pdb_extract helps depositors automatically prepare crystal structure depositions. This software tool extracts information about data collection, phasing, density modification, and the final structure refinement from the output files produced by many applications used for structure determination. The collected information is organized into an mmCIF file that is ready for deposition. Fewer data items have to be manually entered – saving time and minimizing errors.

pdb_extract can be downloaded in source and binary versions for Linux, SGI, SUN, OSF and Mac OSX from **sw-tools.pdb.org**. Source and Linux binary versions of ADIT are also available.

pdb_extract is also part of the CCP4i interface.

PDB_EXTRACT LINKS:

Desktop: **sw-tools.pdb.org/apps/PDB_EXTRACT**

Online: **pdb-extract.rutgers.edu**

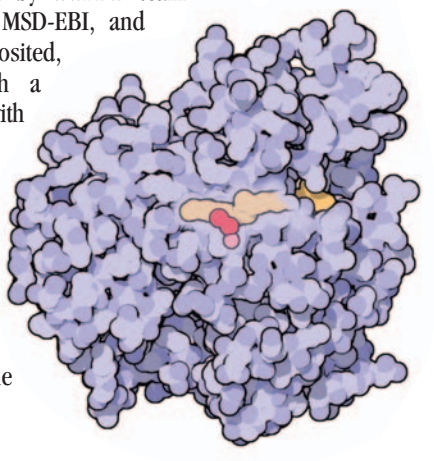
CCP4: **www.ccp4.ac.uk**

PDB Deposition Statistics

In the first half of 2005, 3168 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, 67% were deposited with a release status of HPUB; 17% with HOLD; and 16% with REL.

78% of these entries were determined by X-ray crystallography; 19% were determined by NMR. 80% were deposited with experimental data. 52% released the sequence in advance of the structure's release.



June's Molecule of the Month: Carotenoid Oxygenase
PDB ID 2biw : Kloer, D. P., Ruch, S., Al-Babili, S., Beyer, P., Schulz, G. E.: The Structure of a Retinal-Forming Carotenoid Oxygenase Science *308 pp. 267 (2005)*

Data Query, Reporting, and Access

RCSB PDB Beta Site Features

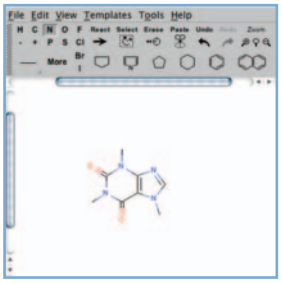
In July 2004, the RCSB PDB released a reengineered beta site (**pdbbeta.rcsb.org**) for public testing. Some of the features of this site are described below.

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

- **Improved Searching and Visualization for Ligands**

Beta site searches can use common ligand names or the identification codes from the Chemical Component Dictionary (formerly called the HET group dictionary). These queries will search ligand names, some synonyms, and class specifications using the Chemical Component Dictionary created by curation efforts of the RCSB PDB team.

Ligand name searching supports partial string matches. For example, searching for 'benz' will return all structures that contain benzene as well as those containing benzamidine. For an exact match, the complete name of the ligand must be entered. Ligand searches can also be performed using the three-character ligand ID in the PDB file (the "HET" record). For example, searching for 'HEM' returns all structures that have a heme ligand.



The PDB can be searched for structures containing the same ligand by drawing a ligand in MarvinSketch **(www.chemaxon.com)**

A recently added search feature is the ability to query for ligands using a SMILES string representation or a 2D structure of the ligand drawn using the MarvinSketch applet. SMILES (Simplified Molecular Input Line Entry Specification) is a comprehensive yet simple nomenclature system for chemicals. A SMILES string represents the valence model of a molecule.

For example, [Fe+2] or [Fe++] is the SMILES string for iron(II) cation; C1=CC=CC=C1 or c1ccccc1 is the SMILES string for benzene; and Nc1ncnc2n(cnc12)C3OC(COP(O)(=O)OP(O)(=O)OP(O)(O)=O)C(O)C3O is the SMILES string for Adenosine-5'-triphosphate.

Exact, substructure, and similarity searches can be performed. An exact search will retrieve PDB IDs associated with ligands whose structures match the SMILES/2D structure exactly. A substructure search will retrieve all PDB IDs associated with ligands that are superstructures of the query. A similarity search returns PDB IDs associated with ligands whose topology is similar to that of the query.

Similarity searches are based on finding molecules similar to the query based on a dissimilarity coefficient, whose value can be set to range from 0 to 1. The default value is 0.3. The dissimilarity coefficient is defined as 1- Tanimoto coefficient. The higher the value, the greater the hits, since

ligands that are only remotely similar to the query are also returned. A lower threshold returns fewer hits. It is more stringent in that it returns ligands that have greater similarity to the query structure.

Both the SMILES string search and the 2D structure search list all the ligands that match the query and the criteria. The matching ligands can be viewed and further explored. A 2D ligand viewer can also be launched from the Structure Explorer page for a single structure.

- **Molecular Viewers**

The beta site features three third-party molecular viewers for interactively visualizing structures: KiNG (Kinemage, Next Generation), written by Ian Davis (**kinemage.biochem.duke.edu/software/king.php**); Jmol, an open source molecule viewer (**jmol.sourceforge.net**); and WebMol, written by Dirk Walther (**www.cmpharm.ucsf.edu/~walther/webmol.html**)

Links to all three viewers are found on each entry's Structure Explorer page. The viewers require a Java-enabled browser, but not any additional plug-ins or helper installations. However, some applets may require the user to accept a security certificate upon first download (click "Yes" or "Always").

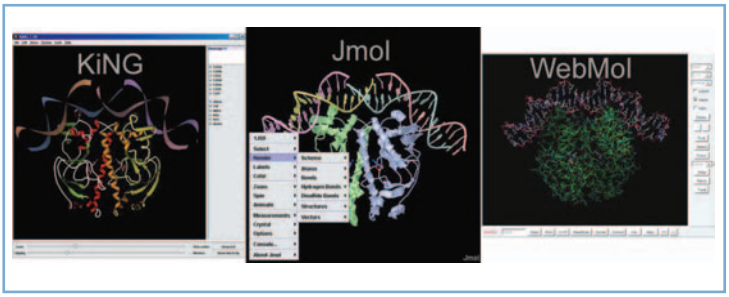
All three viewers offer rich functionality for visualizing molecules, with many options for selecting, rendering, and coloring portions of or entire PDB structures.

KiNG, by default, presents a colored ribbon diagram of the structure. The top menu contains many visualization options, along with additional tools, and the ability to save and later reload the currently displayed view. The right hand panel contains a list of check boxes that determine which molecular entities in the structure file are being displayed. This option extends to the models in NMR structures, and is therefore particularly convenient for comparing multiple NMR models in a single PDB file.

Jmol offers a large number of options for selecting portions of a structure and rendering them in different ways (for example showing a space filled representation of a ligand and a ribbon diagram of the main protein chain). This can be accomplished by right clicking on the applet, and by choosing "Select", "Render", and "Color" from the cascading menu. Other convenient features include the ability to continuously spin the structure, or to visualize the crystal axes or unit cell boundaries.

WebMol presents a stick model of the molecule with several options for coloring (e.g. by chain or by B-factor). Dotted molecular surfaces can also be displayed. Distance matrix plots or Ramachandran plots can be opened in a separate window, which is interactively linked to the display of the molecule.

For further help, information about these viewers is provided – KiNG and WebMol help are available from within the applets, and Jmol help is available at the Jmol home page.



Three viewers available from the RCSB PDB beta website

PDBML/XML Data Uniformity Files

After an extended period of beta testing, the remediated PDB data files from the data uniformity project are now available in PDBML/XML format on the production FTP archive in the following directories:

ftp://ftp.rcsb.org/pub/pdb/data/structures/divided/XML/

ftp://ftp.rcsb.org/pub/pdb/data/structures/divided/XML-extatom/

ftp://ftp.rcsb.org/pub/pdb/data/structures/divided/XML-notatom/

The files in the XML directory contain separate XML tags for each item in the atom-site category. XML-extatom files contain the atom records only, in an alternate format with only one pair of XML tags for each atom. XML-noatoms files contain only the metadata for each structure and no atom records. All files are gzipped (.gz compressed). In each case, the data files are in the usual hash directories according to the middle two characters of the PDB ID (e.g. the files for 100d are in a directory 00).

More information on the PDB data uniformity project is available at **www.rcsb.org/pdb/uniformity**. Comments are welcome at **info@rcsb.org**.

PDB Focus: Redundancy Reduction Cluster Data Available on the PDB FTP Site

The results of the weekly clustering of protein chains in the PDB are posted at **ftp://ftp.rcsb.org/pub/pdb/derived_data/NR/**. These clusters are used in the "remove similar sequences" feature on SearchLite and SearchFields on the PDB web sites.

Files that list the clusters and their rankings at 50%, 70% and 90% sequence identity are available. Smaller rank numbers indicate higher (better) ranking. Chains with rank number 1 are the best representative of their cluster.

The contents of these files and the details of the clustering and ranking are further described at **ftp://ftp.rcsb.org/pub/pdb/derived_data/NR/README** and **www.rcsb.org/pdb/redundancy.html**.

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**.

Access statistics are given below for the primary RCSB PDB website at **www.pdb.org**.

	DAILY AVERAGE		MONTHLY TOTALS			
MONTH	HITS	FILES	SITES	KBYTES	FILES	HITS
Jun 05	258,126	188,615	115,846	223,130,124	5,469,838	7,485,679
May 05	283,870	204,733	131,339	281,893,387	6,142,012	8,516,121
Apr 05	320,485	234,752	143,047	356,340,329	6,573,056	8,973,602

Access Statistics for www.pdb.org

Outreach and Education

RCSB PDB Focus: Help Desks for Beta Site Feedback, Deposition Information, and More

For answers to questions ranging from "how can I deposit my structure" to "how can I create a report about the structures I've found" to "what is DNA?", the RCSB PDB actively maintains several e-mail help desks. Responses are rapidly returned.

- **betafeedback@rcsb.org** receives bugs and comments about the beta web site that is currently undergoing testing (**pdbbeta.rcsb.org**)
- **deposit@rcsb.rutgers.edu** answers questions about deposition and annotation. A FAQ regarding depositing, updating and releasing files is also available (**rcsb-deposit.rutgers.edu/depoinfo/depofaq.html**)
- **info@rcsb.org** responds to questions and comments relating to the navigation of the RCSB PDB. Questions about searches, reports, and using all of the resources available from the RCSB PDB should be sent to this address.

Art of Science Exhibit at Texas A&M and Rutgers

The Art of Science Exhibit at Texas A & M

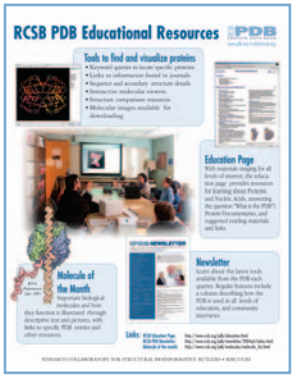
The RCSB PDB's "Art of Science" exhibit was on tour this past spring. It appeared at Texas A&M University's Visual Arts Gallery in the Memorial Student Center (April 13 - May 15, 2005). The show was also part of the American Chemical Society Mid-Atlantic Regional Meeting (MARM) held at Rutgers (May 22-25, 2005).

The Art of Science traveling exhibit looks at the beauty inherent in protein structures. It displays images of molecules in the PDB, including the pictures available from Structure Explorer pages and from Molecule of the Month features. Since its beginnings at a space dedicated to art exhibits at Rutgers University, the show has traveled to many places, including EMBL-Hamburg, Germany; University of Wisconsin-Madison; California State University, Fullerton; Purdue University; and Hyderabad, India. The RCSB 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.**

RCSB PDB Educational Resources Poster

Resources for education available from the RCSB PDB are highlighted on a poster that is available for download (**www.rcsb.org/pdb/education.html**; 8 1/2 by 11 inches).

- The **Molecule of the Month** feature illustrates important biological molecules and how they function through descriptive text, pictures, and links to specific PDB entries and other resources.
- The **RCSB PDB Newsletter** regularly features interviews with members of the community and descriptions of how the PDB is used in all levels of education.
- The **Education Page** provides resources for learning about Proteins and Nucleic Acids, protein documentaries, and suggested reading materials and links.



Poster: Educational Resources at the RCSB PDB

- **RCSB PDB tools for finding and visualizing proteins** are used in the classroom in a variety of ways, including downloading molecular images, exploring the links to information found in journals, and trying different keyword queries to locate specific proteins.

RCSB Meetings: ACA and ISMB

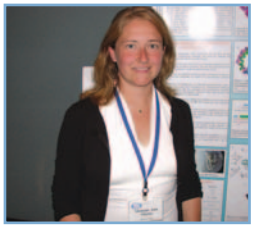
Highlights from the Annual Meeting of the American Crystallographic Association (ACA) held May 28 - June 2, 2005 in Orlando, FL included exhibiting in "Data Alley", along with CCP4 and CCDC. RCSB PDB staff were on hand to answer questions and provide demonstrations of deposition software and the beta site. Annotator Kyle Burkhardt presented a tutorial on using RCSB PDB validation software at the "Workshop on Macromolecular Structure Validation."

The RCSB PDB also exhibited at the 13th Annual Meeting of the International Society for Computational Biology ("Intellegent Systems for Molecular Biology" June 25 - 29 in Detroit, Michigan). Wolfgang Bluhm presented "Structural Bioinformatics Education from the RCSB Protein Data Bank" as part of the Education session.

RCSB Poster Prize Awarded at RECOMB and ACA

Thanks to the students and judges who participated in the recent RCSB PDB Poster awards. Details are available at **www.rcsb.org/pdb/poster_prize.html**. The next award will be presented at the XX Congress of the International Union of Crystallography (August 23-31 2005 in Florence, Italy).

- At the Ninth Annual International Conference on Research in Computational Molecular Biology (RECOMB, May 14-18, Cambridge, MA), the RECOMB & PDB Poster Awards recognizing insight and innovation in structural computational biology went to "Comparative Modeling of Mainly-Beta Proteins by Profile Wrapping" by Andrew V. McDonnell, Matthew Menke, Nathan Palmer, Jonathan King, Lenore Cowen, Bonnie Berger (MIT) and "MAPPIS: Multiple Alignment of Protein-Protein Interfaces" by Alexandra Shulman-Peleg, Maxim Shatsky, Ruth Nussinov and Haim J. Wolfson (Tel-Aviv University).



- At the ACA, the award for best student presentation went to "Safety in Cycling: Novel Redox Proteins from Escherichia coli" by Melanie A. Adams (pictured) and Zongchao Jia (Queen's University).

Melanie A. Adams at the ACA meeting

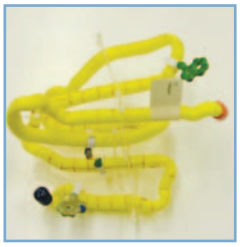
PDB Education Corner:

Wisconsin High School Science Olympiad PROTEIN MODELING Event by Gary Graper, Event Supervisor

GARY GRAPER *is a former biology teacher who taught 35 years at Madison West High School in Wisconsin before retiring from the classroom 2 years ago. Since retirement, he has been promoting the Wisconsin Science Olympiad, facilitating University of Wisconsin College of Engineering outreach to K-12 education, helping teachers develop constructivist approaches to instruction in their classrooms, and advising a high school SMART (Students Modeling A Research Topic) team.*

An understanding of 3D molecular structure and function is at the heart of rapidly expanding fields in the molecular biosciences. Protein Modeling is a new Wisconsin Science Olympiad event in which teams first master the use of a molecular visualization tool to display and analyze a molecular structure, and then create a physical model of that structure using mini-toobers, a free-form modeling media especially designed for this purpose. This event has introduced Science Olympiad competitors in Wisconsin to the value of 3D molecular visualization in understanding protein structure and function, to the PDB and the wealth of information it contains, and to the RCSB PDB's Molecule of the Month's informative articles on relevant and important molecules. The event is an excellent opportunity for high school students to practice inquiry-based science using the tools and methods of research scientists.

Science Olympiad tournaments are rigorous academic interscholastic competitions that consist of over 30 individual and team events which students prepare for during the year. These challenging and motivational events are well-balanced between the various disciplines of biology, earth science, chemistry, physics, computers and technology. The events require knowledge of science facts, concepts, processes, skills and science applications. Students demonstrate an understanding and mastery of science, mathematics, and technology concepts that require not only knowledge and problem solving skills but also the ability to work together as a team. Science Olympiad is devoted to improving the quality of science education, creating a passion for learning science and providing recognition for outstanding achievement in science. One of the major goals of the Science Olympiad is to elevate science education and learning to a level of enthusiasm and support that is normally reserved only for varsity sports programs. At the end of the competition individual medals and team trophies are awarded to the top competitors.



The Science Olympiad Protein Modeling event was part of the Wisconsin Division C (High School) 2005 regional and state (April 2) competitions. For the competition, teams of 1 to 3 students first had to design and build a model of the potassium channel protein (1bl8) that told the story of the protein's function. They utilized RasMol and the Molecule of the Month article, as well as individual research to explore the potassium channel structure and learn how the protein functions. The molecular model was constructed in the weeks prior to the competition using mini-toobers for the backbone and any other creative materials of the students' choice to illustrate the significant structural features of the protein. The models were impounded for scoring on the day of the competition. The scoring rubric was based on accuracy of the physical model, as well as creativity and originality of design. This first part of the competition counted for 40% of the final score.

The second part of the competition consisted of a 50 minute time period during which the teams designed and built a physical model and answered questions about a protein selected from the Molecule of the Month. The modeled structure used at the regional competitions was amino acids #4-31 of the zinc finger molecule (1ZAA), and at the state competition was chain B of the major histocompatibility complex molecule (1HSA). The model was constructed using mini-toobers and clip-on amino acid side chains provided at the competition. In addition, students were provided with RasMol, a PDB file, and the Molecule of the Month feature to guide their model construction and to help answer questions about the protein's structure, function, importance, researchers, when and where the research was published, etc. At the end of the 50 minutes, the model was scored with a rubric for accuracy which counted 30% of final score, and the multiple-choice test was scored which counted for 30% of the final score.

The Wisconsin Science Olympiad Protein Modeling event was designed, organized, and supported by the Center for BioMolecular Modeling (CBM) at the Milwaukee School of Engineering along with myself, a retired science teacher who has utilized the services of the CBM for a number of years to develop curriculum that makes chemistry more interesting and understandable for high school students as well as to support my former school's SMART (Students Modeling A Research Topic) team. The team of CBM members Director Tim Herman, Co-Director Michael Patrick, Jennifer Morris and Shannon Colton and myself wrote the event descriptions and rules, selected molecules to be modeled, planned and taught 4 workshops around the state of Wisconsin, developed scoring rubrics, and ran and judged 2 regional and the state competitions. The mini-toobers and other modeling materials were provided to the fifty teams registered for the event by 3D Molecular Designs. The event was a success at both the regional and state competitions with many competitors expressing

enthusiasm and support for continuation of the event. At the National Science Olympiad held at the University of Illinois-Champaign on May 21, 2005, the event was presented to numerous other state directors with a great deal of interest. It is hoped that arrangements can be made to have other states run the event in 2006, and that it can be a trial event at the 2007 national competition and possibly a regular event at state and national competitions in the future.

References

PDB ID 1bl8. Doyle, D. A., Morais Cabral, J., Pfuetzner, R. A., Kuo, A., Gulbis, J. M., Cohen, S. L., Chait, B. T., MacKinnon, R.: *The structure of the potassium channel: molecular basis of K+ conduction and selectivity.* Science 280 pp. 69 (1998)

PDB ID 1zaa. Pavletich, N. P., Pabo, C. O.: *Zinc finger-DNA recognition: crystal structure of a Zif268-DNA complex* at 2.1 Å. Science 252 pp. 809 (1991)

PDB ID 1hsa. Madden, D. R., Gorga, J. C., Strominger, J. L., Wiley, D. C.: *The three-dimensional structure of HLA-B27 at 2.1 Å resolution suggests a general mechanism for tight peptide binding to MHC.* Cell 70 pp. 1035 (1992)

Related Links

National Science Olympiad: www.soinc.org

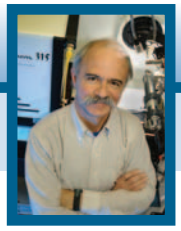
Wisconsin Science Olympiad: wisconsinso.uwstout.edu

Wisconsin Science Olympiad Protein Modeling Event: wisconsinso.uwstout.edu/wsoprotein.html

Center for BioMolecular Modeling: www.rpc.msoe.edu/cbm

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

RasMol: www.umass.edu/microbio/rasmol



PDB Community Focus:

Robert M. Sweet, Brookhaven National Laboratory

ROBERT (BOB) SWEET *is a member of the Biology Department at Brookhaven National Laboratory (BNL), and group leader of the Macromolecular Crystallography Research Resource (PXRR) at the National Synchrotron Light Source (NSLS; funded by BER/DOE and NCRR/NIH). Raised in rural midwestern US, he was educated at Caltech and the University of Wisconsin, Madison. His first exposure to crystallography was at the knee of Dick Marsh at Caltech, where in about 1963 he estimated intensities visually from Weissenberg photographs and calculated his first Fourier synthesis with Beevers-Lipson strips and a Marchant calculator. An automated diffractometer helped to solve a few cephalosporin structures in the lab of Larry Dahl in Madison that provided Bob with a PhD at the beginning of 1970. Then postdoctoral work with David Blow at the MRC Lab in Cambridge gave him an introduction to protein crystallography. Bob also managed to play a small role in the creation of modern oscillation photography in cooperation with Uli Arndt and Alan Wonacott. He spent a decade in Chemistry at UCLA, and has been at BNL since 1983. There, the importance of the NSLS to the PX community has grown steadily: the PXRR now comprises six beam lines, and has contributed to over 280 publications during the last year.*

Q: It's very clear that the PXRR facility will be generating a tremendous number of macromolecular structures, all of which should ultimately be deposited in the PDB. What is your vision for optimal interactions between facilities such as the PXRR and the PDB?

A: Well, to begin with we can remind our users that you're important to us. For six or seven years, our beam-time request form has had a check box (persuasively pre-checked) followed by the words, "Acknowledge your intent to submit the coordinates of the structure derived from this work to the Protein Data Bank." Also, for at least that long, we've had a dream that our data collection and processing programs would create a stream of information in mmCIF format that would represent a major part of the experimental portion of a PDB entry. We have fragments of the data-harvesting code available, but our proudest achievement in this regard is that just a year ago we released our experiment-tracking database, PXDB. This system (please play with it: **www.px.nsls.bnl.gov/database/pxdb_intro.html**) accepts the information from the user's initial application for beam time, logs actual visits, records the identities of specimens, and registers every image taken. We expect it to be hugely useful as Dieter Schneider and Alex Soares get our specimen-mounting robots in place. In summary, we'll grease the rails for you as best we can.

Q: What is the current ratio of "FedEx" data collection to hands-on data collection at the PXRR, and how do you see this changing as time goes on?

A: This program was started by Mike Becker, brought into regular practice by Howard Robinson, and now operated also by Annie Héroux and Alex Soares. (These folks prefer to call it "mail-in," but I think the name you used has a certain ring to it.) At the last attempt to answer this question we found that, integrated over many months, the mail-in scientists have been employing 0.8 of our 6.0 beam lines. It's a growth industry – it has nowhere to go but up; buy stock in it if you can. Another interesting question is, "What fraction of the users follow the traditional cycle of trimester proposals vs. those who gain 'rapid' access?" For five or six years we have had a rapid-access mechanism for some of our less heavily loaded beam lines. It started out as a simple web form, is now an integral part of PXDB, and will eventually be a part of the NSLS user program. Rapid access to at least five of the beam lines is overseen by Anand Saxena, and it can be **really fast**. Essentially, if there's time available and you have crystals, Anand will get you in. So the answer to the question is that something over half of our users come to us this way. It's quick, efficient, and more personalized than you might think. In this context, the FedEx work can be quick: typically a week elapses from the time a dewar of cryocooled crystals arrives until data are reported back to the user and the project is essentially finished.

Q: Rapid on-site data collection, mail-in service, and automatic data reduction and structure solving packages are having a profound effect on macromolecular structure determination. How much crystallography will a researcher need to know to solve structures? And, as an educator continually involved in teaching crystallography, what do you see as the best way to teach the fundamentals of crystallography to the growing base of scientists producing and using the structural data stored in the PDB?

A: I think your first question is really, "How little may a researcher know and still be able to solve a structure reliably?" Well, how much do *you* know about optics and diffraction gratings when you measure an UV/Vis absorption spectrum? Not very much, probably, but you still get really good spectra because the instrument just works. I think that we who devise instruments, software, and methods take this as a model: eventually for some range of "routine" structures, macromolecular crystallography will work about that well. Of course if there are half a dozen anomalies that can give a misleading absorption spectrum (a one-dimensional pattern), then there are at least that many *cubed* for a crystal structure.

But I'm not answering your questions. We've had seven cycles of our RapiData course – **www.px.nsls.bnl.gov/RapiData2005**. We have a number of the software "gods" come to teach firstly data reduction, and then solving of the phase problem. These descriptions are at a fairly high level – some knowledge of diffraction is really necessary. We found fairly quickly that whereas we expected to be judging applicants to the course based on their preparation, instead we find we're hoping they know enough to understand! Four years ago I started teaching a "fundamentals" course as an optional extra day at the beginning of RapiData and at least 3/4 of the students have been coming for the five-hour series of lectures. You're welcome to have a look at the visuals I used during the lectures here: **www.px.nsls.bnl.gov/fundamentals_lecture/**. You'll see that I teach a bit of diffraction theory as related to lens optics, show how a repetitive pattern gives spots, and then move on to reciprocal space. I look briefly at ancient and modern x-ray cameras and diffractometers (students liked the history). Then I develop the expressions for the structure factor and Fourier synthesis. I do a very brief treatment of symmetry, including defining a space group or two and showing how the symmetry of the diffraction pattern comes from the symmetry of the crystal. Then I talk about heavy-atom and direct-methods phasing, and I'm done: all of that in five hours.

I believe this represents the sort of thing the users of our equipment, software, and methods ought to understand in order to have any idea what is going on. The only evidence I have of the usefulness of the approach is that several students each year will say things like, "So that's what that is all about," or "I always wondered how that worked." Certainly no one has said (in our anonymous course evaluations) that it is a waste of time. It would be better, of course, if the same material were presented in a more detailed and leisurely format over ten to twenty lectures back at the university. I'm not experienced with people using but not producing PDB data, so I won't comment on that part of your question.

Q: The PDB is growing at a steadily increasing rate. How have your interactions with the PDB changed over the years and where do you see them going as both macromolecular crystallography and the PDB continue to evolve?

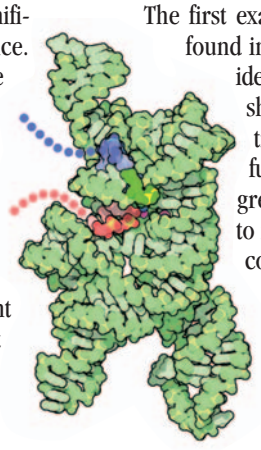
A: Well, a big change is that before 1999 I used to be able to walk across the street to talk with the PDB workers. That changed, of course, when the RCSB took over. Seriously though, the changes are small and incremental. We've had the idea in mind of pipelining information from the data stream to PDB deposition for a long time. We can see that the increased pace makes this even more important. The possibility of having a new synchrotron here (NSLS-II; **www.nsls2.bnl.gov**) gives us an accelerated mission to be ready for the increment in productivity that this will engender. I'm impressed with the ease with which the PDB is providing interchangeability among its file formats from PDB to mmCIF to the more modern XML. This innovation matches our own migration of information-exchange media. I believe we'll be able to communicate easily, and will continue to grow in parallel.

Molecules of the Quarter:

Kinesin, Self-splicing RNA, Carotenoid Oxygenase

The Molecule of the Month series explores the functions and significance of selected biological macromolecules for a general audience. This quarter, kinesin, self splicing RNA, and carotenoid oxygenase were highlighted. The May 2005 issue is excerpted here; the full Molecule of the Month features are available from **www.rcsb.org/pdb/molecules/molecule_list.html**

Self-splicing RNA: In plants and animals, most RNA molecules are made as long precursors that need to be trimmed and reassembled to create the final active molecule. These precursor RNA molecules are composed of exons, which are the important parts, separated by introns, which must be removed. In most cases, the RNA is cut and spliced together by a spliceosome, a molecular machine composed of protein and RNA. In a few cases, however, the RNA can perform the splicing reaction on its own.



The first example, discovered by Thomas Cech, was a ribosomal RNA found in a protozoan. Since then, hundreds of examples have been identified in genome sequences of many organisms. The example shown here, from PDB entry 1u6b, is part of a bacterial transfer RNA that must be spliced before it can adopt its functional form. In the illustration, the large structure in green is the intron, which uses a GTP and two magnesium ions to remove itself. The two exons that will be spliced together are colored red and blue--note that only a small piece of each exon is included in the structure.

PDB ID 1u6b: Adams, P. L., Stahley, M. R., Kosek, A. B., Wang, J., Strobel, S. A.: Crystal Structure of a Self-Splicing Group I Intron with Both Exons. Nature *430 pp. 45 (2004)*

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

- Rutgers, The State University of New Jersey: **rutgers.rcsb.org**
- 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 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