Summer 2007 • Number 34

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

RCSB PDB PROTEIN DATA BANK NEWSLETTER

Weekly RCSB PDB news is available online at www.pdb.org

Contents

Message from the RCSB PDB 1

DATA DEPOSITION AND PROCESSING

Depositing NMR Structures with ADIT-NMR 2
PDB Focus: First Time ADIT Depositors 2
Deposition Statistics 2

DATA QUERY, REPORTING, AND ACCESS

Exploring Ligands in the RCSB PDB Database 2
Using Simple Viewer to Visualize Functional Biological Units . 2
Using PubMed Abstracts to Search the PDB 3
Website Statistics 3

OUTREACH AND EDUCATION

Help Desks Answer Questions about Remediated Data, the RCSB PDB Website, Deposition, and More 3
Structural Genomics News: PSI Highlighted Structures, Technical Advances, and Assessment 3

REMEDIATION PROJECT DETAILS

Accessing the PDB Archive 4
Remediation of the Entire PDB Archive 4
The Chemical Component Dictionary 4
Accessing the Remediated Data from the RCSB PDB 4

EDUCATION CORNER: *Structures* and Other NIGMS Booklets Make Science Accessible 5

MOLECULES OF THE QUARTER 6

PDB COMMUNITY FOCUS: Alex Wlodawer 7

RCSB PDB PARTNERS, MANAGEMENT, AND STATEMENT OF SUPPORT 8

SNAPSHOT: JULY 1, 2007

44320 released atomic coordinate entries

MOLECULE TYPE		EXPERIMENTAL TECHNIQUE	
40730	proteins, peptides, and viruses	37716	X-ray
1760	nucleic acids	6367	NMR
1795	protein/nucleic acid complexes	149	electron microscopy
35	other	88	other
		27026	structure factor files
		3512	NMR restraint files

Participating RCSB Members:
Rutgers • SDSC/SKAGGS/UCSD
E-mail: info@rcsb.org
Web: www.pdb.org • FTP: ftp.wwpdb.org

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

Message from the RCSB PDB

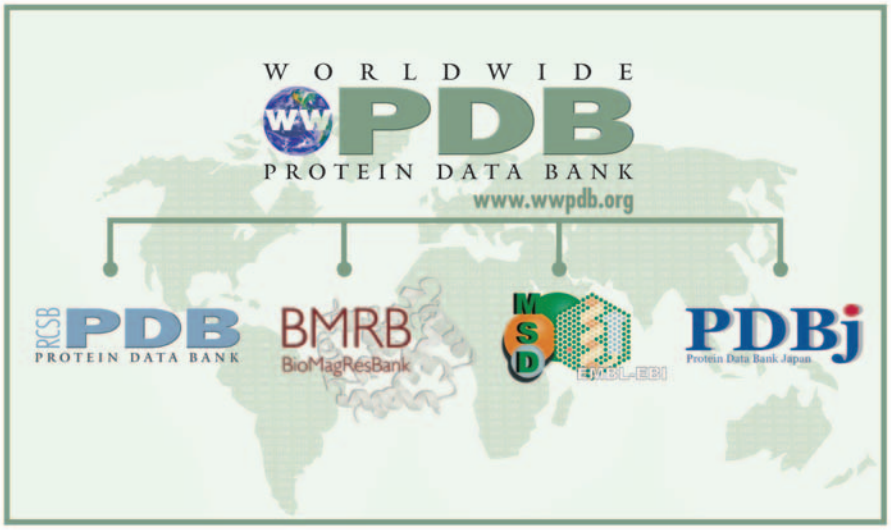
The wwPDB is pleased to announce that the PDB archive (ftp://ftp.wwpdb.org/) is now comprised of remediated data.

In the past, query across the complete PDB archive has been limited by missing, erroneous and inconsistently reported data, nomenclature, and other annotations. The evolution of experimental methods and the techniques used to process these data has introduced various inconsistencies into the PDB archive.

Over the years, wwPDB members – the RCSB PDB, MSD-EBI, PDBj, and BMRB – have worked together to ensure the uniformity of entries archived in the PDB. The entire archive has now been reviewed and remediated with the objectives of improving the detailed chemical description of non-polymer and monomer chemical components; standardizing atom nomenclature; updating sequence database references and taxonomies; resolving any remaining differences between chemical the macromolecular sequences; improving the representation of viruses; and verifying primary citation assignments.

As these corrections were being incorporated, the wwPDB worked with software developers and database maintainers to ensure a smooth transition. This year, the remediated archive was made public for testing, and the resulting feedback was incorporated into the archive. The wwPDB greatly appreciates the efforts of the many people who have taken the time to review and work with these data files, and the advice and discussions with our advisory committees.

Descriptions of this remediation process are available in this newsletter and at the wwPDB site at www.wwpdb.org.



Data Deposition and Processing

Depositing NMR Structures with ADIT-NMR

Users can now deposit NMR structure and experimental data using one tool: ADIT-NMR. Available from batfish.bmrb.wisc.edu/bmrb-adit and nmradit.protein.osaka-u.ac.jp/bmrb-adit, ADIT-NMR can be used to precheck, validate, and deposit NMR structures. Coordinates and constraint data will be processed and released by the RCSB PDB and PDBj, while other NMR spectral data (such as chemical shifts, coupling constants, and relaxation parameters, *etc.*) will be processed and archived by BMRB.

All new NMR depositions at RCSB PDB and PDBj will be submitted using ADIT-NMR. The assignment of PDB/BMRB IDs and the movement of data files between sites is fully automated. More than 100 joint depositions have already been processed through this new system. Any unfinished NMR deposition sessions that were started using ADIT before May 16, 2007 will continue to be available at that site.

Other tools for NMR depositions include:

• **pdb_extract**[1], which minimizes errors and saves time during the deposition process since fewer data items have to be manually entered.

This program extracts key details from the output files produced by many NMR applications (and X-ray crystallographic applications) for use in the deposition process. The program merges these data into macromolecular Crystallographic Information File (mmCIF) data files that can be used with ADIT-NMR to perform validation and to add any additional information for PDB deposition.

pdb_extract can be used via web interface or downloadable workstation from pdb-extract.rcsb.org.

• The **Validation Server**, which lets users check the format consistency of coordinates (PRECHECK) and to create validation reports about a structure before deposition (VALIDATION). These checks can be independently performed by the user. The Validation Server can be used at the RCSB PDB (deposit.pdb.org/validate) and PDBj (pdbdep.protein.osaka-u.ac.jp/validate) sites.

NMR structures may also be deposited using AutoDep at MSD-EBI.

PDB Focus: First Time ADIT Depositors

There are a few steps a depositor can take to make the process of depositing a structure to the PDB quick, easy, and accurate! This is an iterative process. If you encounter problems at a particular step, please make the correction(s) and go through the steps again.

1. Use the **pdb_extract** Program Suite to extract information needed for deposition from output files produced by many structure determination applications.

2. Check your structure with the **Validation Server** to ensure that the data being deposited are accurate and reflect what you intend to submit.

3. Compare your sequence with sequence database references. Any necessary corrections can then be made to your sequence and coordinates (Try BLAST at www.ebi.ac.uk/blast2 or www.ncbi.nih.gov/BLAST).

4. Use **Ligand Depot** to find the proper codes for existing ligands, to link to other entries with a particular ligand, and to search for substructures. If a ligand related to a deposition is not in Ligand Depot, please email the chemical diagram, name, and formula to deposit@deposit.rcsb.org.

5. Deposit your structure using ADIT, using its editor to add any missing information to the deposition.

For a detailed packet of information about first-time deposition, including reprints about validation and Ligand Depot, please send your postal address to info@rcsb.org with the subject line "first time depositor packet".

Deposition Statistics

In the first half of 2007, 4535 structures were deposited to the PDB archive.

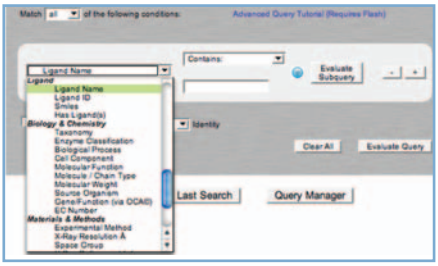
The entries were processed by the wwPDB. Of the structures deposited, 63.8% were deposited with a release status of "hold until publication"; 17.9% were released as soon as annotation of the entry was complete; and 18.3% were held until a particular date.

83.7% of these entries were determined by X-ray crystallographic methods; 16.1% were determined by NMR methods. 84.1% of these structures were deposited with experimental data. 93.7% of the crystal structures were deposited with structure factors; 34.7% of NMR structures were deposited with restraints.

Data Query, Reporting, and Access

Exploring Ligands in the RCSB PDB Database

A ligand name can be entered in the keyword text search at the top of any page from the RCSB PDB website. The Advanced Search query engine can also be used to search for structures based using a ligand's name, ID code, or SMILES string.



The Advanced Search for Ligands.

In addition to reviewing the structures that match the given query constraints, users can select the *Ligand Hits* tab, which lists the ligands known to interact with the structures matching the query. The *Ligand Hits* tab also offers a gallery view of ligand images.

Selecting one of the ligands from this tab returns a summary page with chemical and structural details. The page offers interactive and static views of the ligand. Users may also download "model" coordinates (the experimental coordinates from the first deposition of the ligand) and "ideal coordinates" (generated from the model coordinates and their connectivity) in a variety of formats including CIF, XML, SDF and PDB.

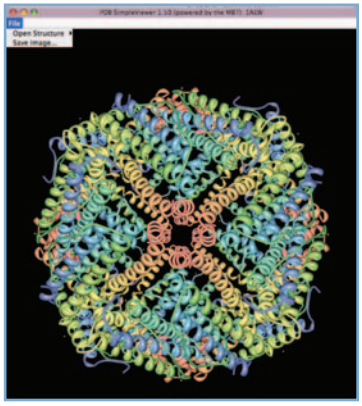
Using Simple Viewer to Visualize Functional Biological Units

When crystallographic structures are deposited in the PDB, the primary coordinate file generally contains one asymmetric unit – a concept that has applicability only to crystallography. For many of these structures, the asymmetric unit represents the functional biological molecule. In other

[1] H. Yang, V. Guranovic, S. Dutta, Z. Feng, H.M. Berman, J. Westbrook (2004) Automated and accurate deposition of structures solved by X-ray diffraction to the Protein Data Bank. *Acta Cryst.* **D60**:1833-1839.

cases, the biological unit can be generated from the asymmetric unit.

In these cases, Protein Workshop can be used to display the asymmetric unit and Simple Viewer can be used to explore the functional biological unit of a structure. Simple Viewer can rotate a structure, zoom in and out, and then save the view of the biological unit as an image file.



The biological unit of 1aew[1] in Simple Viewer.

Simple Viewer tool can be launched from the *Display Options* found on an entry's *Structure Summary* page. Simple Viewer requires Java version 1.4 or greater.

Protein Workshop and Simple Viewer both use the Molecular Biology Toolkit (MBT; J.L. Moreland, A. Gramada, O.V. Buzko, Q. Zhang and P.E. Bourne (2005) The Molecular Biology Toolkit (MBT): A Modular Platform for Developing Molecular Visualization Applications. *BMC Bioinformatics* **6**:21).

Using PubMed Abstracts to Search the PDB

PubMed abstracts are accessible from a published entry's Structure Summary page. The *Abstract* link returns a page with the article title, abstract, keywords, authors, organizational affiliation, journal, and PubMed identifier. The PubMed abstract at NCBI can also be viewed by clicking on the icon next to *Abstract*.

The text box at the bottom of the *Abstract* page can be used to search for related structures in the PDB using any word in the abstract or keyword fields. Terms can be entered into the text box either by typing the word manually or by clicking the mouse over any word in the abstract or the keyword fields.

Selecting the Abstract link (circled) will return a page with the abstract and a search box that can be used to search for structures with keywords found in the abstract.

Website Statistics

Access statistics for **www.pdb.org** are given below for the second quarter of 2007

MONTH	UNIQUE VISITORS	NUMBER OF VISITS	BANDWIDTH
APRIL 07	122,991	293,105	503.15 GB
MAY 07	123,069	300,115	504.13 GB
JUNE 07	104,693	258,107	424.74 GB

[1]P.D. Hempstead, S.J. Yewdall, A.R. Fernie, D.M. Lawson, P.J. Artymiuk, D.W. Rice, G.C. Ford, P.M. Harrison (1997) Comparison of the three-dimensional structures of recombinant human H and horse L ferritins at high resolution. *J.Mol.Biol.* **268**:424-448.

Outreach and Education

Help Desks Answer Questions about Remediated Data, the RCSB PDB Website, Deposition, and More

Electronic help desks are available to support users exploring PDB data.

info@wwpdb.org is available to address questions regarding the remediated PDB archive. The wwPDB appreciates the feedback from users who have examined the Chemical Component Dictionary and files in the remediated archive.

deposit@deposit.rcsb.org answers questions about the deposition and annotation process at the RCSB PDB.

Support pages at **deposit.pdb.org** include a file deposition and release FAQ, an overview of software tools, and tutorials for using ADIT, pdb_extract, the Validation Server, and Ligand Depot.

info@rcsb.org responds to requests relating to the navigation of the RCSB PDB website. Questions about searching, reporting, and using all of the resources available from the RCSB PDB should be sent to this address.

The RCSB PDB help system launches a separate browser window to allow users to access the help information and the website at the same time. It offers detailed topics (including Getting Started, Download Files, Search/Browse the Database, and Results), an index, glossary, and search engine. Select any of the buttons on the website to launch this system.

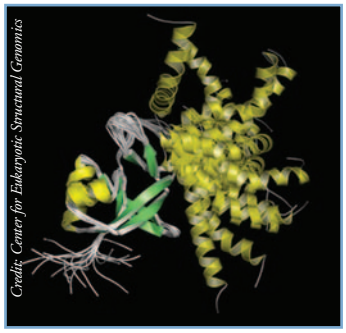
Currently Available Job Openings

The RCSB PDB is currently looking for new people to join our team. Available positions include Scientific Lead, Biochemical Information & Annotation Specialist, Web Developer/ Database Programmer, and a Database Application Programmer.

For more details, please see **www.pdb.org**.

Structural Genomics News: PSI Highlighted Structures, Technical Advances, and Assessment

The RCSB PDB Structural Genomics Information Portal (**sg.pdb.org**) offers online tools, summary reports, and target information related to structural genomics. This site also links to new features from the Protein Structure Initiative (PSI). The PSI *Structures of the Month* highlights recent structures solved by the current centers, while the *PSI Technical Highlight* describes methods developed by the effort's researchers to speed the structure determination process. These features include links to related information, such as PDB structure summaries, published papers, and center websites.



Featured as a June 2007 PSI Structure of the Month, this NMR solution structure of a plant protein may function in host defense. This protein was expressed in a convenient and efficient wheat germ cell-free system.

To date, the overall PSI effort has resulted in nearly 2,500 structures of which about 70 percent share less than 30 percent of their sequence with other known proteins. Methods and tools developed during the first phase of the PSI have been incorporated into the centers' structural genomics pipelines and adopted by structural biology labs throughout the world.

Remediation Project Details

Accessing the PDB Archive

The PDB archive has been remediated by wwPDB members the RCSB PDB, MSD-EBI, PDBj, and the BMRB. It can be accessed from ftp://ftp.wwpdb.org.

New files processed and released into the archive by the wwPDB sites will reflect the new features incorporated as part of this project, including standardized IUPAC[1] nomenclature for all chemical components.

Users may have to download new software to properly view the files with the new nomenclature (*e.g.*, RasMol, Chimera)[2]. Links to resources are available at www.wwpdb.org.

A snapshot of the unremediated PDB archive (as of July 31, 2007) is available at ftp://snapshots.rcsb.org.

Remediation of the Entire PDB Archive

Highlights of the types of information improved through remediation include:

Sequence	Updated references to databases and taxonomies Resolved differences between chemical and macromolecular sequences
Citation	Verified and updated primary citation assignments
Assembly and Virus Information	Improved representation of deposited and experimental coordinate frames, symmetry, and frame transformations
Nucleic Acid Labeling	Deoxy and ribose nucleotides assigned separate chemical definitions
Beamline Data	Beamline and synchrotron facility names have been made consistent with BioSync
Chemical Components	Standardization of chemistry and nomenclature in monomers and ligands

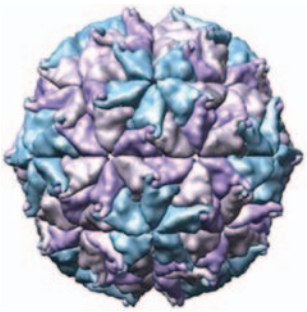


Image of 2bbv[3] created using the remediated data file and the latest version of Chimera (1.2422)

Remediated data are available for each PDB entry in three formats:

• mmCIF (mmcif.pdb.org). All remediation work was done using the PDB Exchange Dictionary (PDBx) that follows the mmCIF syntax.

•PDBML-XML (pdbml.pdb.org). Remediated data files are also available in PDBML-XML format, in a direct translation from the files in mmCIF format.

• PDB File Format (wwpdb.org). The remediated files have been released in PDB File Format version 3.0. This version of the file format incorporates standardized atom nomenclature, and distinguishes deoxyribonucleic acid from ribonucleic acid.

The Chemical Component Dictionary

The Chemical Component Dictionary (formerly known as the "HET dictionary") describes all residues in the PDB, standard and non-standard, and all small molecules. It has been remediated to address the inconsistencies in older dictionary entries that resulted in valence problems, missing model coordinates, redundant ligands, and more.

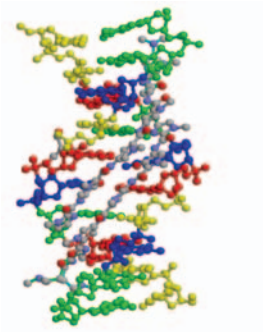


Image of 407d[4] created using the remediated data file and the latest patch to OpenRasmol (2.7.3.1)

The features of this dictionary include:

- Standardized nomenclature
- Model coordinates have been corrected, redundant chemical components obsoleted, and additional definitions for protonated forms are provided.
- Stereochemical assignments, aromatic bond assignments, idealized coordinates, chemical descriptors (SMILES & InChI)[5], and systematic chemical names have been added.

The full Chemical Component Dictionary and the companion Amino Acid Variants Dictionary can be downloaded from remediation.wwpdb.org/downloads.html.

Users can also search for individual chemical components, either by entering the component ID in the form provided, or by browsing by ID. The variant dictionary can also be browsed.

For each chemical component in the dictionary, a summary page provides a 2D chemical diagram and 3D graphic (using Jmol) of the ligand. This page also describes the ligand's physical and chemical features of the ligand. Status information along with links to the component definition in CIF and PDBML/XML formats, model coordinates, idealized coordinates, and chemical diagrams are provided.

Accessing the Remediated Data from the RCSB PDB Website

The latest release of the RCSB PDB website utilizes the data from the wwPDB Remediation Project.

This new site offers:

- Improved searching and reporting capabilities
- Updated sequence references
- Updated primary citation information and links
- Better representations for complex assemblies (such as viruses)
- Access to remediation data and pre-remediation data
- Advanced access to ligand information
- Enhanced sequence details page for each structure

For More Information ...

A variety of documents describing the remediation project are available at wwpdb.org, including format descriptions and further information about what was changed in these files. Links to software resources are also provided. Questions and comments about the remediated data should be sent to info@wwpdb.org.

[1] J.L. Markley, A. Bax, Y. Arata, C. W. Hilbers, R. Kaptein, B.D. Sykes, P.E. Wright and K. Wüthrich (1998) Recommendations for the Presentation of NMR Structures of Proteins and Nucleic Acids *Pure & Applied Chem.* **70**:117-142

[2] see sourceforge.net/projects/openrasmol and www.cgl.ucsf.edu/chimera

[3] 407d: C.L. Kielkopf, S. White, J.W Szewczyk, J.M. Turner, E.E. Baird, P.B. Dervan, D.C. Rees (1998) A structural basis for recognition of A•T and T•A base pairs in the minor groove of B-DNA. *Science* **282**:111-115

[4] 2bbv: J.P. Wery, V.S. Reddy, M.V. Hosur, J.E. Johnson (1994) The refined three-dimensional structure of an insect virus at 2.8 Å resolution. *J.Mol.Biol.* **235**:565-586

[5] D. Weininger (1988) SMILES, a chemical language and information system. 1. Introduction to methodology and encoding rules. *J. Chem. Inf. Comput. Sci.* **28**, 31 - 36 and © The International Union of Pure and Applied Chemistry (2005) IUPAC International Chemical Identifier (InChI) (contact: secretariat@iupac.org)

Education Corner

Structures and Other NIGMS Booklets Make Science Accessible

By Alisa Zapp Machalek

ALISA MACHALEK *is a science writer at the National Institute General Medical Sciences (NIGMS) at NIH. She earned B.S. and M.S. degrees in biochemistry and has research experience in neuroanatomy, biochemistry, agronomy, and breakfast cereal chemistry (at Kellogg's).*

While working on her M.S. at UW-Madison in the early 1990s, she used the PDB almost every day while creating molecular models for the 2nd edition of Principles of Biochemistry *by Lehninger, Nelson & Cox.*

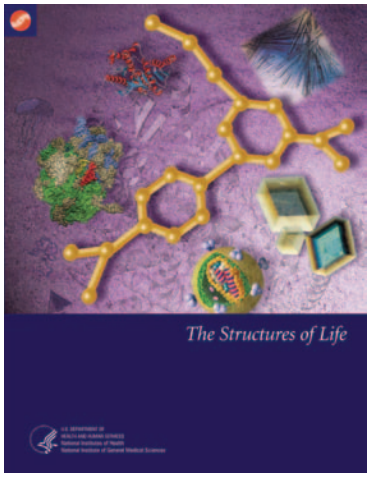
At some point, she decided she liked writing about science better than doing it. She solidified her career choice by enrolling in the graduate science writing program at the University of California, Santa Cruz.

At NIGMS, Alisa writes science education booklets, news and feature articles, profiles of scientists, research highlights for Congress, and occasional policy and publicity documents. She writes on all the areas within the NIGMS mission, including structural biology, computational biology, cell biology, chemistry, genetics, developmental biology, pharmacology, anesthesiology, trauma and burn injury, and wound healing.

As a scientist, you probably don't need to be convinced of the value, importance, and beauty of molecular structures or the thrill of studying them. But try explaining it to the public – or to teenagers.

That's just what *The Structures of Life* seeks to do. This free science education booklet is published by the National Institute of General Medical Sciences (NIGMS), a part of the National Institutes of Health that supports a good chunk of the world's structural biology research.

Naturally, the Protein Data Bank is featured throughout the booklet, both as a source of several images and as the repository into which structural biologists deposit their data to make them freely available to the scientific community.



The Structures of Life, a free booklet about structural biology, will be available in an updated edition this summer.

Why does NIGMS produce science education materials?

The Structures of Life and our many other science education materials help NIGMS show the public how their tax dollars are leading to research advances.

Improving K-12 science education in America is important for many reasons, says Jeremy M. Berg, Ph.D., NIGMS director.

"Of course, part of it is long-range workforce development," he says. "But it's broader than that. The ability to think critically and to solve problems is hugely important for all aspects of society. Many have cited the uncomfortably low math and science scores of American students[4] as evidence that, to remain leaders in the global marketplace, we will need to improve K-12 science education."

Our goal is for NIGMS educational materials to contribute to this effort. We try to encourage an understanding and appreciation of science in all readers by showcasing scientists doing cutting-edge research and explaining its potential implications.

We hope that the materials help inspire some readers to pursue careers in biomedical research. Because role models can be pivotal for young people choosing and pursuing careers, we feature male and female scientists from diverse backgrounds, geographic locations, career stages, and scientific fields.

We also strive to show that scientists have full, interesting lives and unique personalities. In our semi-annual magazine *Findings*, we've written about a crystallographer whose clarinet skill landed him in Carnegie Hall, an NMR spectroscopist who is also a former professional basketball player, a computational biologist who is an expert mountain climber, and many others.

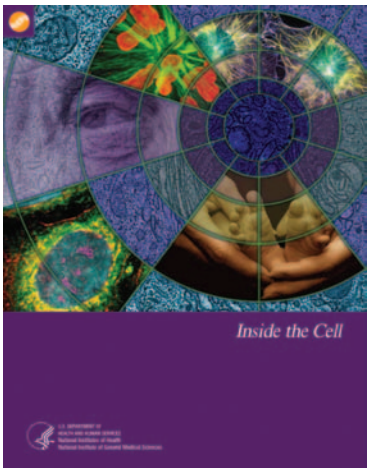
To increase understanding of the nature and importance of basic, untargeted research, we use examples from areas of science within the NIGMS mission, including structural biology, computational biology, cell biology, genetics, pharmacology, and chemistry.

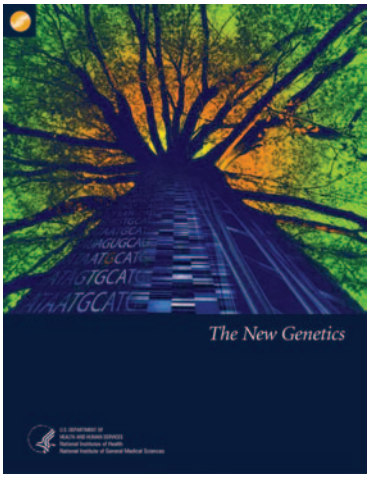
Who uses the materials?

Our booklets are used by teachers, homeschoolers, museums and science personnel, student workshop leaders, science curriculum advisors, and teacher trainers programs around the country.

Most of the materials are geared for a high school audience, but the publications are also used in some advanced middle school classes and introductory college courses.

[4] See results from PISA (Programme for International Student Assessment) at www.pisa.oecd.org. PISA is run by the Organisation for Economic Co-operation and Development, a multinational body dedicated to building strong economies worldwide. PISA tests reading, math, and science skills of 15-year-olds around the globe. In 2003, it also tested real-world problem solving skills.





NIGMS publications include Inside the Cell *and* The New Genetics. *Our newest publication, available this summer, is called* Computing Life *and covers computational biology.*

Here are a few examples of how NIGMS science education publications have been used recently.

• The Massachusetts Institute of Technology distributes NIGMS booklets to the teachers in its Summer Teacher Workshop as examples of exemplary supplementary resources and uses PowerPoint slides from *Findings* to instruct teachers how to incorporate multimedia into their lessons.

• The Arizona Biomedical Research Commission uses the publications to educate its members about the science underlying the grant applications they are reviewing for funding.

• The Distance Learning Unit in Queensland, Australia included part of an NIGMS booklet in its Senior Biology curriculum, which is distributed on CD-ROM and posted online for students who can't attend school because they live in remote areas or are disadvantaged by personal circumstances.

What is available and how can I get them?

In addition to *The Structures of Life* and *Findings*, NIGMS publishes booklets on genetics, pharmacology, cell biology, and biochemistry; a monthly electronic newsletter called *Biomedical Beat*; and a number of fact sheets. We also offer a small but growing collection of images and other multimedia resources on our website.

Our newest publication, available this summer, is called *Computing Life* and covers computational biology.

If you have suggestions about how to improve or use any of our publications, we'd love to hear from you. Contact the NIGMS Office of Communications and Public Liaison at **info@nigms.nih.gov** or 301-496-7301.

RELATED RESOURCES

You can order any of our publications at
publications.nigms.nih.gov/order

Teachers can order classroom sets at
publications.nigms.nih.gov/order/classroom.htm

Educational outlets (museums, science centers, teacher training facilities, *etc.*) can order larger quantities by contacting our office at **301-496-7301** or **info@nigms.nih.gov.**

The Structures of Life (HTML and PDF)
publications.nigms.nih.gov/structlife

Past issues of the journal *Findings* have included articles on computational biologist David Baker (September 2005), biophysicist Dorothee Kern (February 2003), and structural biologist Mavis Agbandje-McKenna (March 2006)
www.nigms.nih.gov/publications/findings

Structure of the Month and *Technical Highlight of the Month* from the Protein Structure Initiative
www.nigms.nih.gov/Initiatives/PSI

Fact Sheet: NIGMS-Supported Structure-Based Drug Design Saves Lives
publications.nigms.nih.gov/factsheets/structure_drugs.html

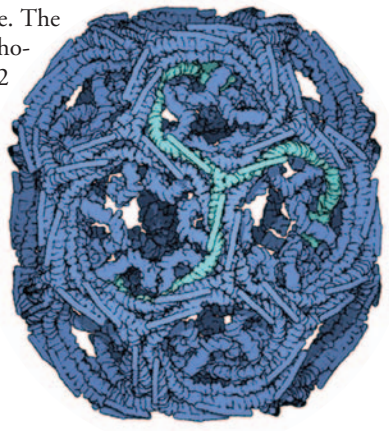
Molecules of the Quarter

Clathrin, Aconitase and Iron Regulatory Protein 1, and Fatty Acid Synthase

The **MOLECULE OF THE MONTH** *series explores the functions and significance of selected biological macromolecules for a general audience. The molecules featured this quarter were clathrin, aconitase/iron regulatory protein 1, and fatty acid synthase. The complete Molecule of the Month features are accessible from the RCSB PDB home page.*

CLATHRIN cages are composed of symmetrical three-legged components called triskelions. The structure shown here, PDB entry 1xi4, is built of 36 triskelia, one of which is highlighted in green. When triskelia snap together in solution, they can interact with enough flexibility to form either 6-sided rings that yield a flatter surface, or 5-sided rings with higher curvature. In a cell, a triskelion floating in the cytoplasm binds to an adaptor protein, linking one of its three feet to the membrane at a time. This triskelion will bind to other membrane-attached triskelia to form a rounded lattice of hexagons and pentagons, reminiscent of the panels on a soccer ball, that pulls the membrane into a bud.

By constructing different combinations of 5-sided and 6-sided rings, vesicles of different sizes may assemble. The structure shown here represents the second smallest possible cage structure, which is actually too small to contain a functional vesicle. It was created in the laboratory by reconstituting triskelions without a lipid vesicle. The smallest clathrin cage commonly photographed, called a mini-coat, has 12 pentagons and only two hexagons. Even smaller cages with zero hexagons probably don't form from the native protein, because the feet of the triskelia are too bulky.



1xi4: A. Fotin, Y. Cheng, P. Sliz, N. Grigorieff, S.C. Harrison, T. Kirchhausen, T. Walz (2004) Molecular model for a complete clathrin lattice from electron cryomicroscopy Nature ***432***:573-579.

PDB Community Focus:

Alex Wlodawer, Macromolecular Crystallography Laboratory, National Cancer Institute

ALEX WLODAWER *received his Ph.D. from the University of California, Los Angeles, in 1974. Having completed his postdoctoral training at Stanford University, he joined the National Bureau of Standards in 1976 and in 1987 moved to the National Cancer Institute, where he is Chief of the Macromolecular Crystallography Laboratory. In 1998-1999 he was an elected Visiting Fellow of Sidney Sussex College, University of Cambridge, UK. He is a member of the American Crystallographic Association and the Protein Society, and has been an elected officer in both organizations. He serves on the editorial boards of the* Journal of Biological Chemistry, Protein Science, The FEBS Journal, *and* Acta Biochimica Polonica. *He was appointed an Adjunct Professor of Biochemistry and Molecular Biology and an Associate Member of the George Washington University Institute of Biomedical Sciences, Washington DC. He was awarded a degree of Doctor Honoris Causa of the Technical University of Lodz, Poland, was elected a Foreign Member of the Polish Academy of Sciences, and is a recipient of the 2006 NCI Mentor of Merit Award.*

Dr. Wlodawer is Chief, Macromolecular Crystallography Laboratory and Chief, Protein Structure Section at the National Cancer Institute in Frederick, MD.

Q: ***In Acta D, you recently expressed the point of view that experimental data for structures solved by X-ray and NMR should be deposited and released under the same policies as coordinate files.[1] Why do you think this is so important?***

A: The question of which crystallographic results should be deposited in PDB and on what schedule has been asked many times, but still does not have the final answer. Here, the experimental data refer to the processed data, *e.g.*, the structure factors in X-ray diffraction, not the raw images. The rules changed very substantially about 8 years ago, when the International Union of Crystallography modified its deposition regulations, and their recommendations became generally accepted by most funding agencies and by scientific journals. The coordinates of published structures must now be deposited in PDB and released upon publication of the relevant papers. However, although structure factors must also be deposited, their release can be delayed by up to 6 months. In a recent Letter to Editor published in *Acta Cryst. D*,[1] I proposed that such a delay should be disallowed. I feel very strongly that the coordinates and structure factors are a matched pair, and one needs the other. The heart of the matter is that scientific results should be useful for the community (I consider description of a structure without the availability of coordinates to be advertising and not science), and verifiable (how can we prove that the structure is correct if not by comparison with the structure factors?). Let me give an example from a paper which I recently reviewed. The authors presented a series of structures of enzyme-inhibitor complexes, with one of the structures repeating a previously published experiment. However, the conformation of the inhibitor reported in the new paper was very different, changing in a substantial way the interpretation of the enzymatic mechanism. Unfortunately, with the diffraction data for the original structure never deposited (against the journal rules!), it was not possible to verify if the differences were real (and thus significant for the understanding of how the enzyme works) or due to errors in the interpretation present in the original paper. This is just one example, but I could cite many more. With the acceleration of the process of structure solution we should not have to wait half a year to verify what we read in the papers, if any doubts are raised. And let us remember that the most interesting results are often the ones that are most controversial.

Q: ***As a member of several editorial boards, what types of information are you looking for when reviewing macromolecular structure papers? Has a journal published a paper, only to be surprised by the validation remarks in the corresponding PDB file? What types of information do you think would be valuable to a reviewer of a paper describing a macromolecular structure?***

A: Oh boy, have we been surprised… I have seen many papers, often published in *Science* and *Nature* (these journals seem to care more about getting the scoop than getting it right) where a look at the PDB files would bring a very unpleasant surprise about the quality of structural work. Whenever I review a paper that describes a structure I look first to see if the coordinates have already been deposited. However, that still tells me very little of what is hidden beyond the accession code. There has been much discussion of whether the reviewers should be given the actual coordinates. In an ideal world they should, but even I am a realist, and I know what is possible, and what is not. However, a minimum of what I would like to see as an Editor is the header of the PDB file and a brief version of the validation report. The former will tell me if the authors were lazy and did not bother to calculate the rmsd's, or disclose what programs were used to solve the structure, where data were collected, *etc.* Many PDB data sets have all such records populated by a uniform answer "NULL". A short validation report would tell me if the structure might have some serious problems. If I see a D amino acid in an otherwise normal protein, or interatomic distances of 0.1 Å, I would like at least to ask the authors some questions before accepting their otherwise brilliant paper.

Q: ***Compared to pharmaceutical companies, what is the National Cancer Institute's approach to focusing structural studies to cancer?***

A: I am not allowed to talk about the policy of NCI without obtaining all sorts of permissions, so I better not delve too deeply into this matter. In general, it is not the mission of NCI to create drugs, but rather to create knowledge that might be the basis for drug development by pharmaceutical companies. We have fewer chemists than even some startup biotech/pharma companies, so not too much should be expected of us in this area. However, we do have some superb scholars doing fundamental research who generate data allowing understanding of the basis of cancer, delineating novel drug targets, creating new treatment methods and protocols, *etc.* Thus our role and that of the pharmaceutical companies should be considered to be different, but complementary.

Q: ***What are your thoughts on the current state of crystallographic education?***

A: What education? As far as I know, there is none. I do not believe that crystallography is still taught as a discipline, at least in the United States. Whether students will be exposed to it rigorously depends entirely on a good will of a faculty member old enough to know what he/she teaches. I am afraid that the education of most young people that actually solve crystal structures is limited to reading the manuals for HKL2000, CCP4, SHELX, COOT, or other black boxes. I am really afraid that when my generation retires, there will be few of the younger people who will be able not only to solve structures, but also understand the methods and develop them further. Hopefully I am wrong (happens often to me), but certainly I am not optimistic in this respect.

---

[1] A. Wlodawer (2007) Deposition of structural data redux. *Acta Crystallogr.* **D63**:421-423.

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 and the Skaggs School of Pharmacy and Pharmaceutical Sciences, University of California, San Diego
9500 Gilman Drive
La Jolla, CA 92093-0537

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

RCSB PDB Management

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 and the Skaggs School of Pharmacy and Pharmaceutical Sciences,
University of California, San Diego
bourne@sdsc.edu

A list of current RCSB PDB Team Members is available from www.pdb.org.

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, the National Institute of Neurological Disorders and Stroke, and the National Institute of Diabetes & Digestive & Kidney Diseases.*

RCSB PROTEIN DATA BANK
www.pdb.org

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

Return Service Requested