

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

Contents

Message from the RCSB PDB.....	1
DATA DEPOSITION AND PROCESSING	
2007 Deposition Statistics.....	2
ADIT _{Beta} Available for Testing.....	2
New Release of pdb_extract Deposition Tool.....	2
Announcement: Experimental Data Will Be Required for Depositions Beginning February 1, 2008.....	2
Structure Deposition Checklist.....	2
DATA QUERY, REPORTING, AND ACCESS	
Website Statistics.....	3
Automated Downloads of PDB Data.....	3
RCSB PDB Focus: Sorting Search Results.....	3
Positions Available at the RCSB PDB.....	3
OUTREACH AND EDUCATION	
Poster Prize Awarded at AsCA.....	4
Flyers Available in Print and Online.....	4
2008 Calendar Now Available.....	4
Web Survey: RCSB PDB Educational Resources.....	4
RCSB PDB Paper Cited More Than 5,000 Times.....	4
EDUCATION CORNER: Fruit-flavored Folding by Teresa MacDonald, Director of Education at The University of Kansas Natural History Museum.....	5
PDB COMMUNITY FOCUS: Protein Modeling at the New Jersey Science Olympiad Regionals.....	6
RCSB PDB PARTNERS, MANAGEMENT, AND STATEMENT OF SUPPORT.....	8

SNAPSHOT: JANUARY 1, 2008

48091 released atomic coordinate entries

MOLECULE TYPE	EXPERIMENTAL TECHNIQUE
44290 proteins, peptides, and viruses	40855 X-ray 6981 NMR
1829 nucleic acids	161 electron microscopy
1938 protein/nucleic acid complexes	94 other
34 other	30057 structure factor files 3793 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 2007 Annual Report explores the advances made in data deposition, query, and outreach by the RCSB PDB during the past year. In particular, the report highlights the release of the data from the wwPDB's Remediation Project that has dramatically improved the data represented within the PDB archive, as evidenced by the higher quality searching and reporting capabilities now possible on the RCSB PDB website and database.

The virus images shown on the Annual Report cover also illustrate one of the many improvements made by the wwPDB Remediation Project. Capsids were once difficult to properly construct, but can now be created directly from their PDB entries.

This report is distributed to the diverse community of PDB users in academia, industry, and education. If you would like a printed copy of this report, please send your postal address to info@rcsb.org.

This newsletter describes some of the 2008 developments already in motion:

- Beginning February 1, the deposition of experimental data (namely, structure factors and/or restraints) will be required when depositing coordinates
- A new version of ADIT, currently in beta testing, will become the production version in early 2008
- A survey regarding the educational resources available from our website is underway
- Several job openings are now available

Other developments and resources will be announced in our online weekly news and future issues of this quarterly newsletter.

Data Deposition and Processing

2007 Deposition Statistics

In 2007, 8127 experimentally-determined structures were deposited to the PDB archive.

The entries were processed by wwPDB teams at the RCSB PDB, MSD-EBI, and PDBj. Of the structures deposited in 2007, 69% were deposited with a release status of "hold until publication"; 20% were released as soon as annotation of the entry was complete; and 11% were deposited with a specific release date.

86% of these entries were determined by X-ray crystallographic methods; 13% were determined by NMR methods. 88% of these depositions were deposited with experimental data.

During the same period of time, 7304 structures were released into the archive.

ADIT_{Beta} Available for Testing



A new version of ADIT developed to improve the accuracy and consistency of data in the PDB is available for testing at deposit-beta.rcsb.org/adit.

The RCSB PDB staff ask that depositors use ADIT_{Beta} to deposit their structures and provide any feedback at deposit@deposit.rcsb.org.

The following features have been added in this version:

- Format checking
 - ADIT_{Beta} will indicate any format errors and provide suggestions for solving them
- Geometry and stereochemistry checking
 - Deposited structures will be automatically validated
- Sequence information
 - ADIT_{Beta} will check for consistency between sequence and coordinates
 - This version provides improved organization of sequence information (e.g., expression tags, mutations)
- Author and Title information
 - Entering author, title, and citation information is easier in ADIT_{Beta}

This version of the tool will become the default version of ADIT early in 2008.

New Release of pdb_extract Deposition Tool



pdb_extract is a program that minimizes errors and saves time during the deposition process by extracting key details from the output files produced by many X-ray crystallographic and NMR 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.

Version V3.004 of pdb_extract has been released, and provides:

- Added support for several new programs, for a total of 34 programs/packages with hundreds of different formats
- Improved usability, with added functions and additional error and warning messages

- Data files that follow the PDB Exchange Dictionary (PDBx) v1.045 and the Protein Data Bank Contents Guide Version 3.1

Complete details are available in the release notes at sw-tools.rcsb.org/apps/PDB_EXTRACT/latestrelease-v3.004.html.

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

Announcement: Experimental Data Will Be Required for Depositions Beginning February 1, 2008

Effective February 1, 2008, structure factor amplitudes/intensities (for crystal structures) and restraints (for NMR structures) will be a mandatory requirement for PDB deposition.

These data must be deposited at a member site of the Worldwide Protein Data Bank (www.wwpdb.org): RCSB PDB (www.pdb.org), MSD-EBI (www.ebi.ac.uk/msd), PDBj (www.pdbj.org), or BMRB (www.bmrwisc.edu).

Data may be released as soon as they have been processed and approved. There is a one-year limit on the length of time a structure and its experimental data can be put on hold, including structures that are on hold until the associated paper is published (HPUB).

This policy was developed as a result of comments and recommendations from the PDB user community, including the Commission on Biological Macromolecules of the International Union of Crystallography and the NMR Task Force, and has been endorsed by the wwPDB Advisory Committee.

Structure Deposition Checklist

It is recommended that depositors have the following items on hand when depositing a structure:

- Contact authors' names (including the Principle Investigator), e-mail addresses, postal addresses, phone and fax numbers
- Title of the deposited structure and any relevant keywords
- Citation information: authors' names, titles, and journal details if these are available
- Macromolecule names
- Biological assembly information
- Ligand names and chemical diagrams
- Sequence and chain ID for each macromolecule, including his tags or cloning artifacts that were not cleaved, and any residues not visible due to disorder
- Source information: scientific names for source organisms, expression systems, or details about synthetically produced molecules

More detailed checklists specific to X-ray, NMR, and electron microscopy (EM) depositions are available from deposit.pdb.org.

Data Query, Reporting, and Access

Website Statistics

The RCSB PDB website at www.pdb.org began to utilize the data from the wwPDB remediation project starting August 1, 2007. Access statistics for this website are given below.

MONTH	UNIQUE VISITORS	NUMBER OF VISITS	BANDWIDTH
AUG 07	87494	225428	380.69 GB
SEP 07	118631	294060	482.76 GB
OCT 07	157581	389647	608.34 GB
NOV 07	156243	373904	662.43 GB
DEC 07	120351	284523	408.06 GB

Automated Downloads of PDB Data from <ftp://ftp.wwpdb.org>

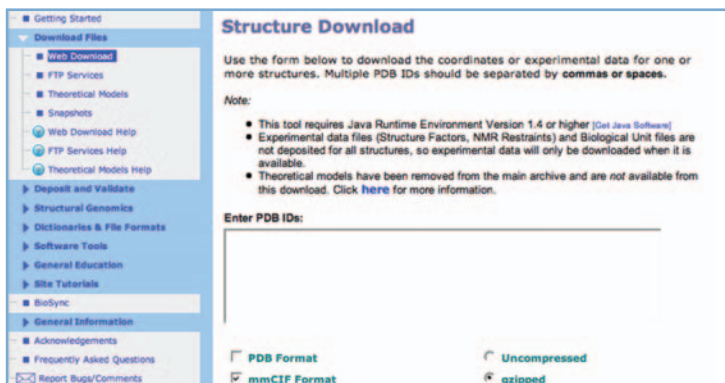
As previously announced, the PDB archive has been moved to <ftp://ftp.wwpdb.org>.

Updated weekly, this location maintains the files from the wwPDB Remediation Project and all newly released files. The archive currently contains approximately 350,000 files, including coordinate data in PDB, mmCIF, and PDBML/XML formats, and experimental data. Since the entire archive requires more than 70 GBbytes of storage, fresh downloads require a substantial amount of time. In December 2007, more than 27 million files were downloaded from <ftp://ftp.wwpdb.org>.

During the same period, approximately 2.4 million files were downloaded from the snapshot of unremediated data at <ftp.rcsb.org>. Users should be aware that this site is no longer updated, and are strongly encouraged to update any automatic scripts or bookmarks to <ftp://ftp.wwpdb.org>.

Data files from the archive can be accessed online in a variety of ways, including:

- The RCSB PDB website offers a tool to download multiple data files at www.rcsb.org/pdb/download/download.do



The online Structure Download tool is accessible from the Download Files section of the left hand menu. PDB IDs can be entered in the box provided.

- URLs for automatic downloads are described at www.rcsb.org/pdb/static.do?p=home/faq.html
- Data files are available for download from each entry's Structure Summary page.

At <ftp://ftp.wwpdb.org/pub/pdb/README>, users will find download information for downloading:

- A single file via ftp
- The entire archive via ftp
- The entire archive via rsync
- All files in a given format (PDB, CIF, XML) via rsync
- All files in a given format (PDB, CIF, XML) via ftp using tar balls

RCSB PDB Focus: Sorting Search Results



Structures matching your search constraints can be sorted according to ID, release date, residue count, resolution, and the rank of how closely they match the query (shown in green).

Following a search that produces multiple entries, the results set can be sorted by choosing 'Sort Results' from the menu on the left hand side of the page.

For most searches, the sorting options include: PDB ID, Release Date, Residue Count, Resolution and Rank (useful with keyword searches).

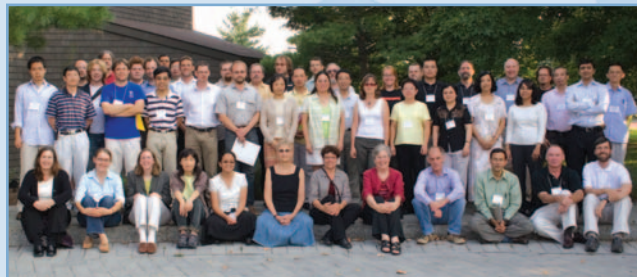
An Advanced Search by sequence (Advanced Search >> Sequence Features >> Sequence (Blast/Fasta)) allows the user to sort results by PDB ID, formula weight and E value.

Positions Available at the RCSB PDB

The RCSB PDB has the following positions available:

- Senior Scientist/Scientific Software Developer (UCSD)
- Lead Web Architect (UCSD)
- Biochemical Information & Annotation Specialist (Rutgers)

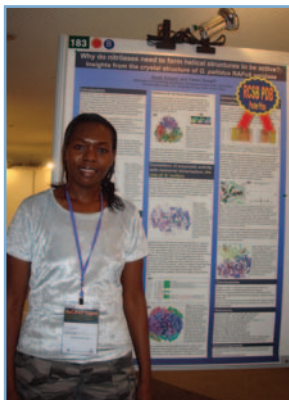
For more information, select the "Job Listings" link at www.pdb.org.



Join the RCSB PDB Team in 2008 (shown here with wwPDB members)

Outreach and Education

RCSB PDB Poster Prize Awarded at AsCA



AsCA poster prize winner
Serah Kimani

Thanks to everyone who participated in the RCSB PDB Poster Prize competition at the 8th Conference of the Asian Crystallographic Association (AsCA) that took place November 4 through 7, 2007 in Taipei, Taiwan.

The RCSB PDB Poster Prize is awarded to the best student poster related to macromolecular crystallography. At AsCA, the judges interviewed the finalists for the prize, and considered the engagement of the student in the work and their understanding of it; the clarity of the presentation in terms of the hypothesis being tested; the appropriateness of the approach; and the justification of the conclusions drawn based on the data presented.

The award went to Serah Kimani for the poster "Why do nitrilases need to form helices to be active?" (Trevor Sewell, Serah Kimani (University of Cape Town, South Africa), and Muhammed Sayed (University of the Western Cape, South Africa)). Serah will receive a copy of *International Tables Volume B - Reciprocal space* and a subscription to *Science*.

Judges: Mitchell Guss (University of Sydney), Sine Larsen (European Synchrotron Radiation Facility), and Mike Lawrence (Walter and Eliza Hall Institute of Medical Research).

Poster Prize Chairman: Jill Trehwella (University of Sydney)

Special thanks to the AsCA organizers and the Program Committee Chairman Se Won Suh for their assistance with organizing the prize.

Congratulations to all of the 2007 RCSB PDB Poster Prize award winners.

Flyers Available in Print and Online

The News & Publications web page, accessible from www.pdb.org, offers links to various RCSB PDB publications, including newsletters and annual reports. Informational brochures describe different educational features, including the Sea of Genes exhibit in the Birch Aquarium at Scripps Research Institute that explored proteins related to underwater creatures.

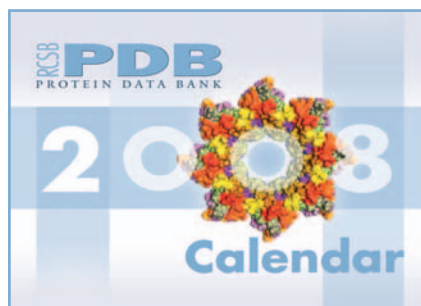
Other brochures help users explore the RCSB PDB. *5 Easy Steps for Structure Deposition* describes the tools that facilitate NMR and X-ray crystal structure deposition and validation for depositors.

A *General Information* trifold provides an overview of the RCSB PDB project, and includes information about data deposition,



data query and reporting, *Molecule of the Month*, structural genomics, wwPDB, and outreach and education resources.

All of these materials can be downloaded from the RCSB PDB site. To receive printed copies of any flyers, please send your postal address and request to info@rcsb.org. Multiple copies may be requested.

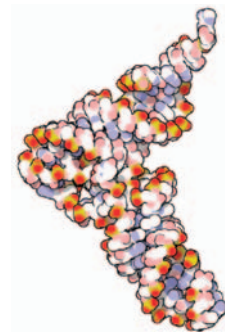


2008 Calendar Now Available

A calendar showcasing PDB structures is now available online. Printed copies are also available via info@rcsb.org.

Web Survey: RCSB PDB Educational Resources

Do you use the *Molecule of the Month*? Teach classes? Use the RCSB PDB when working with students? Then we want to hear from you!



A temporary tattoo of tRNA can be yours!

The RCSB PDB is looking for feedback about the educational resources available from our website, and in particular, the types of educational activities and resources that are of interest to our users.

We have created a short online survey at www.zoomerang.com/survey.zgi?p=WEB226ZPP48MNM that should only take a few minutes to answer.

We greatly appreciate your participation in this survey. As a token of appreciation, we will send temporary tattoos of tRNA to survey respondents who send their postal address to info@rcsb.org. The survey will be closed by March 1, 2008.

RCSB PDB Paper Cited More Than 5,000 Times

According to Essential Science IndicatorsSM ¹, the RCSB PDB primary reference is ranked #4 in the top cited Biology and Biochemistry papers of the past ten years. "The Protein Data Bank,"² published in the 2000 Database Issue of *Nucleic Acids Research*, has been cited more than 5,000 times.

In-cites magazine featured this paper in an interview with RCSB PDB Director Helen M. Berman at www.in-cites.com/papers/HelenBerman.html.

¹ Essential Science IndicatorsSM; www.in-cites.com/rsg/esi

² The Protein Data Bank. (2000) *Nucleic Acids Research*, 28, pp. 235-242. nar.oupjournals.org/cgi/content/abstract/28/1/235



EDUCATION CORNER: Fruit-flavored Folding

Teresa MacDonald, Director of Education at The University of Kansas Natural History Museum

“Frying Pickles and Flying Marshmallows” was one of the news headlines inspired by our museum’s annual science event, titled *Playing With Your Food*.¹ Over six days, more than 4,000 visitors explored science through demonstrations and activities that all used food in some way—such as Cartesian divers, gelatin optics and exploding cornstarch. During all of our events, we offer a range of activities to serve a broad audience, and try to incorporate some more challenging concepts or less familiar science ideas into the visitor experience. During *Playing With Your Food*, we used colored licorice inside napkin rings to demonstrate the ‘tube within a tube’ body plan found in most animals, talked about the biogeography of worms in North America, and used Fruit by the Foot™ to illustrate protein folding.



Demonstration center



Close-up of protein model

We searched for images online that were created with protein modeling software (cartoon images) and followed these to create a three-dimensional model of the tertiary structure of the ovalbumin protein found in eggs. A folded protein model was made using a thin wire frame wrapped in Fruit by the Foot™. An unfolded, twisted mass of Fruit by the Foot™ was used to represent the denatured egg protein. These two models, along with a raw and cooked egg, were used to teach visitors about: (1) what proteins are; (2) how they are made; and (3) the different levels of protein structure.

Visitors are often familiar with some elements of science topics, but can struggle with making connections between, or synthesizing, different pieces of information. One misconception that I encounter on a regular

basis is that DNA is only found in blood, saliva, and gonads because of the many references to crime scene investigations and paternity suits in the popular media. Our events provide an opportunity to make connections between new ideas and the concepts familiar to visitors. We felt that visitors were likely to have heard of proteins and that most would recognize eggs as being a good source of protein, but that the majority of visitors would not have a broader understanding of proteins, such as

TERESA MACDONALD

(tmacd@ku.edu) is the Director of Education at the University of Kansas Natural History Museum and Biodiversity Research Center, and an instructor in the Museum Studies graduate program. She holds a Bachelors degree in physical anthropology, and a Masters degree in vertebrate paleontology. She has over twelve years experience in the field of science education and public understanding of science. Her experience spans five countries on three continents and includes work in museums, science centers, schools and universities. MacDonald is the outreach director for the EPSCoR-funded particle physics education project, Quarked!, and is the Principle Investigator for the NSF-funded Understanding the Tree of Life project.



Tyson Pyle assisting visitors in making their own DNA jewelry



Teresa MacDonald and Dawn Kirchner extracting DNA from strawberries for visitors

¹ Frying pickles and flying marshmallows: museum says it's OK to play with food. *KU news release*, March 7, 2007 www.news.ku.edu/2007/march/7/food.shtml

their varied roles in the body, the relationship between DNA and proteins, or the idea of protein folding.

The Fruit by the Foot™ protein model piqued visitors' interest—they wanted to know what it was and why we would make something like this. We typically began the discussion by asking visitors about what they already knew about proteins and their related knowledge or experiences—e.g., what they knew about DNA coding, whether they had ever cooked eggs, eaten cheese or yoghurt. All visitors, children and adults, had heard of proteins and the majority suggested that “you should eat them them to make you strong.” Few had made any connections between DNA, amino acids, and proteins, or were aware of protein folding. Protein folding was introduced by looking at what happens to egg proteins when you “cook” them—bonds break, proteins unfold, and new bonds form between proteins to produce the familiar hard “egg white.” The model helped to illustrate the secondary—specifically alpha helices and beta pleated sheets—and tertiary structure of proteins. This was then related to the importance of protein folding in studying some human diseases.

Whenever possible, we try to link activities within and between our events. Activities that were related to the protein demonstration included: (1) extraction of DNA from strawberries; (2) DNA jewelry that used colored

beads and pipecleaners to create DNA strands of triplet sequences that coded for letters of the alphabet rather than amino acids; and (3) a Gummy Fish Genetics display which used regular and mini-gummy fish to demonstrate simple Mendelian inheritance. Future links could include antibodies and enzymes, and opportunities for visitors to make their own models.



Selecting a fish from the parent pool for Gummy Fish Genetics.

For more information about The University of Kansas Natural History Museum and Biodiversity Research Center, please see www.nhm.ku.edu.

PDB Community Focus

Protein Modeling at the New Jersey Science Olympiad Regionals



Calmodulin as illustrated in the *Molecule of the Month*.

Many models of the structure calmodulin were built by high school students for the RCSB PDB-sponsored *Protein Modeling* event at the Northern and Central New Jersey Science Olympiad.

Science Olympiad tournaments, which take place across the country, consist of several individual and team events that students prepare for during the year. Medals are awarded for the top finishers in each event and for overall performance. During the competition, teams demonstrate their diverse skills and knowledge in many different events. In *Forensics*, teams identify polymers, solids, and fibers at a crime scene, while in *Write It, Do It*, students compose a description of a structure that will be the only guide used by their other team members to recreate the same shape, sight unseen, with raw materials.

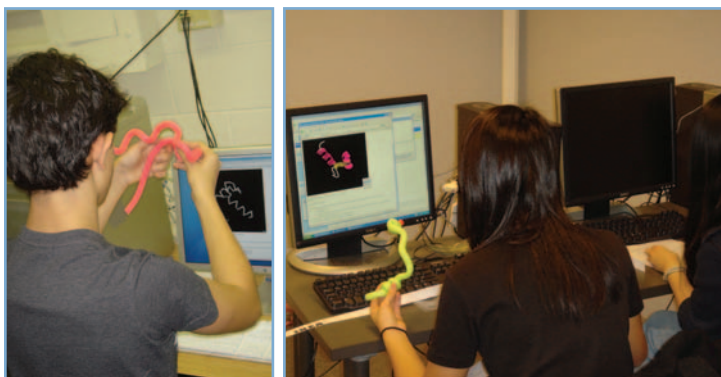
In 2008, *Protein Modeling* is being held as a trial event at Science Olympiads in Florida, Indiana, Massachusetts, New Jersey, and Wisconsin. Team alternates can only participate in trial events, which typically do not count towards the overall score. In New Jersey, scores in protein modeling were used in calculating a team's total score.

This year's protein modeling competition has three components. Students first build a model of the full calmodulin structure (entry 1c1l), and bring it in the morning to be impounded for judging. Teams are

encouraged to include additions and an abstract that help to illustrate the function of calmodulin in this model. This model is worth up to 40 points out of a possible 100.

At the event itself, teams build a portion of PDB entry 1c1l with a Mini-Toober (30 points). They also answer questions in a written exam about the structure, function, importance, and history of the modeled protein (30 points). For all sections of the event, students use the *Molecule of the Month*, the PDB entry, Jmol (jmol.sourceforge.net/), and 1c1l's Structure Explorer page.

In addition to providing the kits, the *Protein Modeling* event in New Jersey is judged by the annotators and computer programmers of the RCSB PDB. They review each structure by comparing it to a 3D model generated directly from the coordinates and using a model built directly from the

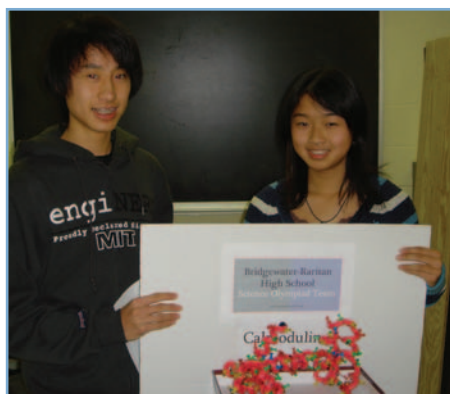


At the event, teams modeled a portion of PDB entry 1c1l.



RCSB PDB judges also met with students to discuss their structures.

structure's PDB file and a predetermined rubric that awards points for accurate depictions of the protein's features. For example, judges look to see if the N- and C-termini are labeled properly and carefully consider the helices of the model. They also consider if the main functional and structural features of the protein are illustrated in the model. The written exam asks questions based upon the entry's Structure Summary page, the *Molecule of the Month* entry, and beyond.



Bridgewater-Raritan's Daniel Zhang and Amy Song with their model of the full structure of calmodulin. Their on-site model of a section of PDB entry 1c1l received a perfect score.



Northern champions Tim Kunisky and Collin Stocks. Their model was enhanced with masking tape calcium ions. The team from Livingston High School also came in first place in the overall competition.

At the Central New Jersey Regional held at Princeton University (January 8, 2008), Bridgewater-Raritan High School came in first; South Brunswick High School, second; and West Windsor-Plainsboro High School North, third.

At the Northern New Jersey Regional held at New Jersey Institute of Technology (January 17, 2008), Livingston High School came in first; Westfield High School, second; and Bergen County Academies, third.

The Science Olympiad is an international non-profit organization devoted to improving the quality of science education, increasing student interest in science and provid-

RCSB PDB Judges Looked For ...

- Does the overall shape of the model resemble the structure?
- Are differences shown in the two domains at the N- and C-termini?
- What are the secondary structure elements of this structure?
- How many helices are there?
- What handedness was used to create the helices?
- Does the model hint at what binds to this protein, and where?

Entries were compared to a model generated directly from the coordinates and a pre-determined rubric.

ing recognition for outstanding achievement in science education by both students and teachers.

The 2008 NJSO (www.njscienceolympiad.org) is presented by the New Jersey Science Teachers Association and the New Jersey Science Education Leadership Association. Special thanks to the Center for BioMolecular Modeling at the Milwaukee School of Engineering (www.rpc.msoc.edu/cbm) for the design of this event. Kits similar to those provided for this event may be purchased from www.3dmoleculardesigns.com.

Protein Modeling State Finals

The state finals will take place March 11, 2008 at Middlesex County College.

Students may bring their model from the regional competition for the prebuild section, or they can build a new model. Onsite, they will build a different section of PDB entry 1c1l.

To prepare, teams should definitely explore the resources at education.pdb.org/olympiad.

Questions about the NJSO Protein Modeling trial event should be sent to buildmodels@deposit.rcsb.org.

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