Spring 2011 • Number 49

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

RCSB PDB PROTEIN DATA BANK NEWSLETTER

www.pdb.org • info@rcsb.org

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

Contents

SNAPSHOT: APRIL 1, 2011

72,104 released atomic coordinate entries

ENTRIES BY MOLECULE TYPE		ENTRIES BY EXPERIMENTAL TECHNIQUE	
66,726	proteins, peptides, and viruses	62,734	X-ray
3,105	protein/nucleic acid complexes	8,819	NMR
2,234	nucleic acids	364	electron microscopy
39	other	33	hybrid
		154	other

RELATED EXPERIMENTAL DATA FILES

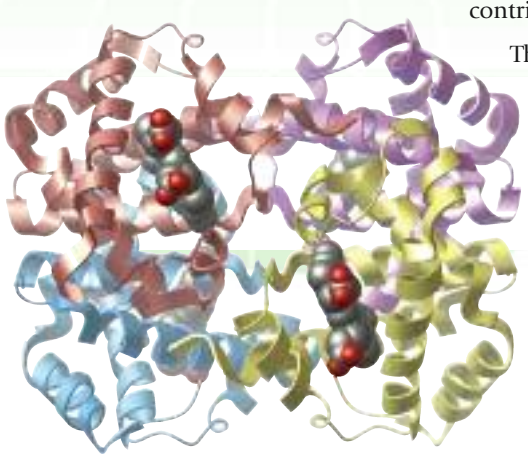
52,142	structure factors
6,118	NMR restraints
28	NMR chemical shifts

A Special Symposium Celebrating the 40th Anniversary of the PROTEIN DATA BANK

October 28 - 30, 2011
Cold Spring Harbor Laboratory

In June 1971, a symposium on *Structure and Function of Proteins at the Three Dimensional Level* was held at Cold Spring Harbor Laboratory. That meeting defined the beginning of the Protein Data Bank as an archive for the experimentally-determined 3D structures of biological macromolecules.

In October 2011, the Worldwide Protein Data Bank (wwPDB) will host a scientific symposium celebrating the 40th anniversary of the inception of the PDB, and the many scientific contributions it archives.



The program will showcase the scientific impact made by structural biology during the past 40 years with a distinguished panel of scientists who have been instrumental in the development of the PDB and structural biology.

The meeting will begin with an evening reception and plenary session on Friday, and conclude with lunch on Sunday.

WORLDWIDE PDB PROTEIN DATA BANK

For information about registration and accommodations, please see the meeting website at http://meetings.cshl.edu/meetings/pdb40.shtml.

Data Deposition and Annotation

Deposition Statistics

In the first quarter of 2011, 2205 experimentally-determined structures were deposited to the PDB archive. The entries were processed and annotated by wwPDB teams at the RCSB PDB, PDBe, and PDBj.

Of the structures deposited, 80.4% were deposited with a release status of "hold until publication"; 17.6% were released as soon as annotation of the entry was complete; and 2.0% were held until a particular date. 92.9% of these entries were determined by X-ray crystallographic methods; 6.2% were determined by NMR methods.

During the same time period, 1914 structures were released in the PDB.

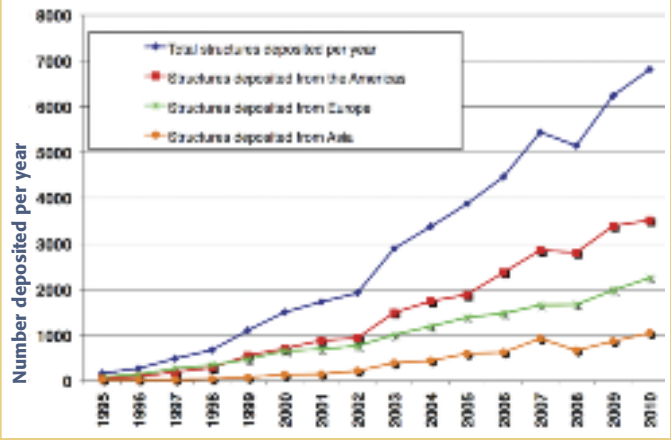
BioSync Synchrotron News

Diamond Celebrates Milestone 500 Structures

Diamond Light Source celebrated their 500th PDB entry by highlighting some of the structures determined along the way at www.diamond.ac.uk/Home/Media/LatestNews/17_03_11.html

Online Guide to High Energy Data Collection Facilities

BioSync provides current information about synchrotrons located worldwide. Visit biosync.sbkb.org to explore the types of services and facilities available at each location and view detailed statistics about related PDB depositions.



Synchrotron structures deposited in the PDB (total and by region). More than 47,000 structures have been solved by the worldwide synchrotron community; close to 23,000 of these structures are from beamlines located in the US. Image from BioSync.

wwPDB News

PDB Archive Version 4.0 to be Released June 2011

The wwPDB has performed an ambitious review of the PDB archive and has created a new set of corrected files that will be released in June 2011. Version 4.0 PDB entries will follow the PDB Exchange Dictionary v.4.0.

The entire archive has been reviewed and remediated with the objective of tackling complex problems. These include the representation of biological assemblies, residual B factors, peptide inhibitors and antibiotics, and entries in the nonstandard crystal frame. A description of the review and resulting changes and corrections is available at www.wwpdb.org.

Time-stamped Copies of PDB Archive Available via FTP

A snapshot of the PDB archive (ftp.wwpdb.org) as of January 3, 2011 has been added to ftp://snapshots.wwpdb.org. Snapshots have been archived annually since January 2005 to provide readily identifiable data sets for research on the PDB archive. The script at ftp://snapshots.wwpdb.org/rsyncSnapshots.sh can create a local copy of a snapshot or sections of the snapshot.

Data Query, Reporting, and Access

Store Personal Annotations with MyPDB

MyPDB lets users create a personalized version of the RCSB PDB accessible from any computer and from PDB*Mobile*.

Using the MyPDB widget in the left-hand menu, users can create new accounts or log in.

- The MyPDB **Saved Query Manager** stores RCSB PDB searches (keyword, sequence, ligand, etc.) and composite queries built with Advanced Search. These saved queries can be run at the click of a button.

Stored searches can be set to run automatically. Email alerts (weekly or monthly) will be sent when matching entries are released.

- **Personal Annotations**. Users can save personal annotations and notes on the Structure Summary tab of any PDB entry, and add structures to a favorites list. The Personal Annotations summary page provides easy access to all of these tagged structures and annotations.
- **User Account**. Personal information (name, email address, account password, country, user type) can be updated at any time. All MyPDB account information is kept private and secure.

New Features at www.pdb.org

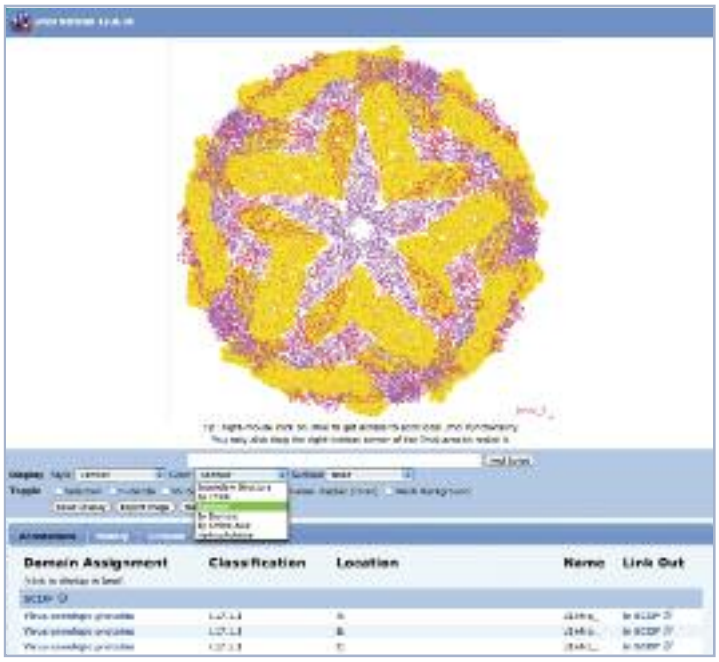
Improved Sequence tab features, new searches and reports, and a special educational view called PDB-101 have been added with the latest website release. See the *What's New* page for details and examples.

Structure Summary Pages

Every PDB entry has a *Structure Summary* page that offers a portal to tools, resources, and related links. Tabs on the initial summary page can be used to toggle between different topics. Users can also access different molecular viewers and external resources from these pages. A few of the many available options are highlighted below.

Quick Jmol Views for Exploring Macromolecular Structures

The Jmol viewer is used by beginners and experts to interactively explore PDB structures.

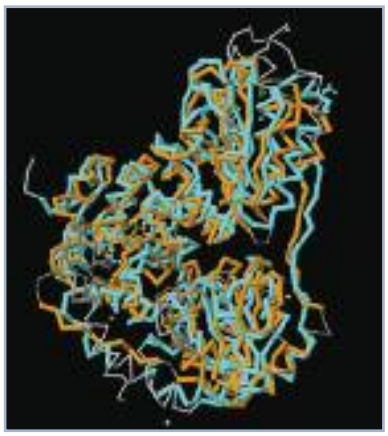


*PDB entry 1k4r[1] as seen in Jmol (**jmol.sourceforge.net**).*

Frequently-used Jmol options are available in pull-down menus and buttons, including the display style, color, and surface. For more detailed usage, click the right mouse button to access the Jmol menu, or enter scripts into the text box. Domain assignments from SCOP, DP, and PDP can also be displayed by selecting from the list provided.

Select View in Jmol from any entry's Structure Summary page to access the interactive viewer.

Structural Neighbors and the 3D Similarity Tab



Comparison of 1q6z[2] (orange) with structural neighbor 3hww[3] (cyan).

Proteins can have various degrees of similarity. If two proteins have highly similar amino acid sequences, it is generally assumed that they are closely related evolutionarily. As the evolutionary distance increases, the degree of similarity usually drops. Even if the sequence similarity is low, proteins may have similar functions and 3D structures. Detecting remote similarities, a core structural bioinformatics technique, is important in the study of functional and evolutionary relationships between protein families.

The RCSB PDB offers tools that quickly identify 3D protein sequence neighbors. For each PDB entry, the **3D Similarity tab** lists the representative entries with 40% sequence similarity that are found using the jFATCAT-rigid[4] algorithm. As an example, look at the 3D similarity tab for entry 1q6z. Representative protein chains are used since calculation of a real all vs. all comparison would require a great amount of CPU time.

Novel domain architectures and unexpected structural similarities can be detected by analyzing structural alignments. As an example, entry 3hww is one of the top ranking structural neighbors of 1q6z (chain A). Clicking on view from the Structure Similarity table will show a summary view of the alignment.

3hww has an RMSD of 3 Å based on the Cα positions, while the two protein chains are only 14% identical by sequence. 1q6z is a benzoylformate decarboxylase (EC number 4.1.1.7), while 3hww is a 2-succinyl-5-enolpyruvyl-6-hydroxy-3-cyclohexene-1-carboxylate synthase (EC number 2.2.1.9). Despite the low sequence identity and divergence in function, the high structural similarity indicates that both proteins evolved from a common ancestor.

Create Publication-Quality Molecular Images

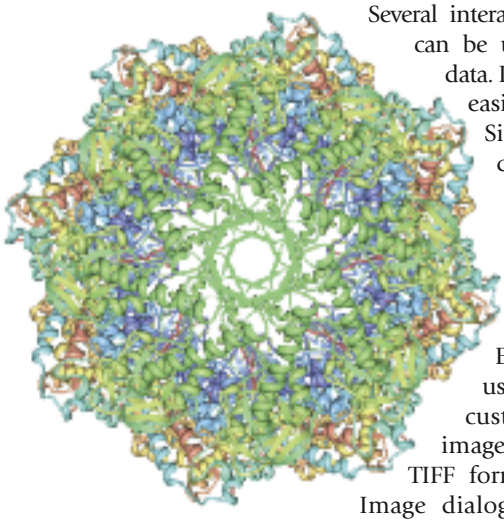


Image of 3izi[5] created using Protein Workshop

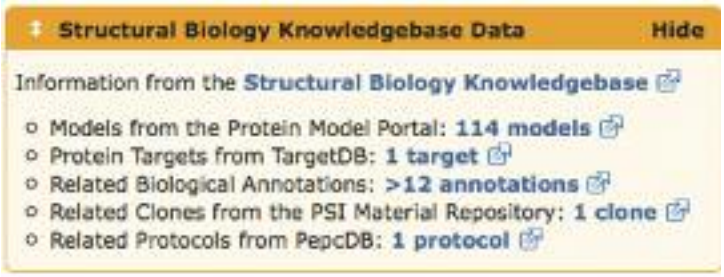
Several interactive, Java-based[6] tools can be used to visualize PDB data. Protein Workshop offers easily customized views; Simple Viewer offers a quick ribbon display; and Ligand Explorer visualizes the interactions of bound ligands in protein and nucleic acids structures.

Each program can be used to create and save custom high-resolution images in JPEG, PNG, and TIFF formats. Using the Save Image dialog box from the File menu, users can specify the width and height of an image in pixels, inches, or millimeters.

Structural Biology Knowledgebase Widget

Structure Summary pages host widget boxes of available data loaded from resources such as SCOP (scop.mrc-lmb.cam.ac.uk/scop) and CATH (www.cathdb.info). These boxes are in orange to indicate the data are annotations from external sources.

A new widget links to related information at the Structural Biology Knowledgebase (SBKB; sbkb.org), a free, comprehensive resource that integrates the results of structural biology and genomics efforts with other publicly available biological information to facilitate further research. The SBKB widget loads data about available models, protein targets, related biological annotations, related clones, and protocols with links to SBKB reports page.



The SBKB widget loads data about available models, targets, biological annotations, and more.

Website Statistics

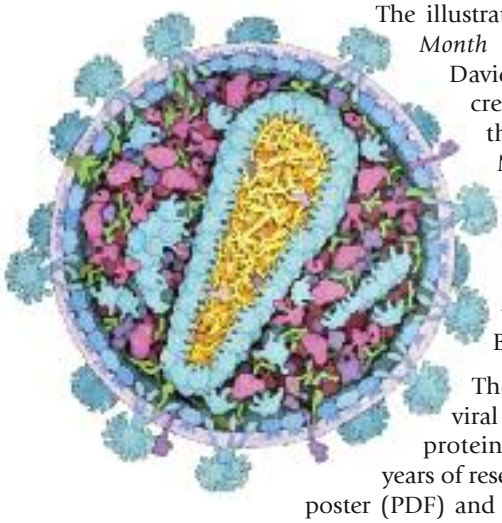
Access statistics for the first quarter of 2011 are shown.

Month	Unique Visitors	Number of Visits	Bandwidth
JANUARY	210986	522651	982.26 GB
FEBRUARY	213862	526896	1114.34 GB
MARCH	239640	598545	1004.99 GB

Outreach and Education

The Structural Biology of HIV: Online Resources for Understanding

A painting of the structural biology of the HIV virus is the focus of a new poster and interactive website.



The illustration, by *Molecule of the Month* author and illustrator David Goodsell, was initially created to commemorate the NIH's 25th Annual Meeting of the *Groups Studying the Structures of AIDS-Related Systems and Their Application to Targeted Drug Design* that was held March 28-30 in Bethesda, MD.

The structural proteins, viral enzymes, and accessory proteins revealed through 25 years of research are highlighted in a poster (PDF) and interactive animation of the virus. Clicking on a protein in the animation reveals a description of the structure and links to the PDB entry.

These features are available from the Educational Resources section of the RCSB PDB. Additional information about HIV-related proteins is found in *Molecule of the Month* features on HIV protease, integrase, and reverse transcriptase.

Meetings and Events

Students of all ages were introduced to 3D macromolecular structure at the *San Diego Science Festival's* **Expo Day** on March 26. Visitors learned about the basic building blocks of life by constructing a model of DNA and translating its sequence into protein. The RCSB PDB's exhibit booth also displayed animations, and offered posters and other materials about these fascinating structures.

Teachers at the NSTA meeting picked up copies of this DNA origami model. Copies can be downloaded from the RCSB PDB's Educational Resources.

RCSB PDB met with researchers and educators while exhibiting at meetings of the *Biophysical Society* (March 5-9; Baltimore, MD) and the *National Science Teachers Association* (NSTA; March 9-12; San Francisco, CA). Attendees stopped by the exhibit booth to meet with staff, see demonstrations of new features, and to take home the latest RCSB PDB materials.

In addition to exhibiting, the RCSB PDB sponsored virus model building at the **Family Science Days** held as part of the annual meeting of the *American Association for the Advancement of Science* (AAAS; February 17-21; Washington, DC).

The RCSB PDB will be participating at the *Annual Meeting of the American Crystallographic Association* (May 28-June 2; New Orleans, LA), the *Annual International Conference on Intelligent Systems for Molecular Biology and 10th European Conference on Computational Biology* (July 17-19; Vienna, Austria), and the *General Assembly and Congress of the International Union of Crystallography* (August 22-30, Madrid, Spain).

The RCSB PDB is also looking forward to the *Special Symposium Celebrating the 40th Anniversary of the Protein Data Bank* (October 28-30, Cold Spring Harbor, NY) described in this issue's *Message from the RCSB PDB*.

Visitors built DNA models as part of Expo Day at Petco Park.

Visitors to the AAAS meeting created virus models.

NJ Science Olympiad Protein Modeling Results

Science Olympiad tournaments consist of a series of events that test student knowledge in biology, earth science, chemistry, physics and technology. In New Jersey, protein modeling was one of the 25 events for high school teams at the regional and state competitions.

Students demonstrate their knowledge of proteins shown to have reprogrammed adult cells into pluripotent stem cells in this event by building 3D models and completing a written exam.

Teams submitted hand-built 3D models of the zinc finger protein found in PDB entry 2wbu on the morning of the event. The models represented the protein backbone, with additional points awarded for structures with details that highlighted important parts of the structure (such as zinc atoms). During the event itself, the students quickly built a model onsite (POU/HMG/DNA ternary complex found in entry 1gt0 at regionals and homeobox protein Nanog in entry 2kt0 at state) and answered questions on a written exam. The *Molecule of the Month*, Jmol, and other resources are used to help prepare for this event.

RCSB PDB members judged this competition, and met with teams at the end of the day to discuss results. 48 teams competed in this event, creating 144 models along the way.

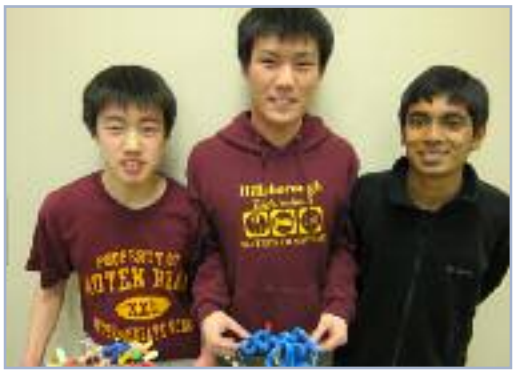
Teams that performed well overall at the regional level participate in the state tournament. The top scoring teams at each event were:

Southern Regional
1. Cherry Hill East (85 points)
2. J.P. Stevens (Team 2, 84)
3. Lawrenceville (81)

Central Regional
1. Hillsborough (88.5)
2. Westfield (86)
3. Union County Vo-Tech (83)

Northern Regional
1. West Windsor-Plainsboro HS South (Team 2, 95.5)
2. West Windsor-Plainsboro HS South (Team 1, 91)
3. Livingston (88.5)

Southern Regional Champions: Cherry Hill East High School



Central Regional Champions (shown at State): Hillsborough High School

Northern Regional Champions: West Windsor-Plainsboro South High School

State Top Scoring Teams
1. J.P. Stevens HS (98.75)
2. Princeton (95)
3. Livingston HS (93.5)

Protein modeling state champions from J.P. Stevens High School (Edison): Andy Shi, Joet Bagga, and Sam Zhang.

Full results are posted, along with information and resources about this NJ event, at education.pdb.org and on Twitter at @buildmodels.

Protein modeling will be one of several events at Science Olympiads across the country. Teams are recognized for their performance in individual events and in the overall tournament. For more information, see cbm.msoe.edu/stupro/so/index.html.

Many thanks to the RCSB PDB judges (Batsal Devkota, Brian Hudson, Buvaneswari Narayanan, Chengua Shao, Huangwang Yang, and Christine Zardecki), the NJ Science Olympiad organizers and volunteers, the host colleges, and to the MSOE Center for BioMolecular Modeling for the materials and design of this event.

Papers Published

Recent articles include:

- The Protein Data Bank: exploring biomolecular structure (2010) *Nature Education* (www.nature.com/scitable) **3**:39.
- Precalculated protein structure alignments at the RCSB PDB website (2010) *Bioinformatics* **26**: 2983-2985 doi: 10.1093/bioinformatics/btq572
- The RCSB Protein Data Bank: redesigned web site and web services (2011) *Nucleic Acids Res.* **39**: D392-D401 doi:10.1093/nar/gkq1021
- Quality assurance for the query and distribution systems of the RCSB Protein Data Bank (2011) *Database: The Journal of Biological Databases and Curation* doi: 10.1093/database/bar003

A list of all RCSB PDB publications can be found in the **News & Publications** section of the website. Citation information is also available.

Education Corner by Richard Tempsick, Academy of Allied Health and Science

Building Protein Structures and the *Molecule of the Month*

Understanding proteins and their 3D shapes presents a real challenge for high school biology students. Just the definitions of proteins and protein folding, available from textbooks and the Internet, can lead students to avoid these topics and scramble for their Xboxes.

For some examples:

Protein folding is the physical process by which a polypeptide folds into its characteristic and functional three-dimensional structure from a random coil.[7]

Most proteins consist of linear polymers built from a series of up to 20 different L-α-amino acids.[8]

All proteinogenic amino acids possess common structural features, including an alpha carbon to which an amino group, a carboxyl group, and a variable side chain are bonded.[8]

Trying to make sense of this information and having a real understanding of the function of a protein is difficult. The complexity of the new protein data that is being published at hyper-speed has left most high school students in the dust. Not only is it hard to explain protein biochemistry, but now we also have the genome, glycome, and proteome. That is why resources like the RCSB PDB are extremely valuable in trying to bring students up to speed and get them excited about the possibilities for the future. The work of Phil McFadden in his protein portraits course was an awesome view of various protein structures.[9] Using protein model competitions is a great way to get students excited about building models, and it shows the fun and creativity of participating in this activity.

Model building is extremely useful and any 3D structure can go a long way in motivating high school students to get involved and excited about proteins. My students look at those structures and think they would like to try to build those sorts of models.

RICHARD TEMPSICK, *B.S., M.A., is a biology teacher at the Monmouth County Academy of Allied Health and Science (AAHS) in New Jersey. AAHS is a 4-year, comprehensive, theme-based high school with a focus on medical science. Tempsick teaches 12th grade* Principles of Cell and Molecular Biology *and 9th grade* Biology. *He is also an adjunct faculty member at Georgian Court University, where he teaches* Principles of Cell and Molecular Biology. *He has conducted research on nitric oxide synthase at Robert Wood Johnson Medical School in the Department of Pharmacology. For the past fifteen years he has been focused on developing the biology curriculum at AAHS and working with students to pursue further education in the biological sciences.*

Richard Tempsick and his Principles of Cell and Molecular Biology *class.*

The RCSB PDB is a resource that encourages model building and has user-friendly tutorials that guide students through various simulations. My favorite feature is the *Molecule of the Month* by David Goodsell. These molecules are presented in such a way that enables students to really understand multiple aspects of a particular protein. The *Molecule of the Month* for April 2003 was RNA polymerase.[10] This entry clearly explains the structure and function of RNA polymerase, and describes how it accurately copies genetic information. High school students typically learn about RNA polymerase and its role in transcribing DNA, but the RCSB PDB takes it one step further

Building pipe cleaner proteins.

by including a section on how RNA polymerase can be poisoned. Students are able to see how a toxin can bind to RNA polymerase and interfere with its function. This, of course, leads to a new discussion on where that toxin would exist in nature, and what happens if you get exposed to it? Is there a treatment? That then leads to the discussion about the death cap mushroom–and now, you're introducing students to fungal biology.[11] When you get the opportunity to link the structure and function of a protein to medical conditions, this gets their attention and they see the significance of the shape of one molecule fitting the shape of another. I recommend the *Molecule of the Month* resource to any teacher who works with proteins and is looking for real practical applications.

A model of green fluorescent protein.

Green fluorescent protein is a very popular protein to work with in high school biology classes. You can easily express the gene for GFP in *E. coli* and watch the bacteria fluoresce in culture. For the next step, you can break open the *E. coli*, run extracts through a column, and see the protein fluoresce as it separates in the matrix. The fractions can be tested for highest protein concentration by scanning for fluorescence. This models the way proteins are worked on by researchers in the lab and the RCSB PDB can be used to explore the structure even further. The *Molecule of the Month* on GFP[12] does a great job of explaining the structure and its potential uses.

My personal favorite structure is the G-protein coupled receptor. The *Molecule of the Month* has columns on G-proteins[13] and adrenergic receptors.[14] This image is our attempt to build a G protein-coupled receptor structure using a tug-of-war rope.

The class builds a G protein-coupled receptor.

The images on the previous page illustrate the fun you can have using pipe cleaners and the PDB as a guide to build protein structures. It might have been way more fun to have baked egg-covered pretzels in the shape of ovalbumin (as Danika Kusuma did in Phil McFadden's course),[9] but we stuck with pipe cleaners.

The common theme, however, that arises in building these proteins is that whatever structure you create (and using whatever materials), you can easily see how domains develop that allow for some other molecule to fit that shape and interact with your protein. This is probably the best message you can impress upon high school biology students, and the RCSB PDB is a resource that gives anyone who teaches or wants to learn about proteins the opportunity to explore that concept.

References

1. R.J. Kuhn, W. Zhang, M.G. Rossmann, S.V. Pletnev, J. Corver, E. Lenches, C.T. Jones, S. Mukhopadhyay, P.R. Chipman, E.G. Strauss, T.S. Baker, J.H. Strauss (2002) Structure of dengue virus: implications for flavivirus organization, maturation, and fusion. *Cell* **108**:717-725.
2. A.K. Bera, N.L. Anderson, M.S. Hasson. High resolution structure of E28A mutant benzoylformate decarboxylase from *Pseudomonas putida* complexed with thiamin thiazolone diphosphate. **doi:10.2210/pdb1q6z/pdb**
3. A. Priyadarshi, E.E. Kim, K.Y. Hwang (2009) Structural and functional analysis of vitamin K2 synthesis protein MenD. *Biochem. Biophys. Res. Commun*. **388**: 748-751.
4. jFATCAT: Based on FATCAT (Y. Ye, A. Godzik (2003) Flexible structure alignment by chaining aligned fragment pairs allowing twists *Bioinformatics* 19: ii246-ii255), BioJava (R.C.G. Holland, T. Down, M. Pocock, A. Prlić, D. Huen, K. James, S. Foisy, A. Dräger, A. Yates, M. Heuer, M.J. Schreiber (2008) BioJava: an open-source framework for bioinformatics *Bioinformatics* **24**:2096-2097), and Jmol (**jmol.sourceforge.net**).
5. N.R. Douglas, S. Reissmann, J. Zhang, B. Chen, J. Jakana, R. Kumar, W. Chiu, J. Frydman (2011) Dual action of ATP hydrolysis couples lid closure to substrate release into the group II chaperonin chamber. *Cell* **144**: 240-252.
6. 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.
7. B. Alberts, A. Johnson, J. Lewis, M. Raff, K. Roberts, P. Walters (2002) *Molecular Biology of the Cell* (Fourth Edition). New York and London: Garland Science.
8. A. Lehninger, D.L. Nelson, M.M. Cox (2005) *Lehninger's Principles of Biochemistry* (Fourth Edition). New York, New York: W. H. Freeman and Company.
9. P. McFadden (2010) Undergraduate Course: Protein Portraits. *RCSB Protein Data Bank Newsletter* **46**:8-11.
10. D.S. Goodsell, **doi: 10.2210/rcsb_pdb/mom_2003_3**
11. D. Arorad (1986). *Mushrooms Demystified: A Comprehensive Guide to the Fleshy Fungi*. Berkeley, California: Ten Speed Press.
12. D.S. Goodsell, **doi: 10.2210/rcsb_pdb/mom_2003_6**
13. D.S. Goodsell, **doi: 10.2210/rcsb_pdb/mom_2004_10**
14. D.S. Goodsell, **doi: 10.2210/rcsb_pdb/mom_2008_4**

RCSB PDB PROTEIN DATA BANK NEWSLETTER

Web: www.pdb.org
Email: info@rcsb.org
FTP: ftp.wwpdb.org

RCSB PDB Partners

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

RUTGERS

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

UCSD SDSC SKAGGS SCHOOL of PHARMACY and PHARMACEUTICAL SCIENCES

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

WORLDWIDE PDB PROTEIN DATA BANK

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. MARTHA QUESADA, Deputy Director
Rutgers, The State University of New Jersey
mquesada@rcsb.rutgers.edu

DR. PHILIP E. BOURNE, Associate 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 Institute of Neurological Disorders and Stroke, and the National Institute of Diabetes & Digestive & Kidney Diseases.*

Return Service Requested

RCSB PROTEIN DATA BANK
www.pdb.org • info@rcsb.org

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