This spring, the RCSB PDB added new features and functionality that modernized the look and feel of www.pdb.org. A full list of features included in this release is available online, with many of these new items highlighted below and in the newsletter.

Quick searches by PDB ID or keyword, author name, structural genomics center, and chemical component name and ID are available from a new pull-down menu offered on every page.

Icons are used to indicate links to help pages (?), molecular viewers ( ), database searches ( ), and external sites ( ). Holding your mouse cursor over a search field or active link will display more information about these options.

Other new features include a Literature Tab View that incorporates PubMed-Central abstracts and images, and a Sequence Similarity Tab View that provides an overview of the sequence clusters for each protein chain in a PDB structure.

We look forward to any comments about these new features!
5 Easy Steps for Structure Deposition

Depositors can follow these steps to make the process of depositing a structure quick, easy, and accurate!

1. **Check the sequence**, which should contain all residues used in the experiment, including expression tags and residues missing due to disorder. Use BLAST\(^1\) to determine the appropriate sequence database references for the proteins or nucleic acids present in the file. Any sequence mismatches should correspond to mutation, variant or expression tags in the submitted sequence.

2. **Check the ligands**. Ligand Expo can be used to see if any of the chemical components in the structure (ligands, drugs, inhibitors, ions, modified residues, etc.) already exist in the Chemical Component Dictionary. If found, use the corresponding 3-character code in the coordinates; otherwise, the new component can be uploaded in ADIT along with the structure.

3. **Prepare data for deposition**. Use pdb\_extract to automatically collect information needed for deposition from the output files produced by many structure determination applications, and SF-Tool to convert the structure factor file format and check the data with SFCHECK.\(^2\)

4. **Validate the structure with the Validation Suite and Server** to ensure that the data being deposited are accurate and reflect what you intend to submit. This program checks the sequence and file format consistency, compares geometrical and chemical interactions to various standards, and reports errors that should be corrected before deposition.

5. **Deposit using ADIT/ADIT-NMR** to check, validate, and edit the PDB data entries.

Deposition is an iterative process. If you encounter problems at any particular step, please make the correction(s) and go through the steps again. For more information, please see deposit.pdb.org.

**Improve the Quality of Your Deposits with SFCHECK**

SFCHECK\(^3\) is integrated with the RCSB PDB’s deposition tools so users can assess the agreement between an atomic model and the corresponding X-ray data before a structure is deposited and released.

SFCHECK generates reports containing R-factors, density correlation, Luzzati plots, Wilson plots, temperature factors, local error estimation by residues, and more. It is a part of the Validation Server and ADIT tools, and is used by the RCSB PDB during the annotation process.

SFCHECK is also part of SF-Tool, a streamlined, web-based tool for validating crystallographic experimental data. In addition to validating model coordinates against structure factor data with SFCHECK, SF-Tool can:

- Easily translate a structure factor file between different formats (mmCIF, CIF, MTZ, CNS/CNX, XPLOR, SHEXL, TNT, HKL2000, SCALEPACK, XSCE, D*Trek, SAINT, other)
- Check for twinned or detwinned data

Depositors are strongly encouraged to use SFCHECK before deposition. For more information, please see deposit.pdb.org.

**How does an HPUB structure get released?**

An HPUB status indicates that a structure will be released when the primary reference is published (“hold until publication”). Once it is confirmed that the corresponding article is available, the structure is released in the weekly update of the PDB.

The wwPDB receives publication dates and citation information directly from a few journals. For most articles, however, the wwPDB searches PubMed and scans the literature for publication information. Citations emailed to deposit@wwpdb.org are also greatly appreciated.

There is a one-year limit on the length of a hold period, including those for HPUBs. If the citation for a structure is not published within the one-year period, depositors will be given the option to either release or withdraw the deposition.

**Deposition Statistics**

In the second quarter of 2009, 2079 experimentally-determined structures were deposited to the PDB archive. The entries were processed by wwPDB teams at the RCSB PDB, PDBj, and PDBi.

Of the structures deposited, 72.7% were deposited with a release status of HPUB; 24.6% were released as soon as annotation of the entry was complete; and 2.7% were held until a particular date. 91.8% of these entries were determined by X-ray crystallographic methods; 7.3% were determined by NMR methods.

1874 structures were released in the PDB during the same period.

**Data Query, Reporting, and Access**

**Literature View: Looking at Structures in PubMedCentral**

The recently-released Literature View aims to provide a broad look at how a given structure has been analyzed and presented in open access journals. For these publications, the full text of the article is available without copyright restriction. The overall intent of this feature is to increase awareness of all publications associated with the structure under study.

---

Accessible by selecting the Literature tab from any Structure Summary page, this view highlights a variety of published articles associated with a PDB entry. The primary citation’s abstract is included and can be used to search for structures with PubMed abstracts containing the same keywords. The Literature View also indicates the other publications included in the entry by the authors in addition to the primary citation (those contained in REMARK 1).

In collaboration with the BioLit project (biolit.ucsd.edu), the Literature View lists any open access articles found in PubMedCentral that contain the entry’s PDB ID—even the papers that do not include the PDB ID’s primary citation in the references. Links to the abstract and copyright information, along with figures and related legends from these articles are displayed.

If the listed open access articles also mention other PDB entries, their image, PDB ID, title, and sequence similarity to the initial structure are included in a table with links to their Structure Summary pages.

**Customizable Structure Summary Pages**

Tabbed Structure Summary pages let users explore information about any given structure. Located above the structure ID and title, each tab offers a different type of information—a general overview, derived data, sequence details, similarity literature, view, biology & chemistry, methods, geometry, and links to external resources. Data are organized on each page in different blocks, or “widgets”. On a structure’s Summary and Derived Data pages, these widgets can be hidden or expanded. The positioning of selected widgets can be moved around on the page. This feature lets users organize these pages to highlight information specific to their interests.

An online video demonstrating how to customize these page layouts is available from the RCSB PDB site in the “What’s New” section.

Customized layouts are automatically stored in a cookie and are retained for all PDB IDs. To switch back to the default layout, users should click on the “Reset View” in the left-hand menu.

_Literature view for 1tim. To explore other Literature Views, try PDB entries 1o4k, 2g12, and 1a1i.

MyPDB: Keep up-to-date with new structures...automatically!

MyPDB notifies users via email when the PDB releases structures that match customized queries. With MyPDB, users can:

- combine and save keyword, sequence, ligand, and other searches
- run searches stored in MyPDB at any time
- receive email alerts weekly or monthly
- access structures directly from the email alerts for further exploration
- utilize the RCSB PDB web interface to refine saved searches

Detailed information on how to register and use MyPDB was described in the Spring 2009 RCSB PDB Newsletter.

Website Statistics

Website access statistics for the second quarter of 2009 are given below.

<table>
<thead>
<tr>
<th>Month</th>
<th>Unique Visitors</th>
<th>Number of Visits</th>
<th>Bandwidth</th>
</tr>
</thead>
<tbody>
<tr>
<td>APRIL 09</td>
<td>178430</td>
<td>404618</td>
<td>875.53 GB</td>
</tr>
<tr>
<td>MAY 09</td>
<td>164084</td>
<td>380493</td>
<td>835.61 GB</td>
</tr>
<tr>
<td>JUNE 09</td>
<td>145958</td>
<td>356857</td>
<td>1079.61 GB</td>
</tr>
</tbody>
</table>

Ligand Expo: Searching and Browsing Features

Ligand Expo (ligand-expo.rcsb.org) can be used to access chemical and structural information about all small molecule components found in PDB entries.

The latest release includes amino acid variants—protonation and deprotonation states—as part of Ligand Expo’s Browse option. For example, when browsing arginine, protonation states ARG_LEO2 and ARG_LFOH_RNH1 will be included in the results list.

Ligand Expo can also be used to search for ligands using:

- files in a variety of formats including PDB, MOL/SDF and the Refmac/Phenix monomer library (mmCIF)
- chemical name
- formula
- SMILES STRING
- 3-character ID code

Ligand Expo is based upon the Chemical Component Dictionary (www.wwpdb.org/ccd.html) maintained by the wwPDB. Ligand Expo’s capabilities have been described in a flyer available for download from www.pdb.org.
Recent and Upcoming Meetings and Presentations

The RCSB PDB has recently exhibited with the PSI Structural Genomics Knowledgebase at the:

- 2009 Experimental Biology meeting (April 18-22) in New Orleans, Louisiana.
- Joint 17th Annual International Conference for Intelligent Systems for Molecular Biology and the 8th European Conference on Computational Biology (June 27-July 2; Stockholm, Sweden). RCSB PDB members also gave presentations at satellite and special interest group sessions.

Exhibition booths and posters will also be presented at the 23rd Annual Symposium of The Protein Society (July 25-29; Boston, MA) and the 2009 Annual Meeting of the American Crystallographic Association (July 25-30; Toronto, Canada).

At the Essentials for Educating Biochemistry and Molecular Biology Undergraduates Symposium sponsored by the American Society for Biochemistry and Molecular Biology (August 5-8; Colorado Springs, CO), the RCSB PDB’s Shuchismita Dutta and Rachel Kramer will run a workshop entitled “Molecular Visualization and Protein Databases (IIA): Tools, Rules and Stories: A Protein Data Bank Workshop Series.” For more information about this workshop, please see www.asbmb.org/Page2.aspx?id=2094.

Congratulations to National Tournament Champions

At the Science Olympiad National Tournament held May 15-16, 2009 at Augusta State University in Augusta, Georgia, New Jersey’s West Windsor-Plainsboro High School South won first place at the protein modeling event.

In 2010, the protein modeling trial event will take place at all of the Science Olympiad competitions held in the United States.

For more information about this event, see the Center for BioMolecular Modeling’s site at cbm.msoe.edu or the RCSB PDB protein modeling site at education.pdb.org.

wwPDB News: Gerard Kleywegt to head Protein Data Bank Europe

Starting July 1, 2009, Gerard Kleywegt will lead the Protein Data Bank Europe (PDBe) project at the European Bioinformatics Institute (Hinxton, UK). During the last 17 years he has been working in Uppsala, Sweden, a center of excellence for biomolecular crystallography, and has developed many tools that are widely used by structural biologists worldwide. He has an extensive publication record and has served on the PDBe Scientific Advisory Board for the last few years. He has also been one of the European representatives on the wwPDB advisory committee. He has a very strong international reputation and is well respected in the structural biology community.

Gerard will be replacing Kim Henrick, who has been the team leader of the PDBe since 2001. During his tenure at the EBI, Kim has made enormous contributions to the PDBe by establishing and leading a strong team that has developed a wide variety of services. Kim was a strong advocate for the formation of the wwPDB in 2003. He has played a key role in developing the standards for representing all the data in the archive, from small molecules to very large biological assemblies, and in leading the remediation efforts. Kim’s deep knowledge of the PDB, his critical eye, and sharp wit will be greatly missed by the wwPDB.

The members of the wwPDB look forward to working closely with Gerard and wish Kim all the best in the next chapter of his career.

Looking at Structures: A Resource for Learning About PDB Data

Where are all the hydrogen atoms in this file? Should I care about the R-factor? Why are there 20 overlapped structures in my file? These questions and many others are explored in the RCSB PDB’s new Looking at Structures.

Using text, images, and interactive Jmols, Looking at Structures intends to help researchers and educators get the most out of the PDB archive. Broad topics include how to understand PDB data, how to visualize structures, how to read coordinate files, and potential challenges in exploring the archive.

A Table of Contents appears on the right side of every page so at any time users can access the individual pages: Biological Units, Dealing with Coordinates, Methods for Determining Structure, Missing Coordinates and Biological Units, Molecular Graphics Programs, Resolution, and R-value and R-free. The addition of future topics is planned.

Looking at Structures is available from the General Education section of the left-hand menu.
San Diego Science Festival

Throughout March and the beginning of April, the San Diego Science Festival (SDSF) brought together students, families, businesses, scientists, and communities for a series of fun (and free) events that highlighted the impact of science and innovation on our lives. Through this program, more than 200,000 people interacted with San Diego’s local scientific community through various school, evening, and outreach programs.

SMALL WONDERS CAVE TOUR

High School students and teachers from all over San Diego County visited the University of California, San Diego (UCSD) as part of the Small Wonders program organized by the SDSF and UCSD. This April 1 event offered a day of exploration into the world of nanotechnology and advanced materials.

The RCSB PDB demonstrated the inner workings of proteins in the Calit2 Cave, an immersive environment that features a 5-walled projected virtual reality room with 68-million pixel resolution (for more information about the Cave, see www.scivee.tv/node/3648).

Students explored the architecture of protein structures, how RNA polymerase transcribes DNA into RNA, and how drugs such as HIV protease inhibitors work. Students and teachers alike were fascinated by both the beauty and the complexity of protein structures. This unique experience certainly sparked a keen interest in science among the students as the organizers had hoped.

EXPO DAY

The finale to the entire SDSF was April 4th Expo Day in Balboa Park. More than 50,000 attendees visited over 200 different booths and watched over 25 science-related performances. The RCSB PDB’s exhibit booth offered hands-on activities, free materials, and demonstrations. In addition to building viruses and exploring proteins with the RCSB PDB, Expo visitors were able to finger paint with algae, play catch with a robot, look through telescopes at sunspots, race remote controlled cars, make test-tube lava lamps, and create musical instruments.
Rutgers Day

On the other side of the country, Rutgers Day was held on April 25 to celebrate all of the exciting activities, programs, and talent at New Jersey’s state university. More than 50,000 people attended and participated in free performances, tours, exhibits, hands-on activities, lectures, demonstrations, and the traditional New Jersey events of the Faraday Physics Lecture, Ag Field Day, and the New Jersey Folk Festival. The hosts from the RCSB PDB showed visitors the 3D structures of proteins and built viruses as part of the Department of Chemistry and Chemical Biology’s live experiments and demonstrations.

Visitors could build virus models using toothpicks and marshmallows.

Q: What do you see as the balance between experimental/wet science and computational science in industry? How about in drug development?

A: Since my experience is in drug discovery and development, I can only comment on that particular aspect of industry. Unfortunately, computational science has been negligent. We rarely know or spend time determining the errors in or the true predictability of the methods we use. As a result, the perception is that experimental/wet science is the gold standard, and computational science is only useful when experiments are too expensive or when the number of experiments is too large. A model cannot be better than the data upon which it is built, so computational science will never “beat” experimental science. We can generate models with errors that are close to the experimental errors, and those models can be very useful in saving time and expense. In some cases, there is a clear need for computational science to "lead" experiments—to make predictions that can contribute to improving or

PDB Community Focus

Gregory Warren, Ph.D. OpenEye Scientific Software, Inc.

Dr. Warren was recently one of the presenters at the Crystallography for Modelers course.

GREGORY WARREN did his graduate studies at the Massachusetts Institute of Technology with Gregory Petsko and Robert Griffin, and his post-doctoral work with Axel Brünger at Yale as part of the CNS development team. He spent eight years at GlaxoSmithKline as a computational chemist/molecule modeler supporting drug discovery. In 2006, he joined OpenEye Scientific Software, Inc. (www.eyesopen.com) as a Senior Applications Scientist where his responsibilities include support of AFITT OpenEye’s X-ray crystallography structure determination application. OpenEye develops large-scale molecular modeling applications and toolkits (programming libraries suitable for custom development). Primarily geared towards drug discovery and design, areas of application include structure generation, docking, shape comparison, charge/electrostatics, chemical informatics and visualization. The software is designed for scientific rigor, as well as speed, scalability and platform independence.

Dr. Warren did his graduate studies at the Massachusetts Institute of Technology with Gregory Petsko and Robert Griffin, and his post-doctoral work with Axel Brünger at Yale as part of the CNS development team. He spent eight years at GlaxoSmithKline as a computational chemist/molecule modeler supporting drug discovery. In 2006, he joined OpenEye Scientific Software, Inc. (www.eyesopen.com) as a Senior Applications Scientist where his responsibilities include support of AFITT OpenEye’s X-ray crystallography structure determination application. OpenEye develops large-scale molecular modeling applications and toolkits (programming libraries suitable for custom development). Primarily geared towards drug discovery and design, areas of application include structure generation, docking, shape comparison, charge/electrostatics, chemical informatics and visualization. The software is designed for scientific rigor, as well as speed, scalability and platform independence.

Dr. Warren did his graduate studies at the Massachusetts Institute of Technology with Gregory Petsko and Robert Griffin, and his post-doctoral work with Axel Brünger at Yale as part of the CNS development team. He spent eight years at GlaxoSmithKline as a computational chemist/molecule modeler supporting drug discovery. In 2006, he joined OpenEye Scientific Software, Inc. (www.eyesopen.com) as a Senior Applications Scientist where his responsibilities include support of AFITT OpenEye’s X-ray crystallography structure determination application. OpenEye develops large-scale molecular modeling applications and toolkits (programming libraries suitable for custom development). Primarily geared towards drug discovery and design, areas of application include structure generation, docking, shape comparison, charge/electrostatics, chemical informatics and visualization. The software is designed for scientific rigor, as well as speed, scalability and platform independence.

Dr. Warren did his graduate studies at the Massachusetts Institute of Technology with Gregory Petsko and Robert Griffin, and his post-doctoral work with Axel Brünger at Yale as part of the CNS development team. He spent eight years at GlaxoSmithKline as a computational chemist/molecule modeler supporting drug discovery. In 2006, he joined OpenEye Scientific Software, Inc. (www.eyesopen.com) as a Senior Applications Scientist where his responsibilities include support of AFITT OpenEye’s X-ray crystallography structure determination application. OpenEye develops large-scale molecular modeling applications and toolkits (programming libraries suitable for custom development). Primarily geared towards drug discovery and design, areas of application include structure generation, docking, shape comparison, charge/electrostatics, chemical informatics and visualization. The software is designed for scientific rigor, as well as speed, scalability and platform independence.
falsifying a particular hypothesis. Computational and experimental science should be complementary, but in practice this does not often happen. Instead, we choose one method and abandon it only when it fails. Given the data, the choice of experimental science over computational science is a reasonable one.

**Q:** What should modelers know about crystallographic data that they currently do not?

**A:** Crystallographic data, like all other experimental data, contains measurement error and that error is measured. The inherent error affects the quality of the molecular model that is built to fit or explain the data. Crystallographic data and the resulting model do not have god-like properties such as infinite precision or perfect accuracy—just because a measurement is presented to three decimal places does not signify that the third decimal place has any meaning. The not-so-funny part is that modelers do understand model quality. We understand that a homology model built from a structure with 15% sequence identity is not as reliable as a model built from of a structure with 95% sequence identity. For some unexplained reason that understanding is not applied to crystallographic data, and too many believe that crystallographic models are infinitely precise.

**Q:** How has the interface between modeling and crystallography changed? What changes are likely to occur?

**A:** In the last 10 years, my perception has been that the interface has gone from one of two separate scientific silos launching data but rarely speaking to a more interactive team effort. Yes, there is still a silo effect because crystallography and modeling groups are only rarely colocated, but it is nothing like what it used to be. It is important to remember that we—in crystallography and modeling—each have customers. Crystallography has critically important data that modeling needs. Modeling uses the data to provide information to the drug discovery team. In all cases, the customer needs to understand the quality of the information being provided so that intelligent and scientifically sound decisions can be made. What changes would I like to see? I would like to see collocation of modeling and crystallography groups, and in the absence of that, consistent communication about the project being worked on and the quality of the data being generated. We need to remember that job security depends, in part, on demonstrated impact. We can have greater impact if we work together.

**Q:** How can crystallographic data best be used in modeling and drug discovery?

**A:** The data can be best used if modelers and crystallographers each had a better understanding of each other’s field. That said, we do not all need to become fully rounded, Renaissance scientists, but a little understanding of the strengths and weaknesses of each discipline would allow the data to be used most effectively. If the crystallographer knows that the modeler will be using a technique that treats the protein as rigid but the data say that the protein is very flexible in the crystal, that information is potentially very important and may cause the modeler to use a different method. I have this unfounded belief that if a scientist understands the strengths and weakness of the data, they will use that information and do the right thing.

**Q:** What are the challenges that computational science and crystallography face with regard to drug discovery?

**A:** I would say the challenge is to have measurable impact. There are examples where crystallography and modeling have had an impact on the discovery and development of a marketed drug. Unfortunately, we need to have a consistent quantifiable effect on the drug discovery process; otherwise, from a business perspective we are not very useful. How can we have a consistent impact? By using the data being generated more efficiently. We need to understand the quality and accuracy of the data and use that information to generate predictions whose quality and accuracy are quantified. An understanding of how good or bad the data are, or a prediction is, allows us to determine when to use the data or seek other options. If crystallography and modeling can provide well-defined information, our impact will be obvious.

---

**A Short Course: Crystallography for Modelers**

On May 7-8, 2009, practicing pharmaceutical and biophysical modelers attended the short course entitled *Crystallography for Modelers* to develop a better understanding of crystal structures and PDB data. The course was held at Rutgers, The State University of New Jersey in Piscataway, NJ and was sponsored by the RCSB PDB.

RCSB PDB members Helen Berman, Shuchismita Dutta, David S. Goodsell, and Cathy Lawson, along with Rutgers Professor of Chemistry and Chemical Biology Joseph Marcotrigiano, described the process of crystal structure determination and provided insight into the workings of the PDB and how “PDB” files are generated. In detailed discussions, participants examined the extensive information (beyond coordinates) available in data files and online, including ligand structures. Topics included proper interpretation of crystal structures, with an emphasis on accuracy, precision, problems and experimental, real, and/or perceived errors.

On the second day, industrial participants Jeffrey A. Bell (Schrödinger, Inc.), Gregory Warren (OpenEye Scientific Software, Inc.), Howard Feldman (Chemical Computing Group Inc.), and Ruben Abagyan (Molsoft LLC) led the class through hands-on software demonstrations.

*Crystallography for Modelers* was organized for the RCSB PDB by Terry Richard Stouch, and offered through the Rutgers Advanced Technology Extension program. The RCSB PDB gratefully acknowledges the additional support from Chemical Computing Group Inc., ACS Division of Computers in Chemistry, and the *Journal of Computer-aided Molecular Design.*

---

Course attendees were able to participate in hands-on activities.
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**

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)

**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.

**RCSB PDB Partners**

**RCSB PDB Management**

**DR. HELEN M. BERMAN,** Director
Rutgers, The State University of New Jersey
berman@rcsb.rutgers.edu

**DR. MARTHA QUSSADA,** 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.