

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

# NEWSLETTER



[www.rcsb.org](http://www.rcsb.org) • [info@rcsb.org](mailto:info@rcsb.org)

Weekly RCSB PDB news is available online at [www.rcsb.org](http://www.rcsb.org)

## Contents

**MESSAGE FROM THE RCSB PDB** ..... 1

**DATA DEPOSITION AND ANNOTATION**

2012 Deposition Statistics ..... 2

wwPDB News ..... 2

**DATA QUERY, REPORTING, AND ACCESS**

2012 Website Statistics ..... 2

Improved Drug Searches and Annotations ..... 2

Explore Different Views of Search Results ..... 2

Map PDB Structures to Full-length Protein Sequences ..... 3

MyPDB: Keep up-to-date with new structures... automatically! ..... 3

Advanced Search: Multiple ID Search ..... 3

Advanced Search: Sequence Motif ..... 4

**OUTREACH AND EDUCATION**

2012 Annual Report ..... 5

Papers Published ..... 5

Explore the Structural Biology of HIV ..... 5

Educational Activities and Lessons ..... 4

Turn Your Computer into a PDB Structure Kiosk ..... 6

**EDUCATION CORNER** by Takanori Nakane, M.D., *Molecular Visualization, Online and On-the-Go* .. 6

**PDB PARTNERS, MANAGEMENT, AND STATEMENT OF SUPPORT** ..... 8

## Message from the RCSB PDB



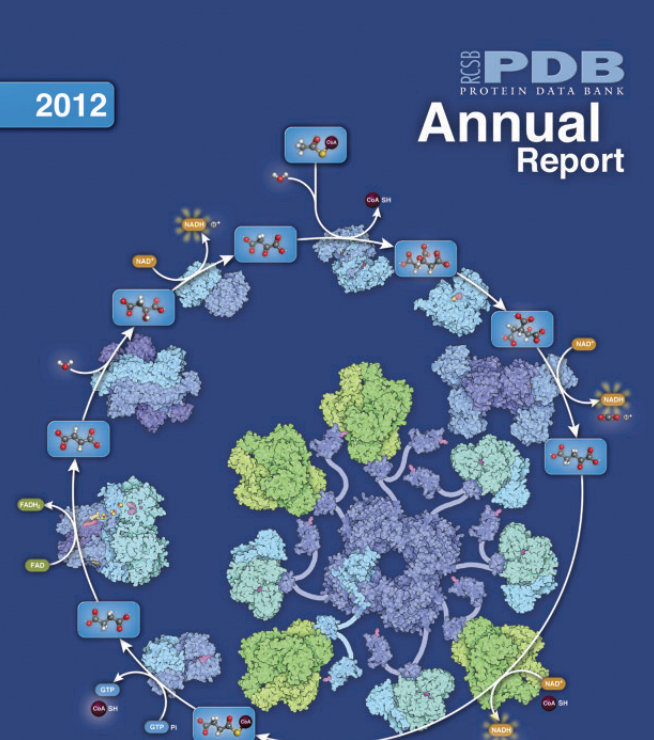
### Newsletter Poll

We'd like to learn more about how our readers use this newsletter.

Do you read the regularly printed version of the newsletter? What types of articles are you interested in reading?

Please take the brief survey online at: [www.surveymonkey.com/s/ZV5W92J](http://www.surveymonkey.com/s/ZV5W92J)

## Annual Report Published



**2012**

RCSB PDB PROTEIN DATA BANK

### Annual Report

The 2012 Annual Report can be downloaded at [www.rcsb.org/ar12](http://www.rcsb.org/ar12)

*The cover highlights the proteins and chemical reactions involved with the citric acid cycle.*

RESEARCH COLLABORATORY FOR STRUCTURAL BIOINFORMATICS  
 Rutgers, The State University of New Jersey  
 University of California, San Diego

## SNAPSHOT: JANUARY 1, 2013

87153 released atomic coordinate entries

ENTRIES BY MOLECULE TYPE		ENTRIES BY EXPERIMENTAL TECHNIQUE	
80681	proteins and peptides	76706	X-ray
3974	protein/nucleic acid complexes	9741	NMR
2474	nucleic acids	489	electron microscopy
24	other	51	hybrid
		166	other

## RELATED EXPERIMENTAL DATA FILES

66157	structure factors
7048	NMR restraints
1679	EMDB electron microscopy maps
806	NMR chemical shifts



This newsletter is printed on recycled paper

## Data Deposition and Annotation

### 2012 Deposition Statistics

In the fourth quarter of 2012, 2385 experimentally-determined structures were deposited to the PDB archive for a total of 9964 entries deposited in the year. 9250 entries were deposited in 2011.

Of all structures deposited in 2012, 81.5% were deposited with a release status of hold until publication; 14.5% were released as soon as annotation of the entry was complete; and 4.0% were held until a particular date. 93.0% of these entries were determined by X-ray crystallographic methods; 5.9% were determined by NMR methods.

8978 structures were released in the PDB archive in 2012. They account for ~10% of the current total holdings of 87153.

### wwPDB News

#### New Reference Dictionary for Biologically Interesting Molecules Released

The wwPDB has reviewed the representation of peptide-like antibiotic and inhibitor molecules in the PDB archive. As a result, a new reference dictionary has been created to help define and represent these biologically interesting molecules.

This Biologically Interesting molecule Reference Dictionary (BIRD) contains chemical descriptions, sequence and linkage information, and functional and classification information as taken from the core structures and from external resources. PDB entries containing these molecules have been annotated using this dictionary, and will contain a corresponding BIRD ID code only in the PDBx-formatted file.

BIRD is available on the wwPDB FTP server adjacent to the Chemical Component Dictionary. The use of BIRD will greatly improve the consistency of peptide-like antibiotic and inhibitor molecules in the PDB. These data reflect the wwPDB's continuing commitment to providing accurate and detailed data to users worldwide.

#### Remediation of Structure Factor Data and Release Plan

Structure factor files have been reviewed for format consistency and correspondence with coordinate data. As a result, around 43,800 structure factor files have been updated to standardize the format and to incorporate data corrections. Changes made are described in the audit.details record of the structure factor file.

Starting January 16, 2013, the wwPDB will release these modified structure factors in weekly batches of 1000.

For more information, see [wwpdb.org](http://wwpdb.org).

#### Experimental Data Release Policy Clarification

Effective February 6, 2013, coordinates and experimental data deposited in the PDB will follow the same release status.

## Data Query, Reporting, and Access

### 2012 Website Statistics

MONTH	UNIQUE VISITORS	NUMBER OF VISITS	BANDWIDTH
JANUARY	242,584	580,079	749.94 GB
FEBRUARY	258,865	628,875	877.04 GB
MARCH	255,442	613,853	825.07 GB
APRIL	258,130	617,748	778.90 GB
MAY	244,865	615,509	794.33 GB
JUNE	206,864	525,815	624.38 GB
JULY	184,117	478,089	697.83 GB
AUGUST	175,961	439,848	484.91 GB
SEPTEMBER	241,088	572,574	755.14 GB
OCTOBER	316,640	723,577	970.88 GB
NOVEMBER	309,021	711,058	1,513.52 GB
DECEMBER	265,140	599,337	1,179.91 GB

### Improved Drug Searches and Annotations

Drug and drug target data from DrugBank ([www.drugbank.ca](http://www.drugbank.ca))<sup>1</sup> have been integrated with the RCSB PDB website.

The simple top bar search can find generic and brand names of drugs. For example, type Zolofit into the simple top bar search to find the related Ligand Summary page.

The Ligand Summary page links to entries containing Zolofit and provides summary information about the drug. The Drug Info widget on this page lists and links to corresponding data from DrugBank (when available), including DrugBank ID, drug name, groups, brand name, descriptions, and more.

### Explore Different Views of Search Results

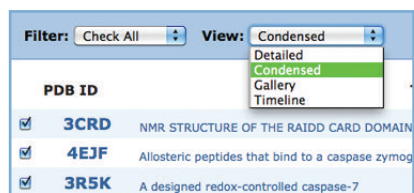
Explore query results through four different views: detailed, condensed, image gallery, and timeline.

In addition to the original query results browser, now called the detailed view, results can be browsed using condensed, timeline or gallery views. The views are synchronized; selecting or deselecting a structure in one view will have the same effect in the others.

The default view was designed to provide detailed information about each entry: a thumbnail image, release date, classification, compound, experimental type, citation, and links related *Molecule of the Month* articles, and more.

The condensed view concisely lists structures, showing only the PDB ID, title, and macromolecule name found in their results.

The gallery view shows only the structure image and PDB ID. Images can be resized by using the options dropdown. The PDB ID links to the Structure Summary page for the entry.



Users can toggle through different views of search results

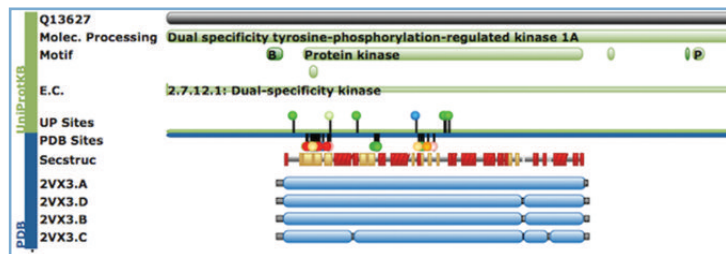
The timeline view takes the current search results and generates a visual timeline ordered by release date, similar to PDB-101's Author Profile feature.

These different views were designed to support the interests of different users.

## Map PDB Structures to Full-length Protein Sequences

The new Protein Feature View visually summarizes how a full-length protein sequence from UniProtKB<sup>3</sup> corresponds to PDB entries. It also loads annotations from external databases (such as Pfam<sup>4</sup> and homology models from the Protein Model Portal.<sup>5</sup> Annotations visualizing predicted regions of protein disorder (computed with JRONN)<sup>6</sup> and hydrophobic regions (as computed using a sliding window approach) are also displayed.

For individual entries, the Protein Feature View is available from the Molecular Description Widget on Structure Summary pages. The example shown for PDB ID 2vx3 illustrates how the ranges of a protein that have been observed in an experiment (in blue) correspond to the full length UniProtKB sequence (in grey). The secondary structure information from the PDB entry is also shown (helices in red, beta strands in yellow).



Protein Feature View from the Structure Summary page for 2vx3<sup>2</sup>

Various features that are known for the UniProtKB sequence are displayed in green as they correlate with regions in the PDB entry. Active sites (from UniProtKB and the PDB entry) are also annotated. Moving the mouse cursor over the "lollipop" displays the residue label. Mousing over the images shown in the "Secstruc" row reveals secondary structure from the PDB entry.

This view can be expanded to map all PDB entries related to a single UniProtKB sequence by selecting the Protein Feature View link shown

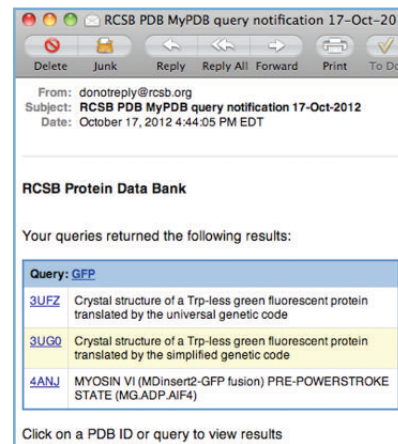
in this widget. By default, a few representative PDB entries are used to give an overview for which regions of the UniProtKB sequence PDB entries are available. Selecting the plus sign or the "Show All" button will expand the view to show all related PDB chains, which can then be sorted by resolution, length, and release date. These Protein View images can be exported as Scalable Vector Graphics (SVG) files.

The PDB to UniProtKB mapping is based on the data provided by the Structure integration with function, taxonomy and sequence (SIFTS) initiative.<sup>7</sup>

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

Create a personalized version of the RCSB PDB website using these MyPDB features:

- Sign up for regular email alerts. Searches saved in MyPDB can be set to run with each update. Email alerts (weekly or monthly) will be sent when new entries matching the search are released in the PDB archive.
- Save searches online. The MyPDB Saved Query Manager stores any type of RCSB PDB structure search: a particular keyword, sequence, ligand, Advanced Search composite query, and more. Run these searches at any time with the click of a button.
- Store personal notes and bookmark structures. Save personal annotations and notes on the Structure Summary tab of any PDB entry. Want to keep a running list of interesting structures? MyPDB can keep a favorites list of PDB entries. The Personal Annotations summary page provides easy access to all of these tagged structures and annotations.



MyPDB will email alerts when structures matching saved queries are released.

All MyPDB account information is kept private and secure, and can be updated at any time.

## Advanced Search: Multiple ID Search



Have a list of PDB IDs you'd like to learn more about? Use Advanced Search to view all of the structures in the Query Results Browser.

From the Advanced Search pull-down menu, select ID(s) and Keywords>PDB ID(s). Enter the list of multiple IDs, which can be separated by commas or white space, including line breaks.

The structures can then be explored using the different query results browser views: detailed, condensed, timeline or gallery.

Users can also link directly to multiple structures in the Query Results Browser using the URL syntax:

[http://www.rcsb.org/pdb/search/smart.do?smartSearchSubtype\\_1=StructureIdQuery&structureIdList\\_1=1MI6+1MVR](http://www.rcsb.org/pdb/search/smart.do?smartSearchSubtype_1=StructureIdQuery&structureIdList_1=1MI6+1MVR)

This example will launch entries 1mi6 and 1mvr in the Query Results Browser; other IDs can be added to the URL.

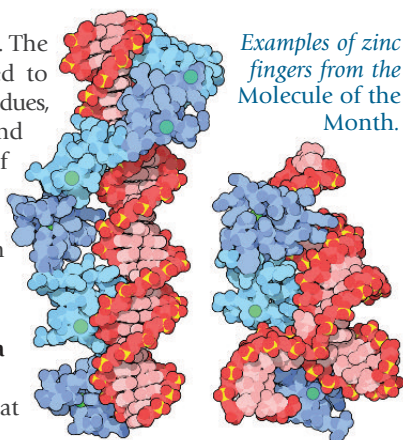


## Advanced Search: Sequence Motif

Advanced Search lets users build queries of specific types of data. Users can query for an exact sequence or for a sequence pattern using regular expression syntax. To look for structures with a particular Sequence Motif, try using one of these techniques with the Sequence Features > Sequence Motif option.

- **Short Sequence Fragments.** The sequence motif search, unlike BLAST or FASTA, can search for short sequence fragments of any size, such as NPPTP
- **Wildcard Searches.** Use an 'X' in the sequence for wildcard searching. For example, XPPXP can be entered to look for SH3 domains using the consequence sequence -X-P-P-X-P (where X is a variable residue and P is proline)
- **Multiples of Variable Residues.** The {n} notation can be used, where n is the number of variable residues. To query a motif with 7 variables between residues W and G, and 20 variable residues between G and L, try WX{7}GX{20}L
- **Ranges of Variable Residues.** The {n,m} notation can be used to indicate ranges of variable residues, where n is the minimum and m the maximum number of repetitions. For example, the zinc finger motif that binds Zn in a DNA-binding domain can be expressed as: CX{2,4}CX{12}HX{3,5}H
- **Motifs at the Beginning of a Sequence.** The '^' operator searches for sequence motifs at the beginning of a protein sequence. Two ways of looking for sequences with N-terminal histidine tags are: ^HHHHHH and ^H{6}
- **Alternative Residues.** Square brackets specify alternative residues at a particular position. To search for a Walker (P loop) motif that binds ATP or GTP, try: [AG]XXXXGK[ST]

The search will look for sequences with A or G, followed by 4 variable residues, then G K, and finally S or T.



Examples of zinc fingers from the Molecule of the Month.

## Outreach and Education

### Join the RCSB PDB Team

We're looking for a Lead Web Architect. Learn more at <http://bit.ly/QSwMk9>

## 2012 Annual Report

The recently published Annual Report highlights the RCSB PDB's accomplishments for 2012. It also reviews topics ranging from deposition statistics to domain-based structural alignments to popular help desk questions and answers.

wwPDB efforts, including the development of the Common Deposition and Annotation Tool, are also highlighted.

This edition's cover maps the proteins and chemical reactions involved with the citric acid cycle.

These bulletins provide a yearly snapshot of RCSB PDB activities and the state of the PDB archive. This edition, the RCSB PDB's thirteenth, is available as a PDF download from [www.rcsb.org/ar12](http://www.rcsb.org/ar12). If you would like a printed copy, please send your postal address to [info@rcsb.org](mailto:info@rcsb.org).



The 2012 Annual Report

## Papers Published

- 1. C. Knox, V. Law, T. Jewison, P. Liu, S. Ly, A. Frolkis, A. Pon, K. Banco, C. Mak, V. Neveu, Y. Djoumbou, R. Eisner, A. C. Guo, D. S. Wishart. (2011) DrugBank 3.0: a comprehensive resource for 'omics' research on drugs. *Nucleic Acids Res.* 39: D1035-1041.
- 2. A. K. Roos, M. Soundararajan, A. C. W. Pike, O. Fedorov, O. King, N. Burgess-Brown, C. Philips, P. Filippakopoulos, C. H. Arrowsmith, M. Wikstrom, A. Edwards, F. Vondelft, C. Bountra, S. Knapp. (2008) The Crystal Structure of the Human Dual Specificity Tyrosine-Phosphorylation-Regulated Kinase 1A. [doi:10.2210/pdb2vx3/pdb](https://doi.org/10.2210/pdb2vx3/pdb)
- 3. UniProt Consortium. (2012) Reorganizing the protein space at the Universal Protein Resource (UniProt). *Nucleic Acids Res* 40: D71-75.
- 4. M. Punta, P. C. Coggill, R. Y. Eberhardt, J. Mistry, J. Tate, C. Bournnell, N. Pang, K. Forslund, G. Ceric, J. Clements, A. Heger, L. Holm, E. L. Sonnhammer, S. R. Eddy, A. Bateman, R. D. Finn. (2012) The Pfam protein families database. *Nucleic Acids Res* 40: D290-301.
- 5. K. Arnold, F. Kiefer, J. Kopp, J. N. Battey, M. Podvinec, J. D. Westbrook, H. M. Berman, L. Bordoli, T. Schwede. (2009) The Protein Model Portal. *Journal of structural and functional genomics* 10: 1-8.
- 6. Z. R. Yang, R. Thomson, P. McNeil, R. M. Esnouf. (2005) RONN: the bio-basis function neural network technique applied to the detection of natively disordered regions in proteins. *Bioinformatics* 21: 3369-3376.
- 7. S. Velankar, P. McNeil, V. Mittard-Runte, A. Suarez, D. Barrell, R. Apweiler, K. Henrick. (2005) E-MSD: an integrated data resource for bioinformatics. *Nucleic Acids Res* 33: D262-265.
- The RCSB Protein Data Bank: new resources for research and education (2013) *Nucleic Acids Research* [doi:10.1093/nar/gks1200](https://doi.org/10.1093/nar/gks1200)
- Update of new tools and resources that provide a structural view of biology for research and education, including improved methods for simple and complex searches of PDB data, creating specialized access to chemical component data and providing domain-based structural alignments.
- Creating a community resource for protein science (2012) *Protein Science* 21: 1587-1596 [doi: 10.1002/pro.2154](https://doi.org/10.1002/pro.2154)
- Discusses how management, data content, infrastructure, and community have influenced the continued development of the PDB. It also analyzes the corpus of data to identify trends in the field of structural biology.
- BioJava: an open-source framework for bioinformatics in 2012 (2012) *Bioinformatics* 28: 2693-2695 [doi:10.1093/bioinformatics/bts494](https://doi.org/10.1093/bioinformatics/bts494)
- Describes modules that provide state-of-the-art tools for protein structure comparison, pairwise and multiple sequence alignments, working with DNA and protein sequences, analysis of amino acid properties, detection of protein modifications and prediction of disordered regions in proteins as well as parsers for common file formats using a biologically meaningful data model.

- Illustrating the Machinery of Life: Viruses (2012) *Biochemistry and Molecular Biology Education* 40: 291-296 doi:10.1002/bmb.20636

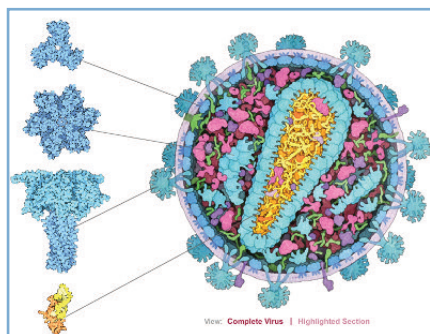
David Goodsell's series continues with illustrations of poliovirus, influenza, and HIV.

- Putting proteins in context (2012) *Bioessays* 34:718-20 doi:10.1002/bies.201200072

David Goodsell describes his thoughts and process for scientific illustrations.

### Explore the Structural Biology of HIV

HIV (human immunodeficiency virus) is composed of two strands of RNA, 15 types of viral proteins, and a few proteins from the last host cell it infected, all surrounded by a lipid bilayer membrane. Together, these molecules allow the virus to infect cells of the immune system and force them to build new copies of the virus. Each molecule in the virus plays a role in this process, from the first steps of viral attachment to the final process of budding.



Visit PDB-101 for this animation and more

The structural proteins, viral enzymes, and accessory proteins of HIV that have been determined are highlighted in a downloadable poster and an interactive, online Flash website. Clicking on a protein in the animation reveals information about the structure and a link to the PDB entry.

The animation and poster are available from PDB-101.

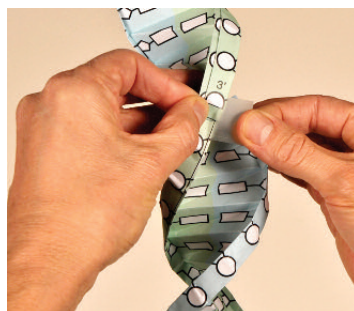
### Educational Activities and Lessons

The PDB-101 website packages together resources that promote exploration in the world of proteins and nucleic acids for teachers, students, and the general public.

The Educational Resources section of PDB-101 offers different materials to help learn about structures in the PDB, including posters, animations, and classroom activities and lessons.

These activities and lessons include:

**Exploring the Structure of tRNA.** Build a paper model of the primary, secondary, and beginnings of the tertiary structure of tRNA. Visit the related activity page to learn more.



Educational materials from PDB-101

**Virus Structure Lecture, Lesson Plans, and Activities.** These resources, aimed at middle and high school classrooms, include templates to make 3D paper models of disease-causing viruses and review questions.

**Green Fluorescent Protein Tutorial.** With this tutorial you will find protein structures using search tools on the RCSB PDB website; use molecular visualization tools to explore the GFP structure and function; and find the GFP gene and view important mutations.

**3D Paper Model of DNA.** Download a PDF to build one complete turn of a DNA double helix, with the major and minor grooves labeled.

**Education Corner.** Other activities and lessons are available from the RCSB PDB Newsletter's Education Corner archives.

To enter this educational view of the RCSB PDB, click on the blackboard PDB-101 logo or its related widget in the left-hand menu. PDB-101 offers easy navigation: select any Molecule of the Month article from the top bar menu or mouse over the upper-left PDB-101 pull-down menu to jump to other sections. Select the blue RCSB PDB logo from the top of the page at any time to return to the main website.

### Turn Your Computer into a PDB Structure Kiosk



Molecules in Motion Kiosk Display

Highlight structures from your lab, institution, or class with the Molecules in Motion Kiosk Viewer. Using a list of PDB IDs, this full-screen animation program will display any structure from different angles and perspectives. It also focuses on any chemical components within the structure. The Java viewer can be downloaded or launched from PDB-101's Educational Resources page.





## Education Corner by Takanori Nakane, M.D.

### Molecular Visualization, Online and On-the-Go

#### Molecules and Computers

I became fascinated by molecules starting in high school, when I found an old edition of Lehninger's *Principles of Biochemistry* on my father's bookshelf. I was quickly captivated by complex structures of biomolecules, and found the cyclization reaction leading from epoxysqualene to lanosterol especially impressive. In this reaction, four rings and eight stereocenters are created in a perfectly controlled manner. Ever since then, I have been deeply immersed in the world of chemistry.

Another enduring interest has been computer programming. When I was an eight year-old boy, my father gave me a very, very old computer and let me play with it freely. The CPU was 8-bit and had a main memory of only four kilobytes! Although it was not powerful at all, I had perfect control over the hardware. Since there wasn't an operating system or any device driver to be concerned about, I was able to read and write anything at any memory address and directly send commands to peripheral chips. It was an invaluable experience and taught me to make best of limited computational resources.

#### Making Structures More Accessible

When I was an undergraduate student, I noticed that biology students tended to learn compounds and enzymes only as names to be memorized, with little attention given to their structures. Similarly, only a few medical students look at the chemical structures of drugs and their target proteins. Even teachers tend to ignore molecular structures, and mention only their names. I believe these are missed opportunities to deepen one's understanding of molecules and to appreciate their beautiful structures. However, once teachers and students are introduced to the study of molecular structures and discover that they are interesting (RCSB PDB's *Molecule of the Month* is a good starting point), they too are able share my fascination with structure. My work is motivated by the idea that by increasing exposure, more people will become familiar with structures. What if more websites showed molecules in an interactive manner?

Another motivation for developing mobile tools is quite personal. Every time I come across a new compound, I search for the structural formula and copy it to my notebook by hand to get familiar with it. With a new protein, I check the PDB to see if its structure is determined. But I am not always in front of my computer!

When you are with your own computer, you can use any software you have. But what if you want to show a structure to your friend or colleague on another machine and you may be unable to install a new program? You need a molecular viewer that works within web browsers. Also, when you are listening to a lecture or a seminar, you might not be near a computer. However, since most of us carry smart phones or tablets with us all day long that use web browsers, I thought that developing a molecular viewer for the web would cover all situations. So my work on GLmol began.

#### GLmol: Molecular viewer for browsers

My work on GLmol, a molecular viewer for web browsers, began about a year ago. It is written in JavaScript alone, requiring no other plugins such as Flash or Java.



**TAKANORI NAKANE** graduated from Kyoto University's School of Medicine in 2010. After passing Japan's national board exam and qualifying as a medical doctor, Nakane decided to enter a four-year PhD course in X-ray crystallography at Kyoto University instead of clinical practice. While working on "wet" experiments, he realized that he is more attracted to "dry" (computational) aspects of science.

Over the last year, Nakane has been developing molecular viewers for Web browsers and mobile devices, including the NDKmol molecular viewer that been incorporated with the RCSB PDB Mobile app. Nakane also contributes to several open-source projects, including PyMOL and Jmol.

Visit [webglmol.sourceforge.jp](http://webglmol.sourceforge.jp) to download Nakane's GLmol (Molecular viewer on WebGL/Javascript) and ESmol & NDKmol (Molecular viewer for Android). Dr. Nakane may be reached at [nakane.t@gmail.com](mailto:nakane.t@gmail.com).

Most websites show structures of compounds and proteins only as static images. Some sites, including the RCSB PDB, use Java-based molecular viewers such as Jmol. While they are relatively fast once launched successfully, loading structures can be very slow because browsers have to first load the Java runtime environment. Occasionally browsers can be unresponsive during initialization. So I wanted to develop a molecular viewer that does not rely on anything but the web browser itself.

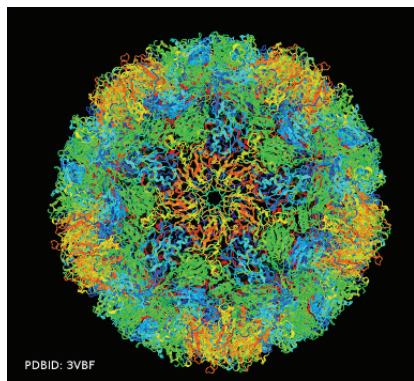
Fortunately, recent browser development has achieved substantial improvement in the speed of Javascript engines. Furthermore, a new enhancement to HTML called WebGL (Web Graphics Library) enables web pages to harness hardware acceleration for 3D rendering. GLmol is based on these technologies.

#### Pursuing Performance

Twenty years ago, molecular visualization required expensive graphics workstations. Today, thanks to powerful processors found in desktop computers, real-time rendering of molecules no longer poses the same challenge. Mobile platforms, however, are an exception. While most computers have four to eight CPU cores running at more than two gigahertz, most tablets have only two cores running at one gigahertz. Their memory is at most one gigabyte. This corresponds to typical desktop computers ten years ago. Therefore, simply porting existing molecular visualization software designed for desktop computers to the mobile platform will not succeed. It potentially might work for relatively small proteins but rendering would be very slow for large molecules. Programs running on web browsers have another limitation even when running on a desktop computer. Since they are written in a scripting language, they run several times slower than native applications.

GLmol was written and designed from scratch. First, I wrote a PDB file reader and a very simple "sphere" renderer, which draws atoms as spheres. Then I implemented "line" and "stick" renderers for bonds. Although this was sufficient for small molecules, it was practically use-

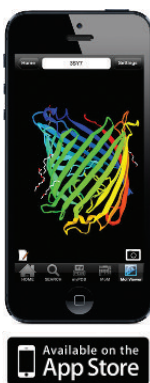
less for large biological macromolecules. As proteins are made up of tens of thousands of atoms, drawing all bonds and atoms quickly fills up the entire screen and makes it visually incomprehensible. Therefore, using a more simplified “ribbon” representation was necessary. However, this is not a trivial endeavor. A ribbon representation traces the protein's backbone, or alpha carbons. To display it as a ribbon, I had to consider what width and thickness to use. The ribbons also had to be smooth; otherwise, beta sheets would wiggle up and down. Of course, there are well-established algorithms for ribbon rendering. But I could not use them without significant modification, because they were too complicated or required too much memory. After trying various approaches, I chose one that gave the best results within the limited capacity of Javascript and WebGL (Web Graphics Library).



GLmol can handle very large molecules such as the enterovirus of PDB ID 3VBF.<sup>8</sup> <http://webglmol.sourceforge.jp/glmol/viewer.html?pdb:3vbf> (enable “biological assembly” in “View” tab)

## Targeting Mobile Devices

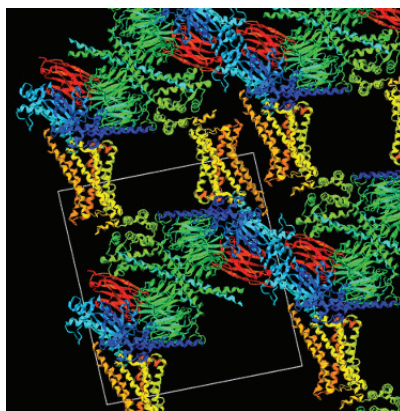
At first I thought GLmol would cover desktop computers (Windows, Mac OS, Linux) and mobile devices (Android, iOS). However, I was disappointed to realize that WebGL was not yet available on most mobile devices. So I rewrote GLmol in C++, ported it to the Android OS platform, and called it NDKmol. I was pleased to find out that NDKmol works smoothly even on my old smart phone running Android 1.6. Then I ported NDKmol's core to iOS, which was adapted into the RCSB PDB Mobile App by Dr. Greg Quinn. NDKmol retains GLmol's very fluid interactivity, and can comfortably display very large molecules.



Outer membrane carboxylate channel seen in PDB ID 3sy7<sup>9</sup> using the NDKmol viewer in RCSB PDB Mobile.

## Biological Assemblies and Crystal Packing

A unique feature of my viewers is the capacity to display biological assemblies and crystal packing based on symmetry operations described in the PDB file. This functionality appears to be missing in most molecular rendering programs for mobile devices, which by default pull data from the coordinate file describing the asymmetric unit, rather than the coordinate file describing the complete biological assembly. This can be very confusing for people without training in structural biology. For example, students might want to look at the structure of potassium channel by the Nobel Prize winner Professor Roderick MacKinnon (PDB ID 1K4C). If they download the asymmetric unit file, they will never see the pore! Fortunately, my viewers use the biological assembly information to display the full structure. But they don't download the coordinate files for the complete biological assembly as that would consume too much memory. Instead, the viewers load the coordinate file describing the asymmetric unit alone, and let the GPU generate the correct symmetry. This process lets my



Crystal packing of adrenergic receptor-G protein complex (PDB ID: 3SN6)<sup>10</sup>. The receptor is in yellow, nanobody in red, G alpha in blue, G beta in green, and G gamma in cyan. <http://webglmol.sourceforge.jp/glmol/viewer.html?pdb:3sn6> (enable “crystal packing” in “View” tab)

viewers display large molecular assembly (like viral capsids) with ease. Similar attention needs to be paid to the effects of crystal packing when examining structures determined by X-ray crystallography. Molecules touch each other in a crystal. This interaction might change the conformation of side chains and flexible loops. By applying crystallographic symmetry operations, my viewers can reconstruct how molecules are packed in the crystal. It's then apparent which part of the molecule might be potentially affected by packing interaction. This may sound very technical, but looking at features such as crystal packing is fun; for example, look at structure 3SN6 by this year's Nobel Prize winner Professor Brian Kobilka. Here the crystal is held together by dense layers of G proteins, nanobody and T4 lysozyme. All of these constituents were necessary to grow well-diffracting crystals of this amazing complex.

## As an open source software

The source code for GLmol and NDKmol are made public under LGPL3 license. GLmol is also released under an MIT license that allows users to adapt the program in any way provided that the copyright notice is kept intact. Dozens of universities, research institutes, e-learning advocates and commercial companies have contacted me to integrate GLmol into their programs and to ask me for feature enhancements. This has been very encouraging and rewarding.

The source code is also very compact and simple. Actually, the entire code for GLmol is smaller than PyMOL's code for cartoon representation alone. The code is a good starting point to learn programming, how to parse PDB files, how to detect bonds, and how to build ribbons from atomic coordinates. You can also learn more general techniques such as OpenGL, HTML5 and multi-touch support. I learned a lot (and continue learning) by reading the source code of open source software and want to keep this cycle going. This is indeed what is happening now. A master course student built a sequence alignment viewer on GLmol. Several educators took GLmol and had summer students implement more features. This is again very stimulating for me.

I have enjoyed developing these molecular viewers, and hope that everyone has a nice time using them!

- X. Wang, W. Peng, J. Ren, Z. Hu, J. Xu, Z. Lou, X. Li, W. Yin, X. Shen, C. Porta, T. S. Walter, G. Evans, D. Axford, R. Owen, D. J. Rowlands, J. Wang, D. I. Stuart, E. E. Fry, Z. Rao. (2012) A sensor-adaptor mechanism for enterovirus uncoating from structures of EV71. *Nat Struct Mol Biol* 19: 424-429.
- E. Eren, J. Vijayaraghavan, J. Liu, B. R. Cheneke, D. S. Touw, B. W. Lepore, M. Indic, L. Movileanu, B. van den Berg. (2012) Substrate specificity within a family of outer membrane carboxylate channels. *PLoS Biol* 10: e1001242.
- S. G. Rasmussen, B. T. DeVree, Y. Zou, A. C. Kruse, K. Y. Chung, T. S. Kobilka, F. S. Thian, P. S. Chae, E. Pardon, D. Calinski, J. M. Mathiesen, S. T. Shah, J. A. Lyons, M. Caffrey, S. H. Gellman, J. Steyaert, G. Skiniotis, W. I. Weis, R. K. Sunahara, B. K. Kobilka. (2011) Crystal structure of the beta2 adrenergic receptor-Gs protein complex. *Nature* 477: 549-555.

## 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  
 Center for Integrative Proteomics Research  
 174 Frelinghuysen 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](http://www.wwpdb.org))

## RCSB PDB Management

**DR. HELEN M. BERMAN**, Director  
 Rutgers, The State University of New Jersey  
[berman@rcsb.rutgers.edu](mailto:berman@rcsb.rutgers.edu)

**DR. MARTHA QUESADA**, Deputy Director  
 Rutgers, The State University of New Jersey  
[mquesada@rcsb.rutgers.edu](mailto: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](mailto:bourne@sdsc.edu)

A list of current RCSB PDB Team Members is available from  
[www.rcsb.org](http://www.rcsb.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*

USA  
 Piscataway, NJ 08854-8087  
 174 Frelinghuysen Road  
 Center for Integrative Proteomics Research  
 Rutgers, The State University of New Jersey

**RCSB PROTEIN DATA BANK**  
[www.rcsb.org](http://www.rcsb.org) • [info@rcsb.org](mailto:info@rcsb.org)