Spring 2009 • Number 41

RCSB PDB PROTEIN DATA BANK

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

NEWSLETTER

Contents

SNAPSHOT: APRIL 1, 2009

56751 released atomic coordinate entries

ΜΟΛΕΧΥΛΕ ΤΨΠΕ	
52444	proteins, peptides, and viruses
1996	nucleic acids
2278	protein/nucleic acid complexes
33	other

ΕΞΠΕΡΙΜΕΝΤΑΛ ΤΕΧΗΝΙΘΥΕ	
48618	X-ray
7776	NMR
230	electron microscopy
127	other

37767 structure factor files
4454 NMR restraint files

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

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

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

Message from the RCSB PDB: Exhibits and a Short Course

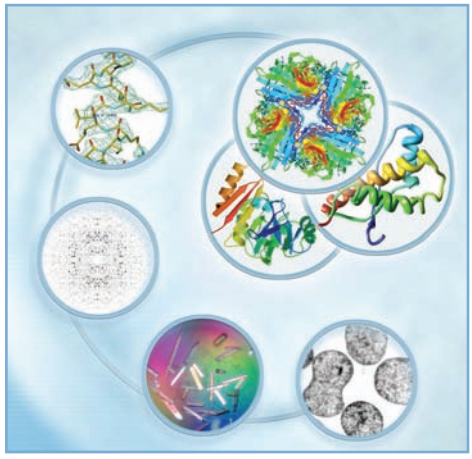
Together with the PSI Structural Genomics Knowledgebase (**kb.psi-structuralgenomics.org**), the RCSB PDB recently exhibited at the 53rd Annual Meeting of the Biophysical Society (February 28 - March 4) in Boston, Massachusetts.

The RCSB PDB and PSI SG KB exhibit booth at the Biophysical Society meeting

The two groups will also be exhibiting at the Experimental Biology meeting (April 18-22) in New Orleans, Lousiana and at the 17th Annual International Conference on Intelligent Systems for Molecular Biology & 8th European Conference on Computational Biology (June 27 - July 2) in Stockholm, Sweden. Please stop by for demonstrations and information, and say hello!

The RCSB PDB will host a short course for practicing pharmaceutical/biophysical modelers looking for a better understanding of crystal structures and PDB data. *Crystallography for Modelers* will be held May 7 & 8 at Rutgers, The State University of New Jersey in Piscataway, NJ.

Do you use protein crystal structure data? Ever wonder why things do not work out as you expect? Are ligand strain energies unreasonably high? Does your drug candidate not appear to make the hydrogen bonds you expect? Is that really a water molecule in the corner of a binding site? Do you wonder why a side chain is in one position rather than another? *Crystallography for Modelers* will offer insight into state-of-the-art research into the quality, errors, and "gotchas" in crystal structure data. Designed by active users of the data, the instructors are RCSB PDB team members, Rutgers researchers, and commercial software developers. It is not an introductory course for crystallography or modeling, but rather a down-and-dirty discussion of crystallographic data, precision, accuracy, and possible errors.



To learn more and register:
rate.rutgers.edu/?q=Crystallography

Crystallography for Modelers *will address issues relating to the "quality" of the data you use. How "good" are the protein structures that you use? What's the error in atomic coordinates? How might you know if something is just plain wrong?*

Data Deposition and Processing

PDB Archive Version 3.15 Released

A newly standardized and enhanced version of the entire PDB archive at ftp://ftp.wwpdb.org has been released.

Files originally released before December 2, 2008 follow PDB Format Version 3.15; files originally released after that date follow Version 3.20. For detailed file format documentation, please see www.wwpdb.org/docs.html.

A time-stamped snapshot of the PDB archive before this release is available from ftp://snapshots.wwpdb.org in the directory 20090316.

Users who maintain local copies of the wwPDB FTP should download the entire archive. Scripts to help in this process are available at www.wwpdb.org/downloads.html.

These data reflect the wwPDB's continuing commitment to providing accurate and detailed data to users worldwide. This release includes improvements and enhancements to the data, including details about the chemistry of the polymer and the ligands bound to it, biological assemblies, and binding sites of ligands and metal ions. An overview is provided at www.wwpdb.org/docs.html.

Questions may be sent to info@wwpdb.org.

Papers Published

Recently published papers describe the wwPDB collaboration and data formats. Other RCSB PDB publications are listed at the RCSB PDB website. Reprints may be requested by contacting info@rcsb.org.

- The PDB format, mmCIF formats, and other data formats (2009) in *Structural Bioinformatics*, second edition, P. Bourne and J. Gu, eds. Hoboken, NJ, John Wiley & Sons, Inc. pages 271-291.
- The Worldwide Protein Data Bank (2009) in *Structural Bioinformatics*, second edition, P. Bourne and J. Gu, eds. Hoboken, NJ, John Wiley & Sons, Inc. pages 293-303.
- Data Deposition and Annotation at the Worldwide Protein Data Bank *Molecular Biotechnology* (2008) [doi:10.1007/s12033-008-9127-7].

Deposition Statistics

In the first quarter of 2009, 2041 experimentally-determined structures were deposited to the PDB archive. The entries were processed by wwPDB teams at the RCSB PDB, PDBe, and PDBj.

Of the structures deposited, 75.9% were deposited with a release status of "hold until publication"; 21.2% were released as soon as annotation of the entry was complete; and 2.9% were held until a particular date. 90.9% of these entries were determined by X-ray crystallographic methods; 8.5% were determined by NMR methods.

During the same time period, 1712 structures were released in the PDB.

CURRENTLY AVAILABLE JOB OPENINGS

The RCSB PDB is looking for:

- **Java Developers**
- **Scientific Software Developers**

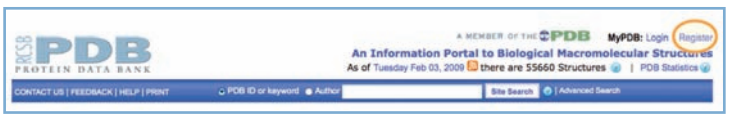
For more details, please see the Job Listings page at **www.pdb.org**.

Data Query, Reporting, and Access

New Feature: Receive Email Alerts When New Structures Match Your Queries with MyPDB

MyPDB is a new feature that regularly sends out emails when structures that match customized queries are released. The matching structures can be accessed directly from the email alerts. Users can also log in to MyPDB to run stored searches at any time.

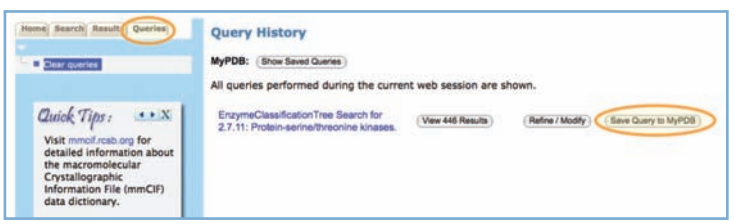
To sign up for MyPDB, users should register using the link at the top right of the RCSB PDB header.



MyPDB registration (circled), login, and logoff links are all found in the top right corner of the RCSB PDB website.

Following registration, an activation link is sent in a confirmation email. After activating the MyPDB account, users can log in at the top of the page. Next, users can review currently saved queries and update MyPDB account information. This page can also be accessed by clicking on the user name at the top of the page.

Saved queries can include combinations of searches by keyword, sequence, ligand, and *Advanced Search*. To store a query in a profile, users should query the RCSB PDB website, such as an *Advanced Search* for structures with Enzyme Commission Number 2.7.11. After evaluating the query, the *Queries* tab in the left menu will temporarily store the search. Selecting *Save Query to MyPDB* option from this tab will store the search in MyPDB.



Queries performed at the RCSB PDB website can be saved in MyPDB for future use. Email notifications can be sent when structures matching saved queries are released.

From the MyPDB page, users can click on the query name and query description to personalize them.

Name	Query Description	Email Notification	Last Run	Next Scheduled Run	Run Query (No Email)	Delete
GFP	gfp	Yes	02/04/2009	02/17/2009	▶	✕
2.7.11	EC Tree Search for 2.7....	Yes	02/11/2009	02/17/2009	▶	✕

Query names, descriptions, and notification schedule can all be customized in MyPDB.

By default, the email notification is turned on. The *Next Scheduled Run* indicates when the query is next run. Email alerts can be set to run weekly (Tuesday afternoon Pacific Time) or monthly (on the first Tuesday).

The email includes the customized query name and a list of matching PDB IDs that have been recently released. Clicking on the query name takes users to the results set at the RCSB PDB website; the results can then be further refined, exported, or further explored. Each PDB ID links to the entry's Structure Summary page.

Questions about this new feature can be sent to info@rcsb.org.

Time-stamped Copies of PDB Archive Available via FTP

Snapshots of the PDB archive (ftp.wwpdb.org) as of January 5, 2009 and March 16, 2009 have been added to ftp://snapshots.rcsb.org. Time-stamped snapshots of the PDB have been archived since 2004. It is hoped that these snapshots will provide readily identifiable data sets for research on the PDB archive.

The script at ftp://snapshots.rcsb.org/rsyncSnapshots.sh may be used to make a local copy of a snapshot or sections of a snapshot. The directories include the experimentally-determined coordinate files that were current at time the snapshote was created. Coordinate data are available in PDB, mmCIF, and XML formats. The date and time stamp of each file indicates the last time the file was modified.

Website Statistics

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

Month	Unique Visitors	Number of Visits	Bandwidth
JANUARY 09	160293	355392	634.85 GB
FEBRUARY 09	168082	373510	767.21 GB
MARCH 09	177316	400950	900.82 GB

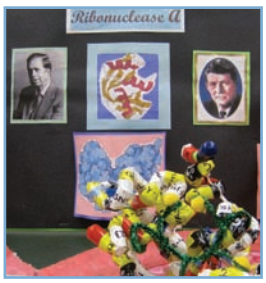
Top Results at the 2009 NJSO Tournaments

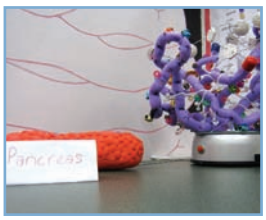
	FIRST PLACE	SECOND PLACE	THIRD PLACE
STATE MARCH 17	BRIDGEWATER-RARITAN HS	WEST-WINDSOR PLAINSBORO HS SOUTH	LIVINGSTON HS
NORTHERN JANUARY 15	WESTFIELD HS	LIVINGSTON HS	MONTVILLE
CENTRAL JANUARY 8	BRIDGEWATER-RARITAN HS	WEST WINDSOR-PLAINSBORO HS SOUTH	PRINCETON HS
SOUTHERN JANUARY 10	MATES ACADEMY	MOORESTOWN FRIENDS SCHOOL	EASTERN REGIONAL HS

Congratulations to all participating teams–there were many great models, abstracts, and responses to the written exam. Special thanks to our judges from the RCSB PDB, the NJ Science Olympiad organizers, and to the MSOE Center for BioMolecular Modeling for the design of this event.

All 2009 results, pictures, and tips for future competitions can be found at education.pdb.org/olympiad.

Outreach and Education





Models built before the competition were judged for accuracy and how well they told the functional story of the molecule.

New Jersey Science Olympiad: Protein Modeling

At Science Olympiad tournaments, teams participate in a variety of events ranging from Egg-O-Naut to Herpetology. Placement in individual events counts towards placement in the overall competition. In New Jersey, protein modeling was a trial event for high school students at all three regional tournaments and at the state finals.

Before the day of competition, teams built a model of the full ribonuclease structure based on PDB ID 1rta. In this model, teams are encouraged to include "additions" that help tell the functional story of ribonuclease. They also provide an abstract/key that describes features highlighted in their model.

At the event itself, teams have 50 minutes to build a section of an assigned PDB structure and answer questions about the structure, function, importance, and history of ribonuclease. For all sections of the event, students are encouraged to use the *Molecule of the Month*, the PDB file, Jmol and the Structure Explorer page. Resources to help students prepare for the event are available at education.pdb.org, including a video that demonstrates how to build a 3D model of a zinc finger.

State Champions (and winner of the Central Regional) from Bridgewater Raritan High School.

The Southern Regional Champions from the MATES Academy team built a portion of ribonuclease using the RCSB PDB website site and MiniToobers (from 3D Molecular Designs, www.3dmoleculardesigns.com).

A judge described how the models were scored with the Northern Regional Champions from Westfield High School.

RCSB PDB annotators judged the protein models built onsite for accuracy.

Education Corner by Gavin Whittaker, Ph.D., Miramodus Ltd.

Molecular Models: Tangible Representations of the Abstract

GAVIN WHITTAKER
(gavin@miramodus.com) is the managing director of Miramodus Ltd., based in Edinburgh, Scotland. He holds graduate and postgraduate degrees from the University of Oxford, and worked as a lecturer in physical chemistry at the University of Edinburgh for several years. He maintains academic and business interests in his research area of microwave-induced chemistry. Miramodus itself, formerly known as Beevers Miniature Models, was formed as a 'spin-out' company from the University in 2007. More information about the business and its models can be found at www.miramodus.com.

An email drops into my inbox from somewhere in the world, from someone that I will probably never meet, asking whether we could produce a model of a protein. I log into a protein structure database and check out the size and shape of this molecule. Over the next few days, mail flies back and forth between us, in which time we determine whether he wants to show every atom, or just the backbone carbons; whether they want prosthetic groups or ligands included in the model; and what colors they would like. The design remit is set and a few weeks later, the custom-made model is shipped off to a (hopefully) delighted customer. Ours is just one possible model type that can be used to represent that protein, but in this case, our ball and spoke model illustrates the protein function in a way that the customer wants for when he has to explain his research to visitors.

The Oxford English Dictionary defines a model as *a representation in three dimensions of some projected or existing structure, or of some material object, artificial or natural, showing the proportions and arrangement of its component parts.*

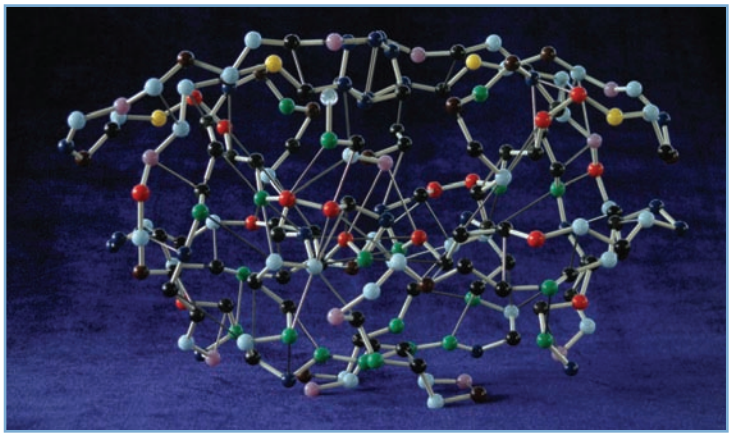
We are familiar with models in our everyday lives, perhaps so much so that we lose sight of what it is that we are seeing. Most people encounter, or even build, models of mechanical structures, buildings, or different forms of transport. Such models are often scaled down versions of larger structures, and in many respects they are very similar to the objects that they represent, being scaled down by factors of only 10 to 1000 times. Molecular models are very different from these engineering models. For a start, we are scaling *up*, typically by a factor of 10^{10}. Furthermore–and this is something that we seldom consider–we are attempting to produce models of structures that are already models, albeit mathematical ones. Those mathematical models themselves are forced to make assumptions and approximations, such as the hydrogen-like nature of orbitals in larger atoms. At best, we are talking about creating models of a model of an approximation; illustrations of descriptions of reality.

Unlike engineering models, molecular models can never be perfect replicas of molecular structures. The difference between the quantum world and the macroscopic world is, in that sense, an unbridgeable gap. When we decide to construct a model of molecular structures, therefore, the first thing that we must do is to decide what the *purpose* of the model is. If a molecular model cannot be in any way a replica, then the purpose of the model must be the representation of a particular aspect of the molecule.

Different types of molecular models illustrate different aspects of the molecules that they represent, but there are two fundamental things that they show: either the position of the nuclei, or the volume occupied by the electrons. The former, typified by ball and rod structures, allows us to visualise the relative positions of the atoms and whether or not there are bonds between them. The latter, almost always space-filling models, allows us to see the volume occupied by the molecules. More often than not, we begin our education in molecular structures at school by drawing structures that comprise element symbols linked by lines or grouped in units such as -CH_3. The great advantage of space filling models at this stage of a science student's education is that they show how the simple 'structures' written on paper are actually representations of electrons held in check within volumes whose shapes are defined by the properties of the nuclei. As we become more familiar with the nature of chemical bonds and electron orbitals, the size and shape of the electron clouds (and hence the shapes of the molecules) become implicit to us in the structures that we write. Indeed, most of the time, it is doubtful that many of us even give a second thought to this process. For the majority of the time, it is therefore sufficient for us to represent the positions of the nuclei–we automatically 'fill in' the volume of the electrons clouds for ourselves. This is akin to a veterinarian seeing a skeleton of an animal and visualising the living creature and the way that it moves. This is obviously tremendously useful–the fact that we can, to some extent, extrapolate the electron clouds of a molecule from the position of the nuclei, almost without thinking, allows us to get more from 'skeletal' structure models than might otherwise be the case. This ability allows us to use the most common form, 'ball and spoke' models, meaningfully in our teaching. Function in chemistry has as much to do with shape and form as anything else, and the fact that it is possible to see right through a ball and rod structure to view atoms on the far side of the model opens up the model's teaching potential enormously. It allows the viewer to clearly see the range of atoms that make up the structure (the constitution) and the manner in which the atoms are joined together (the configuration) to form the shapes that those linkages produce (the conformation). Were we to represent the same structure with space-filling units, it would hardly be possible for us to see beyond the outermost layer of atoms. That we can see through and into the structure allows enormous amounts of information to be conveyed in a simple and elegant form.

This is where we come in. Our particular business is to make those models for people: businesses, museums, academics, individuals–even lawyers.

This is the business that Arnold Beevers started as a unit within the University of Edinburgh, some 30 years ago. Following Arnold's retire-



A model of Proteinase-C, built to show the positions of the α-carbons.

ment, Sheila Gould took over the running of the business for several years. Eventually, Sheila's retirement and that of most of the staff, combined with a shift away from universities running businesses, meant that someone had to take over 'Beevers Models' or it would have simply ceased to exist. This would have been a tragedy–Beevers models grace museums and academic units around the world. In the words of the late Victor Kiam, former owner of Remington; "I liked [the product] so much, I bought the company," and have been running it ever since. We took on new staff, and although we still work in the University of Edinburgh's School of Chemistry, where I lectured for several years, we are now completely separate from it in administrative terms.

Most of our business centers around building ball and spoke models of inorganic structures and small organic molecules. Whilst not always trivial to design, there is little debate over whether to include all the atoms in these structures–it is *precisely* the relations between all of the atoms that people want to see. When we build models of proteins, though, things are a bit more complicated. The number of atoms in an average protein would make the costs astronomical in many cases. However, most representations of proteins are approximations–just as we learn to imply the presence of electrons around an atom in an inorganic structure, we learn to "see" the structure that accompanies such symbols as 'gly','asp', *etc.* When we build protein models, therefore, we often make the same simplification. Taking the alpha carbon as our residue 'centre,' we represent each one with a different coloured ball. By using a scale of 1cm=2Å, we produce models that are small enough to handle, but large enough to show the structure and, where necessary, such things as haem groups or iron-sulphur clusters. We are very aware that the models of proteins that we build, like all protein models, are not a "likeness" of the protein, but they *are* a representation of its structure that the owner can use to see relative distances and forms that cannot be as easily seen on a computer screen.

We are often asked how long we think that the business will last in a world of ever more realistic computer graphics. The fact is that our 3D vision is a far richer phenomenon that the result of two slightly dissimilar images. Depth of field and the ability to refocus in an instant all add to our experience. As a consequence, the three dimensional nature of physical models is much more easily "seen" than even the best computer graphics. Maybe the greatest irony of all, though, is the number of model sales to theoreticians who can't easily visualise or demonstrate the interactions in the molecules that they are studying.

Ultimately, many people simply take pleasure from owning the tangible and elegant object that represents their research interests. So long as people want to have their prized molecules, crystal structures and proteins on display, rather that appearing as a virtual entity on a screen, and as long as people need to have a physical representation of reality to hold, we'll still be here to make the models for them.

PDB Community Focus

Gaël McGill, Ph.D., and Graham Johnson, Molecular Animators

Images of molecules are becoming more and more common in educational and entertainment media. These pictures are often created by computer graphics artists using state-of-the-art programs such as Maya and Cinema4D. However, the methods used to import PDB structures into these advanced programs can be challenging. David Goodsell recently spoke with two molecular graphics professionals to see what is available and what still needs to be done.

Q. ***First off, can you tell me a bit about yourselves and the work you are doing in molecular graphics?***

A. **Gaël McGill**: Cell and molecular biology has been a passion from my middle school days. I came to the USA as an undergraduate specifically to study molecular biology and my dream was to be involved in research. I went on to do my Ph.D. at Harvard Medical School (mostly focusing on cancer signal transduction pathways and apoptosis using a varied mix of cell and molecular biology, biochemistry, animal model/genetics, and screening approaches). Having identified a need in the local academic, medical, biotech and pharmaceutical communities for "scientifically-informed" graphic design and web programming services, I started my company Digizyme (**www.digizyme.com**) in 1999 during my Ph.D. years. On a personal level, this really served as a creative outlet outside of the long hours at the bench. Digizyme has grown to offer more advanced services in recent years–including 3D animation services and even product design and visualization for the biomedical device industry. Over the past few years, I have "reintegrated" into academia in hopes of establishing a full-time team of scientist-animators at Harvard Medical School (where I currently teach Maya molecular visualization classes year-round). I enjoy the variety and challenge of client-driven projects as part of my work at Digizyme, and look forward to the freedom of pursuing larger-scale, longer-term collaborative projects relating to fundamental cell/molecular visualization challenges as an academic.

A. **Graham Johnson**: I graduated from the Department of Art as Applied to Medicine in the Johns Hopkins School of Medicine with a master's degree in Medical and Biological Illustration in 1997. At Hopkins, we studied anatomy and physiology with medical school students while simultaneously sketching and illustrating our dissections along with autopsies and surgeries observed at the hospital. After graduating, I focused on studying molecular and cell biology while illustrating the textbook *Cell Biology* with Tom Pollard and Bill Earnshaw. I began animating this content for other clients and realized that trained medical illustrators could contribute a great deal to this relatively unil-

FOR MORE INFORMATION

The full interview and a glossary terms used in this article are available from the online edition of this newsletter at **www.pdb.org**.

GAËL MCGILL *His website at* **www.molecularmovies.org** *hosts cell and molecular animations, and tutorials on how to import PDB data into programs such as Maya.*

GRAHAM JOHNSON *publishes animations and more at FiVth Medical Media* **(www.fivth.com).**

lustrated subject. It occurred to me that one could simulate most of my animations either by scripting the physics or by translating imported data to a format that my 3D software package could recognize. I began simulating a handful of my client animations, but quickly learned that the out-of-the-box software was not intended for this purpose and could only be used for inaccurate and simple molecular interactions. In 2005, I applied to the Ph.D. program at The Scripps Research Institute to work in the Molecular Graphics Laboratory under Arthur Olson to better understand the content and to communicate directly with a team of talented molecular graphics coders. As I head into my fourth year of the program, I'm finally producing some useable scripts that have made it easier to create illustrations and animations of molecular realms. I hope to distribute many of the tools very soon.

Q. *How do you import PDB structures? Are these tools generally available?*

A. **Gaël McGill**: For the most part, we use existing molecular graphics applications like Chimera and PyMOL to generate geometry files. These are typically exported as VRML (Virtual Reality Modeling Language) and then converted to OBJ format (a common data format for 3D data) before being imported into Maya. We also use MEL (Maya Embedded Language) scripts–either ones already available online (although currently there are not many related to PDB), or in-house/custom ones. Which method we use depends on what we will be doing with the geometry once inside Maya. A great option for bringing in large PDB datasets has been Chimera's "multiscale models" feature. Eventually it would be great to create a similar functionality for creating polygonal models within Maya itself in order to have more control over the output geometry. Still, this type of tool has been very useful in creating animations showcasing large complexes (like entire viruses).

A. **Graham Johnson**: I've written a COFFEE plug-in (Cinema4D's native scripting language) that imports a single PDB file or a list of PDB files directly into my viewer window as a set of points in space (used to generate smooth surface models such as metaballs), CPK spheres, or backbone spline. I'm building a primitive ribbon generator and hope to make the tools available for use within the next year. If I require a more sophisticated surface model, *e.g.*, one colored by electrostatic potential, I'll export it from one of the popular molecular viewers as either a VRML or an OBJ file. Again, for static images, I'll often just export a screen grab from a molecular viewer that offers a style I'm after.

The Synapse Revealed *created by Graham Johnson for the* Howard Hughes Medical Institute Bulletin *©2004*

Q. *What type of molecular imagery is most popular with your clients?*

A. **Graham Johnson**: Because molecular graphics viewers are so user-friendly these days, clients rarely come to me to request an image or movie of a single molecule spinning on a monochromatic background. Most of my clients ask me to generate an editorial image, or to illustrate or animate a process or cell event involving multiple molecule types.

A. **Gaël McGill**: Although it depends on the project, I find that clients (especially biotech and pharma) want images of molecules "in context." In other words, scenery that captures a molecular process of interest but also places it within a cellular landscape. The challenge is to create a still image that captures or suggests a narrative or mechanism. . . essentially an "action shot" in which the visual context of the structure being depicted (and its binding partners) helps to communicate its function.

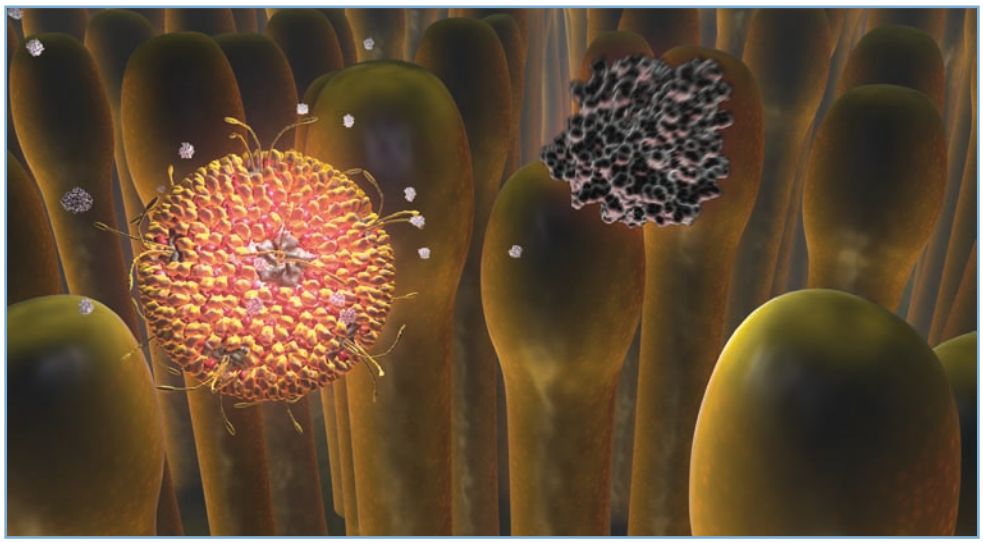
Q. *I imagine that assembly of biologically relevant complexes (such as chromatin or a transcription complex) and modeling their dynamics poses difficult challenges—what types of tools do you use for this?*

A. **Gaël McGill**: This is one of the toughest challenges at the moment (and it does not only apply just to complexes): how does one visually represent the dynamic aspects of proteins based on available (mostly static) data? The ability to create linear morphs between multiple conformational states of a protein using the adiabatic mapping technique (used by Mark Gerstein's method at **www.molmovdb.org**, for example) is very useful to visualize one possible trajectory, but it is only one possible trajectory and it also cannot tackle more complex morphs that involve partial refolding of protein domains. Drew Berry at The Walter and Eliza Hall Institute of Medical Research has pioneered a visual style that suggests the dynamics of proteins, but it would be nice to create animations that are based on actual data for these dynamics (*i.e.*, as opposed to using noise/fractal motions throughout, having vibrations and degrees of flexibility that reflect the protein's actual range of 'thermodynamically-permissible' motion). In packages like Maya, we are currently limited to using pretty basic kinematic tools (*i.e.*, building rigs driven by forward or inverse kinematics) that intrinsically have no knowledge of the molecular structure and its limitations or range of permissible torsion/bending. The software does not even register or warn against impending self-intersections–a problem that we are currently exploring in collaboration with topologists/software developers from the entertainment industry. At the moment (and depending on the target audience), we try to find as many sources of reference data as possible and use them as "inspiration" to create a dynamic representation of a protein or complex. The goal is to find more direct ways of integrating these data into the visualization (inasmuch as it helps communicate crucial parts of the story).

A. **Graham Johnson**: I've attempted to rig a handful of complex builders over the years with out-of-the-box toolsets. Such tools often do a great job of roughing a concept together, but fail when applied to large-scale systems or attempts to accurately simulate the rigor and detail often required for molecular imagery. Years ago, for example, I tried, to stitch together thousands of blocks with pairs of springs to represent the persistence length and flexible backbone of DNA in a plasmid. I animated a twist to see if it would supercoil, but the collision detector would always overload and the system would come to a screeching halt before the DNA could achieve a single twist. I've tried pouring virtual molecules into virtual organelles to fill them with random recipes of non-colliding molecules, but again, the technique has always proved to be slow, limited in volume, and relatively uncontrollable. Most particle generators I've toyed with have similar limitations to their physics simulation. To overcome these challenges, I've begun to construct scripts from scratch that attempt to combine the capabilities of simplified molecular dynamics with the visualization power of commercial 3D software.

Q. ***Are there any resources that you would suggest to artists interested in incorporating PDB structures into their work?***

A. **Gaël McGill**: Other than the fantastic PDB itself (not sure what we would do without it!), I recently launched a free resource for scientists interested in learning 3D software packages for cell and molecular visualization at **www. molecularmovies.org**. One section is a showcase/ directory of some of the web's best cell and molecular movies (organized by scientific topic), and another is dedicated to tutorials and lectures. There are currently hundreds of pages of free tutorials that approach learning Maya in the context of biological visualization. More specifically, several of these tutorials focus on getting PDB data into 3D applications like Maya. Expansion of the site in the near future will also include a "Toolkit" section where animators can share scripts and plug-ins for PDB import (and other tasks related to molecular animation), and a new section that provides a more general directory of visual resources. The idea behind this last section is to find and organize non "narrative-driven" raw data visualizations (*i.e.* like time-lapse movies, MD simulations and other datasets) that animators can use as reference materials to create better visualizations.



Early Events in Reovirus Entry *by Gaël McGill. The full movie can be viewed online at* www.molecularmovies.com/movies/mcgilliwasa_reovirus.html

A. **Graham Johnson**: The updated and integrated Electron Microscopy Data Bank (**emdatabank.org**) offers many low-resolution models of macromolecular structures and has a new online EM viewer. Many files in the PDB exist as low resolution structures with only alpha carbon coordinates published. If you need a rough approximation of the sidechains to generate a teaching model for such a molecule, you can generate or download a pre-generated version from MaxSprout (**www.ebi.ac.uk/Tools/maxsprout**). Lastly, I find the TransMembrane PDB indispensable (**pdbtm.enzim.hu**).

Q. ***Have you had any projects that posed an insurmountable challenge?***

A. **Gaël McGill:** The great thing about cell and molecular visualization is that there is an endless source of topics/mechanisms to visualize and each of these come with their own unique challenges. We may not always use the optimal solution or have the perfect tool available, but there is almost always a creative way to solve the visual representation challenges that emerge. It is one of the aspects of visualization with powerful packages like Maya that make this work so fresh and exciting!

A. **Graham Johnson**: Many projects have and I've often had to truncate my personal goals or compromise with the client to find some workaround because of strict deadlines. In years past, I sometimes had to resort to keyframe animation, hand drawn animation, simple 2D vector animation, and even static imagery to convey a message that could have been most clearly presented as a 3D animated sequence. I simply lacked the technology, skill, or time. Finding out, however, that molecular animation posed more challenges than my other medical illustration jobs directly inspired me to build tool sets to help meet such challenges.

Q. ***What new tools would you like to have?***

A. **Gaël McGill**: As noted above, we are in the process of creating a suite of MEL scripts that can address some of the basic geometry-building tasks for getting PDB data into Maya (without having to resort to molecular graphics software-exported meshes). Once we have this first set of scripts (that just focus on efficient/clean geometry creation), the next step would be to explore the development of programmatically-driven rigging tools for defining the articulation of the models. In other words, to write scripts that not only create Maya-native geometry directly from the PDB but also automatically create a rig that has some inherent motion constraints applied. This is easier said than done and will of course depend on the type of molecular representation (ball & stick versus cartoon for example would have very different 'rules' applied to constrain motion). Having geometry that is more 'self-aware' (and that can at least avoid or warn about self-intersections) would be useful.

A. **Graham Johnson**: I agree that methods for exporting molecular models in styles that are animation-ready would be very helpful to everyone in the molecular illustration field. I would primarily like to see an extension of the PDB that offers biological unit matrices to help users generate pertinent symmetries. This works great in PDB files for viruses that have BIOMT lines in REMARK 350 to describe the transformation and orientation matrices needed to generate a complete virus[1]. More specifically, I'd love to see this for other common cell complexes. How can one generate an *in vivo* microtubule with 13 protofilaments and a proper seam from 1TUB for example? What rotation per *y* translation might a user need to enter to generate an actin filament from an actin monomer? A handful of filamentous files exist in the PDB, but animators can benefit from viewport and render time efficiencies afforded by modern software by cloning a single molecule rather than rendering coordinates from the thousands of copies of 1TUB needed to generate a lengthy microtubule.

A. **Gaël McGill**: Basic collision detection is also not easy to implement at the moment (whether between different parts of the same continuous mesh or between meshes). Some way of integrating electrostatic forces would also be amazing! Better simulation tools would also help us create molecular vistas with some semblance to what is happening *in vivo*. By simulation I don't mean at the same atom-by-atom level that molecular dynamics offers, but something that would drive the stochastic behavior of numerous molecules within a defined volume or environment, for example.

Finally, an area that is ripe for exploration: we need to tap into the full promise of educational gaming and interactive environments by harnessing the power of modern gaming engines. In many cases, the digital assets (models, textures, rigs) used to develop high-end games are created in packages like Maya. So one could easily imagine a scenario where a lot of the work being done to create 'narrative-driven' molecular movies in Maya could be repurposed and adapted to generate interactive molecular environments for educational purposes.

[1]Representation of viruses in the remediated PDB archive (2008) Acta Cryst. D64: 874-882.

RCSB PDB Partners

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

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

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

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

RCSB PDB Management

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

DR. 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.*

RCSB PROTEIN DATA BANK
www.pdb.org

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

Return Service Requested