Winter 2007 • Number 32

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

RCSB PDB PROTEIN DATA BANK NEWSLETTER

Contents

Message from the RCSB PDB 1

DATA DEPOSITION AND PROCESSING

2006 Statistics 2
DOIs Available for Released Entries in the PDB Archive 2
RCSB PDB Focus: The ADIT Help System 2
Annotation at the RCSB PDB 2
RCSB PDB Focus: Depositing New Chemical Components (Ligands) 2

DATA QUERY, REPORTING, AND ACCESS

RCSB PDB Focus: Sorting Search Results and Tabular Reports 3
RCSB PDB Focus: Exploring Domains in Protein Structure . . 3
Searching for Sequence Variants 4
Website Statistics 4
RCSB PDB Focus: External Links 5

OUTREACH AND EDUCATION

PDB Structures on Exhibit at the Birch Aquarium 5
wwPDB Paper Published 5
RCSB PDB Poster Prize Awarded at AsCA 5
Meeting News 5

MOLECULES OF THE QUARTER 6

EDUCATION CORNER: Methods in Structural Biology Course at Cold Spring Harbor Laboratory 7

PDB COMMUNITY FOCUS: Julian Voss-Andreae, Protein Sculptor 9

RCSB PDB PARTNERS, MANAGEMENT, AND STATEMENT OF SUPPORT 12

SNAPSHOT: JANUARY 1, 2007
40933 released atomic coordinate entries

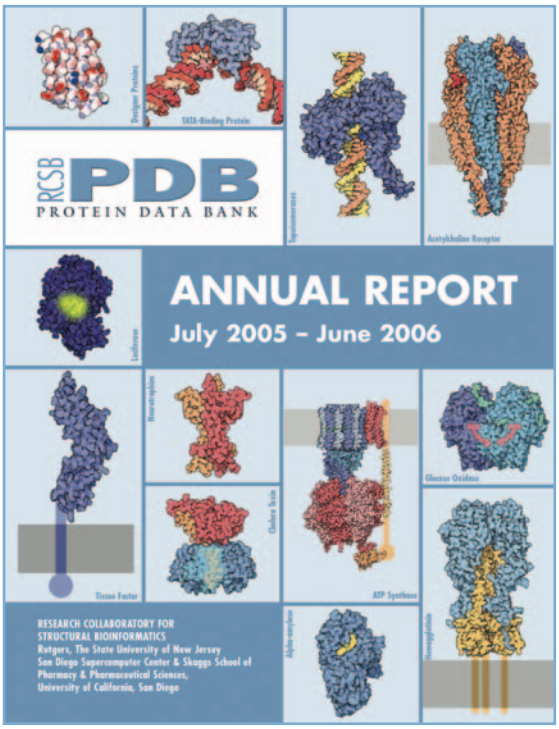
MOLECULE TYPE		EXPERIMENTAL TECHNIQUE	
37537	proteins, peptides, and viruses	34672	X-ray
1687	nucleic acids	6035	NMR
1674	protein/nucleic acid complexes	142	electron microscopy
35	other	84	other
		23871	structure factor files
		3282	NMR restraint files

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



Covering the period of July 1, 2005 – June 30, 2006, the RCSB Protein Data Bank's Annual Report documents the database and website released during this period.

The Annual Report provides background information about the RCSB PDB resource and describes current progress and accomplishments. Available online as a PDF, this snapshot also explores the RCSB PDB's different activities in data deposition, structural genomics, and education.

The cover highlights the structures featured in the Molecule of the Month during the report period. Each installment introduces readers to the structure and function of a particular molecule, and discusses the relevance of the molecule to human health and welfare. Written and illustrated by David S. Goodsell (The Scripps Research Institute), the series is a great place for users of all levels to start exploring the RCSB PDB resource. This cover, along with information about the proteins, is also available as a downloadable PDF flyer.

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

Best wishes for 2007 from the RCSB PDB

Data Deposition and Processing

2006 Statistics

In 2006, 6911 experimentally-determined structures were deposited to the PDB archive.

The entries were processed by wwPDB teams at the RCSB PDB, MSD-EBI, and PDBj. Of the structures deposited in 2006, 71.8% had a release status of "hold until publication"; 16.1% were released as soon as annotation of the entry was complete; and 12.1% were held until a particular date.

86.7% of these entries were determined by X-ray crystallographic methods; 12.8% were determined by NMR methods. 87.5% of these depositions were deposited with experimental data.

During the same period of time, 6910 structures were released into the archive.

DOIs Available for Released Entries in the PDB Archive

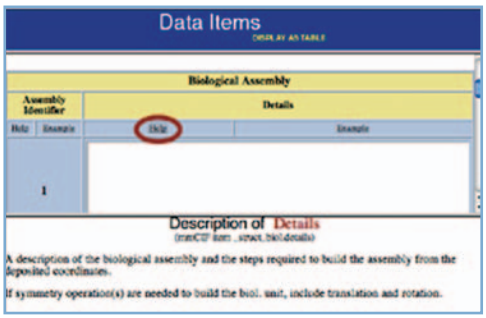
Structures released by the wwPDB into the PDB Archive are now being assigned a Document Object Identifier (DOI). The DOI System is used to identify content objects (such as journal articles, books, and figures) in the digital environment.

The DOIs for PDB structures all have the same format – 10.2210/pdbXXXX/pdb – where XXXX should be replaced with the desired PDB ID. For example, the DOI for PDB entry 4HHB is "10.2210/pdb4hhb/pdb". This links directly to the entry in the PDB file format on the FTP server.

The DOI can be used as part of a URL to obtain this data file (**dx.doi.org/10.2210/pdb4hhb/pdb**), or can be entered in a DOI resolver (such as **www.crossref.org**) to automatically link to pdb4hhb.ent.Z on the main PDB ftp archive (**ftp://ftp.rcsb.org**).

DOIs are automatically registered by the wwPDB when entries are released after the weekly update. They will not be available before a structure's release. Along with the ftp location, the DOIs for PDB entries also include the entry title, the authors, and the deposition date.

RCSB PDB Focus: The ADIT Help System



The ADIT help example for biological assembly details

The deposition tool ADIT (at RCSB or PDBj) includes examples and definitions provided in the PDB Exchange Dictionary as guides for users depositing their structures.

An explanation for each piece of information requested by ADIT can be obtained by selecting the Help button located next to the named data item. This information will appear in the bottom frame. Pressing the Help button in the top frame will display these instructions.

Examples for these items can be obtained by selecting the Example button within the table. This information will appear in the bottom frame.

At any time during deposition, you may view the current state of the entire entry by pressing the PREVIEW ENTRY button.

Questions about ADIT should be sent to **deposit@deposit.rcsb.org**.

Annotation at the RCSB PDB

The question "So, what does an 'annotator' do?" has been answered with the article:

A Biocurator Perspective: Annotation at the Research Collaboratory for Structural Bioinformatics Protein Data Bank Kyle Burkhardt, Bohdan Schneider, Jeramia Ory (2006) *PLoS Comput Biol* **2(10)**: e99

The typical day of an annotator and the challenges facing PDB curators are described. This issue of PLoS Computational Biology also contains an editorial recognizing the efforts of biocurators worldwide and a description of the curation process in use at the Immune Epitope Database and Analysis Resource.

The RCSB PDB is looking for Biochemical Information & Annotation Specialists to curate and standardize macromolecular structures for distribution in the PDB archive. As described in this article, the annotation specialist communicates daily with members of the deposition community, and annotates, releases, and updates entries in the PDB archive. BLAST, PubMed, and other tools are used for the annotation process performed on a linux box.

A background in biological chemistry (PhD, MS, BS or BA) is required. Experience with linux computer systems, biological databases, crystallography, and/or NMR spectroscopy is a strong advantage. The successful candidate should be self-motivated, pay close attention to detail, possess strong written and oral communication skills, and able to meet deadlines. The position is located in Piscataway, NJ.

This position offers the opportunity to participate in an exciting project with significant impact on the scientific community.

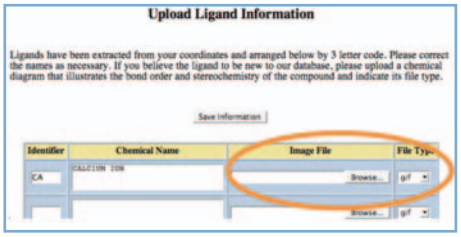
Please send resumes to Dr. Helen M. Berman at **pdbjobs@rcsb.rutgers.edu**.

RCSB PDB Focus: Depositing New Chemical Components (Ligands)

To deposit new ligands, please check Ligand Depot to see if the ligand, drug, ion, non-standard residue, modified residue, group, *etc.* is present in our chemical component dictionary.

If the ligand is present, please make sure that the 3 letter code for the ligand

in your file matches the one used in the chemical component dictionary.



The frame to upload new ligands in ADIT

If the ligand is not present in the dictionary, users can now upload a 2-D figure of the structure as part of the ADIT deposition process in PostScript, TIFF, or GIF formats. From the left-hand categories menu, scroll down and select the "Upload Supplemental Information: Ligand Information" to provide this information. Although the 3 letter code used for the ligand has no specific significance, you may check Ligand Depot and select a code for your ligand that has not been taken, otherwise, one will be selected for you.

Questions about ligands should be sent to **deposit@deposit.rcsb.org**.

Data Query, Reporting, and Access

Sorting Search Results and Tabular Reports

The RCSB PDB offers many ways of looking at the information contained in the database. After searching for a set of structures, users can explore individual structures or examine the whole set by creating reports.

A search results list sorted by Release Date

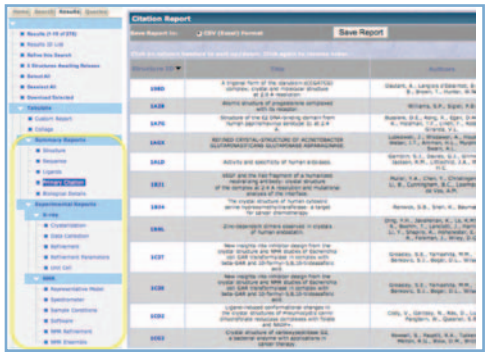
Following a search that produces multiple entries, the results set can be sorted by choosing 'Sort Results' from the menu on the left hand side of the page. Sorting options include: PDB ID, Release Date, Residue Count, Resolution and Rank. An Advanced Search by sequence (Advanced Search>>Sequence Features>>Sequence (Blast/Fasta)) allows the user to sort results by PDB ID, formula weight and E value.

A collage report

Another option for viewing multiple structure results is available from the left menu's "Tabulate" button. It allows the user to create tables of various structural and experimental properties that can be downloaded as CSV files.

These tables can be sorted by clicking on the column headers. Clicking again reverses the sort order.

The "Custom Report" option lets the user select which columns will be included in the report. "Collage" will tile the thumbnails of all of the molecular images.



Citation report for a search results set.

Default reports about structure, sequence, ligands, primary citation, and biological details are available from the Summary Reports option. Experimental reports can also be created for X-ray (crystallization, data collection, refinement, refinement parameters, and unit cell) and NMR (representative model, spectrometer, sample conditions, software, refinement, and ensemble) structures.

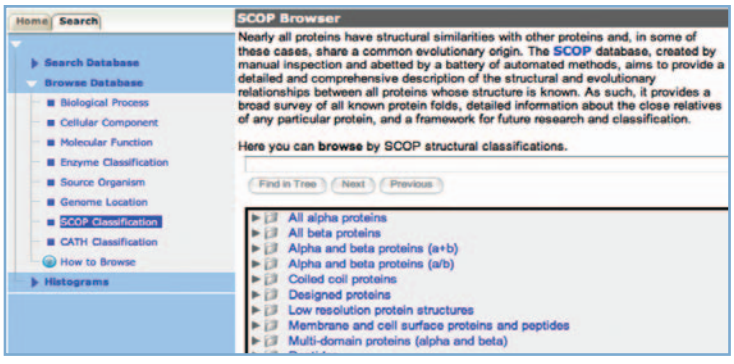
RCSB PDB Focus: Exploring Domains in Protein Structure

Domains can be thought of as the smallest structural units from which proteins are assembled that retain properties of the whole protein, such as a hydrophobic core. In certain cases, domains can also function independently from the rest of the structure. Any given protein structure is comprised of one or more domains from which the overall properties of the protein are derived. Analyzing a protein structure from the point of view of its composite domains is an important, yet not fully solved problem.

The RCSB PDB offers various ways of exploring domains in protein structures:

PDOMAINS (**pdomains.rcsb.org/pdomains**)[1] is a resource centered around the definition and assignment of structural domains in proteins. It offers analysis of existing approaches to domain definition and provides a benchmark dataset to evaluate and cross-compare automatic domain assignment methods.

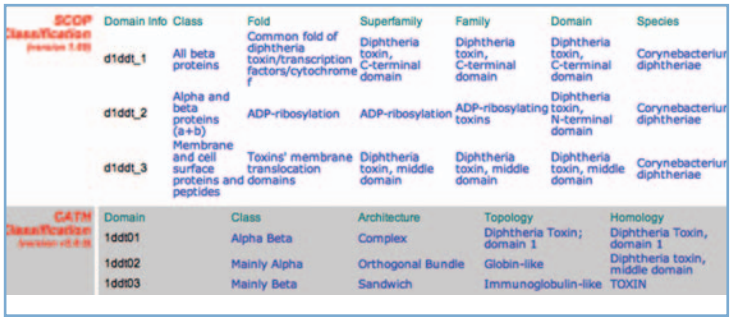
The Browse Database option of the 'Search' tab on the left-hand menu offers a tool to explore the hierarchically organized and curated domain definitions produced by SCOP (**scop.mrc-lmb.cam.ac.uk/scop**) and CATH (**www.cathdb.info**).



The RCSB PDB's SCOP Browser

[1] Holland TA, Veretnik S, Shindyalov IN, Bourne PE. Partitioning protein structures into domains: why is it so difficult? *J Mol Biol.* 2006 361(3):562-90.

Structure Summary pages for individual structures provide links to sets of all structures containing domains similarly categorized by SCOP and CATH. Each category is a link to a result.

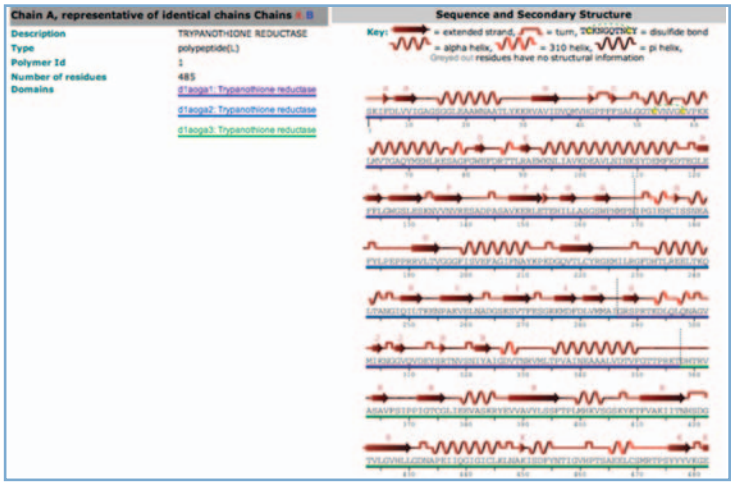


SCOP	Domain Info	Class	Fold	Superfamily	Family	Domain	Species
	d1ddt_1	All beta proteins	Common fold of diphtheria toxin/transcription factors/cytochrome f	Diphtheria toxin, C-terminal domain	Diphtheria toxin, C-terminal domain	Diphtheria toxin, C-terminal domain	Corynebacterium diphtheriae
	d1ddt_2	Alpha and beta proteins (a+b)	ADP-ribosylation	ADP-ribosylation	ADP-ribosylating toxins	Diphtheria toxin, N-terminal domain	Corynebacterium diphtheriae
	d1ddt_3	Membrane and cell surface proteins and peptides	Toxins' membrane translocation domains	Diphtheria toxin, middle domain	Diphtheria toxin, middle domain	Diphtheria toxin, middle domain	Corynebacterium diphtheriae

CATH	Domain	Class	Architecture	Topology	Homology
	1ddt01	Alpha Beta	Complex	Diphtheria Toxin; domain 1	Diphtheria Toxin, domain 1
	1ddt02	Mainly Alpha	Orthogonal Bundle	Globin-like	Diphtheria toxin, middle domain
	1ddt03	Mainly Beta	Sandwich	Immunoglobulin-like	TOXIN

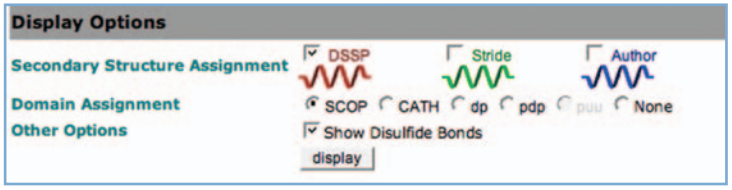
SCOP and CATH information on an entry's Structure Sumary page

Sequence Details pages for each structure illustrate domains aligned with sequence and secondary structure. The colored domain definition links in the SCOP Domains section coincide with the colored bars underneath the sequence to indicate the location of the domains. Boundaries between domains are highlighted further with vertical dashed lines.



View of an entry's Sequence Details page

Users can choose domain definitions according to either SCOP or CATH and the programs DomainParser (compbio.ornl.gov/structure/domain-parser) and PDP (123d.ncifcrf.gov/pdp.html).



Options available from an entry's Sequence Details page

Searching for Sequence Variants

Protein structure sequences are assigned UniProt/SwissProt IDs (UNP/SWS). The new 'Sequence Variants/Non-variants' RCSB PDB feature lets users retrieve all structures with a particular UNP/SWS ID, grouped by the presence or absence of sequence variations (variants or non-variants). Searches for variants will provide structures with post-translational modifications, whereas searching for non-variants will provide occurrences of structures that have at least one identical polypeptide chain.

This query can be run using the Advanced Search option.

- Click on the 'Search' tab on the left-hand menu, and expand the 'Search Database' menu option.
- Click on 'Search Database'>>'Advanced Search'
- Select 'Choose a Query Type'>>'Sequence Features'>>'Sequence (Non/)Variants'
- Enter a structure ID
- Select the desired chain from the pull-down menu
- Select 'No' to retrieve all structures whose sequences do not vary from the reference sequence due to point mutations/insertions/deletions.

The Advanced Search form for Sequence Variants

For example, using 1LJ3 chain A, variant = 'No' will pull up structures of lysozyme with no sequence variations.

Variant = 'Yes' retrieves all structures whose sequences vary from the reference UNP/SWS sequence due to point mutations/insertions/deletions.

Variants and non-variants can also be retrieved from the Structure Summary page for any structure having such variants with the same UNP/SWS ID. From the lefthand menu, select Structure Analysis-> Sequence Variants. The resulting page displays the pairwise alignment of the structure sequence and the UNP/SWS sequence for each structure.

2006 Website Statistics

Access statistics for the year 2006 are given below for the RCSB PDB website at www.pdb.org.

MONTH	UNIQUE VISITORS	NUMBER OF VISITS	BANDWIDTH
JANUARY	82,372	183,705	488.54 GB
FEBRUARY	91,159	195,783	625.25 GB
MARCH	104,495	230,746	527.27 GB
APRIL	145,465	303,735	614.87 GB
MAY	184,052	416,301	654.34 GB
JUNE	218,440	497,835	628.79 GB
JULY	95,899	237,200	549.54 GB
AUGUST	84,884	216,782	542.96 GB
SEPTEMBER	109,812	267,619	570.53 GB
OCTOBER	121,301	278,533	542.23 GB
NOVEMBER	146,272	348,661	727.24 GB
DECEMBER	112,310	259,043	637.95 GB

RCSB PDB Focus: External Links

External links are available for all PDB entries

Structure Summary pages for all structures in the PDB now offer a set of external links. Accessible from the left menu, these links provide further information about the structure under study, such as biochemical pathway information, stereochemistry and ligand binding data.

Outreach and Education

PDB Structures on Exhibit at the Birch Aquarium

The Kiosk at the Birch Aquarium

The Sea of Genes exhibit at the Birch Aquarium (La Jolla, CA) helps to unravel the genetic secrets of life in the ocean through interactive displays that highlight the exciting discoveries of Scripps researchers.

One feature of this exhibition is an interactive kiosk that invites the user to display information about specific proteins found in marine organisms – and in the PDB. This animation is also available (in Flash format) from the RCSB PDB's online Educational Resources page.

This kiosk is the result of a collaboration between the RCSB PDB and the Birch Aquarium (Scripps Institution of Oceanography at University of California, San Diego). The Sea of Genes exhibit will be on display until Spring 2007.

wwPDB Paper Published

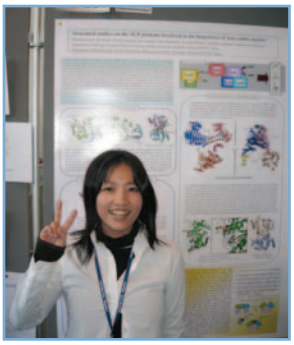
WORLDWIDE PDB PROTEIN DATA BANK

A paper describing the wwPDB – background, data deposition and access information, data uniformity efforts, and more – has been published:

The worldwide Protein Data Bank (wwPDB): ensuring a single, uniform archive of PDB data. Helen Berman, Kim Henrick, Haruki Nakamura, and John L. Markle (2007). *Nucleic Acids Research* **35(Database issue)**:D301-3; doi: 10.1093/nar/gkl971

RCSB PDB Poster Prize Awarded at AsCA

Thanks to everyone who participated in the recent RCSB PDB Poster Prize competition for best student poster related to macromolecular crystallography at the Joint Conference of the Asian Crystallographic Association and the Crystallographic Society of Japan (AsCA; November 20-23 in Tsukuba, Japan). The award went to:



Norika Sumi

Structural studies on the SUF proteins involved in the biogenesis of iron-sulfur clusters. Norika Sumi[1], Kei Wada[1], Shintaro Kitaoka[1], Kei Suzuki[1], Yuko Hasegawa[1], Yoshiko Minami[2], Yasuhiro Takahashi[1], and Keiichi Fukuyama[1]

[1]Department of Biology, Graduate School of Science, Osaka University, Toyonaka, Osaka 560-0043 Japan

[2]Department of Biochemistry, Faculty of Science, Okayama University of Science, Okayama 700-0005 Japan

Many thanks to our judges and organizers.

Judges: Anders Liljas (Chair; Lund University), K. Byrappa (University of Mysore), Mitchell Guss (University of Sydney), Chwan-Deng Hsiao (Academia Sinica), and Genji Kurisu (University of Tokyo)

Organizer: Soichi Wakatsuki (Institute of Materials Structure Science, KEK, Japan)

Meeting News

• Director Helen Berman described "wwPDB: An International Collaboratory for Structural Bioinformatics" at the 20th CODATA International Conference in Beijing, China (October 23-25).

• Berman also gave the talk "Resources for Structural Genomics from the RCSB PDB" at the International Structural Genomics Organization Conference in Beijing (October 22-26). Zukang Feng presented the poster "The RCSB PDB structural genomics portal" at this meeting.

• Co-director Phil Bourne traveled to Cuba to discuss "Protein Structural Data Reveals How Environmental Pressures Shape Evolution" at the 27th Latin-American Conference on Chemistry (October 16-20).

• Bourne then presented "Assigning DOIs to data objects" at the Crossref Annual Member Meeting in Boston, on November 1, 2006.

• The RCSB PDB recently exhibited at the annual conference of the Association of Science-Technology Centers (October 28-31 in Louisville, KY).

The RCSB PDB booth at the ASTC meeting

Molecules of the Quarter:

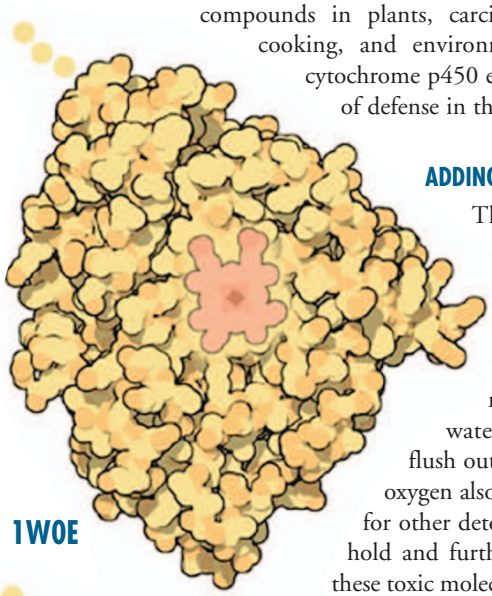
Cytochrome p450, Fibrin, and Transposase

The **MOLECULE OF THE MONTH** *series explores the functions and significance of selected biological macromolecules for a general audience. The molecules featured this quarter were cytochrome p450, fibrin, and transposase. The entire feature on cytochrome p450 is republished here.*

The complete Molecule of the Month features are accessible from the RCSB PDB home page.

Cytochrome p450

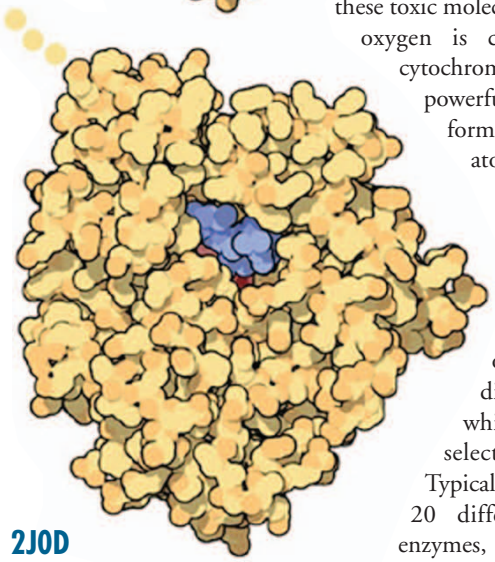
If you have a headache and take a drug to block the pain, you'll notice that the effects of the drug wear off in a few hours. This happens because you have a powerful detoxification system that finds unusual chemicals, like drugs, and flushes them out of your body. This system fights all sorts of unpleasant chemicals that we eat and breathe, including drugs, poisonous compounds in plants, carcinogens formed during cooking, and environmental pollutants. The cytochrome p450 enzymes are our first line of defense in this chemical battle.



1W0E

ADDING OXYGEN

The cytochrome p450 enzymes find unusual molecules and add oxygen atoms to them. In most cases, this has the effect of making the molecule more soluble in water, and thus, easier to flush out of the body. The added oxygen also provides a ready handle for other detoxifying enzymes to take hold and further modify, and destroy, these toxic molecules. This task of adding oxygen is chemically tricky, and cytochrome p450 enzymes use a powerful molecular tool to perform the reaction: an iron atom in a heme group.



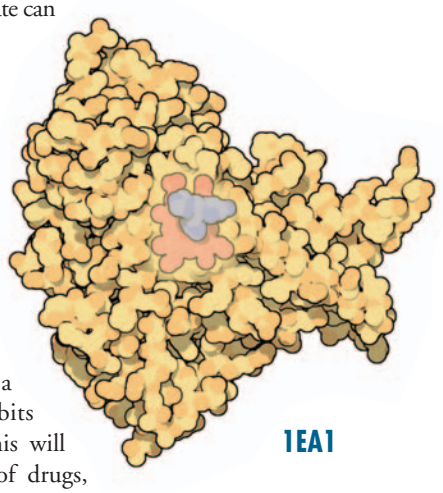
2J0D

P450 EVERYWHERE

Cytochrome p450 enzymes are found in all organisms. Each organism builds several different enzymes, each of which act on a different selection of molecules. Typically, bacteria make about 20 different forms of these enzymes, and we produce about 60. Plants often make hundreds of different forms. This is because plants make unusual pigments and exotic toxins to protect themselves. Many of the reactions needed to make these molecules are performed by specialized cytochrome p450 enzymes.

A DOUBLE-EDGED SWORD

The molecule shown in PDB entries 1w0e and 2j0d is CYP3A4, the cytochrome p450 that plays the major role in drug detoxification in your body. It has been estimated that this enzyme acts on about half of known drugs. For instance, it modifies the antibiotic erythromycin, shown in blue. It also detoxifies such diverse drugs as codeine, diazepam (Valium), paclitaxel (Taxol), and several anti-HIV drugs. In some cases, however, the reaction performed by cytochrome p450 enzymes can cause more harm than good. For example, CYP3A4 is partially responsible for the toxicity of large doses of acetaminophen (Tylenol). The modified form of acetaminophen is dangerously reactive, but it is normally cleared away quickly by other detoxifying enzymes. But with large doses, the reactive intermediate can build up to dangerous levels.



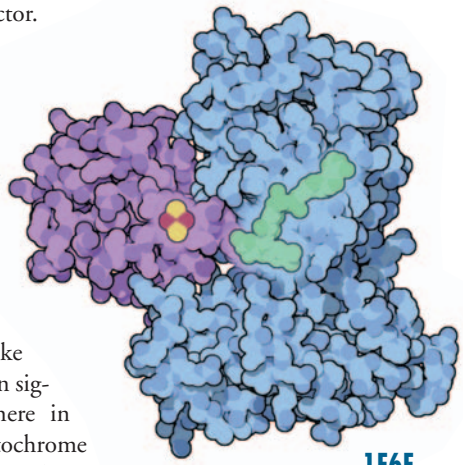
1EA1

PRESCRIPTIONS AND P450

Doctors must be careful to keep the cytochrome p450 enzymes in mind when they prescribe medications. For instance, you may have seen warnings on prescriptions, telling you not to drink grapefruit juice when taking a medication. Grapefruits contain a flavinol molecule that inhibits cytochrome p450 enzymes. This will slow down the detoxification of drugs, which may cause them to have stronger effects than expected by the doctor.

SYNTHETIC WIZARDS

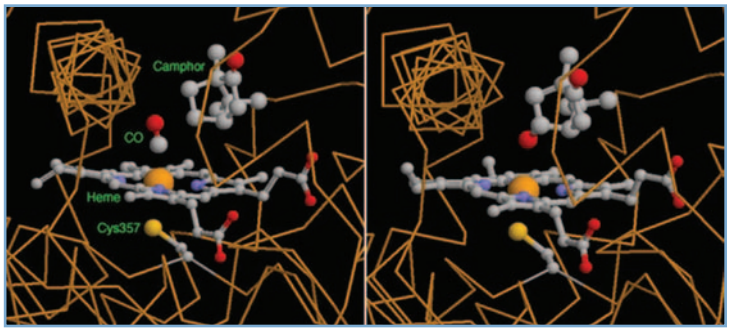
Cytochrome p450 enzymes also play a number of essential roles in the synthesis of normal cellular compounds. For instance, special cytochrome p450 enzymes are built to perform chemical steps in the construction of steroids, vitamins A and D, and lipid-like eicosanoid molecules involved in signaling. The enzyme shown here in PDB entry 1ea1 is a fungal cytochrome p450 that performs a step in sterol synthesis. A similar enzyme in our cells is need-



1E6E

ed for the synthesis of cholesterol. The enzyme complex shown with PDB entry 1e6e provides electrons for the reaction.

EXPLORING THE STRUCTURE



3CPP 1NOO

The best-studied cytochrome p450 enzyme is a bacterial enzyme that adds oxygen to camphor. Two early examples of these enzymes, called cytochrome p450cams, are shown here. On the left (PDB entry 3cpp) is a structure with camphor and carbon monoxide bound in the active site. The carbon monoxide is an inhibitor that poisons the enzyme. It binds to the iron (large yellowish sphere in the middle of the heme) in the same place as oxygen gas. The cysteine amino acid at the bottom activates the iron. The structure on the right (PDB entry 1noo) shows camphor after the reaction, when an oxygen atom has been added (the other oxygen atom is released during the reaction as a water molecule). Looking through the PDB, you can find dozens of other structures of cytochrome p450cam, showing many different molecules bound in the small active site, and showing many different stages in the reaction.

1W0E: P.A. Williams, J. Cosme, D.M. Vinkovic, A. Ward, H.C. Angove, P.J. Day, C. Vonrhein, C, I.J. Tickle, H. Jhoti (2004) Crystal structures of human cytochrome P450 3A4 bound to metyrapone and progesterone *Science* **305**:683-686

2J0D: M. Ekroos, T. Sjogren (2006) Structural Basis for Ligand Promiscuity in Cytochrome P450 3A4 *Proc.Natl.Acad.Sci.USA* **103**:13682-13687

1EA1: L.M. Podust, T.L. Poulos, M.R. Waterman (2001) Crystal structure of cytochrome P450 14alpha-sterol demethylase (CYP51) from *Mycobacterium tuberculosis* in complex with azole inhibitors *Proc.Natl.Acad.Sci.USA* **98**:3068-3073

1E6E: JJ. Muller, A. Lapko, G. Bourenkov, K. Ruckpaul, U. Heinemann (2001) Adrenodoxin reductase-adrenodoxin complex structure suggests electron transfer path in steroid biosynthesis. *J.Biol.Chem.* **276**:2786-2789

3CPP: R. Raag, T.L. Poulos (1989) Crystal structure of the carbon monoxide-substrate-cytochrome P-450CAM ternary complex. *Biochemistry* **28**:7586-7592

1NOO: H. Li, S. Narasimhulu, L.M. Havran, J.D. Winkler, T.L. Poulos (1995) Crystal structure of cytochrome P450-Cam complexed with its catalytic product, 5-exo-hydroxycamphor *J.Am.Chem.Soc.* **117**:6297

PDB Education Corner

Gary L. Gilliland, Methods in Structural Biology Course at Cold Spring Harbor Laboratory

The 2006 class of sixteen students has successfully completed the sixteen-day course X-Ray Methods in Structural Biology. Since 1988, students and instructors have descended upon Cold Spring Harbor Laboratory (CSHL) in New York to study the principles of X-ray crystallography. They then apply these concepts to actual structure determinations through extensive laboratory practicals. Each year, students are selected from a large international pool of applicants from academic, private, and commercial research laboratories. Their previous research experience ranges from that of graduate student to laboratory director. During the course, class begins at 9:00 a.m., with the instructors retiring late in the evening while many of the students work on into the night.

As the PDB user community knows, macromolecular crystallography yields a wealth of unique structural information that is used to further our understanding of biological systems. But learning the skills needed for structure determinations using diffraction methods is not easy. Crystallographic training is most often done in a research laboratory through what could be considered an apprenticeship approach, often with little formal training. Designed for scientists with a working knowledge of

Dr. James Pflugrath teaching students the procedures and techniques required for synchrotron X-ray data collection at the Brookhaven NSLS.

protein structure and function, the X-Ray Methods in Structural Biology course provides those new to the field an opportunity to learn from practicing instructors that are contributing significantly to the current methods of structure determination. The course curriculum was developed to emphasize a hands-on approach that includes crystallizing several proteins and determining one or more structures using state-of-the-art software. This laboratory and computational course focuses on the major techniques used to determine the three-dimensional structures of macromolecules. Topics covered include basic diffraction theory, crystallization (proteins, nucleic acids and complexes), crystal characterization, X-ray sources and optics, crystal freezing, data collection, synchrotron and home-source X-ray data reduction, multiple isomorphous replacement, multiwavelength anomalous diffraction, molecular replacement, solvent flattening, non-crystallographic symmetry averaging, electron density interpretation, molecular graphics, structure refinement, structure validation, coordinate deposition in the PDB and structure presentation.

Prof. Alex McPherson demonstrating the classic macromolecular crystal mounting technique in quartz capillaries used for collecting X-ray diffraction data at room temperature.

Prof. William Furey lecturing on the theory behind the methods used to solve the phase problem needed for the structure solution of biological macromolecules.

The course organizers – Drs. James Pflugrath, Alex McPherson, Bill Furey, and Gary Gilliland – have worked together for more than fifteen years to develop a unique learning experience that takes advantage of the expertise of many crystallographers that are pioneering methods in the field. This year's instructors included Drs. Paul Adams, Martin Caffrey, Serge Cohen, Paul Emsley, Morten Kjeldgaard, Gerard Kleywegt, Leemor Joshua-Tor, Duncan McCree, Randy Read, Dave Richardson, Jane Richardson, Robert Sweet, Tom Terwilliger, Dale Tronrud, David Waugh, and John Westbrook. The benefits to the participants from an instruction team of this quality are immense. Not only are they learning techniques that represent the latest innovations in structure determination, but they are also exposed to the thought processes of individuals who have created the software used in many structure determinations. The CSHL experience is unique in that the students learn and practice at a fundamental level all of the steps involved in a crystal structure determination.

The 2006 CSHL Macromolecular Crystallography was supported by donations and help from the National Cancer Institute, the Department of Energy, the Howard Hughes Medical Institute, the W. M. Keck Foundation, Hampton Research, Inc., Olympus, Rigaku Americas Corporation, Brookhaven National Laboratory, HKL Research, Inc. and Cold Spring Harbor Laboratory. The meetings staff that includes David Stewart, Terri Grodzicker, Wendy Crowly, Barbara Zane, and Andrea Stephenson along with a large supporting cast of other CSHL staff members also need to be thanked for their continuing contribution to making the course experience a rewarding one for the students and the instructors.

GARY GILLILAND, *a former Co-Director of the RCSB PDB, is currently the Director of Structural Biology at Centrocor Research & Development, Inc.*

The daily lectures provide the theory and methods used in parallel with the laboratory work in which the students determine a structure. The class crystallizes a variety of proteins, and then determines their diffraction properties on the CSHL home source while learning how to cryopreserve these fragile crystals. After screening the crystals, which they then store under liquid nitrogen, the students take a field trip to the National Synchrotron Light Source at Brookhaven National Laboratory to collect multiwavelength X-ray data. They return with their raw data and process it at CSHL (if not done while at Brookhaven). They then use the processed data to determine phases and calculate electron density maps that are autointerpreted (if the data is collected to an appropriate resolution) using several approaches. The students then apply interactive graphic techniques to complete the model prior to refinement of the atomic coordinates. The refined coordinates are then analyzed by the students to see if they conform to the principles of protein structure. Finally, the students learn what is required for deposition in the PDB archive.

This year's class has now joined the ranks of the more than 300 former graduates who have gone on to be significant PDB depositors of coordinates and structure factors.

Applications for the next course are due June 15, 2007; please see **meetings.cshl.edu/courses/c-crys07.shtml** for more information.

PDB Community Focus:

Julian Voss-Andreae, Protein Sculptor

JULIAN VOSS-ANDREAE *is a German-born sculptor based in Portland, Oregon. In his youth he painted for a number of years, but then changed course and studied physics at the universities of Berlin and Edinburgh. After participating in a seminal quantum physics experiment in Vienna as part of his graduate research, Voss-Andreae moved to the US in 2000 with his passion for art rekindled. He graduated from the Pacific Northwest College of Art (PNCA) in 2004 with a BFA in sculpture. While still at PNCA, Voss-Andreae developed a novel kind of sculpture based on the structure of proteins, the building blocks of life. Voss-Andreae's work has been commissioned internationally and has been highlighted in journals such as* Leonardo *and* Science.

For more information, please see ***www.julianvossandreae.com.***

Photographs of Voss-Andreae's sculptures, including those shown here, are now part of the RCSB PDB's *Art of Science* traveling exhibit. For more information on hosting this exhibit, please contact info@rcsb.org.

Q: ***Your sculptures are amazing depictions of three-dimensional proteins. How do you decide upon the scale and the materials to use? Are you trying to tell a story about the molecule? Is there a connection between the proteins shown and the wood or metal?***

A: I want to make sculptures that work as metaphors. When I use the structure of a protein as a starting point, it is not only about the beauty of that structure. That can be accomplished equally well with a protein model. I want to create something meaningful beyond that. To give you an example, I created a sculpture based on the structure of collagen (using the coordinate file for 1BKV)[1]. I designed cutouts on each piece (a peptide unit) to reveal the dominant force lines, which is something we see in steel bridges and other steel constructions every day. That way the piece subtly alludes to its role in our body as a structural component. At some point I decided to depart from the molecular structure and opened up the intertwining helices toward the top. This expanded the meaning of the sculpture towards another important aspect most people think of in connection with collagen, aging. Collagen is responsible for our skin's elasticity – its degradation famously leads to the wrinkles that accompany aging. Together with the title "Unraveling Collagen" the piece now has the potential to function as a metaphor for our growth physically as well as mentally, and for our intertwined paths through life.

In another piece, titled "Heart of Steel" (picture on page 11), I was interested in using a literal connection between the chemistry in the protein and the chemistry on the sculpture's surface. I made a complete human hemoglobin (1A3N)[2] out of a certain kind of steel known as 'weathering steel'. This special alloy initially rusts like ordinary steel but eventually stops because the special oxide layer it builds up is not water soluble and thus protects it from further corrosion. I finished the piece with a shiny surface and installed it. Upon its unveiling it was still gleaming, but after a few rain showers the color started to change and within half a year it was dark red. What had completely changed the look of the sculpture is of course the same chemical reaction that occurs when we breathe: Iron binds to oxygen.

Another sculpture is based on a light-harvesting complex (1NKZ)[3], the protein scaffolding containing a subunit of the photosynthesis apparatus in plants (picture on page 10). So I displayed the piece on the floor of a dark empty room with a candle in the center casting shadows of the structures on the wall. That way the sculpture in its environment conveys a feeling of sacredness; it somehow resembles an altar. The flickering shadows of the eighteen alpha helices arranged in two concentric circles are intriguing, because they look a bit like moving plants. It is as if the light-harvesting complex still originates flora, but now with exchanged roles: The macroscopic plants of our world become ephemeral shadows, whereas the microscopic, and ordinarily not perceivable basis for their existence, becomes a tangible object. That is the kind of story I am interested in telling. It is about triggering associations and emotions. I want to make objects imbued with meaning of a poetic kind. My objects should have the potential to evoke metaphors in a scientific context which sometimes seems irreconcilable with the non-rational nature of poetry.

Unraveling Collagen, 2005
Stainless steel, 24" x 135" x 32"

Q: ***How do you decide which proteins to sculpt? Is it the type of protein or the shape (helices, sheets, transitions of structural elements) that interests you?***

A: Sometimes I am fascinated mostly by the structure itself and sometimes more by the conceptual aspects, such as from what organism the protein originates and what biological role it plays there. A recent example for a sculpture based on an especially intriguing structure is a stainless piece portraying microcin J25 (1Q71)[4]. The small peptide resembles a lasso. Its tail folds over and goes through a noose made out of eight amino acids on the other end. An example of

a protein mostly intriguing to me for its conceptual properties is hemoglobin. Ideally, both aspects are equally appealing. The first protein I saw, the green fluorescent protein (GFP), is such a molecule for me. It has a fascinating story as well as one of my favorite structures. It was actually an image of GFP that got me hooked on proteins in the first place while I was still doing science. I recently completed a stainless version that will possibly go on display at the Friday Harbor Laboratories on San Juan Island (Washington State), where GFP was isolated first in the early 1960's.

I don't have a particular preference for certain types of proteins. Concerning secondary structure, I tend to like beta sheets better than alpha helices, because they interact more clearly with their neighbors whereas the helices look more isolated, which is illustrated by the position of the hydrogen bonds in those two motifs. The first large-scale outdoor sculpture I made was, however, a portrait of an alpha helix. This work, which was dedicated to the discoverer of the alpha helix, Linus Pauling, worked sculpturally very well though, because the helix was presented as an isolated object.

I also like cyclic proteins very much. The first one I made was based on the plant cyclotide kalata B1 (1K48)[5]. It is a protein found in an African herbal medicine, with the effect of accelerating labor in childbirth. I made a large table top version of that protein that is now at the University of Queensland in Brisbane, Australia. Dr. David Craik has surrounded the sculpture with real Kalata-Kalata plants in pots, which I find a really beautiful idea.

Light-Harvesting Complex, 2003
Wood, casting resin, 22" x 25" x 25"

Q: ***You wrote a computer program that takes a PDB coordinate file and indicates where the materials should be cut and rotated to build your sculptures. Was there a lot of trial and error in creating your early works?***

A: I wrote the core of that program in 2001 but I keep adding new parts. It allows me to make virtual renditions of the sculptures on the computer to troubleshoot before I build a real one; that worked very well from the start. The large challenge was never really the computer program but the translation into the real sculpture. In addition to inaccuracies there are also always the real, big mistakes we humans tend to make and it is surprising how many mistakes I made in my early work just because of the sheer number of measurements, cuts and connections.

The latest addition to my program further eliminates sources of error by directly generating graphic files that can be used for computer-controlled cutting of metal with a laser beam or a thin water jet. Each peptide unit gets transformed into a four-paneled hinged piece of metal that comes complete with its number laser-etched on it. This approach significantly enhances the precision which allows me to tackle much larger proteins. The error still accumulates due to the one-dimensional nature of proteins, but it is small enough to be corrected on the way. Even though the laser-cutting replaces a lot of the manual labor and eliminates many errors, the construction is still an extremely time consuming process which requires a high level of skill. I still need about one hour of shop time per amino acid for a piece with amino acids the size of a hand.

Q: ***How close to the coordinate file are the final pieces? How important is scientific accuracy to your works?***

A: For pieces consisting of few amino acids, maybe up to 50, my sculptures are typically quite accurate in the sense that the center of each mitered-cut joint corresponds to the position of the carbon alpha atoms in the peptide chain. When the number exceeds about 100, corrections become necessary even though the cutting has an accuracy of a few percent and the connections are similarly precise. Heat distorts the metal and gravity kicks in at some point, resulting in unavoidable deviations from the exact molecular structure. Recently I started using much thicker material which holds up against gravity much better. But still, there are always some residual inaccuracies accumulating. I am not interested in accuracy just to be precise, but in order to capture the character of the protein structure. I don't mind if one end is off by a little bit with respect to the other, but it is unacceptable if some parts of the backbone have an unnatural spacing to their neighbors. If I have a beta sheet for example, I want to make sure that the zigzag pattern on neighboring strands corresponds to each other to capture the character of the fold. It is an important aspect of those pieces that these peptide chains don't just wiggle randomly in space but are in fact shaking around a delicately balanced configuration of minimal energy.

In one case, when I made the sculpture based on kalata B1, I have asymmetrically stretched the coordinates a little bit. I needed some distortion in order to make my angles match up at the ends. But I did it in such a way that the orientation of my square tubing had the same Möbius strip topology as the backbone of the molecule. So it remained a faithful mapping of the topology, but not the coordinates.

When I made a number of alpha helix-based pieces out of wood, I experimented with shortening the length of the peptide units proportional to the natural tapering of the wood, which yielded some very interesting results. For example, I built an alpha helix out of a tree cut into over 100 identical pieces decreasing in size. The condensation of the 30' (9 m) tree into the 10' (3 m) tall sculpture amplified the natural shape of the tree and caused the piece to have a striking resemblance to a human spine. I find such an unexpected emergence of a new level of meaning very interesting. My goal is to create a sculpture that works artistically. In some cases that might require an accurate mapping of the actual geometry, whereas in others it is only the specific departure from the exact geometry that allows it to become a meaningful piece of art.

Q: ***Which do you find more interesting – the shape of the protein, or the manipulation of these materials to make that shape?***

A: I love the structure of proteins and the diverse science behind it. I am stunned by the workings of this machinery of life. The aspect of my work I probably enjoy the most is that I get to explore and learn, and then dream up a sculpture, transforming my knowledge into something sensual and palpable, something that offers me, as well as others, an understanding on

Heart of Steel (Hemoglobin), 2005
Weathering steel and glass, height 5'
Location: 1st Street/"A" Avenue, City of Lake Oswego, Oregon

a deeper level. The most fulfilling is when a completed piece works well on many levels; esthetically, conceptually, and scientifically. And I really love getting feedback about my work from people I would have never met otherwise.

Q: ***How much information about science and technology do you include in your lectures to art students?***

A: In my lectures I talk about everything that fascinates me, and that is typically about 90% science. But I always try to present it as something sensual, with lots of visuals. I never really appeal purely to the intellect, but mostly to my audience's capacity to just watch and wonder. I do that not so much for pedagogic reasons, but because that is the way I function myself. Another aspect is that I always feel that there is a need to fill a gap of knowledge when it comes to science for non-scientists, especially in this country. I want to share my passion about science and show that it can be fun and extremely fascinating.

Q: ***We've noticed that people are interested in protein structures even when they don't know what they are looking at. However, that level of interest seems to greatly increase when they become aware of the specific protein being depicted. A picture becomes more engaging when it is known that it is, for example, hemoglobin. How does this play into your work?***

A: I have noticed that effect, too, especially with people who are generally interested in science. In fact, I used to display my work with labels talking a little bit about the science behind it. But I changed that because often the exact opposite happened, especially in the art world: As soon as you mention "science" people tend to shut off. They often feel uneasy with science and fail to recognize that it is a huge (and hugely important) part of human culture. And therefore, they rob themselves of the opportunity to see a whole world of beauty, the beauty revealed to us through scientific experiments.

Q: ***DNA is an iconic image at this point – it is recognized by almost everyone. Will structures of proteins follow the same path?***

A: I have my doubts. DNA is such a great image because of its simplicity. Take a ladder and twist it. But I do think that the general public is bound to get a clearer visual idea of protein structure in the future. Currently we are witnessing the merging of human technology and naturally evolved technology – biochemistry – and I find it likely that we will end up expanding on this existing technology in the long run instead of completely reinventing the wheel. So people will see many more proteins in the future I think. Right now people already see frequent images of proteins, mostly in cartoons illustrating for example lock-and-key interactions or in popular scientific illustrations of viruses. I think people often fail to understand at this point that all these images actually depict proteins.

Maybe there are certain proteins that have the capacity for an iconic image; I could imagine, for example, the antibody.

Q: ***What are you working on now?***

A: I am currently working on three large commissions: Early next year I will be installing a 7' (2.10 m) stainless steel and glass version of a hemoglobin-based piece for a well-known scientist in Zurich, Switzerland. I have also started to work on a piece for Roderick MacKinnon in New York, who won the 2003 Chemistry Nobel Prize for elucidating the structure and function of the KcsA potassium channel protein (1K4C)[6]. He has commissioned me to create a sculpture inspired by that structure for his home. It will contain a blown glass object representing the internal space of the protein with its seven ion sites. And the new Scripps Research Institute in Florida has recently commissioned me to create a 12' tall stainless steel sculpture based on the antibody molecule which will be installed at the entrance of the new campus in Jupiter in 2008. That piece plays off of a very interesting formal correspondence I discovered between the proportions of the human body in Leonardo's famous "Vitruvian Man" from 1490 and the proportions of the antibody molecule as in the composite model by Eduardo Padlan.[7] I have just finished fine-tuning the design and am in the process of having 1,500 pounds of stainless steel laser-cut.

In addition to working with protein structure, I am very interested in finding sculptural metaphors for some aspects of quantum physics, which was my background before I turned back to art full-time. I recently finished the first version of a sculpture called "Quantum Man", a stylized image of a walking man consisting of about one hundred steel sheets oriented perpendicular to the direction of his motion. This piece was inspired by the question of how the wave function of a moving human would look like and it was covered by *Science* last August. Currently, I am working on a new version of that piece in bronze for a museum in Oregon.

Photograph of Julian Voss-Andreae by LeeAnn Gauthier. Photographs of sculptures by Voss-Andreae.

1. R. Kramer, J. Bella, P. Mayville, B. Brodsky, and H.M. Berman (1999) Sequence dependent conformational variations of collagen triple-helical structure. *Nat. Struct. Biol.* **6**:454-457.

2. J.R. Tame and B. Vallone (2000) The structures of deoxy human haemoglobin and the mutant Hb Tyralpha42His at 120 K. *Acta Crystallogr* D Biol Crystallogr. **56(Pt 7)**:805-11.

3. M.Z. Papiz, S.M. Prince, T. Howard, R.J. Cogdell, and N.W. Isaacs (2003) The structure and thermal motion of the B800-850 LH2 complex from Rps.acidophila at 2.0Å resolution and 100K: new structural features and functionally relevant motions. *J Mol Biol.* **326(5)**:1523-38.

4. K.J. Rosengren, R.J. Clark, N.L. Daly, U. Goransson, A. Jones, and D.J. Craik (2003) Microcin J25 has a threaded sidechain-to-backbone ring structure and not a head-to-tail cyclized backbone. *J Am Chem Soc.* **125(41)**:12464-74.

5. L. Skjeldal, L. Gran, K. Sletten, and B.F. Volkman (2002) Refined structure and metal binding site of the kalata B1 peptide. *Arch Biochem Biophys.* **399(2)**:142-8.

6. Y. Zhou, J.H. Morais-Cabral, A. Kaufman, and R. MacKinnon (2001) Chemistry of ion coordination and hydration revealed by a K+ channel-Fab complex at 2.0 Å resolution. *Nature.* **414(6859)**: 43-48.

7. E.A. Padlan (1994) Anatomy of the antibody molecule. *Mol Immunol.* **31(3)**:169-217. See also **www.umass.edu/microbio/rasmol/padlan.htm**.

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 (www.wwpdb.org)

RCSB PDB Management

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

DR. PHILIP E. BOURNE, Co-Director
San Diego Supercomputer Center and the Skaggs School of Pharmacy and Pharmaceutical Sciences,
University of California, San Diego
bourne@sdsc.edu

A list of current RCSB PDB Team Members is available from www.pdb.org.

STATEMENT OF SUPPORT: *The RCSB PDB is supported by funds from the National Science Foundation, the National Institute of General Medical Sciences, the Office of Science, Department of Energy, the National Library of Medicine, the National Cancer Institute, the National Center for Research Resources, the National Institute of Biomedical Imaging and Bioengineering, the National Institute of Neurological Disorders and Stroke, and the National Institute of Diabetes & Digestive & Kidney Diseases.*

RCSB PROTEIN DATA BANK
www.pdb.org

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

Return Service Requested