Protein Data Bank

Quarterly Newsletter

Release #78 *October 1996*

October 1996 CD-ROM Release

4873 full-release atomic coordinate entries

Molecule Type

4399	proteins, peptides, and viruses
113	protein/nucleic acid complexes
349	nucleic acids
12	carbohydrates

Experimental Technique

138	theoretical modeling
685	NMR
4050	diffraction and other

The total size of the atomic coordinate entry database is 1930 Mbytes uncompressed.

Table of Contents

Internet Sites

WWW .... http://www.pdb.bnl.gov
FTP .... ftp.pdb.bnl.gov

What's New at the PDB

— Joel L. Sussman

A microsymposium entitled "25 Years of PDB" was held at the IUCr XVII Congress and General Assembly in Seattle, Washington on August 11, 1996 to celebrate the twenty-fifth anniversary of the PDB. The symposium was chaired by Joel L. Sussman (Brookhaven National Laboratory, USA and Weizmann Institute of Science, Israel) and co-chaired by Günter Bergenhoff (University of Bonn, Germany).

To open the symposium the chair gave a brief overview of the current status of the PDB and the plans, currently well under way, to convert the PDB into a much richer "Three-Dimensional Database of Biological Macromolecules (3DB)" which is based on an object-oriented approach. Also presented was a status report on the PDB-developed AutoDep, a user friendly submission procedure which uses a WWW-based interface.

Edgar Meyer (Texas A&M University, USA) described the activity at Brookhaven in the years prior to and leading up to the establishment of the PDB at Brookhaven. He emphasized that even in those early days, the concepts of networking and molecular graphics were key components in the creation and development of the PDB. Helen Berman (Rutgers University, USA) discussed the early history of the PDB, focusing on the scientists who played key roles in its establishment, including slides which showed how some of us looked back in the early Seventies. The lecture continued with a presentation on the establishment of the Nucleic Acid Database (NDB) and its relationship to the PDB.

Frank Allen (Cambridge Crystallographic Data Centre, UK) discussed the interrelationship of the PDB and the Cambridge Structural Database, showing the parallel increase in the size of the two databases and predicting that by the year 2010 there may be well over 600,000 protein structures in the PDB. David R. Davies (National Institutes of Health, USA) gave a fascinating lecture describing the impact that the PDB has had in the area of research which relates the structure and function of molecules involved in the immune system. He focused on how PDB structural information has played a key role in the determination of new structures *via* molecular replacement and also how the wealth of PDB information has made it possible to do comparative structural analyses of these molecules. G. Marius Clore (National Institutes of Health, USA) presented a lecture entitled "Improving the Quality of NMR and Crystallographic Protein Structures by Means of a Conformational Database Potential Derived from Structure Databases." He showed the impact that high-resolution X-ray structures are having on developing new conformational potentials for use in NMR structure determination and refinement. It was very exciting to see that Clore's approach is likely to have an enormous impact on X-ray structure determination, particularly when only low-resolution X-ray data are available. In such cases the orientations of the side chains are often uncertain and the new potentials will likely lead to a much better approach than is currently being used. It may also lead to improved techniques for constructing theoretical models and better molecular force fields.

Wolfram Saenger (Free University, Berlin, Germany) began the final lecture of the microsymposium with a gift of flowers for the occasion — "virtual flowers" in the form of a beautiful slide — and proceeded to discuss what happened to fiber diffraction studies in the PDB. Saenger's talk, wonderfully suited to the occasion and full of humor, focused on two subjects: how accurate fiber diffraction studies have been, and still are, in relation to comparable single-crystal studies and what an enormous impact these studies have had on structural biology.

On November 17-21, 1996 there will be a second special celebration entitled "Bioinformatics <---> Structure" in honor of the 25th Anniversary of the PDB and the 10th Anniversary of the Swiss-Prot Database. This will be part of the 24th Katzir Conference to be held in Jerusalem, Israel. This meeting will focus on the relationship of sequence, structure and function, and the different databases that contain this information. Sessions will include talks, poster presentations, and computer demonstrations with Internet access. Details of the meeting, including the full list of invited speakers and the 250 submitted abstracts, are available from URL *http://pdb.weizmann.ac.il/pdb25sw10/*, at the mirror sites listed at that URL, or from PDB's WWW home page.

AutoDep Is Now Available

— Nancy Manning

The Protein Data Bank has released AutoDep Version 1.0, a WWW-based graphical interface to the Deposition Form which leads the author through the entire submission process. AutoDep performs some data verification and validation checks, thus allowing many potential problems to be corrected before the submission arrives at the PDB.

AutoDep may be obtained from the PDB's Web site at URL *http://www.pdb.bnl.gov* under "Submitting Data to the PDB".

The many beta testers of AutoDep agree that it greatly simplifies the process of preparing submissions for the PDB. The interactive form is divided into several sections, each of which leads the depositor through the various needed data items, providing examples and help along the way.

AutoDep's most important features include the following:

- AutoDep may be initially populated from an existing PDB entry or a previous deposition. As easy as pushing a button, AutoDep will enter all the fields from the designated file to the appropriate fields in the form. The author merely has to update fields as necessary to reflect the new structure being submitted.
- Program output files may be merged into AutoDep. For example, the new releases of X-PLOR and SHELXL write refinement details in a format which can be read by the program and used to populate the relevant sections. We are continuing to work with authors of various programs and anticipate increasing numbers of programs will integrate with PDB in this manner.
- Help files, examples, and links to related documentation and useful URLs are provided in every section to support the author during the AutoDep session.
- You may interrupt your AutoDep session at any time and return to it later in the day (or even days or weeks later). Remember to record your session id number and password in order to continue with the same deposition.
- When you are satisfied with the completed Deposition Form, a SUBMIT button is provided that initiates the following:
  - The coordinates are passed through a syntax checker.
  - If they fail, you are asked to correct the problem and resubmit the coordinates.
  - If they pass, you are immediately sent an acknowledgment letter containing your PDB ID code.
  - Your entry enters the PDB processing flow.

The Electronic Deposition Form has been updated to reflect AutoDep. Known as Deposition Form Version 3.30, it is available on our Web server (*http://www.pdb.bnl.gov/* under "Submitting Data to the PDB") and FTP server (*pub/dep_form.txt*). If you are unable to use AutoDep for any reason, you must use this new version of the Dep Form when preparing submissions for the PDB.

Improvements to AutoDep will be made on a continuing basis. If you have any questions about using AutoDep or about the new Deposition Form, please contact our Help Desk by e-mail at *pdbhelp@bnl.gov* or by telephone at 516-344-6356. We gladly welcome your comments and suggestions.

Archive Management

— Enrique Abola and Nancy Manning

— News and Updates

A draft version of The Protein Data Bank Contents Guide: Atomic Coordinate Entry Format Description V. 2.1 is available from *http://www.pdb.bnl.gov/Format.doc/Format.Home.html.* This version contains corrected remark templates, a few additional tokens in the COMPND and SOURCE sections, enhanced examples, and improvements to the language of the text to help clarify the format. No changes to PDB format that would entail going through the formal Format Change procedures have been included.

Changes to Remark Records

As stated in the PDB Contents Guide, changes to REMARK records may be implemented without following the formal change procedure. PDB reserves the right to modify the format of REMARKs, as they have been defined as free text in the Contents Guide. Beginning with version 3.843, X-PLOR(online) introduces changes to the output refinement remark for X-ray studies. Rather than waiting the usual five-month period before implementing these changes, crystallographic entries released after October 20, 1996 will contain the new X-PLOR REMARK 3 template.

Remark records such as REMARK 3 or REMARK 200 contain value-attribute pairs; it is currently the practice of the PDB to use the word NULL to represent values not supplied by the depositor. Following the suggestion of Professor Axel Brünger, we will continue to use NULL in these cases, but will use the word NONE when the attribute is not applicable or when analysis options were chosen so that a value was not calculated. For example, if the refinement was performed without using cross-validation techniques, the free R value token will be assigned the value NONE.

It is envisioned that in the future certain tightly-defined remarks may be converted to unique record types. From that time on, any changes to the format of these records will follow the standard PDB Format Change Policy.

Record Order

MODRES records appear immediately following SEQRES and just before HET in entries. This keeps the het group-related information together. The order was incorrectly stated in the Contents Guide V. 2.0 but has been corrected in the new document.

— Proposed Changes to PDB Format V. 2.0

A number of proposed changes to the existing data format are being presented here for consideration. In accordance with PDB's Format Change Policy, which may be found at URL *http://www.pdb.bnl.gov/format_change_policy.html*, there will be an open sixty-day discussion period during which we will entertain comments and suggestions regarding these changes. Please send comments to Enrique Abola (*abola1@bnl.gov*) or Nancy Manning (*oeder@bnl.gov*). Discussion on the PDB List-server is encouraged as well.

Changes being proposed here, if adopted, will not appear in released entries before March 31, 1997. A public announcement will be made several weeks prior to their appearance in released entries.

Hydrogen Atom Names in Amino Acids

Methylene hydrogen atoms will be labeled as 2HX and 3HX where X is the remoteness indicator of the atom. For example, hydrogen atoms attached to C β of an amino acid will be named 2HB and 3HB. Our current convention is to name these 1HB and 2HB. This change will make PDB more compliant with IUPAC recommendations.

Space Group Symbol for Monoclinic Crystals

The use of the shortened Hermann-Mauguin symbol for monoclinic crystals will be reinstated. This will be applied to crystals in the standard b-unique cell setting. Thus the space group symbol P 21 will be used instead of P 1 21 1. Crystals using other settings will be designated with the full international Hermann-Mauguin symbol (*e.g.*, P 21 1 1).

Representation of Modified Nucleic Acid Residues

Modified nucleic acids will be represented using the same rules that are used by the PDB for representing modified amino acids. We will assign a unique three-letter code for modified residues. For example, we may use BRU for brominated uridine rather than +U. In addition, all atoms belonging to the residue will be grouped together in the coordinate records. Our current practice is to list atoms that modify nucleotides after the TER record.

BIOMOL Update

John P. Rose, University of Georgia, Athens, GA, USA and consultant to the Protein Data Bank (rose@BCL4.biochem.uga.edu).

The PDB maintains a directory of biological macromolecules (BIOMOLs) containing coordinates of the full biological unit (functional molecule) for entries in which the asymmetric unit described in the entry represents only part of the functional molecule. For example, some PDB entries such as hemoglobin contain only one alpha-beta dimer in the asymmetric unit with the rest of the functional tetramer being generated by a crystallographic 2-fold symmetry operation. The corresponding hemoglobin BIOMOL entry would contain coordinates for the complete alpha2, beta2 tetramer.

BIOMOL entries may be obtained from the PDB WWW server (*http://www.pdb.bnl.gov/biomol.html*) or from the PDB anonymous FTP server in the directory */user_group/biological_units*.

As part of the revised PDB format, the PDB has added several new remark types which are used to describe the biological macromolecule. They are: REMARK 300, which describes the biomolecule, and REMARK 350, which describes the transformations required to generate the atomic coordinates for the entire biomolecule from those supplied in the entry. For the hemoglobin example given above, the PDB entry would include the following remarks:

```
REMARK 300
REMARK 300 BIOMOLECULE
REMARK 300 HEMOGLOBIN EXISTS AS AN A1B1/A2B2 TETRAMER. THE BIOMT
REMARK 300 TRANSFORMATION IS THE CRYSTALLOGRAPHIC 2-FOLD.

REMARK 350
REMARK 350 GENERATING THE BIOMOLECULE
REMARK 350 COORDINATES FOR A COMPLETE MULTIMER REPRESENTING THE KNOWN
REMARK 350 BIOLOGICALLY SIGNIFICANT OLIGOMERIZATION STATE OF THE
REMARK 350 MOLECULE CAN BE GENERATED BY APPLYING BIOMT TRANSFORMATIONS
REMARK 350 GIVEN BELOW. BOTH NON-CRYSTALLOGRAPHIC AND
REMARK 350 CRYSTALLOGRAPHIC OPERATIONS ARE GIVEN.
REMARK 350
REMARK 350 APPLY THE FOLLOWING TO CHAINS: A, B
REMARK 350   BIOMT1   1  -1.000000   0.000000   0.000000      0.00000
REMARK 350   BIOMT2   1   0.000000   1.000000   0.000000      0.00000
REMARK 350   BIOMT3   1   0.000000   0.000000  -1.000000      0.00000
```

The PDB has developed a tool (ViewBio) for generating the BIOMOL from the corresponding PDB entry. ViewBio, written in Perl, reads remarks 300 and 350 and expands the coordinate set according to the transformations described in REMARK 350 to generate the biomolecule. This can then be displayed using Rasmol.

The ViewBio script will be available shortly from the PDB. Both Perl and Rasmol may also be obtained from the server. The PDB will continue to expand the list of BIOMOL entries as the older PDB entries are converted to the new PDB format. Anyone desiring a BIOMOL not yet available may contact John Rose at *rose@BCL4.biochem.uga.edu* for assistance.

Staff Changes

— Minette Cummings

The PDB would like to thank David Stampf for his three years of service to the PDB. While with the PDB, Dave served as Sr. Project Manager and Head of the Software Development Group. He greatly enhanced network access to the PDB, especially *via* the PDB WWW Browser. He also designed and helped develop AutoDep, PDB's new Web-based submission tool, which is being introduced this month. Dave has returned to BNL's Computing and Communications Division as Sr. Computer Analyst with the Advanced Technology and Planning Group, and we wish him well in his new position.

We also bid a fond farewell to Judith Callaway, who processed incoming structures with the PDB for seven years. This May Judy earned her Doctorate of Education from Columbia University, and she has decided to devote herself full-time to teaching Biology.

The PDB welcomes several new staff members: first, Jiansheng Jiang from Yale University will be heading up the Software Development Group. His first task will be to implement AutoDep and integrate it into PDB's validation procedures. Dawei Lin, who recently received his Ph.D. in Chemistry from the Institute of Physical Chemistry at Peking University, has begun work with the Database Development Group on new 3D query tools using object-oriented techniques.

Brigitte Sylvain, who earned her BA in Biology from SUNY, Binghamton in May, 1996, has joined the PDB as a Professional Intern. Her year with the PDB will be spent in the Archive Management Group, validating structural entries and learning about the field of bioinformatics. Avraham Bluestone, a senior mathematics major at Brooklyn College, has joined the PDB through the Department of Energy's 1996 Fall Science and Engineering Research Semester Program. Avraham is working with the Database Development Group, developing database methodology using object-oriented techniques.

Christine Metz, currently in her junior year as a chemistry major at SUNY, Stony Brook, began working with the PDB in January, 1996 as a Student Intern through the Suffolk County Community College Women's Studies program. After successfully completing her commitment with this program, Christine was hired by the PDB on a continuing basis to work with both the User Support Group handling the PDB Help Desk and the Archive Management Group preparing newly deposited entries for release.

The summer was a very busy time for the PDB. We had several visitors working with us whom we would like to recognize. Jaime Prilusky, Head of the Bioinformatics Unit at the Weizmann Institute of Science, Rehovot, Israel, and Interim Head of the Database Development Group of the PDB, joined us for eight weeks this summer. He designed the new 3DB Browser and the new PDB Status Browser, both of which were released in August, 1996. Dr. Prilusky continues to work on development of the new relational database based on the Object Protocol Model of V. Markowitz [1].

Wen-Juan Xu, a student of computer science at Suffolk County Community College, was a summer intern with PDB. Wen-Juan built a new orders database for us, using SYBASE and Tk/Tcl.

Clifford Felder of the Weizmann Institute, Rehovot, Israel, was with us for five weeks assisting the Archive Management Group, particularly in the areas of HET groups, documentation, and AutoDep testing. Alexander Faibusovich, also of the Weizmann Institute, was with us for eight weeks helping with system administration issues and working on the CIF structure factor files. Manfred Hendlich of Merck KGaA joined the PDB for two weeks while developing a useful tool for searching the HET group dictionary.

Keith Peters and Jai Zhang both worked during the summer with our Systems Administration Group writing programs to help with entry processing, writing general utility programs, and assisting with hardware support. Keith has begun his freshman year at the California Institute of Technology as a computer science major, and Jai returned to high school as a sophomore. Zach Serber, a recent graduate of Columbia University, helped with some entry processing this summer while a research student working with Joel Sussman in BNL's Biology Department. He has now begun graduate school at the University of Glasgow, Scotland.

In addition Mia Raves and Kurt Giles of the Weizmann Institute, Israel, John Rose of the University of Georgia, USA, S. Swaminathan of the University of Pittsburgh, USA, and James Flanagan and John Spiletic of BNL's Computing and Communications Division continue to help the PDB in our Archive Management, Database Development, and Web Site Management areas.

We thank each of these people for their valuable contribution to PDB.

———

1. I.A. Chen and V.M. Markowitz, An Overview of the Object-Protocol Model (OPM) and OPM Data Management Tools, *Information Systems* **20 (5)**, 393-418 (1995). (Article and related information available at the URL *http://gizmo.lbl.gov/DM_TOOLS/OPM/opm.html*).

PDB CD-ROM Included in Time Capsule

— Pam Esposito

On Friday, October 11, 1996 a new $62.7 million Basic Sciences/Biomedical Engineering building was dedicated at the University of Minnesota. This building characterizes the University's commitment to bring together researchers from many diverse fields to work together as well as to share both resources and knowledge. Research will center around the fields of neuroscience, structural biology, immunology, cellular and molecular biology, and biomedical engineering.

During the dedication of this building a time capsule containing items that are currently at the cutting edge of biomedical research was presented. This time capsule will be sealed within the building and instructions indicate its opening should take place in one hundred years.

Items contained in the capsule include:

- A map of the human genome.
- A kit for drawing blood from a newborn's umbilical cord and storing it for transplantation.
- A "clinical laboratory" on a computer chip — a system of microscopic devices that may someday be used to monitor blood chemistry and deliver drugs from inside the body.
- A breast cancer family tree and a breast cancer (BRCA-1) gene sequence.
- A bioartificial artery made from collagen and lined with human smooth muscle cells.
- A computer model of the Human Immunodeficiency Virus (HIV).
- The PDB four-disk CD-ROM set containing an archive of nearly 5,000 experimentally determined three-dimensional structures of biological macromolecules, which serves a global community of researchers, educators, and students.
- The DNA double helix and J.S. Bach on a computer diskette — a DNA sequence converted into music, followed by digital Bach.

Please see URL *http://www.ahc.umn.edu* for further information on this project.

BioMagResBank (BMRB)

John L. Markley and Eldon L. Ulrich, University of Wisconsin-Madison, Madison, WI, USA (elu@nmrfam.wisc.edu; http://www.bmrb.wisc.edu).

We are pleased to announce that funding for BioMagResBank has been obtained through a three-year grant from the National Library of Medicine. We would like to thank the NMR community, Joel Sussman, Enrique Abola, and the PDB staff for their strong support of BMRB. With sustained funding, we look forward to continuing and strengthening the collaboration between BMRB and PDB. Through this collaboration, our goals are to develop standard formats for archiving and exchanging NMR spectroscopic data on proteins, peptides, and nucleic acids as well as to provide the scientific community with flexible and powerful tools for mining the archived NMR data.

Announcing X-PLOR(online)

Axel T. Brünger, Howard Hughes Medical Institute/ Yale University, New Haven, CT, USA (brunger@laplace.csb.yale.edu; http://xplor.csb.yale.edu/~brunger).
This pre-release version 3.851 is dated October 20, 1996; copyright 1996 by Yale University.

Internet access to X-PLOR(online) version 3.851 is now available, free of charge, for non-profit (academic) users holding a license for X-PLOR version 3.1 from Yale University.

This version is only available *via* the Internet. This is a pre-release and X-PLOR(online) is not yet complete. Minimal documentation consists of tutorial example files. Please monitor the bionet.software.x-plor newsgroup for future updates.

Access instructions, release notes, and example files are available from URL *http://xplor.csb.yale.edu*.

— Features of X-PLOR(online)

X-ray crystallography

- Update of all X-ray crystallographic tutorial files using new crystallographic language for structure factor and map symbolic manipulation.
- Torsion angle molecular dynamics for crystallographic refinement.
- Probabilistic MAD phasing.
- σ_A-weighting for electron density maps with optional cross-validation.
- Difference, anomalous difference, and <n>Fo-<m>Fc electron density maps.
- Cross-validated coordinate error estimates by Luzzati and σ_A method.
- Script files for molecular replacement with multiple molecules.
- Automated water picking procedure.
- New bulk solvent refinement procedure.
- Example for resolution-dependent weighting scheme for refinement.
- Direct rotation function.
- Phased translation function.
- Scalepack/denzo -> X-PLOR(online) conversion program.
- X-PLOR(online) -> PDB deposition script (for crystal structures).

Solution NMR spectroscopy

- J-coupling restraints.
- Proton chemical shift restraints.
- Carbon chemical shift restraints.

Additional Remarks

- For several months, new parameter and topology files for DNA and RNA (dna-rna.top and dna-rna.param) have been available [G. Parkinson, J. Vojtechovsky, L. Clowney, A.T. Brünger, H.M. Berman, New Parameters for the Refinement of Nucleic Acid Containing Structures, *Acta Cryst.* **D52,** 57-64, (1996)].

  These parameters are much improved compared to the older DNA parameters of X-PLOR 3.1. The new parameters can be used in conjunction with the Engh and Huber parameters for proteins (parhcsdx.pro and parhcsdx.pro) [R.A. Engh and R. Huber, Accurate Bond and Angle Parameters for X-ray Protein-Structure Refinement, *Acta Cryst.* **A47**, 392-400 (1991).

  DNA and RNA molecules may require some additional restraints. The arestraints.inp and brestraints.inp files in the *tutorial/generate/* directory provide examples of how to use these additional restraints.

- Electrostatic energy terms have been removed from all crystallographic refinement protocols in order to reduce possible bias. Removal of hydrogens is optional. New parameter files are available with uniform van der Waals radii for a repulsive nonbonded function (*protein_rep.param* and *dna-rna_rep.param*).

The PDB in Undergraduate Education

Virginia B. Pett, The College of Wooster, Wooster, OH, USA (pett@acs.wooster.edu).

The College of Wooster is a liberal arts institution (located in Ohio, not Massachusetts!) with about seventeen hundred undergraduates. This year there are twenty-one seniors majoring in chemistry and eight seniors majoring in biochemistry. Professors and students access the PDB for both teaching and research.

When teaching General Chemistry I use the classroom PowerMac and liquid crystal display panel on the overhead projector to show students macromolecular structures that do not appear in their textbook. For instance, when studying metal ion complexes, I show them the structures of B12 coenzyme and hemoglobin. The capability of the new 3DB Browse [1], located at *http://www.pdb.bnl.gov/cgi-bin/pdbmain*, to limit the PDB search to "biological structure" should expedite finding excellent teaching examples.

The PDB is a valuable source of information for undergraduates in upper level courses. Students in Physical Chemistry have done a molecular dynamics simulation on our Silicon Graphics Indigo with the HIV protease structure from the PDB. We use MidasPlus [2] for display and MacroModel [3] for calculations. My colleague, Montie Borders (*borders@acs.wooster.edu*), assigns a different protein to each pair of biochemistry students. They download the coordinates from the PDB FTP server and display the protein with Kinemage [4], taking particular note of secondary structure, hydrogen bonding patterns, and side chain interactions with heme groups or metal ions. In most cases the students produce a very fine interactive graphics Kinemage.

Recently Roy Haynes (another Wooster colleague) and I developed and tested a teaching module for organic chemistry called "How can we design a therapeutic drug?" The theme of this module is the development of anti-hypertensive drugs which inhibit angiotensin converting enzyme (ACE). Although the structure of ACE has not been determined, researchers used the structures of carboxypeptidase and thermolysin, two other zinc-containing peptidases, to model the active site of ACE. Students do molecular modeling simulations on the small inhibitor molecules and dock an inhibitor into the active site of thermolysin. Several thermolysin-inhibitor complexes are in the PDB.

For further information on teaching modules, please see URL *http://chemlinks.beloit.edu/homepage.html*, the WWW home page for ChemLinks, an NSF-funded consortium of institutions developing new approaches to teaching the first two years of college chemistry. For more innovative ideas on teaching chemistry and also for enhanced versions of Rasmol v2.5 and v2.6, see the ModularCHEM Consortium home page at URL *http://www.cchem.berkeley.edu:8080/MCsquared/*. Marco Molinaro (*molinaro@cchem.berkeley.edu*) and his team have added significant enhancements such as convenient display of bond distances, angles, and torsion angles. For more information see *URL http://hydrogen.cchem.berkeley.edu:8080/Rasmol/*.

All seniors at The College of Wooster do a year-long independent research project in their major field. A number of Wooster seniors have found examples of "structural arginines" in proteins downloaded from the PDB *[Protein Science* **3**, 541-548 (1994)]. Bo-Lu Zhou, a senior at Wooster, found eight new examples by writing a computer program to search the PDB for arginine side chains hydrogen-bonded to main chain N-H groups (International Conference on Protein Folding and Design, April 23-26, 1996, Bethesda, MD, poster 159). The newly added information in the PDB record for biological structure (see related article within this Quarterly Newsletter entitled BIOMOL Update) will enable us to find interactions between symmetry-related subunits that our programmed search would not have previously found. The "R factor" and "resolution" field searches in the extended 3DB Browser to be released in the near future will make it easier to limit automated searches to appropriate structures. As the emphasis upon biochemistry in our chemistry curriculum increases, the PDB will continue to be an important resource for both teaching and research in undergraduate education at The College of Wooster.

1. D.R. Stampf, C.E. Felder, and J.L. Sussman, PDBBrowse - a Graphics Interface to the Brookhaven Protein Data Bank, *Nature* **374**, 572-574 (1995).

2. T.E. Ferrin, C.C. Huang, L.E. Jarvis, and R. Langridge, The MIDAS Display System, *J. Mol. Graphics*, **6(1)** 13-27, 36-37 (1988).

3. F. Mohamadi, N.G.J. Richards, W.C. Guida, R. Liskamp, M. Lipton, C. Caufield, G. Chang, T. Hendrickson, W.C. Still, MacroModel - An Integrated Software System for Modeling Organic and Bioorganic Molecules using Molecular Mechanics, *J. Comput. Chem.* **11**, 440 (1990).

4. D.C. Richardson and J.S. Richardson, The Kinemage: A Tool for Scientific Communication, *Protein Science* **1**, 3-9 (1992).

Exciting New Resources for Education using RasMol and Chime

Eric Martz, University of Massachusetts, Amherst, MA, USA (http://www.bio.umass.edu/immunology/immunology.html; martz@ microbio.umass.edu).

The RasMol home page (*http://www.umass.edu/microbio/rasmol*), created in November, 1995, was described in the January, 1996 issue of the PDB Quarterly Newsletter. It has now been visited by over thirty thousand people — about half from within the USA and half from over sixty other countries. Visitors come from all educational levels, from secondary school through emeritus professors. Interest in translating some RasMol educational scripts into French and Spanish has been expressed (*http://www.umass.edu/microbio/rasmol/nonengli.htm*).

In April, Roger Sayle provided a version of RasMol which displays stereo images. A stereo image of the antianxiety drug alprazolam, which also shows RasMol's ability to render double bonds, may be found at URL *http://www.umass.edu/microbio/rasmol/alprazol.htm*. The PDB-formatted coordinate file for alprazolam was obtained from Dave Woodcock's Chemistry site at Okanagan University College, Canada at the URL *http://www.sci.ouc.bc.ca/chem/molecule/molecule.htm*, which now provides PDB-formatted coordinate files for over five hundred small organic molecules.

Thanks to Margaret Wong (Swinburne University of Technology, Australia) the hypertext form of the RasMol Manual is now up to date with version 2.6 beta-2 and may be found at the URL *http://www.umass.edu/microbio/rasmol/getras.htm#rasmanual*.

Marco Molinaro's UCB-RasMol (University of California, Berkeley, USA), which can load multiple molecules and manipulate them independently, is now available for both Macintosh and Windows (*http://hydrogen.cchem.berkeley.edu:8080/Rasmol/v2.6*).

Educational RasMol scripts for DNA, antibody, and major histocompatibility complex molecules were also mentioned in the January 1996 Quarterly Newsletter. New scripts explore the structure of lipid bilayers and the penetration of water into the gramicidin channel within a bilayer (*http://www.umass.edu/microbio/rasmol/scripts.htm*). They run about fourteen minutes and comprise over two thousand RasMol commands. These scripts were based on molecular dynamics model PDB formatted files kindly provided by Heller, Schaefer, and Schulten [*J. Phys. Chem.* **97**, 8343 (1993)] and by Crouzy, Woolf, and Roux [*Biophys. J.* **67**, 1370, (1994)]. The PDB files are included in the scripts so they can be explored independently in RasMol. Some high-resolution images derived from these bilayer files are also available *at the URL http://www.umass.edu/microbio/rasmol/bilayers.htm*. Don't miss "Five Bakers Dancing".

A new script on protein and hemoglobin structure, together with a paused version of the DNA script, were projected during a full one-hour lecture to a biotechnology class of about one hundred college freshmen. In addition to structure, the theme running through these scripts is that of mutation and the consequences at the protein level (sickle hemoglobin). Freshman Rebecca Pistel wrote the following in her critique:

> "At first I was fairly skeptical about this new method of teaching .... I felt that I would have the tendency to watch ... the "pretty pictures" and not ... focus on the lecture .... However, once the lecture was in full swing, I proved myself very wrong. The graphics clearly caught my attention, and ... I wished to fully understand exactly what I was seeing. I then began to focus attentively on the lecture and became engrossed in the lecture itself."

The existing RasMol scripts can serve as examples for others to follow in writing scripts. Using this approach, Henry Brzeski (University of Strathclyde, UK) has recently provided a script on the lac repressor/operator. Also recently made available are a "Guide to Script Creation for RasMol and Chime" by yours truly as well as an animated "Tutorial on Scripting for Chime and RasMol" by Mitch Miller of MDL Information Systems, Inc. (*http://www.mdli.com/chemscape/chime/example/spttutor/tutorial.html*).

A very exciting educational development has been the creation of Chemscape Chime by Tim Maffett and others at MDLI (*http://www.mdli.com*). MDL announced in August that Chime will be free for all users. Chime produces RasMol-like renderings and has a RasMol-like menu, but works *via* a Web page within Netscape as a "plug-in." In fact, about sixteen thousand lines of RasMol's thirty thousand lines of source code were modified and adapted for use in Chime; however, Chime includes more than eighty thousand lines of source code which are original with MDL. Chime can do several things which RasMol cannot do, including 2D representations, animations of XYZ data files, and execution of scripts from hypertext buttons.

Don't miss the animated thermal motion of an alpha helix found at URL *http://www.umass.edu/microbio/rasmol/scripts.htm*, which has been provided by Raul Cachau.

Chime has enormous educational potential. Dave Woodcock's organic chemistry nomenclature and stereochemistry tutorials *(http://www.sci.ouc.bc.ca/chem/nomenclature/nom1.htm*) use Chime to display many molecules per page. Similar to RasMol, each may be rotated with the mouse or the representation changed, *etc.* Will McClure (Carnegie-Mellon University, USA) has made it possible to click through the twenty amino acids, or the metabolic steps of glycolysis, or the TCA cycle in Chime images (*http://info.bio.cmu.edu/Courses/BiochemMols/BCMolecules.html*). The Medical Education Group at the University of Kansas Medical Center has created an extensive series of Chime script tutorials, *http://www.kumc.edu/research/medicine/biochemistry/bioc800/start.html*, covering amino acids, peptides, secondary structure, domains, antibody, and myoglobin.

In another powerful strategy, one frame on a Web page may be devoted to scrollable hypertext, while an adjacent frame contains a Chime molecular representation. Buttons in the text, when pressed, run scripts to change the content of the graphics frame to illustrate the points made in the text. Examples are Will McClure's tutorial on protein secondary structure (*http://info.bio.cmu.edu/Courses/BiochemMols/BCMolecules.html*), David Marcey's tutorial on DNA polymerase (Kenyon College, USA*, http://www.kenyon.edu/depts/bmb/chime.htm*), and a tutorial on the photosynthetic reaction center by Nixon, Leach, and Rzepa (Imperial College, UK, *http://www.ch.ic.ac.uk/motm/1prc.html*).

Perhaps the most effective educational strategy is to provide interactive quizzes. Will McClure has done this with RasMol-created GIF images of amino acids. You may find this work at *http://info.bio.cmu.edu/Courses/BiochemMols/PQuiz/PQInst.html*). One of his questions is "Which side-chain atom in Trp can participate in hydrogen bonding?" You answer the quiz by clicking on the appropriate atom. Hints are provided when incorrect atoms are clicked, and a correct answer is rewarded with further interesting information.

Links to most of the sites above and many others are on the RasMol home page at *http://www.umass.edu/microbio/rasmol*. Please send URLs for sites which should be linked to the RasMol home page to *emartz@microbio.umass.edu*.

The CATH Classification Scheme of Protein Domain Structural Families

Christine Orengo, Alex Michie, and Sue Jones, University College, London, England; David Jones, Warwick University, Coventry, England; Mark Swindells, Helix Research Institute, Japan; and Janet Thornton, University College and Birkbeck College, London, England (thornton@biochemistry.ucl.ac.uk).

We would like to announce the launch of the new version (1.0) of the CATH Classification Scheme of Protein Domain Structural Families, available from URL *http://www.biochem.ucl.ac.uk/bsm/cath* and soon to be mirrored at the PDB. The approach we have developed to classify protein structures is based on sequence and structure comparisons and heuristic classification. It aims to be largely automated, although some manual checking is required. CATH classifies protein structures by clustering at different hierarchical levels and is based on previously published work [1,2].

CATH is an acronym, where

C = Class

A = Architecture

T = Topological (Fold) Family

H = Homologous superfamily

This scheme provides both a phenetic (class and architecture) and phylogenetic (homologous superfamily) description of protein domains, derived from protein sequences and structures deposited in the PDB.

Following Levitt and Chothia [3] each domain is assigned a class — mainly alpha, mainly beta, or alpha beta (including both alpha+beta and alpha/beta) — automatically using a heuristic algorithm [4]. Each domain is also assigned an architecture, which aims to give an overview of the packing arrangements of the secondary structure, regardless of their sequential connectivity (*e.g.*, all four-alpha bundles are grouped into the same architecture, despite having different topologies). Currently we have identified seventeen beta architectures, ten alpha beta architectures, and only three alpha architectures. At the present time, the assignment of architectures is manual, following clustering into topological or fold (T) families.

The topological family (T) and homologous superfamily (H) for each domain are identified by sequence alignment and structure comparison using the SSAP algorithm [5]. Using predefined cutoffs, the structures are clustered first into homologous (H) families — with either significant sequence similarity or highly similar structures — and then into fold or topological (T) families. The T-level classification is based on the SSAP structures comparison scores and clusters those structures with similar topologies.

Each protein domain is assigned a CATH number describing the classification, which allows facile computer manipulation and sorting. These numbers are similar to the EC numbers for enzyme classification.

The procedure to derive the classification is described briefly as follows:

PROCEDURE		OUTPUT
1. SIFT PDB Entries	⇒	Acceptable Chains
2. Sequence Comparison and Clustering	⇒	Sequence Family (S)
3. Domain Assignment	⇒	Domain Boundaries
4. Recluster Domains by Sequence	⇒	Updated Sequence Families (S)
5. Class Assignment	⇒	Class (C)
6. Structure Comparison and Clustering for Domain Representatives	⇒	Homologous Family (H) Topology Family (T)
7. Architecture Assignment	⇒	Architecture (A)

The CATH classification on the Web is accompanied by complementary data and tools. This includes a **Glossary** of terms, used to describe protein folds and a **Lexicon** which gives a brief description (including text and diagrams) of each major architecture. Summary information for each fold family is available, and in addition there is an attached file of derived data for each PDB entry which describes ligands and secondary structure motifs in tables and schematic diagrams (*e.g.*, LIGPLOT).

We will also provide, in the near future, a CATH server which allows a new structure to be compared against the current database of domains.

There are facilities for searching CATH and options to select summary information at each level (*e.g.*, lists of architectures for a selected class or lists of representatives from each non-homologous family).

1. C.A. Orengo, T.P. Flores, W.R. Taylor, and J.M. Thornton, *Prot. Eng.* **6**, 485-500 (1993).
2. C.A. Orengo, D.T. Jones, and J.M. Thornton, *Nature,* **372**, 631-634 (1994).
3. M. Levitt and C. Chothia, *Nature* **261**, 552-557 (1976).
4. A.D. Michie, C.A. Orengo, and J.M. Thornton, *J. Mol. Biol.* **263**, 2-18 (1996).
5. W.R. Taylor and C.A. Orengo, *J. Mol. Biol.* **208**, 1-22 (1989).

FOLD: a Program for the Analysis and Display of Protein Structures

Darren R. Flower, Astra Charnwood, Loughborough, Leicestershire, UK (darren.flower@charnwood.gb.astra.com).

FOLD is a computer program designed to facilitate the analysis of protein structure [1]. FOLD implements a number of new or enhanced methods for the definition, and subsequent analysis, of hydrogen bonded protein secondary and super-secondary structures. These include methods for the identification of simple features such as hydrogen bonds, α-helices, β-strands, β-bulges, as well as β- and γ-turns. Techniques are also described for the definition and analysis of higher-order structures such as β-hairpins, open β-sheets, and β-barrels.

FOLD assigns secondary structures using pattern recognition methods based on matching different characteristic hydrogen bonding patterns. The program works in a hierarchical fashion — it begins with objective identification of all main chain-main chain hydrogen bonds and proceeds to find patterns characteristic of α- and 3_{10}-helices and β-strands. This allows individual residues to be defined as a particular secondary structure type. Assigned residues are then grouped into elements — helices and strands. In the case of β-strands, having defined a set of β-strands and their connectivity, straightforward techniques drawn from graph theory are used to first partition these strands into sheets and to then analyze their structure.

The topology of a protein β-sheet, the relationship between the sequential ordering of strands, and their hydrogen-bonded connectedness in space, is an important and well-studied property of such structures. Two systems of nomenclature have been proposed to describe β-sheet topology. One is based on following a path through the sequence order of strands and noting their spatial separation within the sheet. The other, lesser-used approach, is based on following a path through the connectedness of neighboring strands and noting their sequence separation. Both sequence-based and connection-based forms of nomenclature are automated within FOLD.

β-sheets possess complex topological properties, including cycles and branches, which are not readily expressed by either form of consecutive notation. To overcome such limitations, a shorthand nomenclature able to express an arbitrary β-sheet topology has been developed [2]. The automated generation of this notation is also implemented within FOLD. Other high level features, such as β-hairpins and β-barrels, are also detected and analyzed automatically by the program. For example, FOLD implements a method for identifying barrels based on ring perception [3].

FOLD is written in standard FORTRAN 77 and runs under both VMS and Unix. FOLD is controlled *via* a simple command line interface through a set of keywords — the parameters controlling the program are fully configurable. The program can operate in either of two modes: individually reading, analyzing, and visualizing structures, or batch processing sets of structures. FOLD is flexible in the type and quantity of its output — data associated with each aspect of the structure analyzed, β-bulges, β-hairpins, sheet topology, *etc*., as well as different types of partial or overall summary, are written to separate self-naming files.

In addition to its considerable textual output, FOLD supports GL-based interactive visualization of hydrogen bonding patterns and secondary structural features. The program has two display styles: an atomic and a schematic mode. The atomic display style is detailed but straightforward, showing the backbone of the protein, color-coded by secondary structure type, together with all of the hydrogen bonds identified by the program. By contrast, in its schematic mode, FOLD seeks to represent the overall structure of a protein in a highly simplified, but aesthetically pleasing, way. In common with most so-called ribbon drawings β-strands are depicted as arrows, α-helices as spiral ribbons, and other structures as a coiling rope or line. This allows the display style and view to be manipulated interactively, although the definitions of secondary structures are those generated automatically by the program.

FOLD may be retrieved *via* anonymous FTP to *guitar.rockefeller.edu* from the directory */pub/jpo/fold.tar.gz*.

1. D.R. Flower, FOLD: Integrated Analysis and Display of Protein Secondary Structure, *J. Mol Graph.* **13**, 377-384 (1995).

2. D.R. Flower, β-sheet Topology: A New System of Nomenclature, *FEBS Letters* **344**, 247-250 (1994).

3. D.R. Flower, Automating the Detection and Analysis of Protein β-barrels, *Prot. Eng.* **7**, 1305-1310 (1994).

A WWW Service System for Automatic Comparison of Protein Structures

Guoguang Lu, Karolinska Institute, Stockholm, Sweden (guoguang@alfa.mbb.ki.se).

As the number of newly-determined protein structures increases explosively, more convenient computer programs for comparing protein topologies and atomic structure are required for structure analysis. We have developed a system for automatic protein structure comparison *via* the WWW using the program TOP [1]. By visiting a URL address where the system has been installed, users can superimpose two protein structures or detect whether a newly-determined structure is similar to any structure currently in the PDB.

The structure superposition is carried out by the TOP program which was written in FORTRAN language compiled under Unix or VAX/VMS systems [1]. After reading coordinate files in PDB format, the TOP program simplifies each protein secondary structure element as a pair of points, systematically searches all the possible alignments of these elements according to several criteria, and refines the initial superimposition by minimization of the distances between Cα atoms. TOP can search to find which proteins among a group of proteins are structurally similar to a probe protein. The comparison group can be the entire PDB. The output generated by the program is the superposition matrix, superimposed coordinates, secondary structure alignment, and sequence alignment based on the three-dimensional structures. The program can also use a compact file which contains only the secondary structure information contained in the PDB so that users are able to perform similarity searching without the PDB on their local computers. This compact database is updated regularly and can be copied from the PDB FTP server. The program is easy to handle and user-friendly.

We have also built up a Web service system supported by several html and perl script files, which provides a Graphical User Interface (GUI) for the TOP program. Without installation of the program on their own system, users can furnish their coordinates to the Internet by the Netscape program (version 2.0 or higher) so that a server computer, with PDB or the compact file installed, can automatically perform the rapid superposition or similarity searching and send the superimposed coordinates or a list of similar structures to the user by e-mail.

There have been several applications of the comparison method. For example, the FAD binding domain of Nitrate Reductase was found to be structurally similar with BARWIN (unpublished results). A circular permutation of the α/β barrel between transaldolase and aldolase was also found by the TOP program [2].

A test Web site is available which includes a detailed description of how to copy and use the program. We welcome other laboratories, especially those who are maintaining and updating the PDB, to install mirror sites of the Web server in order to facilitate protein structure comparisons as a tool.

TOP program URLs include the following:

- Executing the structural comparison *via* WWW and other links:
  http://gamma.mbb.ki.se/~guoguang/webtop/wwwtop.html
- Instructions on how to copy and use the program:
  http://gamma.mbb.ki.se/~guoguang/top.html
- Copy of the Unix version of the program:
  ftp://gamma.mbb.ki.se/pub/guoguang/top.tar.Z
- Copy of the compact secondary structure database of PDB:
  ftp://gamma.mbb.ki.se/pub/guoguang/sndlib.tar.Z
- Instructions on how to mirror WWW server site:
  ftp://gamma.mbb.ki.se/pub/guoguang/webtop.tar.Z

———

1. G. Lu, TOP — An Automatic Topological and Atomic Comparison Program for Protein Structures, *http://gamma.mbb.ki.se/~guoguang/top.html* (1996).

2. J. Jia, W. Huang, U. Schrken, H. Sahm, G.A. Sprenger, Y. Lindqvist, and G. Schneider, Crystal Structure of Transaldolase B from E. Coli Suggests a Circular Permutation of the α/β Barrel within the Class I Aldolase Family, *Structure* **4**, 715-124 (1991).

Note: I am grateful to Dr. Ylva Lindqvist and Professor Gunter Schneider for their suggestions and encouragement. I also wish to thank Dr. Roman Laskowski for permission to merge his program SECSTR into the TOP program.

PROLYSIS: a WWW Server Dedicated to Proteases and Their Inhibitors

Thierry Moreau, University of TOURS, France (moreaut@ univ-tours.fr).

Proteolytic enzymes are now widely known to play an important role in many physiological processes. Although proteases were among the first enzymes to be identified and characterized, the study of proteolytic enzymes is still a very active field of research. Our knowledge of these molecules has increased dramatically over the past decade, as a result of spectacular advances in molecular and structural biology. At the same time, there has been a great deal of research on protease inhibitors and this has provided new openings for biomedical/biotechnological applications based on proteases and/or their inhibitors.

The PROLYSIS Web server was developed to provide a variety of on-line information about proteases and their natural or synthetic inhibitors *(http://delphi.phys.univ-tours.fr/Prolysis).*

Introductory documents on proteases and protease inhibitors provide the non-specialist with a general survey of these molecules. The Enzyme Commission (EC) nomenclature for proteases is included together with a list of 3D structures (and their respective ID codes) of proteases and protease inhibitors available in the PDB.

Users may find data on proteases and natural/synthetic protease inhibitors which are used as biochemical tools (this section is not yet fully implemented). Other sections cover various aspects of the field, such as the prevention of unwanted proteolysis (not yet implemented), the use of synthetic substrates, and assay methods specific to proteases and tight-binding inhibitors.

The “Image gallery” section contains images of representative members of the various classes of proteases, protease inhibitors, and protease-inhibitor complexes. The most interesting structural features of the molecules and/or molecular interactions are highlighted together with a short, explanatory caption.

There are also a few animated images (animated GIFs). However, PROLYSIS will soon use the new interactive tools being developed for the Internet, such as VRML (Virtual Reality Modeling Language) and JAVA applets.

Links are available to various Internet resources dealing with specific aspects of proteases and their inhibitors. These include the Web site for peptidase families (new classification of proteases) implemented in the SwissProt database (University of Geneva, Switzerland), the HIV protease image home page (University of Queensland, Australia), the guide to protein and peptide cleavage recipes (Rockfeller University, USA) to name a few, and links to the major on-line biological journals. The “Newsroom” section contains selected outstanding or fascinating facts about proteases and protease inhibitors.

Along with its role for educational research, PROLYSIS offers the scientific community up-to-date information and images on the latest discoveries such as the 3D structure of the proteasome (contributed by J. Lowe, Max Planck Institute for Biochemistry, Munich, Germany) or the 3D structure of the procathepsins B (contributed by D. Turk, Josef Stefan Institute, Ljubljana, Slovenia, and M. Cygler, Biotechnology Research Institute, Montreal, Canada).

The PROLYSIS server, developed by Dr. Thierry Moreau, a lecturer at the University of Tours, France, is continually being improved. Comments, suggestions, and contributions are welcome at *moreaut@univ-tours.fr.*

Notes of a Protein Crystallographer

Cele Abad-Zapatero, Abbott Laboratories, Abbott Park, IL, USA (abad@abbott.com).

— On the Size, Shape, and Texture of Protein Molecules

Many of the old papers describing the structures of the newly-solved protein structures by X-ray diffraction attempted to describe their overall size and shape by giving the dimensions of the parallelepiped or ellipsoid enclosing them. As more structures were known, attempts were also made to fit these unique, miniature, three-dimensional objects to certain canonical forms such as paraboloids and prolate or oblate ellipsoids of different characteristics. Recently, I saw an article describing the novel structure as a “highly oblate sphere.” These various descriptions reflect the difficulty of describing very irregular objects in terms of the geometrical concepts and dimensions that we use to characterize many of our day-to-day objects.

I have always been fascinated by this problem: how do you mathematically describe irregular objects such as protein structures? The next few paragraphs will be somewhat technical, but I beg you to follow along or skip to where the text is less dense — it will be worthwhile.

In about 1987, I convinced C. T. Lin, a statistician colleague from Abbott Laboratories, to help me tackle the problem from a different viewpoint. We worked together for a few years, and in 1990 we published a paper entitled “Statistical Descriptors for the Size and Shape of Globular Proteins” [C. Abad-Zapatero and C.T. Lin, *Biopolymers* **29**, 1745-1754 (1990)]. I do admit that this is a paper with a cryptic title, published in a journal not directly related to protein structures. If you are interested, I still have a few reprints.

Restricting ourselves to the most “globular proteins”, we computed the frequency distribution of all possible distances (i to j) between C_α carbons of a protein structure and studied the resulting distribution. Using all atoms in the structure didn’t modify the conclusions of the work. As expected, these distributions were skewed, non-Gaussian-like with a sharp spike at 3.8 Å corresponding to the well-known C_α-C_α distances between adjacent residues in the polypeptide chain — we can ignore this information for the remainder of this article.

It was immediately clear that the median or mode of the distribution was an excellent descriptor of the protein (object) size as correlated to the molecular weight of the proteins in the sample or to their radius of gyration (Rg).

Using one statistical trick named the Box-Cox transformation, one could transform the observed distribution to be like a true Gaussian, and the resulting parameter λ (with values ranging from 0 to 1) could be used to measure how far the original distribution deviated from an ideal Gaussian distribution whose λ value is equal to 1. After some analyses we realized that the parameter λ was also a good descriptor of the shape of the original object, similar in character to the axial ratio (a/b) of an ellipse or ellipsoid.

It is well known that the Surface/Volume ratio is a reasonable (*albeit* size dependent) measure of shape, and one can define a size-corrected shape descriptor by the expression:

$$s = (\text{Surface} \cdot \text{Rg})/\text{Volume}$$

resulting in a dimensionless quantity s.

The expected value of s for a solid sphere can be shown to be 2.324, and larger values correspond to more irregularly-shaped globular objects.

The interesting thing is that although there is a clear correlation between s and the λ parameter of the Box-Cox transformation, the observed λ values also reflect a very subtle characteristic of the irregular protein structures — its surface texture.

The description in mathematical terms of the roughness or texture of a surface has been facilitated by the work on fractal geometry by the mathematician Benoit Mandelbrot. He has introduced the Hausdorff’s dimension with values ranging from 2.0, for a perfectly smooth surface, to 3.0 for a infinitely rough surface — so much so that it is essentially a three-dimensional object.

This brief summary of the paper should suggest that most globular protein structures or globular domains within them (and possibly any round, irregular object) could be described by three independent parameters which can be plotted on a three-dimensional Cartesian coordinate system with axes: size (median of the distribution of interatomic distances, ranging from 0 to infinity), shape (2.324 and larger), and texture (2.0 to 3.0). Although size and shape are important, I feel that the most subtle and interesting is texture.

In some literature there have been attempts to relate surface roughness on protein surfaces with functional properties [M. Lewis and D.C. Rees, *Science* **230**, 1163-1165 (1985)], but the original seeds have not yielded any substantial fruits. In my own work, I have come to the conclusion (unpublished) that the roughness of the atomic structure of protein surfaces has two components: the first is the granular texture due to the fact that protein structures are built of small spheres that we call atoms; the second is a smoother-varying component which is the subtle undulations that remain after the atomic component has been removed. It is at this level that active-site cavities and inter-subunit surfaces are relevant.

With our stereoscopic vision and aided by light, our eyes can perceive very readily the properties of size and shape. Yet texture, I contend, is a property that has to be discerned by touch. How can we feel the texture of a protein molecule? Who has touched a three-dimensional model of a protein to feel its texture? One can argue that running the fingers or the palm of the hand over some of the old protein models built of Kendrew parts or CPK components might give you an idea of texture. However, I have a better idea.

For different materials and surfaces, try first running your hand or fingers over marble and glass. Then try the trunk of different types of trees: birches, poplars, pines, elms, maples, oaks, *etc.* and you will have a sense of different roughness or textures.

For the textures of different proteinaceous materials try passing your hand over the surface of sweaters made of different types of wool such as llama, alpaca, or merino wool. Then do the same with shirts or blouses made of silk. Finally, you may try touching the hair of your loved ones: spouses, children, lovers, or friends — you may wonder how protein texture can result in a magic avalanche of chemical reactions in your brain and, simultaneously, a myriad of sensations and feelings in your heart. ■

AFFILIATED CENTERS AND MIRROR SITES

Twenty-eight affiliated centers offer the Protein Data Bank database archives for distribution. These centers are members of the Protein Data Bank Service Association (PDBSA). Centers designated with an asterisk(*) may distribute the archives both on-line and on magnetic or optical media; those without an asterisk are on-line distributors only. Information is given for those Centers which are now also official PDB Mirror Sites.

BIRKBECK
Crystallography Department
Birkbeck College, University of London
London, United Kingdom
Ian Tickle (44-171-6316854)
tickle@cryst.bbk.ac.uk
http://www.cryst.bbk.ac.uk/pdb/pdb.html/

BMERC
BioMolecular Engineering Research Center
College of Engineering, Boston University
Boston, Massachusetts, USA
Nancy Sands (617-353-7123)
sands@darwin.bu.edu
http://bmerc-www.bu.edu/

CAOS/CAMM
Dutch National Facility for Computer Assisted Chemistry
Nijmegen, The Netherlands
Jan Noordik (31-80-653386)
noordik@caos.caos.kun.nl
http://www.caos.kun.nl/

****CCDC***
Cambridge Crystallographic Data Centre
Cambridge, United Kingdom
David Watson (44-1223-336394)
watson@chemcrys.cam.ac.uk

CSC
CSC Scientific Computing Ltd.
Espoo, Finland
Heikki Lehvaslaiho (358-0-457-2076)
heikki.lehvaslaiho@csc.fi
http://www.csc.fi/

EMBL
European Molecular Biology Laboratory
Heidelberg, Germany
Hans Doebbeling (49-6221-387-247)
hans.doebbeling@embl-heidelberg.de
http://www.EMBL-Heidelberg.DE/

EMBL OUTSTATION: THE EUROPEAN BIOINFORMATICS INSTITUTE
Wellcome Trust Genome Campus
Hinxton, Cambridge, United Kingdom
Philip McNeil (44-1223-494-401)
mcneil@ebi.ac.uk
http://www.ebi.ac.uk

PDB Mirror Site:
http://www2.ebi.ac.uk/pdb
Philip McNeil (pdbhelp@ebi.ac.uk)

FUJITSU KYUSHU SYSTEM ENGINEERING LTD.
Computer Chemistry Systems
Fukuoka, Japan
Masato Kitajima (81-92-852-3131)
ccs@fqs.fujitsu.co.jp

FMI
Friedrich Miescher Institute
Basel, Switzerland
Carl David Nager (41-61-697-5678)
carl.david.nager@fmi.ch
http://www.fmi.ch

ICGEB
International Centre for Genetic Engineering and Biotechnology
Trieste, Italy
Sandor Pongor (39-40-3757300)
pongor@icgeb.trieste.it

IGBMC
Laboratory of Structural Biology
Strasbourg (Illkirch), France
Frederic Plewniak (33-8865-3273)
plewniak@igbmc.u-strasbg.fr
http://www-igbmc.u-strasbg.fr

****JAICI***
Japan Association for International Chemical Information
Tokyo, Japan
Hideaki Chihara (81-3-5978-3608)

****MSI***
Molecular Simulations Inc.
San Diego, California, USA
Mark Forster (619-458-9990)
mjf@msi.com
http://www.msi.com/

NATIONAL CENTER FOR BIOTECHNOLOGY INFORMATION
National Library of Medicine, National Institutes of Health
Bethesda, Maryland, USA
Stephen Bryant (301-496-2475)
bryant@ncbi.nlm.nih.gov
http://www.ncbi.nlm.nih.gov/

NCHC
National Center for High-Performance Computing
Hsinchu, Taiwan, ROC
Jyh-Shyong Ho (886-35-776085; ext: 342)
c00jsh00@nchc.gov.tw

NCSA
National Center for Supercomputing Applications
University of Illinois at Urbana-Champaign
Champaign, Illinois, USA
Allison Clark (217-244-0768)
aclark@ncsa.uiuc.edu
http://www.ncsa.uiuc.edu/Apps/CB/

NCSC
North Carolina Supercomputing Center
Research Triangle Park, North Carolina, USA
Linda Spampinato (919-248-1133)
linda@ncsc.org
http://www.mcnc.org

****OML***
Oxford Molecular Ltd.
Oxford, United Kingdom
Steve Gardner (44-1865-784600)
sgardner@oxmol.co.uk
http://www.oxmol.co.uk/

****OSAKA UNIVERSITY***
Institute for Protein Research
Osaka, Japan
Masami Kusunoki (81-6-879-8634)
kusunoki@protein.osaka-u.ac.jp

PEKING UNIVERSITY
Molecular Design Laboratory, Institute of Physical Chemistry
Beijing 100871, China
Luhua Lai (86-10-62751490)
lai@ipc.pku.edu.cn
http://www.ipc.pku.edu.cn

PDB Mirror Site:
http://www.ipc.pku.edu.cn/pdb/
Li Weizhong (liwz@csb0.ipc.pku.edu.cn)

PITTSBURGH SUPERCOMPUTING CENTER
Pittsburgh, Pennsylvania, USA
Hugh Nicholas (412-268-4960)
nicholas@psc.edu
http://pscinfo.psc.edu/biomed/biomed.html

SAN DIEGO SUPERCOMPUTER CENTER
San Diego, California, USA
Philip E. Bourne (619-534-8301)
bourne@sdsc.edu
http://www.sdsc.edu/~bourne

SEQNET
Daresbury Laboratory
Warrington, United Kingdom
User Interface Group (44-1925-603351)
uig@daresbury.ac.uk

****TRIPOS***
Tripos, Inc.
St. Louis, Missouri, USA
Akbar Nayeem (314-647-1099; ext: 3224)
akbar@tripos.com

UNIVERSITY OF GEORGIA
BioCrystallography Laboratory
Department of Biochemistry and Molecular Biology
University of Georgia
Athens, Georgia, USA
John Rose or B.C. Wang (706-542-1750)
rose@BCL4.biochem.uga.edu
http://www.uga.edu/~biocryst

UPPSALA UNIVERSITY
Department of Molecular Biology
Uppsala University
Uppsala, Sweden
Alwyn Jones (46-18-174982)
alwyn@xray.bmc.uu.se

WEHI
The Walter and Eliza Hall Institute
Melbourne, Australia
Tony Kyne (61-3-9345-2586)
tony@wehi.edu.au
http://www.wehi.edu.au

PDB Mirror Site:
http://pdb.wehi.edu.au/pdb
Tony Kyne (tony@wehi.edu.au)

WEIZMANN INSTITUTE OF SCIENCE
Rehovot, Israel
Jaime Prilusky (972-8-9343456)
lsprilus@weizmann.weizmann.ac.il

PDB Mirror Site:
http://pdb.weizmann.ac.il
Alexander Faibusovitch (pdbhelp@pdb.weizmann.ac.il)

Protein Data Bank
Biology Department, Bldg. 463
Brookhaven National Laboratory
P.O. Box 5000
Upton, NY 11973-5000 USA

SCIENTIFIC CONSULTANTS

John P. Rose
University of Georgia
Athens, Georgia, USA

S. Swaminathan
University of Pittsburgh
Pittsburgh, Pennsylvania, USA

Sasha Faibusovich, Clifford Felder, Kurt Giles,
Mia Raves, and Vladimir Sobolev
Weizmann Institute of Science
Rehovot, Israel

PDB STAFF

Joel L. Sussman, Head
Enrique E. Abola, Deputy Head and Science Coordinator
Jaime Prilusky, Interim Head Database Development

Frances C. Bernstein
Avraham Bluestone
Minette Cummings
Betty R. Deroski
Pamela A. Esposito
James Flanagan
Arthur Forman
Jiansheng Jiang
Patricia A. Langdon
Michael D. Libeson
Dawei Lin
Nancy O. Manning
John E. McCarthy
Christine Metz
Regina K. Shea
Janet L. Sikora
John Spiletic
Brigitte R. Sylvain
Dejun Xue

STATEMENT OF SUPPORT

The PDB is supported by a combination of Federal Government Agency funds (work supported by the U.S. National Science Foundation; the U.S. Public Health Service, National Institutes of Health, National Center for Research Resources, National Institute of General Medical Sciences, and National Library of Medicine; and the U.S. Department of Energy under contract DE-AC02-76CH00016) and user fees.

ACCESS TO THE PDB

Main Telephone	1 516-344-3629
Help Desk Telephone	1 516-344-6356
Fax	1 516-344-5751
Help Desk	pdbhelp@bnl.gov
General Correspondence and Depositions	pdb@bnl.gov
WWW Home Page	http://www.pdb.bnl.gov
FTP Server	ftp.pdb.bnl.gov
Network Services	sysadmin@pdb.pdb.bnl.gov
Entry Error Reports	errata@pdb.pdb.bnl.gov
Order Information	orders@pdb.pdb.bnl.gov
User Group	PDBusrgrp@suna.biochem.duke.edu
Listserver Postings	pdb-l@pdb.pdb.bnl.gov
Listserver Subscriptions	listserv@pdb.pdb.bnl.gov
to subscribe, the text of your message should be	subscribe PDB-L Your Name
Deposition Form available from	PDB WWW Home Page and FTP Archives
Newsletters available from	PDB WWW Home Page and FTP Archives

SOFTWARE AVAILABLE VIA ANONYMOUS FTP

Software	FTP Site	Directory Location
PDB-Browse	ftp.pdb.bnl.gov	/pub/pdbbrowse
PDB-Shell	ftp.pdb.bnl.gov	/pub/pdbshell
Getentry	ftp.pdb.bnl.gov	/pub/getentry
RasMol	ftp.pdb.bnl.gov	/pub/other-software/Rasmol
Kinemage	ftp.pdb.bnl.gov	/pub/kinemage
	suna.biochem.duke.edu	/pub
Mirror	ftp.pdb.bnl.gov	/pub/other-software/Mirror
Perl	ftp.pdb.bnl.gov	/pub/other-software/Perl
	ftp.netlabs.com	/pub/outgoing/perl5.0
Tcl	ftp.pdb.bnl.gov	/pub/other-software/Tcl
WHAT_CHECK	ftp.pdb.bnl.gov	/pub/whatcheck
Gopher client	boombox.micro.umn.edu	/pub/gopher
Mosaic client	ftp.ncsa.uiuc.edu	/Web/Mosaic
Netscape client	ftp.netscape.com	/netscape

INTERESTING WWW SITES

Server	URL
ACA (American Crystallographic Association)	http://www.hwi.buffalo.edu:80/ACA
BMCD (The Biological Macromolecule Crystallization Database and the NASA Archive for Protein Crystal Growth Data)	http://ibm4.carb.nist.gov:4400/bmcd/bmcd.html
BMRB (BioMagResBank)	http://www.bmrb.wisc.edu
Brookhaven National Laboratory	http://www.bnl.gov
CCDC (Cambridge Crystallographic Data Centre)	http://www.ccdc.cam.ac.uk
Crystallography Worldwide	http://www.unige.ch/crystal/w3vlc/crystal.index.html
DALI - Comparison of Protein Structures in 3D	http://www.embl-heidelberg.de/dali/dali.html
EBI (European Bioinformatics Institute)	http://www.ebi.ac.uk
EMBL (European Molecular Biology Laboratory)	http://www.embl-heidelberg.de
ExPASy Molecular Biology Server	http://expasy.hcuge.ch
GDB (Genome Data Base)	http://gdbwww.gdb.org
GenBank (NIH Genetic Sequence Database)	http://www.ncbi.nlm.nih.gov/Web/Genbank/index.html
Human Genome Project Information	http://www.ornl.gov/hgmis/
IUCr (International Union of Crystallography)	http://www.iucr.ac.uk
Johns Hopkins University BioInformatics	http://www.gdb.org
mmCIF	http://ndbserver.rutgers.edu/mmcif
MOOSE (Macromolecular Structure Database at San Diego Supercomputer Center)	http://db2.sdsc.edu/moose/
NCBI GenBank	http://www.ncbi.nlm.nih.gov
NDB (Nucleic Acid Database)	http://ndbserver.rutgers.edu
NIH (National Institutes of Health)	http://www.nih.gov
O Home Page	http://kaktus.kemi.aau.dk/
PDB (Protein Data Bank)	http://www.pdb.bnl.gov
Pedro's BioMolecular Research Tools	http://www.public.iastate.edu/~pedro/research_tools.html
PIR (Protein Identification Resource)	http://www.gdb.org/Dan/proteins/pir.html
PROCHECK - to submit a PDB file for analysis	http://www.cryst.bbk.ac.uk/PPS/procheck/test.html
Protein Motions Database	http://hyper.stanford.edu/~mbg/ProtMotDB/
Protein Science	http://www.prosci.uci.edu
Protein Structure Verification-Biotech Server	http://biotech.embl-heidelberg.de:8400 or http://biotech.pdb.bnl.gov:8400
RasMol Home Page	http://www.umass.edu/microbio/rasmol
SCOP: Structural Classification of Proteins	http://scop.mrc-lmb.cam.ac.uk/scop or http://www.pdb.bnl.gov/scop
Swiss-Prot Sequence Database	http://expasy.hcuge.ch/sprot/sprot-top.html
Weizmann Institute, Biological Computing Division	http://dapsas1.weizmann.ac.il
X-PLOR Home Page	http://xplor.csb.yale.edu/

BROOKHAVEN ORDER FORM

Name of User ______________________________ Date ______________

Organization ______________________________ Phone ______________

Address ______________________________ Fax ______________

______________________________ E-mail ______________

____

- ***Price is valid through September 30, 1997***
- ***Price is per CD-ROM set released --- releases occur four times per year***
- ***Facsimile and phone orders are not acceptable***

The Protein Data Bank ***MUST*** receive all three of the following items before shipment can be completed (please send all required items together *via* postal mail — facsimile and phone orders are ***NOT*** acceptable):

1. Completed order form;
2. Mailing label indicating exact shipping address;
3. Payment (using one of the two options below):

--> Check payable to Brookhaven National Laboratory in US dollars and drawn on a US bank. Foreign checks cannot be accepted and will be returned.

--> Original purchase order payable to Brookhaven National Laboratory. After your order is processed, you will be invoiced by Brookhaven National Laboratory. Please indicate exact address to which invoice should be sent:

____

____

____

____

A wire transfer is acceptable only ***AFTER*** we have received an original purchase order from your organization and you have been invoiced by Brookhaven. After receiving Brookhaven's invoice, your bank may send a wire transfer to:

Bank name	:	Morgan Guaranty Trust Company of New York
Account name	:	Brookhaven National Laboratory
Account number	:	076-51-912

Please send all required items together *via* postal mail to:

Protein Data Bank Orders
Biology Department, Building 463
Brookhaven National Laboratory
P.O. Box 5000
Upton, NY 11973-5000
USA

1 Protein Data Bank CD-ROM Set — ISO 9660 Format *$357.25*

(tax and shipping charges not applicable)

This page intentionally left blank.

This page intentionally left blank.

Protein Data Bank
Biology Department, Bldg. 463
Brookhaven National Laboratory
P.O. Box 5000
Upton, NY 11973-5000 USA

(516) 344-3629

First Class