Protein Data Bank
pdb™
Quarterly Newsletter published by
Brookhaven National Laboratory
Protein Data Bank
Release #84 April 1998 Brookhaven National Laboratory

April 1998 CD-ROM Release

7578 Released Atomic Coordinate Entries

	Molecule Type
6723	proteins, peptides, and viruses
298	protein/nucleic acid complexes
545	nucleic acids
12	carbohydrates

	Experimental Technique
183	theoretical modeling
1191	NMR
6204	diffraction and other

1972	Structure Factor Files
429	NMR Restraint Files

The total size of the atomic coordinate entry database is 3.4 GB uncompressed.

Table of Contents

What's New at the PDB 1
Archive Management 2
Local High School Students Visit the PDB 2
3DB Browser™: Tips, Questions and Answers 3
Some WHAT_CHECK Checks Explained 4
Status of the mmCIF Dictionary 5
MOLMOL: A Program for Display and Analysis of Macromolecular Structures 5
Notes of a Protein Crystallographer — Molecular Docking and the Broken Heart 7
Web Sites Referenced in this Newsletter 8
Related WWW Sites 9
Affiliated Centers and Mirror Sites 10
PDB™ Order Form 13
PDB Access, FTP Directory Structure, Consultants, Staff, Support and Instructions to Authors 15

Internet Sites

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

What's New at the PDB

Joel L. Sussman

During the past few months there has been a great deal of discussion in the scientific press about the necessity of changing the PDB's on-hold policy. This policy is based on the decision of the International Union of Crystallography (IUCr), in the late 1980's, which permits depositors of 3D structures of biological macromolecules to ask the PDB to delay release of their coordinates for up to one year following publication of their article in an IUCr journal. The purpose of this policy was, in part, to give a lab the chance to use their structural data for further studies before others, who have not spent years to obtain it, can do the same without any experimental effort. In parallel, funds for certain (industrial) projects are often given with a stipulation that although the results may be published, the coordinates should not be released for a year following publication.

Recently, *Nature Structural Biology* conducted a survey on the Internet which showed that approximately two-thirds of those who responded were in favor of asking the IUCr to eliminate this on-hold policy (http://us.nature.com/survey/nsb_poll.nclk). In parallel, two of the most prominent journals in this field, *i.e., The Journal of Biological Chemistry* and *The Proceedings of the National Academy of Sciences (USA)* (*PNAS*) have changed their policy to require release of coordinates upon publication of papers as a prerequisite for publication. This was stated in an Editorial recently published in *PNAS* **95**(9), pg. iii (1998).

> "In response to a growing consensus in the scientific community [...], the Editorial Board has adopted the following policy: As in the past [...], all authors who submit a paper to Proceedings that describes a new or revised structure must deposit their coordinates in the Protein Data Bank of the Brookhaven National Laboratory or an equivalent public archive. These coordinates, however, must now be released when the articles are published. The up to one-year hold on release is no longer acceptable."

The PDB feels strongly that it is the role of the journals to keep enforcing their rules that data **must** be submitted to the PDB, and that reference to a PDB ID code should be included in the paper, in order to permit publication of the article. To help in making 'release on publication' work smoothly for those journals following this policy, the PDB has developed a 'layered release' approach for submission of entries, as described in the PDB Newsletters of October 1997 and January 1998 (http://www.pdb.bnl.gov/pdb-docs/newsletter.html). It will allow for virtually immediate release of entries onto the PDB Web server without PDB staff intervention. A PDB ID code will be issued immediately, but the depositor can indicate if the data are to be released right away (which the PDB will encourage), upon publication of the accompanying article, or after an 'on-hold' period.

It should be noted that depositors will get immediate feedback at deposition time from a series of validation programs used by the PDB, which will help them to decide whether their data are actually suitable for deposition, or indeed for publication, or still require further work. An entry, which is considered ready by the depositor, will be referred to as 'Layer 1' or 'author approved' entry.

Following release of Layer 1, the PDB staff will process the entry in the same way as at present. This processing will include standardization of nomenclature, and, more importantly, data representation. Most of this work covers issues not fully delegated to software at present. The resulting entry, after author approval, will be loaded on the PDB server as Layer 2. **We strongly believe that such thorough checking and annotation is essential for ensuring the long-term value of the data.**

The PDB is working with the journals which publish articles on macromolecular structures, to coordinate the release of the data by the PDB publication. This should permit easy access to the coordinates in the PDB, which will make it possible for readers to visualize the structure, in 3D, as they are reading the article in a printed journal. We at the PDB are most interested in hearing your comments on these ideas to make structural information more easily and more rapidly accessible to the scientific community.

Archive Management

Enrique Abola and Nancy Manning

Layered Release

In early July, the PDB will convert to the Layered Release protocol that allows for the virtually immediate release of entries to the public. For the first time, depositors will see the output of our processing programs and will be able to take appropriate action before submission to the PDB.

Following submission, the author-approved entry will be released automatically - no corrections will be made by PDB staff. Processing after this initial release will address issues related to any heterogroups, specific processing instructions, and standardization of nomenclature, annotation, and data representation. The resulting entry will replace the first version on our server.

Heterogen groups will be checked against the current PDB Het Dictionary (ftp://pdb.pdb.bnl.gov/pub/resources/hetgroups/het_dictionary.txt) to see if the HET ID and the atom nomenclature used for a group is consistent with the dictionary (e.g., is the GLC group in the coordinate file a glucose molecule as given in the PDB Het Dictionary and are the atoms properly named?). Groups that are not in the dictionary and for which there is no conflict on the HET ID code will be accepted as is and will be checked and standardized as part of the regular processing to be done after the first layer is loaded.

Layered Release has been thoroughly described in this column previously (see the October 1997 and January 1998 PDB Newsletters), and further documentation may be found at our Web site.

PBD ID Codes

It has come to our attention that there may be some confusion as to the correct ID code to be used as proof of data submission to the Protein Data Bank.

Official PDB ID codes are in the form of a number followed by 3 alphanumeric characters, e.g., 2ACE and 1EV8. This ID code alone serves as proof of deposition with the PDB.

Sometimes other numbers are mistakenly thought to be PBD ID codes. PDB's Web-based deposition procedure, AutoDep, assigns a security number in the form of BNL-xxx, or EBI-xxx for those submitted at the EMBL outstation at the European Bioinformatics Institute. The four characters "4TST" are specifically used as a temporary placeholder during AutoDep until the deposition is completed, at which point an actual PDB ID code is assigned. In addition, each set of submitted coordinates is given an internal PDB administrative tracking number in the form of Txxx.

It should be noted that depositors get immediate feedback at deposition time from a series of validation programs used by the PDB, which helps them to decide whether their data are actually suitable for deposition, or indeed for publication, or require further work. Once depositors give their approval *via* AutoDep, a PDB ID code is assigned immediately. **This PDB ID code should be the only PDB number listed in journal articles.**

Until this ID code is assigned, there is no deposition at the PDB.

The PDB 3D Browser (http://www.pdb.bnl.gov/pdb-bin/pdbmain) is updated daily and may be searched by PDB ID code, author name, etc. for proof of submission.

Local High School Students Visit the PDB

Nancy Manning

For the third year in a row, students from the Smithtown High School, Smithtown, New York, visited the Biology Department of Brookhaven National Laboratory. Eighteen DNA Science and Advanced Placement Biology students were accompanied by Smithtown biology teachers Harvey Goldstein and Karen Durand.

The March 27 visit opened with a tour of BNL's Scanning Transmission Electron Microscope (STEM) Facility and a presentation by Joseph Wall, Head of the STEM. Then Joel Sussman reviewed the history of the PDB and discussed the state of structural biology today. A visit to the Genome Sequencing Laboratory and talks by

Smithtown High School science students visiting the PDB.

scientists Jan Kieleczawa and John Dunn on DNA sequencing and the implications of the Human Genome Project rounded out the day. The PDB welcomes this opportunity to meet with students to introduce them to the cutting edge science being carried out at Brookhaven.

3DB Browser™: Tips, Questions and Answers

Jaime Prilusky

Bioinformatics Unit, Weizmann Institute of Science, Rehovot, Israel (lsprilus@weizmann.weizmann.ac.il)

This is a hands-on article, with tips, practical solutions and ideas on how to get the most out of the 3DB Browser™. If you have other tips, requests or solutions, we would like to hear from you.

In the following examples, we will omit the portion of the URL of the actual PDB site, leaving only the common portion. That is, instead of http://www.pdb.bnl.gov/pdb-bin/pdbids we will write.../pdb-bin/pdbids. Just complete the URL with the data for your favorite closerSite© from the list below.

(To have a true hands-on experience, we suggest that you now sit in front of your computer, start your Internet browser, and perform the operations as we describe them. REMEMBER to replace the.../pdb-bin)

Q: How do I link to a PDB entry from my web page?
A: <A HREF=".../pdb-bin/pdbids?id=2ACE">My Favorite Entry</A>

Q: Do you have a way of only searching for 'Pending and Waiting' entries?
A: Yes. .../pdb-bin/whsearch performs a fast search on the 'Pending and Waiting' entries, returning the answer sooner than pdbids.

The .../pdb-bin/whsearch script accepts boolean queries. The following are valid queries for 'Pending and Waiting' entries:

thymidine kinase

thymidine or kinase

thymidine and kinase

Q: Is there an updated list of URLs for the 3DB Browser's package?

A: Yes. This is the current list of URLs. This info is available from http://pdb.pdb.bnl.gov/Sites-bin.html

Argentina	http://pdb.unsl.edu.ar/pdb-bin/
Australia	http://pdb.wehi.edu.au:8181/pdb-bin/
Brazil	http://www.pdb.ufmg.br/pdb-bin/
China	http://www.ipc.pku.edu.cn/pdb-bin/
Germany	http://pdb.gmd.de/pdb-bin/
Israel	http://pdb.weizmann.ac.il/pdb-bin/
Poland	http://pdb.icm.edu.pl/pdb-bin/-bin/
Taiwan	http://pdb.life.nthu.edu.tw/pdb-bin/
UK_CCDC	http://pdb.ccdc.cam.ac.uk/pdb-bin/
UK_EBI	http://www2.ebi.ac.uk/pdb-bin/
USA_BNL	http://www.pdb.bnl.gov/pdb-bin/
USA_UGA	http://bcl10.bmb.uga.edu/pdb-bin/

Q: I would like to create a file with the sequences (in FASTA format) from a subset of entries in PDB. Is there a way of doing this?

A: I am glad you asked. Let's say that you would like to have all the sequences from the PDB entries that have the keyword 'thymidine'. Try the following URL in your browser:

.../pdb-bin/pdbids?kw=thymidine&m=fasta

This strange line tells pdbids to retrieve all PDB entries that have the keyword 'thymidine' (kw=thymidine), in FASTA format (m=fasta).

As you have guessed, we can request a subset of PDB entries based on any field available on 3DB Browser's main page:

.../pdb-bin/pdbids?au=brown&m=fasta (for author equal to 'Brown')

.../pdb-bin/pdbids?tx=snake&m=fasta (for 'snake' anywhere in the text)

With the 'm' parameter you may select other formats for the returned PDB entries selected by the query:

m=dump (a plain list of the PDB ids)

m=short (the short entry, traditional PDB format, no ATOMS)

m=full (the full entry, traditional PDB format)

m=fasta (the sequences in FASTA format from the selected PDB entries)

m=summary (the PDB Browser Atlas page for the selected PDB entries)

The possibilities are endless. Try the following URLs in your browser:

.../pdb-bin/pdbids?tx=snake&m=dump

Q: closerSite© tells me I can access two physically closer sites but they are greatly slower than going to the US.

A: This is correct. closerSite© tells you which mirror sites are geographically closer to you. We cannot, at this time, suggest the fastest site to connect to, due to the dynamic nature of the Internet itself.

Q: How do I know if the closerSite© PDB mirror site I am working with is being updated regularly?

A: Use the.../pdb-bin/info?date query. It will return the date when the site was latest updated. The current options for 'info' are

.../pdb-bin/info?ftp	URL ftp://
.../pdb-bin/info?date	site updated date
.../pdb-bin/info?ent	available number of entries

Q: My favorite database is not being linked from the 3DB Browser. Can you incorporate it?

A: Yes. Please send to us (lsprilus@weizmann.weizmann.ac.il) the URL (http, ftp) of the database and we will try to teach the 3DB Browser to connect to it.

Some WHAT_CHECK Checks Explained

Gert Vriend, Rob Hooft, EMBL

Meyerhofstrasse1, D-69117 Heidelberg, Germany
(Gert.Vriend@EMBL-Heidelberg.de)

Now that the program WHAT_CHECK (Hooft *et al.*, 1996) is routinely used in about a quarter of all X-ray labs, and now that all depositors as a service provided by the PDB (and the EBI) get a WHAT_CHECK report sent to them, it is time to explain the algorithms behind some of the checks so that the structure depositor can better interpret the WHAT_CHECK results. In future issues of the PDB Newsletter we will explain in detail a few of the WHAT_CHECK checks that have not yet been published.

While it has been a long time since the last real errors in WHAT_CHECK were found, suggestions for improvements are still welcome. Fortunately a slowly increasing number of depositors is giving feedback on the WHAT_CHECK reports. We intend to release the next version of WHAT_CHECK in the first half of 1999, so any depositor who wants to see certain improvements implemented should get into contact with us before the end of 1998. Most of the criticism we get is about the length of the report. We cannot help it that WHAT_CHECK does check so many different aspects of macromolecular structures. However, the people of the MSD unit at the EBI are working on a filter script that allows you to flexibly reduce the amount of output. The PDB provides both the entire WHAT_CHECK output and a filtered version within AutoDep, the Web-based submission tool.

One option that keeps provoking confused criticism is the cell-dimension check. How does this check work? Engh and Huber (1991) determined what the perfect bond lengths should be in macromolecules. They extracted this information from peptides and peptide-like structures in the Cambridge Structural Database (CSD) (Allen *et al.*, 1983). Most structures deposited in the CSD are considerably more accurate than most protein structures, and the ideal bond lengths determined by Engh and Huber can therefore, for all practical purposes, be called correct, although some enthusiasts are trying to use very accurately determined protein structures to verify and improve the Engh and Huber dataset.

The option in WHAT_CHECK that checks cell dimensions treats all bond lengths as vectors. The length of each vector is divided by the ideal length. In a far too perfect case all vectors would therefore lie on the unit sphere. In practice, however, there is some natural variation in the bond lengths and if these deviations are randomly distributed then the cell dimensions must be correct (or several errors are made that cancel out; a scenario which we consider highly unlikely).

If all vectors are on average a bit too long, the cell axes must be too long, which in turn most likely results from data processing with an assumed wavelength that is longer than the actual wavelength used in the data. If, for example, all vectors that lie nearly parallel to the A-axis are on average a bit too long but the vectors in other directions have on average length 1.0, then the A-axis must be a bit too long. This type of error is easily made if, for example, the cell dimension is determined from only one or two oscillation films.

An eigenvector analysis determines the principal components of the vector ensemble. A simple least squares procedure determines the cell deformation matrix that would best explain these principal components, and the cell that results from the multiplication of the cell dimensions with this deformation matrix is printed for further inspection by the depositor.

The confusion arises if the suggested better cell dimensions for a perfectly orthorhombic crystal include non-90-degree angles, or if the cell axes of a perfectly cubic cell are suggested to have different lengths. Several crystallographers have e-mailed us that there is a bug in WHAT_CHECK after they saw that the program told them that an angle which is supposed to be 90 degrees would be better if made 89.4 degrees, or something similarly illogical. However, this is not a bug as can easily been seen from the following little experiment.

Assume a cubic cell in which a four helix bundle lies parallel to the A-axis, and assume that there is a typo in the refinement force field because of which the backbone C=0 bond length becomes 1.331 Ångstrom (rather than the normal value of 1.231Å). Don't laugh about this, a well known NMR structure solution program was distributed for a short period with an ideal backbone N-H distance of about 3.0 Ångstrom... WHAT_CHECK uses only one asymmetric unit and not the full cell and thus will realize that all bonds along the A-axis are on average a bit too long and it will thus advise the depositor to make the A-axis a bit shorter. Since the resulting cell is now very far from the ideal cubic case, the reader of the report can notice that something "fishy" is going on. Had WHAT_CHECK used the symmetry of the unit cell, all three axes would have been a "just a bit" too long, but the symmetry would be perfectly conserved, such that the reader is not alerted that the check responds to an error in the force field, and not to a real error in the cell dimensions.

In another example the cell dimensions were suggested to be more than 2 Ångstroms wrong. Since this was not synchrotron data, the crystallographer was a bit surprised. We pushed him to find out what was wrong and he discovered that some wrong parameters were used in the program used to merge a low resolution and a high resolution dataset. So, even though the messages produced by the cell dimension check are not always telling the exact story, a warning by this option normally means that there is at least some kind of trouble.

In a next release of WHAT_CHECK we will, time allowing it, incorporate a few improvements in the cell dimension validation procedure. Systematic bond length errors for certain bond types can, for example, be removed by a regularisation procedure before the cell dimension check is invoked. We also still have to start with the analysis of the normality (normal distribution of the deviations) of individual bond length types and we could implement an iterative scheme that actually changes the cell dimensions while obeying the rules given by the space group. It is even imaginable that other WHAT_CHECK options will be executed twice, once using the deposited cell dimensions and once using the improved cell dimensions. Unfortunately the day has only 24 hours, and there is only so much that can be programmed. However, we always have an open ear for constructive comments from depositors.

So, if you want us to make changes in or additions to WHAT_CHECK, feel free to contact us (Vriend@EMBL-heidelberg.DE). It is likely

that good plans suggested by depositors end up high on the list of things to do. If you are interested in examples of cell dimension validation, you can look at the WHAT_CHECK report for your favorite PDB file in the PDBREPORT database at http://swift.embl-heidelberg.de/pdbreport/.

References:

Hooft, R.W.W., Vriend, G., Sander, C., & Abola, E.E. (1996). *Nature* **381**, 27

Engh, R., Huber, R. (1991) *Acta Cryst.* **A47**, 392-400.

Allen, F.H., Kennard, O., Taylor, R. (1983) *Acc. Chem. Res.* **16**, 146-153.

Status of the mmCIF Dictionary

Paula M. Fitzgerald, Helen M. Berman, John Westbrook, Phil Bourne, Keith Watenpaugh, & Brian McMahon

Parts of this article are taken from the mmCIF server (http://ndbserver.rutgers.edu/mmcif/).

The Crystallographic Information File (Hall, 1991) was created to archive information about crystallographic experiments and results (Hall *et al.*, 1991) and is now the format in which all structures submitted to Acta Crystallographic C are submitted. In 1990, the IUCr formed a working group to expand this dictionary so that it would be able to do the same for macromolecules. This working group was chaired by Paula Fitzgerald (Merck) and included Enrique Abola (PDB), Helen Berman (Rutgers), Phil Bourne (Columbia), Eleanor Dodson (York), Art Olson (Scripps), Wolfgang Steigemann (Max Planck), Lynn Ten Eyck (UCSD), and Keith Watenpaugh (Upjohn).

The original short term goal of the working group was to fulfill the mandate set by the IUCr: to define mmCIF data names that needed to be included in the CIF dictionary in order to adequately describe the macromolecular crystallographic experiment and its results. Long term goals were also determined: to provide sufficient data names so that the experimental section of a structure paper could be written automatically and to facilitate the development of tools so that computer programs could easily interface with CIF data files.

In order to describe the progress of this project and to solicit community feedback, several informal and formal meetings were held. The first meeting, hosted by Eleanor Dodson, convened in April 1993 at the University of York. The attendees included the mmCIF working group, structural biologists and computer scientists. A major focus of the discussion was whether the formal structure of the dictionary that was implemented using the then-current Dictionary Definition Language (DDL 1.0) was adequate to deal with the complexity of the macromolecular data items. Criticisms included the idea that the data typing was not strong enough and that there were no formal links among the data items. A working group was formed to try to address these issues. The second Workshop was hosted by Phil Bourne in Tarrytown, NY in October 1993. The topics at that meeting focused on the development of software tools and the requirements of an enhanced DDL. In October 1994, a workshop hosted by Shoshana Wodak at the Free University of Brussels resulted in the adoption of a new DDL that addressed the various problems that had been identified at the preceding workshops. The dictionary was cast in this new DDL 2 and was presented at the ACA meeting in Montreal in July 1995. This dictionary was open for further community review. The dictionary was placed on a World Wide Web site and community comments were solicited *via* a list server. Lively discussions via this mmCIF list server ensued, resulting in the continuous correction and updating of the dictionary. Software was developed and was also presented on this WWW site.

In January 1997, the mmCIF dictionary was completed and submitted to COMCIFS for review and in June 1997, Version 1.0 was released (Fitzgerald *et al.*, Bourne *et al.*, 1997). A workshop was held at Rutgers University in October 1997, hosted by Helen Berman. Tutorials were presented to demonstrate the use of the various tools that had been developed. There was much discussion about how to proceed with the maintenance and evolution of the dictionary so that it can accommodate new data items and still be compatible with existing software. The method adopted for managing these extensions uses a scientific journal as a model. The proposed extensions are sent to the Editors of the mmCIF Dictionary (Paula Fitzgerald, Editor, Helen Berman, Associate Editor) who send the new definitions to a member of the board of editors for scientific review. These editors have expertise in the various areas covered by the dictionary; they are Phil Bourne, Dale Tronrud, Andy Howard, Joel Sussman, and Frank Allen. Once the definitions are reviewed for their scientific content, they are sent to the Technical Editors, John Westbrook and Herbert Bernstein.

More than 100 new definitions have been proposed since the fall of 1997 and have been reviewed using the procedures outlined. Version 2 of the mmCIF dictionary will contain many of these new definitions and is expected to be released the summer of 1998.

References:

Hall, S.R. (1991). The STAR File: A new format for electronic data transfer and archiving. *J. Chem. Inf. Comput. Sci.* **31**, 326-331.

Hall, S.R., Allen, F.H., and Brown, I.D. (1991). A new standard archive file for crystallography. *Acta Cryst,* **A47**, 655-685.

Fitzgerald, P.M.D., Berman, H.M., Bourne, P.E., McMahon, B., Watenpaugh, K., and Westbrook, J. (1996). "The mmCIF dictionary: community review and final approval", IUCr Congress and General Assembly, August 8-17, *Acta Cryst.*, **A52** Supplement. Seattle, WA. MSWK.CF.06

Bourne, P., Berman, H.M., Watenpaugh, K., Westbrook, J.D., and Fitzgerald, P.M.D. (1997). The macromolecular Crystallographic Information File (mmCIF). *Meth. Enzymol.* **277**, 571-590.

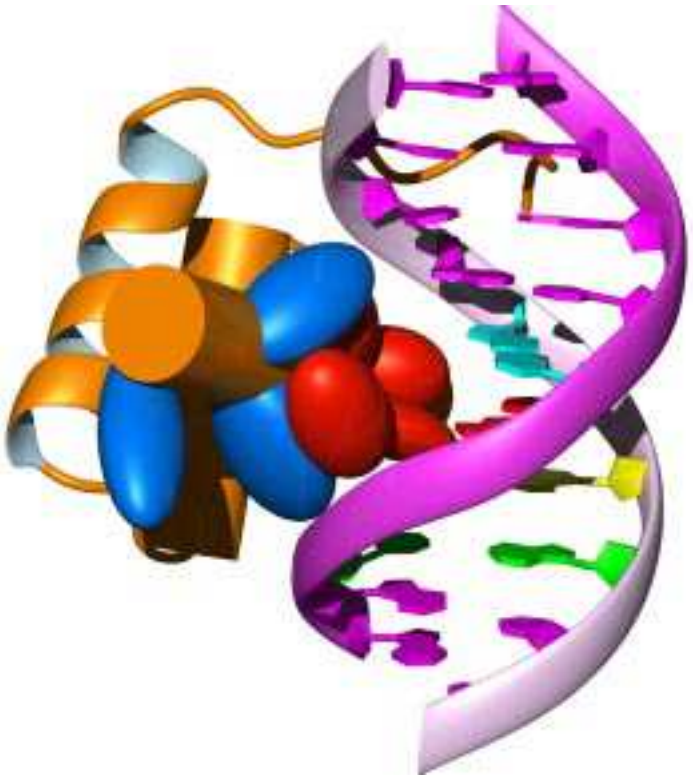
MOLMOL: A Program for Display and Analysis of Macromolecular Structures

Reto Koradi and Martin Billeter

Institute for Molecular Biology and Biophysics, ETH Zurich, Switzerland

MOLMOL (Koradi *et al.*, 1996) is a molecular graphics program for display, analysis, and manipulation of three-dimensional structures of biological macromolecules, with special emphasis on nuclear magnetic resonance (NMR) structures of proteins and nucleic acids. It can be used for evaluating and comparing structures as well as for the generation of high-quality pictures for documentation or publication. The program runs under UNIX and Windows NT/95 and is freely available. It has a fully graphical user interface and supports numerous plot formats.

MOLMOL supports all standard display possibilities. Using the feature that atoms can be displayed as spheres and bonds as cylinders of arbitrary sizes, space-filling (CPK) and ball-and-stick representations can easily be generated. These displays can readily be combined with more advanced display features. Independent choice



NMR structure of an Antennapedia homeodomain-DNA complex (Billeter, et al., 1993) is drawn using cylinders and ribbons for helices, tubes for loops, and ellipsoids for those amino acid side chains that play an important role in contacting the DNA. The DNA is represented by ribbons for the backbone and plates for the sugar rings and the bases.

of display styles for different parts of a molecular structure is made possible by a selection mechanism with a powerful expression syntax. Using this selection principle, the user can also set display attributes such as color, shininess, opacity, etc. for arbitrary subsets of items.

A major focus of MOLMOL is on advanced display possibilities for effective visualization of complex protein structures. One option is schematic drawings of regular secondary structure as ribbons (Richardson, 1981). The program can identify the secondary structure elements automatically (Kabsch and Sander. 1983), but they may also be read from a PDB file or entered by the user. The algorithms for building the ribbons do not assume a specific geometry, such as a certain radius for helices, so they can also be used for other molecules, like DNA double helices.

Beyond representation of secondary structure, any subset of atoms, *e.g.,* an amino acid side chain, a structural unit, or an entire protein domain, can be schematically represented by geometric shapes such as spheres, ellipsoids, cylinders or rectangular boxes. These solids are optimized so that they contain a predetermined percentage of the user-selected atoms, while their volume is minimal. Rings, such as the ones of DNA bases, can also be drawn as solid plates. The dipolar moment of a set of atoms can be indicated by an arrow.

MOLMOL can determine and display various kinds of surfaces, the most popular being contact surfaces (Connolly, 1983). They can be coloured based on the local electrostatic potential, an algorithm for solving the Poisson-Bolzmann equation (Nicholls and Honig, 1990) is built into the program. Parts of surfaces can be cut off, so that they can effectively be combined with other display possibilities.

Very important for publication of quality figures are text labels. MOLMOL allows the interactive definition and manipulation of labels with super-/subscript and Greek letters. They are placed at the proper depth in stereo pictures.

Calculation and analysis possibilities include:

- superpositions and RMSD calculations (for structure comparisons and bundles of conformers from NMR structure calculations)
- hydrogen bonds (tables and/or drawn in structure)
- short distances (tables and/or drawn in structure)
- violations of NMR constraints
- coordinates of missing atoms (*e.g.,* hydrogens in X-ray structures)
- solvent accessible surface
- angles between helix axes

MOLMOL can also draw various kinds of figures for structure analysis, such as Ramachandran plots, contact maps, or graphs that show the distribution of dihedral angles versus the sequence. Some of these plots are especially useful when many structures need to be analyzed, like the result of NMR structure calculations or a molecular dynamics simulation.

While the main focus of MOLMOL is on structure display and analysis, it also contains some possibilities for structure manipulations. Atoms and bonds can be added or removed. Residues can be exchanged, or new sequences built from scratch. Dihedral angles can be rotated interactively. There are also many possibilities for handling distances and constraints obtained from NMR measurements.

The graphical user interface of MOLMOL consists of menus, buttons and dialog boxes. Online help for all approximately 230 commands is available; it can be displayed in text windows or in an external web browser. Experienced users can also type commands on a command line. Menus and buttons can be configured by modifying simple text files, and additional commands can be defined by macros. There is an undo possibility for all commands. The program can read and write coordinates of structures in various common formats, such as PDB.

The program supports various graphics libraries for screen display, the most important one being OpenGL. Various plot formats are supported. Raster files at arbitrary resolution can be saved in TIFF, PNG, JPEG (UNIX) or BMP (Windows) format. PostScript output yields files containing geometric primitives that will make use of the full resolution of the output device. In addition, the program can also produce input files for the public domain ray-tracing program POV-Ray. These files yield high quality figures, with effects such as shadows, reflections, transparency, and texture mapping.

MOLMOL was developed as a joint effort between BRUKER/Spectrospin and the group of Prof. K. Wüthrich at the Institute for Molecular Biology and Biophysics at the ETH Zurich. Further infor-

mation about MOLMOL, including instructions for downloading the source code or executables for various platforms, can be found on the web page http://www.mol.biol.ethz.ch/wuthrich/software/molmol.

References

Koradi, R., Billeter, M. & Wüthrich, K. (1996). *J. Mol. Graphics*, **14**, 51-55.

Richardson, J. S. (1981). *Adv. Protein Chem.*, **34**, 167-339.

Kabsch, W. & Sander, C. (1983). *Biopolymers*, **22**, 2577-2637.

Connolly, M. L. (1983). *J. Appl. Cryst.*, **16**, 548-558.

Nicholls, A. & Honig, B. (1990). *J. Comp. Chem.*, **12**, 435-445.

Billeter, M., Qian, Y.Q., Otting, G., Müller, M., Gehring, J. and Wüthrich, K. 1993). *J.Mol. Biol.* **234,** 1084-1093.

Notes of a Protein Crystallographer — Molecular Docking and the Broken Heart

Cele Abad-Zapatero

Abbott Laboratories, Department of Structural Biology, Abbott Park, IL, USA (abad@abbott.com)

Paul Ehrlich (1854-1915), widely recognized as the founder of modern medicinal chemistry, coined the word “chemotherapy” in 1891 to refer to the curative effect of certain man-made chemical entities. In his view, critical to the success in synthesizing these compounds were the three G’s: Geduld, Geld, Glück (Patience, Money and Luck). After 606 exhausting trials Ehrlich developed an arsenic compound (‘salvarsan’) in 1910 which allowed the effective treatment of syphilis (Mahoney, 1959). This was the first man-made compound that cured an infectious disease in man. Since then the pharmaceutical industry has continued to discover, synthesize, test, develop and market novel ‘drugs’ with a tremendous impact in our societal and individual well-being.

The Food and Drug Administration (FDA) was created in 1938 to enforce the Federal Food, Drug and Cosmetic Act which put tighter restraints on industry practices. The international tragedies resulting from the use of thalidomide in Europe led to the 1962 amendment of the legislature and to drastic restrictions in clinical investigations, and to stringent requirements in any experiment involving human subjects. Consequently, finding truly novel, safe, pharmaceutical entities began to slow down. The cost of putting a new and effective drug in the market had risen to figures between 100 and 200 million dollars, with an elapsed time ranging from 10 to 12 years of devoted research effort distributed between discovery, development and clinical testing. In an atmosphere of diminishing returns, the pharmaceutical industry found itself running low in the three Gs. At the same time during the fifties and sixties, there was a dramatic increase in our knowledge of the enzymatic reactions occurring in the living organisms. This pool of biochemical information added a new dimension to the understanding of the relations between the chemical reactions necessary to sustain life, the enzyme catalysts participating in these reactions, and the small chemical entities that participated and altered these reactions.

It is a tribute to the internal dynamic and ingenuity of the American society that our technical — and somebody might say esoteric — expertise soon found a niche inside the walls of corporate America to aid in the process of inventing and designing pharmaceutical drugs. The premise is self-evident: most biological processes are controlled by protein molecules through their enzymatic or regulatory activity. If you can inhibit those enzymatic processes or if you can modulate their actions, then you can intervene with biological processes of therapeutic value.

During the early 80’s a target of choice was renin, an aspartic proteinase which is part of the cascade regulating blood pressure. A myriad of compounds were synthesized to block the activity of renin and several of those compounds were crystallized and analyzed as complexes with renin, or with the archetypal aspartic proteinase pepsin. In spite of the difficulties of obtaining good quality renin samples to perform the crystallographic studies, the renin target validated the structural approach and the concept of ‘structure-aided drug design’ was a considerable driving force behind the practical applications of protein crystallography.

Strikingly, the epidemic of the late 80’s featured a malignant virus (HIV) that to complete its life cycle needed another aspartic protease known as HIV-protease. This molecular target was a much smaller dimeric aspartic proteinase protein that could be produced in large quantities by DNA recombinant methods. It can be crystallized either by itself or in complex with a multitude of ‘designed compounds’ and dedicated efforts resulted in well-diffracting crystals in laboratories all over the world.

Structures of probably over 300 HIV-protease:inhibitor complexes have been analyzed around the world. I am proud to say that several years ago the compound that turned out to be ritonavir (NORVIR™) went through our laboratory and my colleague Dr. Chang Park provided valuable crystallographic data (PDB entry 1HXW) to the project that designed this successful drug against the AIDS epidemic (Kempf *et al.*, 1995). Initially, a series of inhibitors were designed based on the dimeric structure of the HIV protease (Erickson *et al.*, 1990; 9HVP). This breakthrough was followed by extensive optimization of the activity, physico-chemical properties, metabolic and pharmacological behavior and eventually led to the identification of ritonavir. The introduction of HIV protease inhibitors has resulted in the development of potent combination protocols that succeed in reducing the amount of virus in the blood of AIDS patients to undetectable levels (Kempf *et al.*, 1995).

In spite of this dramatic success, drug design is a very complex endeavor where the parameters and variables to be optimized are related to the molecular interactions between the ‘drug’ and its ‘target’, its toxicity, oral bioavailability, and also to the metabolic degradation or average ‘life time’ of the active compound in the blood of the patient. Well before the three-dimensional structure of the target enzymes was available, quantitative concepts had been defined to address these properties at the molecular level and attempts to correlate these molecular properties with their *in vivo* activity were referred to as Quantitative Structure Activity Relationships (QSAR). Figures for critical parameters such as inhibition constants in the appropriate *in vitro* and *in vivo* assays (Ki, IC50, MIC90, EC50), partition coefficients between n-butanol and water (LogP, CLogP) and others, fill innumerable cells of immense spreadsheets in the minds and personal computers of scientists involved in drug discovery. These quantities are the beacons necessary to successfully navigate the tempestuous waters of any drug-design effort. Crystal structures of the target macromolecules and of their complexes with active molecules provide a very important piece of the puzzle but not the only one. Although drug-design is a team effort *par excellence*, compounds are synthesized by chemists and assayed by biologists; both efforts are of pivotal importance to the final outcome.

Even knowing the three-dimensional structure of receptor:inhibitor complexes, still inadequate are theoretical calculations to obtain reliable numbers for some (*e.g.* Ki) of the macroscopic quantities

defined by the above concepts. Indeed, others (*e.g.* MIC90) cannot be calculated within the structural framework. Amazingly, the simplest component of all, water is the most difficult to incorporate successfully in our calculations. Currently beyond our reach are issues such as absorption by the different tissue barriers, bioavailability, pharmacokinetics and our methods are currently unable to tackle these issues at the molecular level. Super inhibitors (picomolar or better), that we can discern with exquisite detail in our enzyme:inhibitor crystals, fail tomorrow as drug candidates because of permeability, accessibility problems or because of various associated toxicities.

Further challenges lie ahead with the advent of the gargantuan amounts of information provided by the genomics revolution. We are going to find more and more often that strings of characters *in silico* do not reveal themselves as enzymes with an obvious activity *in vivo*. Many of the ones that do will turn out to have subtle and pleiotropic effects in tissues and organs, complicating the issues of target validation, definition, selectivity, and suitability. These unknowns will drive sophisticated biological, genetic and biochemical experiments well into the 21st century.

In addition, the 'hyper-rational' dream of designing chemical entities and predicting their activities *ab initio* will attempt to exploit the sheer computational power of computers such as the 32-node supercomputer (IBM RS/6000*SP) that executed 'Deeper Blue', the landmark computer program that crushed Garry Kasparov in the momentous man-machine chess tournament last year. Unfortunately, the rules that govern the energy (ΔG), enthalpy (ΔH) and entropy ($T\Delta S$) terms of the interactions between the molecules of our dreams and their putative targets are not as well defined as the chess moves. In spite of our ingenious genetic algorithms and our docking searches, the 'value functions' are still difficult to quantify and to translate into positive gains.

I cannot resist finishing these lines about structure-aided drug design with a quotation from the novel 'Written on the Body' by the British writer Jeanette Winterson (born 1959).

"Molecular docking is a serious challenge for bio-chemists. There are many ways to fit molecules together but only a few juxtapositions that bring them close enough to bond. On a molecular level success may mean discovering what synthetic structure, what chemical, will form a union with, say, the protein shape on a tumor cell. If you make this high-risk jigsaw work you may have found a cure for carcinoma. But molecules and the human beings they are part of exist in a universe of possibility. We touch one another, bond and break, drift away on force-fields we don't understand" [...].

The soliloquy comes from the brain of the main character in the novel, as she debates whether staying close to her lover will heal her broken heart or could, in fact, result in an expensively ruinous experiment. I will close these reflections in a similar vein.

We, humans, have made tremendous scientific and technological progress. Some members of our species have walked on the moon, cloned mammals, cleaved an atomic nucleus in half and discovered new planets, galaxies, black holes and other astronomical objects. After 300 years, even Fermat's last theorem has been proven by an inquisitive and determined member of our species. Others have mapped the structure of the common cold virus in atomic detail. Aided by the knowledge of the biological machinery of the AIDS virus at the atomic level, we have been able to design effective drugs to fit molecules of this virulent pathogen and we are treating now what were considered untreatable diseases only a few years ago. Our data base of structural knowledge has expanded enormously and will continue to do so allowing us to make new strides in the constant struggle against aging, disease, pain and deformity.

However, the force fields that govern our interpersonal relationships are beyond our control. The gradients controlling our passions, our desires and our affections are not amenable to our analytical tools. Far away from our rational understanding are the tribal winds that carry hatred, prejudice, bigotry, injustice, and war. The human hearts expand and shrink, thrive or suffer subject to force fields very different from the ones controlling the molecular docking of an inhibitor to its target or a substrate to its enzyme.

References

Mahoney, T. (1959). *The Merchants of Life. An Account of the American Pharmaceutical Industry*. Harper & Brothers, New York.

Kempf, D., Marsh, K.C., Denissen, J.F., McDonald, E., Vasavanonda, S., Flentge, C.A., Green, B.E., Fino, L., Park, C.H., Kong, X-P, Wideburg, N.W., Saldivar, A., Ruiz, L., Kati, W.M., Sham, H.L., Robbins, T., Stewart, K.D., Hsu, A., Plattner, J., Leonard, J.M., Norbeck, D.W. (1995). *Proc. Natl. Acad. Sci.*, **92**, 2484-2488.

Erickson, J., Neidhart, D.J., VanDrie, J., Kempf, D.J., Wang, X.C., Norbeck, D.W., Plattner, J.J., Rittenhouse, J.W., Turon, M., Wideburg, N., Kohlbrenner, W.E., Simmer, R., Helfrich, R., Paul, D.A., Knigge, M. (1990). *Science.* **249**, 527-533.

Web Sites

Referenced in this Newsletter

3DB Browser™
http://www.pdb.bnl.gov/pdb-bin/pdbmain

mmCIF
http://ndbserver.rutgers.edu/mmcif/

MOLMOL: MOLecule analysis and MOLecule display
http://www.mol.biol.ethz.ch/wuthrich/software/molmol/

Nature Structure Biology Survey Results
http://us.nature.com/survey/nsb_poll.nclk

PDB Het Group Dictionary
(ftp://pdb.pdb.bnl.gov/pub/resources/hetgroups/het_dictionary.txt)

PDB Mirror Sites
http://www.pdb.bnl.gov/pdb-docs/mirror_sites.html

PDB Quarterly Newsletter
http://www.pdb.bnl.gov/pdb-docs/newsletter.html

The PDB Report Database
http://swift.embl-heidelberg.de/pdbreport/

Related WWW Sites

Databases

Archive of Obsolete PDB Entries	http://pdbobs.sdsc.edu/
BMRB (BioMagResBank)	http://www.bmrb.wisc.edu
CCDC (Cambridge Crystallographic Data Centre)	http://www.ccdc.cam.ac.uk
EBI (European Bioinformatics Institute)	http://www.ebi.ac.uk
EMBL (European Molecular Biology Laboratory)	http://www.embl-heidelberg.de
ExPASy Molecular Biology Server	http://www.expasy.ch
GDB (Genome Data Base)	http://gdbwww.gdb.org
GenBank (NIH Genetic Sequence Database)	http://www.ncbi.nlm.nih.gov/Web/Genbank/index.html
HIC-Up (Hetero-compound Information Centre Uppsala)	http://alpha2.bmc.uu.se/hicup/
HIV Protease Database	http://www-fbsc.ncifcrf.gov/HIVdb/
Klotho: Biochemical Compounds Declarative Database	http://www.ibc.wustl.edu/klotho/
Library of Protein Family Cores	http://WWW-SMI.Stanford.EDU/projects/helix/LPFC/
Crystal MacroMolecule Files at EBI	http://www2.ebi.ac.uk/msd/macmol_doc.shtml
NCBI (National Center for Biotechnology Information)	http://www.ncbi.nlm.nih.gov
NDB (Nucleic Acid Database)	http://ndbserver.rutgers.edu
PDB (Protein Data Bank)	http://www.pdb.bnl.gov
PIR (Protein Information Resource)	http://www-nbrf.georgetown.edu/pir
Prolysis: A Protease and Protease Inhibitor Web Server	http://delphi.phys.univ-tours.fr/Prolysis/
Protein Kinase Database Project	http://www.sdsc.edu/kinases/
Protein Motions Database	http://bioinfo.mbb.yale.edu/MolMovDB/
RELIBase	http://pdb.pdb.bnl.gov:8081/home.html
SCOP: Structural Classification of Proteins	http://scop.mrc-lmb.cam.ac.uk/scop/
Mirrored at Protein Data Bank	http://www.pdb.bnl.gov/scop/
Swiss-Prot Sequence Database	http://expasy.hcuge.ch/sprot/sprot-top.html
CATH Protein Structure Classification	http://www.biochem.ucl.ac.uk/bsm/cath
Enzyme Structures Database	http://www.biochem.ucl.ac.uk/bsm/enzymes/
PDBsum	http://www.biochem.ucl.ac.uk/bsm/pdbsum

Software-Related Sites

CCP4	http://www.dl.ac.uk/CCP/CCP4/main.html ftp://ccp4a.dl.ac.uk/pub/ccp4
mmCIF	http://ndbserver.rutgers.edu/NDB/mmcif
O Home Page	http://imsb.au.dk/~mok/o/
OPM (Object-Protocol Model) Data Management Tools	http://gizmo.lbl.gov/DM_TOOLS/OPM/OPM.html
RasMol Home Page	http://www.umass.edu/microbio/rasmol/
SHELX Home Page	http://linux.uni-ac.gwdg.de/SHELX
Squid: Analysis and Display of Data from Crystallography and Molecular Dynamics	http://www.yorvic.york.ac.uk/~oldfield/squid/
VMD - Visual Molecular Dynamics	http://www.ks.uiuc.edu/Research/vmd/
X-PLOR Home Page	http://xplor.csb.yale.edu/

Other Resources

Crystallography Worldwide	http://www.unige.ch/crystal/w3vlc/crystal.index.html
BioMoo	http://www.cco.caltech.edu/~mercer/htmls/BioMOOHomePage.html
DALI - Comparison of Protein Structures in 3D	http://www.embl-heidelberg.de/dali/dali.html
NCSA Biology Workbench	http://biology.ncsa.uiuc.edu/
MOOSE (Macromolecular Structure Database at San Diego Supercomputer Center)	http://db2.sdsc.edu/moose
PDB_select: Representative PDBStructures	ftp://ftp.embl-heidelberg.de/pub/databases/protein_extras/pdb_select/recent.pdb_select
PROCHECK - To Submit a PDB File for Analysis	http://www.cryst.bbk.ac.uk/PPS/procheck/test.html
Protein Structure Verification-Biotech Server	http://biotech.embl-heidelberg.de:8400/
Mirrored at Protein Data Bank	http://biotech.pdb.bnl.gov:8400/
Resources for Macromolecular Structure Information	http://www.ucmb.ulb.ac.be/StructResources.html
The Virtual School of Molecular Sciences	http://www.vsms.nottingham.ac.uk/vsms/
Weizmann Institute, Genome and Bioinformatics	http://bioinfo.weizmann.ac.il/

Affiliated Centers and Mirror Sites

Forty-two 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. Official PDB Mirror Sites are marked with a grey bar ( ) and are listed with their sponsoring center.

ARGENTINA

UNIVERSIDAD NACIONAL DE SAN LUIS
Facultad de Ciencias Fisico Matematicas y Naturales
Universidad Nacional de San Luis
San Luis, Argentina
Jorge A. Vila (54-652-22803)
vila@unsl.edu.ar
http://linux0.unsl.edu.ar/fmn

PDB Mirror Site:
http://pdb.unsl.edu.ar
Fernando Aversa (aversa@unsl.edu.ar)

AUSTRALIA

ANGIS
The Australian National Genomic Information Service
University of Sydney
Sydney, Australia
Shoba Ranganathan (61-2-9351-3921)
shoba@angis.org.au
http://www.angis.org.au

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

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

BRAZIL

UNIVERSIDADE FEDERAL DE MINAS GERAIS
Instituto de Ciencias Biologicas
Belo Horizonte, MG - Brazil
Marcelo M. Santoro (55-31-441-5611)
santoro@icb.ufmg.br
Ari M. Siqueira (55-31-952-7470)
siqueira@icb.ufmg.br
http://www.1cc.ufmg.br/

PDB Mirror Site:
http://www.pdb.ufmg.br
Ari M. Siqueira (siqueira@cenapad.ufmg.br)

CANADA

NATIONAL RESEARCH COUNCIL OF CANADA
Institute for Marine Biosciences
Halifax, N.S., Canada
Christoph W. Sensen (902-426-7310)
sensencw@niji.imb.nrc.ca
http://cbrmain.cbr.nrc.ca

CHINA

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)

FINLAND

CSC
CSC Scientific Computing Ltd.
Espoo, Finland
Erja Heikkinen (358-9-457-2433)
erja.heikkinen@csc.fi
http://www.csc.fi

TURKU CENTRE FOR BIOTECHNOLOGY
University of Turku and Abo Akademi University
Turku, Finland
Adrian Goldman (358-2-3338029)
goldman@btk.utu.fi
http://www.btk.utu.fi

FRANCE

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

LIGM
Laboratorie d'ImmunoGenetique Moleculaire
Montpellier, France
Marie-Paule LeFranc (33-04-67-61-36-34)
Lefranc@ligm.crbm.cnrs-mop.fr
http://imgt.cnusc.fr:8104

GERMANY

DKFZ
German Cancer Research Center
Heidelberg, Germany
Otto Ritter (49-6221-42-2372)
o.ritter@dkfz-heidelberg.de
http://www.dkfz-heidelberg.de

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

GMD
German National Research Center for Information Technology
Sankt Augustin,Germany
Theo Mevissen (49-2241-14-2784)
theo.mevissen@gmd.de
http://www.gmd.de

PDB Mirror Site:
http://pdb.gmd.de
Theo Mevissen (theo.mevissen@gmd.de)

MPI
Max Planck Institute for Biochemie Computer Center
Martinsried, Germany
Wolfgang Steigemann (49-89-8578-2723)
steigemann@biochem.mpg.de
http://www.biochem.mpg.de

INDIA

PUNE
Bioinformatics Center University of Pune
Pune, India
A. S. Kolaskar (0212-355039-350195)
Kolaskar@bioinfo.ernet.in
http://bioinfo.ernet.in

ISRAEL

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

PDB Mirror Site:
http://pdb.weizmann.ac.il
Marilyn Safran
(pdbhelp@pdb.weizmann.ac.il)

ITALY

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

JAPAN

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

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

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

THE NETHERLANDS

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

POLAND

WARSAW UNIVERSITY
Interdisciplinary Centre for Modelling
Warszawa, Poland
Wojtek Sylwestrzak (48-22-874-9100)
W.Sylwestrzak@icm.edu.pl
http://www.icm.edu.pl

PDB Mirror Site:
http://pdb.icm.edu.pl
Wojtek Sylwestrzak
(W.Sylwestrzak@icm.edu.pl)

SWEDEN

UPPSALA UNIVERSITY
Department of Molecular Biology
Uppsala University
Uppsala, Sweden
Alwyn Jones (46-18-174982)
alwyn@xray.bmc.uu.se
http://pdb.bmc.uu.se or http://alpha2.bmc.uu.se

TAIWAN

NATIONAL TSING HUA UNIVERSITY
Department of Life Science
HsinChu City, Taiwan
J.-K. Hwang (+886 3-5715131, extension 3481)
or lshjk@life.nthu.edu.tw
P.C. Lyu (+886 3-5715131 extension 3490)
lslpc@life.nthu.edu.tw http://life.nthu.edu.tw

PDB Mirror Site:
http://pdb.life.nthu.edu.tw/
Tony Wu (mirror@life.nthu.edu.tw)

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

UNITED KINGDOM

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

*CCDC
Cambridge Crystallographic Data Centre
Cambridge, United Kingdom
David Watson (44-1223-336394)
watson@ccdc.cam.ac.uk
http://www.ccdc.cam.ac.uk

PDB Mirror Site:
http://pdb.ccdc.cam.ac.uk/
Ian Bruno (mirror@ccdc.cam.ac.uk)

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)

*OML
Oxford Molecular Ltd.
Oxford, United Kingdom
Kevin Woods (44-1865-784600)
kwoods@oxmol.co.uk
http://www.oxmol.co.uk or http://www.oxmol.com

UNITED STATES

*APPLIED THERMODYNAMICS, LLC
Hunt Valley, Maryland, USA
George Privalov (410-771-1626)
George_Privalov@classic.msn.com
http://www.mole3d.com

BMRB
BioMagResBank
University of Wisconsin - Madison
Madison, Wisconsin, USA
Eldon L. Ulrich (608-265-5741)
elu@bmrb.wisc.edu
http://www.bmrb.wisc.edu

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

CMU
Carnegie Mellon/Pittsburgh Supercomputing
Center
Pittsburgh, Pennsylvania, USA
Hugh Nicholas (412-268-4960)
nicholas@psc.edu
http://pscinfo.psc.edu/biomed/biomed.html

*MAG
Molecular Applications Group
Palo Alto, California, USA
Margaret Radebold (650-846-3575)
bold@mag.com
http://www.mag.com

*MSI
Molecular Simulations Inc.
San Diego, California, USA
Stephen Sharp (619-799-5353)
ssharp@msi.com
http://www.msi.com

NCBI
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

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

*PANGEA SYSTEMS, INC.
Oakland, CA 94612
Greg Thayer (510-628-0100)
gregt@pangeasystems.com
http://www.pangea.com

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

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

PDB Mirror Site:
http://BCL10.bmb.uga.edu
John Rose (rose@BCL4.biochem.uga.edu)

This page intentionally left blank.

PDB ™ Order Form

Name of User _____ Date _____

Organization _____ Phone _____

Address _____ Fax _____

_____ E-mail _____

- Price is valid through September 30, 1998
- 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; and
3. Payment (using one of the two options below):

- Check payable to Brookhaven National Laboratory in U.S. dollars and drawn on a U.S. 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: Chase Manhattan Bank
Account name: Brookhaven Science Associates, LLC
Brookhaven National Laboratory
Account number: 615-775942

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

PDB™ Orders
Biology Department, Building 463
Brookhaven National Laboratory
P.O. Box 5000
Upton, NY 11973-5000

One (1) release of the PDB™ on CD-ROM — ISO 9660 Format *$362.45*
Total for four (4) releases *$1449.80*
(tax and shipping charges not applicable)

This page intentionally left blank.

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 ..... 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**

FTP Directory Structure for Entries

The PDB FTP server is updated weekly. Files are available by anonymous ftp to ftp.pdb.bnl.gov.

The PDB releases entries during the early morning hours each Wednesday, New York time. Entries that have been placed on hold by their authors are made available on the first Wednesday following their hold expiration date.

Entry files are found under the directory **pub/pdb/**

all_entries/
coordinate entry files in compressed and uncompressed format

biological_units/
generated coordinates for the biomolecules

current_release/
current database, with entries removed or added since the last CD-ROM

fullrelease/
static copy of the database as found on the last CD-ROM

latest_update/
entries added or removed in the most recent FTP update

newly_released/
entries released since the last CD-ROM

nmr_restraints/
compressed NMR restraint files

obsolete_entries/
withdrawn and/or replaced entries

structure_factors/
compressed structure factor files

fullrelease, newly_released, and current_release are divided into multiple subdirectories.

Scientific Consultants

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

Mia Raves, *Utrecht University, The Netherlands*

Clifford Felder
Kurt Giles
Jaime Prilusky
Marilyn Safran
Vladimir Sobolev
Yehudit Weisinger
Weizmann Institute of Science
Rehovot, Israel

PDB Staff

Joel L. Sussman, *Head*
Enrique E. Abola, *Deputy Head and Head of Scientific Content/Archive Management*
Otto Ritter, *Head of Informatics*

Frances C. Bernstein
Betty R. Deroski
Arthur Forman
Sabrina Hargrove
Jiansheng Jiang
Mariya Kobiashvili
Jiri Koutnik
Patricia A. Langdon
Michael D. Libeson
Dawei Lin
Nancy O. Manning
John E. McCarthy
Christine Metz
Michael J. Miley
Regina K. Shea
Janet L. Sikora
S. Swaminathan
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-AC02-98CH10886) and user fees.

Instructions to Authors

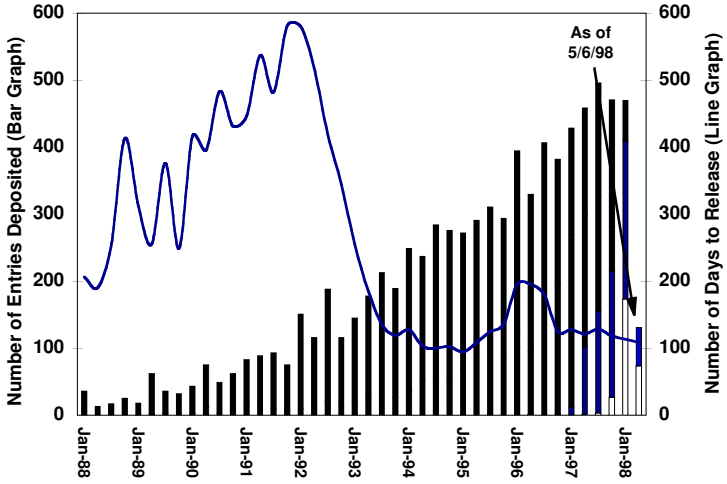
Contributions to the PDB Quarterly Newsletter may be sent
by e-mail or diskette to:
Nancy O. Manning, *Editor*
oeder@bnl.gov
References should be in the format used
by the Journal of Molecular Biology.
Deadlines for contributions are:
March 1, June 1, September 1, and December 1.

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

First Class

Number of Entries Deposited (Bar) and Average Time to Release (Line)

Accumulated and Averaged on a Quarterly Basis



Bar Graph - Number of Entries in the Following Categories:

OnHold	*- (light blue) On-hold per depositor request*
Processing	*- (white) Being processed*
Released	*- (black) Released*

Line Graph - Average Number of Days to Release
The data were accumulated and averaged on a quarterly basis. The average turn-around times for entries now being processed are estimated based on the average of the last 12 months.

Data for the last quarter are accumulated until the date specified on the graph.

See http://www.pdb.bnl.gov/pdb-docs/EntryTurnAround.html for regularly updated plot.

PRINTED ON RECYCLED PAPER