Protein Data Bank

Quarterly Newsletter

Release #72 April 1995

April 1995 PDB Release

3451 full-release atomic coordinate entries (371 new additions)

3200 proteins, enzymes, and viruses
241 nucleic acids
10 carbohydrates

365 structure factor entries
31 NMR experimental entries

The total size of the atomic coordinate entry database is 1175 Mbytes uncompressed.

Table of Contents

page

The latest version of the Electronic Deposition Form should be obtained from the FTP /**pub** directory before depositing data.

What is New at the PDB

– Creating Your Own Mirror Copy of the PDB's WWW Home Page

In our October 1994 and January 1995 Newsletters we described a mirror facility that makes it possible for remote sites with a licensed copy of our database to keep their PDB entry collections up-to-date automatically over the Internet. This was extended to include the indices for the tcl/tk-based PDB-browser, so remote sites that installed mirroring could be sure of searching the up-to-date collection at all times. This was especially important, since earlier last year the PDB started updating its FTP server with fully-released entries every two to three weeks, instead of every three months as had been done previously.

We have released a WWW/Mosaic front-end for our browser as part of our WWW home page. For the first time, this has made it possible for anyone with a workstation or microcomputer connected directly to the Internet, and with a suitable WWW browser such as NCSA Mosaic or NetScape (or even Lynx for those using non-windowing terminals such as a VT100), to directly access and search our entire collection. However, having only a single site to serve the entire Internet community often meant poor service for many of our distant users.

To remedy this situation, we have now made a version of our WWW browser available which can be downloaded and installed locally by remote sites that already mirror the PDB. In this way, users can now set up mirror sites for our PDB WWW home page and browser. This should help to alleviate the network congestion associated with requiring everyone to use a single site. This version of the browser is in the PDB FTP server (**ftp.pdb.bnl.gov**) in the /**pub**/**pdbbrowse**/**WWWBrowse** directory. See the README and INSTALL files in this directory for full instructions on downloading and installing this browser. For details on how to obtain a local copy of the PDB on CD-ROM, see our Order Form in this Newsletter. To learn how you can keep your local PDB collection up-to-date automatically over the Internet, see the article "Mirroring" in the January 1995 Newsletter. Additional information about the WWW browser can be found in this Newsletter in the article "The *Portable* WWW Browser."

– Accessing Remote Databases

To provide easier access to valuable, related databases that are remote from your location, and accordingly may be associated with very poor network access times, the PDB and several other groups are making "mirror" copies of them available. Complementary hyperlinks between databases are being included on each group's WWW home page. One such database to which a number of users in the USA wanted faster access is SCOP: A Structural Classification of Proteins Database, developed at the MRC Laboratory of Molecular Biology and the Cambridge Centre for Protein Engineering (see related article in our January 1995 Newsletter and a review of SCOP by Barton in *TIBS* **19**, 554-555 (1994)). SCOP is available directly from Cambridge (**URL: http://scop.mrc-lmb.cam.ac.uk/scop/**) with mirror copies available around the world, including on the East Coast USA at the PDB (**URL: http://www.pdb.bnl.gov/scop**) and at the NCBI (**URL: ftp://ncbi.nlm.nih.gov/repository/scop/index.html**), on the West Coast USA at Protein Science (**URL: http://www.prosci.uci.edu/scop**), in the Far East (Japan) at PERI (**URL: http://www.peri.co.jp/scop/**) and in the Middle East at the Weizmann Institute Bioinformatics (Israel) (**URL: http://bioinformatics.weizmann.ac.il/localdb.html/**). At the PDB, we plan to establish mirror copies and hyperlink pointers to other important databases to facilitate the transfer of knowledge derived from the PDB.

– *Concerning Homology Model Building of Proteins*

There are a number of different manual and automatic procedures to construct a three-dimensional structural model of a protein by homology. This process is based on aligning amino acid sequence to that of a sufficiently homologous protein whose three-dimensional structure is known. Model building requires not only expensive computer hardware and software, but also expert knowledge of their manipulation. To simplify this procedure for non-experts, Manuel Peitsch has developed a highly automated, knowledge-based protein modelling tool called ProMod: Automated Knowledge-based Protein Modelling Tool, which allows the rapid generation of models for new protein sequences based on the known three-dimensional structure(s) of related family members. This tool, described briefly in M.C. Peitsch and C.V. Jongeneel, *Int. Immunol.* **5**, 233-238 (1993), is accessible through the WWW server Swiss-Model **(URL: http://expasy.hcuge.ch/swissmod/SWISS-MODEL.html)**. It is important to realize that the results of any homology project must be considered tentative, or hypothetical, unless substantiated by independent experimental evidence. As is stated in Swiss-Model's WWW home page, "The results of any modelling procedure are NON-EXPERIMENTAL and MUST be considered with care. This is especially true since there is no human intervention during model building." Swiss-Model is described further in the article "ProMod: Automated Knowledge-based Protein Modelling Tool."

– Joel L. Sussman

The *Portable* ***WWW Browser***

The WWW version of the PDB browser has met with success in many areas of the networked community, with roughly 1,800 accesses per week from around the world. When it was designed, a modular approach was used that allowed for the separation of the user interface, search engine, and file access functions in order for us to adapt quickly to advances in each of those three areas. As a side benefit, this also allowed the browser code to be portable, so that those who have poor network access to the computers at the PDB may be able to have the search engine and file service parts of the browser installed locally and thereby improve access to the PDB archive. This article describes the portable browser. Detailed installation instructions may be found on-line at the PDB.

Since this version of the browser uses the WWW, Dave Stampf, in collaboration with Jaime Prilusky of the Bioinformatics Unit of the Weizmann Institute, set it up to have hyperlinks to the E. C. Database and Entrez Reference Database added automatically to the PDB html file as it is being transmitted to the user's viewer program. This valuable information is thus a mouse click away, permitting an expansion of the research possibilities afforded the user.

In order to run the WWW browser locally, you have to be running an http server. (This is the communications protocol underlying the WWW.) For information on how to do this, please see **http://hoohoo.ncsa.uiuc.edu/docs/Overview.html** for one possible http server. You must also have Perl installed on your system. Once you have a system with a functional http server and Perl, you can install the WWW browser in layers.

– *The Browser Code*

The first layer is a Perl script and its associated help file. This script creates the form with which the user works and constructs, but does not execute, the queries. This is easy to install and uses few system resources.

In order to perform the installation, you need to edit between one and five lines in the Perl script called browse.pl. The lines specify the name and location of the Perl script, the name and location of the help file, the machine and port of the query server (second layer) and the name of the script that retrieves and marks up the files (third layer).

– *The Search Engine*

The second layer consists of a second Perl script and a number of index files. Together, these actually execute the user's query, returning the ident code and perhaps a line of text describing the entry, but not the entry itself. This layer requires much more disk space and some CPU cycles as well, but it is primarily I/O bound. The current version of this search engine uses UNIX dbm files, but we are currently testing the use of SYBASE to perform the same (and more complex!) searches.

This Perl script accepts incoming requests on an unprivileged port and then forks a copy of itself to handle the request and awaits another request. In order to install this part you must have an up-to-date set of the browser's index files installed (and preferably, also mirrored) on your local machine. You must also tell the Perl script the location of these indices in your directory tree. The index files require less than 18 Mbytes of disk space as of April 1, 1995. These files are available from the PDB in compressed format, may be automatically mirrored, and are identical to the index files used in the tk/tcl version of the browser.

– *File Service*

The third and final layer consists of a small number of Perl scripts, together with a few more index files, and requires all of the PDB archive to be on-line locally. Many locations will find this requirement unworkable – those sites can simply install layer one, or layers one and two. Our experience indicates that not all searches culminate in the transfer of a PDB file, so installation of the first two layers may suffice.

To install the third layer, you simply need to identify the location of the files on your system within the Perl scripts and make sure that the index files are installed.

All of the files necessary to install the portable WWW browser, together with more detailed instructions, are contained in **ftp://ftp.pdb.bnl.gov/pub/pdbbrowse/WWWBrowse/INSTALL**. We welcome bug reports and suggestions. Please send these to Dave Stampf (drs@bnl.gov).

Accessing Individual PDB Entries via the WWW

The PDB is providing easy access to all entries using the URL mechanism of the WWW for developers creating new hypertext links to the PDB archive, looking for a faster way to access individual entries, or experimenting with programs written in Fortran, C, or Perl and would like access to entries without having the entire archive locally mounted.

Every entry contained in the PDB is now available *via* the URL

http://www.pdb.bnl.gov/cgi-bin/get-pdb-entry?id=xxxx

where xxxx is the usual PDB ident code (*e.g.*, 1abc). In addition, if your network links are slow, you may wish to try the variation

http://www.pdb.bnl.gov/cgi-bin/get-pdb-entry?id=xxxx&encoding=compress

which will send the file to you in UNIX compressed format, typically saving two-thirds of the network bandwidth.

The first URL should work from any "reasonable" WWW browser. The second is successful with the recent versions of Mosaic (on UNIX) which will uncompress the data it receives, but not Netscape, Lynx, or Get where you will have to save and uncompress the data in a separate command. Of course, if you use either of the above URL's from a UNIX command line, please be sure to quote the string since '?' and '&' are usually special characters in the shell.

Finally, one more parameter is available for either of the lines. If you append the string "&type=view," then the file (either compressed or uncompressed) is transmitted with the MIME Content-type = "chemical/x-pdb." In a manner similar to the action of the PDB WWW browser, this will transmit the file and start the external viewer (*e.g.* RasMol) to display the structure.

In order to distribute the network load evenly around the globe, we will make these scripts available to all PDB redistributors running an http server. We will announce availability of the scripts on the PDB's listserver **(pdb-l@pdb.pdb.bnl.gov)**.

Thanks are due to Steve Brenner who originally asked that we make this public and provided some of the ideas for options. We would be happy to consider any other extensions to this scheme. Please contact Dave Stampf (drs@bnl.gov) with your ideas.

3DBbase – A Relational Database

The PDB is in the process of building a relational database to manage the archive and to give its users a more powerful engine to browse its contents. The schema is presented on PDB's WWW server, and at this time the database that it represents is being made available to the community for evaluation and testing. Access to the data stored in the database will be through PDB's WWW browser.

This database development work is a collaboration that includes the following groups:

The Protein Data Bank, Brookhaven National Laboratory
Bioinformatics Unit, Weizmann Institute of Science
Data Management Tools Group, Lawrence Berkeley Laboratory
The Genome Data Base, Johns Hopkins University

The schema was generated with OPM Schema Editor 3.0, developed by the Data Management Tools Group, Lawrence Berkeley Laboratory.

The database is being implemented on a SYBASE engine. In addition to all coordinate entries found in the PDB, the PDB database includes semantic links to entries found in other biological databases. The first steps in the formation of a federation of biological databases are also being made possible by this work. All PDB database bibliographic citations are stored and maintained on GDB's citation database (CitDB), also built using OPM.

We plan to have a fully operational database system by October 1995. In the interim, all are welcome to test our system, study our schema, and suggest changes and extensions. Further documentation providing details and explanations of the objects presented in the schema, as well as our development plans, will be released in the coming weeks.

The schema can be accessed through the PDB WWW home page **(http://www.pdb.bnl.gov)**.

ProMod: Automated Knowledge-based Protein Modelling Tool

This article was written by Manuel C. Peitsch, Glaxo Institute for Molecular Biology, Geneva Switzerland. It describes a tool that may be of interest to PDB users.

Molecular models of proteins have recently gained much popularity among biochemists and molecular biologists, as they have proven useful in many instances. Indeed, the design of site-directed mutagenesis experiments can be rationalized to a great extent by the use of theoretical protein models. Model building requires not only expensive computer hardware and software, but also expert knowledge of their manipulation. Thus only a limited number of scientists have access to these tools. To overcome these limitations we developed the automated, knowledge-based protein modelling tool ProMod which allows the rapid generation of models for new protein sequences based on the known three-dimensional structure(s) of related family members. This tool is accessible through the WWW server Swiss-Model **(URL: http://expasy.hcuge.ch/swissmod/SWISS-MODEL.html)**.

The Swiss-Model modelling process begins with the identification of suitable template structures based on their sequence similarity with the target sequence. These sequences are then aligned with the target, taking into account the structural similarity between all templates. This step is accomplished with the sequence alignment algorithms BLAST [1], FastA [2], and SIM [3], and the MATCH-3D module of ProMod [4]. The model coordinates are then generated by ProMod, which follows these steps:

- The construction of an averaged framework [5] from the superimposed template structures.
- The generation of atomic coordinates, derived from the averaged framework, using the multiple sequence alignment described above.
- The rebuilding of nonconserved loops (including both insertions and deletions) from their “stems” by structural homology searches through the PDB as described by Greer [6].
- The completion of the main chain using a library of backbone elements (pentapeptides) derived from the best x-ray structures (< 2Å resolution);
- The reconstitution of missing side chains and the correction of existing ones using a library of allowed rotamers [7], and finally, computation of the three-dimensional profile to assess the model quality [8].

  Optimization of bond geometry and relief of unfavorable non-bonded contacts is then performed by 50 steps of steepest descent followed by 500 steps of conjugate gradient energy minimization using CHARMM [9] with the PARAM22 parameter set. A model confidence factor describing the degree of uncertainty linked to each residue is computed during the modelling procedure and occupies the crystallographic B-factor field in the final coordinate file.

Requests can be submitted to the Swiss-Model server through easy-to-fill-out forms using the WWW browsers NCSA-Mosaic or Netscape. Since protein modelling is heavily dependent on the alignment between target and template sequences, Swiss-Model provides two distinct modes of function accessible through two separate forms:

- The First Approach Mode allows the user to submit a sequence or its Swiss-Prot identification code. In this mode, Swiss-Model will go through the complete procedure described above. The First Approach Mode also allows the user to define a choice of pre-selected template structures. The final model and its three-dimensional profile, as well as intermediate results if requested, will be returned to the user *via* e-mail. Among the files sent are the sequence alignment and the command file used by ProMod.
- The Optimize Mode allows the user to recompute a model by submitting altered sequence alignment and ProMod command files, obtained following a First Approach request.

[1] S.F. Altschul, W. Gish, W. Miller, E.W. Myers, and D.J. Lipman, Basic Local Alignment Search Tool, *J. Mol. Biol.* **215**, 403-410 (1990).

[2] W.R. Pearson and D.J. Lipman, Improved Tools for Biological Sequence Comparison, *Proc. Natl. Acad. Sci. U.S.A.* **85**, 2444-2448 (1988).

[3] X. Huang and M. Miller, A Time-Efficient, Linear-Space Local Similarity Algorithm, *Adv. Appl. Math.* **12**, 337-357 (1991).

[4] M.C. Peitsch and C.V. Jongeneel, A 3D Model for the Cd40 Ligand Predicts That it Is a Compact Trimer Similar to the Tumor Necrosis Factors, *Int. Immunol.* **5**, 233-238 (1993); M.C. Peitsch and J. Tschopp, Comparative Molecular Modelling of the Fas-Ligand and Other Members of the Tnf Family, *Mol. Immunol.* in press (1995).

[5] T. Blundell, B.L. Sibanda, M.J. Sternberg, and J.M. Thornton, Knowledge-based Prediction of Protein Structures and the Design of Novel Molecules, *Nature* **326**, 347-352 (1987).

[6] J. Greer, Comparative Modelling Methods: Application to the Family of Mammalian Serine Protease, *Proteins* **7**, 317-334 (1990).

[7] J.W. Ponder and F.M. Richards, Tertiary Templates for Proteins. Use of Packing Criteria in the Enumeration of Allowed Sequences for Different Structural Classes, *J. Mol. Biol.* **193**, 775-791 (1987).

[8] R. Luethy, J.U. Bowie, and D. Eisenberg, Assessment of Protein Models with Three-Dimensional Profiles, *Nature* **356**, 83-85 (1992).

[9] B.R. Brooks, R.E. Bruccoleri, B.D. Olafson, D.J. States, S. Swaminathan, and M. Karplus, CHARMM: A Program for Macromolecular Energy, Minimization, and Dynamics Calculation, *J. Comp. Chem.* **4**, 187-217 (1983).

Notes of a Protein Crystallographer

– Only Refined Proteins go to Heaven

This article was written by Cele Abad-Zapatero, Abbott Laboratories, Abbott Park, IL, USA. He intends to contribute regularly under this heading. If you have comments or suggestions please contact him at abad@randb.abbott.com.

It was a typical manuscript to be refereed: Introduction, Materials and Methods, Results, and Discussion; Crystal Growth, Structure Solution and Refinement, with some discussion as to the biological implications of the refined structure. Yet, after the refinement something caught my attention:

> *"The coordinates have been deposited in the Protein Data Bank with ascension code 1ABC."*

I do not quite recall the entire entry. However, "ascension" was not the right word. It was obviously a misprint, the authors did not mean to write down *that* word. Yet, I could not resist the pun. Was this a freudian slip? Had the authors implied that after so many years of hard work on that particular protein and after the structure had been solved and refined, their results had to ascend to some special "heaven of refined proteins?" Well, yes and no.

In the mid-seventies, the PDB was a depository of the few protein structures that had been solved (a few had even been partially or fully refined!) using the methodology which makes our profession: single-crystal x-ray diffraction. At that time, indeed, only a few proteins made it up there. They were very few for several reasons, the most important of which was that it was relatively difficult still to solve protein structures by the MIR or MR methods. In addition, some crystallographers were reluctant to make their hard-won structural results available to the community at large for a variety of pretexts.

The situation began to change when the journals of our trade (Acta Crystallographica, Journal of Biological Chemistry, Science, Journal of Molecular Biology, and others) began to require that coordinates be deposited at the time of publication. From then on, the PDB has been flooded with new macromolecular structures that have to be "tested for goodness," "processed," and eventually made public to the community of structural biologists gratis or for a modest fee.

The problem has been compounded by an increasing number of macromolecular crystallographers using better hardware to collect data and improved methods for structure solution and refinement. It is to the credit of our profession that structural results from crystals of macromolecules are becoming easier to obtain.

Large numbers do not go very well with protected heavens and thus the PDB is not such an exclusive place anymore. In addition to increasing numbers of structures determined by x-ray diffraction methods, in the last few years coordinates of model structures have also been deposited in the PDB and even the ensembles of structures determined by NMR also ascend to the PDB. All these structural results can now be retrieved within seconds and displayed on any computer workstation. I was astounded when one of my molecular modeler colleagues displayed, within seconds, what amounted to several people-years of effort.

If all these structural jewels have not ascended to heaven, they have certainly been elevated to the realm of eternal ideas or abstractions. We can now have an image of "hemoglobin" with its helical contortions holding oxygen very delicately, like a feather, ready to give it away and to exchange it by carbon dioxide many times per second. Like Plato's images, they are ideal representations of the atomic world beyond our reach. Their shadows appear full of color in our electronic Plato's Caves by virtue of a few dials and a cathode tube.

However, protein crystallographers, like common mortals, have to deal not with the ideal folds but with the individual variants that are often hard to crystallize; their crystals do not necessarily diffract well and take time to solve and refine. Our joy and satisfaction must lie in collecting as many of these beautiful pebbles as possible and in understanding the significance of every detail and variation of these atomic masterpieces.

Newsletter Mailing List Re-initialized

As announced, the PDB re-initialized the Newsletter mailing list after distribution of the January 1995 edition.

Previous subscribers, as well as new subscribers, who wish to receive printed copies of future Newsletters must contact us immediately. A Newsletter Request Form, similar to the one shown in the back of this issue, is available in the file **news_mailinglist** from the FTP /**pub** directory. This form may be completed and returned electronically (send_news@pdb.pdb.bnl.gov) or *via* the postal system to:

PDB Newsletter Mailing List
Chemistry Department, Bldg. 555
Brookhaven National Laboratory
P.O. Box 5000
Upton, NY 11973-5000 USA.

Newsletter Automatically Available

A new PDB electronic mailing list has been established for those wishing to automatically receive a text copy of the PDB Newsletter each quarter *via* e-mail.

To subscribe, send e-mail to **listserv@pdb.pdb.bnl.gov** with the following one-line message:

subscribe NEWSLETTER-L Firstname Lastname

PDB Proposes Changes to ATOM/HETATM Records

PDB is planning changes in the format of the ATOM and HETATM records. The proposed modifications address needs brought to our attention by many users. If acceptable, these changes will take place in approximately two to three months. Along with other changes to our format, as discussed in the article "Revised PDB Format Description," these changes are related to our current efforts to produce PDB entries using CIF.

PDB will introduce these format changes in such a way as to minimize negative impact on existing software. Numerous applications rely on the current coordinate record format, and we wish to give ample notice of any changes. We need input from the community regarding this issue, so please examine the proposal carefully and send us your comments (abola1@bnl.gov).

PDB proposes to use columns 73 - 76 to identify specific segments of the molecule, and columns 77 - 80 to provide element information. Currently, columns 71 - 80 contain the entry's ident code and line number. If elimination of these data presents a problem to any programs, we need to be informed.

Columns 73 - 76 will contain the segment id (SEGID), which will identify specific segments of molecules. The segment can consist of a complete chain or a portion of a chain. The importance of this new field can be readily understood if one considers an antibody structure having two molecules in the asymmetric unit. Since each chain must have a unique chain identifier, the two heavy chains and two light chains cannot currently be labeled to indicate their nature. SEGID's of VH, CH1, CH2, CH3, CL, and VL would clearly identify regions of the chains and the relationship between them. Users of X-PLOR will be familiar with SEGID as used in the X-PLOR refinement application.

SEGID is defined as a string of at most four (4) alphanumeric characters, left justified, *e.g.*, CH86, NASE, VH1.

Columns 77 - 78 will contain the atom's element symbol, right justified, and columns 79 - 80 will indicate any charge on the atom, *e.g.*, MN2+, O1-, H. In the past, hydrogen naming sometimes conflicted with IUPAC conventions. For example, we have not been able to label a hydrogen HG11, but rather have employed 1HG1 in order for it not to be confused with mercury. After adopting the format change, HG11 will be allowed in columns 13 - 16, and hydrogen will be clearly identified in columns 77 - 78; thus columns 13 - 16 will continue to be used to uniquely identify each atom.

Please send your comments on these proposed changes to ATOM and HETATM records to Enrique Abola (abola1@bnl.gov). Again, if acceptable to our users, these changes will be implemented in the near future.

PDB Release Policy

To clarify PDB's policy regarding on-hold entries, the following is included in the acknowledgment letter sent to depositors upon data acceptance:

> *PDB follows the IUCr guidelines which state that coordinates may be held (before release) no longer than one (1) year and structure factors may be held no longer than four (4) years from the date of publication. PDB has chosen to apply the same guidelines to NMR restraints data, allowing a maximum hold of four (4) years. Requests that the PDB delay release of your data (put it on hold) should be submitted at the time of the initial deposition. PDB cannot consider hold requests received more than one week after the date of this acknowledgment letter. A one-time extension of up to six (6) months due to delay in publication can be requested in writing. In no case will a coordinate data set be held for longer than eighteen (18) months from the date of deposition. Deposition of data constitutes your acceptance of PDB's Release Policy.*
>
> *Please note that information on the status of all entries, including those on hold, is available to the public. The PDB Pending and Waiting list, updated daily, is available through anonymous FTP and is searchable using WWW and Gopher. This file gives the status, including hold expiration date if the entry is on hold, of every pending entry. In addition, the full list of on-hold entries, with their hold expiration dates, is available via FTP, WWW, and Gopher. See the file named* ***/pub/on_hold.list****.*

Revised PDB Format Description

The revised PDB Format Description for Atomic Coordinate Entries recently has been released. The purpose of this document is to describe completely the contents of PDB coordinate entry files. Several changes are being introduced to PDB files to make them more explicit to the human reader and more easily computer-parseable. Additionally, once the macromolecular CIF has been adopted, these changes will pave the way for conversion to CIF.

An important enhancement to the current PDB format is in ATOM/HETATM records (see article "PDB Proposes Changes to ATOM/HETATM Records"). Additions to PDB files include new record types, such as TITLE, CAVEAT, KEYWRD, CISPRO, MODRES, DBREF, and SEQADV; introduction of keyword:value pairs in certain records such as COMPND, SOURCE, and REMARK 3; further detailing of the heterogen groups with the new records HETNAM, HETSYN, and HETSIT (list of residues in very close proximity to a given heterogen); the deprecation of footnotes; and restructuring of some REMARK records to make them more machine-accessible.

This revised Format Description will be helpful to several communities. It will assist depositors in preparing entries for deposition, guide software and information resource developers, and help users of the PDB understand the contents of coordinate entries. Ultimately, this document and the enhanced entry format will facilitate the conversion of PDB files into CIF.

This document is available in several formats (HTML, plain text, and PostScript) from FTP, Gopher, and WWW in the /**pub** directory. A more advanced HyperText version is being designed for later release.

A convenient way to peruse the Format Description is to use Mosaic or another WWW interface. Access the PDB home page by opening the URL **http://www.pdb.bnl.gov/**. Move to "NEW! - Revised PDB Format Description in hmtl" where you will see several choices.

WWW Version of the PDB Browser

Previous Newsletter articles described a GUI-based browser utility running under tcl/tk for searching the PDB archives. This browser provided the ability to search text portions of entries using arbitrary regular expressions. Additionally, the browser incorporated graphical tools such as RASMOL and MidasPlus to view selected molecules. The browser was written in a modular fashion, permitting replacement of the search, front-end, or display mechanisms.

A WWW version of the browser was made available to the user community **(http://www.pdb.bnl.gov/cgi-bin/browse)**. This replaces the front end of the browser with popular WWW viewer programs such as Mosaic or Netscape. The WWW front end provides most of the functionality of the original browser while adding the following benefits:

- All searching and access is over the network. One need not install Perl, tcl, and tk. (But, on general principles, you should!)
- The archive being searched is the up-to-date PDB (or a mirrored copy!).
- Hypertext links are provided in the PDB file display to the Enzyme Data Bank as well as to the sequence databases.
- Hypertext links are provided to use the default display program (*e.g.*, RasMol) and pictures from M. Peitsch.
- PDB access is provided for PC, Macintosh, and UNIX computers.

The primary drawback is that the search of the remark fields has been deleted due to the time necessary to complete the searches and the network time-outs that resulted.

Short-term plans include replacing search scripts with a SYBASE database engine, increasing linkages to sequence and reference databases, and taking advantage of video and audio capabilities of WWW. Exporting of the server source code to PDBSA Affiliated Centers, permitting the user community to take advantage of the best network links available, is currently taking place.

Access to PDB

– World Wide Web (WWW)

PDB has a World Wide Web (WWW) server on the computer system **www.pdb.bnl.gov** (130.199.144.1). This server is accessible using the document **URL: http://www.pdb.bnl.gov/**.

Besides including links to the PDB FTP and Gopher servers, the WWW server includes links to many other useful databases and information servers.

– Gopher

PDB has a Gopher server on the system **gopher.pdb.bnl.gov** (130.199.144.1). This server is accessible using a Gopher client connecting to the following link:

Name	=	Protein Data Bank FTP server
Type	=	1
Host	=	gopher.pdb.bnl.gov
Port	=	70
Path	=	1/

As a Gopher client, you may navigate through a hierarchy of directories and documents or ask an index server to return a list of all documents that contain one or more specified words. For instance, you can choose “The PDB Anonymous FTP” after reaching PDB’s Gopher server in order to search and download the same information and coordinate files as through FTP. Alternatively, you can select “An (almost) full-text search of the PDB Bibliographic Headers” in order to search PDB using any keyword **.**

– FTP

PDB has an anonymous FTP account on the computer system **ftp.pdb.bnl.gov** (Internet address 130.199.144.1). Files may be transferred to and from this system using anonymous as the FTP user name and your e-mail address as the password. Besides downloading entries, data files, and documentation, it is possible to upload any files that you may wish to send to PDB, into the special directory /**new_uploads**. Those using VMS may need to place quotes around file names.

– Listserv

PDB has a mailing list devoted to discussions concerning its operation, contents, and access procedures.

To subscribe, send e-mail to listserv@pdb.pdb.bnl.gov with the one-line message: **subscribe PDB-L Firstname Lastname**.

To find out what can be done with this mailing list, send e-mail to the same address (listserv@pdb.pdb.bnl.gov) with the one-line message: **help**.

To send a message to all PDB-L subscribers, e-mail the message to: **PDB-L@pdb.pdb.bnl.gov**.

This page intentionally left blank.

BROOKHAVEN ORDER FORM

Name of User ______________________________ Date ______________

Organization ______________________________ Phone ______________

Address ______________________________ Fax ______________

______________________________ E-mail ______________

____

- ***Price is valid through September 30, 1995***
- ***Price shown is per release - there are four releases 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 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.

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 three required items together *via* postal mail to:

Protein Data Bank Orders
Chemistry Department, Building 555
Brookhaven National Laboratory
P.O. Box 5000
Upton, NY 11973-5000
USA

Protein Data Bank CD-ROM — ISO 9660 Format *$300.00*

(tax and shipping charges not applicable)

NEWSLETTER REQUEST FORM

Current subscribers, as well as new subscribers, who wish to receive printed copies of future Newsletters must contact us immediately. A form similar to the one below is available in the file **news_mailinglist** from the FTP /**pub** directory. It is preferable that this form be retrieved, completed, and returned electronically to:

send_news@pdb.pdb.bnl.gov

If Internet capabilities are unavailable, please complete this form and send it by postal mail to: PDB Newsletter Mailing List, Chemistry Department, Building 555, P.O. Box 5000, Brookhaven National Laboratory, Upton, NY 11973-5000 USA.

Please send me future printed copies of the Newsletter:

prefix:__________ first name:______________________________ middle name:____________________

last name:____ suffix:__________

professional_title:____________________

department_1:______________________ department_2:______________________

institute_1:______________________ institute_2:______________________

po_box:______________________ bldg:______ street_address:______________________

city:______________________ state:______________________

postal_code:______________________ country:______________________

telephone_number:______________________ fax_number:______________________

network_address_1:______________________ network_address_2:______________________

AFFILIATED CENTERS

Twenty-two affiliated centers offer DATAPRTP information for distribution. These centers are members of the Protein Data Bank Service Association (PDBSA). Centers designated with an asterisk(*) may distribute DATAPRTP information both on-line and on magnetic or optical media; those without an asterisk are on-line distributors only.

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

*BIOSYM
BIOSYM Technologies, Inc.
San Diego, California
Laurel Frey (619-546-5509)
rcenter@biosym.com or laurel@biosym.com

BIRKBECK
Crystallography Department
Birkbeck College, University of London
London, United Kingdom
Alan Mills (44-171-6316810)
a.mills@cryst.bbk.ac.uk
http://www.cryst.bbk.ac.uk/PDB/pdb.html

CAN/SND
Canadian Scientific Numeric Data Base Service
Ottawa, Ontario, Canada
Roger Gough (613-993-3294)
cansnd@vm.nrc.ca

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

*CCDC
Cambridge Crystallographic Data Centre
Cambridge, United Kingdom
David Watson (44-1223-336394)
dgw1@chemcrys.cam.ac.uk

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

CINECA
NE Italy Interuniversity Computing Center
Casalecchio di Reno (BO), Italy
Laura Setti (39-51-6599478)
asltc0@icineca.cineca.it

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

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

INN
Israeli National Node
Weizmann Institute of Science
Rehovot, Israel
Leon Esterman (972-8-343934)
lsestern@weizmann.weizmann.ac.il

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

*MAG
Molecular Applications Group
Palo Alto, California
Hilary Jensen (415-473-3039)
hilary@suerte.mag.com

*MSI
Molecular Simulations Inc.
Burlington, Massachusetts
Lance J. Ransom Wright (617-229-9800)
lance@msi.com

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

NCSA
National Center for Supercomputing Applications
University of Illinois at Urbana-Champaign
Champaign, Illinois
Patricia Carlson (217-244-0768)
pcarlson@ncsa.uiuc.edu

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

*OML
Oxford Molecular Ltd.
Oxford, United Kingdom
Steve Gardner (44-1865-784600)
steve@gardner.demon.co.uk

*OSAKA UNIVERSITY
Institute for Protein Research
Osaka, Japan
Yoshiki Matsuura (81-6-879-8605)
matsuura@protein.osaka-u.ac.jp

PITTSBURGH SUPERCOMPUTING CENTER
Pittsburgh, Pennsylvania
Hugh Nicholas (412-268-4960)
nicholas@cpwpsca.psc.edu

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

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

Protein Data Bank
Chemistry Department, Bldg. 555
Brookhaven National Laboratory
P.O. Box 5000
Upton, NY 11973-5000
USA

Statement of Support

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.

To Contact PDB

Telephone 516-282-3629
Facsimile.. 516-282-5751

Internet:

pdb@bnl.gov......................... general correspondence
orders@pdb.pdb.bnl.gov .... order information
sysadmin@pdb.pdb.bnl.gov .. network services
listserv@pdb.pdb.bnl.gov ...... Listserver subscriptions
pdb-l@pdb.pdb.bnl.gov.......... Listserver postings
errata@pdb.pdb.bnl.gov .... entry error reporting

Please include your name, postal mailing address, e-mail address, facsimile number, and telephone number in all correspondence.

PDB Staff

Joel L. Sussman, Head
David R. Stampf, Sr. Project Mgr.
Enrique E. Abola, Science Coordinator
Jaime Prilusky, Interim Head Database Dev.

Frances C. Bernstein	Michael D. Libeson
Judith A. Callaway	Nancy O. Manning
Minette Cummings	John E. McCarthy
Betty R. Deroski	Regina K. Shea
Pamela A. Esposito	John G. Skora
Arthur Forman	Karen E. Smith
Patricia A. Langdon	Dejun Xue

First Class