Protein Data Bank

Quarterly Newsletter

Release #70 — October 1994

October 1994 PDB Release

2921 full-release atomic coordinate entries (249 new additions)

- 2703 proteins, enzymes and viruses
- 208 (DNA, RNA, tRNA)
- 10 carbohydrates

358 structure factor entries
31 NMR experimental entries

The total size of the atomic coordinate entry database is 944 Mbytes uncompressed.

Table of Contents

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

There has been a true revolution in the field of Structural Biology during the past few years, resulting in an enormous increase in the number of laboratories doing structural studies of biological macromolecules to atomic resolution as well as in the rate that these structures can be determined. This is due in part to great improvements in molecular biology techniques and X-ray detectors, greater availability of synchrotrons for protein crystallography, improved tools for refinement and map fitting, and improvements in NMR techniques which make possible structure determination from non-crystalline samples. The resulting avalanche of new biomolecular structural data presents a major challenge for the PDB, and the major revolution in computer networking and database hyperlinking over the past couple of years represents the most promising way to deal with it effectively.

The PDB has been very active in this area. Over the past few years, an anonymous FTP server and archive, and later a gopher hole, were set up at the PDB. In August 1994, a World Wide Web (WWW) home page was set up (**http://www.pdb.bnl.gov**/). This new way of communicating *via* the Internet, recently described in detail [Schatz and Hardin (1994), "NCSA Mosaic and the World Wide Web: Global Hypermedia Protocols for the Internet", *Science* **265**, 895-901], is at present the most convenient way to access the PDB as well as many other biological and chemical databases. Communicating *via* Mosaic provides a very friendly user front end that allows easy access to, and retrieval of files from, the PDB as well as search capabilities for new entries which are pending or on hold. As described in our July 1994 Newsletter, several hundred spectacular images of PDB structures in RGB and GIF formats are available from the PDB FTP server. Created by Dr. Manual Peitsch of the Glaxo Institute for Molecular Biology in Geneva, Switzerland, each one depicts a key aspect of the entry or experiment that it represents. In the future, we plan to add hyperlinks to other biological and chemical databases *via* the WWW server, as well as a home page for the PDB-Browser to replace its current tcl/tk-based front end.

— *Joel L. Sussman*

ZINC - Galvanizing CIF to Work with Unix

The July 1994 Newsletter described the basic structure of a CIF (Crystallographic Information File) and the standard that will probably become the interchange format of the future for the

PDB due to increased amounts of information that it can contain. It was also pointed out that those who are accustomed to working with the PDB format in a Unix environment (with grep, awk, perl, diff, etc.) will not be able to use those skills in dealing directly with CIF.

This article describes a format called ZINC (Zinc Is Not CIF) which is fully accessible to Unix tools and a number of utilities that allow a CIF to be converted into a ZINC and back again, as well as versions of some familiar Unix tools (grep, diff) and some surprising new tools (zincSubset and zincNl) that should make access to CIF much easier.

What is the Problem with CIF and Unix?

CIF defines a format that is generally at odds with Unix tools.

- Many Unix tools are line oriented – they expect related information to be on a single line whereas CIF allows and encourages the use of multiple lines, both by limiting the line length and in the definition of loops.

- Many Unix tools break the lines into fields based on a separator character. Both PDB and CIF formats work against this, the former by having column based fields, the latter by encouraging the liberal use of white space (across line boundaries).

- A number of Unix tools treat the information in files as being position dependent (diff, head, tail, etc.). The PDB file format adapted neatly to this, but CIF allows a much wider variation in the placement of data.

– ZINC

Zinc is not an interchange format as CIF is, but rather a piping format, i.e., a format that makes the contents of a CIF accessible to Unix utilities. Each data line of a ZINC file consists of five tab-separated fields:

block name index value loop-id

The first field is the name of the CIF data block (the data_ prefix is omitted) and is repeated on each line where appropriate. The second field is the name of the data item. The third field is an index specifier, which is empty for non-looped data, and is a zero-based index for looped data. The fourth field is the data item itself. For multiple line CIF data, new lines are replaced by the two characters \n. The backslash character becomes the escape character throughout the ZINC format. The fifth field is a loop identifier. Comments that appear in a CIF are associated with the previous token and are also represented in the ZINC format.

– Existing Tools

A number of tools have been developed to support the user community in using the ZINC format and to access the information contained in a CIF. Most are simple and allow users to modify the code to tackle new problems. Source is provided for all.

- **cifZinc**, obviously the first required tool, converts an existing CIF into a ZINC. This is a C program that converts even the largest CIF in a few seconds.

- **zincCif**, the next most important tool, takes a ZINC and creates a pretty-printed CIF. Very often the pipeline:

  ```
  cifZinc old.cif | zincCif > new.cif
  ```

  will produce a better looking CIF than the original.

- **zincGrep**, greps a ZINC (or a CIF file specified as a command line argument) for a regular expression and returns the block name, data name, index, and value of the match. This is the single most requested tool for dealing with CIF.

- **cifdiff** is a four-line C-shell script that takes two CIFs and determines the differences between them. It will handle CIFs that have been rearranged, and even loops with rearranged columns, and provide only the real differences.

- **zb**, a small (< 200 lines) tcl/tk program that provides a simple GUI front end to a ZINC or CIF, allows users to browse through the contents. Multiple files as well as multiple data blocks can be viewed simultaneously on any X terminal. zb recognizes command line argument file names in the form *.cif as being a CIF and converts it to a ZINC automatically.

- **zincSubset** is another C-Shell script that is very short but very useful. It allows users to generate a custom subset of any ZINC (CIF), simply by listing the data blocks and data names that he or she wishes to include. For example, if you wanted to extract only the names and definitions from the mmCIF dictionary, you would create a file (e.g., defs) with two lines that appear as:

  _name
  _definition

  (preceded and followed by tabs) and run the command:

  ```
  zincSubset defs mmcif94 | zincCif
  ```

- **zincNl**, a perl script, takes a ZINC file and creates a FORTRAN compatible namelist file allowing easy access to any CIF by FORTRAN programs without the need for extensive I/O libraries or reprogramming. As with zb above, it will automatically convert a CIF to a ZINC. It may be used in the following pipeline which extracts the coordinates from a CIF and presents them to a FORTRAN program *via* the namelist mechanism:

  ```
  zincSubset coords datafile | zincNl | myfortran
  ```

  (where coords is a file that simply has three lines with x, y, and z surrounded by tabs).

– A Sample CIF and ZINC

The following CIF illustrates most aspects involved in translating to a ZINC:

```
#
#     A simple CIF
#

data_object

#
#     polygon
#
  _name
;
triangle
;
  loop_
     _x _y
     0.0      0.0
     1.0      0.0
     0.0      1.0

     _num_sides 3
```

In ZINC, this would appear as:

```
          (           0     #
          (           1     #     A simple CIF
          (           2     #
object    (           3     #
object    (           4     #     polygon
object    (           5     #
object    _name             ;\ntriangle\n;
object    _x          0     0.0                 _x
object    _y          0     0.0                 _x
object    _x          1     1.0                 _x
object    _y          1     0.0                 _x
object    _x          2     0.0                 _x
object    _y          2     1.0                 _x
object    _num_sides        3
```

Note that comments "belong" to a data block and are represented with an open parenthesis for the data name. Initially, the data block name is defined to be the null string.

– Code

The formal definition of ZINC and the above mentioned programs are available from PDB using FTP, Gopher, or WWW (**ftp://pub/other-software/Zinc**). Please give it a try. All are invited to submit their favorite scripts that use ZINC. Any comments should be directed to Dave Stampf (drs@bnl.gov).

Mirroring

Earlier this year, PDB started updating the FTP server with fully released entries on a frequent basis – typically every few weeks. PDB users often ask how they can keep their local archive up to date without having to continually check the current holdings of the PDB. Fortunately, a public domain package called *mirror* exists that does exactly this. This program is a perl script that runs on your system and periodically creates an FTP connection to the PDB, determines the difference between your local archive and the PDB (including deletions!) and performs all necessary tasks needed to make them equivalent. We have used this program with the Weizmann Institute of Science in Israel, Turku University in Finland and the European Molecular Biology Laboratory in Germany with great success even over very poor network connections. We encourage the members of our user community who maintain a complete local archive to do the same.

If you wish to try this, here is what you have to do:

0) **Think:** Decide what your local archive will look like. Will it be a virtual image of the PDB? Will it only hold compressed files? Will it use the all_files in one directory scheme, or the 2-character directory scheme? Your space limitations and your usage patterns will determine what is best for you.

1) **Prepare:** Set up your local archive from a 1994 version of the PDB CD-ROM. Be sure the dates on the files match those on the CD. In order to copy a directory on the CD to a local disk without changing the dates, use the following tar pipeline:

```
mkdir /usr/distr      or where you want the
                      files to go

cd /CDROM/distr       or where you are copying
                      the files from

tar cf - . | (cd /usr/distr; tar xvf -)
```

2) **Get mirror:** Copy the mirror software from PDB or elsewhere (/**pub/other-software/Mirror**). There is nothing to compile, but you must have already installed perl and have run the h2ph perl script. Install the mirror program in one of the standard locations.

3) **Configure:** Adapt a copy of the mirror.defaults configuration file. A sample of this file is also on the PDB FTP server.

4) **TEST!!!** If you run mirror -n it will tell you what it will do without doing it. If you are about to transfer a gigabyte over a megabit line, you may wish to reconsider.

5) When you are satisfied, make the mirror run as part of your crontab and forget about it. Your local archive will be kept up to date "automagically"!

As the mirror README says, "Objects in the mirror are closer than they appear!"

Managing the Archives

Careful watchers of the PDB will have noticed significant improvements in the quality of data distributed by PDB. These changes are primarily due to our use of new data processing and validation procedures introduced last year as part of our effort to convert all prerelease entries into fully annotated form. Details of this work were initially discussed during the 1991 ACA meeting in Toledo, Ohio. We then made public commitments to erase the backlog of unprocessed entries by 1993-1994 and announced plans to upgrade older entries by subjecting them to the same processes as those entries recently deposited. Now that we have achieved our primary goal, we have started the task of upgrading older entries. Sections below discuss some recent work aimed at improving the archives and its effect on users and depositors.

Upgrading the contents of the archive will require a number of changes to PDB entries. Some of these changes might effect existing software. New record types, a new HET group dictionary, and an upgrade of the existing format description document are forthcoming. Information missing from existing entries will be added when possible. We will use our current validation suite of programs to check for possible errors not previously reported in the entries. Advisory notices will be added when appropriate. In addition to improving the quality of the data in the archive, these changes will pave the way for the conversion of PDB interchange format to CIF.

– *Keeping Up with Changes*

A number of steps are being taken to keep users and software developers informed of developments that may effect their work. Announcements, documents, and production schedules related to the upgrade and cleanup work are now stored in the FTP directory **/pub/pdb_upgrade**. In addition, all related announcements will be posted to the PDB-L mailing list. Those interested can participate by sending an e-mail message to listserv@pdb.pdb.bnl.gov with the one-line message: **subscribe PDB-L Firstname Lastname**.

– *Informing PDB of Errors and Updating Entries*

We have established an e-mail address that can be used to inform us of errors found in PDB entries or to supply us with information needed to update the contents of these files.

The updating and upgrading of entries will require the cooperation of our depositors and users. There are two areas in which your help would be invaluable at this time:

1. Many entries list papers as TO BE PUBLISHED. Please send us full publication information for any papers that have now been published. Titles often change between the manuscript and publication stages, making it difficult for us to track them.

2. We suspect that there are entries where the sample was prepared using recombinant techniques but this is not noted in the SOURCE records. It is important that our depositors check all their entries and notify us if there are any that are missing this information.

Error and update reports may be sent to errata@pdb.pdb.bnl.gov. Please include the latest REVDAT record found in the entry along with your e-mail message.

– *New Record Types*

A number of suggested new record types related to amino acid or nucleic sequences were introduced in our July 1994 Newsletter. The following is a list of additional record types to be included in forthcoming releases of the PDB:

Record Type	- Purpose
TITLE	- To provide a succinct description of the contents of the entry.
KEYWRD	- To provide additional keywords that can be used to classify the contents of the entry. This record will be used to supplement the classification field provided in the HEADER record.
NAMHET	- To provide the complete compound name for HET groups. This record type will contain information now presented in columns 31-70 of a HET record.
SYNHET	- To provide synonymous compound names for HET groups.

The TITLE record will contain text now found in the COMPND record. It will contain a brief and succinct description of the experiment represented in the PDB entry. A quick look at recently-released entries will reveal that changes have been made to the COMPND record. These changes reflect our attempts to introduce an internal structure to the COMPND record making it easier to parse information from it (e.g., macromolecule name). There is, however, a need to describe the contents of the file in the same way that titles of published articles help users quickly identify interesting papers to read. The TITLE record will serve this purpose.

The KEYWRD record is designed to allow for inclusion of additional keywords that can be used to classify and index PDB entries. This record supplements information now provided in the HEADER record. Thus this record may contain text that gives both functional and structural classification descriptions. A list of valid values will be made available and will be updated periodically.

The NAMHET and SYNHET records are designed to help identify the heterogen groups. NAMHET will be used to provide the complete and systematic name for a HET group. This information is currently found in columns 31-70 of the HET record or in REMARK records for longer names. SYNHET will include a list of synonyms for HET groups. This record type is acutely needed as it is now very difficult to search for even simple groups such as acetic acids which are described as acetate ions in a number of entries. SYNHET should also make it easier for PDB to describe molecules using systematic IUPAC names in the NAMHET record while allowing for use of more common names in SYNHET records. NAMHET and SYNHET records will be included in the HET group dictionary that is in preparation.

– *Record Descriptions*

Record Name: *TITLE*

Cols.	Contents and Description
01 - 06	TITLE
09 - 10	Continuation field (this field will be blank for the first TITLE record in each entry and will be numbered 2, 3, etc., for continuation records).
11 - 70	Succinct description of the experiment.

Example:

```
TITLE    BACTERIOPHAGE T4 LYSOZYME AT HIGH IONIC STRENGTH
```

Record Name: *KEYWRD*

Cols.	Contents and Description
01 - 06	KEYWRD
09 - 10	Continuation field (this field will be blank for the first KEYWRD record in each entry and will be numbered 2, 3, etc., for continuation records).
11 - 70	List of keywords - each keyword or phrase will be separated by a semicolon.

Example:

```
KEYWORD  SERINE PROTEASE; ALPHA/BETA DOMAIN
```

Record Name: *NAMHET*

Cols.	Contents and Description
01 - 06	NAMHET
09 - 10	Continuation field (this field will be blank for the first NAMHET record for each HET group and will be numbered 2, 3, etc. for continuation records).
11 - 14	Non-standard group (heterogen) identifier.
16 - 70	Compound name - in most cases the IUPAC name will be provided.

Example:

```
NAMHET   CMP CYCLIC-3'-5'-CYCLIC MONOPHOSPHATE
```

Record Name: *SYNHET*

Cols.	Contents and Description
01 - 05	SYNHET
09 - 10	Continuation field (this field will be blank for the first NAMHET record for each HET group and will be numbered 2, 3, etc. for continuation records).
12 - 14	Non-standard group (heterogen) identifier.
16 - 70	List of synonyms for the HET group - each synonym name will be separated by a semicolon.

Example:

```
SYNHET   CMP CYCLIC AMP; CAMP
```

– *HET Dictionary*

A dictionary containing descriptions of all HET groups found in PDB entries is currently being constructed. Included in the dictionary are the HET, FORMUL, NAMHET, and SYNHET records for each heterogen. In addition, atom connectivity is described by CONECT records that give atom names. When ready, a copy of the dictionary will be included in the directory /**pub** from the FTP server and will be updated as needed.

The table is being constructed in collaboration with Professor Betty Deroski of Suffolk County Community College and Dr. Clifford Felder of the Weizmann Institute of Science.

Sample description of a HET group:

```
# C4*          HN1      O2-HO2       C2'--C3'
#   \           |        |          /       \
#    C2*--C1*--N1--C1--C2--C3--C4--C1'       C4'
#   /           |           |       \       /
# C3*      O==C--OXT-HXT  N-1,2HN   C6'--C5'
RESIDUE   AHS     50
CONECT      C1'    4 C4   C2'  C6'  H1'
CONECT      C2'    4 C1'  C3' 1H2' 2H2'
CONECT      C6'    4 C1'  C5' 1H6' 2H6'
CONECT      C3'    4 C2'  C4' 1H3' 2H3'
CONECT      C5'    4 C6'  C4' 1H5' 2H5'
CONECT      C4'    4 C3'  C5' 1H4' 2H4'
CONECT      C4     4 C3   C1' 1H4  2H4
CONECT      C3     4 N    C4   C2   H3
CONECT      N      3 C3  1HN  2HN
CONECT      C2     4 C3   O2   C1   H2
CONECT      O2     2 C2   HO2
CONECT      C1     4 C2   N1  1H1  2H1
CONECT      N1     3 C1   C    C1*
CONECT      C      3 O    N1   OXT
CONECT      O      1 C
CONECT      OXT    2 C    HXT
CONECT      C1*    4 N1   C2* 1H1* 2H1*
CONECT      C2*    4 C1*  C3*  C4*  H2*
CONECT      C3*    4 C2* 1H3* 2H3* 4H3*
CONECT      C4*    4 C2* 1H4* 2H4* 4H4*
CONECT      H1'    1 C1'
CONECT     1H2'    1 C2'
CONECT     1H2'    1 C2'
CONECT     1H6'    1 C6'
CONECT     1H6'    1 C6'
CONECT     1H3'    1 C3'
CONECT     1H3'    1 C3'
CONECT     1H5'    1 C5'
CONECT     1H5'    1 C5'
CONECT     1H4'    1 C4'
CONECT     1H4'    1 C4'
CONECT     1H4     1 C4
CONECT     2H4     1 C4
CONECT     1HN     1 N
CONECT     2HN     1 N
CONECT      H3     1 C3
CONECT      H2     1 C2
CONECT      HO2    1 O2
CONECT     1H1     1 C1
CONECT     2H1     1 C1
CONECT      HXT    1 OXT
CONECT     1H1*    1 C1*
CONECT     2H1*    1 C1*
CONECT      H2*    1 C2*
CONECT     1H3*    1 C3*
CONECT     2H3*    1 C3*
CONECT     3H3*    1 C3*
CONECT     1H4*    1 C4*
CONECT     2H4*    1 C4*
CONECT     3H4*    1 C4*
END
HET    AHS  I   4      50
NAMHET     AHS N-ISOLEUCYL-N-CARBOXY-(2-HYDROXY-3-AMINO-4-
NAMHET   2 AHS CYCLOHEXYL-BUTYL) AMINE
SYNHET     AHS AZOHOMOSTATINE DIPEPTIDE ISOSTERE; CP-69,799
FORMUL   2  AHS   C15 H30 N2 O3
```

– *Whom to Contact*

General inquiries pertaining to the contents and management of the archive may be sent to Enrique Abola (abola1@bnl.gov).

New Electronic Deposition Form

A new Electronic Deposition Form which has been tested by a number of depositors is now available for general use. All entries submitted to the PDB must now use this form. This form is expected to simplify preparation of your deposition. It will also ensure that all data necessary for a complete PDB entry is provided in the initial submission.

The new Deposition Form is available from the FTP server in the file /**pub**/**dep_form.txt**. You should retrieve this file, edit it on your computer and return it electronically to the PDB along with all associated data files. We request that you use FTP rather than electronic mail to upload these files. For explicit instructions on uploading to the PDB, see the PDB FAQ available from the FTP server in the files /**pub**/**faq.ps** (PostScript) and /**pub**/**faq.txt** (text). Both the Electronic Deposition Form and the PDB FAQ are also easily obtained *via* Gopher or WWW.

The Electronic Deposition Form begins with a description of the standard structure of a PDB coordinate file and provides guidelines for preparing data for deposition. These guidelines should be studied and followed carefully for the most appropriate representation of your experiment. Additional information not included in PDB's earlier Deposition Form is now being requested. These data items will help our staff significantly in preparing an entry for distribution.

The Electronic Deposition Form leads you through the needed material from your address information through the description of the experiment, crystallographic data, and secondary structure information. Secondary structure information is now being prepared by PDB using the Kabsch and Sander algorithm [Dictionary of Protein Secondary Structure: Pattern Recognition of Hydrogen-bonded and Geometrical Features. *Biopolymers* **22**, 2577-637 (1983)]. This makes submission of HELIX, SHEET, and TURN records optional; however, we will continue to include your secondary structure specifications if you wish to provide this information.

Below is a sample of a completed compound section.

```
COMPOUND (COMPND)

    Present a brief title for the experiment. This is analo-
    gous to the title of a journal article.

    Repeat the block for each molecule of a macromolecular
    complex. Separate multiple synonym names or ligands with
    semicolons. (EC stands for Enzyme Commission.)

Title to be used for this entry: P-hydroxybenzoate hydroxy-
lase mutant with FAD and 2,4-dihydroxybenzoic acid.

    //                          \\
    molecule name      : P-hydroxybenzoate hydroxylase
    synonyms           : PHBH
    EC number          : 1.14.13.2
    engineered mutation: Cys 116 replaced by ser
    ligands            : FAD; 2,4-dihydroxybenzoic acid
    other details      :
    \\                          //
```

User Group News

The PDB User Group is pleased to announce that the directory with coordinates for full biological units is now available from the PDB FTP server. This was the second of our user-initiated initial priorities (the first one was the on-line pending_waiting list for newly-deposited entries), both of which have now been implemented. The User Group greatly appreciates the effort that has been put into generating these files by the PDB staff. These files were generated by Enrique Abola and Mingyu Xu from the PDB and by John Rose of the University of Pittsburgh. We hope that these full-molecule files will be useful for making teaching examples and for study of important subunit interactions by those not using crystallographic software (or, even, for those who are just lazy or can't take the time to sort it out for themselves!).

The directory is available from the FTP server as the file **/user_group/biological_units**. It contains expanded PDB entries with coordinates for the full biological unit (that is, the functional molecule), for cases where there is internal symmetry and only part of the molecule was in the standard PDB entry. For example, standard PDB entries for hemoglobins contain only one alpha-beta dimer, while the entries here contain the full alpha2,beta2 tetramer. To indicate this difference, the files have names starting with bio and ending in .pdb. For example, file pdb1rop.ent contains only the helix-hairpin monomer of ROP protein, while bio1rop.pdb has the full 4-helix bundle. Header information in the file explains its origin and the symmetry elements used to generate it. At present this directory is experimental. It contains many, but not all, of the useful cases (for instance, viruses will be tackled soon).

Please let us know if we have missed a molecule that would be useful to you, or if the header information provided is not sufficient (send e-mail to pdbusrgp@suna.biochem.duke.edu). Also, if these files turn out to be a major help to you, please let us know.

WWW Access to PDB

The World Wide Web (WWW) home page for PDB is accessible using the document URL **http://www.pdb.bnl.gov/**. Starting with links to the PDB FTP and Gopher servers, the WWW home page includes links to many important PDB files and tools, as well as other useful databases and information servers.

WWW is a single, consistent user interface to many of the information retrieval protocols found on Internet today including ftp, telnet, nntp, wais, and gopher. WWW understands the different data formats used by these protocols, including ascii, gif, postscript, dvi, and texinfo. It also adds a new multimedia protocol (HyperText Transfer Protocol or http) and data format (HyperText Markup Language or html).

A very popular WWW client is NCSA Mosaic, freely available for X Windows, Macintosh, and Microsoft Windows platforms. Mosaic clients communicate with WWW servers as well as with more traditional Internet protocols such as ftp, gopher, and wais. Hyperlinks to PDB appear on a growing number of WWW home page locations.

To connect to the WWW home page at PDB from Mosaic, select the FILE pulldown menu item, then OPEN URL. Type **html://www.pdb.bnl.gov**, hit the <return> key or the OPEN button, and the PDB WWW home page will appear. You can choose from many options including FTP and Gopher, search the pending_waiting list, view the latest copies of important documents, visit other databases, and much more.

To add the PDB WWW home page to your Mosaic Hotlist for quick access, select Navigate from the pulldown menu and then Add Current To Hotlist. Or you can select Navigate, Hotlist, and then Add Current.

Information on setting up Mosaic to run on your computer is available *via* anonymous FTP from **ftp.ncsa.uiuc.edu**. Pertinent files are in the following subdirectories:

Mac/Mosaic....................................for Macintosh
PC/Mosaic ....for Microsoft Windows
Web/Mosaic-binaries ....for X Windows
Web/Mosaic-source ....for X Windows

There are README files at various levels to provide information on what is available and on how to set up your Mosaic client. Also available from your bookstore or library is a useful book on the subject: “Mosaic Quick Tour” by Gareth Branwyn.

The following is a list of some interesting and useful home pages accessible using WWW. As you use Mosaic, you will find many other intriguing locations.

URL	*Title*
http://www.nih.gov/	National Institutes of Health
http://expasy.hcuge.ch/sprot/sprot-top.html/	SWISS-PROT Protein Sequence Database
http://www.nih.gov/molbio/	Molecular Biology Databases specifically related to DNA and protein sequence database holdings
http://expasy.hcuge.ch/	ExPASy Molecular Biology server
http://golgi.harvard.edu/biopages.html/	Comprehensive Biosciences Index from Keith Robison
gopher://vm1.hqadmin.doe.gov/1/	Department of Energy Gopher
http://gdbwww.gdb.org/	Genome Data Base
http://kaktus.kemi.aau.dk/	The O Protein Crystallographic Package
http://dapsas1.weizmann.ac.il/	Biological Computing Division at Weizmann Institute of Science
http://ndbserver.rutgers.edu:80/	Nucleic Acid Database Project at Rutgers University
http://www.prosci.uci.edu/	Protein Science Web Server
http://cui_www.unige.ch/meta-index.html/	Search Engines for WWW
http://www.mit.edu:8001/people/mkgray/compre3.html/	Comprehensive list of HTTP sites
http://www.sgi.com/	Silicon Surf from SGI
gopher://pdb.pdb.bnl.gov/11/FTP/pub/pdbbrowse/	PDB-Browse - a browser for Unix systems
gopher://pdb.pdb.bnl.gov/11/Software/PDBShell/	PDB-Shell - a browser for Windows (PC) systems
gopher://pdb.pdb.bnl.gov/11/Software/Procheck/	Procheck information
ftp://pdb.pdb.bnl.gov/pub/other-software/	Other software
http://bnlstb.bio.bnl.gov:8000/	Structural Biology at the BNL Biology Department
http://www.tc.cornell.edu/~richard/AChE.html/	Acetylcholinesterate: Nature’s Vacuum Cleaner
http://www.public.iastate.edu/~pedro/research_tools.html/	Pedro’s Research Tools
http://csdvx2.ccdc.cam.ac.uk/	Cambridge Crystallographic Data Centre
http://www.sander.embl-heidelberg.de/dssp/	The DSSP program and database

Release Schedules

The PDB has been hard at work improving the lag time between quarterly release dates and shipment of the CD-ROM set. Some of the major steps which must take place between a release date and shipment of the CD-ROM include:

- building the release on FTP
- building an image of the CD-ROM files on hard disk
- dumping the image to tape and checking for accuracy
- sending image tape to CD-ROM duplication vendor for pre-mastering
- checking worm checkdisk from vendor for accuracy
- checking final CD-ROM set from vendor for accuracy
- packaging and shipping CD-ROM set

We have had a schedule in place for the past several months that will eventually allow us to ship the CD-ROM at the end of the release month. For instance, CD-ROM shipment for the October 1995 release should begin on October 31. Implementation of these changes will take time since the many steps needed do not allow for gross adjustments in schedules.

What follows are previous and projected CD-ROM shipment dates. As you can see, the scheduled dates versus actual dates sometimes differ, most often due to CD-ROM production problems beyond our control. More importantly, you can see that eventually CD-ROM shipment will begin at the end of the release month.

		scheduled	*actual*
1994	January	04/26/94	04/25/94
	April	07/07/94	07/08/94
	July	09/26/94	09/26/94
	October	12/12/94	na
1995	January	03/03/95	na
	April	05/22/95	na
	July	08/11/95	na
	October	10/31/95	na

Discontinuation of Tape Distribution

Due to the very small number of tapes ordered, PDB has discontinued distribution of these media. We have contacted those few who had previously been interested in tape format, discussed the situation, and all have decided that the CD-ROM would be an acceptable and perhaps even more convenient format.

Revised Newsletter Format

As discussed in both the January and April 1994 Newsletters, PDB has eliminated the tables of newly released and newly deposited entries in this and future Newsletters.

These tables will appear individually in the Full Tables document. The Full Tables document accompanies each order shipped and is available from the FTP server in the /**newsletter** subdirectory in both PostScript and ASCII formats. A printed copy may be obtained upon request.

Newsletter Distribution Changes

Due to the ease of retrieving the Newsletter *via* Internet (FTP, Gopher, and WWW in the /**newsletter** subdirectory in PostScript and ASCII formats) and the very high number of printed Newsletters being distributed *via* the postal system, PDB has decided to re-initialize the Newsletter mailing list. This is expected to take place just after distribution of the January 1995 Newsletter.

Anyone who wishes to receive printed copies of the Newsletter will be able to do so by specifically requesting to be on the Newsletter mailing list. Intentions are to have a form available in January which can be completed by users and returned to us electronically or *via* the postal system. Plans are to have this re-initialized mailing list operational for the April 1995 Newsletter distribution. Please stay tuned for further developments.

New Entry Tracking System

Many depositors, users and journals that refer to PDB have asked us to provide a mechanism for checking the status of entries from depositor's initial contact with PDB to the time when the entry is released. This is now possible due to the establishment of tracking codes, one of which is given to each depositor upon deposition of an entry. This code can be used by depositors, users and journals to track an entry's progress until full release.

Upon receipt of coordinates for an entry, PDB issues a tracking number and sends a letter to the primary and secondary contacts that includes this number. The format of the tracking number is a number preceded by the letter T (*e.g.,* T9999).

The tracking number is published, for every entry that is currently being processed by PDB, in the file **/pub/pending_waiting.list** which is available from the FTP server. If you use Gopher to access PDB, this file is indexed by tracking number, author, and compound name in order to speed searches.

Status for each pending and waiting entry is issued from the following list:

INCOMPLETE	-	if PDB is waiting for additional materials or information from depositor to make a complete deposition
PROCESSING	-	if PDB is checking and verifying entry
DEPOSITOR	-	if entry is with depositor for approval
REVIEW	-	if entry is undergoing final review
REL	-	if entry is in current release
HLD	-	if entry is on hold at present time

This number is not to be construed as an ident code. Its purpose is to speed up our searches, make our responses to inquiries faster and more accurate, and allow depositors to track the progress of their entries.

PDB issues the PDB ident code, required by most journals, only after running some preliminary checking programs and receiving all required documentation, including a (p)reprint of the JRNL reference (or a statement indicting that the entry does not correspond to a specific publication).

Access to PDB

– 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, only into the directory /**new_uploads**. Those using VMS may need to place quotes around file names.

– 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.

– 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.

– 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**.

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 are not 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 *via* postal mail to:

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

CD-ROM — ISO 9660 Format *$300.00*

(tax and shipping charges not applicable)

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
Kathleen Klose (617-353-7123)
klose@darwin.bu.edu

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

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 (+1 31-80-653386)
noordik@caos.caos.kun.nl

*CCDC
Cambridge Crystallographic Data Centre
Cambridge, United Kingdom
David Watson (+1 44-223-336394)
dgwl@chemcrys.cam.ac.uk

CINECA
NE Italy Interuniversity Computing Center
Casalecchio di Reno (BO), Italy
Laura Setti (+1 39-51-6599478)

ICGEB
International Centre for Genetic Engineering and Biotechnology
Trieste, Italy
Sandor Pongor (+1 39-40-3757300)
pongor@icgeb.trieste.itasltc0@icineca.cineca.it

EMBL
European Molecular Biology Laboratory
Heidelberg, Germany
Hans Doebbeling(+1 49-6221-387-247)
hans.doebbeling@embl-heidelberg.de

INN
Israeli National Node
Weizmann Institute of Science
Rehovot, Israel
Leon Esterman (+1 972-8-343934)
lsestern@weizmann.weizmann.ac.il

*JAICI
Japan Association for International Chemical Information
Tokyo, Japan
Hideaki Chihara (+1 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 (+1 886-35-776085; 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 (+1 44-865-784600)
steve@gardner.demon.co.uk

*Osaka University
Institute for Protein Research
Osaka, Japan
Yoshiki Matsuura (+1 81-6-879-8605)
matsuura@protein.osaka-u.ac.jp

Pittsburgh Supercomputing Center
Pittsburgh, Pennsylvania
Hugh Nicholas (412-268-4960)
nicholas@cpwpsca.bitnet

*Protein Science
Princeton, New Jersey
Joseph Villafranca (609-252-3573)
villafranca@bms.com

SDSC
San Diego Supercomputer Center
San Diego, California
Lynn Ten Eyck (619-534-8189)
teneyckl@sdsc.bitnet

SEQNET
Daresbury Laboratory
Warrington, United Kingdom
User Interface Group (+1 44-925-603351)
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: .... +1 516-282-3629
Facsimile:... +1 516-282-5751

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

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

First Class