Protein Data Bank

Quarterly Newsletter

Release #79 January 1997

January 1997 CD-ROM Release

5402 Released Atomic Coordinate Entries

Molecule Type

4787	proteins, peptides, and viruses
217	protein/nucleic acid complexes
386	nucleic acids
12	carbohydrates

Experimental Technique

147	theoretical modeling
780	NMR
4475	diffraction and other

The total size of the atomic coordinate entry database is 2186 Mbytes uncompressed.

Table of Contents

page

What's New at the PDB ... 1
Update on AutoDep ... 2
Archive Management ... 2
— News and Updates ... 3
Preventing Post-AutoDepic Depression ... 4
Archive of Obsolete PDB Entries: A Joint Initiative of the San Diego Supercomputer Center and the PDB ... 4
PDBsum: Summaries and Structural Analyses of PDB Data Files ... 5
The CCP4 Program Suite for Protein Crystallography ... 6
— Introduction ... 6
— Overview of the Suite ... 6
— Recent Changes ... 6
— Availability and Distribution ... 7
PDBCONS Program ... 7
Distance Education in Collaboration with the PDB ... 7
— Introduction ... 7
— Distance Education and PPS95 ... 8
— New Developments: BIOTOOL96 ... 8
— Conclusion ... 8
The 24th Aharon Katzir-Katchalsky Conference: BIOINFORMATICS, November 17 - 21, 1996, Jerusalem, Israel ... 9
— Opening Presentations ... 9
— Panel Discussion ... 9
— Protein Modeling and Design ... 10
— Structures ... 10
— Conclusion ... 11
Notes of a Protein Crystallographer ... 12
— Recuerdos de la Alhambra: Remembrances of the Alhambra ... 12
Web Sites Referenced in the January 1997 PDB Newsletter ... 13
Affiliated Centers and Mirror Sites ... 14
Related WWW Sites ... 16
Order Form ... 17
PDB Staff, Consultants, Support, Access, and FTP Directory for Entries.. 19

Internet Sites

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

What's New at the PDB

— Joel L. Sussman

The 24th Aharon Katzir-Katchalsky Conference, entitled Bioinformatics <---> Structure, in honor of the 25th anniversary of the PDB and the 10th anniversary of the Swiss-Prot Database, was held in Jerusalem, Israel, on November 17-21, 1996. We were extremely fortunate to have the meeting opened by Ezer Weizman, President of the State of Israel, and Professor Ephraim Katchalsky-Katzir, the fourth President of Israel. The plenary lectures by Professors Leroy Hood and Tom Blundell were followed by four days of talks on a variety of topics, many at the cutting edge of research.

Over four hundred participants from about thirty countries participated. The meeting focused on the interplay of molecular and structural biology and the use of computer tools, especially databases, accessible through the World Wide Web. This interplay has had an extraordinary impact on basic research in biology as well as on the rapidly expanding field of biotechnology. Some of the main areas discussed — benefits in the areas of drug design, evolution studies, and macromolecular structure prediction — have emerged from studying the databases of sequences and three-dimensional structures of biological macromolecules and the generalizations which were obtained thereof.

Suggestions were made as to possible future benefits to be derived from these studies, as well as novel ways to examine the enormously large mass of information continuously being accumulated in these databases. Suggestions included several global high-throughput, information-intensive activities already underway: the human genome project; other animal, plant, and microbial genome projects; and the brain project.

What was seen repeatedly throughout the meeting was how valuable 3-D structural databases such as the PDB, the Cambridge Structural Database (CSD), and the Nucleic Acid Database are to scientists in diverse fields extending from genomics, molecular biology, and pharmacology to theoretical and organic chemistry. It was also clear that the demand for this structural information is increasing incredibly fast.

In order to make this information available more rapidly, we plan to institute new procedures at the PDB during the next few months which will permit depositors to "precheck" their depositions before they are submitted. This will aid in making their entries less prone to errors and will also greatly speed up the entire submission process. Authors will then be encouraged, *via* the PDB's new AutoDep system, to release their high-quality entries at the time of submission so that their structural information will be immediately available within PDB's worldwide database. Watch PDB's Web site (http://www.pdb.bnl.gov) for further information pertaining to prechecking.

The AutoDep system, found at http://www.pdb.bnl.gov, was released on October 15, 1996 and accounted for well over 50 percent of the December submissions to the PDB. Scientists depositing entries to the PDB are greatly encouraged to make use of this facility — both simplifying their submission process and making it possible for their entries to be available much more rapidly.

More information pertaining to this Conference may be found in the Volume 1 Supplement issue of "Folding and Design," 1996, Fersht and Cohen, Eds., and in Dr. Ellie Adman's article included in this Quarterly Newsletter. Photographs from the Conference, as well as the abstracts, may be found on the Conference Web page (http://www.pdb.bnl.gov/pdb25sp10/index.html).

Update on AutoDep

— Nancy Manning

PDB's Web-based submission program, AutoDep, was released on October 15, 1996. By the end of December, well over 50 percent of all submissions received by the PDB were being created and submitted through AutoDep.

Below are the AutoDep Submission Rates as of January 27, 1997.

Period	AutoDep versus All Submissions
Oct. 15 - 31, 1996	28%
Nov. 1 - 30, 1996	48%
Dec. 1 - 31, 1996	63%
Jan. 1 - 27, 1997	57%

Depositors have responded enthusiastically to our effort to simplify the submission process. PDB's Help Desk is available to assist our depositors as needed; problems encountered by depositors have been quickly resolved. For a list of bug fixes and updates to the program, see the AutoDep Release Notes.

AutoDep, the Release Notes, and related files and links may be accessed from PDB's Home Page by choosing "Submitting Data to the PDB" then choosing "AutoDep - PDB's New Web-based Submission Tool."

We encourage your comments and suggestions. Please send them to the PDB Help Desk at pdbhelp@bnl.gov.

Archive Management

— Enrique Abola and Nancy Manning

What is a molecule? What is a chain? How many do I have in this structure? Many of our depositors are asking these questions as they prepare their structural entries for submission to the PDB. In order to properly represent the structure in PDB format, and the sequence in the SEQRES records, the correct answers must be determined.

Defining a molecule for use in a database is, of course, not straightforward. The PDB uses an operational definition that typically follows the representation used by the depositor. A molecule in the PDB may be composed of one or more chains — a chain being a sequence of covalently contiguous residues. For example, when hemoglobin contains two alpha and two beta chains in the asymmetric unit, it is represented in the PDB file as one molecule with four chains.

The following is taken from the article "To be Ala or Not to be Ala? That is the Question." by Frances Bernstein, which appeared in the October 1995 PDB Quarterly Newsletter.

The coordinates of a molecule in a PDB file may contain gaps due to ambiguous or uninterpretable regions in electron density maps or due to residues missing for other reasons. PDB distinguishes among the situations which give rise to these gaps. Three cases are illustrated here in which six residues of the known sequence are missing from the coordinate records.

- The atoms of residues 45 - 50 of a chain could not be located in the electron density maps. Residues 45 - 50 are included in the SEQRES records even though no coordinates are present in the entry's ATOM records. The oxygen atom occurring just before the gap is named O. If the refinement procedure required a terminal oxygen atom in the last residue before the gap, either one of the two oxygens is removed or one may be renamed as N of the next residue — the depositor should do this before submitting the coordinates.
- A recombinant experiment was done resulting in residues 45 - 50 being deleted from the protein and residue 51 forming a peptide bond with residue 44. Residues 45 - 50 are not included in the SEQRES records because they were not present in the protein studied.
- Residues 45 - 50 were excised from the protein, resulting in two chains instead of one. Residues 45 - 50 do not appear in the SEQRES records, and the protein will be presented as two separate chains in the PDB file.

The first of these examples illustrates another data representation problem that arises when there are gaps in chains due to disorder. Our current practice assumes that a proper assignment of the element type (*i.e.*, N or O) can be made. It also assumes that proper assignment of the chain direction can be made in spite of the experimental uncertainties. Thus, the atom names given are really arbitrary and misleading. A more appropriate naming practice would be to follow the convention used for GLN and ASN where AE1 and AE2, and AD1 and AD2 are used for the atoms in the carbamoyl group when identification of the O and N atoms is not possible. Please send your comments and suggestions regarding issues and problems discussed here as well as anything else regarding the contents of PDB entries to Enrique Abola (abola1@bnl.gov) or Nancy Manning (oeder@bnl.gov).

For further information, see "Preparing Your Atomic Coordinate Data for Submission" on our Web page, under "Submitting Data to the PDB."

— News And Updates

World Wide Web Site

The PDB's WWW Home Page (http://www.pdb.bnl.gov/) contains two new features that we invite you to explore. The first is the new PDB WWW Bulletin Board, and the second is the Thread of PDB's Listserver. We invite you to make use of the Bulletin Board forum for airing questions and comments relating to the PDB, macromolecular structures, software, and related issues. The Thread is a convenient means of reviewing postings to the PDB Listserver.

Enhanced Remark Texts

On December 20, 1996, the PDB released revised text to remark 500 and added a new remark, number 217, to be used for solid state NMR experiments. The following announcement appeared on the PDB's Bulletin Board:

Listed below are changes to the PDB Contents Guide. Please note that whenever changes to this document are made, we update the format version number. Changes such as those listed below are denoted by a change in the fractional part of the version number. All significant changes will follow our Format Change Policy and will be denoted by a whole number change (*e.g.*, 2.2 -> 3.0). See the Contents Guide for more details.

In response to recommendations from several depositors, we have updated our program which checks for close contacts. Because these are reported in remark 500, PDB has changed the free text field of REMARK 500 when it refers to close contacts. It has been changed from this:

```
REMARK 500
REMARK 500 GEOMETRY AND STEREOCHEMISTRY
REMARK 500 SUBTOPIC: CLOSE CONTACTS
REMARK 500
REMARK 500 THE FOLLOWING ATOMS THAT ARE RELATED BY CRYSTALLOGRAPHIC
REMARK 500 SYMMETRY ARE IN CLOSE CONTACT. SOME OF THESE MAY BE ATOMS
REMARK 500 LOCATED ON SPECIAL POSITIONS IN THE CELL.
REMARK 500
REMARK 500 DISTANCE CUTOFF: 2.2 ANGSTROMS
REMARK 500
REMARK 500 ATM1 RES C SSEQI ATM2 RES C SSEQI SSYMOP DISTANCE
```

to this:

```
REMARK 500
REMARK 500 GEOMETRY AND STEREOCHEMISTRY
REMARK 500 SUBTOPIC: CLOSE CONTACTS
REMARK 500
REMARK 500 THE FOLLOWING ATOMS THAT ARE RELATED BY CRYSTALLOGRAPHIC
REMARK 500 SYMMETRY ARE IN CLOSE CONTACT. SOME OF THESE MAY BE ATOMS
REMARK 500 LOCATED ON SPECIAL POSITIONS IN THE CELL. ATOMS WITH
REMARK 500 NON-BLANK ALTERNATE LOCATION INDICATORS ARE NOT INCLUDED
REMARK 500 IN THE CALCULATIONS.
REMARK 500
REMARK 500 DISTANCE CUTOFF:
REMARK 500 2.2 ANGSTROMS FOR CONTACTS NOT INVOLVING HYDROGEN ATOMS
REMARK 500 1.6 ANGSTROMS FOR CONTACTS INVOLVING HYDROGEN ATOMS
REMARK 500
REMARK 500 ATM1 RES C SSEQI ATM2 RES C SSEQI SSYMOP DISTANCE
```

In addition, because we have received our first structure solved by solid-state NMR, we have added a new standard remark, number 217, which will appear in all solid-state NMR entries.

```
REMARK 217
REMARK 217 SOLID STATE NMR STUDY
REMARK 217 THE COORDINATES IN THIS ENTRY WERE GENERATED FROM SOLID
REMARK 217 STATE NMR DATA. PROTEIN DATA BANK CONVENTIONS REQUIRE THAT
REMARK 217 CRYST1 AND SCALE RECORDS BE INCLUDED, BUT THE VALUES ON
REMARK 217 THESE RECORDS ARE MEANINGLESS.
```

In order to properly annotate the entries, REMARK 4 will now refer to the format as described in Contents Guide version 2.2.

```
REMARK 4
REMARK 4   XXXX COMPLIES WITH FORMAT V. 2.2, 16-DEC-1996
```

The updated Contents Guide, which includes these changes, will be placed on the Web as soon as possible.

Proposed Changes to PDB Format V. 2.0

We would also like to remind our readers of proposed changes to PDB format that first appeared in the October 1996 Quarterly Newsletter.

In accordance with PDB's Format Change Policy, found at URL http://www.pdb.bnl.gov/format_change_policy.html, we are within the sixty-day open discussion period during which we will entertain comments and suggestions regarding these changes. Please send comments to Enrique Abola (abola1@bnl.gov) or Nancy Manning (oeder@bnl.gov). Discussion on the PDB Listserver and Bulletin Board is encouraged as well.

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

Hydrogen Atom Names in Amino Acids

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

Space Group Symbol for Monoclinic Crystals

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

Representation of Modified Nucleic Acid Residues

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

Preventing Post-AutoDepic Depression

Gerard J. Kleywegt, Dept. of Molecular Biology, Uppsala University, Uppsala, Sweden (gerard@xray.bmc.uu.se).

You have finally done it! You have managed to change all the red "X"s on your AutoDep page into green "V"s. Now you hit the Submit button, sit back, and promise yourself a beer. Two minutes later your terminal beeps. E-mail . . . from the PDB. "Must be a confirmation of my submission," you think. Guess again . . . a little daemon sent mail to tell you that the *working* PDB file that you submitted is not acceptable: "PDB's automatic data checking procedures, begun on Monday, November 25, 1996 at 19:10, detected errors in the coordinate data file of your deposition of AutoDep ID BNL-####. Please correct the errors listed below, upload the revised coordinate file, and restart the validation procedure by pressing the 'Send full deposition to the PDB!' button." You pull out what little hair remains on your skull.

A few of the problems that you may encounter are listed below. All of them can be fixed in myriad ways, using anything from a text editor (not recommended) to a PDB file-manipulating program. One such program is MOLEMAN2, which is available free of charge to academics from rigel.bmc.uu.se in the directory pub/moleman2. The manual for the program may be accessed at http://alpha2.bmc.uu.se/~gerard/manuals/moleman2_man.html.

- O uses two columns of the ATOM and HETATM records, which are supposed to be blank (to store the chemical element number); the PDB does not like this. Simply reading and writing the PDB file with MOLEMAN2 will fix this.
- I like to call my pyroglutamate residues PYR, but the PDB gospel dictates that they should be called PCA. Similarly, water oxygen atoms ought to be called "O" rather than "O1." Residue and atom names can both be changed in MOLEMAN2 (PDb NAme command).
- Anything that's not a *bona fide* ATOM must be called HETATM. MOLEMAN2 contains a command (PDb HEtero Deduce) to automatically fix this.
- The PDb CHemical command will fill in the last four columns of your ATOM and HETATM cards, namely the chemical element name and a guess at the formal charge.

MOLEMAN2 was written as a tool-kit for practicing crystallographers who often need to convert between "standard" O PDB files, "standard" X-PLOR PDB files, "standard" CCP4 PDB files and "standard" PDB PDB files. Additional features include:

- The ability to automatically split a PDB file into separate segments for use with X-PLOR; this option will also generate an X-PLOR "generate.inp" file (which even includes patches for any disulfides); and if the infamous OT1 and/or OT2 atoms required by X-PLOR are absent in your file, coordinates will be generated for them automagically by MOLEMAN2.
- An assortment of statistics concerning your model (*e.g.*, pertaining to temperature factors and distances) as well as some simple analysis tools for proteins (Ramachandran plots, *etc.*).
- Numerous options for manipulating, generating, or changing chain names and segment identifiers (also useful prior to depositing models).
- Several coordinate-, B-factor-, and occupancy-manipulation tools; some commands to generate ideal alpha-helices and beta-strands; a few simple sequence-analysis tools; and a number of O-related commands (*e.g.*, for generating dictionaries or water-fitting macros).

Archive of Obsolete PDB Entries: A Joint Initiative of the San Diego Supercomputer Center and the PDB

Helge Weissig and Phil Bourne, San Diego Supercomputer Center, San Diego, CA, USA (bourne@sdsc.edu or helgew@sdsc.edu; http://www.sdsc.edu/PDBobsolete).

As of October 15, 1996, the PDB (http://www.pdb.bnl.gov) holds 4873 released atomic coordinate entries. Although in most cases these entries constitute the original versions of the structure models, several structures in the PDB have been replaced up to four times (*e.g.*, aspartate carbamyltransferase and ferredoxin) and only the latest version is maintained as part of the PDB distribution (*i.e.*, 5AT1 and 5FD1, respectively). Entries withdrawn from the distribution are marked obsolete and the OBSLTE header record is added by the PDB staff. A list of obsolete entries and their successors (which are marked with the SPRSDE header record) is available as the file ftp://ftp.pdb.bnl.gov/pub/resources/index/obsolete.dat.

According to PDB policy, only the primary author who submitted an entry has the authority to withdraw it. Entries can be deleted or replaced by other entries. One obsolete entry may be replaced by multiple new entries and, conversely, several old entries can be replaced by a single new entry. The changes made when a structure is replaced are sometimes minimal and sometimes substantive. Other structures (*e.g.*, the CD, ZN metallothionein isoform II) have been deleted and are no longer available. Below are some of the reasons why it is considered valuable to be able to retrieve obsolete entries:

- So that the correct version of a structure may be related to the original publication.

- So that different versions of the same structure may be compared in a quantitative way.
- So that new software may be directly compared to older software by using the same structures as part of a quantitative test.

Access to replaced structures is now possible through the archive of obsolete PDB entries established at the San Diego Supercomputer Center (http://www.sdsc.edu/PDBobsolete) as a joint initiative with the PDB. Currently, 307 entries are on file and may be searched *via* the Harvest search system developed at the University of Colorado. Full text keyword as well as PDB entry ID searches are supported. The summarizer used for the indexing system has been designed to be expandable to allow more powerful query capabilities to be added in the future. Additionally, the entries within the archive can be browsed by their ID and their header lines.

The chronological information for each entry, its successors, and its precursors is presented in a graphical fashion. This graphic represents an image map, enabling the download of single structures for display with molecular structure viewers such as RasMol. Additionally, comparisons between the actual coordinate entries are possible (works best with frames-capable Web browsers but is also feasible with HTML 2.0-compliant software). Entries are hyperlinked through the OBSLTE and SPRSDE PDB record types.

Limited statistical analysis of the archived entries shows that the average lifetime for these obsolete entries was about four and one-half years. The total lifetimes ranged between almost twenty-one years (1RNS, ribonuclease S) and two months (151C, cytochrome C551). The oldest entry (1B5C, cytochrome B5) was deposited on August 10, 1972 and subsequently replaced by 2B5C on January 16, 1978, which in turn was replaced by 3B5C on January 15, 1991. The latest entry (1TSA, azurin) was deposited with the PDB on September 22, 1995 and replaced by 2TSA on November 8, 1996.

E-mail comments on the archive are welcome and should be directed to bourne@sdsc.edu.

PDBsum: Summaries and Structural Analyses of PDB Data Files

Roman A. Laskowski, Birkbeck College; E. Gail Hutchinson, Alex Michie, and Andrew C. Wallace, University College; and Janet M. Thornton, University College and Birkbeck College; London, England (thornton@ biochemistry.ucl.ac.uk).

We have recently set up a WWW site, called PDBsum, which gives various analyses of the structural features of every entry in the PDB. The analyses include the structural motifs identified by PROMOTIF [1], summary PROCHECK analyses [2], and schematic diagrams of ligands and their interactions, generated by the HBPLUS [3] and LIGPLOT [4] programs. The entries are also cross-linked both to our own and other Web-based databases.

In PDBsum, each PDB entry has its own “page.” In the top left corner is a thumbnail image of the molecule(s) in the PDB file, rendered using Raster3D [5,6]. Protein chains are depicted in a schematic manner according to their secondary structural content; DNA and RNA molecules are shown using a capped-stick representation, while any ligands and metal ions are shown as all-atom models. Different protein and DNA/RNA chains are shown in different colors. For the smaller molecules (and complexes) the schematic diagrams are also available as VRML files which allow you to view and move them interactively using any VRML browser (*e.g.*, SGI’s Webspace).

Below the thumbnail image is a summary of the header information in the PDB file (structure, source, resolution, authors, *etc.*), together with links to various databases. The databases include our own CATH classification of protein domain structural families (described in the October 1996 issue of the PDB Quarterly Newsletter) and the E.C. -> PDB Enzyme Structures Database. The former, available at http://www.biochem.ucl.ac.uk/bsm/cath, allows you to quickly identify proteins which are structurally similar at the domain level to the protein you are viewing. The latter, available from http://www.biochem.ucl.ac.uk/bsm/enzymes, allows you to identify enzymes that are functionally related *via* the hierarchical classification of the E.C. numbering system. Links to other databases include: SWISS-PROT, PDB itself, MMDB at NCBI, the NDB at Rutgers, and SCOP.

Below these links is a summary for each chain and ligand in the PDB file. For each protein chain, a schematic “wiring diagram” shows a one-dimensional plot of the key secondary structural elements along the protein’s sequence. The elements, computed by the PROMOTIF program [1], include helices, strands, β-turns, γ-turns, β-hairpins, and disulphide bridges. Further detailed breakdowns (and schematic diagrams) of each of these items, together with the locations of any β-bulges, β-sheets and βαβ-units are also available *via* the relevant links. The active site residues, as identified by the SITE records in the PDB file, are also shown on the wiring diagrams and are listed in separate links. If the chain comprises more than one domain, the CATH classification is given for each one, and the residues in each domain are shown in different colors. A MolScript diagram of each chain may also be generated as a PostScript picture.

For each ligand identified in the PDB file, the following information is given: a schematic diagram of the molecule (as flattened out by LIGPLOT [4]), a three-dimensional image generated by Raster3D [5,6], a LIGPLOT of the ligand’s interactions with the protein, and links allowing you to quickly find other PDB entries containing the same ligand.

Unfortunately, it is no easy matter to automatically and consistently identify which are the ligands and which are not when processing all the files in the PDB. Hopefully, when all the entries conform to some new standard format, programs that attempt to do this will be more likely to meet with success. Similarly, there is no easy way of automatically determining the hydrogen-bonding capabilities of the ligand’s atoms, so some of the interactions with the protein may be missed by HBPLUS, and consequently be absent from the LIGPLOT diagrams. Nevertheless, much useful information can still be gleaned from the ligand data in the PDBsum entries.

We are currently developing additional Web-based databases covering specific areas of interest (*e.g.*, proteins containing heme-groups) which will include links to and from PDBsum. We would also be interested in linking to any such special-interest databases which may already exist.

The PDBsum database is available from University College London at http://www.biochem.ucl.ac.uk/bsm/pdbsum and is mirrored at the PDB at http://www.pdb.bnl.gov/bsm/pdbsum.

1. E.G. Hutchinson and J.M. Thornton, *Protein Science* **5**, 212-220 (1996).
2. R.A. Laskowski, M.W. MacArthur, D.S. Moss, and J.M. Thornton, *J. Appl. Cryst.* **26**, 283-291 (1993).
3. I.K. McDonald and J.M. Thornton, *J. Mol. Biol.* **238**, 777-793 (1994).
4. A.C. Wallace, R.A. Laskowski, and J.M. Thornton, *Prot. Eng.* ***8***, 127-134 (1995).
5. D.J. Bacon and W.F. Anderson, *J. Mol. Graph.* **6**, 219-220 (1988).
6. E.A. Merritt and M.E.P. Murphy, *Acta Cryst.* **D50**, 869-873 (1994).

The CCP4 Program Suite for Protein Crystallography

Martyn Winn, Synchrotron Radiation Dept., CCLRC, Daresbury Laboratory, Warrington, UK (ccp4@dl.ac.uk; http://www.dl.ac.uk/CCP/CCP4/main.html).

— Introduction

CCP4 (Collaborative Computational Project, number 4) was established in 1979 by a group of protein crystallographers from the United Kingdom to help maintain and extend an adequate set of software for protein crystallography, to discuss new techniques and algorithms, and to educate students in the field. The project is now funded by the UK Biotechnology and Biological Sciences Research Council with additional important contributions from industrial companies. CCP4 distributes to the international community an extensive suite of programs encompassing all aspects of macromolecular crystallography. Specifically, the following areas are covered: data scaling and reduction, isomorphous replacement, molecular replacement, map calculation and manipulation, density modification, structure refinement, and structure analysis. The suite is distributed as source code (mostly Fortran 77 and partly ANSI C) with associated installation procedures for VMS and most Unix platforms. A manual, individual program documentation, and examples are also distributed. The distribution and maintenance of the suite is centralized at Daresbury Laboratory.

In addition to the program suite, CCP4 funds a twice-yearly Newsletter and an annual meeting with associated published proceedings. There is also an active e-mail list to provide help and advice to users, and to stimulate discussion.

— Overview of the Suite

Unlike many other packages, the CCP4 suite is designed to be loosely organized, so that it is easy for different developers to add new programs or to modify existing ones without upsetting other parts of the suite. The suite thus consists of a set of separate programs which communicate *via* standard data files, rather than having all operations integrated into one huge program. This is the approach successfully taken by Unix and now apparently being embraced by some of the large commercial software houses. It has some disadvantages in that complex tasks usually require a script to chain together several programs, therefore, sample scripts are provided to guide the beginner.

There are three basic file formats: namely, MTZ files for reflection data; map files for electron density maps, masks, and images; and PDB files for atomic coordinates. MTZ and map files are binary, while PDB files are ASCII. All files contain essential header information as well as data. Jiffy programs are included in the suite to extract useful information from these files and to interconvert between these formats and other common formats.

The heart of the suite is an extensive and thoroughly-tested set of library routines, covering basic crystallographic and programming operations. In particular, there are routines for reading and writing the standard data files. There are also a small number of machine-specific routines which handle dynamic core allocation, file assignment, *etc*. A data library, containing useful basic data such as space group symmetry operators, atomic form factors, *etc.*, is also provided.

In addition to the libraries, CCP4 distributes approximately one hundred application programs donated by various authors from several countries. New programs are added at a steady rate, as new techniques develop. The philosophy has been to be inclusive; therefore, several programs may be available to do the same task. The components of the whole system are thus a collection of application programs using a standard subroutine library to access standard format files.

— Recent Changes

Version 3.0 of the suite was released in April 1996 followed by version 3.1 in May 1996. The current version 3.2 was released in November 1996. The following new programs were included in these releases:

- CROSSEC .... scattering cross-sections
- GEOMCALC .... geometry calculations
- HGEN .... generate hydrogen positions
- MAKEDIC .... make dictionary entry
- MAPROT .... map manipulations
- MATTHEWS_COEF .... Matthews coefficient
- MTZMNF .... add Missing Number Flags
- RASMOL .... molecular graphics
- REFMAC.......................... maximum likelihood structure refinement
- RESTRAIN .... structure refinement
- SCALEPACK2MTZ......... conversion jiffy
- SOLOMON .... density modification
- SORTWATER .... sort waters
- XDLMAPMAN................. map manipulations
- XDLDATAMAN................ data manipulations

For more details, see the individual program documentation. Victor Lamzin's (EMBL-EBI) program ARP is scheduled for release with the next version. The latest releases also contain numerous improvements to existing programs. The following more general changes to the suite are also taking place.

Missing Data Treatment

It is recommended practice that the set of indices in a data file is made complete within the desired resolution range. The MTZ file will then contain records where there are indices but no measured data for some or all of the data columns. From version 3.0, these columns are flagged MNF (Missing Number Flag). This means that it is easy to estimate completeness, and programs such as REFMAC and SIGMAA can "restore" data estimates where required.

mmCIF Format

It is intended to replace the PDB format for coordinates by the new macromolecular CIF (mmCIF) format. Peter Keller, funded by CCP4, has developed library routines to facilitate the use of mmCIF, and it is planned to convert the application programs in the near future.

Graphical User Interface

In response to demand, it is planned to develop a GUI to the CCP4 suite. A draft specification has been drawn up, which aims to provide a user-friendly interface, while maintaining the flexibility of the existing suite. Liz Potterton, funded by CCP4, has recently begun to work on the implementation.

— Availability and Distribution

The program suite is available free of charge to academic institutes, subject to the return of a license, and then may be obtained by anonymous FTP from ccp4a.dl.ac.uk:pub/ccp4, or supplied on tape for a small handling charge. For more information, including official mirror sites, see http://www.dl.ac.uk/CCP/CCP4/main.html. Interested commercial organizations should contact CCP4 directly at Daresbury Laboratory *via* e-mail to ccp4@dl.ac.uk for separate arrangements.

PDBCONS Program

Ramneek Gupta, Bioinformatics Centre, School of Biotechnology, Madurai Kamaraj University, Madurai 625 021, India (mkubic2@giasmd01.vsnl.net.in).

PDBCONS, a new searcher and browser for the PDB, designed to run with PCs on a DOS platform, has been developed at the Bioinformatics Centre, School of Biotechnology, Madurai Kamaraj University, India.

The program has a built-in index consisting of header and other information from the PDB flat format files which makes the task of narrowing down entries easier on the basis of key fields. One can perform searches based on the ID code, compound name, source of compound, functional class, deposition date, author name (main authors for the PDB entry or even those in the additional references), resolution, R value, *etc*. It is possible to combine conditions and the results of two or more searches. One innovative feature is the ability to search on the basis of sequence motifs. One may specify a portion of a nucleic acid and/or amino acid sequence which it searches in all the macromolecular entries in the database.

Once narrowed down, the results may be conveniently viewed in a windowed screen which attempts to display as much information on the screen as possible. Information accessible from the browsing screen includes basic details (ID code, compound name, source, functional class, a few statistics, authors, *etc*.) as well as remarks for the entry, additional references, and the base sequence of the macromolecule.

Currently the package has built-in indices from the January 1996 release of the PDB, but it can be updated from new PDB files by an auto-convert option built into the package.

Conversion (native PDB file into PDBCONS index) takes an average of ten to twenty seconds per PDB file and searching usually takes about ten seconds depending on the query and the processor used.

PDBCONS is available in two flavors — one for 286s and below, the other for 386s and above. Recommended hardware is a 386 with at least 4MB RAM and MS-DOS version 5.0 or above. The basic program takes 2 MB of hard disk space and indices converted from the PDB January 1996 release (approximately 4100 macromolecules) take about 18 MB of hard disk space including remarks and sequence data.

The program is not yet available on the WWW, but may soon be made available from the PDB. For now, interested users may send e-mail to mkubic2@giasmd01.vsnl.net.in or write to the Bioinformatics Centre, School of Biotechnology, Madurai Kamaraj University, Madurai 625 021, India, for a copy of the program; or contact the PDB Help Desk at pdbhelp@bnl.gov.

Distance Education in Collaboration with the PDB

Peter Murray-Rust, Virtual School of Molecular Sciences, Pharmaceutical Sciences, Nottingham University, Nottingham, UK (pazpmr@unix.ccc.nottingham.ac.uk; http://www.vsms.nottingham.ac.uk/vsms/).

— Introduction

In today's rapidly moving world, where a web-year is measured at three months (!), it is increasingly difficult to keep up, even for "experts." At the same time, macromolecular structure is being recognized as crucial in more and more disciplines; therefore, the demand for high-quality, rapid, relevant instruction as well as guidance is immediately apparent. Few departments, except for the largest, have the resources to offer top-quality courses in macromolecular structure or have the time to keep up to date. Like the informatics world, a bio-year is probably also about three months.

The collaborative approach in biomolecular resources (exemplified by the interlinking of databases from genome to structure) has been an outstanding Internet success, and the PDB has played an important role. As with the other major centers, the PDB has seen its role as providing not only data, but supporting material, as well as human help and guidance. Such centers have an important role, therefore, in helping to provide training and education.

— *Distance Education and PPS95*

The Internet provides new approaches to training and education. Among these are:

- A central (global) repository of key resource material, quality-controlled and updated.
- Re-use of electronic material in different environments for different purposes.
- Direct electronic communication with people regardless of country.
- An open approach to the technology suitable for education, *e.g.*, molecular viewers.

Therefore, last year, Alan Mills (supported by Birkbeck Crystallography Department) and I launched the course, The Principles of Protein Structure, which was:

- Virtual (*i.e.*, no face-to-face meetings).
- Collaborative (anyone was welcome).
- Multimedia (we used RasMol and Kinemage).
- Platform-independent.
- Free.

We appealed to the world biomolecular community for volunteer support and were overwhelmed with the response. The PDB played a major part in this, not only in answering questions, but also in mirroring the course as well as in many other ways. I would like to take this opportunity to give heartfelt thanks.

The course not only used existing material, but provided many opportunities for innovation. Perhaps most valuable was the development of a virtual community, especially through the BioMOO where Gustavo Glusman and Jaime Prilusky, both of the Weizmann Institute, Rehovot, Israel, supported us with a virtual classroom and rooms for all of the amino acids! Some documentation of our discussions may still be found on the Web. And particular thanks to Gail Schuman, our “Webmother” from BNL, who compiled a resource for “PPS people.”

On the technical side, the greatest innovation was the hyperglossary — a collection of protein structure terms such as “alpha-helix” to which we added two- and three-dimensional structures as well as hyperlinks to PDB, MEDLINE, and Klotho (a biochemical compounds declarative database). We see such hyperglossaries as becoming a key resource both in learning materials and in automatic annotation of technical documents — this technology has already been used in other disciplines. For the Virtual Hyperglossary Home Page, please see http://www.venus.co.uk/vhg/. Other less obvious, but equally important developments, were the flexible use of hypermail for managed discussions and chemical/* MIME for downloading and rendering PDB structures.

We learned that the global electronic community can be both exhilarating and frustrating: we received superb contributions with clickable maps, rotatable molecules, *etc.*, from the more than twenty contributors to the hyperglossary. Also learned was the fact that volunteer enthusiasm does not carry very far; therefore, additional courses at Birkbeck have developed into fee-paying, accredited, Certificate courses attracting considerable numbers of students, and some imaginative funding for students from eastern Europe.

— *New Developments: BIOTOOLS96*

I have now moved (virtually) to Nottingham where I am expanding the idea of Virtual Education. This emphasizes continuing education of professional scientists, especially in industry and research organizations (please see http://www.vsms.nottingham.ac.uk/vsms/). The new course entitled Bioinformatics Tools, which uses Java, again supported by the PDB, will explore the impact that object technology will have on bioinformatics and molecular science. We have twenty fee-paying students from several countries and will be concentrating on protein sequence and, if time allows, protein structure. For additional information on BIOTOOLS96, please see http://www.vsms.nottingham.ac.uk/vsms/biotools/.

This course is important because, I believe, object technology — developing very rapidly — will make a major contribution to biomolecular resources. Complex objects can be distributed over several servers (*e.g.*, sequence on one, structure on another, *etc.*), and if properly structured, can be read directly into applications. The PDB and other centers realize this, but of major concern is the fact that very few people are trained in both protein structure and object-oriented technologies such as Java — a rich, highly professional Object-Oriented language, not just for animations. The coming of CORBA will require us to develop object structures specified in Interface Definition Languages (IDLs) in order to use it properly. (My own interest is in developing tools (classes) in the protein/small molecule interface; and, shortly, these classes should be able to support things like substructure search, geometry comparison, and molecular similarity.) Once the basic classes are developed and tested, it should be possible to build substantial applications out of existing tools. For example, a program to do a Ramachandran plot would be only a page long. As with previous courses, the discussions are all public and should have some important lessons for those of us who are developing classes.

— *Conclusion*

Macromolecular structure is a wonderful subject for education — especially visually, often with immediate visual messages, *e.g.*, “where does the inhibitor fit in 4HVP?” Higher and postgraduate education are undergoing massive changes, and electronic resources will be increasingly important. I am confident that the PDB will continue to be seen as a top-quality world resource in this area.

The 24th Aharon Katzir-Katchalsky Conference: BIOINFORMATICS November 17 - 21, 1996 Jerusalem, Israel

Ellie Adman, Dept. of Biological Structure, University of Washington, Seattle, WA, USA (adman@u.washington.edu).

Bioinformatics is coming of age. In 1987 the word "bioinformatics" referred to the immune system repertoire, and by late 1993 it started being used in its current context, which is to describe an emerging field which will impact everyone who explores the richness and complexity of living systems. A birthday celebration was held in Jerusalem on November 17-21, 1996 to celebrate two essential components of the world of bioinformatics. The PDB started twenty-five years ago and at last count had over five thousand entries, each entry the result of an X-ray or NMR experiment to determine the three-dimensional structure of a macromolecule. The Swiss-Prot protein sequence database started only ten years ago and contains over fifty-nine thousand sequence entries. Both the PDB and Swiss-Prot actively promote complete validation and annotation of their entries, providing unique and invaluable resources for biologists seeking to understand complexity in molecular terms. The celebration took the form of a fascinating four-day meeting, bringing together participants who provide primary information, those who build and operate databases, and end users of databases ("bio-informers?").

— Opening Presentations

Ezer Weizman, the President of Israel (and also the nephew of the founder of the Weizmann Institute in Rehovot, where Joel Sussman, the current Head of the PDB, has a laboratory) and Professor Ephraim Katchalsky-Katzir, brother of the late Aharon Katzir-Katchalsky (in whose honor the conference was organized) opened the meeting. Weizman, amid many wonderful stories, emphasized the importance of scientific knowledge for emerging societies by drawing a parallel to how Israel started as a mostly agricultural economy and now has a very technologically-oriented economy. He also spoke of how his young grandson asks his grandparents to have a computer at hand to satisfy his curiosity, suggesting that future generations will take for granted rapid access to information. Katchalsky-Katzir spoke of how "wet" and "dry" science are now so essential: the "wet" science provides enormous amounts of information that the "dry" science can mine for new connections and understandings.

Leroy Hood, Chairman of the Molecular Biotechnology Department at the University of Washington, opened his remarks with a quote from Bill Gates, III, cofounder of Microsoft: "biotechnology and information sciences will be the two dominant scientific and industrial technologies of the twenty-first century." Bioinformatics will be a key tool for biotechnology. Hood elaborated in rich detail the kinds of information that may become available as a result of the collision of information technologies, from the very creative miniaturization of experimental setups and the means to evaluate these experiments. For example, a computer-chip-size addressable array of oligonucleotides of known sequence could be used for screening of genes or gene variants. Systems approaches to understanding biology and disease are being made possible by the wealth of information becoming available from genome projects. For example, prostate cancer may arise as a consequence of combinations of predisposing genes; knowledge of these genes could lead to more effective diagnosis and treatment.

Tom Blundell (University of Cambridge) demonstrated how knowledge of many different protein structures (and software to visualize them) leads to the recognition of structural superfamilies which, in turn, leads to testable hypotheses about related functions. The serum amyloid P component was inferred to be a lectin-like, carbohydrate binding protein from its structural resemblance to known lectins. This realization is leading to possible treatments of pathologies resulting from accumulation of this protein. An overall "relatedness" map derived strictly from the three-dimensional structures can be used with new structures — it is becoming quite rare to uncover a completely new fold.

Joel Sussman described new initiatives of the PDB which include more user-friendly access to coordinate sets and relevant annotations, both for depositors and for end-users. The World Wide Web has made significant impact on the accessibility of the PDB and related information *via* hypertext links. Amos Bairoch (Swiss-Prot) took the opportunity to thank the many contributors and users of Swiss-Prot, many of whom were recently revealed after a community appeal to respond to a financial-support crisis for the database. He emphasized the need for annotated, accurate depositions of sequences and suggested that "super users" who are expert in a particular system or organism are the best, and most useful, depositors of well-annotated sequences. The Director of the Cambridge Crystallographic Data Centre (CCDC), Olga Kennard, described an example of a knowledge-based database being derived from the CCDC and the PDB — a library of hydrogen-bonded contacts organized by functional groups. With such a library, one can ascertain if a new structure reveals novel interactions or ones now commonly known from the vast amount currently in CCDC. Chris Sander (EMBL) described how, with entire genome sequences, one can now look for an overlap of genes to ascertain probable functions for gene products and begin to understand, in molecular terms, the increasing complexity of higher organisms. Expansion in the numbers of genes needed for a particular functionality (the example was initiation/elongation in DNA replication), from five in *Methanococcus jannaschii*, to nine in *Haemophilus influenzae*, to twenty-four in yeast embodies the increase in complexity of organisms. Olivier Lichtarge (University of California) described a marvelous approach to evaluating protein function when a structure is known. Patterns of residue conservation that correlate with functional divergence are mapped onto the protein surface and shown to highlight binding and catalytic sites.

— Panel Discussion

A panel chaired by Otto Ritter (DKFZ, Heidelberg), consisting of representatives from structure, sequence, and genomic databases and a representative of the publishing world, was asked to ignore for the moment the mechanics of databases and to consider issues of what should be in future databases. Tom Blundell, utilizing his recent experience in the Ministry of Science in England, argued that user charges will become necessary, perhaps modeled after the CSD whose income is derived from licenses to countries and fees to commercial users. Access to databases *via* the Web complicates charging and it was generally felt that direct user fees would discourage database users. Databases should be considered as public libraries and be supported by a more widespread mechanism than user fees. Some felt that publishers might be a possible source of support. Matthew Cockerill (Current Biology) pointed out that journals have often served to focus attention on “hot” areas in science and evaluate and referee selected material, a function which is still desirable for dissemination of material on the Web. Indeed, a recurring theme of the meeting was that curated databases were of much greater value than non-curated databases. Value added by validation, annotation and cross-referencing to related databases is key. While much interest focused on Web resources, Meir Edelman (Weizmann) described a program to train and educate scientists at biotechnologically-developing centers where Web resources might still be unavailable. David Eisenberg (UCLA) used as an analogy the books on his shelves — which he found most useful and why — to underscore what he thought databases were about. Those that were just compilations of facts were least useful; fully annotated and cross-referenced ones were most useful. Continuous updating, easy interrogation by people and programs, non-redundancy, and links to other databases were on his list of desirable characteristics for future databases. Leroy Hood stressed the paradigm shift made possible by new data-rich technologies — databases of the future will drive Laboratory Information Management Systems, from the new technologies driven by the Genome project and combinations of heterogeneous databases, and should include visualization tools to enhance their usefulness. Janet Thornton (University College and Birkbeck College) emphasized that the distinction between primary data and derived data should be maintained but also urged that scientific journals should be more receptive to publications of studies which are derived from analysis of database material. Olga Kennard decried the decision by JACS not to accept papers where all “experimental” data is from databases. Others reiterated that knowledge-based building is also science. Ken Fasman (Genome Database, Johns Hopkins University) urged that assuring the fact that databases could cross-reference each other was something the panel members, as potential reviewers of database projects, should insist on. The uses of proprietary or patented data never making it to databases was raised. Hood estimated that one microbial genome sequence per month was not available publicly due to proprietary interests.

— *Protein Modeling and Design*

The half-day devoted to Protein Modeling and Design was enlightening. Manuel Peitsch (Glaxo) described the ExPASy server which provides a comparative modeling service currently containing more than two thousand predicted models in its database. A minimum sequence similarity of 35 percent is required; he estimates that about 30 percent of the unique sequences of Swiss-Prot are modelable. John Moult (University of Maryland) described the results of a modeling competition in 1994 (another competition is underway): while comparative modeling was most successful, reproducing the core structure very well, details in loops which are often most interesting as well as most likely to be unique to a protein, are less well-predicted. Threading is improving but may also introduce false positives in loop regions. Manfred Sippl (University of Salzburg) presented a fascinating contribution, using the database of known structures to define the distribution of oxygen and nitrogen positions in proteins, in lieu of an experimentally inaccessible distribution, to derive the free energy of interaction of these atoms as a function of distance apart. Surprisingly, the energy profile was quite rugged and even exhibited an apparent barrier between the “folded” and “unfolded” distances, likely entropic as these oxygens and nitrogens are attracted to other atoms. Barry Honig (Columbia University) explored the question of building free energies into scoring functions for threading in order to predict structures — concluding that they alone are insufficient. He suggested that sequences may encode kinetics of folding as well as the final folded structure. Steve Bryant (NIH) concluded this session with a description of the powerful ENTREZ Web system for retrieving related information about sequences built on indicators of neighborliness. The structure, sequence, DNA, and MEDLINE databases are evaluated and reorganized according to similarity indices, making it possible to search all of these simultaneously. A fascinating example of its use started with the prediction of the function of a new gene, the obese gene (leptin). Sequence homologies suggested similarity to helical cytokines, which in turn suggested a possible relationship to growth factors and then to growth factor receptors, which later led to a report of a leptin receptor with growth factor receptor similarity.

Graham Cameron (EMBL-EBI) provided a lively description of database evolution — early databases grew out of a need to provide standardized formats for number crunchers and were very much flat — an *ad hoc* solution. Traditional records of scientific information were of high quality, permanent, and citable. New issues resulting from the information explosion include updates, definitive versions, history (patent lawyers most interested here), ephemeral links, and quality control. Secondary functions are to create new knowledge from the databases, but we must beware of the electronic equivalent of mad cow disease! Federations of specialized databases will become more important. As mechanisms evolve, which include interface definition tools, user programs should become simple and more focused on delivering the objective of understanding biological complexity. An example of the development of such interface tools was given by Enrique Abola (PDB). A schema was developed describing all the objects within the PDB and particularly describing relationships among the objects — the ultimate goal being to enable a user to formulate queries which will provide a new understanding of the structures in the databases.

— *Structures*

The session on structures included David Eisenberg’s historical development of judging the compatibility of a sequence with a fold, first *via* evolution (homologies), then *via* physical chemistry (defined environments), and finally combinations that include descriptions of secondary structures. This improved predictability dramatically, but still permitted only about 17 percent of the coding regions of *Mycoplasma genitalium* to be predictable. Ada Yonath (Weizmann) described recent successes with the ribosome structure determination (surely an example of biological complexity) in being able to derivatize a replaceable subunit (L11) and also obtaining crystallizable ribosomes from *Haloarcula marismortui* that diffract to 2.7 Å. Jöel Janin (CNRS) reviewed a fascinating database analysis of buried surface area of macromolecular complexes noting distinct classes: protein/inhibitor or antigen/antibody complexes seem to bury about 1600 $Å^2$; crystal contracts are of the order of 600 $Å^2$, protein/DNA or protein dimers bury approximately 3500-3800 $Å^2$. A contest to predict the docking of a protein and its inhibitor, organized by Natalie Strydnaka (University of Alberta, Canada), showed that predictors identified the correct surface, roughly, but with fairly wide variability around the real site.

Janet Thornton used database analysis to evaluate the relative sizes of clefts in the surfaces of proteins and found that, in most cases, the largest cleft did correspond to an active site. Michael Levitt (Stanford University) described three-dimensional approaches to predict protein shape and topology, which included generating all possible shapes taken by a set of linked points equal to half the number of residues and filtered by rules derived from knowledge of extant protein topologies. Tests against scrambled sequences showed that results were significant. The glyoxalase structure was described by Alex Cameron (Uppsala University) and shown to be similar to dioxygenase and to a bleomycin resistance protein, but with some domains swapped, an increasingly common phenomenon. Ron Unger (Bar Ilan University) described a very novel approach to thinking about protein evolution, suggesting that most folds are the result of evolution from transiently stable interactions to systems with global stability. A lattice model of folding was used to show how that works.

Victor Markowitz (Lawrence Berkeley National Laboratory) gave an excellent presentation on how "Object Protocol Model Data Management Tools" may be used to permit the following: top-down design of new databases, retrofitting of existing databases, means to interconnect databases, and design of complex queries. Development of the 3DB Browser, described by Jaime Prilusky (Weizmann and PDB), is focusing in this way to facilitate complex queries of the Protein Data Bank. Faced with the huge problem of validating the thousands of entries in Swiss-Prot, Rolf Apweiler (EMBL-EBI) described TREMBL, a new approach to keep track of new sequences and to provide a way of smoothly updating Swiss-Prot, checking for redundancy, and annotating with links to other databases.

Shoshana Wodak (Université Libre de Bruxelles and EMBL-EBI) described an analysis of the packing volumes associated with atoms in structures, at first with the intention of using it as a tool for validating structures, but eventually concluding that there are regions of structures (using barnase as a test case) which are cavity-prone. She suggested that volumes need to be subdivided into classes for evaluation. Specialized ("boutique") databases are of enormous use for particular end users — one such example is RELIBase (an object-oriented comprehensive receptor-ligand database) described by Manfred Hendlich (Merck, Germany) which has been designed to facilitate complex queries such as "are there proteins with ligand 'x,' residue type 'y,' with > 70 percent homology, and are there any mutants known that affect binding?" Fortunately, the Merck-supported database will soon become publicly available. Adrian Goldman (Centre for Biotechnology, Turku, Finland) used 124 well-defined representative structures from the PDB to ask how common buried charges are and found that large proteins have a higher percentage of buried charges — greater than 50 percent involved at active sites. Ron Appel (University Hospital of Geneva) described the database for 2D-GELs of proteins separated by isoelectric point and molecular weight. This is perhaps an example of the kind of database that will need to be established and maintained for Leroy Hood's vision of labs on a chip — where quick visual comparisons of experimental results can be made. The 2D-GEL database allows clicking on a particular spot on a gel and retrieving its identity and an index to Swiss-Prot. Osnat Herzberg (University of Maryland Biotechnology Institute) showed how availability of structures in the PDB led to identification of the probable active site of pyruvate phosphate kinase whose structure was determined without a cofactor or substrate present. Structural homology also facilitated understanding of the remarkable multi-enzyme complex of a cellulosome, as well as a protein involved in apoptosis, and the terminator protein of prokaryotic DNA replication (Dirksen Bussiere, Abbott Laboratories). Finally, the results of a drug design project for inhibitors of HIV protease that bypass problems of development of drug resistance was described by Keith Watenpaugh (Pharmacia and Upjohn). New EM studies on the structure of the spliceosome and the elegant structure of the beef heart cytochrome oxidase was reported — the latter all being examples of increasing biological complexity that our experimental tools are capable of handling.

— *Conclusion*

The take-home lesson from this conference is that knowledge is organized information. Knowledge enables us to make testable predictions and eventually usable tools for improving the quality of life on this planet. Valid, annotated, complete, non-redundant, and cross-referenced databases are key to the continued productive organization of information. Just as new technologies such as high-speed computing caused significant advances in structure determination, these sophisticated database tools will drive us rapidly into new ways of asking and answering scientific questions. Access to these tools *via* the WWW (*e.g.*, PDB's 3DB Browser, the ENTREZ browser, ExPASy server, *etc.*) has already catapulted productivity into unprecedented realms. This meeting brought together experts in production of primary data, experts in database design and management, and numerous people who have already made extensive use of databases in very creative ways. Others of us can't wait to do likewise! One wonders if the field isn't ripe for a new "BioInformatics Society" to provide a forum for hashing out standards, methodologies, and end user desires — although perhaps a virtual society already exists *via* the World Wide Web.

An html version of Dr. Adman's report may be found at http://www.bmsc.washington.edu:80/people/adman/pdb.fm.html.

Notes of a Protein Crystallographer

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

— Recuerdos de la Alhambra: Remembrances of the Alhambra

I will certainly not reveal anything extraordinary if I begin these notes referring to the symmetry present in the woodcuts and engravings of the Dutch artist Maurits Cornelis Escher (1898-1972). This fact is very well known to crystallographers, and there are several monographs focusing on this aspect of his work [1]. However, I expect to entice your curiosity by bringing up the issue of the origin, inspiration, and motivation of the symmetrical elements found in the graphic art of this master craftsman.

M.C. Escher was born on June 17, 1898 in Leeuwarden, Northern Netherlands. It was soon evident that he liked to draw and was encouraged by his early teacher, Mr. F.W. van der Haagen, to make prints. On his father's advice Escher went to Haarlem to study architecture. It was one of the faculty members there, the Dutch artist Samuel Jessurun de Mesquita (note the Arab surname), who advised him to leave architecture and concentrate on the graphic arts. In fact, Escher had already realized that he did not like architecture very much and followed his mentor's advice wholeheartedly. He continued his education under the guidance of Jessurun de Mesquita from 1919 to 1922 and afterward, traveled frequently to Southern Europe, particularly Italy and Spain.

In the spring of 1922 he went to Italy, and in the autumn of the same year, he made a brief trip to Spain. This trip to Spain made a tremendous impression on him. From 1923 to 1935 he settled in Italy, making frequent trips to small villages in Northern Italy, Sicily, and Corsica — always taking notes and making sketches of motifs of interest which he later used for prints during the winter. The political situation and the rise of fascism in Italy made his life there less pleasant, and for a short time in 1935 he moved to Switzerland. From May through the end of June 1936, he made his longest study trip along the coast to Italy and Spain. On this voyage, he made detailed copies of the Moorish mosaics in what to me are the two most outstanding monuments of Arab architecture in Western Europe — the Palace of the Alhambra in Granada, and the Mezquita-Cathedral at Córdoba, both in Southern Spain. After a short residence in Ukkel (near Brussels) he returned to the Netherlands in 1941 to settle in Baarn. He passed away in Laren, on March 27, 1972.

After 1937, M.C. Escher became rather sedentary and traveled only for family reasons. His trips did not have any further influence on the material of his work. During a trip to Canada in 1964 to visit his oldest son, he was scheduled to visit the United States to give a series of lectures discussing his methods and techniques. Unfortunately, all of the engagements were canceled due to illness — instead he was admitted to Saint Michael's Hospital in Toronto. However, the intact, fully-prepared text of these lectures, along with the corresponding illustrations, has been published in English [2], and, through them, the reader has a rare glimpse of the artist's techniques and "intentions."

Regarding his interest in the regular division of the plane he wrote [3]:

> "I can't say how my interest in the regular division of planes originated and whether outside influences had a primary effect on me. My first intuitive step in that direction had already been taken as a student [...] in Haarlem. This was before I got to know the Moorish majolica mosaics in the Alhambra, which made a profound impression on me."

In the opening remarks of the first lecture he explains in more detail [4]:

> "Many of the bright-colored, tile-covered walls and floors of the palaces of the Alhambra in Spain show us that the Moors were masters in the art of filling a plane with similar, interlocking figures, bordering on one another without gaps. [...].
>
> What a pity that the religion of the Moors forbade them to make images! It seems to me that they sometimes came very close to the development of their elements into more significant figures than the abstract geometric shapes that they created. No Moorish artist has [...] ever dared (or he didn't hit on the idea) to use as building components concrete, recognizable figures borrowed from nature, such as fishes, birds, reptiles, or human beings. This is hardly believable, for *recognizability* is so important to *me* that I never could do without it.
>
> Another important question is *color contrast*. It has always been self-evident for the Moors to compose their tile-scenes with pieces of majolica in contrasting colors. Likewise, I have never hesitated myself to use color contrast as a means of visually separating my adjacent pattern components."

So, was it the Alhambra that left such a lasting impression on M.C. Escher? It is the last and most exquisite outpost of Arab civilization in western Europe. The last dynasty of Moorish kings reigned there until the incipient nation of Spain, under Queen Isabella of Castille and King Ferdinand of Aragon, conquered it in 1492. The fall of Granada under Christian rule opened the way for the adventure of Columbus and the New World. Atop the most gentle ranges of the Sierra Nevada, the Alhambra looks like "a square, brutal fortress" that does not reveal (nor even hint) at the delicacy of the forms buried in its interior [5]. Inside, the Alhambra is a palace of exquisite, intricate, and delicate beauty. It is a labyrinth of columns, corridors, courts, chambers, and windows — full of geometric and sensuous forms, quivering with the music of water murmuring in the courtyards. I speak from experience when I say that even crystallographers will get intoxicated and dizzy by the ever-present symmetry in the colorful majolica tiles on the walls and floors, carved on the plaster in the ceilings, inscribed on the supporting wooden beams, and surrounding the secluded windows. I can only recommend that you visit it.

As much as music can evoke places and sensations, I will venture that there is a very special composition of music for the classical guitar that does succeed in evoking in my brain the interior of the Alhambra. Musically, it is a very modest piece. In fact, it was intended more as an *étude* than a full composition and is entitled very appropriately *"Recuerdos de la Alhambra*." It was written by Francisco Tárrega (1852-1909), a Spanish guitarist and composer who played a very important role in developing the music and technique for the classical guitar in this century. In doing so, he was instrumental in transporting this delicate instrument from the *tavernas* and *cafés* of 19th century Spain to the concert halls of the world. Classical guitar music is now enjoyed by a multitude of people in the recordings of Andrés Segovia, Narciso Yepes, Julian Bream, Christopher Parkening, and many others.

Being an *étude*, "*Recuerdos*" is not a very exhilarating piece. It is monotonous and repetitive, as symmetry is. Yet, it is a serene and beautiful piece that has become an indispensable composition in the repertoire of the romantic guitar. It pretends to exercise a technique of particular importance to the string instruments and particularly to the classical guitar named "*Tremolo*." It consists of the rapid reiteration of a note or group of notes, resulting in a "quivering" or "tremulant" effect. Not inappropriately, "*tremolo*," has the same root as "tremor" and "*tremuloides*." In its most gentle mode, the tremolo sounds like the quivering of the leaves of the quaking aspens (*Populus tremuloides*) ever present on the slopes of the Rocky Mountains. It is specifically this "quivering" sound which I associate with the murmur of running water of fountains and streams, in combination with the monotonous, symmetrical nature of the music that my brain associates with the serenity and enchantment inside the walls of the Alhambra.

An illustration superimposing a sample of the Moorish symmetrical patterns in Spain with the first page of the musical score of "Recuerdos de la Alhambra" by Francisco Tárrega is available from PDB's Home Page.

______

1. Fantasy & Symmetry. The Periodic Drawings of M.C. Escher. C.H. MacGillavry. Harry N. Abrams, Inc., Publishers, New York. (1976).
2. Escher on Escher. Exploring the Infinity. Translated from the Dutch by Karin Ford. Harry N. Abrams, Inc., Publishers, New York (1989).
3. Ibid, pg. 83
4. Ibid, pg. 25. Parts of these comments are also repeated in the Preface of [1].
5. The Ascent of Man. J. Bronowski. Little, Brown and Co. Boston, Toronto. (1973).

■

WEB SITES REFERENCED IN THE JANUARY 1997 PDB NEWSLETTER

3DB Browser	www.pdb.bnl.gov/cgi-bin/pdbmain
Archive of Obsolete PDB Entries	www.sdsc.edu/PDBobsolete
AutoDep Submissions to PDB	terminator.pdb.bnl.gov:4148/autodep-basepage.html
BioMoo	www.cco.caltech.edu/~mercer/htmls/BioMOOHomePage.html
BIOTOOLS96	www.vsms.nottingham.ac.uk/vsm
CCP4	www.dl.ac.uk/CCP/CCP4/main.html ftp://ccp4a.dl.ac.uk/pub/ccp4
ENTREZ	www3.ncbi.nlm.nih.gov/Entrez/
ExPASy Molecular Biology Server	expasy.hcuge.ch/
SWISS-2DPAGE	expasy.hcuge.ch/ch2d/ch2d-top.html
Swiss-Model	expasy.hcuge.ch/swissmod/SWISS-MODEL.html
SWISS-PROT	expasy.hcuge.ch/sprot/sprot-top.html
Klotho: Biochemical Compounds Declarative Database	www.ibc.wustl.edu/klotho/
MMDB: Entrez's Structure Database	www.ncbi.nlm.nih.gov/Structure/
MOLEMAN2	alpha2.bmc.uu.se/~gerard/manuals/moleman2_man.html ftp://rigel.bmc.uu.se/pub/moleman2
NDB (Nucleic Acid Database)	ndbserver.rutgers.edu:80/
OPM (Object-Protocol Model) Data Management Tools	gizmo.lbl.gov/DM_TOOLS/OPM/OPM.html
PDB Contents Guide and Format Change Policy	www.pdb.bnl.gov/doc_help.html
SCOP	scop.mrc-lmb.cam.ac.uk/scop/
Mirrored at PDB	www.pdb.bnl.gov/scop/
TREMBL	ftp://ftp.ebi.ac.uk/pub/databases/trembl/
The 24th Aharon Katzir-Katchalsky Conference	www.pdb.bnl.gov/pdb25sp10/index.html www.bmsc.washington.edu:80/people/adman/pdb.fm.html
The 1996 Principles of Protein Structure Course at Birkbeck College	www.cryst.bbk.ac.uk/PPS2/index.html
Mirrored at PDB	www.pdb.bnl.gov/PPS2/index.html
Mirrored at Daresbury	www.dl.ac.uk/PPS/index.html
University College London	www.biochem.ucl.ac.uk/
CATH Protein Structure Classification	www.biochem.ucl.ac.uk/bsm/cath
Mirrored at PDB	www.pdb.bnl.gov/bsm/cath
Enzyme Structures Database	www.biochem.ucl.ac.uk/bsm/enzymes/
Mirrored at PDB	www.pdb.bnl.gov/bsm/enzymes/
PDBsum	www.biochem.ucl.ac.uk/bsm/pdbsum
Mirrored at PDB	www.pdb.bnl.gov/bsm/pdbsum
VHG - The Virtual Hyperglossary	www.venus.co.uk/vhg/
VSMS Home Page - The Virtual School of Molecular Sciences	www.vsms.nottingham.ac.uk/vsms/

AFFILIATED CENTERS AND MIRROR SITES

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

RELATED WWW SITES

Databases

Archive of Obsolete PDB Entries	www.sdsc.edu/PDBobsolete
BMRB (BioMagResBank)	www.bmrb.wisc.edu
CCDC (Cambridge Crystallographic Data Centre)	www.ccdc.cam.ac.uk
EBI (European Bioinformatics Institute)	www.ebi.ac.uk
EMBL (European Molecular Biology Laboratory)	www.embl-heidelberg.de
ExPASy Molecular Biology Server	expasy.hcuge.ch
GDB (Genome Data Base)	gdbwww.gdb.org
GenBank (NIH Genetic Sequence Database)	www.ncbi.nlm.nih.gov/Web/Genbank/index.html
HIV Protease Database	www-fbsc.ncifcrf.gov/HIVdb/
Klotho: Biochemical Compounds Declarative Database	www.ibc.wustl.edu/klotho/
Library of Protein Family Cores	WWW-SMI.Stanford.EDU/projects/helix/LPFC/
NCBI (National Center for Biotechnology Information)	www.ncbi.nlm.nih.gov
NDB (Nucleic Acid Database)	ndbserver.rutgers.edu
PDB (Protein Data Bank)	www.pdb.bnl.gov
PIR (Protein Identification Resource)	www.gdb.org/Dan/proteins/pir.html
Prolysis: A Protease and Protease Inhibitor Web Server	delphi.phys.univ-tours.fr/Prolysis/
Protein Kinase Database Project	www.sdsc.edu/kinases/pk_home.html
Protein Motions Database	hyper.stanford.edu/~mbg/ProtMotDB/
SCOP: Structural Classification of Proteins	scop.mrc-lmb.cam.ac.uk/scop/
Mirrored at PDB	www.pdb.bnl.gov/scop/
Swiss-Prot Sequence Database	expasy.hcuge.ch/sprot/sprot-top.html

Software-Related Sites

CCP4	www.dl.ac.uk/CCP/CCP4/main.html ftp://ccp4a.dl.ac.uk/pub/ccp4
mmCIF	ndbserver.rutgers.edu/NDB/mmcif
O Home Page	kaktus.kemi.aau.dk
OPM (Object-Protocol Model) Data Management Tools	gizmo.lbl.gov/DM_TOOLS/OPM/OPM.html
RasMol Home Page	www.umass.edu/microbio/rasmol/
Squid: Analysis and Display of Data from Crystallography and Molecular Dynamics	www.yorvic.york.ac.uk/%7Eoldfield/squid/
VMD - Visual Molecular Dynamics	www.ks.uiuc.edu/Research/vmd/
X-PLOR Home Page	xplor.csb.yale.edu/

Other Resources

Crystallography Worldwide	www.unige.ch/crystal/w3vlc/crystal.index.html
DALI - Comparison of Protein Structures in 3D	www.embl-heidelberg.de/dali/dali.html
MOOSE (Macromolecular Structure Database at San DiegoSupercomputer Center)	db2.sdsc.edu/moose
PDB_select: Representative PDB Structures	ftp://ftp.embl-heidelberg.de/pub/databases/ protein_extras/pdb_select/recent.pdb_select
PROCHECK - To Submit a PDB File for Analysis	www.cryst.bbk.ac.uk/PPS/procheck/test.html
Protein Structure Verification-Biotech Server	biotech.embl-heidelberg.de:8400/
Mirrored at PDB	biotech.pdb.bnl.gov:8400/
Resources for Macromolecular Structure Information	www.ucmb.ulb.ac.be/StructResources.html
Weizmann Institute, Biological Computing Division	dapsas1.weizmann.ac.il/

BROOKHAVEN ORDER FORM

Name of User ____ Date ____________________

Organization ____ Phone ____________________

Address ____ Fax ____________________

____ E-mail ____________________

____

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

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

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

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

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

____

____

____

____

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

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

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

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

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

(tax and shipping charges not applicable)

This page intentionally left blank.

STATEMENT OF SUPPORT

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

ACCESS TO THE PDB

Main Telephone	1 516-344-3629
Help Desk Telephone	1 516-344-6356
Fax	1 516-344-5751
Help Desk	pdbhelp@bnl.gov
General Correspondence	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 has updated the FTP server (ftp.pdb.bnl.gov) in order to have a more standardized directory structure. This will facilitate use of mirror software to keep local copies of the database current. Links have been put in so the server still “looks” the same to users.

Entry files are now found under the directory pub/pdb/.

all_entries/	coordinate entry files in one directory
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.

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

(516) 344-3629

First Class