Quarterly Newsletter published by
Brookhaven National Laboratory
Protein Data Bank

Protein Data Bank
pdb™
Brookhaven National Laboratory

Release #83 **January 1998**

January 1998 CD-ROM Release

6947 Released Atomic Coordinate Entries

	Molecule Type
6151	proteins, peptides, and viruses
268	protein/nucleic acid complexes
516	nucleic acids
12	carbohydrates

	Experimental Technique
168	theoretical modeling
1089	NMR
5690	diffraction and other

1673	Structure Factor Files
400	NMR Restraint Files

The total size of the atomic coordinate entry database is 3.0 GB uncompressed.

Table of Contents

What's New at the PDB 1
Archive Management 2
EBI Now Accepting AutoDep Submissions 3
The 'Intelligent' Search Engine Behind the 3DB Browser™ 4
Request for a Revision of IUCr Policy on Publication and Deposition of Crstallographic Data 5
PDB Computer Services 6
Energy Department Announces New BNL Contractor 6
Exceptional Science Fair Project Uses the PDB 5
Writing Structure Factors in mmCIF using CCP4 7
Protein Topology WWW Site 8
SARF2 – a Program for Comparison of Protein Structures 8
Molecular Docking by Fourier Correlation with FTDOCK 9
MolView and MolView Lite 10
OLDERADO: Extracting Single Structures, Core Atoms and Domains from a NMR-derived Ensemble 11
Notes of a Protein Crystallographer– FRODO, the Electronic Hobbit 11
Web Sites Referenced in the January 1998 PDB Newsletter 13
Affiliated Centers and Mirror Sites 14
Related WWW Sites 16
PDB™ Order Form 17
PDB Access, FTP Directory Structure, Consultants, Staff, Support and Instructions to Authors 19

Internet Sites

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

What's New at the PDB

Joel L. Sussman

Structure factors are the observed experimental data from X-ray crystallographic experiments. They are the basis of the X-ray coordinate entries and as such need to be readily available to and usable by researchers. To facilitate the exchange of this data between scientists, as well as for their deposition and retrieval from the PDB, it was decided to set up a standard format, *i.e.*, a Lingua Franca, for structure factors.

The PDB and a number of macromolecular crystallographers, including the Chairperson of the IUCr Working Group on Macromolecular CIF, Dr. Paula Fitzgerald, and other members of this committee, developed a standard interchange format for structure factors. This standard is in mmCIF format, *i.e.,* the IUCr-developed 'macromolecular Crystallographic Information File'. It was chosen for simplicity of design and for being clearly self-defining. The format is also easy to extend, by simply adding additional tokens as new crystallographic experimental methods or concepts are developed (see the January, 1996 PDB Newsletter and ftp://ftp.pdb.bnl.gov/structure_factors/cifSF_dictionary). The entire mmCIF crystallographic dictionary has recently been ratified by the IUCr's COMCIFS committee (http://ndb.rutgers.edu/NDB/mmcif/).

We have been strongly urging our depositors to submit structure factors with their entries (Baker *et al.*, 1996). We are pleased to report that since the release of the PDB's Web-based deposition tool, AutoDep, in October 1996, that almost two-thirds of the depositions of X-ray structures to the PDB are now accompanied by their structure factors. Dr. Jiansheng Jiang, at the PDB, has converted these recently-deposited structure factors, as well as virtually all the previously-deposited ones, to the standard mmCIF format. The structure factors are available through the PDB's Web-based 3DB Browser™ (http://www.pdb.bnl.gov/pdb-bin/pdbmain), as can be seen on the Browser's 'Atlas' page for each structure.

Over the years, the PDB has observed that one of the most useful reasons for storing structure factors is for the crystallographer who did the experiment to be able to retrieve his/her own data which have been misplaced in their laboratory. In parallel, the fact that this data is now easily available, and in a standard format, has already begun to foster new community-wide efforts at improvements in validation techniques based on the experimental data, *e.g.*, SFCHECK by Vagin, Richelle & Wodak (http://www.sdsc.edu/Xtal/IUCr/CC/School96/), the Uppsala Electron Density Server by Taylor (http://alpha2.bmc.uu.se/valid/density/form1.html), and others.

References:

Baker, E. N., Blundell, T. L., Vijayan, M., Dodson, E., Dodson, G., Gilliland, G. I. & Sussman, J. L. (1996). Crystallographic Data Deposition. *Nature* **379**, 202.

Archive Management

Enrique E. Abola

Layered Release

In the October 1997 PDB Quarterly Newsletter, we discussed the *layered-release* that will allow for a virtually immediate release of entries (to be referred to as the 1st layer) without staff intervention. A PDB ID code will be issued only after the depositor gives approval to release his/her entry either immediately or as soon as it comes off hold. Following this, PDB staff will process the entry as done presently. This processing will include standardization of nomenclature, other annotation, and more importantly, data representation. Most of this work covers issues not now fully delegated to software. The resulting entry will be loaded on our servers as the 2nd layer. The set of mandatory items to be required before data are accepted was discussed in the October 1997 PDB Newsletter. The complete list may be found at our Web site (http://www.pdb.bnl.gov).

Listed below are a series of checks that will be done on an entry as part of the submission process. The first checks will be used to ensure that entries with obvious deficiencies are not released (*e.g.*, duplicate atom records). The other checks will be used to add annotations to the entry. When the coordinates are loaded on the PDB server, a file containing the results of the diagnostic runs will be loaded as well.

The following tests will be done on the entry as part of the submission process. Results of the tests will be provided to the depositor who can then take the appropriate action given the options outlined below before a PDB ID is issued.

1. Diagnostics requiring corrections and re-submission of coordinate data.
   - More than one polypeptide or nucleotide chain assigned the same chain name
   - Heterogen group specified by HET and FORMUL records not present in the ATOM/HETATM records
   - More than 10% of the atoms involved in unusually close crystal packing interactions (this check will also cover the case for which a non-standard space group setting was used and the correct set of symmetry operators were not provided)
   - Violation of atom nomenclature for standard amino and nucleic acids
   - Duplicate ATOM or HETATM records in the same residue with the same atom name or the same coordinates
   - ATOM/HETATM records not correctly formatted
   - Heterogen ID provided in the coordinate file conflicts with the PDB Het Dictionary
2. Diagnostics requiring annotations and/or comments to be provided by the depositors if the data are not corrected. The PDB will insert a CAVEAT record before release.
   - For polypeptides, phi-psi angles for more than 20% of the residues outside the allowable region
   - Unexpected chirality at C-alpha center
3. The following diagnostics are normally used by PDB staff to take a closer look at the data because, by experience, we have found that they may be indicative of unusual structures or possible problems. We will present this information to the depositor and the list will also be included in a file containing the output of our checking runs to be made available to the users. The depositor may, of course, modify the coordinate file during the submission process to correct for possible errors before giving final approval to release the data.
   - RMSD of bond lengths greater than 0.08 Ångstroms from ideal values
   - RMSD of bond angles greater than 5.0 degrees
   - Breaks in the chain (*e.g.*, due to disorder)
   - Differences between amino acid sequences given by the ATOM records and those given in the appropriate sequence database entry
   - Amino acid sequences not reported in any sequence database
   - CIS-peptides and peptide bonds that deviate significantly from the expected trans conformation
   - Individual bond lengths differing by more than 0.1 Ångstroms from standard values
   - Individual bond angles differing by more than 15 degrees from standard values
   - Atoms too close to symmetry axes
   - Atoms involved in unusually close crystal packing interaction
   - Atom occupancies less than or equal to 0.0 or occupancies greater than 1.0
   - Atom occupancies less than 1.0 and for which no alternate location ATOM record is provided
   - Missing residues, missing atoms
   - Thermal factors greater than 100 Å**2
   - Unexpected deviations from planarity
   - Non-standard SCALE matrix
   - OXT atom record in the middle of a chain (flagged as extra atom), typically occurring before a gap in the coordinates
   - R value greater than 30%
   - Free-R value greater than 35%
   - Free-R and R value differ by more than 10%
   - RMSD between atoms related by NCS MTRIX records is greater than 3.0 Ångstroms

Tests which are valid only for diffraction experiments will **not** be applied to entries reporting NMR experiments or model building studies.

Heterogen groups will be checked against the current PDB Het Dictionary. The only check that will be done at this stage of the processing is to see if the HET ID and the atom nomenclature used for a group is consistent with the dictionary (*e.g.*, is the GLC group in the coordinate file a glucose molecule as given in the PDB Het Dictionary and are the atoms properly named?). Groups that are not in the dictionary and for which there is no conflict on the HET ID code will be accepted as is and will be checked and standardized as part of the regular processing to be done after the first layer is loaded.

Complete descriptions of these tests along with a more precise definition of values such as those defining allowable Ramachandran plot regions are provided on our Web pages (http://www.pdb.bnl.gov). Please send your comments and suggestions regarding these tests and/or on the layered-release to abola1@bnl.gov.

Summary of Data Processing Activities for 1997

In 1997 we received 1,844 coordinate sets and released 1,631. This averages out to 153 entries deposited per month which is about 27% more than the 1996 submission rate. On average it took us

119 days to release an entry, which is significantly improved from the 173 days that it took us to release an entry in 1996. A plot giving the growth in data deposition as well as the turn-around time is provided on the back cover of this Newsletter. The plot is accessible *via* our Web Home Page, and is updated after every load.

There were several changes in our procedures that have allowed us to handle the increased rate of deposition while at the same time reducing the amount of time required for processing. Most significant was the release of our AutoDep program in October 1996 that has greatly simplified processing of entries. More than 70% of the entries submitted in 1997 were done through AutoDep.

Starting in November 1997 we initiated a new procedure in which entries are released every Tuesday night. This was done at the request of several Mirror Sites, most of which have programs that automatically generate indices relating PDB entries to other databases. Users wishing to check data loads can visit our Home Page for a list of recently released ID codes.

EBI Now Accepting AutoDep Submissions

The following announcement was posted on several listservers and newsgroups on December 23, 1997, including the PDB Listserver, X-PLOR Listserver, and the O Listserver.

Deposition of 3D Structural Studies of Biological Macromolecules

We are pleased to announce the inauguration of a new deposition site for 3D structural studies of biological macromolecules. Starting on January 5, 1998, authors using the Web-based tool, AutoDep, can submit data either to the European Bioinformatics Institute (EBI), UK or to the Protein Data Bank (PDB) at Brookhaven National Laboratory (BNL), USA. The additional site is expected to significantly facilitate the submission procedure, especially for European researchers.

AutoDep is a Web-based tool originally designed at PDB for automatic submission of macromolecular data into the PDB. Extensive collaboration between EBI and PDB has produced significant changes to the original system allowing for the seamless operation of multiple deposition sites. This includes EBI specifications for standards and protocols to be used in making the code portable and generally more robust.

AutoDep is accessible from the following URLs:

* BNL-PDB	http://www.pdb.bnl.gov
* EBI-MSD	http://autodep.ebi.ac.uk

Those wishing to submit data using the electronic version of the Deposition Form must continue to deposit directly to BNL using e-mail or FTP.

The submission procedure will be identical, and equivalent, at both sites, but PDB ID codes will be issued by BNL. Data submitted at EBI will be forwarded automatically to PDB after depositors have reviewed the AutoDep-generated entry and diagnostics. Final preparation for archiving and release will be done by PDB staff. We encourage depositors to submit not only the structural results, but also their experimental data, *i.e.*, for crystallographers, X-ray structure factors, and for NMR spectroscopists, constraints lists and statistical data describing the calculated NMR conformers and constraints.

Important Notes:

(i) Submissions can be completed only at the site at which they were started.
(ii) The option “Based on a previous submission” may be used to simplify submissions by using an earlier AutoDep session as a template. However, depositors will only have access to their earlier submissions at the site where those submissions were originally made.
(iii) Existing PDB entries may be used as templates by choosing the option “Based on an existing PDB entry”. The full set of entries will be available at either site, irrespective of where the original deposition was made.
(iv) The date of submission for data deposited at EBI will be the corresponding U.S. Eastern Time of the date when submission is completed at EBI.
(v) EBI staff will offer assistance (*via* e-mail: pdbhelp@ebi.ac.uk) up to the point of submission. Once BNL has issued an ID code, correspondence should be directed to BNL (*via* e-mail: pdbhelp@pdb.bnl.gov).

Please note that authors should continue to deposit crystal structures of nucleic acids to the Nucleic Acid Database (NDB) at Rutgers, the State University of New Jersey, USA at URL: http://ndbserver.rutgers.edu:80/NDB/deposition/index.html.

Experimental data related to NMR studies will also be transferred electronically to the BioMagResBank (BMRB) at the University of Wisconsin-Madison, USA for further processing and inclusion into the database (http://www.bmrb.wisc.edu) as well.

Joel L. Sussman
Head, Protein Data Bank
Biology Department
Brookhaven National Laboratory
Upton, NY, USA

Phil McNeil
Head, Macromolecular
Structure Group
EMBL Outstation
European Bioinformatics Institute
Wellcome Trust Genome Campus
Hinxton, Cambridge, UK

The ‘Intelligent’ Search Engine Behind the 3DB Browser™

Jaime Prilusky

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

The new 3DB Browser™ allows the user to rapidly search through the contents of the entire PDB Archive for entries matching certain constraints. A full text search can be made for any string appearing in the text of a PDB entry, excluding the coordinate records. Many specific records can be searched for regular expressions or numerical limits. 3DB Browser gives you the option of saving object sets resulting from queries. This saved set can be used as a starting point for further database operations or as a reference for your work. Every saved set includes the date of the search and the query from which it was generated.

The Search Fields of the 3DB Browser

The main source of information for the 3DB Browser is the data from the Protein Data Bank. This data is highly structured and most of the crystallographers are used to thinking of a piece of data from a PDB entry as belonging to a particular “record” or “field”. It makes sense to use these fields to constrain the search. Searching for ‘rich’ as a keyword has a different meaning than searching for the author Rich.

Search Field	PDB Entry
PDB ID code	Four-character accession code
Keyword	Molecule name, class or family, or related term (HEADER, TITLE, KEYWDS and COMPND fields)
Author	Family name of depositor or author of associated publication (AUTHOR and JRNL fields)
Text query	Any word in the complete PDB text, excluding the field names
FASTA Search	Fasta search of the sequence
Experiment	Method of structure determination
Resolution	A unique value or range of values, in Angstroms (REMARK 2 field)
Space group	Both extended and standard Hermann-Mauguin symbols (CRYST1 field)
Organism	Trivial name, systematic name or expression system (SOURCE field)
Date (lower)	Date entry was released or updated
Date (upper)	Date entry was released or updated
Associated group	Prosthetic group, metal ion, ligand or substrate, or its three letter PDB abbreviation (HET and HETNAM fields)

Examples and Boolean-style Searches

The simplest operation with the browser is to enter one or more words in the "Text query" field and press the "Search" button. The browser engine will come back with those entries from the database that contain or are related to the provided words.

The symbol '*****' can be used as a wild card, to denote a sequence of any number (including 0) of arbitrary characters. Just add a star '*****' at the beginning or end of a word (or both) to 'extend' the search. For example, enter ***tox*** in keywords to retrieve those entries with keywords like neurotoxic and toxin. Wild cards have no meaning in number-only fields, like Resolution and Date.

The Boolean operator **AND** is the default for 3DB Browser, and mandatory (you cannot change it) between fields. If you enter 'ATP' in the Associated group field and 'kinase' in the Keyword field, only those entries matching both constraints are returned.

Inside a given field, you may apply Boolean logical operators at will to the words you enter. The available Boolean logical operators are **AND, OR** and **NOT**. The case is unimportant. The operator **AND** can be represented by '**+**' and the operator **NOT** represented by '**-**'.

For example, 'zinc and (torpedo or snake)' in the Text query field will return those entries that contain either the word torpedo or the word snake, but only where the word zinc is also present.

To Err is Human

One of the main concerns for us, as database-interface developers, is the "false negatives", that is, to not return data after a query, even when the data are available in the database. Frequently this happens because the user was unable to express the query in a way compatible with the search engine, or used words or keywords unknown to the search engine.

3DB Browser deals with this problem by incorporating several automatic and semi-automatic mechanisms to help the user in retrieving the requested data. The request from the user gets filtered and transformed by one or more of the following engines. At the end, the resulting query is the one used for the search.

Engine	Example
american-british	'amoeba' and 'ameba' are equivalent
synonyms	'protease' is equivalent to 'proteinase'
spelling search	based on a dictionary built from the current PDB data, the spelling engine will produce words that are close to the entered one. As an example, entering 'imune' will offer 'immune' as a valid alternative.
soundex search	based on the soundex algorithm that approximates the sound of the word when spoken by an English speaker. Looking for author 'weich' will offer as alternatives: Weiss, Wess, Wyss ...

Inside this section on understanding what the user looks for, we can include the improved search on the CRYST1 record using the short and extended Hermann-Mauguin symbols. You may enter either **'P 1 21 1'** or **'P 21'** in the Space group field and get the same result.

3DB is Just the Starting Point

A search in 3DB brings up a rich Atlas page summarizing additional knowledge related to the entry of interest. The links in this Atlas page carry you to the original sources of information. The number of external sources that 3DB searches and dynamically incorporates into the Atlas pages increases daily. The following table summarizes the external sources currently referenced by 3DB.

Source Name	Short Description
BioMagResBank	Relational Database for Sequence-Specific Protein NMR Data
BLOCKS	Database of conserved regions in groups of proteins
CATH	Protein Structure Classification
Dali/FSSP	Families of Structurally Similar Proteins
EMBL	European Molecular Biology Laboratory
Entrez	NCBI's Documentation database
ENZYME	Enzyme nomenclature database
ESTHER	ESTerases and alpha/beta Hydrolase Enzymes and Relatives
GenBank	NIH genetic sequence database
GDB	Genome Data Base
Kinase	Protein Kinase Database Project
KineMage	Protein Science's Kinemage server
LPFC	Library of Protein Family Cores
MacroMolecule	EBI's Crystal MacroMolecule Files
MMDB	Molecular Modelling Database
NBD	Nucleic Acid Database
OLDERADO	Core, Domain and Representative Structure Database
PDBOBS	Archive of obsolete PDB entries at SDSC

PDBREPORT	Structure verification reports for X-ray structures
PIR	Protein Information Resource
PROSITE	Dictionary of protein sites and patterns
ProtMotDB	Protein Motions Database
scop	Structural Classification of Proteins
SWISS-3DIMAGE	3D images of proteins and other biological macromolecules
SWISS-PROT	Annotated protein sequence database
TREMBL	TRanslation from EMBL

If you know of other sources of information related to PDB that can be incorporated into 3DB's Atlas page, please send an e-mail message to lsprilus@weizmann.weizmann.ac.il.

Support your Local Store

The Protein Data Bank has several mirror sites across the world. These sites have the same data and facilities as in the central PDB server. They are just closer to you, and, frequently, faster to access on the Internet. To help you know your neighborhood, the 3DB Browser incorporates "**closer-site**", an automatic script that detects your location and offers alternative sites that are closer to you (in the network sense).

Drop an e-mail to lsprilus@weizmann.weizmann.ac.il if you are interested in getting the "closer-site" script for your own application.

Request for a Revision of IUCr Policy on Publication and Deposition of Crystallographic Data

Alex Wlodawer

Macromolecular Structure Laboratory, ABL-Basic Research Program, Frederick Cancer Research and Development Center, National Cancer Institute, Frederick, MD, USA (wlodawer@ncifcrf.gov)

Dear Colleagues,

For the last two years, I have been working on trying to change the policies of journals and funding agencies which allow hold periods of up to one year for the coordinates resulting from crystallographic and NMR studies. (See also Sussman 1997.) It is now becoming clear that the best way to accomplish such a change would be to induce IUCr to change their official recommendations (International Union of Crystallography, 1989). Several of us have recently written a letter to Science, which appeared in the January 16th issue (Wlodawer, 1998), suggesting that their policy be modified. It is necessary, however, to involve the largest possible segment of the structural community in this endeavor. For that purpose, we are circulating a petition which will be presented to IUCr. If you agree with the text of the petition below, please send a brief message to me at the e-mail address wlodawer@ncifcrf.gov. You might also wish to send a message if you disagree with the petition and would like to keep the current policy in place. The results of this vote will be reported to the community before any further action is taken.

References:

International Union of Crystallography. Commission on Biological Macromolecules. (1989) Policy on publication and the deposition of data from crystallographic studies of biological macromolecules. *Acta Crystallogr., Sect.A*, **45**, 658. (Policy also in section 11.3 of http://hobbes.gh.wits.ac.za/iucr-top/journals/acta/actaa_notes.html).

Sussman, J. L. (1997) What's new at the PDB. *PDB Q. Newsl.*, No.82, 1.

Wlodawer, A., Davies, D., Petsko, G., Rossmann, M., Olson, A. & Sussman, J. L. (1998) Immediate release of crystallographic data: a proposal. *Science*, **279**, 306.

Petition:

To: Commission on Biological Macromolecules, IUCr

We, the undersigned, would like to request a revision of the IUCr policy on publication and deposition of data from crystallographic studies of biological macromolecules (Acta Cryst. A45, 658 (1989). It is our intention that if the policy gets revised, the new rules will be communicated to granting agencies and to scientific journals, in order to be universally accepted.

The current policy has been implemented on the basis of the discussions which had taken place a decade ago. In the meantime, there has been an incredibly rapid increase in the rate of determination of 3D structures of biomacromolecules, as reflected by the deposition of a new structure in the Protein Data Bank (PDB), on average, every five hours. Unfortunately, in parallel, an increasing proportion of depositors take advantage of the PDB's policy of allowing structures to be kept 'on hold' for up to a year after coordinate deposition. Consequently, as many as 45% of newly deposited structures are not available when the relevant papers are published.

When the issue of deposition was debated by the community ten years ago, the time needed to solve a macromolecular structure was often measured in years, and was rarely less than one year. The time needed for detailed analysis of such structures was also fairly long. The one-year hold on coordinates was therefore instituted to allow the authors to reap the fruit of their tremendous investment of time and effort. Due to recent advances in protein expression and purification, crystallization procedures, X-ray instrumentation, and computer software, the time needed to solve a structure is often shorter than the allowed hold period. In light of such developments, it is very difficult to justify withholding coordinates for any period once the paper has been published.

Biomolecular structure analysis has indeed succeeded in bringing 3D structures to the forefront of molecular biological research. This success has expanded both the interest in and utility of the information being deposited in the PDB. The molecular modeling community has grown and evolved considerably due to the expansion of this source of experimental data. The value of the data rests in their availability to the broader community. Methods are continuously being developed to analyze new structures and their relationships to the collection of existing structures. New uses for these data, such as statistical potentials for folding and threading calculations, and interface recognition tools, are evolving rapidly. No single research group can fully exhaust this wealth of information. The value of the resource grows proportionally to the timeliness of the data and to the number of scientists who have access to them. 3D structural information is also a crucial link elucidating the role of a translated region of a DNA sequence of unknown function.

We feel most strongly that the time has come to change the rules of deposition so as to ensure that the coordinates are released concomitantly with publication of the paper(s) describing the structure. We are convinced that without access to the coordinates, the struc-

tures cannot be utilized for comparison with other proteins, for theoretical analysis or, more and more importantly, for drug design. We propose that coordinates deposited at the PDB should be marked as either "for immediate release" or "to be released upon publication". We also recommend that the maximum hold for primary data, *i.e.*, X-ray structure factors, and NMR-based restraints, be reduced from 4 years to 1 year. These changes would bring macromolecular crystallography into line with the requirements of other fields, such as gene sequencing, which have never allowed extended hold periods.

PDB Computer Services

John McCarthy

PDB's WWW Browser Discontinued

During the final six months of 1997, usage of the PDB's WWW Browser had dropped significantly. Additionally, in December of 1997 the PDB released the latest version of the 3DB Browser™. It has all the features of the WWW Browser plus many more (see "The 'Intelligent' Search Engine Behind the 3DB Browser™" in this Newsletter). For these reasons, the PDB discontinued the WWW Browser in December of 1997. Please remove any bookmarks to it that you might still have.

PDB CD-ROM Files Compression

As was reported in the July 1997 PDB Quarterly Newsletter, the PDB started compressing files on its October 1997 CD-ROM release. The Structure Factor files were compressed allowing the full CD-ROM release to fit on six CDs.

The January 1998 CD-ROM release, most likely, will still fit on six CDs by compressing the Structure Factor files, but in the near future, coordinate entry files will be compressed as well.

As was stated in the July 1997 PDB Newsletter, an effect of compression will be that the filenames will be different. Files that had the PDB ".ent" suffix will have the ".gz" suffix when compressed. Any scripts that read coordinate entry files directly from the CD-ROM will have to be modified to use the new filenames and perform the uncompression as necessary. The PC-based browser PDB-Shell has been updated to be able to read compressed entry files.

The PDB is using the Gnu gzip package to perform the compression and is distributing the Gnu Gunzip package in the CD-ROM set to allow CD-ROM users to perform uncompression.

A questionnaire was sent to all users receiving the July 1997 CD-ROM release requesting their views regarding compression of files on the CD-ROM release. The responses were overwhelmingly in favor of compression.

Energy Department Announces New BNL Contractor

Based on Department of Energy press releases (http://apollo.osti.gov/doe/whatsnew/pressrel/pr97130.html and http://apollo.osti.gov/doe/whatsnew/pressrel/pr98001.html)

Completing a major step in its ongoing effort to improve management and restore confidence at Brookhaven National Laboratory, the Department of Energy announced on November 25, 1997, the selection of Brookhaven Science Associates (BSA) as the new contractor to manage and operate its Long Island, NY, research facility. The BSA team is led by the Research Foundation of the State University of New York on behalf of the State University of New York at Stony Brook and Battelle Memorial Research Institute of Columbus, Ohio.

Secretary of Energy Federico Peña said, "Brookhaven Science Associates has demonstrated leadership at their institutions, and I will look to them to fully integrate safety and environmental protection into scientific research, to accelerate and intensify recent efforts to rebuild community trust, and to achieve overall excellence. Working together, we will make it possible for the laboratory to carry out its mission as a world-class research facility and prove itself a good neighbor to Suffolk County and Long Island."

"This is the fastest competition we have ever held for a management and operating contract, and reflects a new way of doing business at the Department of Energy," he added.

Two proposals from nonprofit-led teams were submitted in response to a July 18, 1997, Request for Proposals. One proposal was from BSA. The other competing team was led by the IIT Research Institute of Chicago, Illinois. Both teams provided excellent proposals and demonstrated their ability to manage scientific, environmental, safety and health, and community involvement initiatives at the laboratory. The selected team demonstrated the best total capability to improve the laboratory's performance.

Stony Brook is a national leader in high energy and nuclear physics. Battelle Memorial Research Institute has operated the department's Pacific Northwest National Laboratory in Richland, Washington, for the last 32 years and has long been a leader in applied science and technology, including environmental, safety and health management. Battelle manages environmental, safety and health activities at the department's Pantex Plant.

BSA committed to exceeding the requirements of the department's Request for Proposal in several areas, including: the department's safety management program that integrates safety into employees' daily work activities; implementing IS0 14001, an international standard for environmental management systems; and instituting a Voluntary Protection Program whereby the lab will subscribe to meeting proven industrial standards for worker safety.

Other immediate organizational commitments include retention of existing salary, benefits and tenure systems for employees and appointment of separate deputy directors for science and operations. In addition, offices for environment, safety and health, environmental management, reactor operations and community involvement will report directly to the laboratory director.

The BSA proposal identified Dr. John Marburger as the new Brookhaven Laboratory director. Dr. Marburger is a distinguished science administrator and served as State University of New York at Stony Brook's president for 14 years. He has committed to integrate laboratory safety with scientific excellence and to regain community trust and confidence.

The department's streamlined process for selecting a new contractor, limited to nonprofit organizations or teams led by nonprofit organizations, was completed in six months, rather than the usual 18, and was developed with extensive community, industrial and academic involvement. The department initially planned to award the contract in mid-November 1997. However, the award was not

signed until January 5, 1998, in order to comply with recently enacted federal legislation requiring 60-day notice to the United States Congress before awarding certain contracts.

BSA will assume responsibility for laboratory operations following a transition period of 55 days after the contract award. Until that time, Associated Universities Inc., the current contractor for Brookhaven, will continue to manage the laboratory.

An introduction to BSA is available at: http://www.pubaf.bnl.gov/pr/BSA.htm. The source selection statement is available at: http://www.ch.doe.gov/bnlseb.

Exceptional Science Fair Project Uses the PDB

Reprinted below is a letter from a father about his son's use of the PDB.

The PDB entry used was 1FKS: FK506 AND RAPAMYCIN-BINDING PROTEIN (FKBP12). The molecular dynamics program referred to was PMD, Parallel Molecular Dynamics (http://tincan.bioc.columbia.edu/pmd/pmd-summary.html), written by Dr. Andreas Windemuth.

25 Nov 1997
Hello Dr. Sussman,

I was at the SuperComputing 97 show last week in San Jose, CA, and I stopped by the Brookhaven National Laboratory booth. I talked about my son who did his high school science fair project last year using the PDB at BNL, and the folks in the booth strongly suggested that I write you a short note describing his work.

My son had a kidney transplant and one result of this is that he takes immunosupressive drugs to prevent organ rejection. While searching the PDB, he found a few entries that were the binding proteins for a commonly used immunosupressive. He copied the structure from the PDB server and then ran the structure through a molecular dynamics program available from Columbia University to generate the full 3D structure of the protein. He experimented with the protein by making single atom changes in one of the 107 amino acids, and had the MD program recompute the structure. He then compared the resulting shape of the new protein to the original protein to estimate the ability of his new protein to bind to the immunosupressive drug. He found that there were several changes he could make to the protein that would appear to have little impact on its binding capabilities, while other simple one-atom changes resulted in a very different 3D structure. He was able to conclude, based upon residual displacements and by using RasMol to visualize the structures, which proteins would still be active in binding to the immunosupressive and which would not.

His science fair project was well received at both his school, Marian High School in Framingham, MA, as well as at the regional science fair at Worcester Polytechnic Institute and at the Massachusetts state science fair at MIT.

This was a wonderfully educational experience for him, and gave him a positive experience in rational drug design. The relevance of this type of science to his daily life was made quite clear to him. The educational capabilities of having this type of data available on the Internet should not be overlooked.

Regards,
Dr. Don Dossa
Digital Equipment Corp

The entire PDB sends best wishes to Dr. Dossa's son for good health in the future.

Writing Structure Factors in mmCIF using CCP4

Peter Keller

European Bioinformatics Institute, Hinxton Hall, Cambridge, CB10 1SD, UK (keller@ebi.ac.uk)

Depositing your MTZ-formatted structure factor data to the PDB using the Web-based AutoDep procedure is quite straightforward: all you need to do is use the 'CIF' output option of the CCP4 program 'mtz2various'. The output file can then be uploaded along with your coordinate file, when you start your submission. The PDB strongly encourages use of this procedure to prepare your structure factor file for submission.

Unlike the other output formats of mtz2various, the mmCIF output will contain every reflection which is present in the MTZ file, even if the structure factor amplitude is the missing number flag, or the reflection is systematically absent for the space group which was finally assigned to the data. Each reflection is flagged in the output according to its status (see the CCP4 documentation on mtz2various for more information).

Other ways to indicate the status of reflections in the output file:

- The 'FREEVAL' option can be used to indicate the test set that was excluded from refinement for calculation of the free R factor.
- If, for some reason or other, you have performed your final refinement against a subset of the data, you can indicate the resolution limits with the 'RESO' option, and a sigma cutoff with 'EXCLUDE SIGP'.

A simple example might look like this:

```
mtz2various hklin sf.mtz hklout sf.cif
OUTPUT CIF data_sf
LABI FP=FNAT SIGFP=SIGFNAT
END
```

A more complicated example:

```
mtz2various hklin sf.mtz hklout sf.cif
OUTPUT CIF data_sf
LABI FP=FNAT SIGFP=SIGFNAT I(+)=INAT SIGI(+)=
SIGINAT FREE=FREEFLAG
FREEVAL 2
RESO 15.0 2.1
END
```

In this case, the reflections for which the FREESET column is 2, have been used as the free R test set, and only data between 15 and 2.1 Ångstroms were used in the refinement. Also, the merged intensities which were input to 'truncate' have been retained in the data file, and are being written to the output mmCIF.

The parameter following 'OUTPUT CIF' must begin with 'data_'. The characters which follow identify the data, and are to some extent arbitrary: they will be changed by PDB staff as appropriate when your submission is processed. If you are unsure what to put here, choose some alphanumeric string which means something to you, such as the MTZ filename or the name of the protein.

The name of the original MTZ file appears within the first few lines of the output file - look for '_audit.creation_method'.

Anomalous data are handled with the DP/SIGDP (and I(-)/SIGI(-)) column assignments.

Protein Topology WWW Site

David R. Westhead[1,4], Daniel C. Hatton[1] and Janet M. Thornton[1,2,3]

[1] *European Bioinformatics Institute, EMBL outstation, Wellcome Trust Genome Campus, Hinxton, Cambridge, CM10 1SD, UK*

[2] *Biomolecular Structure and Modelling Unit, Department of Biochemistry and Molecular Biology, University College, London, WC1E 6BT, UK*

[3] *Laboratory of Molecular Biology, Department of Crystallography, Birkbeck College, University of London, Malet Street, London, WC1E 7HX, UK*

[4] *E-mail: westhead@ebi.ac.uk*

We have recently set up a WWW site (http://tops.ebi.ac.uk/tops) devoted to protein structural topology. The central service offered at this site is an "atlas" of protein topology cartoons in which each PDB entry has a representative topology cartoon. Also available is a "server" facility to which protein structures can be submitted (in PDB file format) for cartoon calculation, and a good deal of information about protein structural topology.

Protein topology cartoons are simple two-dimensional schematic diagrams of protein folds. They represent a fold as a sequence of secondary structure elements (helices and strands) and show the relative position and direction of these elements in the fold. An example cartoon is shown in figure 1. Protein three-dimensional folds can be complicated and difficult to interpret. The aim of topology cartoons is to simplify them so that they can be more easily understood and compared. The simplification afforded by the cartoon in figure 1 is clear.

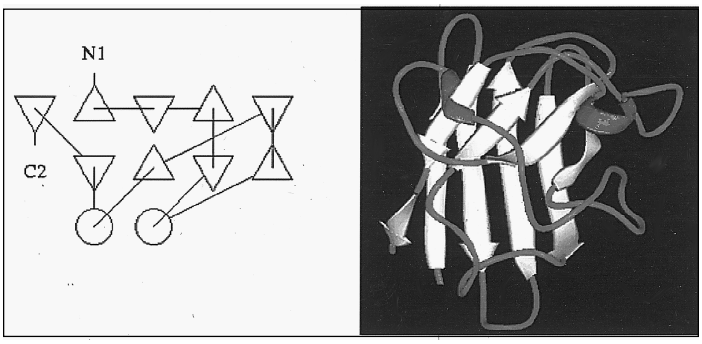


Figure 1. The topology cartoon and 3D structure of superoxide dismutase (1 jcv). The topology cartoon displayed by software available on the WWW site. In the cartoon, triangles represent beta strands and circles helices. The direction of the strands is implied by the orientation of the triangles: "up" (out of the plane of the page) strands are drawn as up triangles and "down" strands as down triangles. The peptide chain runs from N1 to C2. The structure of the fold as a sandwich made of two anti-parallel beta sheets is clear from the cartoon.

The atlas of topology cartoons was generated from the version of the PDB current on July 1, 1997. In order to avoid the generation of many duplicate cartoons, the chains present in the database were first clustered at a sequence similarity threshold of 95%. Chains consisting of nucleic acid sequences were removed, as were protein chains of less than 30 residues. From an original total of 10534 chains this produced 2144 clusters of near-identical sequences. From each cluster a representative TOPS diagram was produced from a single structure. This was chosen to be the highest resolution X-ray structure in the cluster, or an NMR structure if no X-ray structures were available. Within a chain, each structural domain was plotted separately using domain definitions taken from the CATH1 protein structure classification.

The cartoons were generated automatically, in the first instance, using a substantially modified version of the program TOPS 2. While the original version of TOPS would produce satisfactory cartoons for simpler protein folds, it was found to be unable to do so for many more complicated folds. The modifications were necessary in order to increase the success rate of the program sufficiently to make automatic generation of a large number of cartoons a viable proposition. The generation of the atlas of cartoons was viewed as a test of the new version of the program. Each cartoon in the atlas was checked manually with the 3D structure of the protein, and the success rate in producing satisfactory cartoons was found to be 82%. Among the failures were many cartoons which were correct but not aesthetically pleasing, but there were still some complicated folds for which the program failed. The cartoons judged to be failures were corrected by hand editing and included in the atlas.

The atlas is viewed using an applet (a program written in the Java programming language, delivered over the WWW, and run on the client machine). A basic applet using Java version 1.0 simply allows the user to view the cartoons, while users with a WWW browser supporting Java version 1.1 can use a much more functional applet which allows editing and printing of the cartoons. The same applets are used for viewing, editing, and printing cartoons generated at the request of the user by the server facility. Some users with older machines and/or browsers have experienced difficulties with the Java technology and for this reason a purely HTML/GIF version of the atlas will be provided in the near future

We hope to keep the atlas up to date as new structures arrive in the PDB. However, because updates to the atlas require significant effort, we anticipate that there will always be a time lag between structures arriving in the PDB and cartoons being put into the atlas. In this case users will be able to use the server to generate their own cartoons for the new structures. The software used in the generation of the atlas will be made available in some form, and details will be posted on the Web site.

Acknowledgements

We are grateful to Dr. T. P. Flores for giving us the source code for TOPS and allowing us to modify it without restriction. We are also grateful to Dr. C. A. Orengo for providing us with the domain boundary file associated with the CATH1 protein structural domain classification.

References

Orengo, C. A., Michie, A. D., Jones, S., Jones, D. T., Swindells, M. B. & Thornton, J. M. (1997). CATH—a hierarchic classification of protein domain structures. *Structure,* **5,** 1093-1108.

Flores, T. P., Moss, D. S., & Thornton, J. M. (1994). An algorithm for automatically generating protein topology cartoons. *Prot. Eng.* **7,** 31-37.

SARF2 - a Program for Comparison of Protein Structures

Nickolai N. Alexandrov

Amgen, Thousand Oaks, CA, USA (nicka@amgen.com,http://www-lmmb.ncifcrf.gov/~nicka/info.html)

Discovering new similarities in protein structures is an extremely exciting process. It is especially interesting if proteins are not sequentially related and so the structural similarity is completely unexpected. Obviously, when you find a structural resemblance, you have two problems: first, you need to prove that the similarity is significant, and, second, you need to explain the biological meaning of this similarity. Traditionally the significance of the match is demonstrated by an unusually large number of C-alpha atoms which can be superimposed with a small root mean square distance (rmsd). Biological meaning can be explained by the evolutionary relationship of the proteins, similar functional properties, and/or energetic stability of the 3D motif.

There are several programs for protein structure comparison recently reviewed by Gibrat *et al.* (1996). However, finding common motifs in 3D structures is not a trivial problem. Probably the most difficult part here is to think up a measure of similarity between two structures which correlates with biological sense. Usually a similarity between two structures is described in terms of the number of C-alpha atoms and the rmsd between them. Yet, these numbers do not provide an adequate measure of structural similarity. For example, isolated residues with a small rmsd are likely to form a less significant match than a spatial arrangement of continuous backbone fragments.

The program SARF2 (Alexandrov, 1996) detects common motifs in protein structures which consist of similarly-arranged backbone fragments. (The abbreviation SARF stands for Spatial ARrangement of backbone Fragments and was first used by Alexandrov *et al.*, 1992.) There are two kinds of the spatial resemblance: topological and non-topological similarities. Topological equivalence assumes that the fragments in both proteins are connected in the same sequential order. Non-topological similarities are relatively rare. An example of the non-topological structural similarity is four-helical bundle motif, in which helices can be differently connected, but still remain within the same protein architecture. SARF2 is able to detect both kinds of similarities.

The Web site for SARF2 in the Laboratory of Experimental and Computational Biology at the National Cancer Institute (http://www-lmmb.ncifcrf.gov/~nicka/info.html) allows you to compare just two structures. If you want to compare many protein structures, you can download the program from the ftp site (ftp://ftp.ncifcrf.gov/pub/SARF2/) and run it on your SGI or DEC Alpha machine. There are mirror Web sites for SARF2 at the Baylor College of Medicine (http://defrag.bcm.tmc.edu:9503/lpt.html), at the GMD/SCAI in Germany (http://cartan.gmd.de/nick/run2.html), and at the Sanger Centre in England (http://genomic.sanger.ac.uk/).

An important and still open question in protein structure comparison is an evaluation of the significance of the match. One way to solve this problem is to compare the structure of interest with all of the PDB and plot the distribution of the number of matched residues for each structure. The significance of the match can then be measured in the units of standard deviation from the mean of this distribution. This approach has been used by Alexandrov and Fischer (1996) to make a classification of the representative list of protein structures.

A knowledge of the mechanism of protein function is sometimes a useful argument for the significance of the match. Frequently, active sites are surrounded by a similar structural environment, although the protein function can be different. And, vice versa, detection of an unexpected statistically significant structural similarity can lead to new speculations on the mechanism of protein function.

The most interesting structural similarities are those between proteins with low amino acid identities. Understanding the origin of these similarities provides a deeper insight into the mystery of protein folding. The fact that the number of structural classes is smaller than the number of different sequence families encouraged many researchers to apply a variety of sequence-structure compatibility (threading) methods. One of these methods (program 123D), based on the contact capacity potentials, is also presented on the same NCI web site: http://www-lmmb.ncifcrf.gov/~nicka/info.html.

References:

Alexandrov, N. N. (1996). SARFing the PDB. *Protein Eng.* **9,** 727-732.

Alexandrov, N. N., Fischer, D. (1996). Analysis of topological and nontopological structural similarities in the PDB: new examples with old structures. *Proteins,* **25,** 354-365.

Alexandrov, N. N., Takahashi, K., & Go, N. (1992). Common spatial arrangements of backbone fragments in homologous and non-homologous proteins. *J. Mol. Biol.* **225,** 5-9.

Gibrat, J. F., Madej, T., & Bryant, S. H. (1996). Surprising similarities in structure comparison. *Curr. Opinion in Struct. Biol.* **6,** 377-385.

Molecular Docking by Fourier Correlation with FTDOCK

Henry A. Gabb and Michael J.E. Sternberg

Biomolecular Modelling Laboratory, Imperial Cancer Reseach Fund, London, UK (gabb@ibm.wes.hpc.mil, m.sternberg@icrf.icnet.uk)

The ability to predict the binding geometries of biomolecular complexes is becoming increasingly important with the growing number of individual structures deposited in the Protein Data Bank because experimental determination of the structure of biomolecular complexes remains a difficult problem. FTDOCK was developed to address the problem of docking unbound molecules when the structure of the complex is unavailable (*i.e.,* predictive docking). FTDOCK implements the geometric surface recognition algorithm of Katchalski-Katzir and coworkers (Katchalski-Katzir *et al.,* 1992) to dock two macromolecules. The method takes advantage of the fast Fourier transform (FFT) to rapidly search the translational space of two rigidly rotated molecules. An electrostatic function amenable to the Fourier correlation algorithm has been developed in this laboratory that improves the final rank of correctly docked molecules (Gabb *et al.,* 1997). Possible docking orientations are scored for surface complementarity and favourable electrostatics using Fourier correlation theory. For docking starting with unbound coordi-

nates, we have shown that in some systems inclusion of electrostatics is critical to success.

We have used FTDOCK in our laboratory to dock several protein systems for which the coordinates of the complex and the individual subunits are available (Gabb *et al.,* 1997). The test set was comprised of six enzyme-inhibitor and four antibody-antigen complexes. In all but one of our test cases, correctly docked geometries (interface C-alpha root-mean-square deviation less than or equal to 2.5 Ångstroms squared) are found during a complete search of binding space in a list that was always less than 250 complexes and often less than 30. At this point, biochemical information is still necessary to remove incorrect predictions. We found that knowledge of at least one binding site further improved rankings for correct solutions. For six out of nine test cases, a correctly docked complex was placed in the top five predictions. For the other three test cases, two had a correctly docked complex in the top fifteen predictions and the other had a correct answer in the top fifty. Considering that 10^{10} geometries are screened during the global search of binding space, these results are encouraging. When information about the binding site on both molecules is available, a correctly docked complex scored in the top five for eight out of nine test cases. Many of these had the correct answer ranked first in the list of predictions. Even the worst test case had a correctly docked complex ranked 27th.

FTDOCK was developed under Irix 5.3 and 6.2, but the program should run on any UNIX computer. The current version of FTDOCK uses either the fast Fourier transform from Numerical Recipes Software (Press *et al.,* 1986) or the Silicon Graphics CHALLENGEcomplib™ (Silicon Graphics Inc.) to take advantage of the SGI shared memory multiprocessor. However, the latter FFT will also run efficiently on SGI serial computers. A typical docking experiment takes about six hours of CPU time using eight processors in parallel on a SGI Power Challenge. This assumes a rotational increment of 15 (6385 nondegenerate rotations). A typical docking attempt takes 3-4 days on a SGI Indy using the Numerical Recipes FFT rather than the SGI library routine. In some cases, however, a larger increment can be used for the rotational search (Katchalski-Katzir *et al*., 1992). Using an angular deviation of 20 (2629 nondegenerate rotations), for example, reduces the computational time to less than one day on a SGI Indy workstation.

FTDOCK can also be used to dock non-protein systems like nucleic acids or small molecules. Our experiments with non-protein systems have not yet been published. The program can be obtained via our WWW site (http://www.icnet.uk/bmm/software.html).

References

Gabb, H. A., Jackson, R. M. & Sternberg, M. J. E. (1997). Modelling protein docking using shape complementarity, electrostatics, and biochemical information. *J. Mol. Biol.,* **272,** 106-120.

Katchalski-Katzir, E., Shariv, I., Eisenstein, M., Freisen, A. A., Aflalo, C. & Vakser, I. A. (1992). Molecular surface recognition: determination of geometric fit between proteins and their ligands by correlation techniques. *Proc. Natl. Acad. Sci. USA* **89,** 2195-2199.

Press, W. H., Teukolsky, S. A., Vetterling, W. T. & Flannery, B. P. (1986). *Numerical Recipes in Fortran,* Cambridge University Press. Available from Numerical Recipes Software (http://cfata2.harvard.edu/nr/).

Silicon Graphics Inc. (1995) CHALLENGEcomplib™ Science and Math Library. (http://www.sgi.com/Products/Challengecomplib.html).

MolView and MolView Lite

Thomas J. Smith

Department of Biological Sciences, Purdue University, West Lafayette, IN 47907 (tom@bragg.bio.purdue.edu)

MolView is a program to display and analyze atomic structures and MolView Lite is a simple rendering program using the new QuickDraw3D technology. Both freeware applications are currently limited to the Macintosh personal computer, but work is underway to create a version compatible with WIN95.

MolView has a wide variety of options to examine and display atomic structures. Atomic structures can be read into MolView as several types of text files: PDB, O plot files, ChemDraw 3D, and MolView files. Mono and stereo images can be interactively rotated with the mouse, tool palette buttons, or numerical input. Key aspects of the structures can be highlighted by mixing the available display modes: CPK, ribbon, ball&stick, line, and surface stippling. Emphasis has been placed on the user interface so that users unfamiliar with atomic structures can easily create figures and perform analysis while still being able to customize the object's attributes. Users can customize atomic labels and choose atoms for labeling by clicking on the atom or by picking them from a scrolling list. When creating ribbon diagrams, the secondary structure elements are either automatically determined from the structure using psi-phi values and hydrogen bonding patterns, or read from headers of PDB files. The user can color and toggle elements of the ribbon diagrams using a palette of buttons that display the current color, identifies the residue number of the element, and the type of secondary structure in that segment of the protein. The types of analyses that can be performed include distance measurements, 3D structural alignments, Ramachandran plots, Edmunson wheels, hydropathy plots, distance diagrams, B-value figures, hydrogen bonding patterns, and surface plots. More advanced users can also display crystallographic and non-crystallographically related molecules and unit cell boundaries.

There are several options when displaying nucleic acids. A ribbon can be drawn along the phosphoribose backbone and the ring structures can be color coded by filling the rings with colored planes. For presentation and educational purposes, there are several types of files that can be read or written by MolView. When MolView files are updated, colors, some objects, and the new orientation matrix are saved in the file. Line drawings, MOL objects, can be written as separate objects with the color information stored in the header. Using these objects, students can drag and drop files into MolView to view prepared structural lessons. Three different types of QuickTime movies can be created to enhance display performance on older machines or when creating multi-media resources. The various objects and plots can all be written to object-oriented PICT files for publication quality images. Simple line drawings can be saved as DXF files for import into other rendering applications. Finally, all of the various types of molecular objects can be written to QuickDraw3D (3DMF) files where they can be read and interactively rendered by a growing number of applications that run under either MacOS or Windows95 operating systems.

MolView Lite is a simple rendering application that reads and interactively renders the MolView 3DMF output. The image can be written out as a PICT image or QuickTime movie.

There are also several other resources available at the MolView WWW site (http://bilbo.bio.purdue.edu/~tom). Files demonstrating crystallographic and non-crystallographic symmetry are included in the application package. An interactive tutorial can be downloaded that takes the user through examples of the major options and explains many of the terms used in the write-up. The write-up is available in Word (5.1 and 6.0) and HTML formats. Example images and movies are also available at this site.

OLDERADO: Extracting Single Structures, Core Atoms and Domains from a NMR-derived Ensemble

Lawrence A. Kelley and Michael J. Sutcliffe

Department of Chemistry, Leicester University, Leicester, UK (L. Kelley@icrf.icnet.uk, sjm@le.ac.uk).

We have recently developed a WWW server, OLDERADO (On-Line Database of Ensemble Representatives and Domains; http://neon.chem.le.ac.uk/olderado/) (Kelley & Sutcliffe, 1997), which identifies the "best" single structure in a NMR-derived ensemble (Sutcliffe, 1993; Kelley *et al.,* 1996), and determines the "core" atoms across the ensemble and the domain(s) (or rigid body(ies)) to which these belong (Kelley *et al.,* 1997). The database component of OLDERADO has been integrated into the "Atlas" page resulting from a PDB 3DB Browser™ search for a NMR-derived ensemble, and in addition, individual representative MODEL structures for a PDB entry can be downloaded *via* the European Bioinformatics Institute (EBI) (http://www2.ebi.ac.uk/msd/nmr_search.shtml).

OLDERADO consists of two components: (i) a database of NMR-derived ensembles deposited in the PDB, and (ii) the functionality to upload and analyse a user's own ensemble of structures. Generation of the OLDERADO database, and processing of uploaded structures, is performed by two analysis tools: NMRCORE (Kelley *et al*., 1997) and NMRCLUST (Kelley *et al.,* 1996). NMRCORE automatically defines the core atoms and the domains in which these lie. This is achieved using a sorted list of dihedral angle order parameters (Hyberts *et al.,* 1992) to define the core, followed by the definition of the domain(s) which comprise the core using automatic clustering of the variances in inter-atom distances. NMRCLUST automatically clusters ensemble members into conformationally-related sub-families. All structures are superimposed in a pairwise manner and the resulting RMS distance between each pair calculated. These distances are used as a similarity score on which to base the clustering. Average linkage cluster analysis is used in conjunction with a novel penalty function to determine a cut-off in the clustering hierarchy automatically.

At the top of the results page, there is a summary which defines the largest domain and the "most representative" MODEL entry. Under this are two tables - the first detailing (in order of domain size) the core and domain(s), and the second (in order of cluster size) the representative structure(s) and cluster membership. These domains and clusters can be viewed interactively in three-dimensions via the "View Domains" and "View Clusters" buttons, respectively.

OLDERADO has also been integrated into the PDB 3DB Browser - it is accessed *via* the Atlas page if the result of a search is a NMR-derived ensemble. The link gives direct access to the OLDERADO database entry for this PDB entry; the information available is described in the preceding paragraph. Additionally, the OLDERADO methodology enables users to download *via* the EBI (with the aid of Kim Henrick in the Macromolecular Structure Group) an individual MODEL (by default, the "most representative"), rather than the entire ensemble, from an existing PDB entry. In cases where a user requires only a single MODEL, or a set of "representative" models, this reduces the bandwith required for download, reduces the diskspace required on the local machine, and eliminates the need to split a downloaded file.

References

Hyberts, S. G., Goldberg, M. S., Havel, T. F. & Wagner, G. (1992). The solution structure of eglin c based on measurements of many NOEs and coupling constants and its comparison with X-ray structures. *Protein Sci.* **1,** 736-751.

Kelley, L. A. & Sutcliffe, M. J. (1997). OLDERADO: on-line database of ensemble representatives and domains. On Line Database of Ensemble Representatives And DOmains. *Protein Sci.* **6,** 2628-2630.

Kelley, L. A., Gardner, S. P. & Sutcliffe, M. J. (1996). An automated approach for clustering an ensemble of NMR-derived protein structures into conformationally related subfamilies. *Protein Eng.* **9,** 1063-1065.

Kelley, L. A., Gardner, S. P. & Sutcliffe, M. J. (1997). An automated approach for clustering an ensemble of NMR-derived protein structures into conformationally related subfamilies. *Protein Eng.* **10,** 737-741.

Sutcliffe, M. J. (1993). Representing an ensemble of NMR-derived protein structures by a single structure. *Protein Sci.* **2,** 936-944.

Notes of a Protein Crystallographer –

FRODO, the Electronic Hobbit

Cele Abad-Zapatero

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

From early childhood, John Ronald Reuel Tolkien (J.R.R. Tolkien: 1892-1973) was fascinated with languages. When he was five, his mother - who was fluent in Latin, French, and German - taught him to read in all three languages plus her native English. Fatherless since 1896, the family lived in a small rented cottage in the hamlet of Sarehole by the Cole River, far from the smokestacks and soot of Birmingham. The quiet meadows and streams and Sarehole were a haven for Ronald and his younger brother Hilary. There his mother introduced them to botany and inspired in them a love for plants, trees and the beauty of natural landscapes. Nonetheless, change again came to his life abruptly. His mother died in 1904 and the brothers were left under the guardianship of a Catholic priest, Father Francis Morgan, who had a tremendous influence on his education and his life. Tolkien graduated from King Edward VI's school in Birmingham and won an award to attend Oxford University. His interest and passion for languages led him to study philology, specializing in the literary and linguistic tradition of the English West Midlands with extensive knowledge of Anglo-Saxon (or Old English as in *Beowulf*), Middle English (the language of Chaucer), and Finnish, Icelandic, Norse and Germanic mythologies and folklore. He was Professor of Anglo-Saxon at Oxford and a Fellow of Pembroke College from 1925 to 1945, Professor of English Language and

Literature and a Fellow of Merton College from 1945 until his retirement in 1959.

It is impossible to separate Tolkien's academic achievements from his creation of two muti-faceted, highly imaginative, epic stories which had a tremendous influence on the youth of the 1960's all over the world and whose effect still reverberates today. In 1937 he published *The Hobbit* (Tolkien, 1994), which received high acclaim as a fascinating children's story in which he introduced as main characters a 'hobbit' named Bilbo Baggins and a magician of sorts named Gandalf. Tolkien later wrote that the origin of the word hobbit seems to be: "a worn-down form of a word preserved more fully in the language of Rohan: *holbyta* or 'hole-builder'" (Tolkien, 1993b). What is a hobbit? In his own words:

" [..] They are (or were) a little people, about half our height, and smaller than the bearded dwarves. Hobbits have no beards. There is little or no magic about them, except the ordinary everyday sort which helps them to disappear quietly and quickly when large stupid folk like you and me come blundering along, making a noise like elephants which they can hear a mile off. They are inclined to be fat in the stomach; they dress in bright colours (chiefly green and yellow); wear no shoes, because their feet grow natural leathery soles and thick warm brown hair like the stuff on their heads (which is curly); have long clever brown fingers, good-natured faces, and laugh deep fruity laughs (especially after dinner, which they have twice a day when they can get it. Now you know enough to go on with." (Tolkien, 1994, p.3)

The illusion of hobbits as calm, simple people capable of heroic feats caught on quickly and Tolkien was asked to write more adventures of Bilbo Baggins. *The Hobbit* had ended with Bilbo keeping a ring that he had found during his encounter with Gollum, and living happily in the Shire: the idyllic part of Middle-earth where the hobbits lived and that scholars have related to the Sarehole of Tolkien's childhood (Neimark, 1996). The author had no desire to write a sequel. Instead, *The Fellowship of the Ring*, the first volume of the epic trilogy *The Lord of the Rings* was published in 1954. Soon after, the next two volumes appeared: *The Two Towers* and *The Return of the King*. The completed work was a mythological world of monumental proportions in which Tolkien had given life to creatures, kingdoms, wars, calendars, climates, places, landscapes, and seasons to give flesh and blood to the languages spoken by the people of Middle-earth: humans, elves, trolls, goblins, giants, dragons, ents, balrogs, orcs. The hero was Frodo, heir and nephew of Bilbo Baggins, who together with his friend Sam and other companions of the fellowship undertake a quest to destroy the master evil ring of Sauron that Frodo had inherited from his uncle. The appeal of an innocent, gentle creature succeeding in destroying the forces of evil against all odds, in an unspoiled landscape of pristine forests, mountains and lakes was enormous. By 1967, *The Lord of the Rings* had been translated into nine languages with an estimated readership of fifty million people. The graffito: FRODO lives! (Tolkien, 1993a), appeared in the New York subway as testimony to a cultural phenomenon that had opened a magic wonderland of places, characters and events unhindered by the prosaic incidents of our everyday lives. Tolkien had transcended the arcana of scholarly research in obscure languages to create a universal allegory of the constant struggle of good against evil, with strong environmental overtones.

FRODO, the electronic hobbit, had its origins in 1976. Whether the younger generations believe or not, at that time all protein models were built starting from a Cα tracing obtained from markings on an electron density map drawn on small plexiglass sheets stacked up as "mini-maps" (Jones, 1985). From these guide coordinates, detailed atomic models were built at a much larger scale on a Richards optical comparator known in the trade as "Richards Box" or "Fred's Folly" using Kendrew model parts (Richards, 1985). Glass or plastic windows had to be drawn by hand with tracings of the electron density contours at the appropriate scale (2 cm=1 $\mathring{A}$). Atomic coordinates were laboriously extracted from this wire model by tedious and often inaccurate protocols (Salemme, 1985). There was an immediate need for a computerized method that would allow the fitting of an atomic model to the experimental electron density map, and which would remove the tedium and inaccuracies from macromolecular structure determination and refinement (Editorial, 1997).

The idea was floating in the community and several laboratories had initiated projects to achieve that goal. Drs. J. Gassman and R. Huber found a bright young Welsh would-be crystallographer who was interested in living in Munich to develop such a tool, and encouraged him to make it a program useful for the routine operation in a protein crystallography laboratory. Tradition has it that the original program sent data back and forth between a PDP11 and a SIEMENS4004, in a computing environment where many of the programs were named after different hobbits. It was only natural that the central program will be named after the most famous of all the hobbits in Tolkien's trilogy. For obvious reasons, the test version used most of the computing cycles and was called initially SAURON.

FRODO made his appearance in the protein crystallography community twenty years ago in 1978 (Jones, 1978). As for myself, I got to know FRODO very well in 1981 during three beautiful weeks of immersion during the incomparable Swedish spring. Our friendship developed during many nocturnal model-building sessions at the old Wallenberg Laboratory next to the ancient city castle in Uppsala. I must confess that we had our crises, but he was certainly a very friendly hobbit. I was the one to blame for every crisis. Quite often, I failed to understand his prompts or suggestions, and many times his cues made no sense to me. He was always patient, effective and obedient.

You could CHAT (actual FRODO commands in capital letters) with him via a keyboard but the most effective way to communicate was with a tablet and a pen which would allow you to pick and identify atoms, and select different commands from a MENU on the screen. Obedient to the GO command, FRODO would display for you a certain volume of electron density and using well designed commands you could tell him to BREAK certain bonds and cut the protein chain into pieces. These pieces could then be moved with six degrees of freedom (FBRT) to make them fit into the three-dimensional electron density maps which could be rotated at will with dials.

FRODO did not know any protein chemistry, or if he did, he would not explicitly tell you so. It was you who would organize those constellations of points in space into a meaningful protein chain by using the REFInement command. He would faithfully apply the rules of chemistry to certain ZONEs of your spatial points which were covered by your electron density contours. This was a tremendous help when trying to fit those old electron density maps. FRODO was also very handy at modeling exercises by allowing you to create MOLecular objects that you could use either as background while fitting electron density or as objects of study in their own right.

For some time, the rumor (joke) floated in the community that the only documentation for FRODO was "*The Lord of the Rings*". This might have been true, but in his own humble way FRODO proved to be a very useful hobbit and was the ancestor of many other electronic hobbits that are now well settled in our computer underworld. In addition, his faithful friend SAM was always available to insert or delete residues, create a sequence and do all the necessary bookkeeping so that in the end everything was SAVEd in the disk with 'amazing speed' and accuracy. During my visit, FRODO lived in an independent VAX750 computer and his commands were translated into a Vector General VG3400. Later he lived inside many other boxes or hobbit-holes in many other countries. His performance improved as his electronic eyes and hands improved, permitting us to view unimaginable shapes and forms and to examine atomic continents, islands and landscapes of undescribable complexity and beauty. Following his original insights, we can now see atomic crevasses and caves, canyons, rivers, mountain ridges and valleys in different and vivid colours, and subtle hues and shades. FRODO opened for us an atomic underworld that was beyond our reach before. He introduced us to an atomic Middle-world that we could not have imagined without his assistance and that we are just beginning to explore, appreciate and understand.

One could argue that there are no malicious villains in our atomic Middle-world: no Dark Riders or Ringwraiths trying to prevent FRODO from destroying the evil ring. Yet, we routinely encounter, examine, and study molecules with pathogenic and curative properties in our crystals, and a major part of our time is spent trying to understand their interactions with themselves and with others. We are trying to defeat the evil forces of disease, pain and deformity and our operational domain is the atomic Middle-world that FRODO unveiled for us. There are parts of these atomic creatures that we cannot see or cannot fit well in our electron density maps, and that chase us in our sleep like the Dark Riders chased after FRODO and his friends. However, our true Gollum, Shelob and Sauron are uncertainty, lack of knowledge, and especially bias and disorder. Those restrictive forces will always be with us. In the meantime, FRODO will live on in the heart of those of us who -once upon a time- built protein models using mechanical parts and read the coordinates of our structures using a two-dimensional grid and a plumb line. He did so many things for us; he was such good a friend....

References

Editorial (1997), String and sealing wax. *Nature Struct. Biol.* **4,** 961-964.

Jones, T. A. (1978). A graphics model building and refinement system for macromolecules. *J. Appl. Cryst.* **11,** 268-272.

Jones, T. A. (1985). Diffraction methods for biological macromolecules. Interactive computer graphics: FRODO. *Methods Enzymol.* **115,** 157-171.

Neimark, A. E. (1996). *Myth Maker: J. R. R. Tolkien,* pp. 85-86, Harcourt Brace & Co., New York.

Richards, F. M. (1985). Optical matching of physical models and electron density maps: early developments. *Methods Enzymol.* **115,** 145-154.

Salemme, F. R. (1985). Some minor refinements on the Richards optical comparator and methods for model coordinate measurement. *Methods Enzymol.* **115,** 154-157.

Tolkien, J. R. R. (1993a). *The Lord of the Rings Trilogy. Part One: The Fellowship of the Ring.*Intro by Petr Bearle, Authorized edition of the fantasy classic by Ballantine Books, New York.

Tolkien, J. R. R. (1993b). *The Lord of the Rings Trilogy. Part Three: The Return of the King.* Appendix F. Authorized edition of the fantasy classic by Ballantine Books, New York.

Tolkien, J. R. R. (1994). *The Hobbitt.* 2nd Ed. Houghton Mifflin Company, New York.

Web Sites

Referenced in the January 1998 PDB Quarterly Newsletter

BioMagResBank (BMRB) NMR database
http://www.bmrb.wisc.edu

FTDOCK
http://www.icnet.uk/bmm/software.html

IUCr Policy on Publication
http://hobbes.gh.wits.ac.za/iucr-top/journals/acta/actaa_notes.html

mmCIF
http://ndb.rutgers.edu/NDB/mmcif/

MolView
http://bilbo.bio.purdue.edu/~tom

New BNL Contractor
http://www.pubaf.bnl.gov/pr/BSA.htm
http://apollo.osti.gov/doe/whatsnew/pressrel/pr97130.html
http://apollo.osti.gov/doe/whatsnew/pressrel/pr98001.html
http://www.ch.doe.gov/bnlseb

Nucleic Acid Database Submissions
http://ndbserver.rutgers.edu:80/NDB/deposition/index.html

Numerical Recipies
http://cfata2.harvard.edu/nr/

OLDERADO
http://neon.chem.le.ac.uk/olderado/

PDB AutoDep Submissions
http://www.pdb.bnl.gov
http://autodep.ebi.ac.uk

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

PMD, Parallel Molecular Dynamics
http://tincan.bioc.columbia.edu/pmd/pmd-summary.html

Program 123D
http://www-lmmb.ncifcrf.gov/~nicka/info.html

Representative MODEL Structures for a PDB Entry
http://www2.ebi.ac.uk/msd/nmr_search.shtml

SARF2
http://www-lmmb.ncifcrf.gov/~nicka/info.html
ftp://ftp.ncifcrf.gov/pub/SARF2/
http://defrag.bcm.tmc.edu:9503/lpt.html
http://cartan.gmd.de/nick/run2.html
http://genomic.sanger.ac.uk/

SFCHECK
http://www.sdsc.edu/Xtal/IUCr/CC/School96/

SGI CHALLENGEcomplib(tm)
http://www.sgi.com/Products/Challengecomplib.html

Structure Factor mmCIF Dictionary
ftp://ftp.pdb.bnl.gov/structure_factors/cifSF_dictionary

TOPS
http://tops.ebi.ac.uk/tops

Uppsala Electron Density Server
http://alpha2.bmc.uu.se/valid/density/form1.html

Affiliated Centers and Mirror Sites

Forty affiliated centers offer the Protein Data Bank database archives for distribution. These centers are members of the Protein Data Bank Service Association (PDBSA). Centers designated with an asterisk(*) may distribute the archives both on-line and on magnetic or optical media; those without an asterisk are on-line distributors only. Official PDB Mirror Sites are marked with a grey bar ( ) and are listed with their sponsoring center.

ARGENTINA

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

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

AUSTRALIA

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

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

BRAZIL

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

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

CANADA

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

CHINA

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

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

FINLAND

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

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

FRANCE

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

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

GERMANY

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

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

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

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

ISRAEL

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

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

ITALY

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

JAPAN

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

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

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

THE NETHERLANDS

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

POLAND

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

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

SWEDEN

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

TAIWAN

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

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

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

UNITED KINGDOM

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

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

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

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

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

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

SEQNET
Daresbury Laboratory
Warrington, United Kingdom
User Interface Group (44-1925-603351)
uig@daresbury.ac.uk
http://www.seqnet.dl.ac.uk

UNITED STATES

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

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

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

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

MAG
Molecular Applications Group
Palo Alto, California, USA
Margaret Radebold (415-846-3575)
bold@mag.com
http://www.mag.com

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

NCBI
National Center for Biotechnology Information
National Library of Medicine
National Institutes of Health
Bethesda, Maryland, USA
Stephen Bryant (301-496-2475)
bryant@ncbi.nlm.nih.gov
http://www.ncbi.nlm.nih.gov

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

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

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

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

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

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

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

Related WWW Sites

Databases

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

Software-Related Sites

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

Other Resources

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

PDB ™ Order Form

Name of User _____ Date _____

Organization _____ Phone _____

Address _____ Fax _____

_____ E-mail _____

- Price is valid through September 30, 1998
- Price is per CD-ROM set released — releases occur four times per year
- Facsimile and phone orders are not acceptable

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

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

   - Check payable to Brookhaven National Laboratory in U.S. dollars and drawn on a U.S. bank. Foreign checks cannot be accepted and will be returned.
   - Original purchase order payable to Brookhaven National Laboratory. After your order is processed, you will be invoiced by Brookhaven National Laboratory. Please indicate exact address to which invoice should be sent:

     _____

     _____

     _____

     _____

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

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

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

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

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

This page intentionally left blank.

Access to the PDB

Main Telephone ..... +1-516-344-3629

Help Desk Telephone ..... +1-516-344-6356

Fax ..... +1-516-344-5751

Help Desk ..... pdbhelp@bnl.gov

General Correspondence ..... pdb@bnl.gov

WWW Home Page ..... http://www.pdb.bnl.gov

FTP Server ..... ftp.pdb.bnl.gov

Network Services ..... sysadmin@pdb.pdb.bnl.gov

Entry Error Reports ..... errata@pdb.pdb.bnl.gov

Order Information ..... orders@pdb.pdb.bnl.gov

User Group ..... PDBusrgrp@suna.biochem.duke.edu

Listserver Postings ..... pdb-l@pdb.pdb.bnl.gov

Listserver Subscriptions ..... listserv@pdb.pdb.bnl.gov
to subscribe, the text of
your message should be **subscribe PDB-L Your Name**

FTP Directory Structure for Entries

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

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

all_entries/
coordinate entry files in compressed and uncompressed format

biological_units/
generated coordinates for the biomolecules

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

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

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

newly_released/
entries released since the last CD-ROM

nmr_restraints/
compressed NMR restraint files

obsolete_entries/
withdrawn and/or replaced entries

structure_factors/
compressed structure factor files

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

Scientific Consultants

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

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

PDB Staff

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

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

Statement of Support

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

Instructions to Authors

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

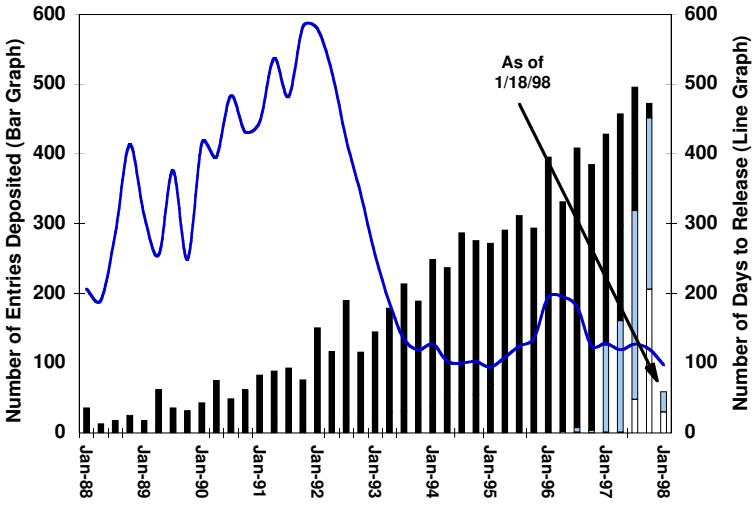
Protein Data Bank
Biology Department, Bldg. 463
Brookhaven National Laboratory
P.O. Box 5000
Upton, NY 11973-5000 USA

Telephone +1-516-344-3629 Fax +1-516-344-5751

Protein Data Bank pdb™ Brookhaven National Laboratory

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

Accumulated and Averaged on a Quarterly Basis



Bar Graph - Number of Entries in the Following Categories:

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

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

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

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

First Class

PRINTED ON RECYCLED PAPER