Protein Data Bank

Quarterly Newsletter

Release #75 *January 1996*

January 1996 CD-ROM Release

4162 full-release atomic coordinate entries

Molecule Type

3777	proteins, peptides, and viruses
88	protein/nucleic acid complexes
285	nucleic acids
12	carbohydrates

Experimental Technique

126	theoretical modeling
566	NMR
3470	diffraction and other

The total size of the atomic coordinate entry database is 1566 Mbytes uncompressed.

New Phone/Fax Numbers

Telephone 516-344-3629
Fax 516-344-5751

Table of Contents

Internet Sites

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

What's New at the PDB

Over the past two years we at the PDB have heard numerous reports of X-ray structure factors being lost, misplaced, or on some kind of tape or even punch cards that could not be read anymore. Frequently the crystallographers who actually collected the data or other members of their lab turn to the PDB for this data. As the structure factors are directly derived from the raw data measured in a crystallographic experiment, we feel that it is of the utmost importance that this data be deposited along with the coordinates in the PDB archive.

A great deal of discussion on this matter took place at the November 1995 International Seminar-cum-School on Macromolecular Crystallographic Data in Calcutta, India, sponsored by the International Union of Crystallography (IUCr). Virtually all participants felt that the deposition of the structure factors along with the coordinates is essential for the following reasons:

- Rigorous validation of the structure determination results can only be carried out using both atomic parameters and experimental structure factor amplitudes.
- Archiving of this data will ensure their preservation and continued accessibility (see article in our October 1995 Newsletter entitled PDB Structure Factor Files in CIF – A Proposal).

Some participants at this Seminar-cum-School felt that a number of crystallographers would be reluctant to submit their structure factors as they may want to continue their refinement before letting another group work on their data. This should not be a problem because the current policy of both the IUCr and the PDB provides crystallographers with the option of delaying the release of atomic parameters for up to one year and structure amplitudes for up to four years from the date of publication.

A short "Letter to the Editor" was composed by several participants which has now been sent to journals in which three-dimensional structural studies of macromolecules are published (see article entitled Letter to the Editor on Crystallographic Data Deposition). This letter urges journals to require deposition of not only atomic coordinates but also structure factors. The PDB feels that this is a very important development and urges members of the crystallographic community to encourage journals to follow this policy. One way in particular that this can be accomplished is that referees of journal articles should insist that the journal require deposition of coordinates and structure factors by the authors as a requirement for publication. We would welcome any thought you may have in this regard and would be pleased to publish these thoughts in future Newsletters.

– Joel L. Sussman

Revised Entry Format Description

The draft revision of the PDB Format Description has been finalized and is now available as Protein Data Bank Contents Guide: Atomic Coordinate Entry Format Description Version, Version 2.0. The full text is over one hundred pages and is accessible through PDB's home page on the WWW and *via* FTP in the /pub directory.

The PDB wishes to thank all those who made comments and suggestions about the draft document which has been available on the WWW.

Entries released after April 15, 1996 will comply with Version 2.0 of the Contents Guide. Conversion of older entries to this format will begin in the fall of 1996.

Letter to the Editor on Crystallographic Data Deposition

The following "Letter to the Editor" has been sent to journals in which X-ray crystallographic structure determinations of macromolecules are published.

A formal discussion of the archival journal requirements for data deposition was held at the November 1995 International Seminar-cum-School on Macromolecular Crystallographic Data in Calcutta, India.

The current policy of the International Union of Crystallography (IUCr) is that upon publication of a crystal structure determination of a macromolecule, the atomic parameters used or represented in the publication must be deposited in the Brookhaven Protein Data Bank. The deposition of structure amplitudes is recommended but not required. The policy provides crystallographers with the option of delaying the release of atomic parameters for one year and of structure amplitudes for up to four years from the date of publication. Participants strongly supported this policy and felt it should be strictly applied by the journals (referees).

Recent developments in X-ray crystallographic experimental and refinement techniques and the huge expansion in computing power and networking, however, necessitate the review of deposition arrangements.

It was noted that the new validation procedures are much more effective but require the experimental structure amplitudes as well as the atomic parameters. In addition, the technical arrangements for deposition, analysis, and validation of macromolecular crystal structures are now much easier.

The undersigned consider it vital for the macromolecular crystallographers to respond to these developments in their deposition practices. We recommend, therefore, that publication of macromolecular crystal structures should be accompanied by deposition of atomic parameters and also structure amplitudes. Amongst the many reasons identified for this practice, the following two are critical:

- Rigorous validation of the structure determination results can only be carried out using both atomic parameters and experimental structure amplitudes. It is important that journals ensure that referees have sufficient information to prevent incorrect structures being published.

- Archiving of this data will ensure they are not lost. There were numerous reports at this Meeting of data being lost. This most probably reflects a general problem in the crystallographic community.

Edward N. Baker - Member of IUCr Executive Committee and Member of the IUCr Commission on Biological Molecules
Department of Chemistry and Biochemistry
Massey University
Palmerston North
New Zealand

Tom L. Blundell
ICRF Unit of Structural Molecular Biology
Department of Crystallography
Birkbeck College
Malet Street
London, WC1E 7HX
England

Mamannamana Vijayan - Chairman of IUCr Commission on Biological Molecules
Molecular Biophysics Unit
Indian Institute of Science
Bangalore 560012
India

Eleanor Dodson - Member of IUCr Electronic Publishing Committee
Department of Chemistry
University of York
York, YO1 5DD
England

Guy Dodson - Previous Chairman of IUCr Commission on Biological Molecules
Department of Chemistry
University of York
York, YO1 5DD
England

Gary L. Gilliland, Associate Director
Center for Advanced Research in Biotechnology
9600 Gudelsky Drive
Rockville, MD 20850
USA

Joel L. Sussman - Head, Protein Data Bank
Departments of Biology and Chemistry
Brookhaven National Laboratory
Upton, NY 11973
USA
and
Department of Structural Biology
Weizmann Institute of Science
Rehovot 76100
Israel

Change in Policy for Deposition of Nucleic Acid Structures Determined by X-ray Crystallography

This article was written by Helen M. Berman, Head, Nucleic Acid Data Base, Rutgers University, Piscataway, NJ, USA ***(berman@dnarna.rutgers.edu)*** *and Joel L. Sussman, Head, Protein Data Bank, Brookhaven National Laboratory, Upton, NY, USA* ***(jls@bnl.gov)****).*

Starting January 1, 1996 data for crystal structures of oligonucleotides should be deposited directly with the Nucleic Acid Database (NDB). Once the data are processed they will be forwarded to the PDB for deposit in the central single archive. This will simplify current procedures and make the data on nucleic acids available more quickly. Protein/nucleic acid complexes and all NMR structures should continue to be deposited at the PDB. All crystal structure data for DNA and RNA will continue to be available from both the NDB and the PDB.

To deposit the data, submit the coordinates, structure factors, and current PDB deposition form to: **deposit@ndbserver.rutgers.edu**.

A preprint of the related manuscript should be sent by fax to 908-445-5958 or by postal mail to Dr. Anke Gelbin, The Nucleic Acid Database, Department of Chemistry, Rutgers University, POB 939, Piscataway, NJ 08855, USA.

Current PDB Submission Procedures and Requirements

There are three essential elements of a complete PDB deposition:

- PDB-formatted coordinate data.

- A completed up-to-date version of our Electronic Deposition Form which you can pick up off of the PDB WWW home page **(http://www.pdb.bnl.gov)**, download from our FTP server **(ftp.pdb.bnl.gov)**, or request *via* e-mail from the PDB. Please fill out this Deposition Form using an on-line editor rather than by hand, as we run a program on the form to generate a preliminary header for your final PDB entry. As our older, non-electronic Deposition Forms are more difficult to fill out and because those forms are not as complete, we request that you no longer send the old forms to us. Our new Form has been available for about one and a half years, and we prefer that you use it.

- Copies of all relevant preprints and reprints OR a copy (print-out) of your submitted manuscript with the PDB Tracking Number it relates to prominently noted. As questions frequently arise about this, we would like to stress that IF YOUR PAPER IS NOT AT THE PREPRINT STAGE YET, a regular xerox or print-out of the submitted manuscript fulfills this requirement. If there is no manuscript in progress, please indicate so in the Journal (JRNL) section of the Deposition Form and the print requirement will be waived.

  Referenced papers can be sent electronically using FTP or e-mail, by fax to 516-344-5751, or by postal mail to:

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

  FOR DEPOSITORS NEEDING TO OBTAIN AN ID CODE AS QUICKLY AS POSSIBLE, WE SUGGEST YOU E-MAIL OR FAX YOUR MANUSCRIPT.

  References are kept completely confidential and are used only to aid us with the processing of your entries and to ensure that we reference related papers correctly in the entries themselves.

The PDB now issues ID codes to depositors as soon as we receive the above three items. For more information on PDB submissions, please contact Minette Cummings at **pdb@bnl.gov**.

Replacement Data Procedure

When sending replacement data to the PDB, please send an accompanying e-mail to **pdb@bnl.gov** indicating the related Tracking Number or ID code as well as a listing of all files sent.

O! O! Oops!

*This article was written by Gerard J. Kleywegt, Department of Molecular Biology, Biomedical Centre, Uppsala University, Uppsala, Sweden **(gerard@xray.bmc.uu.se)**.*

OOPS is a little utility program for people who use O to rebuild their crystallographic protein models. Although the program itself is absolutely trivial, it offers two major benefits:

1. It focuses the crystallographer's attention on the trouble spots in the current model.

2. It can be instructed to skip residues which appear to be perfectly in order, thereby saving much rebuilding time.

Basically, OOPS is a simple filter. As input it takes a number of files produced by O (or other programs) which contain information about the quality of the current model on a per-residue basis. The major output is a set of rebuilding macros for O which will take the crystallographer on a journey past all residues which may need attention because they scored poorly in one or more of the quality tests. The macros are "chained", which means that when one has finished rebuilding a suspicious residue, a click of the mouse will take the crystallographer to the next suspect.

At present, OOPS can check the following quality indicators (amongst others):

- Bad pep-flips (a measure for the distance between a peptide oxygen orientation and those encountered in the database).

- Bad real-space-fit values (the correlation or R-factor between calculated and 2Fo-Fc density for any or all atoms in a residue).

- Bad rotamer side chain (RSC) fit values (a measure of how well the side chain conformation resembles that of a rotamer).

- Too high and too low temperature factors and occupancies.

- Bad ϕ, ψ angle combinations.

- Poor peptide planarity.

- Poor Cα chirality.

In addition, the current model can be compared to the previous one (in terms of displacements, temperature factor, and occupancy changes, as well as changes in the main and side chain torsion angles). Moreover, up to ten user-defined criteria can be used (*e.g.*, quality of the geometry from X-PLOR, or the number of bad contacts from Procheck).

The output of OOPS consists of:

- Statistics for most of the used quality indicators.
- Plot files for some of the criteria (as a function of residue number).
- A list of potentially bad residues, plus their faults.
- A list of the violation counts for each quality criterion.
- A set of O rebuilding macros.
- A small file with PDB REMARK records pertaining to the quality of the current model. If the current model is the final one, these records can be included in the PDB Deposition Form.

Using OOPS requires some O datablocks to be prepared in advance (however, there is an O macro available to do most of that work for the user as well). Running the program takes only a few minutes. The result in terms of speed-up of the rebuilding process is well worth this small effort. Also, OOPS makes it less likely that residues with serious errors in them are overlooked and may therefore help improve the quality of the model.

OOPS is one in a series of "O-dalisques", *i.e.*, programs that work in conjunction with O. The OOPS program runs on SGI, ESV, and DEC ALPHA/OSF1 workstations. For more information, contact Gerard Kleywegt *via* e-mail **(gerard@xray.bmc.uu.se)**.

Rasmol's New Home Page (http://www.umass.edu/microbio/rasmol)

*This article was written by Eric Martz, Department of Microbiology, University of Massachusetts, Amherst, MA, USA **(emartz@microbio.umass.edu)**.*

RasMol has a new home page dedicated to its popularization and distribution, educational uses of RasMol, and access to atomic coordinate (PDB) data files for macromolecules and other free molecular visualization resources. The introductory documents on this home page are written for the scientific public (biologists not specializing in protein structure, high school teachers, *etc.*). Don't miss the personal history of the origin of RasMol by Roger Sayle (author of RasMol)! An e-mail list has also been set up for announcement of new releases of RasMol and discussion of its use (details on the home page).

RasMol is a highly capable and amazingly fast program for molecular visualization. It runs on Windows, MacIntoshes, and unix systems *via* X-windows. The University of California, Berkeley MultiChem Facility offers an enhanced version of RasMol (now available for the MacIntosh; Windows version under development). This can display several molecules at once and move them relative to each other. RasMol is the generous gift to the scientific public from Roger Sayle, the University of Edinburgh, and Glaxo Research and Development (Sayle's employer, Greenford, UK), who are continuing to upgrade and distribute it free.

The RasMol page offers the first publicly available PDB formatted file for an intact antibody molecule, provided by Eduardo Padlan of the National Institutes of Health. Other contributions of wide interest would be welcomed (lipid bilayers, T-cell antigen receptor complexes with MHC:peptide, *etc.*). Several RasMol scripts are provided. One, designed to introduce RasMol to general biological science audiences/classes, shows the power and glory of RasMol by illustrating several aspects of the structure of the DNA double helix. Scripts for more specialized audiences/classes treat the structure of antibody and antibody interaction with antigen, as well as the structure of the major histocompatibilty complex and its binding to peptide antigens. Links are offered to class WWW pages at various universities which make extensive use of RasMol. These include New York University's "Mathematics & Molecules" aimed at K-12 groups, organic chemistry (Virginia Polytechnic Institute and State University), biochemistry (University of California, Santa Barbara and Carnegie-Mellon University), immunology (University of Massachusetts, Amherst), and chemotherapy and drug design (Leeds University, UK). As this issue went to press, the RasMol home page was receiving visits from nearly one thousand people per week, from over thirty countries.

The RasMol home page and e-mail list were set up by Eric Martz, a professor at the University of Massachusetts, Amherst, MA, USA **(emartz@microbio.umass.edu)**. Eric welcomes suggestions of additional resources/improvements for the RasMol page.

New Server at Weizmann's Bioinformatics Unit: Blocks

*This article was written by Jaime Prilusky, Bioinformatics Unit, Biological Services, Weizmann Institute of Science, Rehovot, Israel **(lsprilus@inherit1.weizmann.ac.il)**.*

Detection and verification of protein sequence homology is now available at the BLOCKS WWW Server at the Bioinformatics Unit, Biological Services, Weizmann Institute of Science, Rehovot, Israel at URL **http://bioinformatics.weizmann.ac.il/blocks**. This Server is provided in collaboration with Jorja Henikoff **(jorja@howard.fhcrc.org)** from the Fred Hutchinson Cancer Research Center, Seattle, Washington, USA.

Blocks are multiply-aligned, ungapped segments corresponding to the most highly conserved regions of proteins.

Block Searcher, Get Blocks, and Block Maker are aids to detection and verification of protein sequence homology. They compare a protein or DNA sequence to a database of protein blocks, retrieve blocks, and create new blocks, respectively.

– The Blocks Database

The blocks for the BLOCKS database are made automatically by looking for the most highly conserved regions in groups of proteins represented in the PROSITE database. These blocks are then calibrated against the SWISS-PROT database to obtain a measure of the chance distribution of matches. It is these calibrated blocks that make up the BLOCKS database. The WWW versions of the PROSITE and SWISS-PROT databases that are used on this Server are located at the ExPASy WWW Molecular Biology Server of the Geneva University Hospital and the University of Geneva **(http://expasy.hcuge.ch)**.

The blocks created by Block Maker are created in the same manner as the blocks in the BLOCKS database but with sequences provided by the user. Results are reported in a multiple sequence alignment format without calibration and in the standard BLOCK format for searching.

Internet Course in Principles of Protein Structure

*This article was written by John Walshaw, Jacky Turner, and David Moss, Crystallography Department, Birkbeck College, University of London, London WC1E 7HX, UK **(pps2@mail.cryst.bbk.ac.uk)**.*

The pilot year of an undergraduate level course in Principles of Protein Structure, the first international multimedia science-education course to be taught entirely *via* the Internet, finished in June 1995. Over 150 students, teachers, and advisors from twenty-seven countries were involved, and about seventy students finally “graduated”. The course, a collaboration with the Texas-based Virtual School of Natural Sciences, was pioneered at Birkbeck by Peter Murray-Rust of Glaxo-Wellcome and Alan Mills now of Venus Internet Ltd.

Beginning on January 15, 1996 an updated PPS course will run — this time as a one-year, part-time Advanced Certificate course accredited by Birkbeck College (University of London) and available worldwide to those with access to a suitable networked computer. The course comprises three units (approximately one per term): bioinformatics, protein structure, and a dissertation. Students will be familiarized with the biological Internet, including some technical issues behind the genomic, protein sequence, and structural databases.

The core is the study of protein structure, progressing from fundamentals to recent developments and current research. The course material on the WWW is public to all, whether they are involved in the course or not. In the second half of the course, students will produce a hypertext dissertation to be mounted on the WWW.

An equally important component is the interaction between students, teachers, and an international network of volunteer consultants with a range of specialist interests. We have had a fantastic response from these volunteers.

Students are divided into small e-mail discussion groups, each with a tutor from the Crystallography Department who holds discussions and tutorials on the week’s work. Students are to spend approximately six hours per week on the course. We are very fortunate in having several rooms in the BioMOO at the Weizmann Institute for “virtual” tutorials and the enthusiastic cooperation of both Brookhaven National Laboratory **(http://www.pdb.bnl.gov/PPS2/index.html)** and Daresbury Laboratory **(http://www.dl.ac.uk/PPS/index.html)** who are providing mirror services.

The academic level is formally that of the final year of an undergraduate degree, but inevitably more specialized. The Internet medium allows us to tailor the course to the interests, needs, and talents of individual students. There is a wide range of backgrounds among the students currently enrolled. Many are professional research scientists or Ph.D. students, who will enrich the course by sharing their expertise. (Unfortunately the regulations do not allow current undergraduate students to register for the course, except in the case where the student’s university decides to incorporate it as one credit unit in the undergraduate degree.)

The Advanced Certificate course consists of three 11-week terms beginning on January 15, 1996 and ending in October 1996. The cost is 186 pounds sterling for EU students and 500 pounds sterling worldwide. Enrollment is now taking place.

Anyone interested in the course, either as a student or a consultant, please refer to **http://www.cryst.bbk.ac.uk/PPS2/** or send e-mail to **j.mcgill@mail.cryst.bbk.ac.uk**.

Training Courses in Parallel Programming and High Performance Computing

As part of the Supercomputing Resource for Molecular Biology (SRMB) programme, the European Molecular Biology Laboratory in Heidelberg, Germany is offering training courses in Parallel Programming and High Performance Computing to European researchers in molecular biology.

Participants will receive instruction in writing message passing and data parallel programs using PVM, MPI, and Fortran90/HPF. General tuning and performance optimization techniques, suitable for today’s high performance RISC microprocessors, will also be covered. All topics are complemented by hands-on training sessions, using the parallel supercomputer facilities at EMBL-Heidelberg.

This course is open to European Molecular Biologists at the advanced post-graduate level with research interests in sequence analysis, image processing, structural refinement, protein design, and molecular dynamics.

Visitors from EU and associated countries will have travel and accommodation expenses funded by an EU HCM/ALSI grant. Visitors from other EMBL member states will be supported by funds from EMBL.

The next course is scheduled for February 19-23, 1996. Please see **http://www.embl-heidelberg.de/Services/srmb/pphpc_course/** for information from EMBL.

For additional information on the SRMB programme, visit URL **http://www.embl-heidelberg.de/Services/srmb/** or write to:

SRMB Secretary
Biological Structures and Biocomputing Programme
European Molecular Biology Laboratory
Postfach 10.2209
D-69012 Heidelberg
Germany

Phone: +49 6221 387 271
Fax: +49 6221 387 306
E-mail: SRMBadmin@EMBL-Heidelberg.de.

The International School-cum-Seminar on Macromolecular Crystallographic Data

This article was written by Geoffrey B. Jameson, Department of Chemistry and Biochemistry, Massey University, Palmerston North, New Zealand ***(G.B.Jameson@massey.ac.nz)****.*

The International School-cum-Seminar on Macromolecular Crystallographic Data (ISMCD) was held at the Saha Institute of Nuclear Physics in Calcutta, India, from November 16-20, 1995. Over 150 participants were present from fourteen countries, and included representatives from the Protein Data Bank (Joel Sussman and S. Swaminathan), the Nucleic Acid Database (John Westbrook), the Biological Macromolecule Crystallization Database (Gary Gilliland), the International Union of Crystallography Editorial Office (Brian McMahon), and the mmCIF project (Philip Bourne).

Three broad themes relating to macromolecular crystallographic data were interwoven throughout the ISMCD. The first related directly to databases of macromolecular structure. The current status and future plans of the PDB and the NDB were presented. The PDB, in particular, is growing exponentially and very rapidly. Ease of deposition, efficient validation of data, and improved and more sophisticated access were addressed. The Crystallographic Information File (CIF, now widely accepted and used in the small molecule community) and the macromolecular version (mmCIF) provide a standardized format and content for data deposition that, in cooperation with the community of programmers, should facilitate data deposition and data validation. Eleanor Dodson described the overall CCP4 structure and then focused on refinement strategies and indicators for reliability and precision of refined structures. A lively discussion, chaired by Guy Dodson, ensued on data deposition and withholding of deposited data, especially structure factors. Even with a one-year hold on coordinates and a four-year hold on structure factors, concerns were expressed about both the loss of intellectual property, especially to entities not prone to releasing their own data, and the consequences resulting from detection of major errors. Others stressed the importance of depositing data in the PDB as a guard against loss of data as well as the responsibility of scientists (and of journals) to make available raw data ensuring that (re)analysis of and comparisons among known structures and the development of new insights are not stifled or restricted by non-availability of data.

The importance of having available at least the coordinate data from the PDB was underscored by the second theme of the conference — the mining of databases for deeper insight into macromolecular structure and function. Tom Blundell described superfamilies containing proteins of weak sequence homology but similar structure and general function, a topic examined in detail by Ted Baker for the four-helix bundle structure of cytochromes c' and by M. Vijayan for quaternary association of lectins. A complementary perspective was offered by Guy Dodson of the diverse and unrelated structure types of the hydrolases in which catalytic triads share a common mechanism of nucleophilic attack on amides, esters, and related substrates. Databases underpin molecular modelling for structure prediction (Tom Blundell and an alphabet soup of programs for proteins, M. Bansal and N. Yathindra for oligonucleotides); for structure solution by molecular replacement (Jorge Navaza, M.R.N. Murthy, and K. Suguna); for structure refinement and drug design (T. Bhat and Jose Varghese); for virtual reality approaches (N. Seshagiri); and for designer mutants (R. Varadarajan and P. Balaram). Importantly, databases house information which in an individual structure lacks statistical significance, but which when repeated in many structures acquires validity, as exemplified by the C(α)-H..O=C(main chain) hydrogen bonds found in beta sheets (V. Pattahbi) and by the rarity of “unusual” conformations (P. Balaram). In a series of hemoglobin structures, the use of a common database of restraints and refinement protocols revealed small but systematic and significant structural effects of pH, with implications for the mechanism of cooperativity (Guy Dodson).

The final theme involved additions of new structures to macromolecular databases. Recent oligonucleotide structures were presented by M. Sundaralingam (a novel U-U C-H..O hydrogen bond), C. Betzel, and N. Gautham. Recent protein structures, additional to the ones mentioned above, included bovine cytochrome c oxidase — a monumental crystallographic achievement (T. Tsukihara), beta-lactoglobulin revisited (Maria Bewley), a Kunitz-type chymotrypsin inhibitor (J.K. Dattagupta), two new species of lactoferrin (T.P. Singh), a double mutant of D-xylose isomerase for which quantum chemical analysis provided insight into the origins of metal specificity for effective catalysis (Monica Fuxreiter), and the complex of acetylcholinesterase with the snake neurotoxin fasciculin (Joel Sussman).

The scientific content was outstanding, from which only a selection has been presented above. Location in India and generous sponsorship from the IUCr provided opportunities for attendance at an international conference by students and post-doctoral fellows for whom attendance otherwise would have been difficult. The strength, depth, and vitality of structural biology in India was evident. Finally, the success of the Conference was ensured by the organizing team headed by J.K. Dattagupta (Saha Institute of Nuclear Physics in Calcutta) and M. Vijayan (Indian Institute of Science in Bangalore) through their meticulous attention to logistical detail and an extraordinary level of hospitality extended to all participants.

Targets for Protein Structure Prediction Results Now Available on WWW

This article was written by Tim Hubbard, Centre for Protein Engineering, MRC Centre, Cambridge, UK and Anna Tramontano, Istituto di Ricereche di Biologia Molecolare, Pomezia, Rome, Italy ***(th@mrc-cpe.cam.ac.uk)****.*

Between July and September 1995, announcements were sent to the PDB's Listserver mailing list inviting the submission of protein sequences of unknown structure as prediction targets for the IRBM practical course entitled Frontiers of Protein Structure Prediction.

This Workshop was held from October 8-17, 1995 at the Istituto di Ricereche di Biologia Molecolare (IRBM). One hundred thirteen target submissions had been received by then and these were automatically analyzed to screen for homologies to known structures and provide raw material for the course. Of these, twelve were predicted during the Workshop, at different levels of detail.

The results of the analysis carried out on each of the 113 target proteins, the detailed reports on the twelve predictions, a short description of all the methods used, and the documentation provided by each teacher are all now publicly available at the URL **http://www.mrc-cpe.cam.ac.uk/irbm-course95**/. A large amount of general documentation written for the course is also available at this URL.

Our sincere thanks to all those who took the time to fill in the forms to submit their sequences and apologies to those whose sequences were not worked on during the course due to the limited time available.

Notes of a Protein Crystallographer

– The Ballad of the 2.8 Å Structure of SBMV

This article was written by Cele Abad-Zapatero, Abbott Laboratories, Abbott Park, IL, USA ***(abad@abbott.com)****.*

Most people would associate the term "ballad" with past achievements which could go back as far as the origins of story telling. It is certainly unusual to read or even hear this word associated with contemporary scientific events. Even more so for the solution of the three-dimensional structure of a macromolecule: they are so commonplace nowadays. The scientific journals are inundated with beautiful color pictures advertising on their cover the solution of yet another protein structure.

However, in the early nineteen seventies a modern research project in structural biology reminded me of the mighty feats of medieval heros. For nine long years, several generations of valiant postdocs led by Professor M.G.R. struggled to adapt the methodology of protein crystallography to the solution of the atomic structure of the first icosahedral virus particle. There were two other groups working on a similar endeavor. One at Harvard led by Professor Steve Harrison, focused on Tomato Bushy Stunt Virus (TBSV) — at the time the best structurally characterized virus. A second in Uppsala, Sweden, trying to solve the structure of Satellite Tobacco Necrosis Virus (STNV), a very small satellite virus, under the *aegis* of Professor Bror Strandberg.

The achievement established new methodology which is currently used; and the results obtained opened intriguing lines of research on the structure, function, and evolution of viruses. As I worked on the project, the stanzas of a ballad came to my mind as the most natural way to express my admiration for the feat of all the participants. If you do not hear from the other groups, it is not because their achievement was less significant. Absolutely not; they simply did not have their *balladeer*.

The virus of our story is Southern Bean Mosaic Virus (SBMV), a humble RNA-containing plant virus which infects bean plants in the South of the United States. Neither SMBV nor its relative TBSV were ever as famous as the animal viruses that are fashionable today as human pathogens. However, small (approximately 300 Å in diameter), nonenveloped, single-stranded, RNA plant viruses like they, were easy to obtain in gram quantities from a few infected plants. In addition, they were easy to crystallize and consequently they were the object of a concerted effort to obtain their atomic structure by X-ray diffraction methods.

The icosahedral symmetry of small spherical viruses had been proposed by Watson and Crick in the early fifties, and the detailed arrangement of the proteins in the capsid on the surface was described by Caspar and Klug in their classic 1962 paper. Nonetheless, there was yet no atomic model for an icosahedral virus particle. As initially proposed by M.G.R., the crucial factor in the determination of the structure was the presence of several identical copies of the polypeptide chain in the asymmetric unit. In those days the major hurdle was to devise algorithms and programs which would allow averaging of enormous electron density maps, containing many millions of grid points, over the redundant copies in the asymmetric unit.

One of the most striking results of the structure determination of SBMV was the similarity of folds between the protein capsids of SBMV and TBSV, and later STNV. This unifying principle has had an enormous impact in understanding the structure, function, evolution, and diversity of a wide spectrum of viruses.

Someday I may have the time and space to print the entire ballad, including the music. For this occasion, I would like to include only a few stanzas to give the reader a sense of the innuendoes of the text and of its rhythm and flow.

M.G.R. began to work on small viruses soon after the structure of LDH had been published by taking a sabbatical leave in Uppsala, in the laboratory of Bror Strandberg.

LDH has now been solved
I must find something to do
Rossmann fold has been proposed
I'll take a sabbatical leave (repeat)
And I'll look at the STNV.

Soon after the sabbatical leave, he started to work on SMBV. Afternoon tea was a ritual at that time in the lab where the work progress was discussed.

Shall we start by growing some crystals?
It's only a matter of weeks
After that we can write some proposals
For the future of SBMV (repeat)
You should drink your afternoon tea.

Although M.G.R.'s dream was to solve viruses *ab initio*, he soon realized that heavy atom derivatives were a safer route at the time. Of course, he continued to sail in Lake Freeman, Indiana.

Heavy-atoms must now be found
Playing chemists is all we must do
One alone will be safer ground
For the structure of SBMV (repeat)
I'm sailing the Indiana sea.

After eight years of work, the atomic model of SBMV slowly grew as a metallic sculpture made up of Kendrew parts in a forest of rods within a Richard's Box. M.G.R. kept bumping his head against the top of the box, so he purchased a hard hat which he rigorously wore for his building sessions with me.

Eight years have already passed
Many people have done their best
I won't say the struggle has finished
For the structure of SBMV (repeat)
I'll buy a helmet for me.

The fold of SBMV turned out to be a β-barrel almost identical to the one already described by Steve Harrison for the structure of TBSV.

After so many years of labor
All we have is a barrel of sheet
Steve H. did us a favor
With the structure of TBSV (repeat)
We can trace our SBMV.

The entire original text of the ballad was sung at a party at the Rossmann's residence to celebrate the structure solution of SBMV. The melody was adapted from a song by Pete Seeger that I heard on the radio one beautiful autumn morning on my way to the lab. The entire ballad was meant to be an homage to all participants in the project of the three-dimensional structure of SBMV. Many of them I have met through the years at meetings and conferences. I shared with others hours of effort, frustration, and excitement in the basement of the Lilly Hall of Life Sciences at Purdue University, a unique laboratory whose day-to-day routine is still masterly orchestrated by Sharon Wilder. To all of them (the unsung heros of this ballad) and to many others who participated in less visible ways, I would like to express my deep appreciation: Sherin Abdel-Meguid, Toshio Akimoto, J.E. (Jack) Johnson, Andrew G.W. Leslie, Ivan Rayment, Michael Rossmann, Ira Smiley, Dietrich Suck, Tomitake Tsukihara, and Mary Ann Wagner. I am just a modest minstrel, the troubadour of this epic feat.

Personnel Changes

We wish to welcome Janet Sikora to our staff at PDB. Janet accepted a position as Office Services Assistant last August and will be performing various administrative duties.

Karen Smith has left the PDB in order to devote herself full-time to the caring of her new son, born last September. We would like to sincerely thank Karen Smith for her four years of dedication and hard work with the PDB. Karen was known to our depositors as the receiver of all depositions. Our very best wishes go to her and her family for the future.

Karen's responsibilities are presently being handled by Minette Cummings. Depositions sent by e-mail to **pdb@bnl.gov** or transmitted by FTP to **ftp.pdb.bnl.gov** will be received and acknowledged by Minette.

New Telephone Number Exchange

The PDB has new telephone numbers as the result of a recent service upgrade at Brookhaven National Laboratory. All phone numbers now begin with the area code 516 and exchange 344. For example, the PDB's main number is now 516-344-3629. ■

New PDB Handout

A handout is now available which lists important PDB numbers and addresses, software available *via* anonymous FTP, and sites on the World Wide Web of interest to our users. This handout was first produced during the summer of 1995, when PDB participated in several scientific meetings and workshops. As part of our participation we produced a handout the size of one half sheet of paper. These handouts were so well received that we are keeping them up-to-date and will have them with us at future meetings.

Below is the February 1996 version of the card.

Who: Protein Data Bank, Brookhaven National Laboratory

What: Database of 3-dimensional structures of biological macromolecules including atomic coordinates, crystallographic and NMR experimental data, primary sequence, and secondary structure information. Links to other bio-information databases are also maintained.

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

TO REACH THE PDB

Telephone	1 516-344-3629
Fax	1 516-344-5751
Help Desk	pdbhelp@bnl.gov
E-mail and Depositions	pdb@bnl.gov
PDB WWW Home Page	http://www.pdb.bnl.gov
FTP Server	ftp.pdb.bnl.gov
Gopher Server	gopher.pdb.bnl.gov
PDB User Group	PDBusrgrp@suna.biochem.duke.edu
PDB Listserver Subscriptions	listserv@pdb.pdb.bnl.gov
to subscribe	subscribe PDB-L Your Name
PDB Listserver Postings	pdb-l@pdb.pdb.bnl.gov
Network Services	sysadmin@pdb.pdb.bnl.gov
Deposition Form	on PDB WWW home page
Newsletter	on PDB WWW home page
Electronic Newsletter Mailing List	listserv@pdb.pdb.bnl.gov
to subscribe	subscribe NEWSLETTER-L Your Name
PDB CD-ROM Order Form	on FTP server as /pub/orderfrm.ps

SOFTWARE AVAILABLE VIA ANONYMOUS FTP

PDB-Browse	ftp.pdb.bnl.gov	/pub/pdbbrowse
PDB-Shell	ftp.pdb.bnl.gov	/pub/pdbshell
RasMol	ftp.pdb.bnl.gov	/pub/other-software/Rasmol
Kinemage	ftp.pdb.bnl.gov	/pub/kinemage
	suna.biochem.duke.edu	/pub/MACprograms or /pub/PCprograms or /pub/UNIXprograms
Mirror	ftp.pdb.bnl.gov	/pub/other-software/Mirror
Perl	ftp.pdb.bnl.gov	/pub/other-software/Perl
	ftp.netlabs.com	/pub/outgoing/perl5.0
Tcl	ftp.pdb.bnl.gov	/pub/other-software/Tcl
WHAT_CHECK	ftp.pdb.bnl.gov	/pub/whatcheck
Gopher client	boombox.micro.umn.edu	/pub/gopher
Mosaic client	ftp.ncsa.uiuc.edu	/Web/Mosaic
Netscape client	ftp.netscape.com	/netscape

INTERESTING WWW SITES

American Crystallographic Association	http://nexus.hwi.buffalo.edu/ACA
BioMagResBank	http://www.bmrb.wisc.edu
BMCD-The Biological Macromolecule Crystallization Database and the NASA Archive for Protein Crystal Growth Data	http://ibm4.carb.nist.gov:4400/bmcd/bmcd.html
Brookhaven National Laboratory, Biology Department	http://bnlstb.bio.bnl.gov:8000
Cambridge Crystallographic Data Centre	http://csdvx2.ccdc.cam.ac.uk
Crystallography Worldwide	http://www.unige.ch/crystal/w3vlc/crystal_index.html
DALI-Comparison of Protein Structures in 3D	http://www.embl-heidelberg.de/dali/dali.html
EBI-European Bioinformatics Institute	http://www.ebi.ac.uk
ExPASy Molecular Biology Server	http://expasy.hcuge.ch
Genome Data Base (GDB)	http://gdbwww.gdb.org
International Union of Crystallography	http://www.iucr.ac.uk
Johns Hopkins University BioInformatics	http://www.gdb.org
mmCIF	http://ndbserver/rutgers.edu:80/mmcif
National Institutes of Health	http://www.nih.gov
NCBI GenBank	http://www.ncbi.nlm.nih.gov
Nucleic Acid Database	http://ndbserver.rutgers.edu
Pedro's BioMolecular Research Tools	http://www.public.iastate.edu/~pedro/research_tools.html
Protein Data Bank (PDB)	http://www.pdb.bnl.gov
Protein Identification Resource (PIR)	http://www.gdb.org/Dan/proteins/pir.html
Protein Motions Database	http://hyper.stanford.edu/~mbg/ProtMotDB/
Protein Science	http://www.prosci.uci.edu
Protein Structure Verification-Biotech Server	http://biotech.embl-heidelberg.de:8400 or http://biotech.pdb.bnl.gov:8400
RasMol Home Page	http://klaatu.oit.umass.edu:80/microbio/rasmol
SCOP-Structural Classification of Proteins	http://scop.mrc-lmb.cam.ac.uk/scop or www.pdb.bnl.gov/scop
Swiss-Prot Sequence Database	http://expasy.hcuge.ch/sprot/sprot-top.html
Weizmann Institute, Biological Computing Division	http://dapsas1.weizmann.ac.il

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

BROOKHAVEN ORDER FORM

Name of User ____________________ Date ____________________

Organization ____________________ Phone ____________________

Address ____________________ Fax ____________________

____________________ E-mail ____________________

____

- ***Price is valid through September 30, 1996***
- ***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;
3. Payment (using one of the two options below):

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

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

____

____

____

____

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

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

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

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

1 Protein Data Bank CD-ROM Set — ISO 9660 Format *$332.26*

(tax and shipping charges not applicable)

AFFILIATED CENTERS

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

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

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

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

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

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

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

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

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

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

INN
Israeli National Node
Weizmann Institute of Science
Rehovot, Israel
Leon Esterman (972-8-9343934)
lsestern@weizmann.weizmann.ac.il

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

*MAG
Molecular Applications Group
Palo Alto, California
Hilary Jensen (415-473-3039)
hilary@suerte.mag.com
http://hyper.stanford.edu/~Mag/

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

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

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

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

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

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

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

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

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

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

Statement of Support

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

Contacts

Telephone 516-344-3629
Fax 516-344-5751

Internet:
help desk .... pdbhelp@bnl.gov
general correspondence ..... pdb@bnl.gov
depositions .... pdb@bnl.gov
order information .... orders@pdb.pdb.bnl.gov
network services .... sysadmin@pdb.pdb.bnl.gov
Listserver subscriptions ...... listserv@pdb.pdb.bnl.gov
Listserver postings .... pdb-l@pdb.pdb.bnl.gov
entry error reporting .... errata@pdb.pdb.bnl.gov

Staff

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

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

Internet Sites

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

First Class