Protein Data Bank

Quarterly Newsletter

Release #81 *July 1997*

July 1997 CD-ROM Release

6098 Released Atomic Coordinate Entries

Molecule Type

5397	proteins, peptides, and viruses
241	protein/nucleic acid complexes
447	nucleic acids
12	carbohydrates

Experimental Technique

161	theoretical modeling
923	NMR
5013	diffraction and other

The total size of the atomic coordinate entry database is 2544 Mbytes uncompressed.

Table of Contents

What's New at the PDB 1
Local High School Students Visit PDB 2
pdb2cif 2
Crystal Macromolecule Files Generated at the European Bioformatics Institute 3
On the Importance of Being Disordered 3
The ExPASy WWW Server - A Tool for Proteome Research 5
IMGT, the International ImMunoGeneTics Database 7
Development and Validation of a Genetic Alogorithim for Flexible Ligand Docking 8
XScape: an Integrated Environment for Macromolecular Crystallographic Computing 9
My Experience in BNL's Professional Associates Program 10
Notes of a Protein Crystallographer —The 1.8 Ångstroms Structure of Scientific Revolutions 10
Staff Changes 11
Affiliated Centers and Mirror Sites 12
Web Sites Referenced in the July 1997 PDB Newsletter 14
Related WWW Sites 15
PDB Order Form 17
PDB Staff, Scientific Consultants, Statement of Support Access to the PDB, and FTP Directory Structure for Entries 19

Internet Sites

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

What's New at the PDB

– Joel Sussman

During the last few years, there has been a significant change in the attitude of journals regarding the deposition of three-dimensional macromolecular structural information into the PDB. Presently, virtually all journals, even the most prestigious ones, list deposition of coordinates into the PDB as one of the requirements for publication. For example, a recent editorial in Nature **379**, 191 (1996) stated that "For both structures and sequences, public databases are now well established, and easily accessed for both the registration and the supply of data. From now on, Nature will hold to the principle that such information should be made available to the public as condition of publication. No paper reporting such data submitted after the date of this issue will be accepted without a reference either to a database accession number, or to an Internet address at which the data can be accessed from the time of publication. In imposing those conditions (which apply equally to publicly and privately funded researchers), Nature will wish to be assured that the game is being played fairly: that registered structures, for example, will include all atoms rather than simply a carbon backbone. But if a community widely supports a database that allows access to be restricted for a specified period, as do the structural biologists, so be it."

With the advent of easy access to the PDB *via* the World Wide Web (http://www.pdb.bnl.gov/), journal editors, as well as readers, can easily verify that the structure in a paper has in fact been deposited. If this is not the case, we at the PDB strongly urge the community to continue to follow up and contact journal editors that they should enforce their journal policy, as was recently discussed (see Giles, April 1997 PDB Quarterly Newsletter).

In parallel, we feel that journals should not only insist that coordinates be deposited but also the experimentally-measured X-ray structure factors and NMR NOEs. A short 'Letter to the Editor' by Baker, Blundell, Vijayan, Dodson, Dodson, Gilliland, and Sussman was sent to several journals in which three-dimensional structural studies of macromolecules are regularly published. This letter urged the journals to require deposition not only of atomic coordinates but also of the experimental data. The PDB feels that this is a very important development and urges members of the community to encourage journals to adhere to this policy. One way in particular in which this can be accomplished is that referees of journal articles should insist that the journal require not just deposition of coordinates but also of structure factors by the authors as a precondition for publication.

The PDB staff is making the process of submitting coordinates and structure factors as simple as possible, and, in fact, since the introduction of AutoDep, well over 70% of the entries that are submitted *via* this Web-based tool include structure factors. These structure factors are being put on the PDB server in the standard mmCIF format as described in the October 1995 PDB Newsletter and at

ftp://pdb.pdb.bnl.gov/pub/pdb/structure_factors/cifSF_dictionary. We would welcome any thoughts you may have in this regard, and would be pleased to publish these thoughts in future Newsletters.

Local High School Students Visit PDB

Nancy Manning, PDB, and Elaine Champey, Chemistry Teacher, Smithtown High School, Smithtown, NY, USA.

On April 15, 1997, sixteen students and three teachers from the Smithtown High School, Smithtown, New York, visited the Biology Department of Brookhaven National Laboratory. This trip was arranged by two of the school's science teachers, Harvey Goldstein and Elaine Champey, for students studying DNA Science and Advanced Placement Biology. It was the second year that science students from Smithtown toured BNL. The PDB welcomes the opportunity to meet with students to introduce them to the cutting edge science being carried out at Brookhaven.

The visit was nicely described by Ms. Champey, who sent the following letter to me afterward.

Dear Nancy,

I would like to take this opportunity to thank you for giving our Smithtown High School students a wonderful tour of the Protein Data Bank at the Brookhaven National Laboratory on April 15th. Once again, the program you coordinated for our DNA class was fascinating and extremely informative.

It was great to be greeted by Louise Hanson, who has been an active member of our Industry Advisory Board here at Smithtown High School. Dr. Hanson spoke to the students regarding protease inhibitors, drugs used in AIDS therapy. This information comes at an ideal time since our students have just completed an AIDS Awareness Week at the high school. This helped them to understand some of the most recent advances made in the laboratory and their significance at the clinical level.

Next, we heard from Jeanne Wysocki, who shared her research with the students. The students were very impressed with the work being done in the Genome Sequencing Lab, and her tour of the facility. Several students are looking forward to being able to participate in some way this summer.

The students were captivated by Joel Sussman, Head of the Protein Data Bank. It was a real treat for the students to be able to hear a lecture from such a renowned scientist. Dr. Sussman gave the students the history of the PDB and shared with them some of the current uses of the PDB. He also shared his research on acetylcholinesterase and related it to their lives. The students enjoyed your computer demonstration, Nancy. "Nature's Vacuum Cleaner" was an informative MPEG movie they will now be able to reference at any time.

Finally, Joe Wall, Head of BNL's Scanning Transmission Electron Microscope (STEM) Facility, and Martha Simon shared their research with the students, and were extremely informative. Frank Kito gave the students an interesting tour of the STEM facility.

Both Harvey Goldstein, our AP Biology and DNA instructor, and I were extremely impressed with the manner in which all the scientists spoke to our students, treating them as graduate students from a university, rather than as high school students.

We would like to extend our heartfelt appreciation to all the people who spent time with us during the tour. And, of course, Nancy, we want to especially thank you for the time you devoted to making our experience at Brookhaven so rewarding.

Every time you open the doors for students, you invite them to consider the wonderful opportunities that await them in the world of science and technology. This is the second year you have provided our students with that opportunity, and we thank you.

Sincerely,

Elaine Champey, Chemistry Teacher

pdb2cif

Frances Bernstein, PDB, and Herbert Bernstein, Bernstein and Sons, 5 Brewster Lane, Bellport, NY, USA (yaya@bernstein-plus-sons.com).

pdb2cif (CIF Applications. pdb2cif: Translating PDB Entries into mmCIF Format, Herbert. J. Bernstein, Frances C. Bernstein, Philip E. Bourne, accepted) is a program to translate PDB coordinate entries to the1996 mmCIF format as defined in the mmCIF dictionary 0.9.01 (version 0.8.02 presented by Paula M. Fitzgerald, Helen M. Berman, Philip E. Bourne, Brian McMahon, Keith D. Watenpaugh, John Westbrook at the 17th IUCr Congress and General Assembly, Seattle, WA, USA, 8-17 August 1996, Abstract E1226). Version 0.9.01 of the mmCIF Dictionary has been approved by the IUCr Committee for the Maintenance of the CIF Standard (COMCIFS).

The Crystallographic Information File (CIF) is an IUCr project which has been very successful in increasing productivity in the handling of small molecule files. The macromolecular CIF project (mmCIF) is an IUCr project to bring the same benefits to macromolecular structural studies and to provide room for growth in the handling of ever larger and more complex datasets. mmCIF format uses a tag-value style of presentation and has very little sensitivity to the ordering of the information. Information is presented either in tag-value pairs or in column-headed tabular form. Tags are distinguished from values by an initial underscore. Information is constrained to 80-column lines, but spacing between fields is arbitrary.

The PDB coordinate entry format is quite different, using a fixed-field presentation of information in a precisely defined order. Even more importantly, there are substantive conceptual differences in representation of macromolecules in PDB format and in mmCIF. Thus the translation of PDB entries to mmCIF is not just a matter of reformatting lines. It is necessary to understand both formats thoroughly before one can attempt to translate between them.

Anyone interested in learning more about CIF and mmCIF should consult the IUCr Web page at http://www.iucr.ac.uk and the mmCIF home page at http://ndbserver.rutgers.edu/NDB/mmCIF.

In addition, actually comparing a PDB entry to the mmCIF version of that entry is very useful in understanding the similarities and differences in representation.

pdb2cif translates from PDB coordinate entry format to mmCIF format. In the current implementation of pdb2cif, all valid PDB record types are converted, but most PDB REMARK records are carried forward as text, rather than being parsed any further. The resulting entries are substantially compliant with the mmCIF dictionary 0.9.01 which conforms to Dictionary Definition Language version 2.

pdb2cif has been installed as an output option in the PDB's 3DB Browser available from http://www.pdb.bnl.gov and as part of the SDSC MOOSE database at http://db2.sdsc.edu/moose.

Both of these browsers allow the user to search all available PDB entries. The desired entry can then be retrieved in PDB format or in mmCIF format (*via* pdb2cif).

The code of pdb2cif is freely available to anyone. The latest source code of pdb2cif is available at any of the following WWW servers:

- http://www.sdsc.edu/pb/pdb2cif/pdb2cif
- http://ndbserver.rutgers.edu/NDB/mmcif/software
- http://www.ebi.ac.uk/NDB/mmcif/software
- http://ndbserver.nibh.go.jp/NDB/software

Any questions or problems related to the program pdb2cif should be sent to Herbert J. Bernstein (yaya@bernstein-plus-sons.com).

Crystal MacroMolecule Files Generated at the European Bioinformatics Institute

Kim Henrick, EMBL Outstation, European Bioinformatics Institute, Hinxton, United Kingdom.

A crystallographic experiment on a particular MacroMolecule yields a set of coordinates that are not independent of the crystallographic symmetry (Space Group and unit cell). The deposited coordinates in a PDB entry are those atoms required to be used in refinement against the observed experimental data (*i.e.*, the structure factors) but may not necessarily completely describe the actual molecule under study.

For a PDB entry the deposited coordinates usually consist of the contents of the asymmetric unit (ASU), *i.e.*, that fraction of the crystallographic unit cell which has no crystallographic symmetry, and from which the coordinates of the whole crystal system may be generated. There are several possibilities for a relationship between this set of crystallographically unique coordinates and the actual MacroMolecule which was studied:

- The contents of the ASU are the complete MacroMolecule.
- The contents of the ASU consist of more than one copy of the MacroMolecule that was studied.
- The contents of the ASU require crystallographic symmetry operations to be applied to generate the complete MacroMolecule(s).
- A combination of the above.

An automatic procedure has been devised to recognise where multiple copies exist, and/or where symmetry is required to generate coordinate sets that describe the particular MacroMolecule studied in the X-Ray diffraction experiment. The procedure skips NMR and THEORETICAL MODEL entries. A check step is run and the results of various procedures including accessibility calculations are presented for a user to determine if the generated oligomer is a true representation of the experimentally-determined MacroMolecule.

The procedure is automatic, and may well give the wrong assembly. For example a dimer is generated when the species is monomeric. However, even for these cases the false-dimer may contain information of interest.

For PDB entries of virus samples, the deposited coordinates are normally not the contents of the asymmetric unit, but rather the coordinates for the unique non-crystallographic repeat. For these entries, including those with cylindrical polar symmetry (*e.g.*, 2tmv); helical virus entries (*e.g.*, 3ifm); and those that show icosahedral symmetry, the deposited coordinates are expanded, where appropriate, to files containing the complete virion (or a significant number of repeat units), together with files giving the coordinates for the 5-fold, 23-fold and unique set of contacting chains.

A link to the new files can br found on the summary page (as part of the Data retrieval sub-section) produced from a search using the PDB 3DB Browser, at the EBI PDB Mirror Site.

For a known PDB entry identifier, access to the files can be gained *via* the URL http://www2.ebi.ac.uk/pdb/cgi-bin/macmol.pl?filename=1abc where 1abc refers to the PDB ID code of interest.

For this set of files, the description given for a MacroMolecule is not accurate. For example an anti-body/antigen complex is described as a HETERO-TRIMERIC-COMPLEX. For many Protein/Nucleic Acid Complexes, where entries contain more than one protein chain and more than one nucleic acid chain, the protein chains may form a dimer (oligomer) in the absence of the nucleic acid chains or the Entry may consist of several protein molecules complexed to the same nucleic acid chains but lack protein-protein interaction. For some of these Entries the Check step has been deliberately removed. We are currently trying to give an acceptable naming scheme to such assemblies of bio-polymers that can also be derived automatically.

A full document on the Crystal MacroMolecule Files is available from the URL http://www2.ebi.ac.uk/pdb/macmol_doc.html.

On the Importance of Being Disordered

A. Keith Dunker, Department of Biochemistry & Biophysics, Washington State University, Pullman, WA, USA; Zoran Obradovic and Pedro Romero, Department of Electrical Engineering and Computer Science, Washington State University, Pullman, WA, USA; Charles Kissinger, and J. Ernest Villafranca, Agouron Pharmaceutical, Inc., 3565 General Atomics Court, San Diego, CA, USA.

We recently carried out predictions of disordered regions in proteins [1]. By 'disordered' we simply mean lack of coordinates in PDB files. Lack of uniformity in describing and representing such regions in the PDB was a serious problem: we hope this situation can be improved by drawing attention to, and hopefully stimulating interest in, regions of missing electron density. Following is a brief account explaining our reasons for being interested in these invisible regions.

The static view of protein-ligand interactions stresses the importance of protein functional groups being correctly positioned so that the appropriate ligand becomes bound. For such rigid molecules, the lock-and-key analogy [2] has proven to be apt. In other cases, ligand binding leads to significant conformational change, in which case the interaction is described by the term induced fit [3]. Induced fit interactions are often viewed to involve shifts of structured domains such as that observed when the hexokinases bind their substrates [4].

Many proteins just don't fit the two simple pictures described above. Proteins containing visible, ordered regions mixed with invisible, disordered regions are common. In addition, many proteins remain refractory to crystallization, despite the plethora of successful crystallizing conditions now known. Failure to crystallize can be due to the presence of a disordered region. Just in case a region of disorder is impeding crystallization, a common strategy is to carry out multiple crystallization trials on proteins with random deletions from the amino and/or carboxy termini. As a result of such strategies, a steadily increasing number of proteins in PDB are fragments, not whole proteins.

Disordered regions are more than an annoyance to structural biologists. Such unstructured domains are known to be involved in function, even molecular recognition. Involvement of an unstructured domain in a molecular recognition event appears contrary to the long-held ideas of complementary surface fitting for binding. However, such disordered regions often become ordered as the proteins associate with their cognate molecules (*e.g.*, "induced folding"), so fitted-surfaces do appear as the key to molecular recognition even for disordered regions, at least for those characterized to date.

We have identified numerous examples of disorder-to-order transitions upon complex formation so far, with the number increasing almost weekly. In some of these examples, the disorder-to-order transitions visualized by x-ray diffraction or NMR have been further investigated by protease digestion. Protease digestion provides a useful test because disordered regions can be sensitive and then become resistant upon the formation of order. Digestion studies also confirm that a given region of disorder is dynamic enough for its backbone to be engulfed by the active site of the protease.

The well-characterized disorder-to-order transitions upon binding include the following types of molecular interactions: enzyme with substrate; receptor with ligand; protein with protein; protein with RNA; and protein with DNA. Thus, disorder-to-order transitions are involved in a wide range of molecular recognition events.

Schulz [5] and Petsko and coworkers [6] independently recognized that a disorder-to-order transition upon binding would reduce the affinity of a given interaction compared to an otherwise identical interaction between two rigid structures. That is, the free energy used to fold the disordered region would take away from the overall binding free energy. Spolar and Record [7] carried out thermodynamic studies confirming that the disorder-to-order transitions inferred from the x-ray data do indeed occur in solution under physiological conditions.

In his little-noticed paper (computer searches reveal just two citations), Schulz [5] alone articulated a further refinement: disorder-to-order transitions allow the biologically advantageous combination of high specificities coupled with low affinities.

We have extended the ideas of Schulz [5] and Petsko and his collaborators [6] to show that disorder-to-order transitions upon binding lead in principle to an uncoupling of affinity and specificity, thus allowing natural selection to operate separately on these two aspects of molecular recognition. In hindsight, such an uncoupling of affinity and specificity would appear to be essential for fine-tuning the myriad of interconnected molecular recognition events that comprise the living state.

The usual argument against functional disordered regions within the cell is that they would be destroyed by proteases. Compartmentalization, even molecular-level compartmentalization by chaperone-like proteins, could obviate this problem. Chaperones and similar proteins, which provide transient protection during folding, could also provide long-term protection of disordered regions.

If disordered regions are so critical to the evolution of molecular recognition, they should have distinctive amino acid sequences. To test this, we used the absence of coordinates in PDB files and other reported x-ray structures to identify 67 disordered regions from 53 proteins having a total of 1,350 amino acids. Amino acid compositions of these disordered regions were found to be different from those of ordered regions from the same proteins; for example, the disordered regions had clearly lower contents of tryptophan, tyrosine, and cysteine, a somewhat lower content of leucine, a somewhat higher content of alanine and a clearly higher content of serine. Compared to the ordered regions, the disordered regions were also found to have larger negative hydropathies using the Kyte and Doolittle scale [8], larger flexibilities using the B factor-based indices of Vihinen and co-workers [9], and reduced values in "plot similarity" graphs [10] developed from alignments of related sequences. Evidently, amino acid composition and sequence determine disorder as well as order in native proteins.

The amino acid composition and sequence information from the 67 disordered regions was used to train and test several neural networks for the prediction of disordered regions. Success rates for these neural networks ranged from 69 to 74% for a two-state prediction: ordered or disordered [1]. Using these networks to compare the sequences of proteins in the PDB with those in sequence-only databases leads to the conclusion that proteins in nature are significantly richer in disordered regions, especially long and very long disordered regions, as compared to the proteins in the PDB.

Difficulties in crystallizing proteins with long disordered regions could account for the reduced levels of such proteins in the PDB as compared to the sequence data bases. The apparent commonness of disordered regions in the proteins in nature is entirely consistent with our proposal that such regions are critical for fine-tuning the complex networks of interacting molecules by natural selection.

There has been a tendency to ignore regions of disorder as a mere nuisance. However, all of the above suggests that disordered regions of protein merit elevation to the status of a category of protein structure, with all the implications connoted by such a status [11].

Our neural network (and especially next generation improvements of it) will have practical utility in identifying proteins that are unlikely to crystallize due to long regions of disorder. Alternatively, the prediction of disorder could lead to the specification of protein fragments that have improved probabilities for crystallization.

If there is sufficient interest, we will propose a session on disordered regions in proteins for the Pacific Symposium on Biocomputing, to be held the first week of January, 1999. Possible themes for the session would be the characterization and prediction of disordered regions in proteins.

Independently of any meeting, we would like to have discussions with crystallographers, NMR spectroscopists, their collaborators, and others about the nature of the disorder in proteins having missing electron densities or highly mobile regions. Hopefully, these discussions will encourage researchers to pay more attention to the disordered parts of proteins under study and thereby provide additional information about such regions. Does a given invisible region of an x-ray structure exhibit high mobility by NMR? Are such

regions associated with function? Do such regions exhibit protease sensitivity? Additional information of this type might lead to improved understanding and prediction of disordered regions.

If you are unsuccessfully trying to crystallize a protein and are interested in subjecting it to a disordered regions prediction, if you would be interested in attending a meeting on protein disorder, or if you just want to discuss disordered regions further, please contact us by e-mail at dunker@mail.wsu.edu or consult our Web site at http:// www.biochem.wsu.edu/disorder.

Acknowledgments: we wish to thank Greg Petsko and Georg Schulz for comments and encouragement at various stages of the work and SteveThompson and Susan Johns, of the Center for Visualization, Analysis and Design in the Molecular Sciences at Washington State University, for their helpful instructional and infrastructure support.

Comment: while preparing this note, but after our title was selected, an article appeared that was called "The importance of being unfolded" [12]; this News and Views commentary from Nature discusses the possible functional role of unfolded (disordered) protein, especially for a protein called FlgM, but uses a very different perspective from the one presented herein. As a result of this News and Views article, we applied our disordered region predictor [1] to the FlgM sequence; the resulting strong prediction of disorder for essentially the entire protein suggests that the FlgM sequence shares features in common with the regions of missing electron density in the PDB files that were used to train the neural networks.

1. P. Romer, Z. Obradovic, J.E. Villafranca, and A. K. Dunker, Identifying Disordered Regions in Proteins from Amino Acid Sequence. Proceedings of the IEEE Conference on Neural Networks,**1**, 90-95 (1997) .
2. E. Fischer, Einfluss der Configuration auf dir Wirkungder Enzyme. Ber. Dt. Chem. Ges. **27**, 2985-2993 (1894).
3. D.E. Koshland, Application of a Theory of Enzyme Specificity to Protein Synthesis. Proc. Natl. Acad. Sci. USA **44**, 98-104 (1958).
4. C. M. Anderson, F. H. Zucker, and T. A. Steitz, Space-filling Models of Kinase Clefts and Conformation Changes. Science, **204**, 375-380 (1979).
5. G.E. Schulz, Nucleotide Binding Proteins. In: Molecular Mechanism of Biological Recognition, Elsevier/North-Holland Biomedical Press, 79-94 (1977).
6. T. Alber, W.A. Gilbert, D.R. Ponzi, and G.A. Petsko, The Role of Mobility in the Substrate Binding and Catalytic Machinery of Enzymes.In: Mobility and function in Proteins and Nucleic Acids. Ciba Foundation Symposium **93**, 4-24 (1982).
7. R. S. Spolar and M. T. Record Jr., Coupling of Local Folding to Site-Specific Binding of Proteins to DNA. Science **263**, 777-784 (1994).
8. J. Kyte and R.F. Doolittle, A Simple Method for Displaying the Hydropathic Character of a Protein. J. Mol. Biol. **157**, 105-132 (1982).
9. M. Vihinen, E. Torkkila and Riikonen, Accuracy of Protein Flexibility Predictions, PROTEINS: Structure, Function, and Genetics, **19**,141-149 (1994).
10. Genetics Computer Group, Program Manual for the Wisconsin Package, Version 8, 575 Science Drive, Madison, Wisconsin, USA 53711 (1994).
11. G. Lakoff, Women, Fire and Dangerous Things: What Categories Reveal about the Mind. The University of Chicago Press, Chicago (1987).
12. K. W. Plaxco and M. Gross, The Importance of Being Unfolded, Nature **386**, 657-658 (1997) .

The ExPASy WWW Server- a Tool for Proteome Research

Amos Bairoch, Department of Medical Biochemistry, University of Geneva, Geneva, Switzerland (bairoch@medecine.unige.ch); Ron D. Appel, Laboratoire d'Imagerie Moleculaire et Bioinformatique, Division d'Informatique Medicale, Geneva University Hospital, Geneva, Switzerland (ron.appel@dim.hcuge.ch); and Manuel C. Peitsch, Geneva Biomedical Research Institute, Glaxo Wellcome R & D, Geneva, Switzerland (mcp13936@ggr.co.uk).

ExPASy is a World Wide Web server (http://www.expasy.ch) providing access to a variety of databases and analytical tools in the field of protein science known as proteome research. It was developed jointly by a team which includes members from the Geneva University Hospital, the University of Geneva and the Geneva Biomedical Research Institute (GBRI). It first started to operate in August 1993 (and seems to have been the first WWW server to be established in the field of life sciences) and has been running without interruption since that date. As of early May 1997, it had been accessed 13 million times by a total of 374,000 computer hosts from 110 countries. We describe here the information resources and tools that are available on ExPASy.

—Databases

ExPASy is the main host for the following databases that are partially or completely developed in Geneva:

- SWISS-PROT [1], a curated protein sequence database which strives to provide a high level of annotation such as the description of the function of a protein, its domain structure, post-translational modifications, and variants; a minimal level of redundancy; and a high level of integration with other databases.
- SWISS-2DPAGE [2], a database of proteins identified on two-dimensional polyacrylamide gel electrophoresis (2-D PAGE). SWISS-2DPAGE contains data from a variety of human biological samples as well as from Escherichia coli and Saccharomyces cerevisiae.
- PROSITE [3], a database of biologically significant sites, patterns and profiles that helps to reliably identify to which known protein family a new sequence belongs.
- SWISS-3DIMAGE [4], a database of high quality annotated images of biological macromolecules with known three-dimensional structure.
- ENZYME [5], a repository of information related to the nomenclature of enzymes.

- CD40L base [6], a collection of clinical and molecular data on the CD40 ligand defects leading to Hyper-IgM syndrome.
- SeqAnalRef [7], a bibliographic reference data bank relative to papers dealing with sequence analysis.

Each of the above databases have their own home pages which can be directly reached from the home page of ExPASy, from which a variety of access options are available. These options allow the user to display and retrieve specified subsets of the database. For example, from the home page of SWISS-PROT there are options that allow searching by description, accession number, author, citation or a full text search. To complement these options we will very soon implement an SRS [8] version 5 server that will allow complex searches to bemade on any fields of the combination of SWISS-PROT and TrEMBL [1] databases. A huge variety of documents are available with SWISS-PROT; these documents are all browsable from ExPASy and are enhanced by a variety of hyperlinks.

All the databases available on ExPASy are extensively cross-referenced to other molecular biology databases or resources all over the world. For example, SWISS-PROT is cross-referenced to 30 different databases (EMBL/Genbank, PDB, Medline, MIM, FlyBase, MGD, SGD, SubtiList, MaizeDB, proDom, etc.).

—Tools

We have developed over the last three years an extensive collection of software tools for the analysis of protein sequences. These tools can all be accessed from ExPASy:

- Swiss-Model [9], an automated knowledge-based protein modelling server. Swiss-Model is able to build three-dimensional models of proteins whose sequences are closely related to those of proteins of known structure.
- AACompIdent [10], a tool for the identification in SWISS-PROT of a protein by its amino acid composition.
- AACompSim compares the amino acid composition of a SWISS-PROT entry with all other entries in the database.
- TagIdent and MultiIdent [11], these two related programs identify proteins using a variety of experimental information such as the pI, the molecular weight, the amino acid composition, partial sequence tags, and peptide mass fingerprinting data.
- PeptideMass calculates the theoretical masses of peptides generated by the chemical or enzymatic cleavage of proteins so as to assist in the interpretation of peptide mass fingerprinting.
- ScanProsite scans a sequence against all of the patterns in PROSITE or a pattern against all sequences in SWISS-PROT.
- ProtParam calculates physico-chemical parameters of a protein sequence such as the composition, the pI, the extinction coefficient, etc.
- ProtScale computes and represents the profile produced by any amino acid scale on a selected protein. Some 50 predefined scales are available, the default being the Doolittle and Kyte hydrophobicity scale.
- Translate, a tool to translate a nucleotide sequence to an amino acid sequence.
- RandSeq, a random protein sequence generator.
- Swiss-Shop, a sequence alerting system for SWISS-PROT that allows you to automatically obtain new sequence entries relevant to your field(s) of interest.

Some of the above tools (Swiss-Model, Swiss-Shop, AACompSim) report their results back by e-mail while the others display them directly on-line. It must also be noted that an important feature of tools such as TagIdent, MultIdent and Peptide Mass is that they use the annotations of the SWISS-PROT entries to take into account post-translational modifications as well as sequence variants to perform their predictions.

These tools are all listed on a page of ExPASy that also offers links to many other useful programs for the analysis of protein sequences available elsewhere on the Web. We notably have links to the tools provided by our colleagues from the bioInformatics group at ISREC in Lausanne (http://ulrec3.unil.ch). They have developed a WU-BLAST similarity search server, ProfileScan (to scan a sequence against the profiles in PROSITE), TMpred (to predict transmembrane regions) and interfaces to the SAPS (Statistical Analysis of Protein Sequences) and COILS (prediction of coiled coil regions) programs.

—Other interesting ExPASy features

An indexed, digitized, and clickable version of the Boehringer Mannheim 'Biochemical Pathways' poster is available on the server. It allows the user to navigate through the graphical representation of metabolic pathways and is directly linked to the ENZYME database.

A wide variety of information concerning 2-D PAGE is available from ExPASy. This includes the full description of experimental protocols; a tutorial for the Melanie II 2-D PAGE analysis software package; and WORLD-2DPAGE, an index of all known 2-D PAGE databases, WWW servers, and related services.

The ExPASy server is the home of SWISS-PdbViewer [12], an application that runs on Mac and under Windows-95 that offers a wide range of options to visualize and manipulate protein structures. It can be also used as a WWW helper application for the display of PDB-formatted entries. SWISS-PdbViewer can be downloaded directly from ExPASy or from PDB.

One must not forget that science can also have a lighter side, so we hope that users will take the time to take a small pause from the hectic pace of modern research and visit 'Swiss-Quiz' where they can have a chance to win some Swiss chocolate (real, not virtual!) when they successfully answer a quiz in the field of molecular biology or 'Swiss-Jokes', where they have access to a collection of jokes and aphorisms from the fields of life and computer sciences.

The mass of information available to life scientists on the Web has completely changed the way that biological data is accessed and processed. It has created many opportunities but also brought new dangers. The most critical problem is the inability of researchers to distinguish useful and up-to-date sources of information from sites that provide either 'fossilized' or low quality data. To partially address this problem, we have created and are maintaining a list of molecular biology resources on the Web. This list is known as "Amos' WWW links page".

—Conclusion

The team that develops ExPASy is committed to bringing to their users top quality information services in the field of protein sciences. We hope that in the coming years we will be able to add many new features to those that already exist. One of the ongoing developments is to sponsor the establishment of mirror versions of ExPASy in different parts of the world.

—Acknowledgment

We thank the government of the State of Geneva for their support of the ExPASy server through a special donation that allowed us to acquire a new machine in the beginning of 1997.

References

1. A. Bairoch, R. Apweiler, Nucleic Acids Res. **25**, 31-36 (1997).
2. J.-C. Sanchez, R.D. Appel, O. Golaz, C. Pasquali, F. Ravier, A. Bairoch, D.F. Hochstrasser, Electrophoresis **16**, 1131-1151 (1995).
3. A. Bairoch, P. Bucher, K. Hofmann, Nucleic Acids Res. **25**, 217-221 (1997).
4. M.C. Peitsch, T.N.C. Wells, D.R. Stampf, J.L. Sussman, Trends Biochem. Sci. **20**, 82-83 (1995).
5. A. Bairoch, Nucleic Acids Res. **24**, 221-222 (1996).
6. L.D. Notarangelo, M.C. Peitsch, Immunol. Today **17**, 511-516 (1996).
7. A. Bairoch, Comput. Appl. Biosci. **7**, 268-268 (1991).
8. T. Etzold, A.V. Ulyanov, P. Argos, Meth. Enzymol. **266**, 114-128 (1996).
9. M.C. Peitsch, Biotechnology **13**, 658-660 (1995).
10. M.R. Wilkins, E. Gasteiger, J.-C. Sanchez, R.D. Appel, D.F. Hochstrasser, Current Biol. **6**, 1543-1544 (1996).
11. M.R. Wilkins , C. Pasquali, R.D. Appel, K. Ou, O. Golaz, J.C. Sanchez, J.X. Yan, A.A.Gooley, G.Hughes, I. Humphrey-Smith, K.I.Williams, D.F. Hochstrasser, Bio/Technology **14**, 61-65 (1996).
12. N. Guex and M. Peitch, PDB Quarterly Newsletter **77**, 7 (July 1996).

IMGT - the International ImMunoGeneTics Database

Marie-Paule Lefranc, LIGM (Laboratoire d'ImmunoGénétique Moleculaire) CNRS, Montpellier II University, Montpellier, France (lefranc@ligm.crbm.cnrs-mop.fr).

The molecular synthesis of the Immunoglobulin (Ig) and T cell Receptor (TcR) chains is particularly complex and unique since it includes biological mechanisms such as DNA molecular rearrangements in seven loci (three for Ig and four for TcR) located on four different chromosomes in the human, nucleotide deletions and insertions at the rearrangement junctions, and hypermutations in the Ig loci. The number of potential protein forms of Ig and TcR is almost unlimited. Owing to the complexity and high number of these sequences, the specialized database IMGT was created in 1992 by Marie-Paule Lefranc at Montpellier, France (http://imgt.cnusc.fr:8104).

IMGT, the international ImMunoGeneTics database, is an integrated database specialising in Immunoglobulins, T cell Receptors and Major Histocompatibility Complex (MHC) molecules of all vertebrate species. IMGT includes two databases: LIGM-DB (for Ig and TcR) and MHC/HLA-DB. IMGT comprises expertly-annotated sequences and alignment tables. LIGM-DB contains more than 23,000 Immunoglobulin and T cell Receptor sequences from 92 species. MHC/HLA-DB contains class I and class II Human Leucocyte Antigen alignment tables. IMGT works in close collaboration with the EMBL database and, since August 1996, the IMGT/LIGM-DB content follows closely the Ig and TcR EMBL one.

All Ig and TcR sequences are assigned with IMGT keywords. The general keywords, indispensable for the sequence assignments, are described in an exhaustive and non-redundant list, and are organized in a tree structure, whereas specific keywords are specifically associated to particularities of the sequences (orphon, pseudogene...) or to diseases (leukemia, lymphoma, tumor...). The whole list of keywords can be reached at http://imgt.cnusc.fr:8104/textes/LECT/kw.html.

The description of the Ig and TcR sequences, at the DNA and protein level, relies on an extensive list of labels for the structural and functional motifs established by LIGM. More than one hundred sixty labels were shown to be necessary for an accurate annotation. The list of labels with their corresponding definition and main schemas are available at http://imgt.cnusc.fr:8104/textes/LECT/labeldef.html.

A program, DNAPLOT, available at http://www.genetik.uni-koeln.de/dnaplot/ and from IMGT, has been developed by H. H. Althaus and W. Müller from IFG (Köln,Germany), to generate and display sequence alignments, and to propose assignment of rearranged or expressed variable Ig or TcR genes to the potential germline genes. The objective of IMGT is also to provide immunologists and geneticists with a unique nomenclature per locus which will allow extraction and comparison of data for the complex B and T cell antigen receptor molecules, whatever the species. In a first step, data concerning the human Ig and TcR genes have been standardized and maps of loci and tables with IMGT nomenclature, correspondence to other gene designations and gene functionality are available from the IMGT home page. Even more importantly, a unique numbering for Ig and TcR V-region of all species has recently been set up which will greatly facilitate the comparison of sequences for polymorphism, mutation, evolution and structural analysis.

IMGT is maintained with SYBASE as the relational DBMS at CNUSC, Montpellier, France. Biologists' needs were taken into account for the development of the interface WWW-SYBASE, and for the IMGT on-line consultation which allows users to create very specific and structured queries combining aspects of relational data-bases and hypertext. Requests can be performed through distinct modules that allow one to classify search criteria type. At the end of a run, a number of resulting sequences is proposed and it is then possible either to look at the solutions, or to add new conditions to modify the results, keeping in memory the previously selected criteria. There are several ways to retrieve the results; in particular it is possible to extract specific coding regions from the query-resulting sequences.

Flat files are produced in collaboration with EMBL-EBI and are available on the EBI anonymous ftp server and are also distributed with many other databases on the EMBL CD-ROM. Names of entries remain the EMBL accession number. IMGT/LIGM-DB flat file typical entries provide LIGM expertise: standardized LIGM keywords appear in KW code lines, complement to definition in DE code lines and sequence description with LIGM labels in FT code lines. Core data, as well as cross-references to other databases in DR code lines are kept from EMBL. Flat file format allows IMGT/LIGM-DB data to be compatible with the most efficient software for information

retrieval or data manipulation such as the browser SRS. SRS and EMBL-EBI anonymous ftp server can be reached from http://imgt.cnusc.fr:8104/textes/services.html.

The information provided by IMGT is of much value to clinicians and biological scientists in general and has important implications in medical research (repertoire in autoimmune diseases, AIDS, leukemias, lymphomas), therapeutical approaches (antibody engineering), genome diversity and genome evolution studies. IMGT is designed to allow common access to all immunogenetics data. This approach is based on a very tight collaboration with EMBL for the nucleotide sequence data, with SWISS-PROT for the protein sequence data, and with IGD for providing a user friendly interface for the mapping and genetic data. Particular attention will be given to the establishment of cross-referencing links to other databases pertinent to the users of IMGT.

IMGT is developed by LIGM (Montpellier, France) in collaboration with CNUSC (Montpellier, France), EMBL-EBI (Hinxton, UK), ICRF (London, UK), IFG (Köln,Germany), BPRC (Rijswijk, The Netherlands) and EUROGENTEC S.A. (Seraing, Belgium), Information is available at http://imgt.cnusc.fr:8104/textes/info.html.

IMGT is funded by the European Union's BIOMED1 and BIOTECH programmes, the CNRS (Centre National de la Recherche Scientifique), and the MENESR (Ministére de l'Education Nationale, de l'Enseignement Supérieur et de la Recherche). Subventions have been received from Association pour la Recherche sur le Cancer, Association de Recherche sur la Polyarthrite, Fondation pour la Recherche Médicale, Groupement de Recherche et d'Etude sur les Génomes and the Région Languedoc-Roussillon.

V.Giudicelli, D. Chaume, J.Bodmer, W.Müller, C.Busin, S. Marsh, R. Bontrop, M. Lemaitre, A. Malik, and M.-P. Lefranc, Nucleic Acids Research, **25**, 206-211 (1997).

Development and Validation of a Genetic Algorithm for Flexible Ligand Docking

Gareth Jones, Peter Willett, Department of Information Studies and Krebs Institute for Biomolecular Research, University of Sheffield, Western Bank, Sheffield S10 2TN, UK(http://panizzi.shef.ac.uk/gareth; gareth.jones@sheffield.ac.uk); Andrew R. Leach, Glaxo-Wellcome Medicines Research Centre, Stevenage, UK; Robin Taylor, Cambridge Crystallographic DataCentre, Cambridge, UK.

Prediction of small molecule binding modes to macromolecules of known three-dimensional structure is a problem of paramount importance in rational drug design (the "docking" problem). This problem in molecular recognition is extremely demanding. Not only are powerful search engines, capable of solving multiple minima problems, required, but an appreciation of the process of molecular recognition is required to generate suitable target functions.

Genetic Algorithms (GAs) are novel optimisation algorithms which emulate the process of Darwinian evolution in order to solve complex search problems. Guided by the mechanics of evolution, successive generations of populations of artificial creatures called chromosomes search the fitness landscape of a problem to determine optimum solutions. These algorithms have proved capable of yielding approximately optimal solutions given complex, multimodal, non-differential and discontinuous search spaces. The application of the GA to problems in computational chemistry and biology is the subject of much investigation.

A GA for docking conformationally flexible ligands into partially flexible protein sites called GOLD (Genetic Optimisation for Ligand Docking) has been developed. To reduce the search space and couple the GA fitness function to the chromosome encoding, hydrogen bonds are explicitly encoded in the fitness function. In order to predict binding modes successfully, an attempt has been made to quantify the ability of common substructures to displace water from the receptor surface and form hydrogen bonds.

Validation of such algorithms is typically achieved using the PDB. The PDB contains a growing number of small-molecule protein complexes. The ligand is extracted from the complex and the docking algorithm used to predict the binding mode of the ligand. The predicted binding mode can then be compared with the crystallographically observed binding mode. As GAs are non-deterministic in nature, a number of docking runs are required to elucidate a binding mode. Typically, GOLD is run twenty times, though a solution is normally found in under ten runs. The twenty solutions are ranked using GOLD's scoring function and the predicted binding mode is then the solution with the highest GOLD score.

PDB entry 3DFR can be used to illustrate the effectiveness of GOLD. In this structure NADPH and methotrexate are complexed to DHFR. Both NADPH and methotrexate were extracted from DHFR and GOLD was used to dock NADPH into DHFR. Although this ligand is large and highly flexible the GOLD prediction is within 1.1 Ångstroms (root mean square deviation of heavy atoms) of the experimental result.

In order to validate GOLD comprehensively, the algorithm has been tested on a data set of 100 complexes extracted from the PDB. These complexes were selected by the CCDC on the basis of ligand diversity and of pharmacological interest. When used to dock the ligand back into the receptor, the algorithm achieved a 71% success rate in identifying the experimental binding mode. The results produced by the program and the coordinates of the test-systems used are now freely available on the internet (http://panizzi.shef.ac.uk/gareth/gold/gold.html). Provided the user has the chime molecular viewer from MDL(http://www.mdli.com/chemscape/chime), 3D models of GOLD predictions overlaid on experimental binding modes may be observed. In addition to the original 100 complexes, results for a further 34 complexes are available.

We have found the PDB essential to the validation of GOLD and hope that the extensive testing indicates the effectiveness of our approach. However, the PDB is heavily biased towards peptide and peptide-like ligands and more drug-like ligands are required for further development of these programs. Additionally, there is a shortage of high quality structures with precise ligand co-ordinates. We believe that the design of this algorithm provides an insight into the mechanism of molecular recognition.

1. D.E. Goldberg, Genetic Algorithms in Search Optimisation and Machine Learning, Addison-Wesley, Reading MA, (1989).

2. L. Davis (Ed.) Handbook of Genetic Algorithms, Van Nostrand Reinhold, New York, (1991).
3. G. Jones, P. Willett, and R.C. Glen, J. Mol. Biol. **245**, 43-53 (1995).
4. G. Jones, P. Willett, R.C. Glen, A.R.L. Leach, R.J.Taylor, Mol. Biol. **267**, 727-748 (1997).

XScape: an Integrated Environment for Macromolecular Crystallographic Computing

David Wild and S. Choe, Structural Biology Laboratory, The Salk Institute, La Jolla, CA, USA (wild@sbl.salk.edu, choe@sbl.salk.edu).

Despite the advances in macromolecular crystallography over the last two decades, the user interface to macromolecular crystallographic software has been neglected until fairly recently. Much of the crystallographic software used for macromolecular structure determination still retains the "card image" input format common in the early 1970's when it was first written. More recently, a number of projects have been undertaken either to develop completely new packages based on the X-window system or to develop X-window based interfaces to some existing crystallographic software. There remains, however, a need for user-friendly integrated software packages for macromolecular crystallography which cover all aspects of structure solution, from data reduction to model building and refinement. Although tools to perform all these steps do exist, they are distributed between a number of software packages, each with its own data file formats, command conventions and philosophy.

This article describes our approach to developing "XScape", an integrated environment for crystallographic computing. It does not necessitate the development of new packages, but aims to provide graphical user interfaces and data visualization tools for a number of existing and widely used macromolecular crystallographic software packages, and combines these behind a common graphical front-end. Initially selected programs from the Collaborative Computing Project 4 (CCP4) suite have been interfaced, using the commercial software development environment AVS/Express. Our prototype has focussed on interfaces to the CCP4 programs most commonly used in the solution of crystal structures by multiple isomorphous and molecular replacement methods.

An important feature of the CCP4 suite is that individual programs may be combined in different ways to accomplish particular tasks. Input for each program is provided by a series of control cards comprising a leading keyword plus additional keyword/value pairs or numerical parameters. Since many crystallographic calculations involve a series of steps, using different programs, these are often chained together in a script or command file. We have emulated this approach by using the 'data flow network' architecture of the AVS/Express environment.

Data flow visualization environments, such as AVS, allow modular software components to be connected together into executable data flow networks or directed acyclic graphs to construct a data visualization application. Data flow networks are built interactively using a visual programming interface, by choosing icons representing individual modules from a palette and connecting them together in a workspace.

Graphical user interfaces for individual modules or complete networks of modules may be constructed with the help of built-in user interface builders. In addition to uses in molecular visualization, the dataflow model also reflects the nature of crystallographic computing using a loosely organized package such as CCP4. Each module represents a single CCP4 program, and networks of programs may be constructed to perform particular tasks. In XScape, both these 'task' networks or individual programs may be executed from a standard Motif pull down menu bar.

The function of the GUI is not only to simplify program input, but also to present program output to the user. The program output, read by an encapsulated CCP4 'Xloggraph' module, may be displayed directly in a text browser widget on-screen. A future development is planned that will filter this output to extract summary information, which can then be presented in text widgets in the GUI. The output from a number of CCP4 programs already has tabular statistical information marked with keywords to allow extraction and display as graphs by the Xloggraph utility. Graphical information from these tables is passed directly to the AVS/Express graph viewing tools and displayed in the integrated 2D/3D viewer, providing visual monitoring of the progress of the calculations.

Another advantage of integrating crystallographic programs into the AVS/Express environment is that a number of prepackaged graphical tools are available for the visualization of 2D and 3D data types, including an integrated 2D/3D viewer. The results of crystallographic calculations, such as electron density and Patterson maps, may be visualized with these graphical tools; for example, as 3D solid or wire frame isosurfaces or as 2D contoured sections, and combined with renderings of the molecular structure. A number of modules are available which facilitate the interactive display of any arbitrary contoured section of an electron density or Patterson map. The section number, axis and contour levels may all be varied interactively with sliders and the output displayed in the integrated viewer window either as 2D contoured sections or as part of a 3D object which can be rotated and translated with the mouse. Hard copy Postscript output may also be obtained. The 3D visualization tools may also be used to display rotation and translation function maps.

In addition to providing interfaces to the most commonly used CCP4 programs, we plan to integrate other Fortran crystallographic programs which expect "card image" input data and produce "line printer" output into the same environment, thus providing alternative methods for the same task. Since the proposed software environment can readily incorporate different programs with minimal reconstruction, its major advantage over other "stand-alone" crystallographic and biocomputing packages lies in its flexibility with respect to future expansion.

Finally, the AVS/Express Developer's Edition permits a compiled application to be built which can be distributed as a stand-alone system without the need for AVS/Express to be installed on the user's machine. Potential users will only need to obtain a low-cost run time license. Please contact the authors for further details.

An expanded version of this article, including figures, may be found at http://sbl.salk.edu/~wild/IUCR/IUCR.html.

My Experience in BNL's Professional Associates Program

Brigitte Sylvain, PDB

Brookhaven National Laboratory's Professional Associates Pre-employment Program is a career-oriented program for graduates who have earned a Bachelor's degree in a scientific field. The program is designed to enhance the diverseness of the Laboratory's staff by increasing the number of under-represented minorities (Black, Hispanic, and Native American), people with disabilities, and women in various professional fields. The one-year program provides the Associates with an opportunity to gain experience and acquire skills which will make them more competitive in the general labor market and when competing for placement at Brookhaven.

The minimum requirements for participation are a Bachelors degree with good academic standing, plans to attend graduate or professional school, and interest in long-term employment at Brookhaven. With the guidance of the Professional Associates Program coordinator, Frances V. Ligon, colleagues of the Associate, and supervisors experienced in the Associates' particular area of interest who serve as mentors, Professional Associates benefit from a strong support network. The program is for one year and evaluations are given throughout.

Each Professional Associate receives an individually-structured program that enhances their background and exposes them to various issues pertaining to their chosen field. Professional Associates are encouraged to do research on their own and to acquire as many skills as possible in order to be a successful candidate in the labor market.

As an Associate with the Protein Data Bank, I am able to finally connect the ideologies of my degree to reality. After completing my Bachelors degree in Biology, the biological concepts I learned are now challenged daily at the PDB. The PDB challenges me to develop deductive reasoning skills, efficient analytical problem solving strategies, public speaking skills and interpersonal skills. As a member of the Archive Management Group, I validate and normalize newly-submitted structural data. I strive to resolve issues such as those pertaining to sequence, stereochemistry, close contacts, secondary structure, and het groups, and I communicate with depositors regarding their entry and issues that need to be addressed. The atmosphere at the PDB is exciting, especially since the PDB is at the cutting edge of science. Since my arrival I have prepared and given two presentations to staff members of the PDB, and I continue to learn about structural biology and bioinformatics.

The Professional Associate experience at BNL is unique for each participant. For example, Onarae V. Rice, a Professional Associate this year in the Medical Department, focuses on body composition with debilitating diseases. Mike Marshall, a former Professional Associate in the Computing and Computer Division (CCD), is now a full-time employee in the CCD with the Computer Security group and the Visualization/Multimedia team.

Learning the tools for a successful career is vital, and the Professional Associate's Program introduces the concept that the true challenge in any career arises when there is the demand to apply your knowledge to the unknown.

For more information about the program, contact:

Frances V. Ligon, Program Coordinator
Brookhaven National Laboratory,
Bldg 185A Diversity Office Upton, NY 11973 USA
e-mail: ligon1@bnl.gov

Notes of a Protein Crystallographer

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

—The 1.8 Ångstrom Structure of Scientific Revolutions

As the reader might have already inferred, I have a philosophical slant in my interests and perceptions. Therefore, it will not come as a surprise that in my graduate student days at the University of Texas at Austin (under the mentorship of Profs. M. L. Hackert and J. L. Fox), I came across a copy of the now classic book "The Structure of Scientific Revolutions" by T. S. Kuhn [1]. With my still crude English-reading skills, I read it from cover to cover and realized immediately that it was a milestone-not only in understanding the way science was performed ona day-to-day basis, but also in putting in the right perspective the accumulated knowledge resulting from the labors of scientists and researchers in different areas. I was a student of protein crystallography at the time and, as a joke, I wrote on the blackboard the title of the book with '1.8Å' inserted just before the word 'structure'.

Yet, the content of the book laid dormant for many years in some arcane crevices of my brain. Over the past few years, I have come to reflect again over its contents in relation to the so-called revolution in the biological sciences originated by the determination of the three-dimensional structures of DNA and proteins at atomic resolution, the discovering of the Genetic Code and the advances in understanding the machinery of the cell at the molecular level [2]. The common use of DNA recombinant technology and the availability of a multitude of protein structures as determined by X-ray diffraction and NMR methods can be taken as tangible proof of such a revolution.

As practitioners of the trade that is responsible for the majority of those atomic masterpieces, we have been very successful in producing structures of ever increasing complexity and size in amazingly short periods of time. There is an informal record circulating in the community of three months from an isolated target gene to a complete three-dimensional structure determination of the associated protein product, well refined and with 'excellent geometry'.

The future looks even brighter with the availability of third generation synchrotron sources (ESRF, APS, and Spring 8), CCD detectors and almost unlimited computing power, combined with superb software tools that our community has developed over the years. Soon, we will be able to slice biochemical and enzymatic events at the microsecond or even nanocond time scale[3].

Our results and methods are being used to expedite and rationalize the design and development of new drugs and diagnostic

tests for the patient community at large, and enzymes with novel properties are being engineered by random and directed mutations of the three-dimensional structure of the naturally occurring variants.

And yet, as conscientious professionals we also realize that our tools and methods have limitations such as [4]:

- Many interesting target proteins do not seem to crystallize in our laboratories or many never be crystallized.
- The quality and extent of the diffraction patterns from our crystals never seem to be adequate to get answers to many of the interesting biochemical or biological questions.
- Our electron density maps, and consequently our refined protein structures, have disordered regions that cannot be determined unambiguously.
- Our refined models have only limited information regarding the mobility and dynamics of the macromolecular structures we work on.

In the domains of structure, flexibility, and dynamics of macromolecules in solution, the NMR community has also made tremendous strides in adding, expanding, and complementing the knowledge derived from x-ray diffraction. They also face their intrinsic limitations and challenges and what their ingenuity and efforts will yield in the future can only be a matter of conjecture.

Through our education, we have all been exposed to the limitations imposed in our knowledge of the atomic world by the Heisenberg uncertainty principle governing the simultaneous determination of the position and momentum of an electron in motion. Of course, you will argue that our results do not correspond to this domain of knowledge but some work on the migration of carbon monoxide in myoglobin [3] may point in that direction. Could it be that the latter three of the above limitations will impose an analogous uncertainty principle for our atomic description of the biochemical phenomena responsible for some of life processes?

I do not mean to imply the existence of an *'élan vital'* but only that our methods and probes into the atomic structure of the biological matter have limitations.

Our efforts, individually and as a community, have undoubtedly resulted in an on-going scientific revolution. The instruments of this revolution are the tools (hardware and software) that our community has developed to examine the atomic structure of biological macromolecules, aided by the dedicated efforts of molecular biologists and protein biochemists, supporting our work and complemented by the tools that the NMR community has developed on its own.

The icons of this revolution are the myriad of well refined three-dimensional structures of macromolecules of paramount biological importance displayed in various scientific journals, many of them well beyond the ones dedicated to crystallographic work. These icons are deposited in the Protein Data Bank and have already made their appearance in the current textbooks of Biology, Biochemistry and Medicine. They are already part of what T. S. Kuhn would call 'normal science', distant already from the 'revolutionary science' that we pioneered only a few years ago.

Therefore, we must admit that, like any other revolution, ours is also limited in extent and scope. Whether this limitation is inherent to nature or only restricted by the tools we use, it is still a matter for the future to decide.

1. T.S. Kuhn, The Structure of Scientific Revolutions, 2nd enlarged Ed., University of Chicago Press (1970).
2. H.F. Judson, The Eighth Day of Creation, Jonathan Cape, Ltd. London (1979).
3. V. Srajer, T.Y. Teng, D. Bourgeois, W. Wulff, and K. Moffat, Photolysis of the Carbon Monoxide Complex of Myoglobin; Nanosecond Time-resolved Crystallography, Science **274,** 1726-1729 (1996).
4. G.J. Kleywegt and T. Alwyn Jones, Where Freedom is Given, Liberties are Taken, Structure **3**, 535-540 (1995).

Staff Changes

The PDB is pleased to welcome Dr. S. Swaminathan as a senior member of the Archive Management Group. He holds a joint scientific appointment in the Biology Department of Brookhaven National Laboratory where he will continue his research into the structure of toxins and superantigens.

Swami's experience in protein crystallography will enable him to spearhead our work in structure validation. He comes to us following 17 years at the Crystallography Department of the University of Pittsburgh and the Biocrystallography Lab at the VA Medical Center, Pittsburgh, PA and is well-known to our user community.

The PDB wishes a fond farewell to Pam Esposito, who has been with us since 1988. She has published as well as managed the distribution of the PDB Quarterly Newsletter, coordinated the PDB Service Agreements, managed the distribution of the CD-ROMs and tapes containing the PDB, and provided administrative support to the PDB. She leaves us for a position with the new BNL-RIKEN Center, an international collaboration doing spin-polarized proton physics research at BNL's Relativistic Heavy Ion Collider (RHIC).

We'd like to thank Pam for her contribution to the PDB over the years. We wish her much success and happiness in her new position.

We also wish to welcome two students who are doing summer internships with the PDB. Mariya Kobiashvili is a sophomore at the State University of New York at Stony Brook. She is a Women in Science and Engineering (WISE) student, recipient of an NSF Scholarship, and plans to major in biochemistry. Mariya will be handling the PDB Help Desk this summer as well as preparing some new PDB documentation. She is also working on a project involving the upgrading of sequence information in older PDB entries.

Sabrina Hargrove comes to the PDB through the Suffolk County Community College Women's Internship Program. She is a sophomore majoring in computer technology, with plans to continue her education in computers at SUNY Stony Brook after graduation. Sabrina is assisting the Archive Management Group with administrative and clerical responsibilities. ■

AFFILIATED CENTERS AND MIRROR SITES

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

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

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

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@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)

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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.icb.ufmg.br

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)

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

WARSAW UNIVERSITY
Interdisciplinary Centre for Modelling
Warszawa, Poland
Wojtek Sylwestrzak (48-22-658-4901)
W.Sylwestrzak@icm.edu.pl
http://www.icm.edu.pl

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

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

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

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

WEB SITES REFERENCED IN THE JULY 1997 PDB NEWSLETTER

An Avs/Express interface to CCP4	http://sbl.salk.edu/~wild/IUCR/IUCR.html
Chemscape Chime Molecular Viewer from MDL	http://www.mdli.com/chemscape/chime
Complete MacroMolecule(s) for a PDB entry	http://www2.ebi.ac.uk/pdb/cgi-bin/macmol.pl?filename=labc
Crystal MacroMolecule Files - Description	http://www2.eb.ac.uk/pdb/macmol_doc.html
Disorder Web Site, Department of Biochemistry & Biophysics, WSU	http://www.biochem.wsu.edu/disorder
DNAPLOT Search Page	http://www.genetik.uni-koeln.de/dnaplot/
ExPASy Molecular Biology Server	http://www.expasy.ch
GOLD: Genetic Optimisation for Ligand Docking	http://panizzi.shef.ac.uk/gareth/gold/gold.html
IMGT - International ImMunoGeneTics Database	http://imgt.cnusc.fr:8104
ISREC Bioinformatics Group	http://ulrec3.unil.ch
IUCr - International Union of Crystallography	http://www.iucr.ac.uk
mmCIF Home Page	http://ndbserver.rutgers.edu/NDB/mmCIF
mmCIF Software Tools	http://ndbserver.rutgers.edu/NDB/mmcif/software
mirrored at EBI, UK	http://www.ebi.ac.uk/NDB/mmcif/software
mirrored at NIBH, Japan	http://ndbserver.nibh.go.jp/NDB/software
Moose - Macromolecular Structure Database	http://db2.sdsc.edu/moose/
PDB Home Page	http://www.pdb.bnl.gov
pdb2cif Program	http://www.sdsc.edu/pd/pdb2cif/pdb2cif
Structure Factor CIF Dictionary	ftp://pdb.pdb.bnl.gov/pub/pdb/structure_factors/cifSF_dictionary

RELATED WWW SITES

Databases

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

Software-Related Sites

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

Other Resources

Crystallography Worldwide	www.unige.ch/crystal/w3vlc/crystal.index.html
BioMoo	www.cco.caltech.edu/~mercer/htmls/BioMOOHomePage.html
DALI - Comparison of Protein Structures in 3D	www.embl-heidelberg.de/dali/dali.html
MOOSE (Macromolecular Structure Database at San Diego Supercomputer Center)	db2.sdsc.edu/moose
PDB_select: Representative PDB Structures	ftp://ftp.embl-heidelberg.de/pub/databases/protein_extras/pdb_select/recent.pdb_select
PROCHECK - To Submit a PDB File for Analysis	www.cryst.bbk.ac.uk/PPS/procheck/test.html
Protein Structure Verification-Biotech Server	biotech.embl-heidelberg.de:8400/
Mirrored at PDB	biotech.pdb.bnl.gov:8400/
Resources for Macromolecular Structure Information	www.ucmb.ulb.ac.be/StructResources.html
The 1996 Principles of Protein Structure Course at Birkbeck College	www.cryst.bbk.ac.uk/PPS2/index.html
Mirrored at PDB	www.pdb.bnl.gov/PPS2/index.html
Mirrored at Daresbury	www.dl.ac.uk/PPS/index.html
The Virtual School of Molecular Sciences	www.vsms.nottingham.ac.uk/vsms/
Weizmann Institute, Genome and Bioinformatics	bioinfo.weizmann.ac.il/

This page intentionally left blank.

BROOKHAVEN ORDER FORM

Name of User _____ Date _____

Organization _____ Phone _____

Address _____ Fax _____

_____ E-mail _____

_____

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

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

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

→ Check payable to Brookhaven National Laboratory in 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:

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

1 Protein Data Bank CD-ROM — ISO 9660 Format *$357.25*
(tax and shipping charges not applicable)

This page intentionally left blank

SCIENTIFIC CONSULTANTS

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

Sasha Faibusovich, Clifford Felder, Kurt Giles, Jaime Prilusky,
Mia Raves, Vladimir Sobolev, and 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
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
John Spiletic
S. Swaminathan
Brigitte R. Sylvain
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.

ACCESS TO THE PUB

Main Telephone	1 516-344-3629
Help Desk Telephone	1 516-344-6356
Fax	1 516-344-5751
Help Desk	pdbhelp@bnl.gov
General Correspondence	pdb@bnl.gov
WWW Home Page	http://www.pdb.bnl.gov
FTP Server	ftp.pdb.bnl.gov
Network Services	sysadmin@pdb.pdb.bnl.gov
Entry Error Reports	errata@pdb.pdb.bnl.gov
Order Information	orders@pdb.pdb.bnl.gov
User Group	PDBusrgrp@suna.biochem.duke.edu
Listserver Postings	pdb-l@pdb.pdb.bnl.gov
Listserver Subscriptions	listserv@pdb.pdb.bnl.gov
to subscribe, the text of your message should be	subscribe PDB-L Your Name

FTP DIRECTORY STRUCTURE FOR ENTRIES

The PDB has updated the FTP server (ftp.pdb.bnl.gov) in order to have a more standardized directory structure. This will facilitate use of mirror software to keep local copies of the database current. Links have been put in so the server still "looks" the same to users.

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.

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

516-344-3629

First Class