PDB NEWSLETTER

Number 2 ◆ APRIL 1999 ◆ RCSB

Published quarterly by the Research Collaboratory for Structural Bioinformatics

CONTENTS

PDB SNAPSHOT

9631 released atomic coordinate entries

MOLECULE TYPE

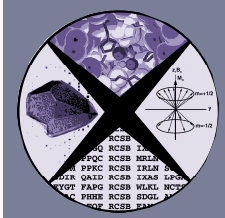
8551 proteins, peptides, and viruses
403 protein/nucleic acid complexes
655 nucleic acids
12 carbohydrates

EXPERIMENTAL TECHNIQUE

214 theoretical modeling
1504 NMR
7913 diffraction and other
2716 structure factor files
527 NMR restraint files

RCSB

SDSC: www.rcsb.org
RUTGERS: rcsb.rutgers.edu
NIST: rcsb.nist.gov
E-MAIL: info@rcsb.org
FTP: ftp.rcsb.org
BNL: www.pdb.bnl.gov



The Research Collaboratory for Structural Bioinformatics (RCSB) is a non-profit consortium dedicated to improving our understanding of the function of biological systems through a study of 3-D biological macromolecular structure.

Message from Helen M. Berman

Welcome to the second Research Collaboratory for Structural Bioinformatics (RCSB) newsletter. The communication with PDB users and the development of RCSB PDB tools has made the past several months very exciting. We really appreciate your feedback, which guides the further development of the system.

On April 1, 1999, the 10,000th structure (1CD3) was released by the Protein Data Bank (PDB). This event occurred in the midst of several other important dates in the development of the RCSB and the PDB.

In March, it was decided by the NSF, Brookhaven National Laboratory (BNL), and RCSB that the timetable for the transition of the PDB should be accelerated. The RCSB will assume full responsibility for the NSF-funded PDB on July 1, 1999 — a full three months ahead of schedule. At that point, the RCSB will have responsibility for all PDB operations formerly carried out by BNL.

The details of the transition are described in this newsletter and on our Web site at **http://www.rcsb.org/pdb/transition_status.html.**

This progress is possible because all aspects of the project involving the RCSB, from deposition of structural data, through query and distribution, to long-term archival and clean-up of original data have proceeded smoothly and because of the cooperation of the RCSB and BNL staff.

The summer promises to be as busy as the last few months as we prepare to launch the third phase of the query software. We will also have an exhibit booth at the International Union of Crystallographic (IUCr) meeting — please stop by and let us know how we are doing!

We look forward to the next 10,000 structures.

Helen M. Berman ✦

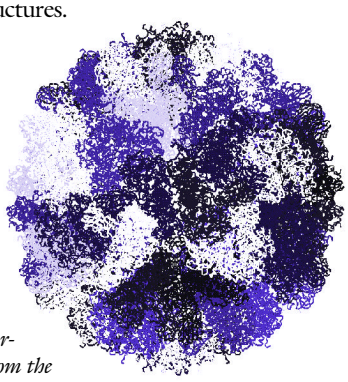


Figure 1: 1CD3, PDB's 10,000th structure. T.Dokland, R.A.Bernal, A.Burch, S.Pletnev, B.A.Fane, M.G.Rossmann: The Role of Scaffolding Proteins in the Assembly of the Small, Single-Stranded, DNAVirus FX174, to be published, 1999. This figure shows the full FX174 viral coat, built from 60 icosahedrally arranged copies of the protein subunits deposited in this entry. Note: This count of 10,000 structures includes both the currently released and obsolete entries available from the PDB.

Summer Events

The PDB will have an exhibition booth at the IUCr meeting to provide computer demonstrations of the query and deposition software, and to meet with PDB users. Please stop by and say hello!

IUCr Congress and General Assembly, August 4-13,1999,Glasgow, Scotland.A PDB Users Group Meeting will be held at this meeting on August 7 at 12:45 PM in the ALSH Room. ✦

Transition Update

Due to the advances made by the PDB project in data deposition and processing, and query and distribution, the RCSB will be taking over complete responsibility for the PDB starting July 1, 1999.

Data Deposition and Data Processing

The transition goals for data deposition and data processing have been met. Since January 27, 1999, data deposited at the PDB have been processed by the RCSB staff using the AutoDep Input Tool (ADIT). Virtually all of the structures received have been processed to completion, regardless of their release status. It does not matter if a structure is intended for immediate release or to be held until publication — structures are processed as they are deposited.

ADIT, the deposition tool developed by the RCSB, has successfully completed its beta test. ADIT is available for the deposition of all structures at **http://pdb.rutgers.edu/** or by clicking on the DEPOSIT link on the RCSB home page, **http://www.rcsb.org/.**

Distribution

The responsibility for the distribution of the data has been successfully transferred to the RCSB. Since February 3, 1999, the RCSB has been distributing the PDB FTP archive. Both users and mirrors access the PDB data from the RCSB Web and FTP sites.

The FTP site will continue to be maintained in its current form at **ftp://bnlarchive.rcsb.org** so that the mirrors of the BNL site will be able to maintain operation. In addition, an improved RCSB FTP file structure will be available at **ftp://ftp.rcsb.org** on May 3, 1999.

Traffic to the BNL Web site after June 30, 1999, will be redirected to the RCSB Web site.

New mirrors of the RCSB Web site will soon be established in Japan and the UK. Additional mirrors of the RCSB PDB Web and FTP sites will be added in the future.

The 1999 Third Quarter PDB CD-ROM will be produced by the RCSB.

Query

Phase I of the Web query interface (SearchLite) has been providing convenient search and analysis capabilities for one or more structures.

Phase II developments in the RCSB's query capability include a user interface for commonly used structure selection criteria and the ability to search the status of structures deposited with the PDB. These Phase II developments will be available from May 15, 1999, and are described in an article in this newsletter.

Phase III, which will incorporate even more advanced query options, will appear in the coming months. ◊

Structure Validation at The RCSB PDB

Introduction to Structure Checking by the RCSB

In order to provide the community with high quality data, the RCSB has developed a number of tools that are used to deposit and process X-ray and NMR structures. Deposition is accomplished through ADIT, which collects all the information necessary for data processing. Data processing involves checking various aspects of the structure and the data collected through ADIT.

Some of these checks are available through the Validation Server, a new tool that allows the user to check the format of coordinate and structure factor files and to perform a variety of validation tests on the structure prior to deposition in the PDB. Available at **http://pdb.rutgers.edu/validate/**, these checks can be done independently by the user.

Validation involves two steps: the coordinate format check and structure validation. The coordinate format precheck produces a brief report identifying any changes that need to be made in a data file in order to obtain a validation report.

Structure validation presents the user with a validation summary letter that contains a collection of structural and nomenclature diagnostics, including bond distance and angle comparisons, chirality, close contacts, and sequence comparisons. This letter is designed to determine features of a structure that may require some special attention (e.g., close contacts) and to present this information in a concise summary report. The validation letter is produced by the RCSB program MAXit[1].

In addition, the Validation Server presents an Atlas summary page, graphic images, and diagnostic reports from a variety of programs: PROCHECK v3.4.4[2], PROCHECK-NMR[3], NUCHECK[4], SFCHECK[5], and CURVES 5.1[6]. Future versions of the Validation Server will include output from WHAT IF[7].

Once a structure is deposited to the PDB, it is checked using the same procedures that were available in the Validation Server and many corrections are made automatically. Other aspects of the

[1] *Macromolecular Exchange and Input Tool, a program originally developed by the Nucleic Acid Database Project to assist in the processing and curation of macromolecular structure data.*

[2] *R.A.Laskowski, et al., "PROCHECK: A Program To Check the Stereochemical Quality of Protein Structures,"* J. Appl. Cryst., *26 (1993): 283-291.*

[3] *R.A.Laskowski, et al., "AQUA and PROCHECK-NMR: Programs For Checking The Quality of Protein Structures Solved by NMR,"* J. of Biomol. NMR, *8 (1996): 477-486.*

[4] *Z.Feng, J.Westbrook, H.M.Berman, Rutgers University, New Brunswick, NJ, (1998): NDB-407.*

[5] *A.A.Vaguine, J.Richelle, S.J.Wodak,* Acta Cryst. D, *55 (1999): 191-205.*

[6] *R.Lavery and H.Sklenar, "The Definition of Generalized Helicoidal Parameters and of Axis Curvature in Irregular Nucleic Acids,"* Biomol. Struct. Dynam., *6 (1998): 63-91.*

[7] *G.Vriend, "WHAT IF: A Molecular Modeling and Drug Design Program,"* J. Mol. Graph. *8 (1990): 52-56.*

structure require expert judgment to provide fully processed PDB entries. This is done by the data curators in collaboration with the depositors.

Structure Checks

In this section, we outline the results contained in the validation summary letter for both NMR and X-ray crystal structures.

1. Covalent Bond Distances and Angles

Covalent bond distances and angles for macromolecules are compared against standard values. For proteins, these are taken from Engh and Huber[8]. The standard values for nucleic acid bases are taken from Clowney, et al.[9], and for nucleic acid sugar and phosphates from Gelbin, et al.[10] Bonds and angles related to hydrogens are not checked.

For each type of bond (e.g., N-CA, N-C) or angle (e.g., N-CA-C, CA-CB-CG), the RMS deviation of that bond or angle (V_actual) relative to the standard value (V_standard) is:

$$RMSD = (\Sigma(V\_actual - V\_standard)^2/N)^{1/2}$$

where N = the number of individual angles or bonds of a particular type included in the summation.

V_actual for a particular bond is listed as 4*RMSD violation if:

$$(V\_actual - V\_standard) > 4*RMSD$$

In addition, the validation summary provides the total average deviations from standard dictionaries. This offers an overall measure of agreement with the standard values.

Other methods exist to make this comparison; however, this approach tends to highlight only serious outliers. The selection of the cutoff value of 4*RMSD is used to maintain compliance with previous PDB practice. In consultation with the community and in response to various comments, including those of our advisors, we plan to change this value to 6*RMSD to ensure the reporting of truly serious outliers.

2. Stereochemical Validation

All chiral centers of proteins and nucleic acids are checked for correct stereochemistry. Violations of standard stereochemistry are reported for both proteins and nucleic acids using the following method:

```
Neighboring atoms a, b and c of the chiral center
       form vectors V_a,V_b and V_c with the center.
```

The chiral volume is:

$$VC = V\_a * V\_b * V\_c$$

If the sign of the actual chiral volume is different from the standard chiral volume, a chirality violation is listed.

3. Atom nomenclature

The nomenclature of all atoms is checked for compliance with the current PDB standard. In some cases, this nomenclature is not in complete agreement with IUPAC [11] standards. This is particularly true for hydrogen atoms. Correspondence of PDB hydrogen nomenclature with the nomenclature used by IUPAC and many refinement programs is available at **http://www.bmrb.wisc.edu/ref_info/atom_nom.tbl**. In the near future, the hydrogen nomenclature will be brought into compliance with IUPAC standards.

In addition, particular attention is paid to the nomenclature of hydrogens on the ND2 atoms of Asn residues and/or the NE2 atoms of Gln as well as the NH1 and NH2 atoms of Arg residues for agreement with the standard for E/Z orientation presented by the IUPAC[11]. For nucleic acids, the atom labeling of O1P/O2P atoms are checked against the convention defined by the IUBMB[12].

During processing, the nomenclature of all of the above atoms is adjusted if necessary.

4. Close contacts

MAXit calculates the distances between all atoms within the asymmetric unit of crystal structures and the unique molecule of NMR structures. For crystal structures, contacts between symmetry-related molecules are checked as well. These checks include ligand and solvent molecules in addition to the macromolecular structure. Atoms less than 2.2 Angstroms apart are listed as close contacts.

Interactions of atoms forming standard bonds, defined through PDB LINK records, or related by 1-4 contacts, are not listed as close contacts.

In crystal structures, atoms that have full occupancy and lie on special positions are listed as having close contacts to indicate that a lower occupancy is appropriate. Atoms with full occupancy related by a crystallographic symmetry element will be listed as having close contacts if they are less than 2.2 Angstroms apart.

If disulfide bridges are denoted in the coordinate file with SSBOND records, they will not be listed as close contacts.

In data annotation, close contacts corresponding to metal coordination are represented as LINK records in the PDB file.

5. Ligand and Atom Nomenclature

The names of residues and atoms are compared against the nomenclature used in the PDB dictionary: **ftp://ftp.rcsb.org/pub/pdb/data/monomers/het_dictionary.txt**, for all ligands as well as standard residues and bases. Unrecognized ligand groups are flagged and any discrepancies in known ligands are listed as extra or missing atoms.

When structures are processed, residue and atom nomenclature for existing HET groups are corrected to follow the residue and atom naming convention that is given in the PDB HETgroup dictionary at **ftp://ftp.rcsb.org/pub/pdb/data/monomers/het_dictionary.txt**. For new ligands, a residue name is assigned with the preference given to that provided by the author and the ligand is compared topologically against the dictionary, to find similar molecules. Such similar molecules are used to create the most appropriate atom nomenclature for the new group.

[8] *R.A.Engh and R.Huber, "Accurate Bond and Angle Parameters for X-ray Protein structure refinement,"* Acta Crystallogr, *A47 (1991): 392-400.*

[9] *L.Clowney et al., "Geometric Parameters in Nucleic Acids: Nitrogenous Bases,"* J. Am. Chem. Soc., *118 (1991) 509-518.*

[10] *A.Gelbin et al., "Geometric Parameters in Nucleic Acids: Sugar and Phosphate Constituents,"* J. Am. Chem. Soc., *118 (1996) 519-529.*

[11] *J.L.Markley, et al., "Recommendations for the Presentation of NMR Structures of Proteins and Nucleic Acids,"* Pure & Appl. Chem., *70 (1998): 117-142.*

[12] *C.Liebecq,* Compendium of Biochemical Nomenclature and Related Documents, *2d ed., Portland Press: London and Chapel Hill, 1992.*

We are currently standardizing the existing HET group dictionary in order to make it more usable both for ourselves and the community. Significant effort is being made to classify the contents of the dictionary and to correct errors. For example, the same group may exist with multiple names or similar groups may have widely varying atom nomenclature. As soon as the new dictionary is completed, it will be made publicly available.

6. Sequence Comparison

The sequence given in the PDB SEQRES records is compared against the sequence derived from the coordinate records. This information is displayed in a table where any differences or missing residues are marked.

During structure processing, the sequence database references given by DBREF and SEQADV are checked for accuracy. If no reference is given, a BLAST[13] search is used to find the best match. Any conflicts between the PDB SEQRES records and the sequence derived from the coordinate records are resolved by comparison with various sequence databases. Residues in disordered regions modeled as alanines are switched both in the SEQRES and in the coordinate section to their true residue names. In general, the sequence and coordinates are made to reflect the sequence of the protein studied, even if it was not possible to model every region.

7. Distant waters

MAXit calculates the distances between all water oxygen atoms and all polar atoms (oxygen and nitrogen) of the macromolecules, ligands, and solvent in the asymmetric unit. Waters further than 3.5 Angstroms are listed in the validation report. Thus, second, third, etc. hydration shell waters are excluded from the list. Water-water groupings which are as a whole distant from the macromolecule or ligands are included in the listing. Waters further than 5.0 Angstroms are listed in the PDB file.

For X-ray crystal structures, the validation summary also lists distant water (> 3.5 Angstroms from polar atoms of the macromolecules, ligands, or solvent of the asymmetric unit) that can be moved through the application of symmetry operations to be closer to the asymmetric unit. For example, if the closest contact that, the oxygen of a water molecule makes with a polar atom of the macromolecules, ligands, or solvent of the asymmetric unit is 5.5 Angstroms, and a symmetry operation (for example, 2_456 in spacegroup P 21) would place this water in closer proximity to a polar group of the asymmetric unit, then the water will be relocated.

For all structures, the Atlas Summary presents molecular graphic images so that the overall appearance of the structure can be checked. For X-ray crystal structures, molecular graphic images are generatedin GIF and VRML for the asymmetric unit and for crystal packing. For NMR structures, molecular images are generated for the structure and the ensemble.

Any questions regarding the use or the content of the Validation Server should be directed to **help@rcsb.rutgers.edu**.◊

[13] *Z.Zhang, et al., "Protein sequence similarity searches using patterns as seeds,"* Nucleic Acids Res., *26 (1998): 3986-3990.*

Phase II of PDB Query

RCSB plans call for three releases of the PDB Web query interface in the first year of RCSB operation, that is through October 1, 1999. Phase I provided a simple text search interface called SearchLite, as described in the previous newsletter **http://rcsb.rutgers.edu/pdb/newsletter/1999q1/article6.shtml**. While valuable, since it covers all possible search terms contained in a PDB file, the SearchLite interface does not address questions like "what structure(s) has author brown had a hand in solving since 1990?" Phase II provides this level of query capability through the "Comprehensive Query" option. Comprehensive query enables you to customize the query form to include the query terms that you wish to access. For the above query, you would use an "Author" and "Deposition Date After" field. Explicit query fields are available for:

- General information (e.g., PDB HEADER fields, authors, deposition date)
- Crystallographic information (e.g., resolution, space group, all cell parameters)
- Experimental (refinement)
- Sequence (complete using FASTA and short string)
- Features (molecular weight, secondary structure content, EC number)

Part of the RCSB mandate is to better annotate structures, both those already in the database and those being deposited. The first step in this direction has been a structure classification into enzyme, protein-containing, DNA-containing, RNA-containing, carbohydrate-containing, and glycoprotein-containing entries. This classification can be included in a Phase II query.

Another feature of the Phase II query capability is the ability to determine the status of an entry. Since structures are generally processed within two weeks of receipt, the need for processing status is limited. What remains important is finding structures that are on hold and determining when they will become available. A separate query status page permits the user to search for entries on hold by name, PDB id, or release date.

Phase II query will be available from May 15, 1999.

Phase III query, due for release later in the year, will extend these Phase II features and include the ability to query non-redundant sets of data based upon both sequence and similar fold. ◊

A Fast Topology-Based Structural Comparison Service

Users are invited to use an updated release of a fast topology-based structural comparison service at **http://tops.ebi.ac.uk/tops/compare.html**. The service uses TOPS topology diagrams and patterns, and is part of the suite of programs available on the TOPS web-site at **http://tops.ebi.ac.uk**. TOPS topology diagrams are also linked from the PDB via PDBSum in the "Other Sources" section of the "Structure Explorer" pages.

This service permits comparison of user-supplied target protein domains with the TOPS Atlas (a representative set of 3000 domains), or the entire PDB (over 18,000 domains, as of January

1999). Alternatively, users can supply the descriptions of two protein domains and compare them to each other.

The system also performs an analysis based on possible common motifs contained in the target file, plus some information about strand pairs and secondary structure types. The output per match comprises a distance, and can optionally be annotated with the matching residue numbers, and the common topological pattern.

This structure comparison method is not based on aligning atomic coordinates of residues, but on making a structural alignment of secondary structure elements (helices and strands) using additional information about relationships between strands in sheets, etc., plus chiralities. The method works best for structures containing beta sheets, but an attempt will be made to align all-alpha structures if there is sufficient chirality information.

The advantage of the method is that individual comparisons are fast — a comparison of one structure against the Atlas takes on average 1-3 minutes of CPU time, and from under 10 minutes to one hour against the entire PDB (on a DEC Alpha). However, other structural comparison techniques based on making structural alignments at the atomic coordinate level are likely to be more accurate.

Uploaded files are deleted after the comparison has been made.

PDB file format: Files will be pre-processed by the DSSP program under Unix. If your input file(s) cannot be processed by DSSP, then the comparison will fail. You may remove any information (e.g., the amino-acid sequence) which will not abort DSSP, and thus preserve some anonymity of your data. You are strongly advised to edit your PDB file so that it only contains the description of the domain in which you are interested. Multi-domain files will take longer to process than single-domain files; also, the comparison databases contain single domain descriptions. Jobs submitted may be subject to queueing in the case of heavy processor usage.

Please mail comments to David Gilbert at **drg@ebi.ac.uk**, who is responsible for maintaining the service as part of his ongoing research project at EBI as a part-time visiting researcher. The original research effort was supported by an EPSRC grant for DRG as a Visiting Research Fellow at EBI while on sabbatical from City University during 1998.

David Gilbert, *drg@ebi.ac.uk*
Department of Computing, City University, London, UK and EBI (visiting)
David Westhead, *westhead@bmb.leeds.ac.uk*
School of Biochemistry and Molecular Biology, Leeds University, UK
Janet Thornton, *thornton@biochem.ucl.ac.uk*
Department of Biochemistry, University College London, and Crystallography Department, Birkbeck College, and European Bioinformatics Institute

The Most Wanted Structures in the PDB

The BNL PDB Web logs from the last six months of 1998 were analyzed in order to determine which structures were the most broadly accessed. A structure was considered to be "broadly accessed" if it was frequently investigated across a spectrum of Web pages and activities, namely, saving the PDB file to disk in 3DB and PDB Lite, viewing the full coordinate set, and viewing the structure in RasMol and Chime.

A structure had to be one of the 100 most frequently accessed entries by each of these actions in order to be selected. The specific URLs corresponding to these user actions are listed below with tubulin (PDB id 1tub) as an example:

1. Viewing the full coordinates (3DB only)
http://www.pdb.bnl.gov/pdb-bin/send-pdb?id=1tub

2. Saving to a disk (3DB only):
http://www.pdb.bnl.gov/pdb-bin/save-pdb?id=1tub

3. Saving to a disk (PDB Lite only):
http://www.pdb.bnl.gov/pdb-docs/litesave.html?v1=1TUB

4. Viewing with Chime(3DB & PDB Lite)
http://www.pdb.bnl.gov/pdb-bin/ccpeek?id=1TUB

5. Viewing with RasMol(3DB & PDB Lite)
http://www.pdb.bnl.gov/pdb-bin/send-ras?id=1tub

A total of 17 PDB entries met this criteria. They are listed below in alphabetical order by id:

```
'1A3N','HEMOGLOBIN' Deposited: 01/22/1998
'1AG2','MAJOR PRION PROTEIN' Deposited: 03/31/1997
'1ALK','ALKALINE PHOSPHATASE (E.C.3.1.3.1)' Deposited:
       03/03/1993
'1AMO','OXIDOREDUCTASE' Deposited: 06/17/1997
'1AO6','SERUM ALBUMIN' Deposited: 07/18/1997
'1AOI','HISTONE H3, HISTONE H4, HISTONE H2A, HISTONE
       H2B/DNA COMPLEX' Deposited: 07/03/1997
'1ATN','ACTIN,DEOXYRIBONUCLEASE I (DNASE I)'
       Deposited: 03/08/1991
'1BL8','POTASSIUM CHANNEL PROTEIN' Deposited:
       07/23/1998
'1BMF','BOVINE MITOCHONDRIAL F1-ATPASE' Deposited:
       03/13/1996
'1D66','GAL4/DNA COMPLEX' Deposited: 03/06/1992
'1GC1','ENVELOPE GLYCOPROTEIN GP120, SURFACE GLYCOPRO-
       TEIN CD4, ANTIBODY 17B (LIGHT CHAIN), ANTI-
       BODY 17B (HEAVY CHAIN)' Deposited: 06/15/1998
'1HHO','HEMOGLOBIN A (OXY) (ALPHA CHAIN), HEMOGLOBIN
       A (OXY) (BETA CHAIN)' Deposited: 06/10/1983
'1IR3','INSULIN RECEPTOR, PEPTIDE SUBSTRATE'
       Deposited: 09/22/1997
'1KZU','LIGHT HARVESTING PROTEIN B-800/850' Deposited:
       08/31/96
'1OCC','CYTOCHROME C OXIDASE' Deposited: 04/18/1996
'1TUB','TUBULIN' Deposited: 09/23/1997
'2ACE','ACETYLCHOLINESTERASE' Deposited: 06/23/1996
```

It was decided not to rank PDB entries solely by their number of hits because that number may not always provide an accurate representation of user interest in an individual entry. An example of the problem is illustrated by Cytochrome B5, id 3b5c. This entry did not meet these selection criteria, but did have a very high hit number. Cytochrome B5 was viewed in Rasmol more than 25 times as often as it was viewed using Chime. One explanation for this discrepancy is that Cytochrome B5 is used in a Web-based RasMol tutorial by Gail Rhodes of the University of Southern Maine at **http://www.usm.maine.edu/~rhodes/RasTut/index.html**.

An alternative analysis, carried out for the final three months of 1998, was recently circulated by Eric Martz to the PDB discussion group and may be found at **http://www.rcsb.org/pdb/lists/pdb-l/199904/msg00001.html.**

For more general information on Web log analysis, please see the following excellent articles:

"Interpreting WWW Statistics" by Doug Linder
http://gopher.nara.gov:70/0h/what/stats/webanal.html

"Making Sense of Web Usage Statistics" by Dana Noonan
http://www.piperinfo.com/pl01/usage.html

"Getting Real about Usage Statistics" by Tim Stehle
http://www.wprc.com/wpl/stats.html

"Why Web Usage Statistics are (Worse Than) Meaningless" by Jeff Goldberg: **http://www.cranfield.ac.uk/stats/** ◊

Citing the Protein Data Bank

Several users of the PDB have recently asked what is the best way to cite structures that they have obtained from the PDB. In the same time-frame, some members of the crystallographic community have also expressed the concern that when a structure is downloaded from the PDB for analysis, it is sometimes the PDB rather than the workers who were responsible for the original structure determination that is referenced.

We share these concerns and have provided the following guidelines (also available at **http://www.rcsb.org/pdb/citing.html**) for citing structures downloaded from the PDB:

The contents of PDB are in the public domain, but it is expected that the authors of an entry as well as the PDB be properly cited whenever their work is referred to. Structures used from the PDB should be cited with the PDB id and the JRNL reference.

For example, structure 102l should be referenced as:

```
PDB ID: 102l.
```

D.W.Heinz, W.A.Baase, F.W.Dahlquist, B.W.Matthews, "How Amino-Acid Insertions are Allowed in an Alpha-Helix of T4 *Lysozyme,"* Nature, *361 (1993): 561.*

The Brookhaven National Laboratory PDB should be referenced:

F.C.Bernstein, T.F.Koetzle, G.J.Williams, E.E.Meyer Jr., M.D.Brice, J.R.Rodgers, O.Kennard, T.Shimanouchi, M.Tasumi, "The Protein Data Bank: A Computer-based Archival File For Macromolecular Structures," J. of. Mol. Biol., *112 (1977): 535.*

The Research Collaboratory for Structural Bioinformatics PDB should be referenced with the WWW address: **http://www.rcsb.org/pdb/.** ◊

Developments in the PDB Newsletter

Submitted Articles

In the future, the RCSB PDB newsletter will carry submitted articles, such as Gilbert, Westhead, and Thornton's article in this issue.

Our publication policy is that submitted articles have a clear scientific or news content that is of interest to PDB users. Example article topics include databases for structural biology, software and new calculation methods for structure determination and analysis, and the results of structure analysis. News of importance to users of the PDB includes announcements of new databases and resources for the user community, releases and updates to software for structure determination or analysis, and policy changes by major funding agencies. Articles from commercial vendors will be considered for publication provided that they conform to these guidelines. Pieces that appear to be purely advertising (whether from a commercial or academic source) will not be published.

Articles for the PDB newsletter may be submitted by e-mail at any time to John Badger at **badger@sdsc.edu** in plain text form. These articles will be circulated for review by members of the RCSB PDB; external reviewers may also be called upon if necessary to evaluate articles. Questions regarding the appropriateness of an article or other issues related to publication in the PDB newsletter should also be submitted to John Badger.

Publication of a submitted article in the PDB newsletter will not in anyway constitute an endorsement of the materials it contains, and any views expressed in submitted articles are considered solely those of the authors.◊

Newsletter Distribution

The RCSB PDB newsletter is currently distributed electronically (HTML, PDF and plain text formats) as well as by postal mail to maintain all of the channels of communication that had previously been provided by BNL's PDB newsletter.

Our Web site archives the HTML and PDF versions of the newsletter. We believe that the Web is the most convenient and accessible source for most PDB users. The HTML version provides quick access, while the PDF version offers high quality and easy printing. This PDF formatted version also provided the basis for creating our own printed copies.

The newsletter is also distributed as a plain text e-mail message, since this provides a channel with low bandwidth and instant communication. The current e-mail distribution list now contains approximately 550 recipients and is still growing. A form for subscription to the e-mail distribution list is available at **http://www.rcsb.org/pdb/forum.html**. Alternatively, you may subscribe by e-mailing the message "subscribe news" (without quotation marks) to **majordomo@rcsb.org**.

At the present time, printed copies of the RCSB PDB newsletter are also mailed to addresses on a subscription list that was provided to us by BNL. This list contains almost 600 addresses for individual scientists and institutions.

At a time when many journals are moving away from print towards electronic distributions, it is appropriate to reconsider the necessity of the printed mailings of the PDB newsletter. For example, a recent article in Nature, devoted considerable discussion to the future and the survival of printed journals in the biological sciences (Nature, Vol 397:195-200). As a concrete example of the changes that are beginning to take place, the Biophysical Journal has moved to an electronic distribution for Biophysical Society members.

We will use our participation in scientific meetings over the next few months as an opportunity to survey PDB users on the level of interest in maintaining the hard copy distribution to individual subscribers and to determine ways in which our other distribution channels might be improved. ◊

RCSB PDB Meetings, Presentations, and Staff

Planned RCSB Presentations 1999

MEETING/SEMINAR	DATE/LOCATION/ATTENDEE
Gorden Conference on Computational Aspects of Biomolecular NMR	June 5-11, 1999 *Borga, Italy*
Cambridge Healthtech Institute's Eighth Annual Meeting "Bioinformatics and Genome Research"	June 14-15, 1999 *San Francisco, CA* Phil Bourne
Eleventh Conversation in Biomolecular Stereodynamics	June 15-16, 1999 *Albany, NY* Helen M. Berman
Thirteenth Symposium of the Protien Society	July 24-28, 1999 *Boston, MA* Christine Zardecki and Bohan Schneider
International Union of Crystallography Congress and General Assembly	August 4-13, 1999 *Glasgow, Scotland* Helen M. Berman, Phil Bourne, and John Westbrook

RCSB Presentations (January 1, 1999-June 22, 1999)

MEETING/SEMINAR	DATE/LOCATION/ATTENDEE
Finding the Path: Issues of Access to Research Resources *National Science and Technology Council Committee on Science, Subcommittee on Biotechnology*	January 27-28, 1999 *Washington, DC* Helen M. Berman
Argonne National Laboratory	February 11, 1999 *Chicago, IL* Helen M. Berman
NIGMS Structural Genomics Targets Workshop	February 11-12, 1999 *Washington, DC* Helen M. Berman
Biophysical Society Annual Meeting	February 13-17, 1999 *Baltimore, MD* Helen M. Berman
The Association of Biomolecular Resource Facilities '99: Bioinformatics & Biomolecular Technologies	March 19-22, 1999 *Durham, NC* Helen M. Berman and Phil Bourne
The Sealy Center Fourth Symposium on Structural Biology	March 19-21, 1999 *Galveston, TX* Helen M. Berman
Data Mining Course	May 12-23, 1999 *Erice, Italy* Helen M. Berman and Phil Bourne
American Crystallographic Association Annual Meeting	May 22-27, 1999 *Buffalo, NY* Helen M. Berman, Phil Bourne, Gary Gilliland, and John Westbrook
Biochemistry and Molecular Biology 1999 Meeting	May 16-20 1999 *San Francisco, CA* Christine Zordecki
Biotechnology Industry Organization: Bio99	May 16-20, 1999 *Seattle, WA* Gary Gillilad, Phoebe Fagan, Greg Vasquez

RCSB Project Team

The overall operation of the RCSB PDB is managed by the RCSB Project Team. Technical and scientific support is provided by the RCSB Members.

Dr. Helen M. Berman
Department of Chemistry
Rutgers University
610 Taylor Road
Piscataway, NJ 08854-8087
732-445-4667
Fax: 732-445-4320
berman@rcsb.rutgers.edu

Dr. John Westbrook
Department of Chemistry
Rutgers University
610 Taylor Road
Piscataway, NJ 08854-8087
732-445-4290
Fax: 732-445-4320
jwest@rcsb.rutgers.edu

Dr. Gary Gilliland
Center for Advanced Research in Biotechnology
National Institute of Standards and Technology
Mail Stop 8310
Gaithersburg, MD 20899-8310
301-975-2629
Fax: 301-330-3447
gary.gilliland@nist.gov

Phoebe Fagan
Center for Advanced Research in Biotechnology
National Institute of Standards and Technology
9600 Gudelsky Drive
Rockville, MD 20850
301-738-6285
Fax: 301-738-6280
phoebe.fagan@nist.gov

Dr. Peter Arzberger
San Diego Supercomputer Center
University of California, San Diego
9500 Gilman Drive
La Jolla, CA 92093-0505
619-534-5079
Fax: 619-822-0948
parzberg@sdsc.edu

Dr. Phil Bourne
San Diego Supercomputer Center
University of California, San Diego
9500 Gilman Drive
La Jolla, CA 92093-0537
619-534-8301
Fax: 619-822-0873
bourne@sdsc.edu

BNL PDB Staff

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

Dr. Betty R. Deroski
Dr. Arthur Forman
Sabrina Hargrove
Dr. Jiansheng Jiang
Mariya Kobiashvili
Patricia A. Langdon
Dr. Dawei Lin
Nancy O. Manning
John E. McCarthy
Regina K. Shea
Janet L. Sikora
Lu Sun
Dr. S. Swaminathan
Dejun Xue

Statement of Support *The RCSB's PDB is supported by funds from the National Science Foundation, The Office of Biology and Environmental Research at the Department of Energy, and two units of the National Institutes of Health: The National Institute of General Medical Sciences and the National Library of Medicine.*

PDB Staff

RCSB Members

NIST

CARB
9600 Gudelsky Drive
Rockville, MD 20850

Dr. Hillary S. Gilson
301-738-6106
Fax: 301-738-6280
hillary.gilson@nist.gov

Dr. Diane Hancock
301-738-6110
Fax: 301-738-6280
diane.hancock@nist.gov

Dr. Narmada Thanki
301-738/-6204
Fax: 301-738-6280
narmada@carb.nist.gov

Dr. Michael Tung
301-738-6284
Fax: 301-738-6280
michael.tung@nist.gov

Dr. Greg Vasquez
301-738-6209
Fax: 301-738-6280
gregory.vasquez@nist.gov

Dr. Lynn Young
301-738-6283
Fax: 301-738-6280
lynn.young@nist.gov

UCSD/SDSC

SDSC
UC San Diego
9500 Gilman Drive
La Jolla, CA 92093

Dr. John Badger
619-822-0936
Fax: 619-822-0873
badger@sdsc.edu

Terri Gaughn
619-822-0882
Fax: 619-822-0873
pdbadmin@sdsc.edu

Dr. Douglas S. Greer
619-534-8357
Fax: 619-534-5113
dsg@sdsc.edu

Dr. Michael Gribskov
619-534-8312
Fax: 619-822-0873
gribskov@sdsc.edu

Dr. Glen Otero
619-822-0974
Fax: 619-822-0873
gotero@sdsc.edu

Anne Kuller
619-534-5171
Fax: 619-822-0873
akuller@sdsc.edu

Dr. Arcot K. Rajasekar
619-534-8378
Fax: 619-534-5077
sekar@sdsc.edu

Shawn Strande
619-534-5133
Fax: 619-534-5117
strande@sdsc.edu

Dr. Lynn F. Ten Eyck
619-534-5141
Fax: 619-534-5113
teneyckl@sdsc.edu

Dr. Helge Weissig
619-534-8399
Fax: 619-822-0873
helgew@sdsc.edu

Dr. Kenneth Yoshimoto
619-822-0886
Fax: 619-822-0873
kenneth@sdsc.edu

RUTGERS

Department of Chemistry
Rutgers University
610 Taylor Road
Piscataway, NJ 08854-8087

Kyle Burkhardt
732-445-2200
Fax: 732-445-4320
kburkhar@rcsb.rutgers.edu

Dr. Zukang Feng
732-445-0103
Fax: 732-445-4320
zfeng@rcsb.rutgers.edu

Dr. Anke Gelbin
860-685-2777
Fax: 860-685-2211
ankeg@rcsb.rutgers.edu

Dr. Shri Jain
732-445-0103
Fax: 732-445-4320
sjain@rcsb.rutgers.edu

Dr. Rachel Kramer
732-445-2200
Fax: 732-445-4320
kramer@rcsb.rutgers.edu

Victoria Petrova
732-445-0103
Fax: 732-445-4320
victoria@rcsb.rutgers.edu

Dr. Bohdan Schneider
732-445-0103
Fax: 732-445-4320
bohdan@rcsb.rutgers.edu

Dr. Kata Schneider
732-445-0103
Fax: 732-445-4320
kata@rcsb.rutgers.edu

Christine Zardecki
732-445-0103
Fax: 732-445-4320
zardecki@rcsb.rutgers.edu

Protein Data Bank

University of California, San Diego
San Diego Supercomputer Center, MC 0537
9500 Gilman Drive
La Jolla, CA 92093-0537

ADDRESS SERVICE REQUESTED