Quarterly Newsletter published by
Brookhaven National Laboratory
Protein Data Bank

Release #87 *January 1999*

Protein Data Bank pdb™

Brookhaven National Laboratory

January 1999 CD-ROM Release

9179 Released Atomic Coordinate Entries

	Molecule Type
8143	proteins, peptides, and viruses
381	protein/nucleic acid complexes
643	nucleic acids
12	carbohydrates

	Experimental Technique
206	theoretical modeling
1452	NMR
7521	diffraction and other

2424	Structure Factor Files
522	NMR Restraint Files

The total size of the atomic coordinate entry database is 4.3 GB uncompressed.

Table of Contents

What's New at the PDB 1
Deposition of Structure Factors at the Protein Data Bank 3
PDB World Wide Web Mirroring System 3
Proposal: PDB Depositors Club
Morton Kjeldgaard 5
EBI-MSD 7
Validation of Sugars in the PDB 8
Notes of a Protein Crystallographer-
A Crystal in Time 9
Affiliated Centers and Mirror Sites 12
BNL PDB Access, FTP Directory Structure, Consultants,
Staff, Support 14

Internet Sites

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

RCSB PDB
WWW ..... http://www.rcsb.org
FTP ..... ftp.rcsb.org

What's New at the PDB

Joel L. Sussman

On Oct 1, 1998, the following announcement was made by Rutgers University:

"NEW BRUNSWICK/PISCATAWAY, N.J. - The Research Collaboratory for Structural Bioinformatics (RCSB; http://www.rcsb.org/), a consortium composed of Rutgers, the State University of New Jersey; the University of California at San Diego; and the National Institute of Standards and Technology (NIST), has received a $10 million, five-year award from the National Science Foundation (NSF), the Department of Energy (DOE), and two units of the National Institutes of Health: the National Institute of General Medical Sciences (NIGMS) and the National Library of Medicine (NLM). The award will enable the RCSB to operate and significantly extend the capabilities of the Protein Data Bank (PDB), a critical tool for unlocking the secrets of biological systems in pharmaceutical and medical research."

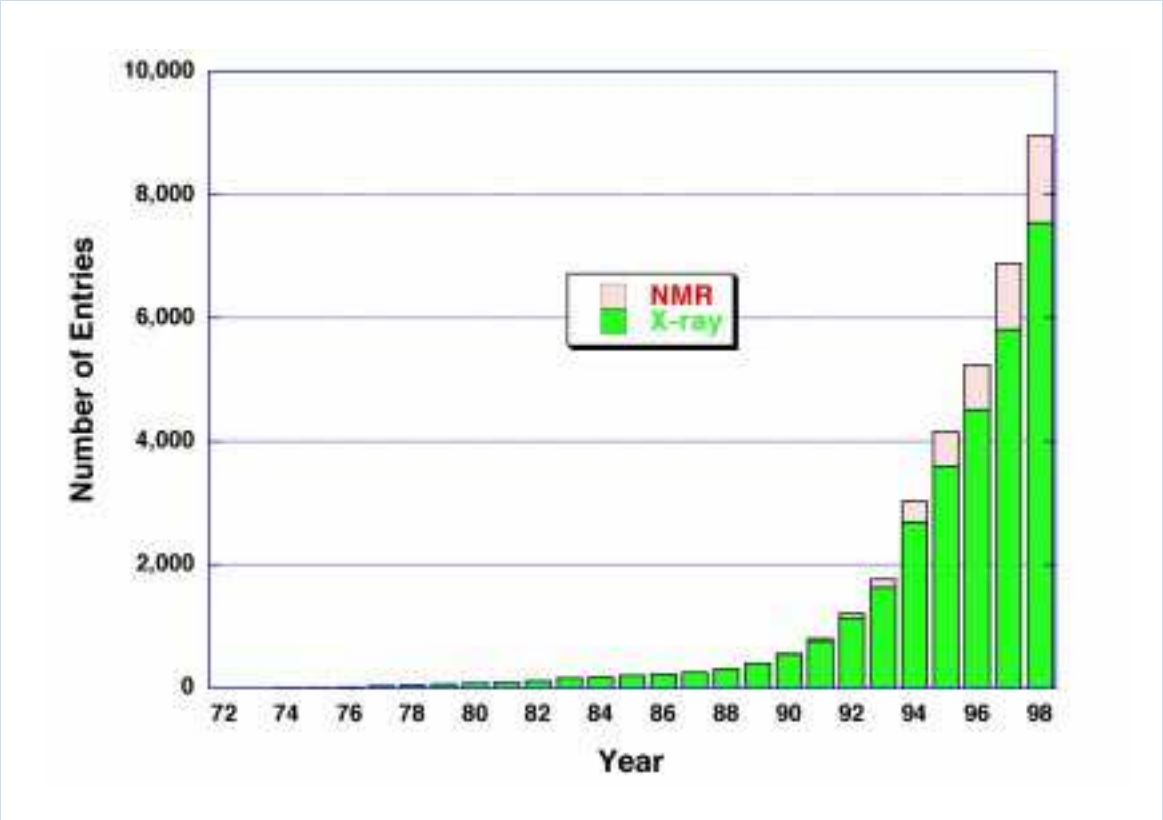
Needless to say, we at the PDB wish the RCSB all the best in continuing the 27-year tradition of the Brookhaven Protein Data Bank. The PDB is at present a major international resource used by scientists, educators and students throughout the world. During the past few years, we at the PDB, in collaboration with many others, have greatly enhanced this resource into a very user-friendly and powerful tool for bridging the gap between the 3D structure and the genome worlds (Sussman, J. L. [1997]. "Bridging the Gap" *Nature Struct. Biol.* **4**, 517). Some examples of this can be seen:

• PDB's AutoDep procedure, which has made deposition of structural data to the PDB much easier, and, more importantly, much richer in information and more accurately checked before release of the data. It has also made uploading coordinates, structure factors and NMR restraints files very simple for the depositors.

• Results of the 'Layered Release Protocol' have exceeded our best expectations, with the number of new entries being requested to be 'on-hold' now down to only ~20% (and still going down) as contrasted to well over 75% just a year ago (Sussman, J. L. [1998]. "Protein Data Bank Deposits" *Science* **282**, 1991).

• The fact that the PDB is now receiving structure factors for a very high percentage of the structures determined by X-ray crystallography (Jiang, J., Abola, E. & Sussman, J. L. [1999]. "Deposition of structure factors at the Protein Data Bank" *Acta Cryst.* **D55**, 4, and reprinted in this Newsletter).

• The close interaction that the PDB now has with most journals relevant to structural studies to ensure deposition in the PDB (and release) of coordinates as a prerequisite for acceptance of manuscripts (see *e.g.,* editorials in: *Proc. Natl. Acad. Sci. USA* [1998] **95**, pg. iii; *Nature* [1998] **394**, 105; *Science* [1998] **281**, 175).



PDB coordinate entries available per year.

Numerous close interactions/collaborations with scientists from around the world has yielded beneficial results for the entire community. This has resulted in the PDB becoming a truly international endeavor, *e.g.*:

First remote PDB deposition site has been established in Europe at the EBI.

Improvement in handling of ligands and Het groups for both deposition and retrieval of information *via* programs developed by M. Hendlich (University of Marburg, Germany) and the CCDC (Cambridge, UK).

PDB Lite & 'Noncovalent Bond Finder' (E. Martz, University of Massachusetts, USA).

The user-friendly way of accessing the PDB *via* the 3DB Browser (developed in close collaboration with Dr. Jaime Prilusky, Bioinformatics Unit, Weizmann Institute of Science, Israel) has already become the standard for several online journals pointing to the PDB atlas pages of structures. In fact, the information presented there is in some ways clearer and easier to read than the methods sections in some journal articles.

The close interaction with the BioMagResBank (BMRB, Univ. of Wisconsin) for the handling of NMR structural data.

The fact that industrially determined 3D structures are now being deposited to the PDB, even without publication, has been made possible *via* the close collaboration between the PDB and the HIV Protease Database (developed by Alexander Wlodawer, at the NCI, Frederick, MD and Jiri Vondrasek at IOCB, Prague, Czech Republic, www.-fbsc.ncifcrf.gov/HIVdb).

The 17 mirror sites in 13 countries around the world now provide easy and fast local access to the PDB web pages.

This work has been carried out by a most dedicated and talented staff at the PDB, led by Enrique Abola, Deputy Head of the PDB, together with Betty Deroski, Arthur Forman, Sabrina Hargrove, Jiansheng Jiang, Mariya Kobiashvili, Pat Langdon, Michael Libeson, Dawei Lin, Nancy Manning, John McCarthy, Christine Metz, Otto Ritter, Regina Shea, Janet Sikora, Lu Sun, Subramanyam Swaminathan and Dejun Xue. In addition, John Rose (Univ. of Georgia), Mia Raves (Utrecht Univ.), Clifford Felder, Kurt Giles, Jaime Prilusky, Marilyn Safran, Vladimir Soboev (Weizmann Institute of Science), Kim Henrick (EBI), Gert Vriend (EMBL-Heidelberg), Barry Honig (Columbia Univ.), and Axel Brünger (Yale Univ.) have provided invaluable support throughout the years. The PDB Advisory Board and the BNL administration together with the BNL Chemistry and Biology Departments have been an invaluable resource over the years. I wish to express my great appreciation and respect for this team, which has constantly shown enormous initiative and professional capability in all their endeavors.

Deposition of Structure Factors at the Protein Data Bank

Jiangsheng Jiang, Enrique Abola and Joel L. Sussman

The following article appeared in Acta Cryst. (1999) **D55**, and is reprinted with permission.

The Protein Data Bank (PDB) has long made available the experimental data which were used to determine the 3D structures in the database. In recent years more and more depositors and users of the PDB have come to appreciate the importance of reliable access to such fundamental data. The deposition of the experimental data, along with the coordinates is essential for the following reasons:

(1) Rigorous validation of the structure determination results can only be carried out using both atomic parameters and experimental structure factor amplitudes.

(2) Archiving of this data will ensure their preservation and continued accessibility.

Whether or not to require that the experimental data be deposited concomitantly with the structure data has been hotly discussed recently in the scientific press [Baker, Blundell, Vijayan, Dodson, Dodson, Gilliland & Sussman (1996). *Nature (London),* **379**, 202] and on the internet [EBI/MSD Draft Consultative Document for Deposition of Structure Factors, http://croma.ebi.ac.uk/msd/Policy/sf.html].

At present more than 50% of the X-ray diffraction submissions are being deposited with their associated structure factors (see Table 1), compared to 25% four years ago. This increase is probably partly due to the ease of uploading the files *via* our WWW-based submission tool, *AutoDep*, and the fact that this tool is available both in the USA at BNL (PDB deposition site at http://www.pdb.bnl.gov) and in Europe at the EBI(EBI deposition site at http://www2.ebi.ac.uk/pdb). The PDB strongly encourages all researchers to deposit their structure factors at the time of coordinate submission. Furthermore, we actively encourage journals to require their submission as a prerequisite for publication. [Sussman (1996) *Protein Data Bank Quart. Newslett.* No. 75, p. 1, at ftp://pdb.pdb.bnl.gov/newsletter/newsletter96jan/newslttr.txt].

In order to facilitate the use of deposited structure factors, we at the PDB, together with a number of macromolecular crystallographers and the IUCr Working Group on Macromolecular CIF, developed a standard interchange format for structure factors [PDB Structure Factor mmCIF at ftp://pdb.pdb.bnl.gov/pub/pdb/structure_factors/cifSF_dictionary; *Protein Data Bank Quart. Newslett.* No. 74, p. 1 (1995), at ftp://pdb.pdb.bnl.gov/newsletter/newsletter95oct/newslttr.txt]. This standard is the mmCIF format, *i.e.* the IUCr-developed Macromolecular Crystallographic Information File. It was chosen for its simplicity of design and for being clearly self-defining. The format is also easy to expand, as new crystallographic experimental methods or concepts are developed, by simply adding additional tokens. The entire mmCIF crystallographic dictionary (http://ndb.rutgers.edu/NDB/mmcif) has recently been ratified by the IUCr's COMCIFS committee.

The PDB has written a program to quickly and easily convert structure factors, as output by the most frequently used crystallographic programs, into the mmCIF format. This tool, which also converts binary *CCP4* MTZ files, will be accessible through the *AutoDep* program following final testing. MTZ files, which are useful in individual labs, are not appropriate for archival purposes. This is because particular groups arbitrarily attach different labels to the MTZ columns.

During the past year, the PDB has converted virtually all the old structure-factor files to this standard format and is keeping up-to-date on all new submissions. As of November 1998, there ~2 000 structure factor files released in the structure factor mmCIF format (PDB mmCIF structure-factor files can be found at ftp://pdb.pdb.bnl.gov/pub/pdb/structure_factors/CIF_format), with an additional ~1 300 'on-hold' for up to four years according to the IUCr policy (see IUCr deposition policy at http://www.iucr.org/iucr-top/journals/acta/actad_notes.html). The structure factors are also available through the PDB's WWW-based *3DB Browser* (http://www.pdb.bnl.gov/pdb-bin/pdbmain). This can be seen on the browser's atlas page for each structure.

The ready availability of structure-factor files in a standard format has made it possible for any scientist to validate a structure in the PDB *versus* its experimentally observed data. There are now some excellent tools available for this, such as SFCHECK (http://www.iucr.org/iucr-top/comm/ccom/School96/pdf/sw.pdf) and the *Uppsala Electron Density Server* (http://alpha2.bmc.uu.se/valid/density/form1.html). The PDB has also observed that one of the most popular uses for these stored structure factors is for the crystallographer who did the experiment to be able to retrieve his/her own data which have been misplaced in their laboratory.

Table 1
PDB structure factor (SF) submission.

Year	Number of X-RayStructure Submissions	Number of SF Submissions (%)
1994	804	205 (25.0)
1995	963	343 (36.0)
1996	1124	546 (49.0)
1997	1484	932 (62.8)
1998*	1616	868 (53.7)
Total	5991	2894 (48.3)

* As of November 24 1998.

PDB World Wide Web Mirroring System

Dawei Lin, John Spiletic, and Nancy O. Manning

PDB's World Wide Web server is the major tool used to access the three dimensional macromolecular structural information archived at the PDB. Thousands of times a day, scientists, students and other users around the world visit the PDB to browse and access this data. In order to meet the need for rapid access worldwide, a global network of seventeen mirror sites has been established.

The information on PDB's web server changes frequently. New information is generated on a daily basis. Synchronizing the PDB and its mirror sites to provide exactly the same services while requiring minimum human involvement is a necessary but nontrivial task. We developed an automatic web mirroring procedure to solve

this problem. The procedure is based on ftp mirroring technology. It has been used by the mirror sites and PDB for approximately two years.

The development and mirroring procedures are shown in Figure 1. The numbered steps are explained as follows:

1.HTML pages and CGI codes are developed and tested on the development server in the source code control area.

2.The working code and HTML pages are copied to a read-only area, which can be mirrored by test servers.

3.The updated information is mirrored onto an internal test server, which is in a different area than the development area. It has its own directory tree. The internal server is used to test if the relative links and the mirror procedure are working. People are asked to test the web pages and the function of CGI scripts.

4.After everything is tested, the files are copied outside the firewall to an account that is available to the mirror sites.

5.All the mirror sites and the PDB use exactly the same mirroring procedure to update our web servers.

Specific areas on the httpd server are dedicated to PDB web activities. All the HTML pages and CGI scripts are in the /pdb-docs/ and /pdb-bin/ directories, respectively. There are also index files and local configuration files in /PDB-support/. This avoids confusing PDB applications with other applications on the same server, which would complicate the mirror procedure.

Relative links are used in all the HTML pages and the HTML pages generated by the scripts. For example, to create a hyperlink to the 3DB Browser in the file named index.html,

```
<a href="/pdb-bin/pdbmain"> 3DB Browser</a>
```

is used instead of

```
<a href="http://www.pdb.bnl.gov/pdb-bin/
pdbmain"> 3DB Browser</a>.
```

The advantage of relative links is that pages copied to the mirror sites' machines will point to local resources without having to be edited locally. This is one of the key points in automating the web mirror procedure.

To make relative links work properly, the mirror sites maintain a local configuration file. The configuration file reflects the local directory tree and available resources. PDB provides a generic template, and mirror sites modify it according to their set up. This configuration file is excluded from the automatic mirroring procedure to avoid being overwritten by the original template file. Changes to the configuration files are sent to mirrors by e-mail one week in advance, to be included manually.

To avoid duplication and allow easy maintenance of the resources, PDB's web and ftp servers share some files. All mirror sites support both web and ftp servers. When a hyperlink points to a file on the ftp server, a Server Side Include (SSI) script is used in order to access the local ftp server of each mirror site. Its function is to use configuration variables to dynamically generate a path to the local file. A sample perl code is shown below:

```
#!./perl
require "PDB-local.pl";
print "Content-type: text/html", "\n\n";
$id = "<A HREF=\"$PDB'ftpServer/";
print "$id";
```

PDB-local.pl is the configuration file used by mirror sites and the PDB to specify the local directory tree structure and resources. $PDB'ftpServer is the variable that generates the local ftp server name. For instance, at the PDB, $PDB'ftpServer is equal to "ftp://ftp.pdb.bnl.gov". The hyperlink to the Het Group Dictionary file in the actual HTML file is:

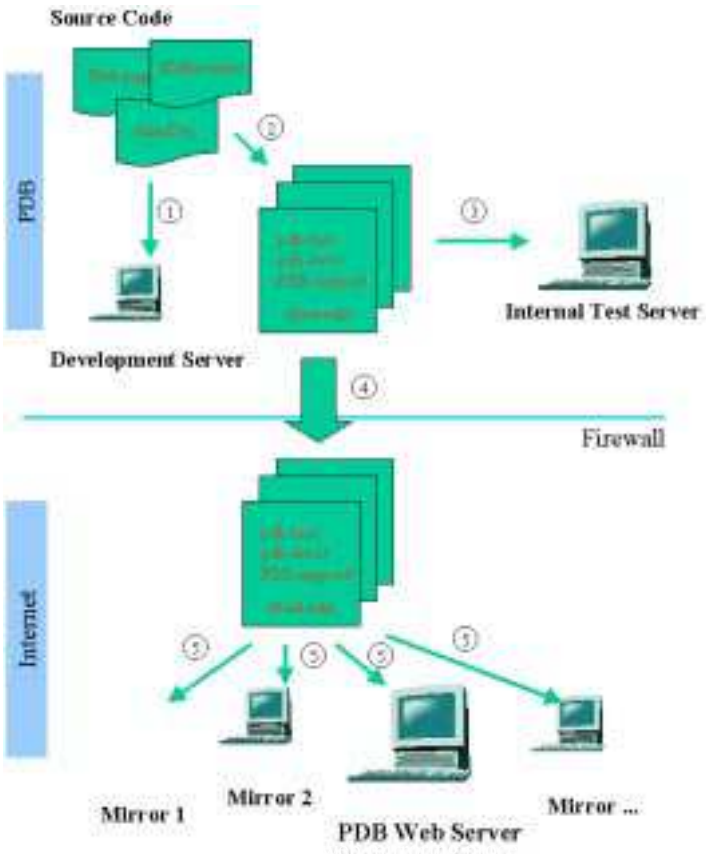
```
<!–#exec cgi="/pdb-bin/pdb_ftp.pl"–>pub/re-
sources/hetgroups/het_dictionary.txt"> Het Group
Dictionary</a>.
```

When a user requests this link, the web server will parse the SSI script pdb_ftp.pl and translate the above link to

```
<a href="ftp://ftp.pdb.bnl.gov/pub/resources/
hetgroups/het_dictionary.txt"> Het Group Dic-
tionary </a>.
```

Clicking on this link returns the file from the PDB ftp server. The same thing happens at each mirror site. The mirror's server substitutes $PDB'ftpServer with its local ftp server name.

HTML pages and CGI scripts are put into a read-only account available to mirror sites. Mirror sites use the ftp mirror tool, mirror.pl, to mirror the updated information from this account. For security reasons, this account is not an anonymous ftp account, but requires a password for access. In addition, this account can only be accessed by ftp. This process can be made as a cron job to fully automate the update procedures. Although the procedure is automatic, an e-mail message is sent to mirror sites for update verification.



Acknowledgement: We would like to thank the EBI and other mirror sites for their suggestions, their support, and their help in making the PDB easily available to our users.

PDB Mirror Sites

Argentina

University of San Luis **pdb.unsl.edu.ar**

Australia

ANGIS - Australian National Genomic Information Service, Sydney **molmod.angis.org.au/pdb/**

The Walter and Eliza Hall Institute of Medical Research, Melbourne **pdb.wehi.edu.au/pdb/**

Brazil

ICB-UFMG, Instituto de Ciencias Biologicas, Universidade Federal de Minas Gerais **www.pdb.ufmg.br**

China

Institute of Physical Chemistry, Peking University, Beijing **www.ipc.pku.edu.cn/npdb/index.html**

France

Institut de Génétique Humaine, Montpellier **pdb.igh.cnrs.fr/**

Germany

GMD, German National Research Center for Information Technology, Sankt Augustin **pdb.gmd.de/**

India

Bioinformatics Centre, University of Pune **202.41.70.33/**

Israel

Weizmann Institute of Science, Rehovot **pdb.weizmann.ac.il/**

Japan

Institute for Protein Research, Osaka University **www2.protein.osaka-u.ac.jp/**

Poland

ICM - Interdisciplinary Centre for Modelling, Warsaw University **pdb.icm.edu.pl/**

Taiwan

National Tsing Hua University, HsinChu **pdb.life.nthu.edu.tw**

United Kingdom

Cambridge Crystallographic Data Centre, Cambridge **pdb.ccdc.cam.ac.uk**

EMBL Outstation, European Bioinformatics Institute, Hinxton **www2.ebi.ac.uk/pdb**

United States

Bio Molecular Engineering Research Center, Boston University, MA **www.pdb.bu.edu**

North Carolina Supercomputing Center, Research Triangle Park, NC **pdb.ncsc.org**

University of Georgia, Athens, Georgia **pdb.bmb.uga.edu**

Proposal: PDB Depositors Club

Morten Kjeldgaard, Institute of Molecular and Structural Biology, Aarhus University, Aarhus, Denmark (mok@imsb.au.dk)

The following letters appeared on several crystallographic discussion groups on November 5 and 6, 1998, and are reprinted with permission of the authors.

Dear Colleague,

As you are probably aware, the Protein Data Bank is moving from Brookhaven National Laboratory to the RCSB which is a consortium of three academic research institutions in the United States.

It is the opinion of many crystallographers worldwide that the PDB at Brookhaven has been improving tremendously the last few years under the leadership of Prof. Joel Sussman. Therefore, the decision to move the database came as a surprise to many crystallographers, especially outside the USA. Many have felt that this is just another case of an arrogant "we-pay-for-it-so-we-can-do-what-we-want" attitude.

However, life goes on, and whatever frustrations one might have over the decision, it has been made and we need to make the best of it.

One thing that is of concern to me, and several other crystallographers I have talked to, is the question of whether the PDB is gradually being taken over by bioinformticists. Although the representation of crystallographers in the RCSB is presently strong through the involvement of the Berman and Gilliland groups, this question is relevant because a major part of the grant proposal (http://rcsb.rutgers.edu/pdb/docs/grant/toc.html) describes various databases that are to be created from the deposited structural models. The deposition process itself, and the maintenance of an archive, is not emphasized very much in the grant proposal.

Problems with Brookhaven PDB

To be honest, we have to admit that there have been problems with the Brookhaven PDB. First of all, the reluctance of the Brookhaven team to modernize the PDB format and to remove oddities like the HETATM cards, the inconsistencies of the files (different versions of the format exists), and other weirdnesses that have caused programmers to age before time. Second, the question (or solution to the question) of the large number of bookkeeping errors that exist in the current database has not been addressed, at least not in public. The "one million errors in protein structures" controversy initiated by a Nature letter (Hooft *et al.* (1996), Nature **381**, 272) was surprisingly co-authored by a prominent Brookhaven PDB coworker. The discussion following (Jones *et al.* (1996), Nature **383**, 18-19) revealed that a large fraction of that million errors were actually errors in the files themselves, not in the structural models.

The deposition of coordinates in the Brookhaven PDB has been vastly facilitated during the last couple of years through the introduction of the "Autodep" system, but many crystallographers have been alienated by the very tight ties that have evolved lately through the checking procedure as implemented in the WhatCheck program. It seems that the Brookhaven PDB deposited it's responsi-

bility for entry checking with a single programmer who has implemented his own home-grown, more-or-less arbitrary and/or empirical checking schemes. The depositor has often been faced with kilobytes of "error" report most of which actually reflect errors or misconceptions in the software and not of the structure. I'd love to see the WhatCheck report on a future 5A ribosome model. Kidding aside, this situation has of course not been satisfactory to the vast majority of crystallographers.

New PDB

Last week at the Cold Spring Harbor Course on Macromolecular Crystallography, the RCSB group leaders gave presentations presenting the New PDB, followed by a critical discussion, where a few crystallographers in the audience aired their frustrations. What are the Americans doing to the PDB? We have all contributed to the database! How will the deposition of models and data be handled in the future? How will the validation be carried out? Are there plans for including important data (refinement dictionaries, for example) in the database? Are there plans for collaboration with international institutions? And so on...

My impression from the discussion was that the New PDB are quite willing to listen to the crystallographic community in building a future service. To the question of international partners, the European Bioinformatics Institute (EBI) was mentioned, but it was hinted that the position of this institution had not yet been clarified and that they might want to initiate an alternative service. The EBI does not really represent the European crystallographers anyway.

PDB Depositors Club

I propose the formation of a "PDB Depositors Club", not only to maintain the interests of the people who *deposit* information in the PDB, but also to act as a sparring partner for the New PDB. I imagine that the club could be a "grassroot movement", first as a discussion forum on the internet and a Web page (volunteers?). Later we could perhaps have mini-workshops and get-togethers at various international crystallography meetings. To get things started, I have established a mailing list where we can have the discussion. To join, send an email to pdb-depositor-request@imsb.au.dk with the word "subscribe" in the subject line. At the time of this writing, the mailing list has one only subscriber (guess who) so you'll have to subscribe if you wish to follow the (I hope) upcoming discussion. Postings to the list should be sent to pdb-depositor@imsb.au.dk. Please wait a few days before submitting anything to the pdb-depositor list, otherwise not many people will see it.

Below, I have detailed my views on a number of topics that I think are relevant to the PDB depositors:

Deposition of structural models
Deposition of diffraction data
Deposition of NMR data
On hold period for release of data
International funding of PDB

Deposition of structural models

The typical misunderstanding by bioinformaticists is the conception that the atomic model is a *representative* of the data. As any crystallographer knows, this is not the case. The atomic model is an *interpretation* of the data, and this is a very important distinction. Many models deposited in the PDB have been built from low-resolution diffraction information, and actually represent much more information than was originally present in the diffraction data. This is a regrettable fact arising because we always choose to represent molecules by the coordinates of the atom centroid and a displacement parameter. One could of course represent each residue as a characteristically shaped "blob", which would be more appropriate when the crystals only exhibit limited resolution. But for the sake of lazy convenience, and because "blob-refinement" programs have not yet been written (and to the benefit of bioinformaticists who know how to write programs that read PDB format files), we choose to build models that contain coordinates for each atom. It is the responsibility of the user of this information to judge how accurate it is. The crystallographic community needs to discuss what level of checking is necessary and relevant, and how it should be carried out.

Deposition of diffraction data

If we want to record the ever-growing body of crystallographic data, it is imperative that we start thinking seriously about the deposition of diffraction data, and all relevant information associated with this. In the old PDB, as well as the New PDB, the overwhelming emphasis concerns the structural models. To be provocative, one might say that the coordinates are completely irrelevant from an archiving point of view, they are merely a convenience to the users. The *real* and *important* information stemming from a diffraction experiment are the structure factors and phases, *including* the derivative data. If we need to redo a structure 50 years from now using new and improved methods, that is the information we need to use. The PDB depositors club would be a good forum to discuss these things, and to come up with proposals for guidelines.

Deposition of NMR data

Not being an NMR spectroscopist, I will leave this for other people to comment.

On hold period for release of data

Recently, *Science* joined the group of scientific journals that require the crystallographers to release the coordinates at the time of publication. The voluntary on-hold period of one year that many researchers in the field have used is not accepted any more. This policy is the result of an intense lobbying effort by many bioinformaticists and a few crystallographers. An Internet poll conducted by *Nature Structural Biology* gave 855 votes in favor of release on publication and 410 against. An additional poll was later conducted, asking whether the voter was actually a coordinate depositor or not. I never saw the result of that poll, but I have no doubt what the result must have been.

In the perfect world, we are all happy, singing, and dancing on the grass, and we would be happy to give away our most important information. However, in the real world, a structure determination often represents a great investment economically, and years of work. It is reasonable, that the crystallographers, if they think that non-disclosure of coordinates of a specific project is important, should have a limited time to make use of those coordinates. We all know the hectic weeks and days before a structure paper is sent off to the journal. There is not much time to discover the interesting features of the structure. In our lab, we have adopted the policy of putting the release of coordinates on-hold for one year, but releasing them to anybody who asks for them. The new policy of the structural biology journals will not result in a speed-up of the release of coordinates, but rather a slow-down in the writing of pa-

pers. Not all journals however, favor the release-on-publication policy, and The Biophysical Journal, for example will *not* adopt the practice. This topic is also important and interesting for a discussion in the Depositors Club.

International funding of PDB

The data in the PDB has been determined by the entire international community of crystallographers and NMR spectroscopists. Therefore, scientists of all nationalities have a natural interest in the functioning and well being of the PDB archive. This resource needs to be secured for the future. It would be a natural development to attempt a full international funding and governing of the data repository. We have to face it: it is great for crystallographers all over the world that the US government has supported the PDB so far, but if we want influence, we need to contribute more than data. These remarks cover the archiving function of the PDB. Creation of *databases* is in my opinion a separate task from *archiving* as this represents a service to Internet users. Anyone with the desire to create a relational database can acquire the archive and get on with it. We need to have this important discussion.

This letter is already too long. I hope you will appreciate it as an introduction to discussion, and that it will be useful to you and your colleagues in establishing your views on the matter. If you have received it more than once, it is because I have submitted it to a few mailing lists, so please accept my apologies.

Sincerely yours,

Morten Kjeldgaard

The EMBL European Bioinformatics Institute, Macromolecular Structure Database group (EBI-MSD), Hinxton, UK (msd@ebi.ac.uk)

Firstly, we welcome any input from the crystallographic community and comments on suggested directions to proceed in. However, a great deal of work has been and is being done on the topics that Morten has put forward for his suggested “PDB Depositors Club”

The EBI-MSD is aware of its responsibility to the macromolecular structure determination community and welcomes input from both producers and consumers of structural data. The EBI-MSD group is developing a deposition system that is based on commercial database and web interface software and although this is at an advanced development stage again input is welcomed.

Morten mentions the relationship between the PDB (RCSB) and the Macromolecular Structure Database group at the European Bioinformatics Institute.

re: whatever impression was gained at the Cold Spring Harbor Course on

EBI-MSD is not part of the US RCSB, but is working in close cooperation with the RCSB. The NSF’s request for proposals to run the PDB explicitly required the winner to cooperate with EBI-MSD.

re: Deposition of structural models

The EBI-MSD group attends the meetings of the EU supported network CT96-0189 : CRITQUAL : Coordinator Wilson (York), Jones (Uppsala), Kaptein (Utrecht), Lamzin (EMBL-HH), Thornton (London), Vriend (EMBL-HD), Wodak (Brussels).

Future validation of submissions will be based upon the conclusions produced by this group - their draft report is due soon and the EBI-MSD will base validation and validation filters upon the report from the CRITQUAL Network.

re: The crystallographic community needs to discuss what level of checking is necessary and relevant, and how it should be carried out.

This is of course true, and a great deal of discussion is already under way. In Europe the EU supported network CT96-0189: CRITQUAL: Coordinator Wilson (York), Jones (Uppsala), Kaptein (Utrecht), Lamzin (EMBL-HH), Thornton (London), Vriend (EMBL-HD), Wodak (Brussels) has initiated discussion at various meetings; in particular, the ECM17 satellite meeting August 1997, aand further discussions made up a major part of the EBI/CCP4 workshop in September this year. The paper published by the network in *J.Mol.Biol.* this year also addresses the question of “what level of checking is necessary”. (Who checks the checkers? Four validation tools applied to eight atomic resolution structures. EU 3-D Validation Network. (1998) *J.Mol.Biol.* **276,** 417-436.)

The depositors club should provide an excellent forum for further discussion, and dissemination of ideas.

re: Deposition of diffraction data

This is already required, but not policed effectively enough (see Ted Baker’s IUCr letter).

The EBI-MSD group has initiated a major change in the submission of crystallographic data that has been given international support from most of the authors of the software used in macromolecular crystallography (see http://www2.ebi.ac.uk/msd/Harvest/report.html).

For example, the authors of CNS have written a deposition macro that writes a harvest file and is now ready in the latest version and includes all structure factor and dictionary information.

Other examples are both the CCP4 and the ESRF beam line software that is currently under development to meet EBI-MSD suggestions for data capture.

re: Deposition of NMR data

The EBI-MSD group have completed a full macromolecular relational database representation that includes details for an NMR macromolecular experiment and are working closely with the RCSB (NDB) and the BMRB. We have contact with the proposed new CCP within the UK for NMR and with the IUPAC initiative to define the tags required to define spectra including NMR.

re: whether the PDB is gradually being taken over by bioinformaticists

This point is not such an evil as indicated. Crystallographers are not necessarily the best judges of how to organise and archive and setup the retrieval of all the information contained within a PDB entry. The creation of a relational database requires domain knowledge but also requires database technology that can cope with a global view of all the entries. Crystallographers are usually not interested in data base organisation, per se. However with the increasing number of structures available some hierarchy has to be set up to allow efficient retrieval and usage. It is important to have consistency in atom naming, description of biological units, etc. The

EBI-MSD group has experience in Crystallography, NMR structure determination and software development. The group has access to a strong database development team to integrate 3D structure data into all the database development carried out both at the EBI and for EBI partnerships throughout the world.

re: bookkeeping errors that exist in the current database have not been

The EBI-MSD group in collaboration with the PDB and now with the RCSB and other groups are undertaking a major cleanup of all the PDB entries to create a set of files that are globally consistent and internally consistent and will be in a single format. Enormous progress has been made for this undertaking and the result will be ready as a complete set before the PDB shuts down at BNL. This cleaned up version of the PDB files will remove perhaps all of the errors from the existing PDB files with the exception of the few coordinate errors.

re: the EBI does not really represent the European crystallographers

The EBI-MSD does not formally represent structural research groups in Europe, it does however have close contact with European crystallography through CCP4 and the EMBL has set up a senior advisory panel of European scientists to work with the EBI-MSD group. The EBI-MSD group is in part funded by EU money, and has the relevant skills for helping to update the PDB. It is not clear who does represent the European community; there are European representatives on the PDB Advisory board (Keith Wilson currently) and the PDB is advised by the IUCr on crystallographic questions. Ted Baker, the current President of the IUCr also sits on the advisory board. Through the ECM and local crystallographic associations it is possible to have considerable influence on both the PDB and the Journals.

Please send comments to msd@ebi.ac.uk.

The EBI-MSD Macromolecular Structure Database Group: Kim Henrick, Peter Keller, John Irwin, John Ionides, Geoff Barton.

Validation of Sugars in the PDB

Tirso Pons, Daan van Aalten, Gert Vriend, European Molecular Biology Laboratory, Heidelberg, Germany (Gert.Vriend@embl-heidelberg.de)

After the release of dictionaries that allow for the refinement of sugar groups around 1990, the number of sugars and sugar-like residues in the PDB shows an increase as function of time (see Fig. 1 for a plot of the number of sugar residues deposited in the PDB as function of the year). Unfortunately, a certain fraction of these sugars are deposited with fancy names that bare hardly any resemblance with their chemical nature. There are even a few sugars deposited with the name of another sugar. At present the number of sugar residues in the PDB is still small (approximately 4299), so we should think *now* about the deposition of sugars in the PDB, because once this number is 10 times bigger, we will never find anybody crazy enough to go back to all old PDB files and modify them. The authors of this note are not sugar chemists, so don't expect any solutions from us; we merely describe the problem.

From a validator's point of view, sugars are much more complex than amino acids. Every atom in amino acids has a fixed chirality, but in sugars about every carbon is chiral leading to a plethora of diastereoisomers. Sugars can occur as chair or boat and a whole series of conformations in-between. Worst of all, they can be linear and circular, and the circular form sometimes isn't even unique [*e.g.* 1]. Additional problems are created by the fact that two sugars can use more or less every pair of OH groups to form a glycosidic bond. In proteins we have, rather logically, decided that N-Ca-C=O forms the backbone of one residue. Sugar residues, however, link up in an almost symmetrical manner. Two OH groups together split off one water and the two sugar rings are connected with one oxygen between them. Without detailed knowledge about the underlying chemistry it is not possible to decide which residue this oxygen belongs to. We looked at all 34 PDB files that contain at least two linked glucose units that were called GLC, and counted how often the bridging oxygen administratively belongs to the previous unit, and how often to the next one. The results (109 times to the previous and 51 times to the next sugar residue) indicate that the depositors have not treated this aspect of the deposition randomly, but a much higher consistency nevertheless seems desirable. The last problem we want to address here is that sugars are sometimes deposited backwards. For proteins the rule is that the N-terminal residue comes first, the residue it gave its oxygen to in the di-peptide formation process becomes the second residue, etc. Similar rules exist for nucleic acids. For proteins and nucleic acids these rules followed rather naturally from our understanding of the biosynthesis. Surely, if protein synthesis had started at the C-terminal end, all proteins would have been deposited in the PDB with their sequence order inverted. The authors of this article do not know much about sugar synthesis, but think that it is time to discuss the topic of the order in which sugars ought to be deposited. The 1–4 sugar linkage is the most common in the PDB. We found a few cases where glucose chains are deposited in a 4–1 direction. We do not know if this inverted sugar order expresses a real chemical difference, *i.e.,* if the biosynthesis took place in the direction as indicated in the PDB file. Fact is that we found glucoses linked up in two different directions with all administrative parameters identical; a true validator's nightmare.

At present, most molecular graphics programs will read sugars as a series of connected clumps of atoms. The last decade has seen an increase in the number of articles describing all kinds of aspects of protein structures. These articles are normally based on the study of a large series of PDB files. We have so far seen only a relatively small number of studies about aspects of sugars and protein – sugar interactions [*e.g.,* 2–5]. It seems likely, however, that the number of such articles will grow with the number of PDB depositions that contain sugar residues. It seems equally likely that the authors of those articles would be greatly helped if they could actually read the PDB files into a program that deals with sugars in a structured manner.

We have written a program that, given the coordinates of a small molecule, returns a string that encodes the atom types, bonds, bond types, chiralities and ring closures of that molecule in a unique character string (a so-called MOLDES). The MOLDES for glucose is given in Fig. 2. A full explanation of MOLDES strings is beyond the scope of this article, and has been published before [6]. A WWW based server that converts atomic coordinates into MOLDES strings is available (http://swift.embl-heidelberg.de/prodrg_serv/). These strings are much like smiles strings [7], but better computer readable, albeit much less human readable. The advantage of this program is that the input atoms do not need to have the correct names,

as long as the names of all atoms start with the Medeleev symbol. The program also does not care what name the depositor has given to the residue. It does matter, though, that the bond lengths and bond angles agree with the hybridization of the atoms. We have made a library of seven MOLDES strings. Using these, we can correctly detect about 44% of all sugars in the PDB. We intend make a library of MOLDES strings that covers all sugars deposited in the PDB (about one hundred strings would suffice to detect more than 98% of all sugars in the PDB). This would, however, be a lot of work, and it would be nice if some committee consisting of sugar chemists, crystallographers and NMR spectroscopists and the PDB staff could sit together and derive a set of guidelines for the conversion of the IUPAC rules for carbohydrate nomenclature [8] to the more practical PDB entries. The problems mentioned above should definitely be addressed if we ever want to be able to validate sugars that are deposited in the PDB (See Fig. 3).

1) Drew K.N., Zajicek J., Bondo G., Bose B., Serriani A.S., ^{13}C-Labeled aldopentoses: detection and quantitation of cyclic and acyclic forms by heteronuclear 1D and 2D NMR spectroscopy. Carbohydr. Res. 1998, in press.

2) Perez, S., Kouwijzer M., Mazeau K., Engelsen S.B., Modeling polysaccharides: Present status and challenges. J. Mol. Graphics **14**: 307-321, 1996.

3) Quiocho F.A., Carbohydrate-binding proteins: tertiary structures and protein-sugar interactions. Ann. Rev. Biochem. **55**: 287-315, 1986.

4) Vyas N.K., Atomic features of protein-carbohydrate interactions. Curr. Opin. Struct. Biol. **1**: 732-740, 1991.

5) Elgavish S., Shaanan B., Lectin-carbohydrate interactions: different folds, common recognition principles. Trends Biochem. Sci. **22**: 462-467, 1997.

6) Van Aalten, D.M.A., Bywater, R., Findlay, J.B., Hendlich, M., Hooft, R.W.W., Vriend, G., PRODRG a program for generating molecular topologies and unique molecular descriptors from coordinates of small molecules. J.Comp.Aid.Mol.Des. **10**: 255-262, 1996.

7) Weininger, D., Smiles, a chemical language and information system. J.Chem.Inf. Comput.Sci., **28**: 31-36, 1988.

8) McNaught, A.D., Nomenclature of carbohydrates. Adv. in Carb. Chem. and Biochem. **52**: 43-177, 1997.

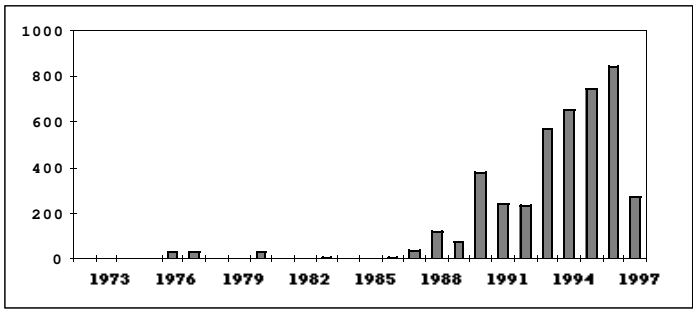


Figure 1. Plot of the number of sugar residues deposited in the PDB as a function of year.

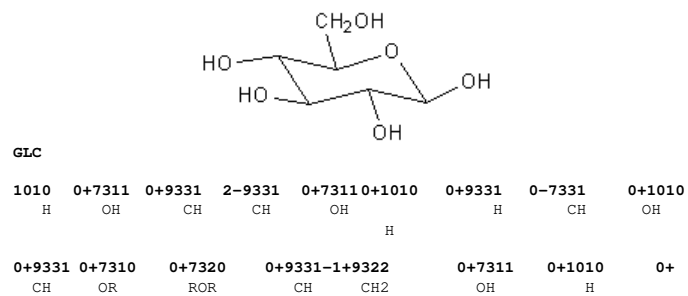
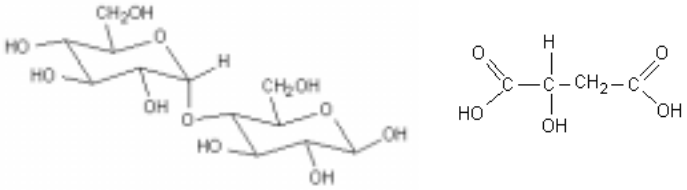


Figure 2. The MOLDES string for glucose

a) Compound name: **MAL**



The same 3 letter code has been used for maltose (left) in PDB entries 1cdg and 1cxe, for D-malate (right) in 2scs and 4scs, for L-malate in 1scs and 3scs and for malonate ($C_3H_4O_4$) in 1at1 and 2at1.

b) Glucose (left) and mannose (right) are C2 epimers. Mannose is called GLC in (for example) 1dog, 1gah and 1gai.

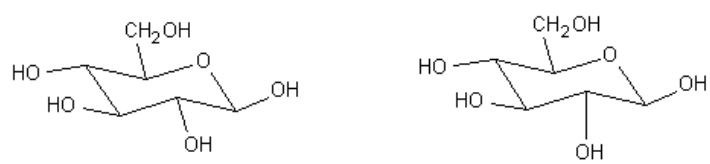


Figure 3.Two nomenclature problems.

Notes of a Protein Crystallographer – A Crystal in Time

Cele Abad-Zapatero

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

Except for the time that it takes to solve our crystal structures, the variable 'time' does not play an important role in the professional life of protein crystallographers. The variables upon which we concentrate all of our efforts are the spatial coordinates (x,y,z), either within our electron density maps or as the triads placing in space the atoms of our chemical models. As many other people have argued before, our science and our results are static. Only the temperature factors associated with atoms or groups of atoms in the crystal give a glimpse of the incessant motion of our atomic universe.

During my postdoctoral years at Purdue University, I was fortunate to meet and become friends of a very special person whose main

interest was time and its relation to the study of biological clocks. At that time, Arthur T. Winfree was a well known and respected figure in the field and had just published a major book entitled 'The Geometry of Biological Time' (1). The monograph was an amazing compendium of observations and mathematical models of what was known at the time about biological clocks. We used to have lunch together at some of the local eateries in West Lafayette and these social encounters were my first introduction to the fascinating world of the circadian regularities in living organisms. Unfortunately for me, he left Purdue University for warmer climates soon after my arrival, and as a good-bye present gave me a copy of his monograph with the following dedication

"For CAZ, crystallographer of space

From ATW, crystallographer of time"

I was very intrigued by those few words and inspired by Arthur's personality and approach to science. Dr. Winfree went on to gain recognition for his iconoclastic and imaginative research in cardiac arrhythmias and was awarded a well deserved McArthur fellowship. During the ensuing postdoctoral years at Purdue and during my non-existent spare time, I read some sections of the book and tried to grasp the fundamentals of the field. Naturally I failed, but a few years later I rediscovered the theme in a simpler, more descriptive and artistic, version of the original monograph entitled 'The Timing of Biological Clocks' (2). In its new reincarnation, the universe of circadian rhythms swallowed me for about two or three months and even though I have certainly not mastered the field, during my readings I discovered a fundamental theme that percolates through the biological clocks of many living systems. This fundamental observation was baptized by Arthur T. Winfree as the 'Time Crystal' and was first discovered in a species of the fruit fly *Drosophila pseudoobscura*. In addition to the word crystal, there is also an anecdotal and historical connection to protein crystallographers. The data showing the first *time crystal* were plotted on perspex sheets in Cambridge, England, in the same workshop where the pioneers of protein crystallography built stacks of electron density maps to visualize the early protein structures (Fig. 1, left).

I am by no means an expert in the field of biological clocks but I'll try to introduce the basic concept of the *time crystal* to our community for two reasons. First, as a small homage to another outstanding scientist and friend of mine in a field different from ours. Second, as an inspiration to the new generations of protein crystallographers. Nowadays, when our trade has become so streamlined, some of the old timers might even say, almost effortless; when new structures are solved and refined at an ever faster and alarming rate; when perhaps the new generations are wondering why did they get into protein crystallography in the first place. Now, I would like to point out to them that they should look for inspiration in solving problems related to the interface between our static structures and the quintessential dynamic process of life. How do biological clocks work at the molecular level? What is the structure of the essential molecular components? How do the physico-chemical properties of the microscopic cellular milieu produce this circadian dance in so many living systems: from the rhythmical glow of *Gonyaulax* cells, to the eclosion of a population of eggs in *Drosophila*, and to the collective rhythm of the flowers of the Kalanchoë plant?

The existence of an internal clock in many different biological systems with an approximate period of 24 hours (circadian) has been well established (see for instance the two books mentioned). Of interest for the discussion is the fact that within the pupal case of the fruit fly *Drosophila* (rice-like structures where larvae await their flight to adulthood), the brain of the larvae keeps time and dictates the exact moment of eclosion of each particular individual. In this state, the motionless pupa is a self-contained system which does not exchange any food or excreta with its surrounding environment. In nature and in a laboratory that is exposed to 24-hour cycle of equal days and nights, the emergence of the individual, or eclosion, occurs in the first hours of daylight. Typically, the adult individuals emerge from their pupal cases in bunches or bursts, the timing of which is a reflection of the internal clocks.

Even though the pupal cases do not exchange matter with the surrounding environment they are subject to external stimuli, especially light. Rearing the larvae under constant light suppresses the ticking clocks and it turns out that these clocks are blind to red or even yellow light, but are extremely sensitive to blue light. The eclosion of an entire population of pupae can be put in synchrony experimentally by collecting them in bright fluorescent lights and them put in red or yellow environment. However, even a brief exposure to a perturbing light penetrating their eternal darkness offsets the timing of all the subsequent bursts, as though the incoming photons had reset the original phase of the internal clock from its old value to a new phase. This new phase depends also on the intensity of the perturbing light.

I apologize for the lengthy preparation but I could not explain the time crystal existing in biological clocks without introducing the identity and meaning of the three axes: x-axis horizontal, old phase of the circadian clock (hours); y-axis, vertical new phase (hours), and z-axis (into the page), stimulus duration in seconds. When Dr. Winfree plotted in perspex sheets the summary of several hundreds of experiments of perturbed eclosion events in *Drosophila* larvae he found a repeated pattern that he labeled the time crystal (Fig. 1). The three-dimensional plot showed how the new phase induced by the perturbation was dependent on the pre-existing – old – phase and on the intensity of the external stimulus; it displayed a 2_1 screw axis in the singularity point where the switch between odd and even resetting takes places (Fig. 2). Time, space and the limitations of my own knowledge prevent me from discussing the subtleties of this pattern that has been found in many other circadian clocks when the old phase is reset by different external stimuli to a new phase. I do encourage the reader to read some of the details in the books that I have introduced.

Thus, crystalline symmetry is found not only in the geometrical patterns that we are so accustomed to in our everyday experience. It has also been unveiled in the internal works of dynamical processes which are essential to living systems. If I were to be a young macromolecular crystallographer again, it is within this domain that I would look for new scientific puzzles. It is the causal connection between our static structures and the rhythms of life that intrigues me. Well, perhaps the new generations will move one step further in this direction, now that our friends the molecular and cell biologists have cut a trail towards some of the proteins responsible for these circadian clocks (3).

References:

1. The Geometry of Biological Time. (1980). A. T. Winfree. Springer-Verlag. Biomathematics Monograph no. 8. New York, Heidelberg, Berlin.

2. The Timing of Biological Clocks. (1987). A.T. Winfree. Scientific American Library. Distributed by W. H. Freeman and Company. New York, Oxford, England.

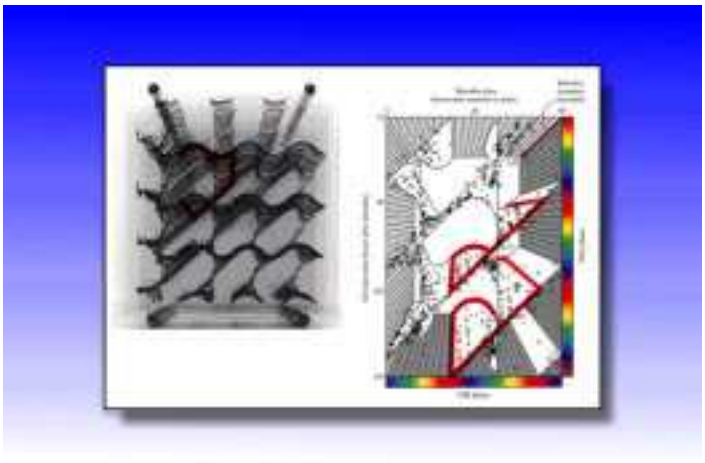


Fig.1 (left)

The time crystal of Drosophila pseudoobscura *spanning almost four cycles in each direction. These data were first plotted in three dimensions in 1971 at the Medical Research Council's Laboratory of Molecular Biology. This is a photo of the original perspex model: old phase increases to the right, stimulus duration increases from background to foreground, and the new-phase measurements increase upward.*

(right)

Computer-plotted version of the image on the left. Six unit cells are represented from the upper left corner with the mathematically defined ideal curves. Data points represent emergence peaks. Here, old phase and new phase are represented as before but stimulus duration increases into the background. The helical edge of a pair of unit cells of the crystal is outlined in color.

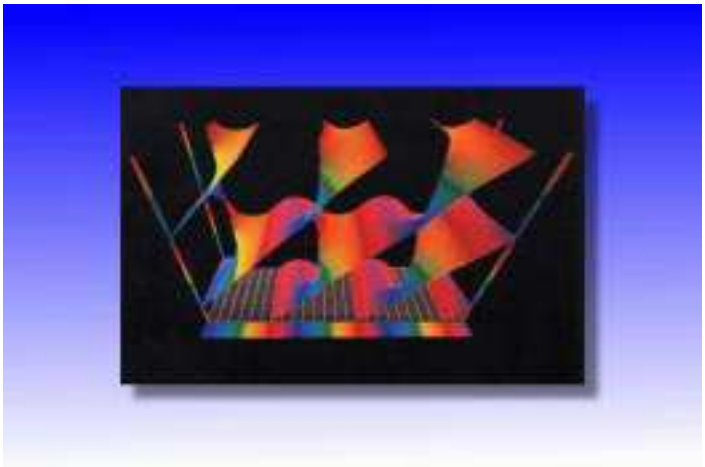


Fig. 2

This three-dimensional graph summarizes thousands of experimental measurements of the eclosion times friut flies after an exposure to a stimulus that doubles seven times from foreground to background.

Figures reprinted and adapted from Reference (2) pages 90 and 82 respectively. Figures reprinted with permission from. W.H. Freeman and Co.

Affiliated Centers and Mirror Sites

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

ARGENTINA

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

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

AUSTRALIA

ANGIS
The Australian National Genomic Information Service
University of Sydney
Sydney, Australia
Shoba Ranganathan (61-2-9351-3921)
shoba@angis.org.au
http://www.angis.org.au

PDB Mirror Site:
http://molmod.angis.org.au/pdb/
Shoba Ranganathan (mmnclp@angis.org.au)

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

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

BRAZIL

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

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

CANADA

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

CHINA

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

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

FINLAND

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

FRANCE

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

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

PDB Mirror Site:
http://pdb.igh.cnrs.fr/
Denis Pugnére (pdbhelp@igh.cnrs.fr)

GERMANY

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

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

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

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

MPI
Max Planck Institute for Biochemie Computer Center
Martinsried, Germany
Wolfgang Steigemann (49-89-8578-2723)
steigemann@biochem.mpg.de
http://www.biochem.mpg.de

INDIA

PUNE
Bioinformatics Center University of Pune
Pune, India
A. S. Kolaskar (0212-355039-350195)
Kolaskar@bioinfo.ernet.in
http://bioinfo.ernet.in

PDB Mirror Site:
http://202.41.70.33/
A.S. Kolaskar (kolaskar@bioinfo.ernet.in)
Sunita Jagtap (sunita@bioinfo.ernet.in)

ISRAEL

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

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

ITALY

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

JAPAN

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

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

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

PDB Mirror Site:
http://www2.protein.osaka-u.ac.jp/
Masami Kusunoki
(kusunoki@protein.osaka-u.ac.jp)

THE NETHERLANDS

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

POLAND

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

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

SINGAPORE

BIOINFORMATICS CENTRE
National University of Singapore
Singapore - 119074
Tan Tin Wee (65-774-7149)
tinwee@bic.nus.edu.sg

TAIWAN

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

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

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

UNITED KINGDOM

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

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

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

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

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

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

UNITED STATES

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

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

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

PDB Mirror Site:
http://www.pdb.bu.edu/
Esther Epstein (esther@darwin.bu.edu)

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

*MAG
Molecular Applications Group
Palo Alto, California, USA
Margaret Radebold (650-846-3575)
bold@mag.com
http://www.mag.com

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

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

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

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

PDB Mirror Site:
http://pdb.ncsc.org/
Linda Spaminato (info@ncsc.org)

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

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

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

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

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

Access to the BNL PDB

Main Telephone ..... +1-516-344-3629
Help Desk Telephone ..... +1-516-344-6356
Fax ..... +1-516-344-5751
Help Desk ..... pdbhelp@bnl.gov
General Correspondence ..... pdb@bnl.gov
WWW Home Page ..... http://www.pdb.bnl.gov
FTP Server ..... ftp.pdb.bnl.gov
Entry Error Reports ..... errata@pdb.pdb.bnl.gov
Order Information ..... orders@pdb.pdb.bnl.gov

FTP Directory Structure for Entries

The PDB FTP server is updated weekly. Files are available by anonymous ftp to ftp.pdb.bnl.gov and on the Web at http://www.pdb.bnl.gov.

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

all_entries/
coordinate entry files in compressed and uncompressed format

biological_units/
generated coordinates for the biomolecules

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

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

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

layer1
layer 1 entries in compressed and uncompressed format

layer2
layer 2 entries in compressed and uncompressed format

ndb
entries received from NDB in compressed and uncompressed format

newly_released/
entries released since the last CD-ROM

nmr_restraints/
compressed NMR restraint files

obsolete_entries/
withdrawn and/or replaced entries

reports
all report files

structure_factors/
compressed structure factor files

current_release, fullrelease, layer1, layer2, ndb, and newly_released are divided into multiple subdirectories

Scientific Consultants

John P. Rose, *University of Georgia, Athens, Georgia, USA*
Mia Raves, *Utrecht University, The Netherlands*
Barry Honig, *Columbia University, New York City, NY*
Goran Neshich, *Embrapa/Cenargen and BBNet/BBRC, Brazilia, Brazil*
Gert Vriend, *European Molecular Biology Laboratory, Heidelberg, Germany*
Manfred Hendlich, *University of Marburg, Germany*
Jiri Vondrasek, *Czech Academy of Sciences, Prague, Czech Republic*
Alexander Wlodawer, *NCI-FCRDC, Frederick, MD*
Eric Martz, *University of Massachusetts, Amherst, MA*
Peter Murray-Rust, *University of Nottingham, Nottingham, UK*
Eldon L.Ulrich, *University of Wisconsin - Madison, Wisconsin, USA*

Clifford Felder
Kurt Giles
Harry M. Greenblatt
Jaime Prilusky
Marilyn Safran
Vladimir Sobolev

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*

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
Regina K. Shea
Janet L. Sikora
Lu Sun
S. Swaminathan
Dejun Xue

Statement of Support

The PDB is supported by a combination of Federal Government Agency funds (work supported by the U.S. National Science Foundation; the U.S. Public Health Service,National Institutes of Health, National Center for Research Resources, National Institute of General Medical Sciences, and National Library of Medicine; and the U.S. Department of Energy under contract DE-AC02-98CH10886) and user fees.

Protein Data Bank
pdb™
Brookhaven National Laboratory

Biology Department, Bldg. 463
P.O. Box 5000
Upton, NY 11973-5000 USA