SCREENING FOR CRYSTALLIZATION CONDITIONS AND ROBOTICS

Acta Cryst. (1994). D**50**, 408-413

Biological Macromolecule Crystallization Database, Version 3.0: New Features, Data and the NASA Archive for Protein Crystal Growth Data

BY GARY L. GILLILAND AND MICHAEL TUNG

Center for Advanced Research in Biotechnology of the Maryland Biotechnology Institute and National Institute of Standards and Technology, 9600 Gudelsky Drive, Rockville, MD 20850, USA

DOROTHY M. BLAKESLEE

Standard Reference Data Program, National Institute of Standards and Technology, Gaithersburg, MD 20899, USA

AND JANE E. LADNER

Center for Advanced Research in Biotechnology of the Maryland Biotechnology Institute and National Institute of Standards and Technology, 9600 Gudelsky Drive, Rockville, MD 20850, USA

(*Received* 15 *November* 1993; *accepted* 14 *February* 1994)

Abstract

Version 3.0 of the NIST/NASA/CARB Biological Macromolecule Crystallization Database (BMCD) includes crystal and crystallization data on all forms of biological macromolecules which have produced crystals suitable for X-ray diffraction studies. The data include summary information on each of the macromolecules, crystal data, crystallization conditions and comments about the crystallization procedure if it varies from the traditional methods employed for crystal growth. The database-management software maintains continuity with previous versions providing similar search procedures and displays. Version 3.0 of the BMCD includes protocols and results of crystallization experiments undertaken in space. These new data are comprised of both the NASA Protein Crystal Growth Archive, which includes information on all NASA-sponsored protein crystal growth experiments, and data describing other internationally sponsored microgravity macromolecule crystallization studies. The entries for the space growth crystallization experiments contain the crystallization protocols, apparatus descriptions, flight summary data, indication of success or failure of the experiments, references, *etc*. Other new features of the BMCD include the addition of crystallization procedures for small peptides and cross references to other structural biology databases.

Introduction

The NIST/NASA/CARB Biological Macromolecule Crystallization Database (BMCD)* contains crystal data and the crystallization conditions which have been compiled from the literature (Gilliland, 1988; Gilliland & Bickham, 1990). The earlier and current versions of the BMCD include entries for all classes of biological macromolecules for which diffraction quality crystals have been obtained. These include proteins, protein-protein complexes, nucleic acids, nucleic acid-nucleic acid complexes, protein-nucleic acid complexes and viruses.

The primary motivation for creating the BMCD was to develop crystallization strategies (Gilliland & Davies, 1984; Gilliland, 1988; Gilliland & Bickham, 1990). Even now, after more than 50 years of experience in the production of diffraction quality crystals, it is still true that each new biological macromolecule requires the development of its own crystallization protocol. Although over the years a number of strategy suggestions have been put forth (McPherson, 1976; Blundell & Johnson, 1976; Carter & Carter, 1979; McPherson, 1982; Gilliland & Davies, 1984; Gilliland, 1988; Gilliland & Bickham, 1990), there is still no 'universally' accepted strategy.

The BMCD may be used for developing a variety of strategies (Gilliland & Bickham, 1990). These include strategies for (1) the crystallization of a previously crystallized biological macromolecule; (2) crystallization of a modified or mutant biological macromolecule for which the unmodified or wild-type biological macromolecule has been reported; (3) crystallization of a biological macromolecule that is homologous to previously crystallized macromolecule(s); and (4) *de novo* crystallization of a biological macromolecule that was not previously crystallized.

The development of a productive crystallization strategy will yield crystals, but the quality of X-ray diffrac-

* Abbreviations: BMCD, Biological Macromolecule Crystallization Database; PCG, protein crystal growth.

tion limits the information which can be obtained from the ensuing crystallographic studies. Any improvement in the diffraction quality of the crystals will, therefore, enhance the productivity of the structural investigation. Recently, a number of crystallization experiments undertaken in space indicate microgravity has a positive influence on the crystal growth process, *i.e.* an improvement in the diffraction quality of crystals of biological macromolecules (DeLucas *et al.*, 1991). Many additional crystallization experiments will be required to ascertain the reasons for the observed differences in the diffraction resolution observed between earth- and space-grown crystals.

The BMCD, in addition to including crystallization data reported in the literature, has expanded its role and now contains the NASA Protein Crystal Growth Archive. This includes the crystallization data generated from studies carried out in a microgravity environment supported by NASA. Data from other crystallization experiments carried out under microgravity sponsored by other international space agencies are also included. This paper describes the current data structure of the BMCD along with new software features which improve the utility of the system. The current hardware and software requirements for implementing and accessing the database are also reported.

BMCD data

The data contained in the BMCD can be divided into three major categories - data concerning the macromolecules, data concerning the crystals and summary information. Macromolecule data are included for biological macromolecules for which crystals suitable for diffraction studies have been obtained. An entry is considered a separate macromolecule if it has a unique amino-acid or nucleic acid sequence. For example, a site-directed mutant protein with a single amino-acid change from the wild-type protein is considered a separate macromolecule entry. Analogously, a crystal entry extracted from the literature must have unique unit-cell constants; crystal entries for crystal forms with cell constants which are nearly isomorphous with previously reported crystals are included if the author describes significant differences in the intensity distribution of the diffraction pattern from that previously reported. Macromolecule, crystal entries are each assigned a four-character alphanumeric identifier beginning with M and C for macromolecule and crystal, respectively.

The NASA PCG Archive data mentioned below contains entries patterned after the crystal entry for each experiment carried out in microgravity. Thus, a single crystal form of a macromolecule may have multiple entries, one for each microgravity experiment. Each of these crystallization entries is also given a four-character alphanumeric identifier beginning with C.

Macromolecule entry

The data in a macromolecule entry include, as previously described (Gilliland & Bickham, 1990), the preferred name of the macromolecule and other aliases. They also include biological source information which consists of the common name, genus/species, tissue, cell, and organelle from which the macromolecule was isolated. Attempts have also been made to include this information for recombinant proteins expressed in a foreign host. The subunit composition and molecular weight are also included. This information consists of the total number of subunits, the number of each type of distinct subunit, the total molecular weight and the molecule weight for each type of individual subunit. (A subunit of a biological macromolecule entity is defined as a part of the assembly which is associated with another part by non-covalent interactions. For example, the two oligomeric nucleic acid strands of a double-stranded nucleic acid fragment are considered as two subunits.) A typical macromolecule entry is illustrated in Fig. 1.

Crystal entry

The data in each crystal entry include the crystallization procedure, the crystal data, crystal morphology, and complete references. The crystallization procedure includes the crystallization method, the macromolecule concentration, the temperature, the pH, the chemical additives in the growth medium and the length of time required to produce crystals of a size suitable for diffraction experiments. If the crystallization deviates from standard protocols, a description of the procedure is provided. The crystal size and shape are given along with the diffraction quality. If crystal photographs or diffraction pictures are published, the appropriate references are indicated. The crystal data include the unit-cell

Macromolecule Name: **ME#:** M135

immunoglobulin-Fab IgG (Jel 42)

Aliases:

anti-HPr Fab fragment

genus: **tissue:** ascites fluid

species: **cell:**

common name: mouse **organelle:**

Total Molecular Weight: 50000 **Total No. Subunits:** 2

Remarks:

Jel 42 is a monoclonal antibody specific for histidine-containing protein, a small phosphocarrier protein required for sugar transport in Escherichia coli.

Total Number of Crystal Entries: 2

Fig. 1. A representative example of a biological macromolecule entry M135 in the BMCD.

dimensions (a, b, c, α, β, γ), the number of molecules in the unit cell (Z), the space group and the crystal density. A crystal entry for the macromolecule entry illustrated in Fig. 1 is shown in Fig. 2.

Summary information

The summary information provides an easy mechanism for browsing through the data contained in the BMCD. The user can browse through the complete list of macromolecule names. The summary information also tabulates the number of macromolecules and crystal forms for each source, prosthetic group, space group, chemical addition and crystallization method. In addition, the listing of complete references can be obtained for a particular author or for a key word or phrase. The BMCD also provides a listing of general references concerning all aspects of crystal growth. For convenience, the references have been divided into categories which include reviews and books, articles concerning procedures, and references concerning nomenclature. Complete titles are included. A remark is often included to emphasize important aspects of a reference which may not be evident from the title.

Crystal entry CE#: C1N9

a	b	c	alpha	beta	gamma	Z	Space Group	CS
117.48	66.56	67.31	90.00	118.70	90.00	4	P2<1>	Monoclinic

Method(s) used for the crystallization:

batch method

Macromolecule concentration: 10.000 mg/ml

pH: 8.000

Chemical additions to the crystal growth medium:

polyethylene glycol 6000	14.0000	%
sodium chloride	0.2000	M
phosphate buffer	0.0500	M

% solvent in crystals: 47.000 %

Diffraction limit: 3.500 Angstroms

Crystal dimensions in mm: 0.7 0.7 0.5

Reference for the photograph or drawing:

R1WE

Reference's designation of this crystal form:

{100},{011},{01(bar)1}

Comments:

Diffraction data to 3.5 A was collected on a single native crystal over a period of 16 days.

References:

R1WE:

Prasad, L; Vandonselaar, M; Lee, JS; Delbaere, LTJ (1988) J Biol Chem, 263(5), 2571-2574.

"Structure determination of a monoclonal Fab fragment specific for histidine-containing protein of the phosphoenolpyruvate:sugar phosphotransferase system of Escherichia coli."

Fig. 2. A representative example of a crystal entry C1N9 for macromolecule entry M135 in the BMCD. It should be noted that the concentrations of chemical additives use four decimal places to accommodate the range of values of concentrations in the database, not to indicate the precision of the reported data.

NASA PCG archive

The NASA Protein Crystal Growth (PCG) Archive contains a complete description of all of the biological macromolecule crystallization experiments which have been performed on the NASA Space Shuttle missions. Summaries of the results of these experiments are included with the crystallization procedures. Along with the NASA data, the BMCD includes microgravity crystallization experiments sponsored by other space agencies. The data for each microgravity crystallization experiment include the crystallization protocol stored and displayed in the same manner as other BMCD crystal entries and information about the space flight. Additional information includes the mission designation, dates of the flight, duration of microgravity, crystallization apparatus, number of experiments, investigators, general remarks concerning the experiments and complete reference for publications resulting from the studies. An example of a microgravity crystallization entry is shown in Figs. 3(a) and 3(b). This crystal form of this entry is nearly identical to that shown in Fig. 2. The crystal identifiers are different, however, since separate entry identifiers are given to each microgravity experiment.

Also added to the BMCD are new summary displays of information concerning microgravity crystallization experiments. The user can obtain a list of the microgravity missions for which data are included. There are lists of the acronyms and names of sponsors supporting microgravity crystallization experiments and the different apparatuses used in microgravity crystallization experiments sponsored by these agencies. In addition, a list of the missions on which the apparatuses have been flown, along with detailed descriptions of the apparatuses, are also available. A list of all references discussing theory or experimental results of crystallization of biological macromolecules in microgravity is included.

BMCD functions

The BMCD provides a menu-driven interface for querying the database. Searches can include all of the data or can be restricted to that associated with the NASA PCG Archive and other microgravity experiments. Depending upon the type of query, a parameter value (numerical or text) or a range of values (numerical) is requested. The results of searches are displayed at the monitor and can be written to an ASCII file or printed as a report.

The primary mechanism for querying the BMCD is the *Search and Display* option of the main menu

Microgravity data for crystal: C1NA

Flown on STS-31 04/24/90 to 04/29/90

Time at microgravity: 4.00 days

Equipment used on flight:

Vapor Diffusion Apparatus

Number of experiments flown with these crystallization conditions: 5

Investigators: Delbaere, LTJ(leader); Quail, JW; Vandonselaar, M; Prasad, L; DeLucas, LJ; Bugg, CE

Comments on microgravity experiments:

Comparison of intensity data with those from an equal-sized earth-grown crystal, showed microgravity-grown crystal diffracted better in every resolution range (significant to 2.7A, 3A for earth-grown crystal). Space crystal data were better at highest resolution range than data from earth-grown crystal ten times the size in volume.

(*a*)

Crystal entry CE#: C1NA Microgravity Data

a	b	c	alpha	beta	gamma	Z	Space Group	CS
117.48	66.56	67.31	90.00	118.70	90.00	4	P2<1>	Monoclinic

Method(s) used for the crystallization:

vapor diffusion

Macromolecule concentration: 10.000 mg/ml

Temperature: 22.0 degrees C

pH: 8.000

Crystal growth time: 4.50 days

Chemical additions to the crystal growth medium:

polyethylene glycol 8000	16.5000	%
sodium chloride	0.2000	M
phosphate buffer	0.0500	M
sodium azide	0.0200	%

Vm of crystal: 2.300 cubic A/dalton

Diffraction limit: 2.700 Angstroms

Diffraction life time: 24.000 hours

Crystal dimensions in mm: 0.3 0.3 0.3

Reference's designation of this crystal form:

{100}, {011} and {01(bar)1}

Comments:

One crystal was obtained from this experiment.

References:

RG6K:

DeLucas, LJ; Smith, CD; Carter, DC; Twigg, P; He, XM; Snyder, RS; Weber, PC; Schloss, JV; Einspahr, HM; Clancy, LL; McPherson, A; Koszelak, S; Vandonselaar, M; Prasad, L; Quail, JW; Delbaere, LTJ; Bugg, CE (1992) Adv Space Res, 12, 393-400.

"Protein crystal growth aboard the U.S. Space Shuttle Flights STS-31 and STS-32."

RG6L:

Quail, JW; Vandonselaar, M; Prasad, L; Delbaere, LTJ; DeLucas, LJ; Moore, K; Bugg, CE (1992) Proceedings Spacebound'92 (Ottawa, Canada), 88-90.

"Protein Crystallization Aboard Space Shuttle flights STS-31, STS-48 and IML-1, and Aboard the MIR Space Station."

RG6M:

Delbaere, LTJ; Vandonselaar, M; Prasad, L; Quail, JW; Birnbaum, GI; DeLucas, LJ; Moore, K; Bugg, CE (1993) Proceedings of the 31st Aerospace Sciences Meeting (Reno,NV) January 1993, paper AIAA-93-0718.

"Protein Crystallization Aboard the Space Shuttle and the Mir space Station."

(*b*)

Fig. 3. A representative example of the microgravity crystal data for macromolecule entry M135. (*a*) Flight information and (*b*) NASA PCG Archive crystal entry C1NA.

Table 1. *Database search parameters*

Macromolecule	Crystallization conditions
1. Macromolecule name	11. Crystallization method
2. Biological source	12. Macromolecule concentration
3. Molecular weight	13. Crystallization temperature
4. Subunit composition	14. pH of crystallization
5. Prosthetic group	15. Crystal growth time
6. Multiple crystal forms	16. Chemical additions to crystallization solution
Crystal data	Reference
7. Space group	17. Author
8. Unit-cell dimensions	18. Year
9. Z, molecules/unit cell	19. Journal
10. Crystal density	
	Database
	20. Database cross reference

followed by the *Search Database* option of the *Search and Display* menu. From this menu either a single-parameter search or a multiple-parameter search can be selected. These searches can be applied to all of the data or restricted to the NASA PCG Archive. A database-search parameter (Table 1) is then selected and a search is initiated. Search results are stored in a temporary *User File.*

Single-parameter search

Data retrieved using a single-parameter search are stored in a temporary *User File.* These entries are stored as the identifiers of macromolecule and crystal entries which contain data matching the search parameter value(s). Subsequent searches in this mode append additional identifiers of data-matched entries to the file. For example, if an initial search to find all crystal entries grown using ammonium sulfate was followed by a second search for all crystal entries grown using sodium sulfate, all entries found for both searches would be contained in the *User File.* This is equivalent to a query which uses a logical OR; entries included in the *User File* would be for those crystals which had been grown using ammonium sulfate or sodium sulfate.

Multiple-parameter search

Like the single-parameter search mode, the initial parameter search creates a temporary *User File* for the identifiers of macromolecule and crystal entries that match the search parameter value(s). However, in this case, subsequent searches are restricted to entries contained within the temporary *User File.* Entries that do not match the subsequent search parameter value(s) are eliminated from the *User File.* Therefore, entries remaining in the user file match all queries. For example, if an initial search seeks all crystals reported in space group $P1$ and a second search restricts the crystal entries to those that are grown between pH 3.0 and 4.0, the crystal entries remaining in the user file are those that are both grown with a pH between 3.0 and 4.0 and have

space group *P*1. This is equivalent to a query which would implement a logical AND.

For both the single- and multiple-parameter search modes, once the first parameter is selected, the user may query for as many values (numerical or text) of the parameter as necessary. The results from each search are appended to the *User File*. For example, all members of a protein family might require a number of selections from the macromolecule name list. Subsequent parameter selection in the multiple-parameter search will eliminate entries that do not match the search conditions.

Data display and User File management

How the search results are displayed to the user is controlled by the *Display* menu which is accessed from the *Search and Display* menu. The *Display* menu provides a mechanism to view the macromolecule and crystal data for all of the entries which are contained in the *User File*. There are several different summaries which display only selected information from the entries contained in the *User File*. Additionally, a specific macromolecule or crystal entry may be displayed by directly entering the appropriate identifier.

The search results are stored in a temporary *User File* which can be restored, saved or deleted by using the appropriate BMCD utility feature accessed from the *Manage User File* option of the *Search and Display Data* menu. This provides a convenient means for the user to save the results from specific searches. Restoring a *User File* allows the display or printing of the results from previous work sessions.

New data and features of BMCD Version 3.0

The Version 3.0 release of the BMCD contains, besides the NASA PCG Archive, new data from the literature and a number of new features not present in previous releases. There are now more than 2300 crystal forms of over 1500 biological macromolecules included in the database. This represents data reported through mid 1993. One of the goals of the BMCD is to maintain a complete database of all of the crystal forms of macromolecules for which diffraction studies have been reported since the first report by Bernal & Crowfoot (1934). A number of crystal forms of biological macromolecules reported during this time period have not yet been included. The new data also include many crystal entries for peptides which are, in general, more difficult to crystallize than proteins. This new release provides for the first time cross references to two other structural biology databases, the Brookhaven Protein Data Bank (Bernstein *et al.*, 1977) and the Nucleic Acid Database (Berman *et al.*, 1992). Searches for BMCD entries which resulted in three-dimensional structures of biological macromolecules can be made through queries using parameter 20 in Table 1, Database cross reference.

Almost all of the references in Version 3.0 contain titles of the articles or books, whether they are general references or those relating to a specific crystallization protocol. There has also been a restructuring of the reference data to facilitate maintenance and utility of this resource of the BMCD. There are now more than 2900 references included in the BMCD. The references can be searched by author or by any word or phrase that may be contained in the title.

An important functional aspect of the BMCD is that all data summaries and/or search results can be directed into ASCII files. These can subsequently be taken into a word-processing program and examined and edited.

BMCD environment*

The BMCD software requires an IBM-AT† class or PS/2 PC (80286–80486 chip) with an operating system of PC-DOS or MS-DOS 2.1 or greater. The program requires 640 kilobytes of memory, a hard disk with 3.7 megabytes of available storage, and a PC-compatible printer. It has been written and compiled with Clipper‡ in order to run as an independent program. All of the data files are in dBASE III Plus§ format, and all of the index files are in the Clipper indexing file format. The BMCD is supplied with a users' guide (Ladner, Blakeslee & Gilliland, 1993) which describes the installation and features of the software. It also contains a number of complete examples illustrating how to execute various features of the program.

Future BMCD releases

The continuing primary goal of the BMCD is to maintain an error-free, up-to-date, complete crystallization database and NASA PCG Archive. Although the data include information reported in 1993, there are still crystallization reports from previous years that need to be included. We encourage BMCD users to report omissions in the data so that they can be included in future releases. We also encourage users to report errors or omissions in current entries so that corrections can be made. The addresses, telephone numbers and e-mail

---

* Questions concerning the availability and acquisition of a copy of the NIST/NASA/CARB Biological Macromolecule Crystallization Database should be directed to: Ms Joan Sauerwein, National Institute of Standards and Technology, Standard Reference Data Program, Gaithersburg, MD 20899, USA.

† Certain commercial equipment, instruments and materials are identified in this paper in order to specify the experimental procedure. Such identification does not imply recommendation or endorsement by the National Institute of Standards and Technology, nor does it imply that the materials and equipment identified are necessarily the best available for the purpose.

‡ Clipper is a registered trademark of Nantucket Corporation, 12555 West Jefferson Boulevard, Los Angeles, CA 90066, USA.

§ dBASE III Plus is a registered trademark of Borland International, Incorporated.

addresses of contacts who can correct the data or address problems related to the database are listed in Appendix C of the *Users' Guide* (Ladner *et al.*, 1993). The number of crystallization and structure papers in the literature continues to increase at a rate which will very shortly make updating the BMCD difficult. Part of the difficulty stems from the fact that crystallographic studies of a large number of variants (site-directed mutants) of an increasing number of proteins are being undertaken. These structural studies support protein folding, protein engineering and drug design programs. We are considering options to restrict BMCD entries to a parent molecule for a family of isomorphous variants, but to include references for mutants. Mutant protein crystals which are not isomorphous with wild-type crystals would be considered as separate entries. Long-range plans include moving the BMCD to UNIX platforms with provisions for network access to the continuously updated files. This will allow the crystallographic community wider access to the data, encourage users to develop their own analysis tools, and strengthen the communication between the authors and the users which hopefully will result in a more effective and useful resource.

The authors would like to acknowledge the assistance of Ms X. R. Dong of CARB in the acquisition of new literature data included in Version 3.0 of the NIST/NASA/CARB BMCD. The authors would also like to thank Dr J. Flippin-Anderson of the Naval Research Laboratory for her interest in expanding the BMCD to include peptide crystallization data. We would also like to thank Professor H. Berman of Rutgers for her interest in cross referencing the BMCD data with that found in other structural databases and for references on crystallographic reports of nucleic acids. The authors would like to thank the Microgravity and Science Application Division of NASA for their support and encouragement for expanding the BMCD to include the NASA PCG Archive.

References

BERMAN, H. M., OLSON, W. K., BEVERIDGE, D. L., WESTBROOK, J., GELBIN, A., DEMENY, T., HSIEH, S.-H., SRINIVASAN, A. R. & SCHNEIDER, B. (1992). *Biophys. J.* **63**, 751-759.

BERNAL, J. D. & CROWFOOT, D. (1934). *Nature (London)*, **133**, 794.

BERNSTEIN, F. C., KOETZLE, T. F., WILLIAMS, G. J. B., MEYER, E. F. JR, BRICE, M. D., ROGERS, J. R., KENNARD, O., SHIMANOUCHI, T. & TASUMI, M. (1977). *J. Mol. Biol.* **112**, 535-542.

BLUNDELL, T. L. & JOHNSON, L. N. (1976). *Protein Crystallography*. New York: Academic Press.

DELUCAS, L. J., SMITH, C. D., SMITH, W., VIJAY-KUMAR, S., SENADHI, S. E., EALICK, S. E., CARTER, D. C., SNYDER, R. S., WEBER, P. C., SALEMME, F. R., OHLENDORF, D. H., EINSPAHR, H. M., CLANCY, L. L., NAVIA, M. A., MCKEEVER, B. M., NAGABHUSHAN, T. L., NELSON, G., MCPHERSON, A., KOSZELAK, S., TAYLOR, G., STAMMERS, D., POWELL, K., DARBY, G. & BUGG, C. E. (1991). *J. Cryst. Growth*, **110**, 302-311.

CARTER, C. W. JR & CARTER, C. W. (1979). *J. Biol. Chem.* **254**, 12219-12223.

GILLILAND, G. L. (1988). *J. Cryst. Growth*, **90**, 51-59.

GILLILAND, G. L. & BICKHAM, D. (1990). *Methods: A Companion to Methods in Enzymology*, Vol. 1, pp. 6-11. New York: Academic Press.

GILLILAND, G. L. & DAVIES, D. R. (1984). *Methods Enzymol.* **104**, 370-381.

LADNER, J., BLAKESLEE, D. & GILLILAND, G. L. (1993). *NIST/NASA/CARB Biological Macromolecule Crystallization Database*. Version 3.0. *Users' Guide*. US Department of Commerce, National Institute of Standards and Technology, Standard Reference Data Program, Gaithersburg, MD, USA.

MCPHERSON, A. JR (1976). *Methods Biochem. Anal.* **23**, 249-345.

MCPHERSON, A. (1982). *Preparation and Analysis of Protein Crystals*. New York: John Wiley.