

# The Biological Macromolecule Crystallization Database: crystallization procedures and strategies

Gary L. Gilliland,\* Michael Tung  
and Jane E. Ladner

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

Correspondence e-mail: gary.gilliland@nist.gov

The Biological Macromolecule Crystallization Database (BMCD) archives crystallization data from published reports for all forms of biological macromolecules that have produced crystals suitable for X-ray diffraction studies. The information includes the crystallization conditions, crystal data, comments about the crystallization procedure and information on the biological macromolecule or biological macromolecule complex. Crystallization procedures, including fast screens and more general procedures, can be developed effectively using this web-based resource (<http://www.bmcd.nist.gov:8080/bmcd/bmcd.html>).

Received 8 April 2002

Accepted 15 April 2002

## 1. Introduction

Structural biologists initiating crystallographic studies of biological macromolecules or their complexes require the production of single crystals with favorable diffraction properties. This is true for the more traditional hypothesis-driven research as well as the large-scale efforts in structural genomics that are just getting under way (Burley, 2000). The methods and protocols employed for the discovery of crystallization conditions for new biological macromolecules and those that have been previously crystallized rely, for the most part, on the successes of the past. Today, it is routine to begin crystallization trials with commercial kits that screen a broad range of reagent combinations with diverse solution properties. These kits incorporate reagent combinations that worked effectively in the past. If this 'fast-screen' technique (Jancarik & Kim, 1991) does not result in crystals, then a systematic approach is frequently attempted (McPherson, 1976, 1982, 1999; Carter & Carter, 1979). The fast-screen approach is used with biological macromolecules that have never been crystallized and it is common practice to use it with macromolecules that have been crystallized in the past even before trying to reproduce the conditions that were reported in the literature. The success or failure of a crystallographic structure-determination project depends on the outcome of such experiments. Therefore, one might argue that the rapid and successful solution to a crystallization problem would be more likely by inclusion of information from as many other successful efforts as possible. It was with this in mind that the Biological Macromolecule Crystallization Database (BMCD) was created. It contains comprehensive crystallization information for all classes of biological macromolecules for the development of crystallization strategies (Gilliland, 1988; Gilliland & Bickham, 1990).

The first stand-alone version of the NIST/CARB (National Institute of Standards and Technology/Center for Advanced Research in Biotechnology) BMCD was released in 1989 as a

PC database for which both the software and data were distributed. A year later, a second version with a significant increase in data content was released (Gilliland & Bickham, 1990). Recently, the data in the BMCD has been further expanded and the database ported to a UNIX platform to provide a web-accessible resource available at <http://www.bmcd.nist.gov:8080/bmcd/bmcd.html> (Gilliland *et al.*, 1996). The current version of the BMCD includes 3547 crystal entries from 2526 biological macromolecules. Here, we describe the contents of the BMCD and provide examples of how it can be used effectively to assist in the production of crystals for biological macromolecules.

## 2. Crystallization data and references

The BMCD entries contain macromolecule and crystallization data for only those macromolecules that have crystallized in forms suitable for diffraction studies and that have been reported in the literature. The crystal entries contain the crystallization conditions required to reproduce the crystals and the crystal data that identifies each crystal form. It also has other data resources that include links to other web-based databases, general information and references describing crystallization and related techniques.

### 2.1. Biological macromolecule data

Each entry contains information that defines the biological macromolecule, including its name and other aliases and its biological source, both the scientific and common names. The latter is hierarchical in nature and includes the tissue, cell and organelle from which the macromolecule was isolated. Source information is also included for recombinant proteins expressed in foreign hosts. The total number of subunits defines the active biological assembly of the macromolecule present in the crystal lattice. The subunit name, number present and molecular weight are also provided. The total molecular weight of the biological unit and remarks about the macromolecule that may be pertinent to the crystallization complete the description. Biological macromolecule subunits are defined as components of the assembly that associate by non-covalent interactions. For example, mammalian glutathione S-transferases are dimeric, having two tightly associated identical or closely related subunits; for nucleic acids, the two polynucleotide strands of a double-stranded molecule are considered as two subunits. A representative macromolecule entry for subtilisin GX (Gilliland *et al.*, 1987) is shown in Fig. 1.

### 2.2. Crystallization and crystal data

The details of each crystal entry include the crystal data, crystal morphology, the experimental crystallization protocol and complete references. The crystal data include the unit-cell parameters ( $a$ ,  $b$ ,  $c$ ,  $\alpha$ ,  $\beta$ ,  $\gamma$ ), the number of molecules in the unit cell ( $Z$ ), the space group and the crystal density. The crystal size and morphology are given, along with the diffraction quality. If crystal photographs or diffraction

## Macromolecule -- subtilisin GX (M0MH)

**Molecule Name :** subtilisin GX

**Alias Names :**

**Genus:** Bacillus

**Species:** sp. GX6644

**Total Molecular Weight:** 25300 **Total No. Subunits:** 1

**Prosthetic Group :**

**Total Number of Crystal Entries:** 1

1. [Crystal C0ZY: A=38.4, B=70.3, C=73.5, alpha=90, beta=90, gamma=90](#)

### Figure 1

A representative example of a biological macromolecule, subtilisin GX, entry M0MH in the BMCD.

### Crystal Data (C0ZY).

**Molecular Name:** subtilisin GX.

Unit cell dimensions					
A	B	C	alpha	beta	gamma
38.4	70.3	73.5	90	90	90
					Z
					4

**Space group:**

P2<1>2<1>2<1> - Orthorhombic

**Crystal Density:** 0

**Diffraction limit (Angstroms):** 2

**Diffraction life time (Hours):** 0

**Crystal dimensions in mm:** 0.1 x 0.3 x 1

**Crystal habit:** 20d h0l; 24.5d 0kl -R0T7

**Reference for the crystal photograph:** R0T7

### Crystallization Data.

**Method(s) used for the crystallization:**

vapor diffusion on plates or slides

vapor diffusion in hanging drops

Macromolecule concentration range	11.7	0	mg/ml
Temperature range	7	0	° C
pH value	7	0	
Crystal growth time	30	0	Day

**Chemical additions in the reservoir:**

Chemical Additive	Range	Unit
ammonium sulfate	45	0 %
HEPES	0.05	0 M

*Note: In the above tables, 0 indicates that no data is available.*

### References and Comment.

**Comment:**

Crystals suitable for X-ray diffraction studies were produced by macroseeding.

**References:**

- [R0T7: J Biol Chem 262, 4280-4283, 1987](#) By Gilliland *et al.*

### Figure 2

A representative example of a crystal entry C0ZY for the subtilisin GX entry M0MH in the BMCD.

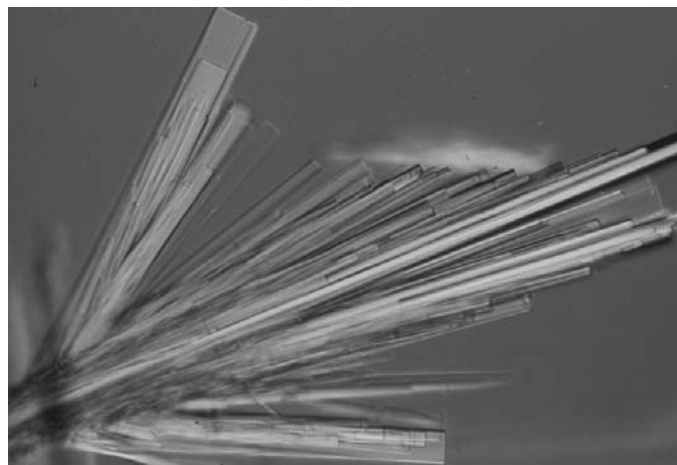
pictures were published, the appropriate references are indicated. The experimental details include the crystallization method, the macromolecule concentration, the temperature, the pH, the chemical additives to the growth medium and the length of time required to produce crystals of a size suitable for diffraction experiments. A description of the procedure is provided if the crystallization protocol deviates from methods that are in general use (McPherson, 1999). Cross-references to two other structural biology databases, the Protein Data Bank (Berman *et al.*, 2000; Berman, Battistuz *et al.*, 2002) and the Nucleic Acid Database (Berman *et al.*, 1992; Berman, Westbrook *et al.*, 2002), are provided if the corresponding entries have been identified. Shown in Fig. 2 is the crystal entry for the macromolecule entry shown in Fig. 1. Crystals of subtilisin GX grown using these conditions are shown in Fig. 3.

### 2.3. Reference and summary information

Access to the summary information provides a mechanism for gaining insight into crystallization methods and experimental results. This information includes a comprehensive list of macromolecule names, tabulations of the number of macromolecules and crystal forms for each source, prosthetic group, space group, chemical addition and crystallization method, and access to the complete reference list. The references can be directly queried for matches with an author or with a phrase or keyword. The BMCD also provides a general reference list of publications dealing with all aspects of crystal growth. These references have been sorted into categories, which include reviews and books, articles concerning procedures and references concerning nomenclature. All references include complete titles and may have remarks added if important aspects of a reference need to be emphasized, especially if the title may not reflect why it was included.

### 3. Crystallization strategies and procedures

Over the last 25 y, a number of systematic crystallization procedures and strategy suggestions for biological macro-



**Figure 3**  
Photomicrograph of the orthorhombic crystal form of subtilisin GX grown from solutions containing the reagents shown in Fig. 2 (Gilliland *et al.*, 1987).

molecules have been put forward (see, for example, McPherson, 1976, 1982, 1999; Blundell & Johnson, 1976; Carter & Carter, 1979; Gilliland & Davies, 1984; Gilliland, 1988; Gilliland & Bickham, 1990; Jancarik & Kim, 1991; Gilliland *et al.*, 1994, 1996). These and other strategies are all based on the successful experiences of crystallographers in the production of suitable crystals for diffraction studies. Most current strategies employ a version of the fast screen first popularized by Jancarik & Kim (1991). Fast screens are sets of experiments that use premixed solutions that have frequently produced crystals. Crystals are often found quickly in such experiments, but if failure occurs then there is a need for a more general approach. Thus, a combination of both fast screening and more general approaches (Gilliland *et al.*, 2001) is used by many laboratories

### 3.1. Reproducing reported procedures

Structural biologists engaged in protein engineering, rational drug design, protein stability and other studies of proteins whose structures have been previously determined often find themselves dealing with macromolecules whose crystal structures have been solved by other laboratories. The BMCD contains the information needed to reproduce the crystallization conditions for many biological macromolecules reported in the literature. The approach described here can be extended to the crystallization of sequence variants, chemically modified derivatives or ligand–biological macromolecule complexes. The reported crystallization conditions of the native macromolecule are the starting points to initiate the crystallization trials. The crystallization of the biological macromolecule may be simple to reproduce, but differences in the isolation and purification procedures, reagents and crystallization methodology of different laboratories can dramatically influence the results. A first attempt using the published conditions often produces crystals, but more often than not the result is poor or no crystals are grown. The crystallization conditions in the database should then be considered as a good starting point for the search or optimization that will require experiments that vary pH, macromolecule and reagent concentrations and temperature, along with the crystallization method.

### 3.2. Fast screens

Within a few short years after the introduction of the fast screen by Jancarik & Kim (1991), almost all attempts to crystallize a protein began with experiments from a screen of one form or another. The basic idea for screening was put forth by Carter & Carter (1979) in their discussion of the use of incomplete factorial experiments to limit the search for crystallization conditions. The advantage of this over a straightforward combinatorial approach is in the reduction of thousands of possible experiments to as few as 50–100. This turns out to be an enormous saving of time, reagents and materials. After the rise in popularity of the original fast screen (Jancarik & Kim, 1991), a number of screens were developed and even commercialized (*e.g.* Cudney *et al.*, 1994).

These early screens were quite general and applicable to a wide range of biological macromolecules and fast screens, based on specific classes of molecules such as RNA, soon developed (Scott *et al.*, 1995). Developing novel screens is ongoing even today.

The BMCD is an ideal tool to facilitate the development of screens for general or specific classes of macromolecules. Earlier reports have provided examples of developing specific screens using BMCD data on Fabs (Gilliland & Bickham, 1990; Gilliland *et al.*, 1997), acid proteases (Gilliland *et al.*, 1996) and endonucleases (Gilliland *et al.*, 2001). To develop a screen requires tabulating all of the crystallization conditions from entries that are related to one another. When this is performed, as in the three cases mentioned above, the data usually indicate that using a small set of reagents over a limited pH, temperature and protein concentration range will have a high probability of finding crystallization conditions for the majority of the macromolecules. Then, a small number of experiments can be derived for the screen using the principles developed by Carter & Carter (1979).

### 3.3. General procedures

A number of general procedures for the crystallization of biological macromolecules that take advantage of the BMCD have been reported (Gilliland, 1988; Gilliland & Bickham, 1990; Gilliland *et al.*, 1994, 1996, 2001). The most recently published procedure for soluble proteins incorporates both the fast-screen approach mentioned above along with a general one (Gilliland *et al.*, 2001). This procedure is illustrated in Fig. 24.4.7.1 of Gilliland *et al.* (2001). Briefly, the purified protein is prepared for the crystallization trials by first concentrating it to 10–25 mg ml<sup>-1</sup> and then dialyzing it into 0.005–0.025 M buffer at a neutral pH or at a pH required to maintain solubility of the biopolymer. Other macromolecule-stabilizing agents such as EDTA and/or dithiothreitol may be included at low concentrations. Once the protein has been prepared, commercial or customized fast screens are carried out using vapor-diffusion experiments. If crystals are obtained, experiments that optimize the crystallization parameters (pH, ionic strength, temperature, *etc.*) are then carried out, or microseeding or macroseeding is employed if optimization proves difficult (McPherson, 1982, 1999).

As shown in the above-mentioned figure, if the fast-screen experiments prove fruitless, a more systematic approach is then undertaken. An analysis of the BMCD data reveals that out of the large number of reagents used as precipitating agents, a small set accounts for the majority of the crystals observed. The pH range for all crystals is quite large, but most proteins crystallize between pH 3.0 and 9.0. Even though temperature can be an important factor, crystallization experiments are usually set up at room (~293 K) or cold room (279 K) temperatures. Protein concentration varies quite markedly, but it appears that investigators typically use >10 mg ml<sup>-1</sup>. Experiments are then set up to explore incrementally these parameters. In parallel, or if the crystallization trials just described are unsuccessful, another set of experi-

ments can be carried out that include the addition of small quantities of ligands, products, substrate, substrate analogs, monovalent or divalent cations, organic reagents *etc.* to the crystallization mixtures. If this does not prove successful, additional reagents may be selected with the aid of the BMCD and new experiments initiated. In addition to the procedure described above, experiments at reduced ionic strength should be carried out. An analysis of the BMCD data reveals that ~10% of the soluble proteins crystallize at low ionic strength (<0.2 M ionic strength).

### 4. BMCD and the future

The BMCD is currently being upgraded to ORACLE<sup>1</sup> to ensure that it will continue to be available for years to come. Along with the change in database software, a major new data release is scheduled for late summer to early fall of 2002. Direct deposition of data by the user community is also being considered. This is especially important with the accelerating pace of crystallography resulting from the structural genomics initiatives (Burley, 2000). The capabilities of the web resource will be expanded to include tools to facilitate the development of crystal strategies for new crystallization problems. The BMCD will also be further integrated with other structural biology web resources to address the structural biology challenges of the future.

### References

- Berman, H. M., Battistuz, T., Bhat, T. N., Bluhm, W. F., Bourne, P. E., Burkhardt, K., Feng, Z., Gilliland, G. L., Iype, L., Jain, S., Fagan, P., Marvin, J., Ravichandran, V., Schneider, B., Thanki, N., Padilla, D., Weissig, H., Westbrook, J. D. & Zardecki, C. (2002). *Acta Cryst.* **D58**, 899–907.
- 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.
- Berman, H. M., Westbrook, J., Feng, Z., Gilliland, G., Bhat, T. N., Weissig, H., Shindyalov, I. N. & Bourne, P. E. (2000). *Nucleic Acids Res.* **28**, 235–242.
- Berman, H. M., Westbrook, J., Feng, Z., Iype, L., Schneider, B. & Zardecki, C. (2002). *Acta Cryst.* **D58**, 889–898.
- Blundell, T. L. & Johnson, L. N. (1976). *Protein Crystallography*. New York: Academic Press.
- Burley, S. (2000). *Nature Struct. Biol.* **7**(Suppl.), 932–934.
- Carter, C. W. Jr & Carter, C. W. (1979). *J. Biol. Chem.* **254**, 12219–12223.
- Cudney, B., Patel, S., Weisgraber, K., Newhouse, Y. & McPherson, A. (1994). *Acta Cryst.* **D50**, 414–423.
- Gilliland, G. L. (1988). *J. Cryst. Growth*, **90**, 51–59.
- Gilliland, G. L. & Bickham, D. (1990). *Methods*, **1**, 6–11.
- Gilliland, G. L. & Davies, D. R. (1984). *Methods Enzymol.* **104**, 370–381.

<sup>1</sup> Certain commercial materials, instruments and equipment are identified in this manuscript in order to specify the experimental procedure as completely as possible. In no case does such identification imply a recommendation or endorsement by the National Institute of Standards and Technology, nor does it imply that the materials, instruments or equipment identified is necessarily the best available for the purpose.

- Gilliland, G. L., Howard, A. J., Winborne, E. L., Poulos, T. L., Steward, D. B. & Durham, D. R. (1987). *J. Biol. Chem.* **262**, 4280–4283.
- Gilliland, G. L., Tung, M., Blakeslee, D. M. & Ladner, J. (1994). *Acta Cryst.* **D50**, 408–413.
- Gilliland, G. L., Tung, M. & Ladner, J. (1996). *J. Res. Natl Inst. Stand. Technol.* **101**, 309–320.
- Gilliland, G. L., Tung, M. & Ladner, J. (1997). *Proceedings of the IUCr School on Crystallographic Databases*.
- Gilliland, G. L., Tung, M. & Ladner, J. (2001). *International Tables For X-ray Crystallography*, Vol. F, *Macromolecular Crystallography*, edited by M. G. Rossmann & E. Arnold, pp. 669–674. Dordrecht: Kluwer Academic Publishers.
- Jancarik, J. & Kim, S.-H. (1991). *J. Appl. Cryst.* **24**, 409–411.
- McPherson, A. Jr (1976). *Methods Biochem. Anal.* **23**, 249–345.
- McPherson, A. (1982). *Preparation and Analysis of Protein Crystals*. New York: Wiley.
- McPherson, A. (1999). *Crystallization of Biological Macromolecules*. New York: Cold Spring Harbor Laboratory Press.
- Scott, W. G., Finch, J. T., Grenfell, R., Fogg, J., Smith, T., Gait, M. J. & Klug, A. (1995). *J. Mol. Biol.* **250**, 327–332.