International Tables for Crystallography (2006). Vol. F, Chapter 24.4, pp. 669–674.

24.4. The Biological Macromolecule Crystallization Database

BY G. L. GILLILAND, M. TUNG AND J. E. LADNER

24.4.1. Introduction

The crystallization of a biological macromolecule is the first step in determining its three-dimensional structure by X-ray crystallographic techniques. In crystallizing macromolecules, empirical procedures are used that take advantage of the knowledge gained from past successes. The solution properties of a macromolecule, determined by factors such as shape, size, conformational stability and surface complexity, directly relate to if and how it will crystallize. Usually, when a crystallization study is initiated, limited information is available on the properties of the macromolecule. Thus, a series of experiments are carried out that vary parameters such as pH, temperature, ionic strength and macromolecule concentration. The number of experiments required for success is variable. In many cases the search ends quickly either because the right choices were made early or because crystallization occurs over a broad range of conditions. Unfortunately, in other cases, a large number of experiments are required before the discovery of crystallization conditions, and in some cases no crystallization conditions are found regardless of how many experiments are performed.

After more than 50 years of experience in the production of diffraction-quality crystals, there is still no generally accepted strategy for searching for the crystal-growth parameters for a biological macromolecule. However, a number of systematic procedures and strategy suggestions have been put forth (*e.g.* McPherson, 1976; Blundell & Johnson, 1976; Carter & Carter, 1979; McPherson, 1982; Gilliland & Davies, 1984; Gilliland, 1988; Gilliland & Bickham, 1990; Gilliland *et al.*, 1994, 1996; McPherson, 1999). These and other strategies are all based on the successful experiences of the authors and of other investigators 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 failure results in the need for a more general approach.

The motivation for the creation of the Biological Macromolecule Crystallization Database (BMCD) was to provide comprehensive information to facilitate the development of crystallization strategies to produce large single crystals suitable for X-ray structural investigations (Gilliland & Davies, 1984). The earlier and current versions of the BMCD (Gilliland, 1988; Gilliland & Bickham, 1990; Gilliland *et al.*, 1994, 1996) 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.

24.4.2. History of the BMCD

The BMCD has its roots in work that was initiated in Dr David Davies' laboratory at NIH in the late 1970s and early 1980s (Gilliland & Davies, 1984). Working on a variety of frustrating protein-crystallization problems, a large body of crystallization information was extracted from the literature. This eventually led to a systematic search of the literature and a compilation of data that included almost all of the crystallization reports of biological macromolecules available at the time. In 1983 the data, as an ASCII file, were submitted to the Protein Data Bank (Chapter 24.5) for public distribution. The data included the crystallization conditions for 1025 crystal forms of more than 616 biological macromolecules.

In 1987, with assistance from the National Institute of Standards and Technology (NIST) Standard Reference Data Program, the data were incorporated into a true database and distributed with software that made it accessible using a personal computer. The database was released to the public in 1989 as the NIST/CARB (Center for Advanced Research in Biotechnology) Biological Macromolecule Crystallization Database, version 1.0 (Gilliland, 1988). In 1990, a second version of the software and data for the PC database was released (Gilliland & Bickham, 1990), and in 1994 the BMCD began including data from crystal-growth studies carried out in microgravity (Gilliland *et al.*, 1994). Recently, the BMCD has been ported to a UNIX platform to take advantage of the development of network capabilities that give the user community access to the most recent updates and allow rapid implementation of new features and capabilities of the software (Gilliland *et al.*, 1996).

24.4.3. BMCD data

The BMCD contains both data extracted from the literature defining the macromolecules and data describing the crystallization and crystal form. Macromolecule data are included for biological macromolecules for which crystals have been obtained that are suitable for diffraction studies. Crystal entries must have unique unit-cell constants. Both macromolecule and crystal entries are assigned a four-character alphanumeric identifier, beginning with M or C for macromolecule and crystal, respectively.

Macromolecule data. Each macromolecule entry includes the name of the macromolecule and other aliases. Each entry includes biological source information that includes 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 molecular weight for each type of individual subunit. (A subunit of a biological macromolecule entity is defined as a part of the assembly that is associated with another part by non-covalent interactions. For example, haemoglobin has four subunits, two α-globins and two β-globins, and the two oligomeric nucleic acid strands of a double-stranded nucleic acid fragment are considered as two subunits.) A representative macromolecule entry is illustrated in Fig. 24.4.3.1.

Crystallization and crystal data. The data in each crystal entry include the crystal data, crystal morphology, the experimental details of the crystallization procedure and complete references. The crystal data include the unit-cell dimensions (a, b, c, α, β, γ), the number of molecules in the unit cell (Z), the space group and the crystal density. 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 experimental details include the macromolecule concentration, the temperature, the pH, the chemical additives to the growth medium, the crystallization method 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. Cross-references to two other structural biology databases, the Protein Data Bank (Chapter 24.5) and the Nucleic Acid Database (Berman *et al.*, 1992), are given if the identifiers are known. One of the crystal entries for the macromolecule entry illustrated in Fig. 24.4.3.1 is shown in Fig. 24.4.3.2.

24.4.4. BMCD implementation – web interface

The BMCD is a web-accessible resource available to the crystallographic community through the website at http://wwwbmcd.nist.gov:8080/bmcd/bmcd.html. The current version of the BMCD includes 3547 crystal entries from 2526 biological macromolecules. The web interface provides an easy mechanism for browsing through the data contained in the BMCD. The user can examine the complete list of macromolecule names and tabulations of 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 is available along with a set of general references concerning all aspects of crystal growth.

The web interface offers a number of ways to query the database. For example, the results of precomputed queries are available through the tabulations of chemical additives, space groups, crystallization methods and prosthetic groups mentioned above. The tables provide links to lists of macromolecules and crystal forms that match the database query for these parameters. Allowing the user to enter specific parameter values that must be matched provides another method of searching the database. Examples of this include querying for macromolecules based on their molecular weight. A user may also search for crystal forms of macromolecules that crystallize at a particular temperature, macromolecule concentration and pH. A range of values or a single value of one or all of the parameters may be used to limit the search. For references, queries for specific authors, keywords and journal information are allowed.

24.4.5. Reproducing published crystallization procedures

The BMCD contains the information needed to reproduce the crystallization conditions for a biological macromolecule reported in the literature. This is an activity performed by many laboratories engaged in protein engineering, rational drug design, protein stability and other studies of proteins whose structures have been previously determined. The crystallization of sequence variants, chemically modified derivatives, or ligand–biological macromolecule complexes can sometimes be considered problems that fit into this category. Usually, the reported crystallization conditions of the native macromolecule are the starting points for initiating 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. The crystallization conditions in the database should be considered 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.

The crystals for one of the isozymes of glutathione S-transferase of rat liver grown from conditions reported in the literature (Sesay *et al.*, 1987) are used to illustrate these points. The original crystallization conditions were for an enzyme isolated from rat liver (entries M0P3 and C13R). However, the enzyme used in the crystallization trials was cloned and expressed in *Escherichia coli*. The crystals of the natural enzyme were grown in 3 to 5 days from vapour-diffusion experiments at 4 °C, with droplets containing a protein concentration of 11.3 mg ml^{-1}, 0.46% β-octylglucoside, 30–37% saturated ammonium sulfate and 0.1 *M* phosphate buffer, pH 6.9 equilibrated against well solutions containing 60–74% ammonium sulfate.

The recombinant enzyme required an optimization of these conditions to produce large single crystals (Ji *et al.*, 1994). The recombinant protein crystallized best at 4.0 °C, with droplets containing a protein concentration of 12 mg ml^{-1}, 0.2% β-octylglucoside, 20–25% saturated ammonium sulfate, 1 m*M* EDTA and 0.025 *M* TrisHCl, pH 8.0 equilibrated against well solutions containing 40–50% ammonium sulfate. Both crystallization protocols required the presence of 1 m*M* (9*R*,10*R*)-9-*S*-glutathionyl-10-hydroxy-9,10-dihydrophenanthrene, a product inhibitor. The recombinant enzyme, β-octylglucoside and ammonium concentrations were adjusted. The pH was varied, with the largest crystals being found at pH 8.0. Thus, TrisHCl was substituted for the phosphate buffer. EDTA was also included as an additive; its

Macromolecule -- subtilisin BPN' : prodomain (M1MT)

Molecule Name : subtilisin BPN' : prodomain
Alias Names :
Genus: Bacillus
Species: amyliloquefaciens
Common Name: recombinant
EC Number : 3.4.21.14
Catalytic Reaction : hydrolysis of proteins and peptide amides
Total Molecular Weight: 36500 **Total No. Subunits:** 2
1. Subunit Molecular Weight: 28000; Subunit Number: 1
2. Subunit Molecular Weight: 8500; Subunit Number: 1

Prosthetic Group :
Remarks : During biosynthesis of subtilisin, the 77 amino acid prodomain is autocatatalytically cleaved.

Total Number of Crystal Entries: 2

1. Crystal C2CJ: A=71.9, B=96.7, C=48.9, alpha=90, beta=90, gamma=90
2. Crystal C2CK: A=74.1, B=77.85, C=57.65, alpha=90, beta=90, gamma=90

Fig. 24.4.3.1. A representative example of a biological macromolecule, subtilisin BPN′: prodomain, entry M1MT in the BMCD.

Crystallization Data for Molecule - subtilisin BPN' : prodomain

Crystal Data (C2CK).
Molecular Name: subtilisin BPN' : prodomain.

Unit cell dimension						
A	B	C	Alpha	beta	gamma	Z
74.1	77.85	57.65	90	90	90	4

Space group:
P2<1>2<1>2 - Orthorhombic
Crystal Density: 0
Diffraction limit (Angstroms): 2
Diffraction life time (Hours): 0
Crystal dimensions in mm: 0.1 x 0.5 x 0.5

Crystallization Data.
Method(s) used for the crystallization: vapor diffusion in hanging drops

Macromolecule concentration range	0	0	mg/ml
Temperature range	0	0	° C
pH value	4.6	0	
Crystal growth time	0	0	Day

Chemical additions in the reservoir:

Chemical Additive	Range		Unit
ammonium sulfate	2	0	M
Sodium acetate	0.1	0	M

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

Reference for Crystal - C2CK.

Reference (R2NE).
Reference Title: "The prosegment-subtilisin BPN' complex:crystal structure of a specific 'foldase'."
Author Names: Gallagher, T; Gilliland, GL; Wang, L; Bryan, P;
Reference Source: Structure 3, 907- 914, 1995.
Reference also for: Crystal C2CJ |

Other Information Related to the Crystal.
Protein Data Bank: 1SPB
NDB:

Fig. 24.4.3.2. A representative example of a crystal entry C2CK for the subtilsin BPN′: prodomain, entry M1MT in the BMCD.

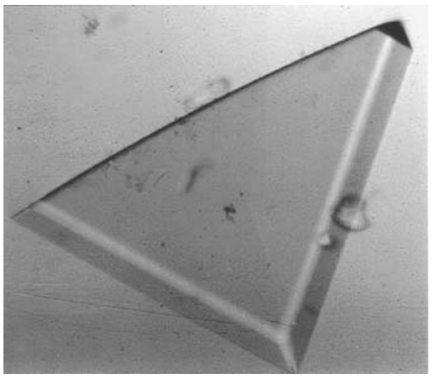

Fig. 24.4.5.1. Crystal of recombinant rat liver glutathione *S*-transferase (Ji *et al.*, 1994) grown from optimized conditions based on the data in BMCD entries M0P3 and C13R.

absence or presence did not affect the crystallization. Crystals of the recombinant enzyme grew within 5 to 10 days (Fig. 24.4.5.1).

24.4.6. Crystallization screens

With the introduction of the fast screen by Jancarik & Kim (1991), almost all attempts to crystallize a protein begin with experiments based on a screen of one form or another. The first screen was based on the ideas put forth by Carter & Carter (1979) in their discussion of the use of incomplete factorial experiments to limit the search, the experience of the investigators themselves, and the experience of others. A number of screens have been developed and even commercialized (*e.g.* Cudney *et al.*, 1994). The first screens were quite general and applicable to a wide range of biological macromolecules, but fast screens based on specific classes of molecules such as RNA soon developed (Scott *et al.*, 1995).

The BMCD is an ideal tool for facilitating the development of screens for general or specific classes of macromolecules. For example, if it were desired to produce a screen for endonucleases, a quick search of the PDB would provide the information in Table 24.4.6.1. An examination of the crystallization conditions of the endonucleases reveals that crystals are grown using a protein concentration ranging from 2.5 to 12.9 mg ml^{-1}, ammonium sulfate, sodium phosphate, or polyethylene glycol 400 to 8000 as precipitants at 4 to 20 °C between pH 4.5 and 8.3. A variety of buffers and standard biochemical additives are also used. From an examination of these parameters, a small subset of the crystallization experiments comprising an endonuclease screen could be developed.

24.4.7. A general crystallization procedure

The use of the BMCD has been incorporated into more general procedures required for the crystallization of a biological macromolecule that has never been crystallized (Gilliland, 1988; Gilliland & Bickham, 1990; Gilliland *et al.*, 1994, 1996). One such general procedure is shown in Fig. 24.4.7.1. The example below illustrates how the data in the BMCD were used to develop this general procedure for soluble proteins. The BMCD can be used to develop analogous procedures for other classes of biological macromolecules. Briefly, in this procedure the purified biological macromolecule is concentrated (if possible) to 10 to 25 mg ml^{-1} and dialysed into 0.005 to 0.025 *M* buffer at a neutral pH or at a pH required to maintain solubility of the biopolymer. Other stabilizing agents such as EDTA and/or dithiothreitol may be included at low concentrations to stabilize the biological macromolecule during the crystallization trials.

Once the protein has been prepared, commercial or customized fast screens are carried out using vapour-diffusion experiments. If crystals are obtained, X-ray diffraction studies are initiated, but frequently small or poor-quality crystals are observed. Experiments that systematically vary the crystallization parameters (pH, ionic strength, temperature *etc.*) are then carried out. Micro- or macroseeding may also be required to optimize crystal growth (McPherson, 1982, 1999).

If the fast screens produce no crystals, a more systematic approach can be undertaken that is based on the data contained in the BMCD. 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 (~20 °C) or cold-room (6 °C) temperatures. Protein concentration varies quite markedly, but it appears that investigators typically use >10 mg ml^{-1}.

After examining the data in the BMCD, the precipitating agents, ammonium sulfate, polyethylene glycol 8000, 2-methyl-2,4-pentanediol and sodium–potassium phosphate might be selected for the initial crystallization attempts, and experiments might be restricted to a pH range of 3.0 to

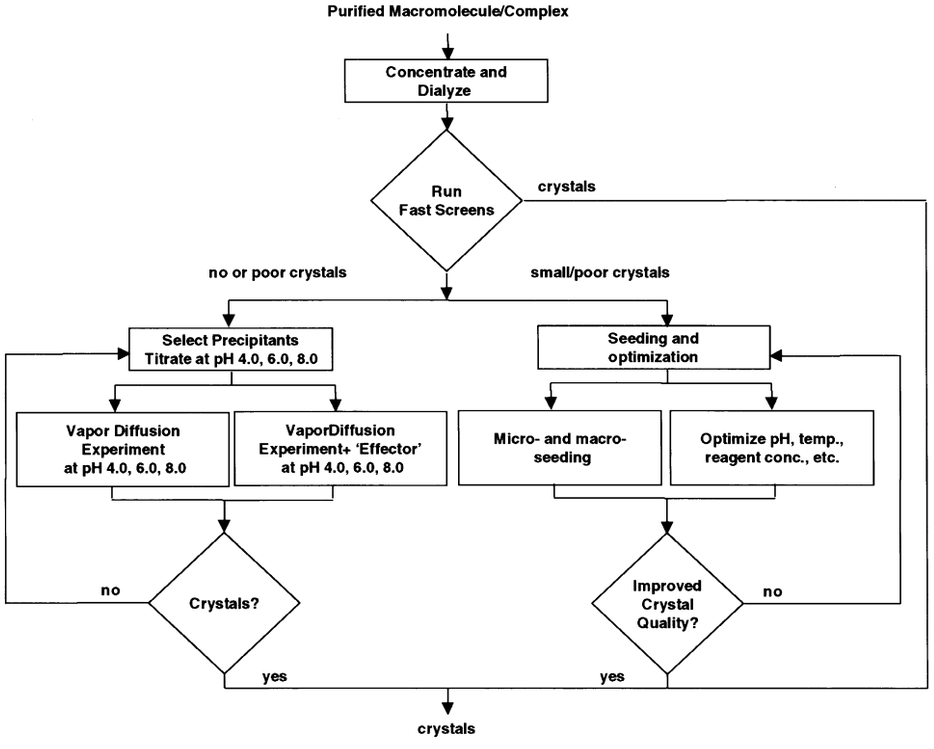

Fig. 24.4.7.1. A general crystallization strategy based on the data contained in the BMCD.

Table 24.4.6.1. *Crystallization conditions for endonucleases*

Endonuclease	Crystal data (space group, unit cell)	Crystallization method	Protein concentration (mg ml^{-1})	Chemical additives to reservoir	pH	*T* (°C)	Reference(s)
*Bam*HI	*C*2 76.4, 46.0, 69.4 Å, 110.5°	Vapour diffusion with microseeding	12.5	10% Glycerol, 0.02 *M* potassium phosphate, 12% polyethylene glycol 8K	6.9	20	Newman *et al.* (1994)
*Bam*HI:12bp DNA	$P2_12_12_1$ 108.8, 81.9, 68.8 Å	Vapour diffusion	10.8	5% Glycerol, 12% polyethylene glycol 8K, 0.15 *M* potassium chloride	6.9–7.6	22–24	Strzelecka *et al.* (1994)
*Crf*10I, type II restriction	*I*222 64.5, 81.3, 119.7 Å	Vapour diffusion	6.0	1.0 *M* Ammonium acetate, 0.075 *M* MES	6.5–7.5	20	Bozic *et al.* (1996)
*Eco*RV, type II restriction	$P2_12_12_1$ 58.2, 71.7, 136.0 Å	Vapour diffusion	3.3–5.3	0.18 *M* Sodium chloride, 10% polyethylene glycol 4K	7.0–7.8	20–24	D'Arcy *et al.* (1985); Winkler *et al.* (1993)
*Eco*RV, type II restriction:11-mer DNA	*P*1 49.4, 50.2, 64.0 Å, 96.5, 109.1, 108.1°	Batch	6.4–12.9	0.0043 *M* Phosphate buffer, 0.1073 *M* sodium chloride, 0.00040 *M* EDTA, 0.00040 *M* dithiothreitol, 0.0043 *M* cacodylate, 0.8–1.4% polyethylene glycol 4K	6.0–7.5	22–24	Kostrewa & Winkler (1995)
*Eco*RV, type II restriction:cognate DNA	$C222_1$ 60.2, 78.4, 371.3 Å	Vapour diffusion	n/a	0.1 *M* Sodium phosphate, 0.08 *M* sodium chloride	6.4–6.9	19–22	Winkler *et al.* (1991, 1993)
*Eco*RV, type II restriction:non-cognate DNA	$P2_1$ 68.4, 79.6, 6.4 Å, 104.6°	Vapour diffusion	7.0–10.0	0.1 *M* Sodium chloride, 0.02 *M* MES	6.4–6.9	19–22	Winkler *et al.* (1991, 1993)
*Eco*RV, type II restriction:product DNA	*P*1 49.3, 50.3, 63.9 Å, 96.7, 108.8, 108.4°	Microbatch	6.4	0.0043 *M* Phosphate buffer, 0.1073 *M* sodium chloride, 0.0004 *M* EDTA, 0.0004 *M* dithiothreitol, 0.0043 *M* cacodylate, 1–2% polyethylene glycol 4K, 0.0043 *M* magnesium chloride	6.0–7.5	4–22	Kostrewa & Winkler (1995)
II, DNA repair [4Fe–4S]	$P2_12_12_1$ 48.5, 65.8, 86.8 Å	Dialysis and macroseeding	n/a	5.0% Glycerol, 0.0003 *M* sodium azide, 0.1 *M* sodium chloride, 0.005 *M* HEPES	7.0	15	Kuo, McRee, Cunningham & Tainer (1992); Kuo, McRee, Fisher *et al.* (1992)
*Pvu*II	$P2_12_12$ 84.2, 106.2, 46.9 Å	Vapour diffusion	2.5	20–50% Saturated ammonium sulfate	5.0	18	Athanasiadis & Kokkinidis (1991)
*Puv*II:cognate DNA	$P2_12_12_1$ 95.8, 86.3, 48.5 Å	Vapour diffusion	9.6	0.0001 *M* EDTA, 2.5–3.6% polyethylene glycol 4K, 0.0155–0.0225 *M* sodium acetate	4.5	16	Balendiran *et al.* (1994)

Table 24.4.6.1. *Crystallization conditions for endonucleases* (*cont.*)

Endonuclease	Crystal data (space group, unit cell)	Crystallization method	Protein concentration (mg ml^{-1})	Chemical additives to reservoir	pH	T (°C)	Reference(s)
*Ruv*C specific for Holliday junctions	$P2_1$ 72.8, 139.6, 32.4 Å, 93.0°	Microdialysis	8.0	0.05 *M* TrisHCl, 7.5% glycerol, 0.001 *M* EDTA, 0.001 *M* dithiothreitol, 0.3–0.4 *M* sodium cloride	8.0	22–24	Ariyoshi *et al.* (1994)
V, mutant E23Q	$P2_1$ 41.4, 40.1, 37.4 Å, 90.4°	Vapour diffusion	10.0	0.05 *M* Potassium chloride, 0.008 *M* sodium cacodylate, 15% polyethylene glycol 400	4.5–8.0	4	Morikawa *et al.* (1995)
V, mutant R3Q	$P2_1$ 41.4, 40.7, 37.4 Å, 90.1°	Vapour diffusion	10.0	0.05 *M* Potassium chloride, 0.008 *M* sodium cacodylate, 15% polyethylene glycol 400	4.5–8.0	4	Morikawa *et al.* (1995)
V	$P2_1$ 41.4, 40.1, 37.6 Å, 90.01°	Vapour diffusion	10.0	0.05 *M* Potassium chloride, 0.008 *M* sodium cacodylate, 15% polyethylene glycol 400	4.5	4	Morikawa *et al.* (1988, 1992, 1995)
V, mutant E23D	$P2_1$ 41.7, 40.2, 37.1 Å, 92°	Vapour diffusion	10.0	0.05 *M* Potassium chloride, 0.008 *M* sodium cacodylate, 15% polyethylene glycol 400	4.5–8.0	4	Morikawa *et al.* (1995)
Sm_I	$P2_12_12_1$ 69.0, 106.7, 74.8 Å	Vapour diffusion	10.0	0.01 *M* TrisHCl, 1.2–1.6 *M* ammonium sulfate	8.3	4	Bannikova *et al.* (1991)
Extracellular	$P2_12_12_1$ 106.7, 74.5, 68.9 Å	Dialysis	8.0	1.0–1.7 *M* Ammonium sulfate, 0.05 *M* sodium phosphate	6.0	4	Miller *et al.* (1991)
Restriction, *Fok*I:20pb DNA	$P2_1$ 65.6, 119.3, 71.5 Å, 101.4°	Vapour diffusion with macroseeding	10.0	1.1 *M* Ammonium sulfate, 0.5 *M* MES, 0.2 *M* potassium chloride, 0.0005 *M* dithiothreitol, 0.0005 *M* EDTA, 5% glycerol	6.0	20	Hirsch *et al.* (1997); Wah *et al.* (1997)

9.0 and temperatures of 6 and 20 °C. Then a small amount (10 µl) of the protein is titrated with each of the selected reagents (McPherson, 1976) at pH 4.0, 6.0 and 8.0 at both cold-room and room temperatures. This establishes the concentration ranges for the reagents for setting up hanging-drop (or any other commonly used technique) experiments. Next, separate sets of experiments that would sample the pH range in steps of 1.0 and reagent concentrations near, at and above what might induce precipitation of the protein would be set up at temperatures of 6 and 20 °C. The assessment of the results of experiments after periodic observations may show (for example by an abrupt precipitation at a particular reagent concentration, pH and/or temperature) a need for finer sampling of any or all of the parameters near the observed discontinuity. In parallel, or if the crystallization trials just described are unsuccessful, another set of experiments can be carried out that include the addition of small quantities of ligands, products, substrate, substrate analogues, monovalent or divalent cations, organic reagents *etc.* to the crystallization mixtures. If this does not prove fruitful, additional reagents may be selected with the aid of the BMCD and new experiments initiated.

In addition to the procedure described above, a set of experiments at reduced ionic strength should be considered. The BMCD shows that about 10% of soluble proteins crystallize at low ionic strength (< 0.2 *M*). Thus, microdialysis experiments that equilibrate the protein solutions against low ionic strength over time in a stepwise manner over a pH range of 3.0 to 9.0 in steps of 0.5 to 1.0 should also be undertaken. It is also worthwhile to do microdialysis experiments at or near the protein's isoelectric point, a point at which a protein is often the least soluble. As with the vapour-diffusion experiments mentioned above, if crystallization does not occur, the introduction of small quantities of ligands, products, substrate, substrate analogues, monovalent or divalent cations,

organic reagents *etc.* to the crystallization mixtures may facilitate crystal growth. Also, in analogy to the vapour-diffusion experiments, the search may be expanded to finer increments of pH if results warrant.

24.4.8. The future of the BMCD

The BMCD will continue to be available for years to come. New data will be incorporated annually and direct deposition of data by the user community is being considered. 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 integrated with other structural-biology web resources to address the structural-biology challenges of the future.

Acknowledgements

The authors would like to acknowledge the assistance of Ms X. R. Dong of CARB in the acquisition of literature data. 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.

24.2 (*cont.*)

Berman, H. M., Westbrook, J., Feng, Z., Gilliland, G., Bhat, T. N., Weissig, H., Shindyalov, I. N. & Bourne, P. E. (2000). *The Protein Data Bank*. *Nucleic Acids Res.* **28**, 235–242.

Bernstein, F. C., Koetzle, T. F., Williams, G. J., Meyer, E. E., Brice, M. D., Rodgers, J. R., Kennard, O., Shimanouchi, T. & Tasumi, M. (1977). *Protein Data Bank: a computer-based archival file for macromolecular structures*. *J. Mol. Biol.* **112**, 535–542.

Bourne, P., Berman, H. M., Watenpaugh, K., Westbrook, J. D. & Fitzgerald, P. M. D. (1997). *The macromolecular Crystallographic Information File (mmCIF)*. *Methods Enzymol.* **277**, 571–590.

Brünger, A. T. (1992). *X-PLOR*. Version 3.1. *A system for X-ray crystallography and NMR*. Yale University Press, New Haven, CT, USA.

Clowney, L., Jain, S. C., Srinivasan, A. R., Westbrook, J., Olson, W. K. & Berman, H. M. (1996). *Geometric parameters in nucleic acids: nitrogenous bases*. *J. Am. Chem. Soc.* **118**, 509–518.

Feng, Z., Hsieh, S.-H., Gelbin, A. & Westbrook, J. (1998). *MAXIT: macromolecular exchange and input tool*. NDB-120. Rutgers University, New Brunswick, NJ, USA.

Feng, Z., Westbrook, J. & Berman, H. M. (1998). *NUCheck*. NDB-407. Rutgers University, New Brunswick, NJ, USA.

Gelbin, A., Schneider, B., Clowney, L., Hsieh, S.-H., Olson, W. K. & Berman, H. M. (1996). *Geometric parameters in nucleic acids: sugar and phosphate constituents*. *J. Am. Chem. Soc.* **118**, 519–528.

Grzeskowiak, K., Yanagi, K., Privé, G. G. & Dickerson, R. E. (1991). *The structure of B-helical C-G-A-T-C-G-A-T-C-G and comparison with C-C-A-A-C-G-T-T-G-G: the effect of base pair reversal*. *J. Biol. Chem.* **266**, 8861–8883.

Lavery, R. & Sklenar, H. (1989). *Defining the structure of irregular nucleic acids: conventions and principles*. *J. Biomol. Struct. Dyn.* **6**, 655–667.

Parkinson, G., Vojtechovsky, J., Clowney, L., Brünger, A. T. & Berman, H. M. (1996). *New parameters for the refinement of nucleic acid-containing structures*. *Acta Cryst.* D**52**, 57–64.

Sayle, R. & Milner-White, E. J. (1995). *RasMol: biomolecular graphics for all*. *Trends Biochem. Sci.* **20**, 374.

Schneider, B., Neidle, S. & Berman, H. M. (1997). *Conformations of the sugar–phosphate backbone in helical DNA crystal structures*. *Biopolymers*, **42**, 113–124.

Scott, W. G., Finch, J. T. & Klug, A. (1995). *The crystal structure of an all-RNA hammerhead ribozyme: a proposed mechanism for RNA catalytic cleavage*. *Cell*, **81**, 991–1002.

Vaguine, A. A., Richelle, J. & Wodak, S. J. (1999). *SFCHECK: a unified set of procedures for evaluating the quality of macromolecular structure-factor data and their agreement with the atomic model*. *Acta Cryst.* D**55**, 191–205.

Westbrook, J. (1998). *AutoDep input tool*. NDB-406. Rutgers University, New Brunswick, NJ, USA.

24.3

Abola, E. E., Sussman, J. L., Prilusky, J. & Manning, N. O. (1997). *Protein Data Bank archives of three-dimensional macromolecular structures. Methods Enzymol.* **277**, 556–571.

Allen, F. H., Davies, J. E., Galloy, J. J., Johnson, O., Kennard, O., Macrae, C. F., Mitchell, E. M., Mitchell, G. F., Smith, J. M. & Watson, D. G. (1991). *The development of versions 3 and 4 of the Cambridge Structural Database system. J. Chem. Inf. Comput. Sci.* **31**, 187–204.

Allen, F. H., Kennard, O., Watson, D. G., Brammer, L., Orpen, A. G. & Taylor, R. (1987). *Tables of bond lengths determined by X-ray and neutron diffraction. Part 1. Bond lengths in organic compounds. J. Chem. Soc. Perkin Trans.* 2, pp. S1–S19.

Bernstein, F. C., Koetzle, T. F., Williams, G. J. B., Meyer, E. F., Brice, M. D., Rodgers, J. R., Kennard, O., Shimanouchi, T. & Tasumi, M. (1972). *The Protein Data Bank: a computer-based archival file for macromolecular structures. J. Mol. Biol.* **112**, 535–547.

Bruno, I. J., Cole, J. C., Lommerse, J. P. M., Rowland, R. S., Taylor, R. & Verdonk, M. L. (1997). *IsoStar: a library of information about nonbonded interactions. J. Comput.-Aided Mol. Des.* **11**, 525–537.

Hall, S. R., Allen, F. H. & Brown, I. D. (1991). *The crystallographic information file (CIF): a new standard archive file for crystallography. Acta Cryst.* A**47**, 655–685.

Hayes, I. C. & Stone, A. J. (1984). *An intermolecular perturbation theory for the region of moderate overlap. J. Mol. Phys.* **53**, 83–105.

Kennard, O. & Allen, F. H. (1993). *3D search and research using the Cambridge Structural Database. Chem. Des. Autom. News*, **8**, 1, 31–37.

RCSB (2000). The Protein Data Bank. Research Collaboratory for Structural Bioinformatics, Department of Chemistry, Rutgers University, Piscataway, NJ, USA (http://www.rcsb.org).

Sayle, R. (1996). *The RASMOL visualiser*. Glaxo Wellcome Research, Stevenage, Hertfordshire, England.

24.4

Ariyoshi, M., Vassylyev, D. G., Iwasaki, H., Fujishima, A., Shinagawa, H. & Morikawa, K. (1994). *Preliminary crystallographic study of Escherichia coli RuvC protein. An endonuclease specific for Holliday junctions. J. Mol. Biol.* **241**, 281–282.

Athanasiadis, A. & Kokkinidis, M. (1991). *Purification, crystallization and preliminary X-ray diffraction studies of the PvuII endonuclease. J. Mol. Biol.* **222**, 451–453.

Balendiran, K., Bonventre, J., Knott, R., Jack, W., Benner, J., Schildkraut, I. & Anderson, J. E. (1994). *Expression, purification, and crystallization of restriction endonuclease PvuII with DNA containing its recognition site. Proteins*, **19**, 77–79.

Bannikova, G. E., Blagova, E. V., Dementiev, A. A., Morgunova, E. Yu., Mikchailov, A. M., Shlyapnikov, S. V., Varlamov, V. P. & Vainshtein, B. K. (1991). *Two isoforms of Serratia marcescens nuclease. Crystallization and preliminary X-ray investigation of the enzyme. Biochem. Int.* **24**, 813–822.

Berman, H. M., Olson, W. K., Beveridge, D. L., Westbrook, J., Gelbin, A., Demeny, T., Hsieh, S.-H., Srinivasan, A. R. & Schneider, B. (1992). *The Nucleic Acid Database: a comprehensive relational database of three-dimensional structures of nucleic acids. Biophys. J.* **63**, 751–759.

Blundell, T. L. & Johnson, L. N. (1976). *Protein crystallography*. New York: Academic Press.

Bozic, D., Grazulis, S., Siksnys, V. & Huber, R. (1996). *Crystal structure of Citrobacter freundii restriction endonuclease Cfr10I at 2.15 Å resolution. J. Mol. Biol.* **255**, 176–186.

Carter, C. W. Jr & Carter, C. W. (1979). *Protein crystallization using incomplete factorial experiments. J. Biol. Chem.* **254**, 12219–12223.

Cudney, B., Patel, S., Weisgraber, K., Newhouse, Y. & McPherson, A. (1994). *Screening and optimization strategies for macromolecular crystal growth. Acta Cryst.* D**50**, 414–423.

D'Arcy, A., Brown, R. S., Zabeau, M., van Resandt, R. W. & Winkler, F. K. (1985). *Purification and crystallization of the EcoRV restriction endonuclease. J. Biol. Chem.* **260**, 1987–1990.

Gilliland, G. L. (1988). *A biological macromolecule crystallization database: a basis for a crystallization strategy. J. Cryst. Growth*, **90**, 51–59.

Gilliland, G. L. & Bickham, D. (1990). *The Biological Macromolecule Crystallization Database: a tool to assist the development of crystallization strategies. Methods Companion Methods Enzymol.* **1**, 6–11.

Gilliland, G. L. & Davies, D. R. (1984). *Protein crystallization: the growth of large-scale single crystals. Methods Enzymol.* **104**, 370–381.

Gilliland, G. L., Tung, M., Blakeslee, D. M. & Ladner, J. E. (1994). *Biological Macromolecule Crystallization Database, version 3.0: new features, data and the NASA archive for protein crystal growth data. Acta Cryst.* D**50**, 408–413.

Gilliland, G. L., Tung, M. & Ladner, J. (1996). *The Biological Macromolecule Crystallization Database and NASA Protein*

24.4 (*cont.*)

Crystal Growth Archive. J. Res. Natl Inst. Stand. Technol. **101**, 309–320.

Hirsch, J. A., Wah, D. A., Dorner, L. F., Schildkraut, I. & Aggarwal, A. K. (1997). *Crystallization and preliminary X-ray analysis of restriction endonuclease FokI bound to DNA. FEBS Lett.* **403**, 136–138.

Jancarik, J. & Kim, S.-H. (1991). *Sparse matrix sampling: a screening method for crystallization of proteins. J. Appl. Cryst.* **24**, 409–411.

Ji, X., Johnson, W. W., Sesay, M. A., Dickert, L., Prasad, S. M., Ammon, H. L., Armstrong, R. N. & Gilliland, G. L. (1994). *Structure of the xenobiotic substrate binding site of a glutathione S-transferase as revealed by X-ray crystallographic analysis of product complexes with the diastereomers of 9-(S-glutathionyl)-10-hydroxy-9,10-dihydrophenanthrene. Biochemistry*, **33**, 1043–1052.

Kostrewa, D. & Winkler, F. K. (1995). *Mg^{2+} binding to the active site of EcoRV endonuclease: a crystallographic study of complexes with substrate and product DNA at 2 angstroms resolution. Biochemistry*, **34**, 683–696.

Kuo, C. F., McRee, D. E., Cunningham, R. P. & Tainer, J. A. (1992). *Crystallization and crystallographic characterization of the iron–sulfur-containing DNA-repair enzyme endonuclease III from Escherichia coli. J. Mol. Biol.* **227**, 347–351.

Kuo, C. F., McRee, D. E., Fisher, C. L., O'Handley, S. F., Cunningham, R. P. & Tainer, J. A. (1992). *Atomic structure of the DNA repair [4Fe–4S] enzyme endonuclease III. Science*, **258**, 434–440.

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.

McPherson, A. Jr (1976). *The growth and preliminary investigation of protein and nucleic acid crystals for X-ray diffraction analysis. Methods Biochem. Anal.* **23**, 249–345.

Miller, M. D., Benedik, M. J., Sullivan, M. C., Shipley, N. S. & Krause, K. L. (1991). *Crystallization and preliminary crystallographic analysis of a novel nuclease from Serratia marcescens. J. Mol. Biol.* **222**, 27–30.

Morikawa, K., Ariyoshi, M., Vassylyev, D. G., Matsumoto, O., Katayanagi, K. & Ohtsuka, E. (1995). *Crystal structure of a pyrimidine dimer-specific excision repair enzyme from bacteriophage T4: refinement at 1.45 Å and X-ray analysis of the three active site mutants. J. Mol. Biol.* **249**, 360–375.

Morikawa, K., Matsumoto, O., Tsujimoto, M., Katayanagi, K., Ariyoshi, M., Doi, T., Ikehara, M., Inaoka, T. & Ohtsuka, E. (1992). *X-ray structure of T4 endonuclease V: an excision repair enzyme specific for a pyrimidine dimer. Science*, **256**, 523–526.

Morikawa, K., Tsujimoto, M., Ikehara, M., Inaoka, T. & Ohtsuka, E. (1988). *Preliminary crystallographic study of pyrimidine dimer-specific excision-repair enzyme from bacteriophage T4. J. Mol. Biol.* **202**, 683–684.

Newman, M., Strzelecka, T., Dorner, L. F., Schildkraut, I. & Aggarwal, A. K. (1994). *Structure of restriction endonuclease BamhI phased at 1.95 Å resolution by MAD analysis. Structure*, **2**, 439–452.

Scott, W. G., Finch, J. T., Grenfell, R., Fogg, J., Smith, T., Gait, M. J. & Klug, A. (1995). *Rapid crystallization of chemically synthesized hammerhead RNAs using a double screening procedure. J. Mol. Biol.* **250**, 327–332.

Sesay, M. A., Ammon, H. L. & Armstrong, R. N. (1987). *Crystallization and a preliminary X-ray diffraction study of isozyme 3–3 of glutathione S-transferase from rat liver. J. Mol. Biol.* **197**, 377–378.

Strzelecka, T., Newman, M., Dorner, L. F., Knott, R., Schildkraut, I. & Aggarwal, A. K. (1994). *Crystallization and preliminary X-ray analysis of restriction endonuclease BamHI–DNA complex. J. Mol. Biol.* **239**, 430–432.

Wah, D. A., Hirsch, J. A., Dorner, L. F., Schildkraut, I. & Aggarwal, A. K. (1997). *Structure of the multimodular endonuclease FokI bound to DNA. Nature (London)*, **388**, 97–100.

Winkler, F. K., Banner, D. W., Oefner, C., Tsernoglou, D., Brown, R. S., Heathman, S. P., Bryan, R. K., Martin, P. D., Petratos, K. & Wilson, K. S. (1993). *The crystal structure of EcoRV endonuclease and of its complexes with cognate and non-cognate DNA fragments. EMBO J.* **12**, 1781–1795.

Winkler, F. K., D'Arcy, A., Blocker, H., Frank, R. & van Boom, J. H. (1991). *Crystallization of complexes of EcoRv endonuclease with cognate and non-cognate DNA fragments. J. Mol. Biol.* **217**, 235–238.

24.5

Berman, H. M., Olson, W. K., Beveridge, D. L., Westbrook, J., Gelbin, A., Demeny, T., Hsieh, S. H., Srinivasan, A. R. & Schneider, B. (1992). *The Nucleic Acid Database – a comprehensive relational database of three-dimensional structures of nucleic acids. 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). *The Protein Data Bank. Nucleic Acids Res.* **28**, 235–242.

Bernstein, F. C., Koetzle, T. F., Williams, G. J., Meyer, E. E., Brice, M. D., Rodgers, J. R., Kennard, O., Shimanouchi, T. & Tasumi, M. (1977). *Protein Data Bank: a computer-based archival file for macromolecular structures. J. Mol. Biol.* **112**, 535–542.

Bourne, P., Berman, H. M., Watenpaugh, K., Westbrook, J. D. & Fitzgerald, P. M. D. (1997). *The macromolecular Crystallographic Information File (mmCIF). Methods Enzymol.* **277**, 571–590.

Brookhaven National Laboratory (1998). *AutoDep.* Version 2.1. Brookhaven National Laboratory, Upton, NY, USA.

Dodge, C., Schneider, R. & Sander, C. (1998). *The HSSP database of protein structure-sequence alignments and family profiles. Nucleic Acids Res.* **26**, 313–315.

Feng, Z., Hsieh, S.-H., Gelbin, A. & Westbrook, J. (1998). *MAXIT: macromolecular exchange and input tool.* NDB-120. Rutgers University, New Brunswick, NJ, USA.

Feng, Z., Westbrook, J. & Berman, H. M. (1998). *NUCheck*. NDB-407. Rutgers University, New Brunswick, NJ, USA.

Gibrat, J.-F., Madej, T. & Bryant, S. H. (1996). *Surprising similarities in structure comparison. Curr. Opin. Struct. Biol.* **6**, 377–385.

Gilliland, G. L. (1988). *A Biological Macromolecule Crystallization Database: a basis for a crystallization strategy. J. Cryst. Growth*, **90**, 51–59.

Hogue, C. W., Ohkawa, H. & Bryant, S. H. (1996). *A dynamic look at structures: WWW-Entrez and the Molecular Modeling Database. Trends Biochem. Sci.* **21**, 226–229.

Holm, L. & Sander, C. (1998). *Touring protein fold space with Dali/FSSP. Nucleic Acids Res.* **26**, 316–319.

Hooft, R. W. W., Sander, C. & Vriend, G. (1996). *Verification of protein structures: side-chain planarity. J. Appl. Cryst.* **29**, 714–716.

Kabsch, W. & Sander, C. (1983). *Dictionary of protein secondary structure: pattern recognition of hydrogen-bonded and geometrical features. Biopolymers*, **22**, 2577–2637.

Laskowski, R. A., Hutchinson, E. G., Michie, A. D., Wallace, A. C., Jones, M. L. & Thornton, J. M. (1997). *PDBsum: a web-based database of summaries and analyses of all PDB structures. Trends Biochem. Sci.* **22**, 488–490.

Laskowski, R. A., MacArthur, M. W., Moss, D. S. & Thornton, J. M. (1993). *PROCHECK: a program to check the stereochemical quality of protein structures. J. Appl. Cryst.* **26**, 283–291.

Laskowski, R. A. & Wallace, A. C. (1998). *Enzyme Structures Database.* http://www.biochem.ucl.ac.uk/bsm/enzymes/.

Lee, B. & Richards, F. M. (1971). *The interpretation of protein structures: estimation of static accessibility. J. Mol. Biol.* **55**, 379–400.

Markley, J. L., Bax, A., Arata, Y., Hilbers, C. W., Kaptein, R., Sykes, B. D., Wright, P. E. & Wüthrich, K. (1998). *Recommendations for the presentation of NMR structures of proteins and nucleic acids.*