Development of Crystallization Strategies Using the Biological Macromolecule Crystallization Database

G. L. Gilliland, M. Tung, and J. E. Ladner

Center for Advanced Research in Biotechnology of the University of Maryland Biotechnology Institute and National Institute of Standards and Technology, 9600 Gudelsky Dr., Rockville, Maryland, U. S. A.
gary@ibm3.carb.nist.gov
http://indigo15.carb.nist.gov/carb/carb.html

Abstract

The NIST/NASA/CARB Biological Macromolecule Crystallization Database (BMCD) contains crystal data and crystallization conditions for biological macromolecules abstracted from the literature. Each entry consists of information describing the biological macromolecule crystallized and crystal data and the crystallization conditions for each crystal form. The BMCD also serves as the NASA Protein Crystal Growth Archive in that it contains protocols and results of crystallization experiments undertaken in microgravity (space). The BMCD was designed as a tool to help X-ray crystallographers in the development of protocols to crystallize biological macromolecules. Strategies for four general categories of crystallization problems are discussed that include the production of diffraction quality crystals (1) for a biological macromolecule reported in the literature, (2) for a chemically modified, or sequence, variant of a macromolecule that has been successfully crystallized, (3) for a macromolecule that is homologous to a previously crystallized macromolecule, and (4) for a macromolecule that is not related to one that has been previously crystallized.

1 Introduction

The three-dimensional structure determination of a biological macromolecule by x-ray crystallographic techniques requires the production of large single crystals. The ability of a biological macromolecule to crystallize is directly related to its solution properties, determined by factors such as size, shape, surface complexity, and conformational stability. The search for solution conditions that produce crystals suitable for diffraction studies requires not only finding suitable chemical agents that induce and sustain crystal growth but also that those parameters such as protein concentration, ionic strength, temperature, pH, etc., be sampled over a wide range of values. This could require many thousands of experiments even if only a few chemical agents are used and if parameters are varied with course increments. The number of experiments required for success is dependent on the macromolecule and the choices made by the investigator. If crystallization occurs over a broad range of conditions, or if the right choices are made in the initial experiments, the search ends quickly. In other cases many experiments are required to discover crystallization conditions, and occasionally no crystallization conditions are found, no matter how many experiments are done.

Several systematic procedures and strategy suggestions have been put forth to try to decrease the number of experiments needed to produce suitable crystals [i.e., 1-11]. Nevertheless, no universal strategy for searching for the crystal growth parameters for a biological macromolecule has gained widespread acceptance. Carter and Carter [3] were the first to propose the use of the incomplete factorial method, a sparse matrix sampling technique, to reduce the number of experiments needed to carry out in the crystallization discovery process. The method [3,9] provides the means to determine a minimal number of experiments for efficiently sampling the multidimensional parameter space. This procedure has recently been expanded to an iterative protocol that gives larger weights to parameters from experiments that have produced positive results [10]. Further experiments are based on the weighted parameter set, and the procedure is repeated until crystals suitable for diffraction experiments are obtained.

Recently, the sparse matrix sampling techniques mentioned above have been used to develop experimental procedures called *fast screens* first popularized by Jancarik and Kim [12]. These procedures employ a set of premixed solutions to set up crystallization trials that test a variety of precipitants, buffers, and additives over a wide range of pH. The solutions are ones that have frequently produced crystals. Several new general purpose and specific fast screens have been reported [i.e., 13-15].

Despite the strategy used to search for crystal growth conditions, all are based on the successes of many scientists that have produced suitable crystals for diffraction studies. The Biological Macromolecule Crystallization Database (BMCD), described here, catalogs and summarizes the information concerning crystallization that is available in the literature. It is in essence a large compendium of the successes of countless crystallization trials. The data and software that constitute the BMCD are described, and illustrations of how the database can be used to develop crystallization strategies for a variety of crystal growth problems are also presented.

2 The BMCD

The NIST/CARB (Center for Advanced Research in Biotechnology) BMCD database was created in 1989 with assistance from the National Institute of Standards and Technology (NIST) Standard Reference Data Program [6]. At this time, the BMCD software was developed as an independent program for personal computers (PCs). Access to the data was provided by menu driven software written and compiled with Clipper.[1,2] The database files were in dBase III Plus [3] format, and the index files were in the Clipper indexing format. The compilation of data consisted of the crystallization conditions of 1,025 crystal forms of 616 biological macromolecules. The information, which included most of the crystallization protocols of biological macromolecules in the literature through the end of 1982, had been previously deposited in a formatted ASCII file in the Brookhaven Protein Data Bank [16].[4]

In 1991 the second version of the NIST/CARB BMCD software and data were released [7]. The data included 1,465 crystal forms of 924 biological macromolecules. The BMCD became the NASA Protein Crystal Growth (PCG) Archive in 1994 and incorporates data from crystal growth studies supported by NASA [17]. The new software and data were released as Version 3.0 of the NIST/NASA/CARB BMCD. The number of crystal entries was increased to 2,218 for 1,465 biological macromolecules.

Version 3.0 of the BMCD has been ported to a UNIX platform to take advantage of the development of network capabilities that employ client-server tools [18]. This implementation of the BMCD uses the POSTGRES database management system [19] and World Wide Web/NCSA Mosaic client-server protocols [20]. The network version provides most of the features of the earlier PC versions of the BMCD. It also gives the user community access to more recent updates, and it allows the rapid incorporation of new features and capabilities of the software. The number of crystal forms has been increased to 3,101 for 2,166 macromolecules in release 3.1 of the BMCD software.

[1] 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.

[2] Clipper is a registered trademark of Nantucket Corp., 12555 West Jefferson Boulevard, Los Angeles, CA 90066.

[3] dBASE III Plus is a registered trademark of Borland International, Inc.

[4] Gilliland, unpublished data (1983).

Macromolecule -- transferase, glutathione S-, 3-3 isozyme **Entry:** M03C
Molecule Name: transferase, glutathione S-, 3-3 isozyme
Common Name: rat
Tissue: liver
E.C. Number: 2.5.1.18
Catalytic Reaction: RX + glutathione = HX + R-S-G
Total Molecular Weight: 52000 **Total No. Subunits:** 2
Remarks:

Total Number of Crystal Entries: 1
Click on Entry for Crystallization Conditions.
1. S.G.=5: A=88.24, B=69.44, C=81.28, alpha=90, beta=106.01, gamma=90.

Return to the *Molecule Name List.*

Any comments or bug reports regarding the BMCD should go to
mike@ibm4.carb.nist.gov

Figure 1 The BMCD macromolecule data for the rat liver 3-3 isozyme of glutathione S-transferase comprising macromolecule entry M0P3.

3 BMCD Data

Crystallization data for peptides, proteins, protein-protein complexes, nucleic acids, nucleic acid-nucleic acid complexes, protein-nucleic acid complexes and viruses reported in the literature are included in the BMCD. It is required that unit cell constants for the crystals must have been reported. In addition, the BMCD is now serving as the NASA Protein Crystal Growth Archive. The crystallization data generated from ground-based and microgravity crystallization studies supported by NASA are included as data entries.

The BMCD information is divided into three major categories: biological macromolecule, crystallization, and summary data. Each biological macromolecule included in the database has a unique biopolymer sequence, and analogously, each crystal must have unique unit cell constants or a unique intensity distribution of the diffraction pattern. The BMCD macromolecule and crystal entries include only those data elements reported in the literature.

3.1 Biological Macromolecule Entry

The data elements of a biological macromolecule entry provide a unique description of the biopolymer or biopolymer complex. A four-character alphanumeric identifier beginning with the letter M is assigned to each biological macromolecule entry. Every entry includes the macromolecule name and aliases and information concerning the biological source consisting of the common name, genus-species, tissue, cell, and organelle from which the macromolecule was isolated. The subunit composition and molecular weight consist of the total number of subunits and the total molecular weight, and the number and corresponding molecular weights of each type of distinct subunit associated with the macromolecule. A subunit of a biological macromolecule or complex is defined as a part of the assembly associated with other parts by noncovalent interactions. The name of any prosthetic group associated with the biological macromolecule is also included. The EC number [21] and catalytic reaction are provided if the macromolecule is an enzyme. For cases where special features of the macromolecule are important, or might influence, its crystallization behavior, general remarks are provided. A macromolecule entry for an enzyme, rat liver glutathione S-transferase [22], is illustrated in Figure 1.

Crystal Data for Molecule - transferase, glutathione S-, 3-3 isozyme

---

Crystallization Data.
Unit cell dimensions:

A	B	C	alpha	beta	gamma		Z
88.24	69.44	81.28	90	106.01	90	.	4

Space group: C2
Crystal system: Monoclinic
Vm of crystal (cubic A/dalton):

Methods used for the crystallization:
vapor diffusion in hanging drops
Macromolecule concentration (mg/ml): 11.3 - 11.3
Temperature (degree C): 4 - 4
pH: 6.9 - 6.9
Crystal growth time (days): 3-5

Chemical additions to the crystal growth medium:

potassium phosphate	0.01 -		M
beta-octylglucoside	0.46 -		%
ammonium sulfate	60 -	74	%

Diffraction limit (Angstroms): 2.2
Diffraction life time (Hours):
Crystal dimensions in mm: 0.2, 0.2, 0.7
Reference for Crystal Photograph:
Reference's designation for this crystal form:

References:
R0WP
Sesay, MA; Ammon, HL; Armstrong, RN.
"Crystallization and a preliminary X-ray diffraction study of isozyme 3-3 of glutathione S-transferase from rat liver:" (1987) J Mol Biol, 197, 377-378.

Comments:
The protein solution contained 0.0034 M of product inhibitor (9R,10R)-9,10-dihydro-9-(S-glutathionyl)-10-hydroxyphenanthrene.
Return to the *Molecule Name List.*

Figure 2 The BMCD crystal entry C13R, a representative example of one of the 3 crystal forms of the rat liver 3-3 isozyme of glutathione S-transferase, macromolecule entry M0P3 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 data.

3.2 Crystal Entry

Each unique crystal form of a biological macromolecule characterized by diffraction studies is described as a separate crystal entry. The crystal entry data includes the crystallization procedure, the crystal data, crystal morphology, and complete references. Each crystal entry has a four character alphanumeric identifier beginning with the letter C. The details of the experiments required to reproduce the crystallization procedure compose data elements of the crystal entry. These consist of the crystallization method, the temperature, the pH, the chemical additives to the growth medium (stabilizers, buffer, and precipitant), the length of time for crystals to grow to a useful size, and the macromolecule concentration. If the crystallization method deviates from standard protocols [4], details of the experimental procedure are provided in the comments. The crystal size, shape, and diffraction limit are given along with references for published crystal photographs or diffraction pictures. The unit cell dimensions (a, b, c, α, β, γ), the space group, the number of molecules in the unit cell (Z), and the crystal density constitute the crystal data. Data entries are cross-referenced to other structural biology databases such as the Nucleic Acid Database [23] and the Brookhaven Protein Data Bank [16]. A crystal entry is shown in Figure 2 corresponding to one of the three crystal forms of the glutathione S-transferase entry illustrated in Figure 1.

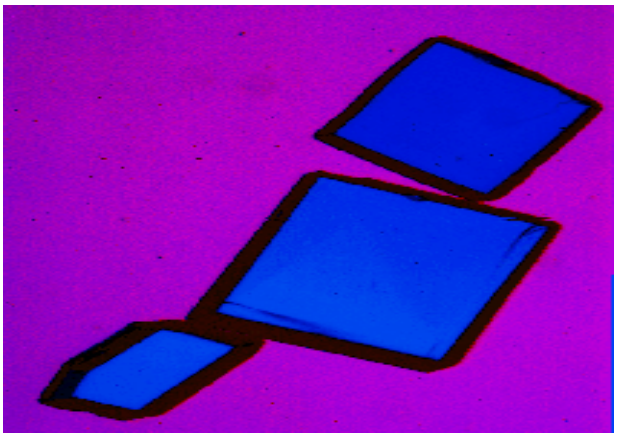

Figure 3 A photomicrograph of crystals of the 3-3 isozyme of rat liver glutathione S-transferase grown using the conditions contained in the BMCD (C13R/M0P3) [22]. The crystals are approximately 0.5 mm in their longest dimension.

A photomicrograph of the crystals grown using these conditions is shown in Figure 3.

3.3 Summary Information

The summary information provides a convenient mechanism for browsing the BMCD data. A complete list of macromolecule names links the user directly to the macromolecule and crystal entries. Tabulations of the number of macromolecules and crystal forms for each biological source, prosthetic group, space group, chemical

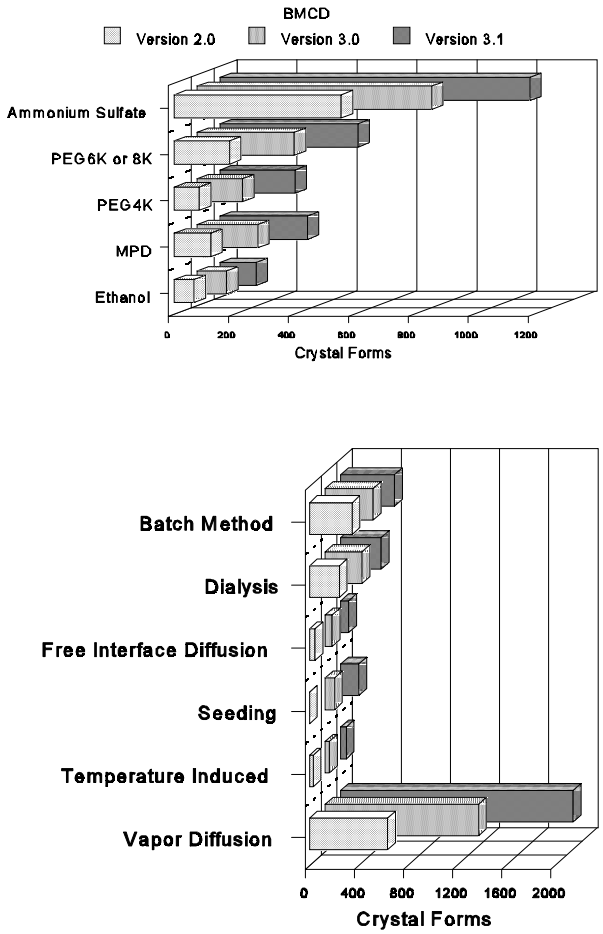

Figure 4 A display of Versions 2.0, 3.0, 3.1 of the BMCD summary information for (top) chemical additions to the crystallization solutions and (bottom) crystallization methods.

MACROMOLECULE ENTRY		
Macromolecule Name	Biological Source	Molecular Weight
Subunit Composition	Prosthetic Group	

CRYSTAL ENTRY		
Crystal data	**Crystallization Conditions**	**Reference**
Crystal System	Crystallization Method	Author
Space Group	Macromolecule Concentration	Year
Unit Cell Dimensions	Crystallization Temperature	Journal
Z, Molecules/Unit Cell	pH of Crystallization	Title
Crystal Density	Crystal Growth Time	
	Chemical Additions	

Table 1 The BMCD Search Parameters

addition, and crystallization method are available, and a listing of complete references can also be displayed. The panels in Figure 4 graphically illustrate part of the summary data for the chemical addition and crystallization method tabulations for BMCD Versions 2.0, 3.0, and 3.1.

The summary information furnishes a listing of all the BMCD references. A listing of general references concerning all aspects of crystal growth is also provided. These references have been subdivided into categories that include reviews, books and articles concerning procedures. Remarks are sometimes added to emphasize important aspects of an article that may not be evident from the title.

4 Interacting with the BMCD

The BMCD provides direct access to summary listings and the database through the user interface. Database searches may be carried out for data elements of any of the categories listed in Table 1. Depending upon the element selected, a parameter value (numerical or text) or a range of values (numerical) is requested. The interface provides mechanisms that allow searching for multiple parameters, thus, quite complex searches may be made. A listing of the macromolecule and crystal entries that meet the search criteria are displayed. The user then has the option of displaying the data associated with a particular entry. The search results can be written to an ASCII file or printed.

5 Developing Crystallization Strategies

The BMCD from its inception has been designed as a tool to help in the development of crystallization strategies [6]. Crystallization problems range from reproducing published procedures to crystallizing a newly isolated biological macromolecule that has never been crystallized [7]. Four general categories of problems that present themselves are discussed below. These include developing strategies for producing diffraction quality crystals of a biological macromolecule (1) that has been reported in the literature, (2) that is a chemically modified or a sequence variant of a macromolecule that has been successfully crystallized, (3) that is homologous to a previously crystallized macromolecule, and (4) that is not related to a biological macromolecule previously crystallized.

5.1 Previously Crystallized Macromolecules

The information required to reproduce the crystallization conditions for a biological macromolecule that has been reported in the literature is contained in the BMCD. The reported crystallization conditions are the starting points to initiate the crystallization trials. The crystallization of the biological macromolecule may be quite routine, but differences in the isolation and purification procedures, reagents, and crystallization methodology of laboratories can influence dramatically the reproducibility, and therefore, the outcome. The crystallization conditions in the database should be considered as an initial starting point that may require experiments that vary pH, macromolecule and reagent concentrations, and temperature along with the crystallization method.

	1	2	3	4	5	6	7
A	59%SAS pH 6.7	62%SAS pH 6.7	65%SAS pH 6.7	68%SAS pH 6.7	71%SAS pH 6.7	74%SAS pH 6.7	77%SAS pH 6.7
B	59%SAS pH 6.8	62%SAS pH 6.8	65%SAS pH 6.8	68%SAS pH 6.8	71%SAS pH 6.8	74%SAS pH6.8	77%SAS pH 6.8
C	59%SAS pH 6.9	62%SAS pH 6.9	65%SAS pH 6.9	68%SAS pH 6.9	71%SAS pH 6.9	74%SAS pH6.9	77%SAS pH 6.9
D	59%SAS pH 7.0	62%SAS pH 7.0	65%SAS pH 7.0	68%SAS pH 7.0	71%SAS pH 7.0	74%SAS pH 7.0	77%SAS pH 7.0
E	59%SAS pH 7.1	62%SAS pH 7.1	65%SAS pH 7.1	68%SAS pH 7.1	71%SAS pH 7.1	74%SAS pH7.1	77%SAS pH 7.1

Figure 5 Initial crystallization trials for the 3-3 isozyme of rat liver glutathione S-transferase (C13R/M0P3). SAS represents saturated ammonium sulfate. The experiments would be set up at 4°C with using an initial protein concentration of 10 mg/ml in 0.01 M potassium phosphate buffer at the indicated pH with 0.5% β-octylglucoside.

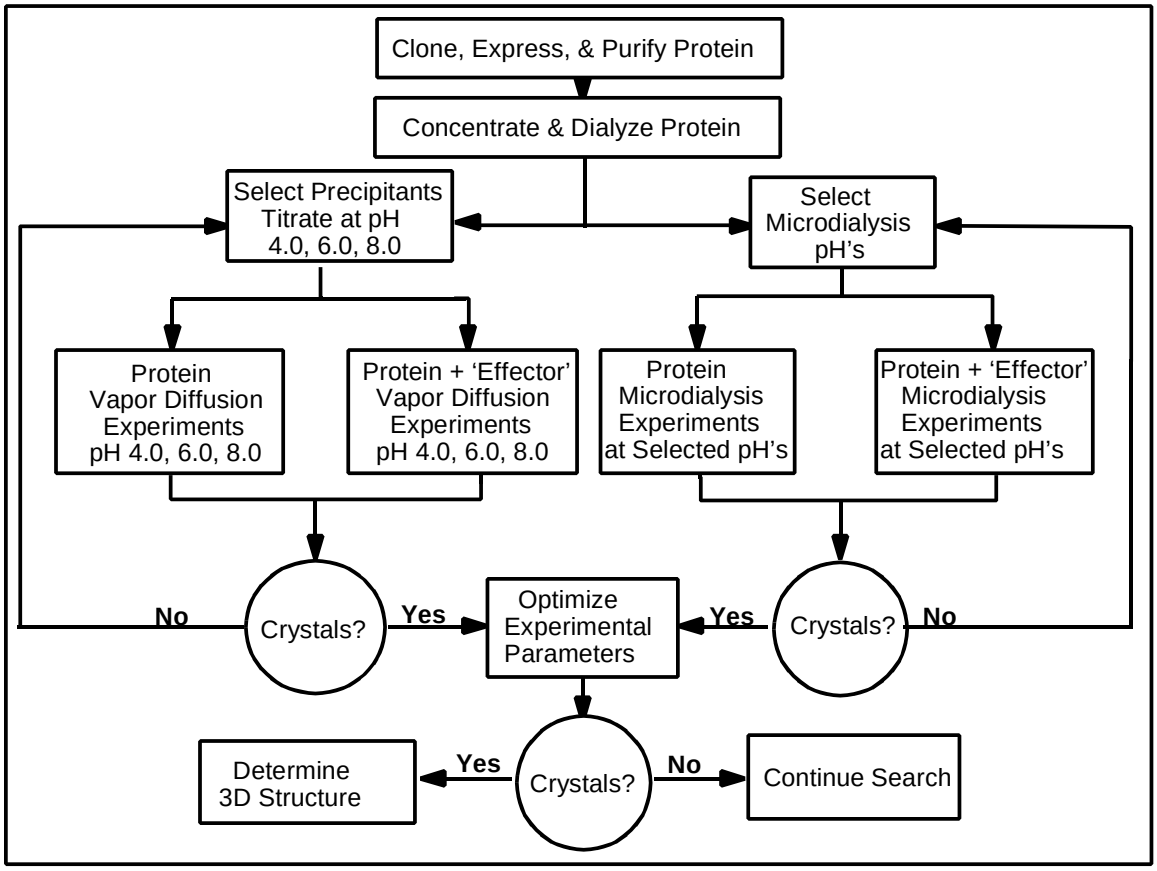

Figure 6 A general strategy for finding the crystallization conditions for a recombinant protein based on data in the BMCD.

For example, if it were desired to reproduce the crystals of rat liver glutathione S-transferase shown in Figure 3 [22] (entry M0P3/C13R shown in Figures 1 and 2), a series of initial hanging-drop experiments bracketing the reported conditions might be carried out. The protein was crystallized using the hanging-drop vapor diffusion technique carried out at 4 °C. The protein was concentrated to 11.3 mg/ml in 0.01 M potassium phosphate buffer at pH 6.9 with 0.46% (w/v) β-octylglucoside. Crystals were grown over a broad concentration range of precipitant, 60 to 74% saturated ammonium sulfate. An initial set of experiments shown in Figure 5 that vary pH from 6.7 to 7.1 and the ammonium sulfate concentration from 59 to 74% is suggested. In the first experiments the temperature and concentrations of the protein, phosphate buffer and β-octylglucoside would be fixed at the reported values. Depending upon the results of these first trials, additional experiments varying the temperature and the concentrations of protein and the additives would be considered. Once crystals are found, further experiments that use a finer grid of pH and ammonium sulfate concentration, vary temperature, protein and additive concentrations should be carried out to optimize the crystal growth procedure.

5.2 Variant Macromolecule or Ligand-Biological Macromolecule Complexes

Once the structure of a biological macromolecule has been determined, probing the relationship of structure to function of a macromolecule requires the structures of sequence variants, chemically modified macromolecules, or ligand-biological macromolecule complexes. Frequently the altered biopolymer sequence, chemical modification, and presence of the ligand do not interfere with crystal packing interactions, or significantly alter the conformation, or influence the solution properties of the macromolecule. If this is the case, the crystallization conditions of the biological macromolecule may be similar to the wild type, unmodified or the unligated biological macromolecule. The recent structure determination of 5 single-site mutant structures of the rat liver glutathione S-transferase [24] provides a good example. All five of the mutant proteins were crystallized under nearly identical conditions to the wild type protein (M0P3/C13R). However, this is not always the case. Crystallization problems may arise that produce unsuitable crystals or prevent the protein from crystallizing, and a search for unique conditions may be required such as that outlined below.

5.3 Homologous Biological Macromolecules

Crystallographic studies are frequently initiated with a biological macromolecule related through sequence (and presumably structural) homology to a set of macromolecules that have previously been crystallized. Analysis of the tabulation of the crystallization information for the members of a macromolecule family in the BMCD can be used to limit the parameters for the initial crystallization trials. For example, the crystallization conditions for the 114 Fab crystal entries in Version 3.1 of the BMCD include the reagents listed in Table 2. Perhaps not surprisingly the crystals are grown using a limited set of precipitants: sulfate and phosphate salts, 2-methyl-2,4-pentanediol, and a variety of polyethylene glycols. Polyethylene glycols have appeared to have the most success. About two thirds of the crystal forms of the immunoglobulin fragments are produced using one form or another of polyethylene glycol. Crystals are grown over a pH range of 3.6 to 8.5, a protein concentration range of 3.0 to 23.8 mg/ml, and a temperature range of 4 to 22.5 °C. Thus, these reagents and parameters represent a logical starting point for initiating a limited set of crystallization trials. If no crystals were obtained using this information, the more general procedure outlined next would be warranted.

Alternatively, the reagents listed in Table 2 and parameters could be used to develop a "fast screen" specific for Fab's. Most of the experiments for such a screen would use polyethylene glycol as the precipitant, 0.1 M buffer at various pH's from 3.6 to 8.5, and with and without 0.1 to 0.3 M sodium chloride present. A limited number of other experiments that use sulfate or phosphate salts and 2-methyl-2,4-pentanediol as the precipitants, and other atoms would also be added to increase the chances of success. The number of experiments with specific reagents and pH's would be based on the frequency of which they appear in the BMCD data set. A modification

	Chemical Reagent	No.Molecules	No.Crystals		Concentration (M)		pH		
Precipitants	ammonium sulfate	26	29		0.80	3.50			
	magnesium sulfate	1	1		1.70				
	2-methyl-2,4-pentanediol	5	5		15%	47%			
	phosphate	1	2		1.80		7.0		
	polyethylene glycol 300	1	1		?				
	polyethylene glycol 400	1	1		?				
	polyethylene glycol	1	1		15.0%				
	polyethylene glycol 600	2	4		2.0%	39.0%			
	polyethylene glycol 1000	1	1		24.0%				
	polyethylene glycol 3350	6	6		10.0%	30.0%			
	polyethylene glycol 3400	1	1		32.0%				
	polyethylene glycol 4000	11	12		4.0%	20.0%			
	polyethylene glycol 6000	7	14		12.0%	25.0%			
	polyethylene glycol 8000	17	27		4.8%	22.5%			
	polyethylene glycol 10000	3	5		9.0%	17.0%			
	potassium phosphate	1	1		1.20		8.8		
	sodium-potassium phosphate	1	1		1.5				
Buffers	ammonium acetate	2	2		0.02	0.20	5.5		
	bis-Tris	1	1		0.01		5.5	6.5	
	cacodylate	2	2		0.014	0.05	4.2		
	citrate	2	2		0.005		4.5	7.0	
	HEPES/sodium HEPES	2	2		0.01	0.05	7.2	7.5	
	imidazole/imidazole-HCl	6	6		0.05	0.10	7.0	7.4	
	imidazole-citrate	2	2		0.15	0.20	6.5	6.6	
	imidazole-malate	6	14		0.10	0.20	5.5	8.1	
	MES-NaOH	1	1		0.003		5.5	6.5	
	MOPS	1	1		0.025		6.8	7.0	
	phosphate buffer		4	7		0.04	0.10	5.8	8.0
	PIPES	1	1		0.023		6.0		
	potassium phosphate	12	19		0.002	0.1	5.4	8.5	
	sodium acetate/acetate	9	9		0.03	0.10	3.6	5.6	
	sodium cacodylate	6	7		0.005	0.10	5.6	6.5	
	sodium phosphate	8	11		0.01	0.10	5.7	8.0	
	Tris	8	10		0.013	0.10	6.5	8.5	
	Tris-maleate	2	2		0.10		3.5	6.5	
Additives	p-aminobenzenearsonic acid		1	3		0.004	0.010		
	ammonium sulfate	1	1		0.01				
	calcium acetate	1	1		0.02				
	copper (II) chloride	2	2		0.003				
	dioxane	2	2		6.0%				
	ethanol	1	1		2.0%				
	glycerol	2	2		5.0%	20.0%			
	lithium chloride	1	1		0.40	0.87			
	manganese (II) chloride	1	1		0.05				
	2-mercaptoethanol	1	1		0.002				
	2-methyl-2,4-pentanediol	1	1		5%				
	β-octyl glucoside	1	1		0.005%				
	phosphate	1	2		0.10		5.0	7.0	
	polyethylene glycol 200	1	1		5.0%				
	polyethylene glycol 600	1	1		2.0%				
	polyethylene glycol 10000	4	6		0.01%	17.0%			
	polyethylene glycol 20000	1	1		0.01%				
	sodium azide	17	28		0.003	0.2%			
	sodium chloride	20	31		0.02	0.30			
	sodium citrate	3	3		0.025	0.10	6.0		
	t-butanol	1	1		2.0%	3.0%			
	zinc sulfate	2	2		0.015	0.03			

Table 2 Chemical reagents used in the crystallization of Fab fragments recorded in the BMCD.

of the incomplete factorial experimental design protocol of Carter and Carter [3] would have applicability in producing a few experiments that adequately represent the observed data. The reagent concentrations for such experiments would be based on the upper limit of concentrations found in the BMCD.

5.4 General Crystallization Procedure

A general procedure for the crystallization of a unique biological macromolecule that has never been crystallized has been developed that incorporates the use of the BMCD [6,7]. The procedure for a soluble recombinant protein is illustrated in Figure 6. Once a protein is cloned, expressed, and purified, it is concentrated (if possible) to at least 10 mg/mL and preferably to a concentration of 25 mg/mL or higher. It is then dialyzed into 0.005 to 0.025 M buffer at neutral pH, or at a pH required to maintain the solubility of the protein. Stabilizing agents such as EDTA and dithiothreitol may be included at low concentrations. Next, two different types of experimental procedures are carried out concurrently, one employing the vapor diffusion method, and the other microdialysis.

After selecting from three to ten of the reagents that have been most successful at inducing crystallization from the data summary in the BMCD (Figure 4), vapor diffusion experiments are set up. The data in the BMCD is also used to set the limits for parameters such as pH and temperature. A distribution of the number of crystal forms for all biological macromolecules versus pH and temperature is shown in the panels of Figure 7. The procedure illustrated in Figure 6 uses three different pH's, 4.0, 6.0 and 8.0, and experiments are done at both 20 and 6 °C corresponding to room and coldroom temperatures, respectively. Before setting up the vapor diffusion experiments, the protein's precipitation point is determined by adding small aliquots of concentrated buffered reagents (0.1 M buffer) to a small droplet (10 µL) of the protein in a stepwise manner [4]. The droplet is sealed in an airtight chamber and observed after 10 min after each addition of 0.5 to 1.0 µL of the reagent. Distilled water can be added to the droplet to attempt to redissolve the protein if precipitation occurs. The titration experiments should be done at both room and coldroom temperatures to set the limits of reagent concentrations for the vapor diffusion experiments.

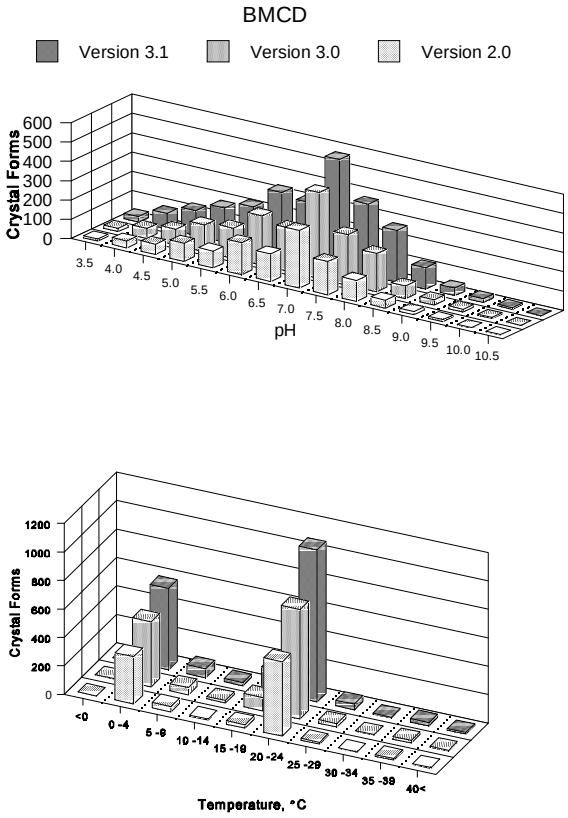

Figure 7 The distribution of the number of crystal forms in Versions 2.0 3.0, and 3.1 of the BMCD as a function of (top) pH and (bottom) temperature.

After the precipitation points have been found, two sets of vapor diffusion experiments covering the range of reagents and pH conditions are initiated at both room and coldroom temperatures. Droplets of the protein solution (5 µL) are mixed with equal volumes of a reservoir solution and placed in sealed chambers that allow the droplets to equilibrate with 0.5 to 1.0 mL of reservoir solutions. At each of the selected pHs, successive experiments are initiated that increase the precipitant concentration to a value greater than that required to

precipitate the protein in the titration experiment. For example, if a protein precipitates at 50% saturated ammonium sulfate, experiments that equilibrate droplets against 44, 47, 50, 53, and 55% saturated ammonium sulfate are appropriate. Equivalent vapor diffusion experiments with protein solution containing an effector (ligand, cofactor, metal ion, etc.) can be carried out in parallel.

An analysis of the data contained in the BMCD reveals that about 10% of the soluble proteins crystallize at low ionic strength (< 0.2 M) [5]. The procedure outlined in Figure 6 suggests that a set of microdialysis experiments equilibrating the protein solutions against low ionic strength should be carried out. The initial experiments should be carried out over a pH range of 3.0 to 9.0 in steps of 0.5 to 1.0, again, at both room and coldroom temperatures. Microdialysis experiments should also be carried out at, or near, the protein's isoelectric point, a point at which a protein is often the least soluble. The introduction of small quantities of ligands, products, substrate, substrate analogs, monovalent or divalent cations, organic reagents, etc. to the crystallization mixtures may facilitate crystal growth. Thus, it is recommended that a second set of experiments that include additives be run in parallel to those described above.

Once the experiments have been set up, periodic monitoring of the droplets is required. Crystals may appear within a few hours, or it may take up to many months. Further crystallization trials are suggested if no crystals appear within three weeks. Another set of reagents should be selected from the BMCD for the additional vapor diffusion experiments, and the titrations carried out before setting up a new set of experiments. Alternatively, experiments with finer increments in reagent concentrations, pH and/or other temperatures can be done. For the microdialysis experiments additional experiments also at finer increments of pH and at other temperatures should be carried out. Other additives or effectors should also be considered for both the vapor diffusion and microdialysis experiments.

If crystals are discovered, the parameters that are influencing crystal growth should be optimized. This requires additional experiments that introduce small perturbations in the pH, temperature, and the reagent and protein concentrations. Seeding experiments should also be considered to improve crystal size or reproducibility. Optimization may be necessary to improve the diffraction quality of the crystals. If the crystals diffract well, the three-dimensional structure determination can be initiated, but if the crystals do not diffract, or diffract poorly, the search should be continued.

6 The Future of the BMCD

The BMCD's primary goal will continue to be an error-free and up-to-date crystallization database and NASA PCG Archive. The availability of the BMCD on the Internet will facilitate the distribution and error correction of the data. Alternate ways for users to search and display the data will continue to be added as the software develops. Direct links to related entries in the BMCD and other related databases are being developed. Software tools that will help the user in designing the experiments for a particular biological macromolecule are being developed. These tools will incorporate the principles outlined in the discussion of crystallization strategies above.

References

[1] A. McPherson, Jr., "The Growth and Preliminary Investigation of Protein and Nucleic Acid Crystals for X-Ray Diffraction Analysis," Methods Biochem. Anal., 23, pp. 249-345, 1976.

[2] T. L. Blundell and L. N. Johnson, Protein Crystallography, Academic Press, New York, 1976.

[3] C. W. Carter, Jr. and C. W. Carter, "Protein Crystallization Using Incomplete Factorial Experiments," J. Biol. Chem., 254, pp. 12219-12223, 1979.

[4] A. McPherson, Preparation and Analysis of Protein Crystals, John Wiley, New York, 1982.

[5] G. L. Gilliland and D. R. Davies, "Protein Crystallization: the Growth of Large-Scale Single Crystals," Methods Enzymol., 104, pp. 370-381, 1984.

[6] G. L. Gilliland, "A Biological Macromolecule Crystallization Database: A Basis for a Crystallization Strategy," J. Crystal Growth, 90, pp. 51-59, 1988.

[7] G. L. Gilliland and D. Bickham, "The Biological Macromolecule Crystallization Database: A Tool for Developing Crystallization Strategies," Methods: A Companion to Methods in Enzymology, Vol. 1, pp. 6-11, New York: Academic Press, 1990.

[8] Stura, E. A., A. C. Satterthwait, J. C. Calvo, D. C. Kaslow and I. A. Wilson, "Reverse screening," Acta Cryst. D 50, pp. 448-455, 1994.

[9] Carter, C. W. Jr. (1990). " Efficient factorial designs and the analysis of macromolecular crystal growth conditions."

Methods: A Companion to Methods in Enzymology Vol. 1, pp.12-24. New York: Academic Press.

[10] Shieh, H.-S., Stallings, W. C., Stevens, A.M. & Stegeman, R.A. (1995). " Using sampling techniques in protein crystallization." Acta Cryst. D51, 305-310.

[11] Kingston, R., Baker, H. M. & Baker, E. N. (1994). " Search and Design for Protein Crystallization Based on Orthogonal Arrays. " Acta Cryst. D51, 429-440.

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

[13] Cudney, B., Patel, S., Weisgraber, K., Newhouse, Y. & McPherson, A. (1994). " Screening and optimization strategies for macromolecular crystal growth. " Acta Cryst. D50, 414-423.

[14] Zeelen, J. Ph., Hiltunen, J. K., Ceska, T. A. & Wierenga, R. K. (1994). " Crystallization experiments with 2-enoyl-coA hydratase, using an automated 'fast-screening' crystallization protocol. " Acta Cryst. D51, 443-447.

[15] 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.

[16] 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). " The protein data bank: A computer-based archival file for macromolecular structures." J. Mol. Biol. 112, 535-542.

[17] Gilliland, G. L.,Tung, M.,Blakeslee, D. M. & Ladner, J. (1994). " The Biological Macromolecule Crystallization Database, version 3.0: New features, data, and the NASA archive for protein crystal growth data." Acta Cryst. D50, 408-413.

[18] Comer, D. & Stevens, D. (1994). Internet working with TCP/IP, Vol III.: Client Server Programming and Applications. Englewood Cliffs, NJ: Prentice Hall.

[19] Stonebraker, M., Hanson, E. & Hong, C.-H. (1986). " The design of POSTGRES," in Proc. 1986 ACM-SIGMOD Conference on Management of Data, Washington, D.C. .

[20] Schatz, B. R. & Hardin, J. B. (1994). " NCSA Mosaic and the World Wide Web: global hypermedia protocols for the Internet." Science, 265, 895-901.

[21] Webb, E. C. (1992). Enzyme Nomenclature 1992, San Diego: Academic Press.

[22] 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.

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

[24] Xiao, G., Liu, S., Ji, X., Johnson, W. W., Chen, J., Parsons, J .F., Stevens, W. W., Gilliland, G. L. & Armstrong, R. N. (1996). " First-sphere and second-sphere electrostatic effects in the active site of a class mu glutathione transferase." Biochemistry, 35, 4753-4765.