Acta Cryst. (1995). **D51**, 305–310

Using Sampling Techniques in Protein Crystallization

BY HUEY-SHENG SHIEH, WILLIAM C. STALLINGS, ANNA M. STEVENS AND RODERICK A. STEGEMAN

Searle/Monsanto, BB4K, Department of Medicinal and Structural Chemistry, 700 Chesterfield Parkway North, Chesterfield, MO 63198-0001, *USA*

(*Received* 30 *March* 1994, *accepted* 14 *October* 1994)

Abstract

The crystallization of homogeneous or highly purified macromolecules depends on many variables such as precipitant, pH, choice of buffer, protein concentration, temperature, the participation of different mono- and divalent ions, as well as the presence of minute amounts of detergent and organic molecules. Finding the best combination among these many parameters is a multi-variable optimization problem. This kind of problem can be treated mathematically by sampling techniques. We have used this technique for protein crystallization. The iterative procedure starts with random sampling, followed by quantitative evaluation and cycles with weighted sampling. A simple procedure, derived from this concept and called MON48, has been successfully applied to many protein crystallization problems.

Introduction

The three-dimensional structures of macromolecules are important in understanding their biological functions. It is now well recognized that a rate-limiting step in crystallographic structure determination is often the growth of well ordered large single crystals; this has become the focus of a number of national and international meetings [see *Journal of Crystal Growth* (1991). Vol. 110]. Details of protein crystal growth are not well understood and involve many different non-independent variables. These include the choice of precipitant, buffer and temperature as well as more subtle factors such as the presence of detergents, other organic molecules or specific inorganic cations and anions. Traditionally, all these factors are explored systematically until a favorable combination, optimizing crystal growth, is identified; this process consumes not only time, but also large quantities of purified protein which, even with recombinant DNA techniques, are not always available.

Mathematically, protein crystallization can be considerd a multi-variable optimization problem. This kind of problem can be treated by random sampling or Monte Carlo techniques. Carter & Carter (1979) originated this concept and applied the incomplete factorial search method in protein crystallization. The Carters' original work has been extended and successfully implemented in many laboratories (Betts, Frick, Wolfenden & Carter, 1989; Carter, 1990; Abergel, Moulard, Moreau, Loret, Cambillau & Fontecilla-Camps, 1991; Jancarik & Kim, 1991). We have used the Carters' incomplete factorial method to develop an approach which is more systematic, automatic and iterative. Our mathematical treatment comprises approaches similar to those presented earlier by Carter & Carter (Carter, 1990, 1992), but was developed independently. A simple procedure which is based on these concepts has resulted; we call this procedure MON48 which requires 2 mg of protein with 48 setups.

The whole procedure consists of a random sampling method, a quantitative evaluation procedure and a weighted sampling method. It can be divided into nine steps.

(1) Selection of a crystallization variable table.

(2) Use of the random sampling technique to generate a set of trial conditions.

(3) Adjustments for insolubility or immiscibility problems which may be encountered in any trial condition.

(4) Setup of the trial crystallization conditions.

(5) Use of a defined scoring system to record the results of the trials.

(6) Use of a least-squares procedure to calculate the weight of each variable.

(7) New experiments to re-sample the variables based on their weights.

(8) Iteration of steps 3 to 7 until suitable crystals are obtained.

(9) Entry into a crystallization database management system for record keeping and data analysis.

A system of Fortran programs that handle different steps has been written. With the user-friendly program *XDBAM* as the crystallization database management system, crystallization conditions and results can be stored efficiently, retrieved rapidly and analyzed systematically.

The entire procedure is simple to carry out and easily adapted to additional crystallization variables. It has been designed to facilitate automatic iteration of experimental results. Most importantly, starting with the conditions described in Table 1 (see below), it has been used successfully to grow crystals of more than 20 proteins since it was introduced in the middle of 1991.

ISSN 0907-4449 ©1995

Table 1. *Protein crystallization variables selected for MON48*

Protein concentration: 10 mg ml^{-1} Temperature: 296 K

Salts
(1) 0.2 *M* Calcium acetate [$Ca(OAc)_2$]
(2) 0.2 *M* Lithium sulfate (Li_2SO_4)
(3) 0.2 *M* Magnesium chloride ($MgCl_2$)
(4) 0.2 *M* Sodium chloride (NaCl)
(5) 0.2 *M* Zinc acetate [$Zn(OAc)_2$]

Buffers
(1) 0.1 *M* Phosphate/citrate (P/C) pH 3.0
(2) 0.1 *M* Acetate (ACET) pH 4.5
(3) 0.1 *M* Citrate (CIT) pH 5.5
(4) 0.1 *M* 2-(*N*-morpholino)ethanesulfonic acid (MES) pH 6.0
(5) 0.1 *M* Cacodylate (CACO) pH 6.5
(6) 0.1 *M* Tris-(hydroxymethyl)aminomethane (TRIS) pH 7.0
(7) 0.1 *M* *N*-2-Hydroxyethylpiperazine-*N'*-2-ethanesulfonic acid (HEPES) pH 7.5
(8) 0.1 *M* Imidazole (IMID) pH 8.0
(9) 0.1 *M* TRIS pH 8.5
(10) 0.1 *M* 2-(*N*-Cyclohexylamino)ethanesulfonic acid (CHES) pH 9.5
(11) 0.1 *M* 3-(Cyclohexylamino)-1-propanesulfonic acid (CAPS) pH 10.5
(12) 0.1 *M* CAPS pH 11.0
(13) 0.1 *M* Di-/tribasic phosphate (P/P) pH 11.5

Precipitants
(1) 5% Isopropanol (iPrOH)
(2) 35% 2-Methyl-2,4-pentanediol (MPD)
(3) 10% Polyethylene glycol (PEG) 3350
(4) 10% PEG 8K
(5) 50% Saturated ammonium sulfate (SAS)
(6) 1.0 *M* Ammonium phosphate [$(NH_4)_3PO_4$]
(7) 2.5 *M* Sodium bromide (NaBr)
(8) 50% Saturated sodium citrate (SSC)

Procedure

The choice of crystallization variables is a difficult step. While the goal is to explore as many conditions as possible, the focus should be on conditions suitable for a particular crystallization problem; thus, we avoid conditions known to precipitate the sample or cause its denaturation. The crystallization variables can be chosen based on biochemical data accumulated for a particular protein, or from previous experience with other proteins (Abergel *et al.*, 1991). An example of a variable set is given in Table 1 which also defines the chemical abbreviations used throughout. In this variable set, protein properties are not considered; it serves as a very simple and non-specific approach and is also the basis of MON48. There are three variables: precipitants, buffers, and additional salt additives. The temperature and protein concentration are held constant in the trials. For each variable, there are a series of factors: five for salt additive, 13 for buffer and eight for precipitant.

A computer program, *RANSET*, employing a random-number generator is used next to select one factor for each variable to form one condition. A set of trial conditions is created in this manner. Additionally, the program also ensures that all factors within a variable are generated with nearly equal frequencies. The set MON48, with 48 conditions, was created in this way (Table 2). Some adjustments were required to keep some combinations of salt, buffer and precipitant in solution.

Table 2. *RANSET result for MON48*

No.	Salts	Buffers	Precipitants	Adjustments
1	None	P/P 11.5	10% PEG3350	
2	NaCl	P/C 3.0	5% iPrOH	
3	$MgCl_2$	CAPS 10.5	50% SAS	30% SAS
4	$Ca(OAc)_2$	ACET 4.5	2.5 *M* NaBr	
5	Li_2SO_4	HEPES 7.5	50% SAS	
6	NaCl	CAPS 10.5	2.5 *M* NaBr	
7	$MgCl_2$	MES 6.0	2.5 *M* NaBr	
8	Zn $(OAc)_2$	MES 6.0	5% iPrOH	
9	$Ca(OAc)_2$	CACO 6.5	2.5 *M* NaBr	
10	Li_2SO_4	ACET 4.5	2.5 *M* NaBr	
11	None	CAPS 10.5	10% PEG 8K	
12	Li_2SO_4	TRIS 7.0	10% PEG 8K	
13	NaCl	CAPS 11.0	35% MPD	
14	None	TRIS 7.0	5% iPrOH	
15	$MgCl_2$	P/P 11.5	50% SSC	
16	Li_2SO_4	P/C 3.0	2.5 *M* NaBr	
17	$Ca(OAc)_2$	TRIS 7.0	10% PEG 8K	
18	None	IMID 8.0	10% PEG 3350	
19	None	P/P 11.5	1.0 *M* $(NH_4)_3PO_4$	
20	None	CHES 9.5	50% SSC	
21	$Ca(OAc)_2$	TRIS 7.0	10% PEG 3350	
22	Li_2SO_4	IMID 8.0	10% PEG 8K	
23	$MgCl_2$	TRIS 8.5	35% MPD	
24	NaCl	IMID 8.0	1.0 *M* $(NH_4)_3PO_4$	
25	None	TRIS 8.5	1.0 *M* $(NH_4)_3PO_4$	
26	$Zn(OAc)_2$	ACET 4.5	10% PEG 8K	
27	$MgCl_2$	HEPES 7.5	50% SAS	30% SAS
28	$MgCl_2$	MES 6.0	50%SAS	30% SAS
29	Li_2SO_4	ACET 4.5	50% SAS	
30	$MgCl_2$	P/C 3.0	50% SSC	
31	$MgCl_2$	IMID 8.0	10% PEG 8K	
32	$Ca(OAc)_2$	MES 6.0	35% MPD	
33	$MgCl_2$	HEPES 7.5	35% MPD	
34	$MgCl_2$	HEPES 7.5	50% SSC	
35	None	CIT 5.5	1.0 *M* $(NH_4)_3PO_4$	
36	NaCl	TRIS 8.5	10% PEG 3350	
37	NaCl	TRIS 8.5	1.0 *M* $(NH_4)_3PO_4$	
38	NaCl	CHES 9.5	10% PEG 3350	
39	$Zn(OAc)_2$	CIT 5.5	50% SSC	
40	NaCl	CHES 9.5	1.0 *M* $(NH_4)_3PO_4$	
41	None	CHES 9.5	35% MPD	
42	None	P/C 3.0	5% iPrOH	
43	None	CACO 6.5	10% PEG 3350	
44	NaCl	CIT 5.5	50% SSC	
45	$Zn(OAc)_2$	CACO 6.5	5% iPrOH	
46	$Ca(OAc)_2$	CACO 6.5	5% iPrOH	
47	$Ca(OAc)_2$	CAPS 11.0	50% SAS	30% SAS
48	$Zn(OAc)_2$	CIT 5.5	5% iPrOH	

We used the hanging-drop method for each setup (McPherson, 1982). The protein stock is usually prepared at a concentration of 10 mg ml^{-1} with other necessary stabilizing reagents. The pH at which the stock solution is prepared is one at which the protein is known to be stable, active and soluble; the buffer concentration of the protein stock solution is kept low, usually 20 m*M*, relative to that of the reservoir where the buffer concentration is typically 0.1 *M*. A numbering scheme is used for scoring the results: the closer the result to a well ordered single crystal, the higher the score. An example of the scoring system is given in Table 3. The quantitative score for each experiment i, called Q_{oi}, can be approximated by a linear combination of individual factors, F_{ji}, in such a way that,

$$Q_{ci} = \sum_{j=1}^{14} \beta_j F_{ji}, \quad (1)$$

Table 3. *Crystal quality scores used for MON48*

1 Unfavorable (such as yellowish) precipitate or no precipitate (clear solution)
2 Phase separation
3 Regular granular precipitate
4 Birefringent precipitate or microcrystals
5 Rosettes or spherulites
6 Needles (growth in one dimension)
7 Plates (growth in two dimensions)
8 Single crystals (growth in all three dimensions) with no dimension larger than 0.2 mm
9 Single crystals (gowth in all three dimensions) with dimensions larger than 0.2 mm

where β_j is the weighting coefficient for factor j. Its sign and magnitude indicates the extent to which each factor contributes positively or negatively toward maximization of the score, Q. F_{ji} is the representation for each factor j used in experiment i. In the present case, there are 14 factors: five for the salts, one for the buffer and eight for the precipitants. F_{ji} for the salts and the precipitants are defined as $F_{ji} = 1$ (absence) and $F_{ji} = 2$ (presence). For the buffer, F_{ji} assumes values between 1 an 13 which are dependent on the identity of the buffer (Carter & Carter, 1979).

We use a least-squares procedure to solve for the 14 β's in the MON48 experiment. The quantity to be minimized is,

$$W = \sum_{i=1}^{48}(Q_{oi} - Q_{ci})^2. \qquad (2)$$

If we set the first derivative with respect to each weighting coefficient to zero, we will have 14 equations from (2),

$$\sum_{j=1}^{14} \beta_j \left(\sum_{i=1}^{48} F_{ki}F_{ji} \right) = \sum_{i=1}^{48}(Q_{oi}F_{ki}). \qquad (3)$$

Therefore, the 14 β's can be evaluated.

A computer program, *QCHK*, incorporates a routine to evaluate all the coefficients in (3) and a matrix-inversion routine to solve the equations. When β values from all of the factors are tabulated, the relative importance of each individual factor can be readily understood. Factors with positive β values should be enhanced in future experiments and ones with negative values should be reduced or eliminated. This is a branching point. If results are inconclusive, *e.g.* the initial scores in the first round of experiments are all less than 3, the same variables can be re-randomized with the conditions tried previously excluded, or a new variable table can be generated and sampled. If initial experiments indicate conditions where crystals with scores of 4 or higher grow, at least two approaches are possible. Simple inspection of the β values can suggest modifications of the initial conditions which would lead logically to growth of better crystals. However, an automatic and iterative process is proposed here. In the cases when large variable tables are used or when different concentrations of reagents are included as independent factors, an automatic iterative process using a weighted sampling technique may efficiently eliminate some of the experiments required to achieve optimal crystallization conditions. The inclusion of various concentrations of a fixed reagent will also have the advantage of providing the direction for fine tuning the crystallization parameters.

The philosophy of the program *RANSET* is to sample randomly crystallization variables; each factor has an equal weight and occurrence. In a weighted sampling mode, although randomness in the sampling remains a key component in the technique, the weight of each factor is also considered. A program *WTRAN* was written to include weights in the design of new crystallization scans. The weights are derived from the β values of preceding scans.

In *WTRAN*, we apply this idea in the following manner. If the weighing factor is arbitrarily set equal to 1 when β is equal to its minimum value $\beta_{\min}$, and equal to 11 when β is equal to its maximum value $\beta_{\max}$, then the weighting scheme can be set to,

$$\omega_j = \{[(\beta_j - \beta_{\min}) \times 10]/(\beta_{\max} - \beta_{\min})\} + 1. \qquad (4)$$

An integer value, representing the relative weight for each factor, is defined by rounding the right-hand side of the following,

$$n_j = \left[\omega_j / \left(\sum_{j=1}^{14} \omega_j \right) \right] \times 100. \qquad (5)$$

14 ranges of numbers 1 to 100 are then assigned to 14 factors, (1) to (n_1) for factor 1, ($n_1 + 1$) to ($n_1 + n_2$) for factor 2, ($n_1 + n_2 + 1$) to ($n_1 + n_2 + n_3$) for factor 3, ..., ($n_1 + n_2 + n_3 + \ldots + n_{13} + 1$) to 100 for factor 14. The range of the numbers assigned to each factor becomes representative of the factor's relative weight. Values from the random-number generator corresponding to these ranges then key the assignment of a particular factor to a new experiment in subsequent iterations. Hence, the occurrence of each factor depends on its weight.

This is by no means the only way to perform the weighted sampling technique. The number range, from 1 to 100, can be increased or decreased based on the number of parameters; the weight range can also be changed based on the range of observed β values. In an automatic iterative process, the weighted sampling step (*WTRAN*) replaces step 7 described in the *Introduction*. Manual intervention will be minimized in this step.

Another important part in protein crystallization is record keeping. Often, thousands of crystallization conditions can be explored before a condition which produces good crystals will emerge. Record keeping and analysis of the results becomes a significant component of the labor involved in a crystallization project. A program, *XDBAM* (crystal database analysis and management), was written to facilitate these tasks. *XDBAM* uses VT100 screen controlling routines to create a user-friendly menu-driven environment with screen-

editing capability. Database files can be either direct-access unformattted or ASCII-formatted files. Each data record contains 22 items: a crystallization code for archiving, the crystallization tray and well reservoir, the protein concentration, the identity of the protein preparation, the protein buffer, the precipitant identity and concentration, the buffer identity, the pH and buffer concentration, the identity and concentration of up to four additives, the temperature, the result (score) and the crystal code. The data file is automatically dated whenever it has been modified. Most of the commands can be applied at both group and individual levels; this allows for efficient data entry, retrieval, modification and reorganization.

Application

The MON48 procedure has been used in many protein crystallization experiments in our laboratory and elsewhere. Here we illustrate the general procedure and show a sample analysis for a typical MON48 crystallization experiment. The proteins are usually derived from recombinant-DNA techniques and purified extensively. The purity of the proteins in our laboraotry is judged by a single band on SDS–PAGE (sodium dodecyl sulfate–polyacrylamide gel electrophoresis) gels and one major band on IEF (isoelectric focussing) gels with both gels stained by Coomassie blue. The proteins are concentrated by microfiltration. The microconcentrators used are Centricon from Amicon and the refrigerated centrifuge was a Sorvall RT6000B from DuPont. With a Centricon size 30 (5×10^3 rev min^{-1}, at 277 K) 2 ml of protein solution can usually be concentrated to 25–50 µl in 30–60 min. The protein concentration is estimated by the absorbance at a wavelength of 280 nm with the calculated extinction coefficient based on the formula by Gill & von Hippel (1989) for denatured proteins.

The crystallization droplet consists of 4 µl of protein solution and 4 µl of the MON48 reservoir solutions. Since 48 droplets are used for the procedure, approximately 2 mg of protein are needed for each trial if the protein concentration is 10 mg ml^{-1}. In the particular procedure described below, the trays were all set up and allowed to equilibrate at ambient temperature ($\simeq$ 296 K). The trays are checked and recorded every 3–4 d. The resulting Q values are recorded at least two weeks after their set up. If no Q value is greater 3 at this step, it means a total failure and one should go back to re-randomize the conditions or change the crystallization variables and then randomize. However, in most of the cases tried in our own and in other laboratories, the procedure has usually yielded some form of crystalline material.

The Q values are then input into the program *QCHK* for least-squares fitting. The weight, β, of each factor is evaluated. There are two choices at this stage. The automated procedure, with the accumulated weighting information, can be used to re-randomize the conditions or, the weighting information can be used to manually refine the crystallization conditions toward their optimal values. Since the experiment (MON48) began with a small set of variables and a limited range of precipitant concentrations, the automatic reiterative procedure is not necessary. By using finer scans in the precipitant, pH and salt ranges, by adding minute quantities of organic solvents, by adjusting the protein concentration, and by allowing the experiments to equilibrate at new temperatures, optimal conditions for the growth of large well ordered single crystals are often identified.

We now show the results from a typical MON48 crystallization experiment where we have explored conditions suitable for the growth of crystals of the complex of bovine somatotropin and the extracellular domain of its receptor. Table 4 reports Q values from this trial and Table 5 shows the evaluation of the results. By analogy with procedures described by Carter & Carter (1979), comments based on the sign of β are attached to each factor to facilitate the analysis. The comments give indications as to which conditions should be explored in the next round of crystallization experiments. Although the β values tend to have high estimated standard deviations due to the limited number of experiments, Table 5 has indicated that $(NH_4)_3PO_4$ and PEG 3350 are the preferred precipitants and NaBr is the least desirable precipitant. Table 4 clearly indicates that finer scans around condition 36 are also appropriate.

Table 4. *Resulting crystal scores (Q) of bSTbp on MON48 protein crystallization setups*

Condition No.	1	2	3	4	5	6	7	8	9	10	11	12
Q	1	3	3	3	3	1	3	3	3	3	1	4
Condition No.	13	14	15	16	17	18	19	20	21	22	23	24
Q	3	3	4	4	4	4	4	3	4	3	3	4
Condition No.	25	26	27	28	29	30	1	32	33	34	35	36
Q	4	2	3	3	3	3	3	3	3	4	4	**8**
Condition No.	37	38	39	40	41	42	43	44	45	46	47	48
Q	4	3	4	3	3	3	3	3	3	3	3	3

Table 5. *bSTbp Crystallization remarks*

j	Factor	β_j*	Remark
1	$Ca(OAc)_2$	0.759 (0.569)	Presence is better
2	Li_2SO_4	1.043 (0.641)	Presence is better
3	$MgCl_2$	0.694 (0.510)	Presence is better
4	NaCl	0.576 (0.445)	Presence is better
5	$Zn(OAc)_2$	0.400 (0.618)	Presence is better
6	Buffer pH	−0.038 (0.052)	Lower pH is better
7	iPrOH	−0.169 (0.466)	Absence is better
8	MPD	−0.170 (0.528)	Absence is better
9	PEG 3350	0.727 (0.444)	Presence is better
10	PEG 8K	−0.674 (0.536)	Absence is better
11	SAS	−0.360 (0.550)	Absence is better
12	$(NH_4)_3PO_4$	0.930 (0.447)	Presence is better
13	NaBr	−0.930 (0.534)	Absence is better
14	SSC	0.115 (0.486)	Presence is better

* The values in parentheses are estimated standard deviations.

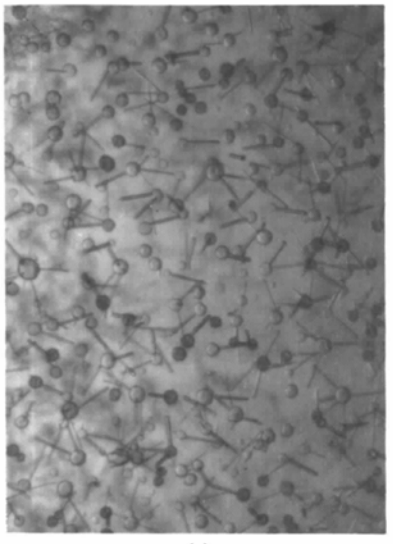
Often the tuned condition for large crystal growth is not very different from the MON48 condition. In the case of the complex of *C. albicans* myristoyltransferase (cNMT) with myristoyl coenzyme A (MCoA), small needle crystals grew from condition 17 of MON48 (Fig. 1*a*). When temperature, and precipitant and salt concentrations were adjusted, larger crystals of the complex grew (0.1–0.3 mm in each dimension, as shown in Fig. 1*b*).

Table 6 presents an extended crystallization variable list which covers more concentration ranges. If this variable list is used, we believe the reiterative procedure, using weighted sampling techniques, could eventually yield optimal conditions for crystal growth. In the cases where the crystallization variables are specified in different senses, such as ionic strength, positive and negative ionic radii as well as chaotropic and detergent properties, their weight evaluations should not be different from the way they are being used now.

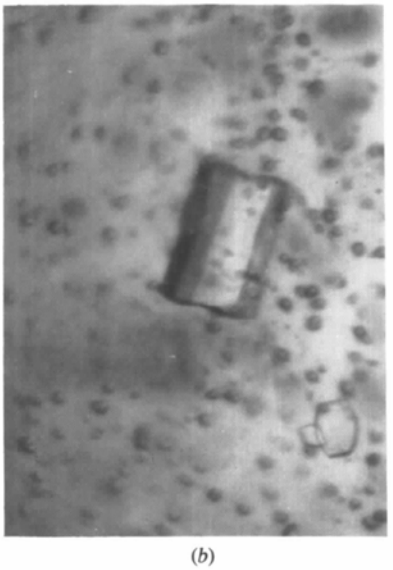
Table 6. *Extended list of variables for protein crystallization*

Protein concentration, 5, 10, 15, 20 mg ml^{-1}
Temperature, 277, 288, 296, 303 K
Buffers, same as given in Table 1

Salts	Precipitants
(1) 0.2 *M* $Ca(OAc)_2$	(1) 5% iPrOH
(2) 0.5 *M* $Ca(OAc)_2$	(2) 10% iPrOH
(3) 0.2 *M* Li_2SO_4	(3) 25% MPD
(4) 0.5 *M* Li_2SO_4	(4) 40 % MPD
(5) 0.2 *M* $MgCl_2$	(5) 8% PEG 3350
(6) 0.5 *M* $MgCl_2$	(6) 12% PEG 3350
(7) 0.5 *M* NaCl	(7) 8% PEG 8K
(8) 1.5 *M* NaCl	(8) 12% PEG 8K
(9) 0.2 *M* $Zn(OAc)_2$	(9) 30% Sat. SAS
(10) 0.5 *M* $Zn(OAc)_2$	(10) 50% Sat. SAS
	(11) 1.0 *M* $(NH_4)_3PO_4$
	(12) 1.5 *M* $(NH_4)_3PO_4$
	(13) 2.0 *M* NaBr
	(14) 3.5 *M* NaBr
	(15) 30% SSC
	(16) 50% SSC



(*a*)



(*b*)

Fig. 1. The crystals of cNMT.MCoA complex: (*a*) needle crystals directly from MON48 condition 17; (*b*) rectangular crystals from slightly modified condition [288 K, 0.4 *M* $Ca(OAc)_2$, 12% PEG 8K and 0.1 *M* MES pH 6.4–6.6].

Concluding remarks

Random sampling is a powerful approach in the solution of multi-variable problems. A simple procedure, MON48, has been developed by applying this concept to protein crystallization. With limited setups, a substantial number of crystallization variables can be examined and evaluated. In the past few years, this procedure has led to many protein crystals in Monsanto and Washington University laboratories. These include 5-enolpyruvoylshikimate 3-phosphate synthase, *N*-myristoyl-transferase, leukotriene A_4 hydrolase, bovine somatotropin receptor complex, bovine placental lactogen, a variant of interleukin-3 as well as sarcosine oxidase, trimethylamine dehydrogenase, hydroxyacid oxidase (Scott White & F. Scott Mathews, private communication) and domain I of *Ascaris* hemoglobin (Kloek, Yang, Mathews & Goldberg, 1993). Since this procedure was introduced at the 1991 meeting of the American Crystallization Association and at the 1991 Pitsburgh Diffraction Conference, good results have also been obtained at Cornell University (Jun Liang, private communiation) and Pittsburgh University (B. C. Wang, private communication).

To analyze the results of a typical MON48 crystallization experiment, we have employed a least-squares treatment familiar to crystallographers. Nevertheless, the

statistical significance of the results (Table 5) can be questioned. While an analysis of the variance could be used to validate the conclusions, our partially empirical approach does not rely on this. We have worked to keep the sample size small in order to allow for crystallization experiments with a limited supply of highly purified protein. Currently, perhaps the most valuable aspect of the MON48 procedure in our laboratory has been in the use of *RANSET* to generate new randomized crystallization scans which can be tailored to a particular protein.

In the future, when limitless supplies of correctly folded proteins can be produced routinely, human intervention may be minimized for the growth of good crystals as robotized systems could be used both to set up the crystallization experiments and to give optical feedback with automated scoring of the results. Systematic evaluation and analysis of the results would continue to provide the basis for the next crystallization experiment. This feedback mechanism could make the whole procedure strictly iterative and completely automatic.

Finally, this method should also be applicable to the crystallization of compounds other than proteins. By careful selection of starting crystallization variables to suit particular molecules (for example, oligonucleotides and nucleic acids almost always crystallize near neutral pH), one should be able to adapt this method to any crystallization. Indeed, the procedure is now being used by others (Helen Berman, private communication; Donald Voet, private communication) with modified crystallization variables for nucleic acids and oligonucleotides.*

We are indebted to Dr R. J. Fletterick of the University of California, for critically reading and reviewing this manuscript. We also thank Professor W. K. Shiue of Southern Illinois University, for stimulating discussions.

*The programs described in this paper are available from Huey-Sheng Shieh.

References

ABERGEL, C., MOULARD, M., MOREAU, H., LORET, E., CAMBILLAU, C. & FONTECILLA-CAMPS, J. C. (1991). *J. Biol. Chem.* **266**, 20131–20138.

BETTS, L., FRICK, E., WOLFENDEN, R. & CARTER, C. W. JR (1989). *J. Biol. Chem.* **264**, 6737–6740.

CARTER, C. W. JR (1990). *Methods – A Companion to Methods in Enzymology*. Vol. 1, pp. 12–24. New York: Academic Press.

CARTER, C. W. JR (1992) In *Crystallization of Proteins and Nucleic Acids: a Practical Approach*, edited by A. DUCRUIX & R. GIEGÉ, pp. 47–71. Oxford: IRL Press.

CARTER, C. W. JR, BALDWIN, E. T. & FRICK, L. (1988). *J. Cryst. Growth*, **90**, 60-73.

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

GILL, S. C. & VON HIPPEL, P. H. (1989). *Anal. Biochem.* **182**, 319–326.

JANCARIK, J. & KIM, S.-H. (1991). *J. Appl. Cryst.* **24**, 409–411.

KLOEK, A. P., YANG, J., MATHEWS, F. S. & GOLDBERG, D. E. (1993). *J. Biol. Chem.* **268**, 17669–17671.

MCPHERSON, A. (1982). *Preparation and Analysis of Protein Crystals*, p. 96. New York: John Wiley.