

Quantitative Analysis in the Characterization and Optimization of Protein Crystal Growth

BY CHARLES W. CARTER JR AND YUHUI YIN

Department of Biochemistry and Biophysics, Campus Box 7260, University of North Carolina at Chapel Hill, Chapel Hill, NC 27599-7260, USA

(Received 21 October 1993; accepted 31 January 1994)

Abstract

Protein crystal growth often depends on the combination of many different factors. Some affect protein solubility directly; others may act indirectly by causing conformational changes. Systematic characterization of these factors can be important for generating good crystals. It can also provide useful insight into the biochemical behavior of the protein being crystallized. Here we focus on statistical methods to achieve these two objectives. (1) Characterization of a protein system by analyzing patterns of crystal polymorphism under different levels of biochemical parameters, such as ligands and pH. Tests of the reproducibility of crystal growth experiments indicate that quantitative scales of crystal quality can be statistically significant. Analysis of variance for a replicated, full-factorial design in which four factors were tested at two levels has been used to demonstrate highly significant, biochemically relevant, two-factor interactions strongly implicating pH and ligand-dependent conformational changes. (2) Optimization of crystal growth *via* response-surface methods. 'Minimum predicted variance' designs provide for efficient response-surface experiments aimed at constructing quadratic models in several dimensions. We have used such models to improve crystal size and quality significantly for three forms of *Bacillus stearothermophilus* tryptophanyl-tRNA synthetase. In one case we can now avoid having to increase the size by repeated seeding, a difficult procedure that also produces unwanted growth of satellite crystals. Graphs of two-dimensional level surfaces reveal a number of ridges, where the same result is obtained for many combinations of the factors usually varied when trying to improve crystals. An important inference is that it may be better to sample simultaneously for the effects of protein concentration and supersaturation. For a system involving only one crystallizing agent, supersaturation can be approximated as the *product* of protein and precipitant concentrations. Use of this search direction significantly improves the performance of response-surface experiments. Advantages of growing crystals at stationary points of their response surfaces include better crystals and higher reproducibility, since crystal growth at stationary points is insulated from the deleterious effects of experimental fluctuations. This arises because the derivatives of the response are by definition zero with respect to the experimental variables.

Quantitative analysis of appropriately designed crystal growth experiments can thus be a powerful way to characterize complex and interacting biochemical dependencies in macromolecular systems and optimize parameters important to the crystallography.

1. Introduction

Crystallization of a macromolecule is a complex, multi-equilibrium process that depends on an active interplay between intermolecular forces in solution and the surface properties and conformation of the macromolecule (Weber, 1991). The best value for an experimental variable affecting these equilibria nearly always depends on the settings of other variables. Understanding the joint functional dependence of crystal growth on different factors is, therefore, an invaluable tool for making the best crystals. Designed factorial experiments and quantitative analysis of variations in crystal growth behavior provide a natural way to deal with this multidimensionality. The elements of experimental design and analysis described here are summarized in Table 1.

Any multidimensional experimental problem entails a *factorial design*, in which N different 'factors' are varied simultaneously in a series of $M > N$ experiments, in order to determine how they affect the result. By *quantitative analysis*, we mean estimating and interpreting parameters that minimize the sum of the squares of differences between observed results, referred to here as Q_{obs} , and predictions, Q_{calc} , based on an appropriate mathematical model for crystal growth. Mathematical models can be useful in three distinct phases of a crystal growth project: *screening*, *characterization* and *optimization*. Using such models entails an experimental design, and a quantitative score, Q_{obs} , attached to each experiment, as well as the model itself.

1.1. Experimental design

Different experimental designs, and different models are appropriate, depending on which objective is to be met (Table 1). Most straightforward is the full-factorial design. All possible combinations of the different factor levels are tested in a full-factorial experiment, so they are usually limited to a small number of factors that are often tested at only two different 'levels'. A full-factorial design is both necessary and sufficient to

Table 1. *Designed experiments for different contexts*

Definitions for terms used in the text, together with the three different contexts in which quantitative analysis has proven useful to us. The type of design matrix and mathematical model used for fitting the observed data are associated with specific objectives typically encountered in a crystal growth or other experimental project requiring identification of suitable conditions, characterization of the effects and optimization.

Factorial design:	M experiments with simultaneous variation of $N < M$ factors
Quantitative analysis:	Minimizing $\left\{ \sum_{j=1}^{M(\text{expts})} Q'_{\text{obs}} - Q'_{\text{calc}} \right\}$
Objective	Design/model
Detection of important main effects and interactions	Incomplete factorial
	$Q_{\text{calc}} = \beta_0 + \sum_i^{N(\text{factors})} \beta_i F_i + \epsilon \quad \text{(I)}$
Verification of these inferences	Replicated, full factorial
	$Q_{\text{calc}} = \beta_0 + \sum_i^{N(\text{factors})} \beta_i F_i + \sum_{i=1}^{N-1} \sum_{j>i}^N \beta_{ij} F_i F_j + \dots \text{higher terms} + \epsilon \quad \text{(II)}$
Optimization of crystal growth conditions	Response surface
	$Q_{\text{calc}} = \beta_0 + \sum_i^{N(\text{factors})} \beta_i F_i + \sum_{i=1}^{N-1} \sum_{j>i}^N \beta_{ij} F_i F_j + \sum_{i=1}^N \beta_{ii} F_i^2 + \epsilon \quad \text{(III)}$

estimate jointly and unambiguously the importance of synergistic effects between all possible combinations of factors. Full-factorial experiments (Carter, Doublie & Coleman, 1994) are among the most powerful ways to verify and characterize quantitatively synergistic effects between more than one factor.

Neither screening nor optimization are typically done with full-factorial designs. Rather, they involve differ-

ent strategies for *sampling* the space covered by the full-factorial design. A full-factorial screening design typically would involve many more experiments than are practical, and some subset must be chosen to sample it. Optimization, on the other hand, requires a different sampling strategy, whose objective is to construct an accurate functional model or 'response surface', for the behavior of the system. Because they require efficient sampling, screening and response-surface designs are often selected from a large number of potential designs generated by a computer program according to specific criteria, illustrated schematically in Fig. 1.

Efficient covering of the entire experimental space is a key requirement for screening. Incomplete factorial designs, also called 'sequence-levelled experimental designs' (SLED's, Carter, C. W., personal communication) were introduced to detect the most important factors and their interactions from screening experiments. Their use has been described previously (Carter & Carter, 1979; Carter, Baldwin & Frick, 1988; Carter, 1990, 1992). Factor levels are chosen randomly and then balanced to achieve nearly uniform sampling (Fig. 1a) and preserve the ability to detect large main effects and two-factor interactions with minimal confounding (Carter, 1992). We recently illustrated the power of these designs quantitatively by using them for phase permutation experiments in which the sampling fraction was roughly $1/(N^{1/2})$ (Doublie, Xiang, Bricogne, Gilmore & Carter, 1994)

A *response surface* is an analytical approximation to how the system actually responds to the input variables. It can be used to locate stationary points that may be optima, and hence to find the best conditions for a desired result. Response-surface experiments are sampled

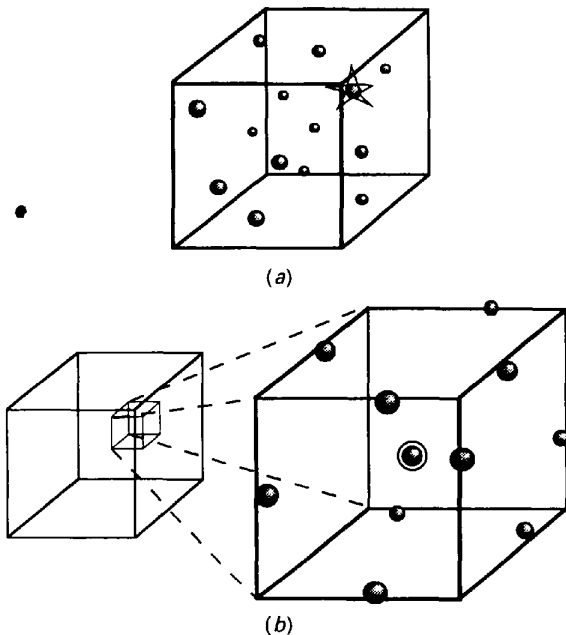


Fig. 1. A comparison between the experimental designs appropriate to (a) sampling an unknown experimental space (screening) and (b) modeling the response surface of an already partially characterized process (optimization).

by different criteria from those used in screening (Fig. 1b). If one already knows something about where the best result might be obtained, it is no longer sensible to scatter the experimental test points uniformly throughout the space. Therefore, points are selected that will have maximal impact on the accuracy of the parameters of the response-surface function and hence on the coordinates of stationary points.

The problem of how to choose the best points for the individual tests has received considerable attention recently in the mathematical literature (Conway & Sloane, 1987). One way to define 'maximal impact' is to require that the experimental points produce a model with a *minimum integrated prediction variance* (Hardin & Sloane, 1993), or a minimum scatter between the observed and calculated values of the dependent variable. It can be shown for a quadratic model that these consist of some number (one or more, depending on the total number of experiments) near the suspected optimum, together with others spread around close to the boundary of the experimental volume. Often the individual tests are placed in a hypercube (Fig. 1b), as was the case for our design (Table 3), which was prepared for us by Neil Sloane, using his program *GOSSET* (Hardin & Sloane, 1993). A simple analogy is that of fitting the slope and intercept of a straight line through a set of sample points, which works best if a majority of the sampling points lie close to the low and high ends of the interesting range of the independent variable, x_i . Since Hardin-Sloane designs minimize the integrated prediction variance, they also reduce the number of experiments necessary to fit the surfaces with a desired precision. Therefore, they maximize the efficiency of an optimization search much as incomplete factorial experiments do for screening experiments (Carter, Baldwin & Frick, 1988; Carter, 1990, 1992).

1.2. Scoring crystal growth: converting impressions into quantitative measures

To be quantitatively useful a score must faithfully reflect what we are looking for. Crystal growth experiments have a predictable variability. One quickly develops an intuitive feeling from the appearance of a precipitate that some conditions are 'close' to producing crystals, while others seem straight away to be disastrous. We have found that crystal growth experiments are quite reproducible, and that this intuition, although crude, can be surprisingly accurate. A simple approach to scoring is to *quantitate this intuition*. It is surprisingly robust. The most dangerous mistakes – underrating a microcrystalline precipitate and overrating a crystal with nice morphology but poor diffraction characteristics – can be avoided. All precipitates can be assayed by streak seeding (Stura & Wilson, 1990), and diffraction diagrams for all crystals will normally be characterized. Thus, consistent and accurate scoring

systems can be based initially on microscopic analysis and supplemented by further tests. More sophisticated assays for 'crystallizability', based on hydrodynamic measurements of the thermodynamics of aggregation behavior, have considerable potential utility (Kam, Shore & Feher, 1978; Baldwin, Crumley & Carter, 1986; Kadima, McPherson, Dunn & Jurnak, 1990; Mikol, Hirsch & Giegé, 1989, 1990; Feher & Kam, 1985; Wilson, 1990; George & Wilson, 1994).

Most experiments described here were evaluated using our earlier semi-quantitative system that distinguished three different categories of precipitates (floculent, granular and spherulites); three different categories of crystals, corresponding to dominant growth in one, two, or three dimensions; and three different levels of diffraction quality (Carter & Carter, 1979). We were able to utilize the full scale in most cases because most of the crystal forms have now been characterized rather thoroughly by X-ray diffraction.

Optimization experiments are far easier to score. Diffraction quality is highly correlated with crystal volume, asymmetry and uniformity. We scored our most recent response-surface experiments for each of these: crystal volume was estimated from the product of the three principal axis lengths, and the asymmetry ratio, W/L = width/length, was the ratio of the smallest dimension to the largest dimension measured with a microscopic reticle. Uniformity was assessed from the relative quality of the crystal faces and the presence or absence of satellite crystals.

1.3. Mathematical models and inference; main effects, interactions and response surfaces

Regression methods of statistical inference involve construction of a predictive model for the result, based on the contributions of the different factors. Model parameters are estimated by minimizing the sum of the squares of differences between Q_{obs} and Q_{calc} . We have used three types of models (Table 1). Models (II) and (III) are derived from the first by adding specific terms, which are indicated in bold face. The linear model used in our original study (Carter & Carter, 1979), included only main effects [Table 1, (I)]. For the full-factorial design in Table 2 we added all the multi-factor interaction terms [Table 1, (II)]. Quadratic models [Table 1, (III)] supplement the general linear model with all possible squared terms, together with the two-factor interactions. By substituting new variable names for the squared and interaction terms, these models become special cases of the linear model (I).

In all three models, ϵ is the residual error to be minimized, and the constant value, β_0 , is the mean value of all experiments in the design. Each β_i coefficient in the linear model (I) is the average amount by which the presence of factor, F_i , raises or lowers the score from the overall average. Higher-order β_i coefficients have

similar significance for models (II) and (III). Since many important parameters have non-linear effects it is not surprising that crystal growth behavior can be modeled more effectively by multivariate quadratic functions. These are the simplest functions that assume stationary points, which can be either maxima, minima, or saddle points, on interior points of the range of independent variables. Stationary points can be determined analytically by partial differentiation and equating the gradient to zero. Coordinates of stationary points that are optima provide estimates for the factor levels giving the best result. This is the basis of the response-surface method (Box, Hunter & Hunter, 1978).

1.3.1. *Replication and the analysis of variance.* Key to the exploitation of full-factorial and response-surface methods is the quantitative analysis of the different behaviors observed as experimental conditions are varied. In order to assess the significance of effects detected from the variation in a factorial experiment they must be compared with the variance of the experimental observations. These estimates are necessary for the analysis of variance (ANOVA; Box, Hunter & Hunter, 1978), which in turn provides an indication of which effects are statistically significant. In short, one wants to know the signal-to-noise ratio. An essential requirement for using ANOVA is a reasonable estimate for the experimental error, including the reproducibility of the scoring system. *Replication*, or carrying out each experiment two or more times, provides a direct estimate for the variances of individual tests. The quest for significance has led us increasingly to use carefully replicated experiments, in which each point is sampled with at least duplicate experiments.

1.4. *Summary of results*

Our purpose here is first to review work to be described elsewhere (Carter, Doublé, & Coleman, 1994) demonstrating the possibility of scoring crystal growth experiments quantitatively, and then to describe the use of Harden-Sloane response-surface designs to improve several inferior crystal forms. The examples are from the study of *Bacillus stearothermophilus* tryptophanyl-tRNA synthetase (TrpRS), a dimeric enzyme with identical subunits, $M_r = 37\,000$ Da (Barstow, Sharman, Atkinson & Minton, 1984).

A full-factorial experiment described previously (Table IV in Carter & Carter, 1979) showed that pH changes and the presence of different ligands led to extensive crystal polymorphism. Quantitative analysis of a twofold replication of that design showed that the quality of the crystals associated with this polymorphism was due primarily to the pH and to the synergistic effects of pH plus tryptophan and ATP plus tryptophan. Thus, ligand-dependent conformational changes are likely to be an underlying cause of the polymorphism. Molecular flexibility probably plays a key role in catalysis and/or

recognition for this enzyme family (Carter, 1993), so the analysis also helped to establish connections between crystal growth conditions and function.

Two different crystal forms of TrpRS have given us trouble with reproducibility, inadequate volume and/or unsuitable morphology. Using response-surface experiments we have shown that these difficulties arose because we were not growing them under optimal conditions. We have used a 20-experiment Harden-Sloane response-surface design for four variables to improve significantly our conditions for growing both difficult crystal forms and to characterize a third form, which had always given good crystals.

In the course of learning to use response-surface methods, we also realized that two aspects of what we were doing were likely to be generally useful. First, it turns out to be very useful to vary simultaneously the protein concentration and the *product* of the protein and precipitant concentrations as two of the four 'factors'. When there is only a single crystallizing agent, these represent the size of the initial 'reservoir' of soluble protein and the initial supersaturation. Using the Harden-Sloane matrix in this way significantly improved the performance, enabling us to detect the presence of an optimum condition well outside the range of supersaturation tested previously. Second, we realized that working at stationary points helps insulate crystal growth from deleterious and irreproducible experimental fluctuations.

2. Materials and methods

2.1. *Enzyme sources and crystallization conditions*

Enzyme used in the factorial experiment was obtained in a crude fraction (Atkinson *et al.*, 1979) or from a cloned source (Carter, 1988), purified (Carter & Green, 1982) and crystallized from 4.2 mg ml^{-1} protein solutions in $2.1\text{ M K}_2\text{HPO}_4$ at pH = 6.2 or 7.6, with combinations of the ligands tryptophan and ATP as described (Carter & Carter, 1979). Conditions for response-surface experiments to optimize individual crystal forms are described in §4.

2.2. *Reproducibility of crystal growth conditions*

Crystal growth depends on a number of factors that are especially hard to control experimentally. The pH of a solution containing a high salt concentration is difficult to measure accurately. A number of various factors that affect nucleation (and growth) are themselves quite likely to interact strongly. One example of the latter problem is the fact that the solubilities of crystals and the crystallizing agent, potassium phosphate, are both temperature dependent. Thus, it is hard to characterize accurately the absolute effects of any of these factors without making a complete phase-diagram analysis. For these reasons, several decisions were made at the outset

to limit the response-surface studies to a protocol that was as reproducible and well characterized as possible. The following aspects of this protocol are notable.

2.2.1. *pH control.* Individual phosphate stock solutions of the specified pH were made up at 3.0 M in $\text{KH}_2/\text{K}_2\text{H}$ phosphate by first determining empirically a nomograph relating the mole fraction of each salt and the pH that resulted when the 3.0 M stock was diluted to 2.1 M, as measured with single point standardization, and without intermediate washing of the pH electrode with deionized water between determinations for different buffers. All pH values were adjusted either with 12 N phosphoric acid or 12 N KOH, after they had been made up to volume (variations from the desired pH were never more than 0.02 pH units, and the changes in phosphate concentration caused by addition of the acid or base were negligible).

2.2.2. *Protein concentration.* All experiments were carried out in either 5 or 35 μl perspex dialysis cells obtained from Cambridge Repetition Engineers, Ltd (Cambridge, England). Use of dialysis, rather than vapor diffusion for equilibration, increased our control over reproducibility because both the pH and the refractive index could be monitored and controlled, and because the protein concentration remained constant. However, an important and irreducible source of variation in protein concentration was introduced by the variability of the indentation of the dialysis membrane into the well following application of the membrane and O-ring. It was assumed that this variation applied nearly equally to all concentrations in the designs. Recent work has shown that vapor diffusion poses no special difficulty (Lee Kuyper, unpublished results).

2.2.3. *Precipitant concentration.* All crystallizing solutions (approximately 2.1 M in phosphate) were titrated with deionized water immediately before setting up the experiments, in order to achieve a refractive index of 1.3735 (5) for all pH values. This meant that the molar concentration of salt varied smoothly with the pH, being somewhat higher at lower pH's and lower at higher pH's. For the third response-surface experiment this was not possible as the precipitant concentration was explicitly varied as a component of the composite independent variable, $[\text{protein}] \times [\text{precipitant}]$ (§4.2).

2.2.4. *Pre-incubation.* During the work, we found that a pre-incubation at sub-saturating salt concentrations was essential for optimal crystal growth. Accordingly, all experiments were dialyzed for 24 h at approximately half strength, before adding the final equilibrating solution.

2.2.5. *Temperature.* All experiments at the same temperature were carried out in the same Linbro plate and in environmental chambers at the stated temperature, except for room-temperature experiments, which were monitored several times daily for temperature fluctuations. These precautions could assure a temperature constancy of only $\pm 0.5^\circ\text{C}$.

2.2.6. *Other.* All experiments contained PMSF (0.1 mM), β -mercaptoethanol (0.2% by volume) and toluene (0.05% by volume) as previously described. Some of the microphotographs were generously taken for us by Professor Peter Petrusz in the UNC Department of Cell Biology and Anatomy.

2.3. Statistical calculations

Calculations were carried out using the computer programs SYSTAT (Wilkinson, 1987) and Theorist (Bonadio, 1990) as implemented on a Macintosh SE computer. Examples are summarized in Tables 2 and 5. Levels of the four factors of the full-factorial design in Table 2 were coded as -1 for the lower level (low pH or temperature, absence for tryptophan or ATP) and 1 for the higher level.

2.3.1. *Contrasts.* Preliminary examination of a factorial experiment involves calculating contrasts for main effects and interactions. A contrast is defined as the average difference between those experiments treated at the two levels.* For two-factor interaction effects (columns E-J) the contrast is the difference between the average effects of one factor at the two different levels of the other factor (Box, Hunter & Hunter, 1978; Carter, 1990). A large two-factor interaction contrast indicates likely non-additive or synergistic effects between the two factors. Higher-order interaction contrasts (columns K-P) are also pairwise comparisons between multi-factor products having the same parity, since the signs in columns E-P are the products of the appropriate signs in columns A-D.

2.3.2. *Statistical significance.* Although easy to calculate, the statistical significance of the contrast sums themselves cannot be determined directly from their magnitudes; it must be evaluated by comparing their magnitudes with the experimental error. This can be done conveniently by examining the corresponding analysis of variance (ANOVA) tables. ANOVA has a simple geometrical representation if we consider the ensemble of the M experimental results as a vector in M dimensions, whose elements are the experimental results, $\{Q_{\text{obs},j}\}$. The corresponding vector of calculated values, $\{Q_{\text{calc},j}\}$, is decomposed into an average vector, $\{Q_{\text{ave},j}\}$, and a treatment vector, $\{Q_{T,j}\}$, whose components are the contributions calculated from the model. The F -ratio test compares the lengths of $\{Q_{T,j}\}$ and the residual vector, $\{Q_{R,j}\} = \{Q_{\text{obs},j} - Q_{\text{calc},j}\}$, given by the corresponding treatment and error 'sums of squares' corrected for the number of degrees of freedom (Box, Hunter & Hunter, 1978). The P value is the probability of

* An equivalent definition of the contrast, and the one used in SYSTAT, is the average difference between experiments treated at the high level and the overall average for all experiments. The contrasts in Table 2 were obtained by multiplying the average score in column 5 for each treatment by the sign in the appropriate column (A-P), summing and dividing by 8, which is the number of treatments at each of the two levels.

Table 2. 2⁴ Factorial experiment for tryptophanyl-tRNA synthetase ligand complex crystals

The design is keyed to the results in Fig. 2 via the letters in parentheses after the experiment number. Experimental scores for two independent replicates from different dates are given in columns *Q* and *R*, the mean and variance in columns *S* and *T*, and contrasts for all effects and interactions with their *F* ratios and probabilities under the null hypothesis in the bottom three rows.

Expt	A pH	B Temp	C Trp (<i>W</i>)	D ATP	E pH × Temp	F pH × <i>W</i>	G pH × ATP	H Temp × <i>W</i>	I Temp × ATP	J <i>W</i> × ATP	K pH × Temp × <i>W</i>	L pH × Temp × ATP	M pH × <i>W</i> × ATP	N Temp × <i>W</i> × ATP	P pH × Temp × <i>W</i> × ATP	Q <i>Q</i> * 2/78	R <i>Q</i> † 1/91	S (<i>Q</i>)	T Var‡
1 (<i>a</i>)	-	-	-	-	+	+	+	+	+	+	-	-	-	-	+	5	5	5.0	0.0
2 (<i>b</i>)	-	+	-	-	-	+	+	-	-	-	+	+	-	+	-	7	7	7.0	0.0
3 (<i>i</i>)	-	-	-	+	+	+	-	+	-	-	-	+	+	+	-	6	6	5.0	2.0
4 (<i>j</i>)	-	+	-	+	-	+	-	-	+	-	+	-	+	-	+	3	2	2.5	0.5
5 (<i>e</i>)	-	-	+	-	+	-	+	-	+	-	+	-	+	+	-	2	3	2.5	0.5
6 (<i>f</i>)	-	+	+	-	-	-	+	+	-	-	-	+	+	-	+	2	3	2.5	0.5
7 (<i>m</i>)	-	-	+	+	+	-	-	-	+	+	+	+	-	-	+	6	6	6.0	0.0
8 (<i>n</i>)	-	+	+	+	-	-	-	+	+	-	+	-	-	+	-	1	3	2.0	2.0
9 (<i>c</i>)	+	-	-	-	-	-	-	+	+	+	+	+	-	-	-	8	8	8.0	0.0
10 (<i>d</i>)	+	+	-	-	+	-	-	-	+	+	-	-	+	+	+	6	6	6	0.0
11 (<i>k</i>)	+	-	-	+	-	+	-	+	-	-	+	-	-	+	+	6	3	4.5	4.5
12 (<i>l</i>)	+	+	-	+	+	-	+	-	+	-	-	+	-	-	-	6	6	6.5	0.5
13 (<i>g</i>)	+	-	+	-	-	+	-	-	+	-	-	+	-	+	+	7	4	5.5	4.5
14 (<i>h</i>)	+	+	+	-	+	+	-	+	-	+	-	-	-	-	-	9	9	9.0	0.0
15 (<i>o</i>)	+	-	+	+	-	+	+	-	-	+	-	-	+	-	-	9	9	9.0	0.0
16 (<i>p</i>)	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	9	9	9.0	0.0
Contrast	3.13	-0.13	0.13	-0.13	1.0	1.75	0.25	0.0	-1.0	1.75	0.88	1.13	-0.13	-0.88	-1	5.75	5.5	5.625	0.938
ANOVA																			
<i>F</i> ratio	83.3	0.13	0.13	0.13	8.53	26.13	0.53	0.0	8.53	26.13	6.53	10.8	0.13	6.53	8.53				
Probability	9.6	0.720	0.720	0.720	0.010	6.7	0.476	1.000	0.010	6.7	0.021	0.005	0.720	0.021	0.010				
	× 10 ⁻⁸					× 10 ⁻⁵				× 10 ⁻⁵									

* Experimental scores taken from Carter & Carter (1979).
 † Experimental scores taken from a replicate experiment performed as described in the text.
 ‡ The sample variance, $s^2 = 1/(N - 1) \sum (Q_i - \bar{Q})^2$.

achieving an *F* ratio of a given magnitude under the null hypothesis, i.e. that the modeled contribution is insignificant. Therefore, it provides an estimate for the significance of the model.

Analysis of variance for the data in Table 2 was carried out using the four factors as categorical, rather than as continuous variables. Replication of the design (16 experiments × 2 = 32 tests) provided an additional 16 degrees of freedom for the error sum of squares, permitting a much better estimate for the experimental error. This improved error estimate was crucial. It had the effect of reducing significantly the denominator of the *F* ratios, making it possible to identify as significant the two two-factor interactions in columns *F* and *J*. ANOVA calculations for response-surface experiments are similar in principle, but the independent variables are continuous, rather than categorical.

2.3.3. Multiple comparisons. Whenever more than one significance test is being applied, and especially when many are applied at the same time, individual tests will tend to be distributed normally. So when comparing *F* ratios for many effects, chances increase that one or more tests will be significant at a given confidence level even under the null hypothesis. Protection against concluding that a factor is significant when it is actually insignificant can be secured by strengthening the decision criterion. One strengthened criterion, the Bonferroni criterion (Neter & Wasserman, 1974) involves dividing the *P* value by the number of effects compared.

2.3.4. Selection of models. The mathematical models in Table 1 are general, in the sense that they include all possible terms. In practice, not all terms are significant in a given situation, and a more useful and significant model can usually be found by eliminating some terms. Selecting the best set of independent variables is essentially the problem of identifying only real effects, so as not to fit the experimental error. For this reason it is typically one of the most difficult tasks in regression analysis (Neter & Wasserman, 1974). It involves balancing the predictive power (the multiple correlation coefficient, *R*, and its square) against the statistical significance (the probability value of the *F* ratio) estimated from the analysis of variance. We do this first by a stepwise regression algorithm, which selects the factors one at a time in the order of their impact on *Q*_{calc}, adjusting the fitted residual at each step to reflect the contributions to the observed variance from the factors already selected. Stepwise algorithms can also be run in the reverse direction by eliminating factors with the smallest impact on the model. Models resulting from this process can subsequently be refined manually, by adding or deleting factors based on *t* tests. An example is provided §4.1.

2.3.5. Examining response surfaces. The quadratic surfaces evaluated from the models of type (III) (Table 1) provide a powerful window on crystal growth. The complete response surfaces are multi-dimensional and cannot be represented easily except by choosing constant values

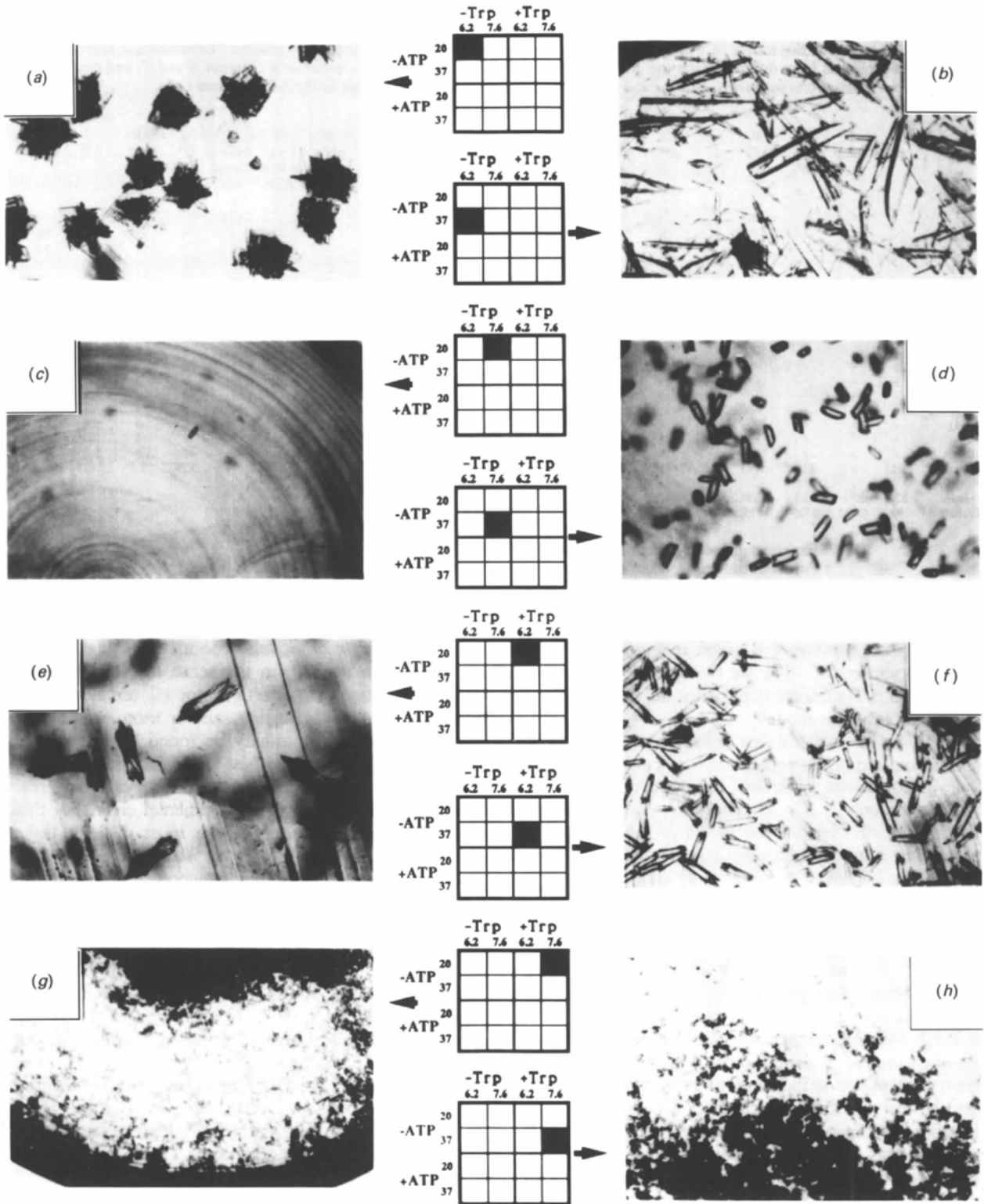


Fig. 2. Microphotographs of the 16 different results from the second replicate of the 2^4 full-factorial experiment summarized in Table 1. Factor treatments are coded by the accompanying diagrams, as indicated by bold arrows to either side. Experiments are grouped according to the ligands present: (a)-(d) had no ligand; (e)-(h) had only tryptophan; (i)-(l) had only ATP; and (m)-(p) had both tryptophan (2 mM) and ATP (5 mM).

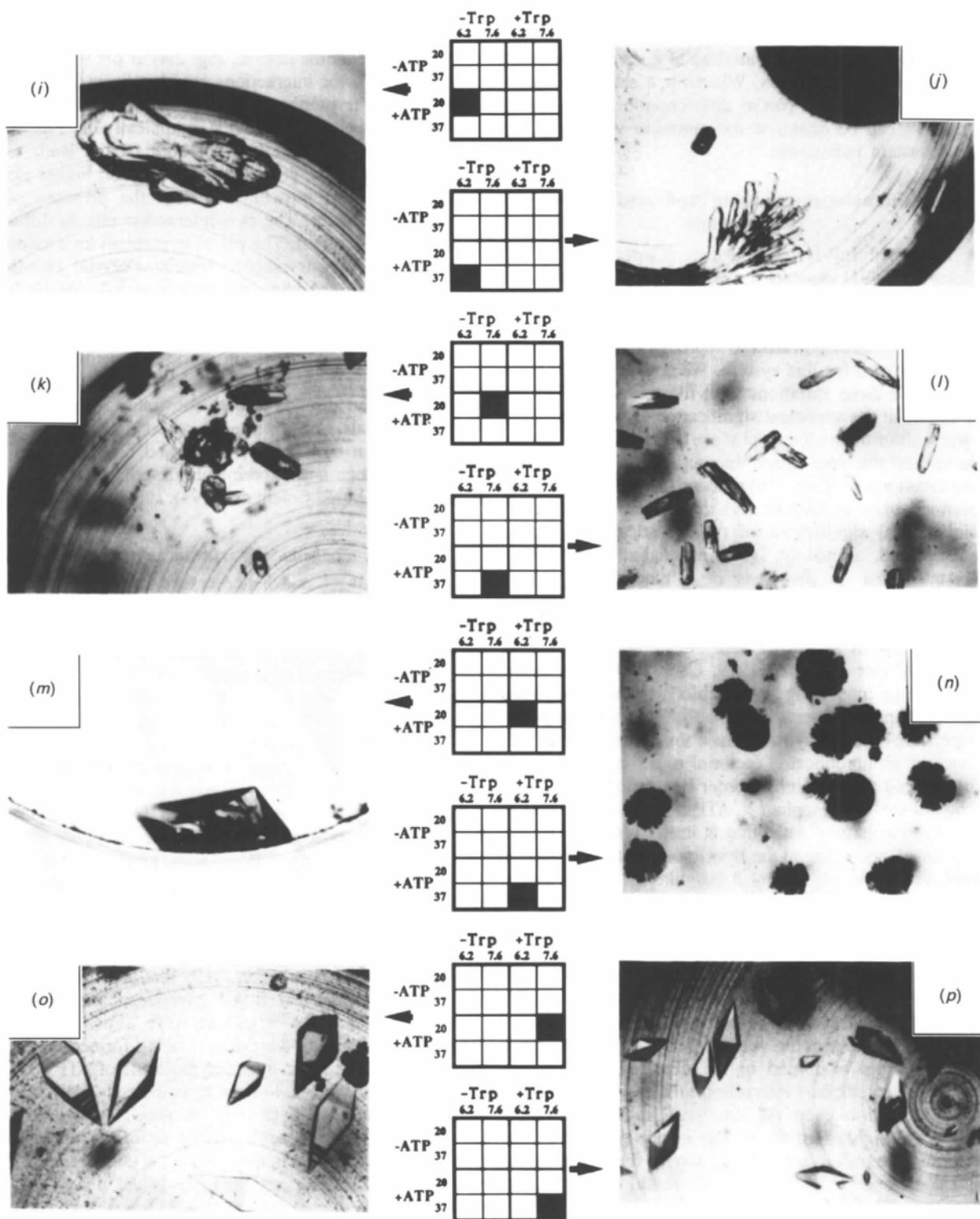


Fig. 2 (cont.) Within each of these four blocks, the two top figures were obtained at pH = 6.2, the two lower ones at pH = 7.6, and the two on the left were obtained at 293 K, while the two on the right at 210 K. Hence, for any particular ligand the pH effect can be assessed by comparing the two rows, whereas the temperature effect can be assessed by comparing the two columns. The magnification is such that the sides of the coding box represent almost exactly 1 mm.

for some of the variables (for example, the coordinates of stationary points) and plotting the dependent variable against two independent variables at constant levels for the remaining parameters. Wherever a stationary point can be identified by partial differentiation these 'level surfaces' can be drawn at the optimum values for the two constant parameters.

3. Characterization using replicated factorial experiments

A replicated full-factorial design (Carter, Doublie & Coleman, 1994) showing the effects of pH, temperature, tryptophan and ATP on the crystallization of TrpRS is shown in Fig. 2 and Table 2. The key observation is that these four parameters lead to reproducible variations in crystal growth for this system. We sought to define the sources of these variations with five goals in mind: (i) to establish the statistical significance of all 15 possible effects (four main effects, plus six two-factor, four three-factor and one four-factor interactions); (ii) to identify important main effects; (iii) to identify synergistic effects between pairs of factors; (iv) to evaluate the potential biochemical significance of the statistically important effects; and, ultimately, (v) to rationalize the different crystal forms on the basis of a biochemical model consistent with the available data.

The two two-factor interactions highlighted in Table 2 stood out in an earlier analysis of the unreplicated design (Carter & Carter, 1979; Carter, 1990). Because they are so interesting from a biochemical standpoint, we were quite interested in establishing their statistical significance. One way to estimate the experimental variances in an unreplicated factorial experiment is to use the contrast for the highest order interaction. The pH \times temperature \times tryptophan \times ATP interaction in Table 2 has a rather large F ratio that is itself significant at the 85% level, according to the Bonferroni criterion. Since it may actually be significant it provides an unreasonably large estimate for the true experimental error. So we repeated the experiment as exactly as possible in order to obtain a more realistic estimate of the error.

3.1. Detecting synergistic effects

TrpRS crystal growth depends in a highly significant way only on a specific biochemically relevant subset of the factors examined in the experiment. The pH, the pH \times tryptophan interaction and the tryptophan \times ATP interaction exert the strongest influences. These effects are highly significant. The F -ratio probability for rejecting the null hypothesis for any of the 15 contrasts at a confidence level of, say, 0.99 would be 0.01/15 or 0.00067. All three of the effects highlighted in Table 1 are, therefore, significant at better than the 0.999 level (their P values are all less than or equal to 6.7×10^{-5}). A good illustration of the pH main effect is shown in Figs. 2(m)–2(p). Large, uniform, tetragonal type IV

crystals grow reproducibly at pH = 7.6 (Figs. 2o and 2p), irrespective of the temperature, but only at one of the two temperatures (293 K, Fig. 2m) at pH = 6.2.

Both two-factor interactions are significant independently of the tryptophan and ATP main effects, which are small. They are illustrated graphically in Fig. 3. In both cases, the high level of both factors leads to the best crystals: for growth with tryptophan higher pH is better, and for growth with ATP the presence of tryptophan is better. The two interaction effects differ qualitatively, however. The pH \times tryptophan interaction is such that adding tryptophan improves crystal growth only at high pH. Adding tryptophan at low pH leads to poorer crystals. The synergistic effect is therefore in the same direction as the main effect of pH. The ATP \times tryptophan interaction works differently; either substrate by itself has an unfavorable impact on crystal growth, and the absence of both substrates gives almost as good crystals as the presence of both together. The two interaction effects were confirmed by supplementary observations and experiments described elsewhere (Carter, Doublie & Coleman, 1994).

3.2. Drawing biochemically relevant conclusions

Lattice contacts, and hence crystal integrity, are apt to be sensitive to conformation changes. There is considerable evidence that the significant effects identified in Table 2 arise indirectly, *via* effects of pH and ligand binding on conformational equilibria.

3.2.1. *Dependence of type IV crystal growth on the synthesis of Trp-2'(3')-ATP.* Tryptophan and ATP act synergistically to produce the best TrpRS crystals (Fig. 3b). From an early stage it appeared that enzyme-catalyzed Trp-2'(3')-ATP synthesis was essential for the growth of type IV crystals (Coleman & Carter, 1983). Inhibiting acyl transfer with inorganic pyrophosphate prevented growth of type IV crystals, producing instead small crystals with a type II* habit. Pyrophosphate would be expected to promote pyrophosphate exchange of bound tryptophanyl adenylate *via* reversal of amino-acid activation, thereby competing with acyl transfer. We verified this conclusion with the enzyme in solution by showing that 2 mM pyrophosphate inhibits acyl transfer to tRNA^{Trp} and to ATP. Thus, in cases where tryptophanyl-5'-adenylate is being formed on the enzyme, acyl transfer to a second molecule of ATP must occur to produce enzyme-bound tryptophanyl-2'(3')-ATP in order for type IV crystals to grow. In order to affect crystal growth, acyl transfer probably causes a conformational change in the enzyme. A conformational change accompanying acyl transfer is the most likely explanation for the strong Trp \times ATP interaction.

3.2.2. *pH dependence of type I, type II* and type IV crystal stability.* In keeping with the strong pH \times Trp interaction, the growth and diffraction quality of the different crystal forms have markedly different

sensitivities to pH changes. Type II* crystals containing tryptophan can be grown and are stable only above about pH = 6.8. Below pH = 6.5 they shatter violently. Type IV crystals, on the other hand, grow over a wide pH range and are stable to heavy-atom derivatization in the pH range 5.2–7.8, provided the ATP is replenished weekly. Deterioration of type II* crystals at low pH may also result from the effect of pH on the protein conformation. Tryptophanyl ATP, on the other hand, remains bound to type IV crystals at high pH (Coleman & Carter, 1983), consistent with the stability of these crystals both at low and high pH.

The strong pH \times Trp interaction is also consistent with evidence that pH is linked to conformational changes related to substrate-binding affinity in tRNA synthetases (Hélène, Brun & Yaniv, 1971; Lam & Schimmel, 1975). For the valyl and isoleucyl enzymes, (both class I enzymes, as is TrpRS) lower pH favors conformations with higher affinity for tRNA substrates; higher pH favors conformations with higher affinity for amino acid. These precedents provide a natural rationalization for the pH \times Trp interaction and, in particular, the fact that the best crystals containing tryptophan grow at high pH. Protons (low pH) may stabilize an enzyme conformation compatible with the enzymology of acyl transfer, *i.e.* that found in type IV crystals, and that conformation may be incompatible with type II* lattice packing. These effects, reinforced by analysis of crystal packing in the different crystal forms (Carter, Doublé & Coleman, 1994), provide a coherent rationalization for the different TrpRS polymorphs in terms of ligand- and pH-linked conformational equilibria related to catalysis.

These conclusions confirm our original premise that quantitative analysis of crystal polymorphism caused by different, functionally significant factors can provide biochemical insight. They emphasize the usefulness of

Table 3. *Hardin–Sloane minimum integrated variance design matrix for four factors*

This design was prepared specifically for use in the experiments reported here by N. J. A. Sloane, using *GOSSET* (Hardin & Sloane, 1993). Matrix entries should be interpreted as: 0 = center, -1 = the low end, and 1 = the high end of the variable range. The same design has been used repeatedly in different contexts, by assigning the matrix entries to different parameters and/or ranges. The three response-surface experiments described here were carried out as described in the accompanying Table 4.

Experiment	Variable 1	Variable 2	Variable 3	Variable 4
1	0.000	-0.056	0.000	-0.250
2	0.000	-0.056	0.000	-0.250
3	0.000	1.000	0.000	-0.250
4	0.000	-1.000	0.000	-1.000
5	1.000	-0.007	0.116	1.000
6	-1.000	-0.007	-0.116	1.000
7	0.210	0.108	-1.000	-1.000
8	-0.210	0.108	1.000	-1.000
9	-1.000	1.000	1.000	-0.250
10	1.000	1.000	-1.000	-0.250
11	1.000	-1.000	-1.000	-0.250
12	1.000	1.000	1.000	-0.250
13	0.492	-1.000	1.000	1.000
14	-0.492	-1.000	-1.000	1.000
15	-1.000	1.000	0.577	-1.000
16	1.000	1.000	-0.577	-1.000
17	0.669	1.000	-1.000	1.000
18	-0.669	1.000	1.000	1.000
19	-1.000	-1.000	-1.000	-1.000
20	1.000	-1.000	1.000	-1.000

systematic evaluation of macromolecular crystal growth conditions, especially in cases where such polymorphic behavior is evident.

4. Optimization with response-surface experiments and quadratic models

A central goal in studying macromolecular crystal growth is to produce the best possible crystals for high-resolution diffraction studies. Our previous work showed that it is possible to identify important factors

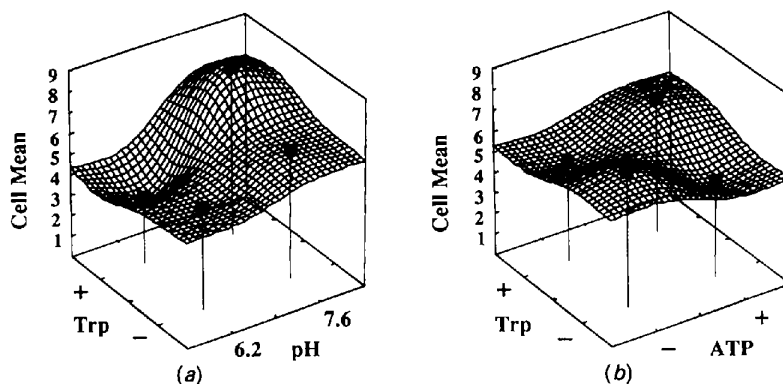


Fig. 3. Graphic representation of the significant two-factor interactions from Table 1. Cell mean refers to the average score for the four cells from Table 2 having the same combination of signs for the two factors. These surfaces are *not* analytical functions. They were prepared using an inverse-squared distance-smoothing algorithm provided by the three-dimensional graphing module in *SYSTAT* (Wilkinson, 1987) to fit the four plotted cell means. (a) The tryptophan \times ATP interaction. The presence of tryptophan leads to significantly better crystals at high pH and to significantly worse crystals at low pH. (b) The tryptophan \times ATP interaction. Growth in the presence of only one of the two substrates leads to significantly poorer crystals than can be grown when both are present or when both are absent.

Table 4. Variable and range assignments for response-surface experiments

Experimental parameters, together with their high and low levels, chosen for individual response-surface experiments performed using the matrix in Table 3.

Type I experiments	Type IV experiments	Type II* experiments
6.40 (-1) < pH < 7.2 (1)	7.20 (-1) < pH < 8.0 (1)	7.5 (-1) < [Prot][PPNT] < 9.5 (1) (mg ml ⁻¹)(M)
4.4 (-1) < [Prot] < 5.2 (1) (mg ml ⁻¹)	3.8 (-1) < [Prot] < 4.6 (1) (mg ml ⁻¹)	4.0 (-1) < [Prot] < 6.4 (1) (mg ml ⁻¹)
0.25% (-1) < [PEG] < 1.0% (1)	0.25% (-1) < [PEG] < 1.0% (1)	0.35% (-1) < [PEG] < 0.42% (1)
287 K (-1) < Temp < 303 K (1)	287 K (-1) < Temp < 303 K (1)	287 K (-1) < Temp < 310 K (1)

influencing crystal growth by using factorial screening experiments. Work reviewed in the previous section showed that our quantitative scoring system was reproducible enough that important effects could be demonstrated convincingly by statistical analysis of a replicated full factorial experiment. The surprisingly small mean sample variance in column *T* in Table 2, 0.938, suggested that our scoring system might be used to optimize crystal growth by response-surface methods (Box, Hunter & Hunter, 1978).

At the same time it became apparent that several of the different TrpRS crystal polymorphs exhibited serious crystal growth problems that impeded structural studies. Tetragonal (type IV) crystals of the selenomethionine analog of the synthetase, which we were using for phase determination, were always smaller than corresponding crystals of the native protein. Optimization in this case might produce larger crystals of the Se-Met protein. Triclinic (type I) crystals, grown from ligand-free enzyme, were prone to polymorphism of a different order, producing at least three different habits, none of which could be thoroughly characterized, because crystals were too small and grew irreproducibly. Finally, the monoclinic form (type II*; Carter, Crumley, Coleman, Hage & Bricogne, 1990), presented two different problems. First, they never grew larger than about 20 × 20 × 200 μm, about 100 times smaller than is required for X-ray photography or data collection. Second, repeated macroseeding (Thaller *et al.*, 1981, 1985) led to extensive growth of satellite crystals on the seed, greatly lowering the effective yield of good crystals.

For these reasons, we decided to try to optimize the crystal growth conditions using response-surface methods. We were greatly stimulated by Neil Sloane, who provided the experimental design after pointing out to us the important differences between factorial experiments used when screening for unknown determinants and those used for optimization, once a clear idea about the important determinants of the response surface is available (Fig. 1).

4.1. Response surfaces for the tetragonal and triclinic crystal forms.

4.1.1. Use of the design matrix. Specifications for 20 sampling points in our response-surface experiments are

encoded in the Harden-Sloane (H-S) design matrix in Table 3, which exhibits the properties illustrated in Fig. 1(b). The point [0,0,0,0] represents the mean value, and +1 and -1 the highest and lowest levels used for each factor. There are two identical experiments essentially at the origin and 18 others distributed within the faces of a hypercube. This particular matrix was generated with the restriction that the temperatures coded in column 5 are 287, 294 and 303 K because these temperatures were accessible in our laboratory.

Similar response-surface experiments were carried out for all three TrpRS crystal forms. The independent variables for the quadratic surface to be modelled for crystal forms I and IV included the pH, temperature, protein concentration, and the concentration of an additive, polyethylene glycol (PEG 400), previously suspected to be beneficial for growth of single type IV crystals (Coleman, 1988; Carter, Doublie & Coleman, 1994). The assignments are summarized in columns 1 and 2 of Table 4. Temperature and PEG concentration were varied over the same ranges. The two experiments differed, however, in the specific ranges of the pH and protein concentration, because the analysis of the complete factorial experiment in Table 2 had shown that lower pH's were better for crystals grown in the absence of tryptophan.

4.1.2. Optimal quadratic models. The type I and type IV experiments were performed first, almost simultaneously; they are summarized in Tables 5 and 6, and Figs. 4-6. Experimental scores similar to those used for the full-factorial design were initially used to estimate parameters for the complete quadratic model (III). However, in both cases, it became clear that selecting subsets of variables improved the statistical significance (*F* ratio) without decreasing the multiple correlation coefficient for prediction. Both models reproduce the observed variability very well - the multiple regression coefficients and their squares are close to 1.0 - and are statistically very significant - the *F* ratios have very low probability under the null hypothesis.

The response-surface experiment for type I crystals is presented below, step by step, and in some detail to illustrate the process of building and analyzing the model.

The selection and range of independent variables assigned to the H-S matrix are given in the first column

Table 5. Regression statistics and analysis of variance for the type I Hardin-Sloane response surface experiment

The dependent variable, Q_{obs} , was fitted to calculated values, Q_{calc} , based on the model in Table 6 for the 20 experiments in Table 3. Coefficients of the model are given in column 2. The model provided an excellent set of predictions, superimposed on the observed values in Fig. 4. The correlation coefficient between Q_{obs} and Q_{calc} is 0.958, and the standard error of the estimated Q_{calc} is 0.821 or about 10%. The coefficients are all significantly larger than their standard errors, as indicated by the individual t -tests (Student's t = coefficient/standard error), whose probability values are given in the last column. The F ratio given in the ANOVA table and its P value are similar tests referring to the model as a whole. To illustrate choosing between competing models, when the $[Protein]^2$ term was included with this model, the correlation coefficient changed very little ($0.958 \Rightarrow 0.959$), but the F ratio decreased to 12.645, and the t -test for the β_{C^2} had a P value of 0.7. All of these statistics improve significantly with the twice-replicated experiments with the modified experimental search directions and quantitative scoring used for the data in Table 6 and Fig. 7.

Parameter	Coefficient	Std. error	Std. coeff	Student's t	$P(2\text{-tail})$
β_0	7.29	0.40	0.000	18.07	0.16×10^{-8}
β_{pH}	0.67	0.26	0.24	2.61	0.024
β_C	-0.45	0.24	-0.17	-1.85	0.091
β_T	1.02	0.23	0.39	4.46	0.96×10^{-3}
β_{pH^2}	-1.32	0.43	-0.27	-3.08	0.010
β_{T^2}	-2.58	0.42	-0.57	-6.12	0.75×10^{-4}
$\beta_{pH \times C}$	1.34	0.29	0.43	4.58	0.79×10^{-3}
$\beta_{T \times C}$	-1.03	0.28	-0.33	-3.74	0.003
$\beta_{pH \times T}$	0.73	0.30	0.21	2.42	0.034

Analysis of variance					
Source	Sum-of-squares	Degrees of freedom	Mean-square	F ratio	P
Regression	83.53	8	10.44	15.47	0.57×10^{-4}
residual	7.42	11	0.68		

Table 6. Quadratic models for type I and type IV TrpRS crystal growth response surfaces

Models were selected following a stepwise multiple regression with the parameters of a complete quadratic model (III), followed by an empirical procedure of eliminating weak predictors to maximize the multiple correlation coefficient and its square while minimizing the probability, P , of the corresponding F ratio. The standard error of the estimate in these examples is based on a unitless score ranging from 0 to 9.

Type I model

$$Q = \beta_0 + \beta_{pH}pH + \beta_T T + \beta_C C + \beta_{pH^2}pH^2 + \beta_{T^2}T^2 + \beta_{pH \times T}pH \times T + \beta_{pH \times C}pH \times C + \beta_{T \times C}T \times C$$

Regression statistics		Analysis of variance			
R	R^2	Std error of estimate	Degrees of freedom	F ratio	p
0.958	0.918	0.821	11	15.474	0.6×10^{-4}

Type IV model

$$Q = \beta_0 + \beta_{PEG}PEG + \beta_T T + \beta_C C + \beta_{PEG^2 \times T}PEG^2 \times T + \beta_{C^2}C^2 + \beta_{PEG \times T}PEG \times T + \beta_{T \times C}T \times C$$

Regression statistics		Analysis of variance			
R	R^2	Std error of estimate	Degrees of freedom	F ratio	p
0.864	0.747	2.177	7	5.064	0.007

of Table 4. The 20 experiments were each performed only once in this case. They were scored using the same qualitative scheme used to score the experiments in the full-factorial experiment. The scores were input to a SYSTAT data file together with the H-S matrix.

Using the MGLH (multiple regression, general linear hypothesis) module in SYSTAT, tests were carried out

using the stepwise regression algorithm in the forward direction starting from a model containing all 15 terms of equation (III), Table 1 (four each of the main effects and the quadratic terms, plus the seven two-factor interactions). The ultimate model turned out to use only nine of these 15 parameters. Student t tests were examined to verify the statistical significance of each parameter in the model (Table 5). None of the terms involving PEG 400

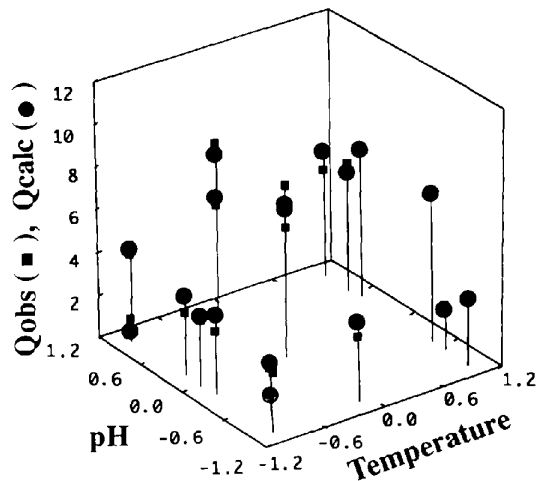


Fig. 4. The quadratic model gives excellent predictions for the observed results. Comparison between the observed data points (Q_{obs}) for the Hardin-Sloane design performed for type I crystals and those calculated from the fitted response-surface formula given in Table 5 (Q_{calc}). The 20 points are displayed on the two-dimensional surface spanned by pH and temperature, which shows the most important variation and in which the most important optimization takes place (see Fig. 6). Multiple points along a given spike, therefore, represent different values for the two variables ([PEG] and [protein]) not spanned by that surface.

had any significant effect on the score, and the protein-concentration terms were only marginally important. The concentration did, however, interact with both the pH and temperature, which were dominant in defining the surface.

The model itself is presented in Table 6, together with regression and ANOVA statistics. Agreement between observed and predicted scores, illustrated graphically in the pH \times temperature plane (Fig. 4), is convincing; the multiple correlation coefficient between Q_{obs} and Q_{calc} is 0.959. The F ratio, 15.47, has a probability of 0.000057 under the null hypothesis.

Partial derivatives with respect to pH, concentration and temperature were evaluated from the model expression given in Table 6. Equating these three derivatives to zero and solving for the coordinates (pH_{opt} , C_{opt} and T_{opt}) of the optimum gave rise to the prediction that crystals grown at these values would be better than any of those observed in the H-S experiment. Verification of that prediction is shown in Fig. 6. Examination of the level surfaces at the stationary point (Fig. 5a) shows that the

pH \times temperature surface has a clear maximum, whereas the pH \times concentration and temperature \times concentration surfaces are ridges in which nearly optimal crystals appear over a wide range of different combinations of the two factors.

The model for type IV crystals was obtained and analyzed in the same manner, but was qualitatively quite different. The pH proved to be unimportant, and was omitted from the model. Use of PEG 400, however, was quite important. An improved fit was obtained by including a term for the interaction between the temperature and the square of the PEG effect. This term, $T \times \text{PEG} \times \text{PEG}$, suggests an interaction between the optimum value for the PEG concentration and the temperature. It is possibly related to the observation that substitution of selenomethionine produces crystals that are unstable at lower temperatures, in marked contrast to the behavior of native protein crystals (Doublie & Carter, 1994). Both observations may arise from the involvement of hydrophobic lattice packing forces in stabilizing type IV crystals.

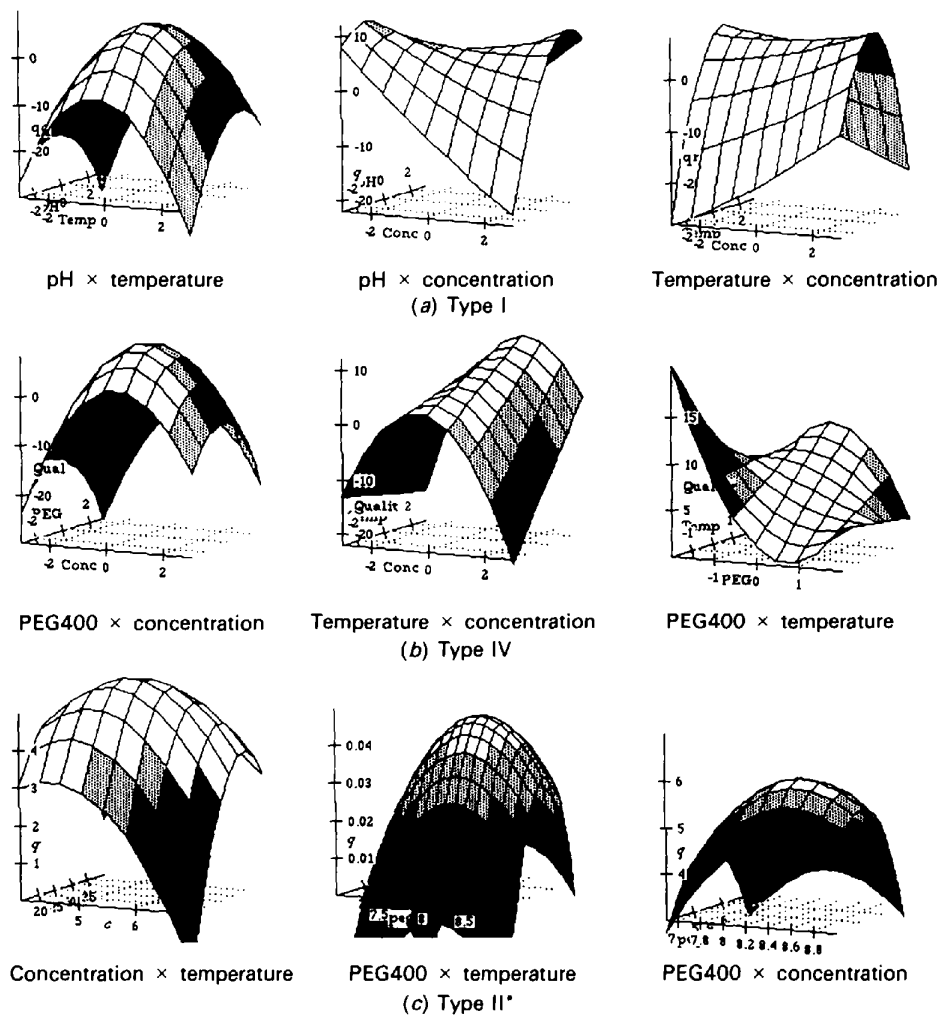


Fig. 5. Representative two-dimensional level surfaces for the response surfaces in Tables 5 and 6. Values for the two additional variables were set at the stationary points by partial differentiation and solution of the resulting simultaneous equations.

4.1.3. *Level surfaces.* The level surfaces have interesting properties (Fig. 5) that are frequently encountered in response-surface studies (Box, Hunter & Hunter, 1978, ch. 15). As noted above, ridges result when different combinations of two factors produce essentially the same result. They indicate that the natural coordinate directions chosen for the experiment are not 'fundamental' factors, but that some combination of coordinate directions may represent a fundamental factor (Box, Hunter & Hunter, 1978). For example, the temperature \times concentration surfaces for both crystal forms I and IV are ridges. Both protein concentration and temperature influence the supersaturation ratio, which does have a fundamental influence on crystal growth because it determines the concentration of nucleation centers. The ridge in this case probably represents combinations of concentration and temperature with the same supersaturation ratio.

4.1.4. *Prediction and confirmation.* Two of the level surfaces show maxima: pH \times temperature for type I and PEG 400 \times concentration for type IV crystals. Thus, the stationary points make definite predictions

with respect to optimization of crystal growth. Both predictions were verified by subsequent experiments. The predicted optima for type IV crystals produced a modest increase in crystal size of the native protein, but unfortunately did not noticeably increase the size of SeMet protein crystals. Predicted optima produced a dramatic improvement in the habit, uniformity and homogeneity of type I crystals (Fig. 6). We have never seen single type I crystals with so favorable a morphology.

4.2. *[Protein] \times [precipitant] as a search direction: optimization of type II* crystals*

The preliminary success with the first two response-surface experiments encouraged us to try to increase the volume of type II* crystals by this procedure. This experiment differed in several important ways from the two previous ones.

We set up duplicate experiments for each sample point in the design, in order to enhance the precision, statistical significance, and hence the accuracy of the model parameters.

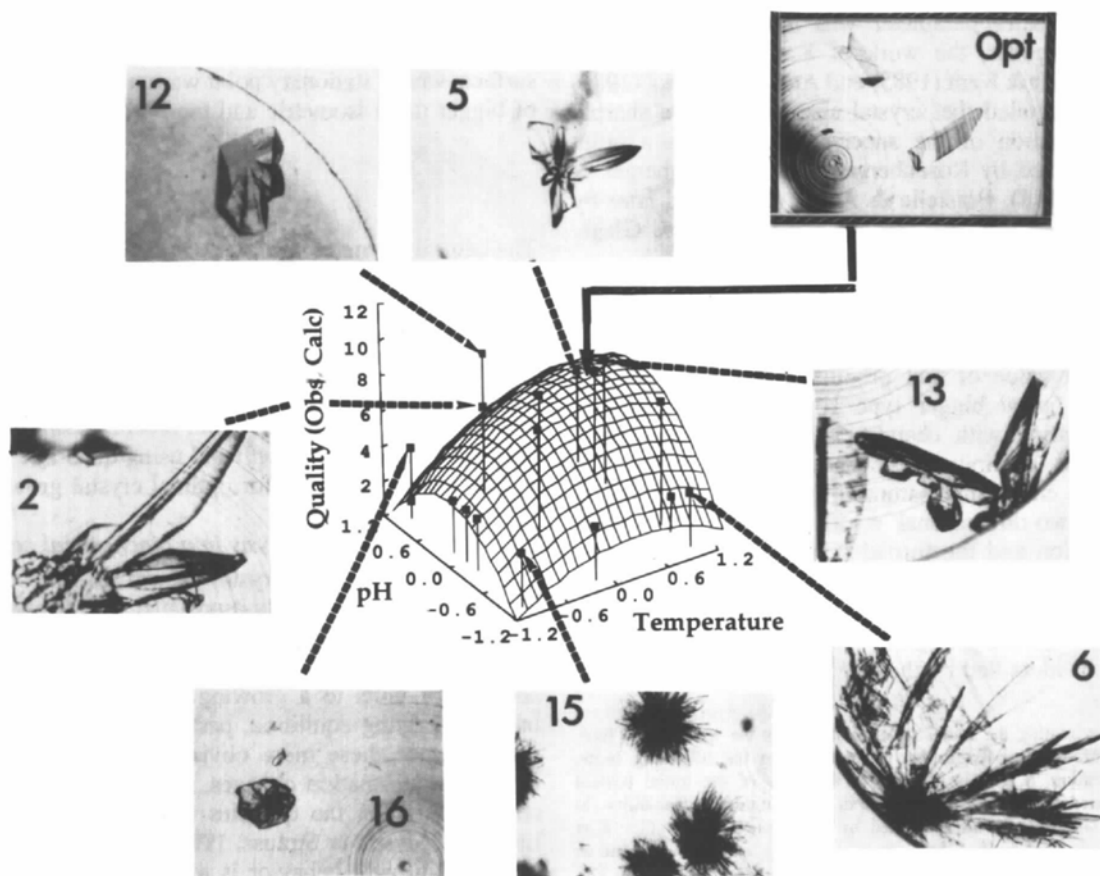


Fig. 6. The type I crystal response-surface experiment, together with its verification. Numbered examples around the periphery denote the numbered design points from Table 3 and show the data from which the surface in the center was constructed. That surface is the same as the first surface in Fig. 5(a). The spikes in the surface represent the observed scores for the individual experiments. Opt represents a verification experiment performed at the stationary point of the surface, which represents an optimum score. Crystals of such size and uniform dimensions had never been seen previously for this crystal form.

We used three different scores to evaluate the results. Since type II* crystals had always been too small, and had a tendency to grow as needles, we scored experiments separately on the basis of the total volume and the ratio of width to length, estimated by microscopic examination. Since type II* crystals also tend to sprout satellite crystals, we used a new score, called 'uniformity', to quantify this tendency in each experiment.

We changed one of the search directions, replacing the pH by the product of the protein and precipitant concentrations as the independent variable in column 1 of the design matrix. We reasoned that an optimal pH could be inferred from the strong, positive, two-factor interaction between pH and tryptophan, which implied that the type II* crystals should be grown at a higher pH. Not having to optimize the pH freed up one column to be assigned to a new factor.

Varying the product of the two concentrations systematically was an attempt to explicitly vary the supersaturation. In each of the two previous response-surface models, we had observed that the temperature \times concentration level surfaces were ridges, suggesting that some other parameter was a more fundamental factor. Moreover, the work of Kam, Shore & Feher (1978), Feher & Kam (1985) and Ataka & Tanaka (1986) had demonstrated that crystal size was often a sharply peaked function of the supersaturation ratio, a point also addressed by Rosenberger (1986), Rosenberger & Meehan (1988), Boistelle & Astier (1988) and later by others (Riès-Kautt & Ducruix, 1992; Mikol & Giegé, 1992).

These studies showed that crystal growth experiments often produce similar results along a hyperbola in the [protein] \times [precipitant] plane, characterized by a constant value of that product. Since we had been unable to grow bigger type II* crystals by varying supersaturation with changes in protein concentration at constant precipitant concentration, we decided to search the entire supersaturation zone, sampling points along a two-dimensional region spanned by protein concentration and the [protein] \times [precipitant] product as summarized in Table 4, column 3.†

This new search direction proved very useful indeed (Fig. 7). Previously, type II* crystals grown under the conditions identified by the original incomplete factorial

search were similar to those observed in experiments 5 and 20. Systematic titration of these conditions to change the initial supersaturation by changing the TrpRS concentration produced only a slightly bigger maximum size. This modest increase can be rationalized, *post hoc*, from the more complete, two-dimensional response surface. This one-dimensional search missed a considerably better optimum achievable at a *higher* initial [protein] \times [precipitant] product (Opt1 and Opt2). When we tried growing crystals in two different experiments at this point, they were substantially better in all three respects. They were more isometric; they were larger by two orders of magnitude; and they were free of the satellite crystals that plagued all our previous attempts to prepare good type II* crystals. These crystals were the first of this type that have ever grown large enough the first time to give strong X-ray diffraction patterns.

It is rather important to notice that this optimum size was obtained *outside* the range of variables we used in the response-surface determination in the first place. This can be seen in terms of the traditional solubility diagram in Fig. 8. The maximum of the fitted surface lies at a [protein] \times [precipitant] product well above the range of hyperbolae that bounded our experiment. Despite this limitation, fitting the sampled points produced a useful surface, whose stationary point was an accurate predictor of bigger more isometric and more uniform crystals.

5. Discussion

The central theme of this work is that quantitative scoring of appropriately designed crystal growth experiments is of more than academic interest because it opens up a powerful range of analytical methods based on multiple regression and the analysis of variance. Two different practical advantages have been described: making quantitative inferences about biochemically relevant influences on crystal growth and using quadratic models to find stationary points for optimal crystal growth.

5.1. Crystal growth analysis in a biochemical context

One often thinks of crystal growth in terms only of the equilibria most closely associated with formation of intermolecular contacts - prenuclear aggregates, critical nuclei and so on (Mikol & Giegé, 1992) - and with the addition of units to a growing crystal face. However, other underlying equilibria, particularly ligand binding, can influence these more obvious processes indirectly through conformation changes, giving rise to considerable variation in the qualities of the crystals (Schutt, Lindberg, Myslik & Strauss, 1989). Normally, the study of crystal growth behavior is aimed at gaining control over this variability in order to make better crystals. The variability itself is not considered as a source of biochemical information about the macromolecule under study.

† Approximating the initial supersaturation by the [protein] \times [precipitant] product can be further rationalized on the following basis. Supersaturation, S , can be defined as the ratio of the initial protein concentration to the equilibrium protein concentration or solubility. In turn, the solubility is often defined by the relation $\ln S = C - K_s m$ (Arakawa & Timasheff, 1986), where C and K_s are constants, and m represents a measure of the chemical potential of the precipitant. The minus sign for K_s implies that the ratio, [protein]/exp $\{C - K_s m\}$ can be rewritten as [protein]exp $\{K_s m\} \simeq$ [protein] $\{1 + K_s m\}$. In practice, $K_s m \gg 1$, so $S \simeq$ [protein] $K_s m$. For more than one precipitant, and assuming to a first approximation that $\ln S = C - \sum_{i=1}^n K_i m_i$ gives a more general formulation, $S =$ [protein] $\prod_{i=1}^n \exp(K_i m_i) \simeq$ [protein] $\prod_{i=1}^n (1 + K_i m_i)$.

However, this situation can be turned around: variability in crystal growth behavior actually also carries information about biochemically relevant conformational equilibria, provided it can be shown to be caused by factors that change the behavior of the macromolecule. A full factorial experiment, combined with the analysis of variance, can provide evidence for such causal relationships. As noted above, ligand-binding- and pH-dependent conformational changes have previously been described for aminoacyl-tRNA synthetases. Conformational changes would be expected to affect protein crystal growth. A factorial experiment provided statistically significant evidence that TrpRS crystal growth behaves in a manner consistent with that literature, and hence that representative stages of the TrpRS mechanism have been trapped in the different crystal forms.

5.2. Stationary points, quality control and sources of residual variability

A 'stationary point' of any function is, by definition, a combination of the independent variables where the result does not change much for small fluctuations in any direction away from that point. This crucial concept from response-surface methodology underscores

other important practical advantages of locating stationary points besides that of growing optimum crystals. Working at stationary points for any process tends to insulate it from fluctuations in other uncontrolled or uncontrollable factors. Locating and using stationary points may therefore lead to corollary benefits, such as better quality control, higher success rates and greater cost effectiveness for processes that depend on getting reproducible crystals (crystallization in microgravity). Moreover, in the case of microgravity, such an effort could help clarify its real benefits, as distinct from those that result from random fluctuations in how experiments were set up.

There is evidence supporting this view in the patterns of residual variability in the experiments illustrated in Table 2 and Fig. 2. Some of the conditions, notably number 15 and number 16 produced results that were indistinguishable by any criteria on both occasions (Figs. 2o and 2p) and have invariably produced the same results under many replications of the same conditions, throughout our work on that structure. Other conditions, for example, numbers 3, 8, 11 and 13 (Figs. 2i, 2n, 2k and 2g) gave results that were, qualitatively, distinctly different in the two different replicates, and which we have verified in many trials also give more variable

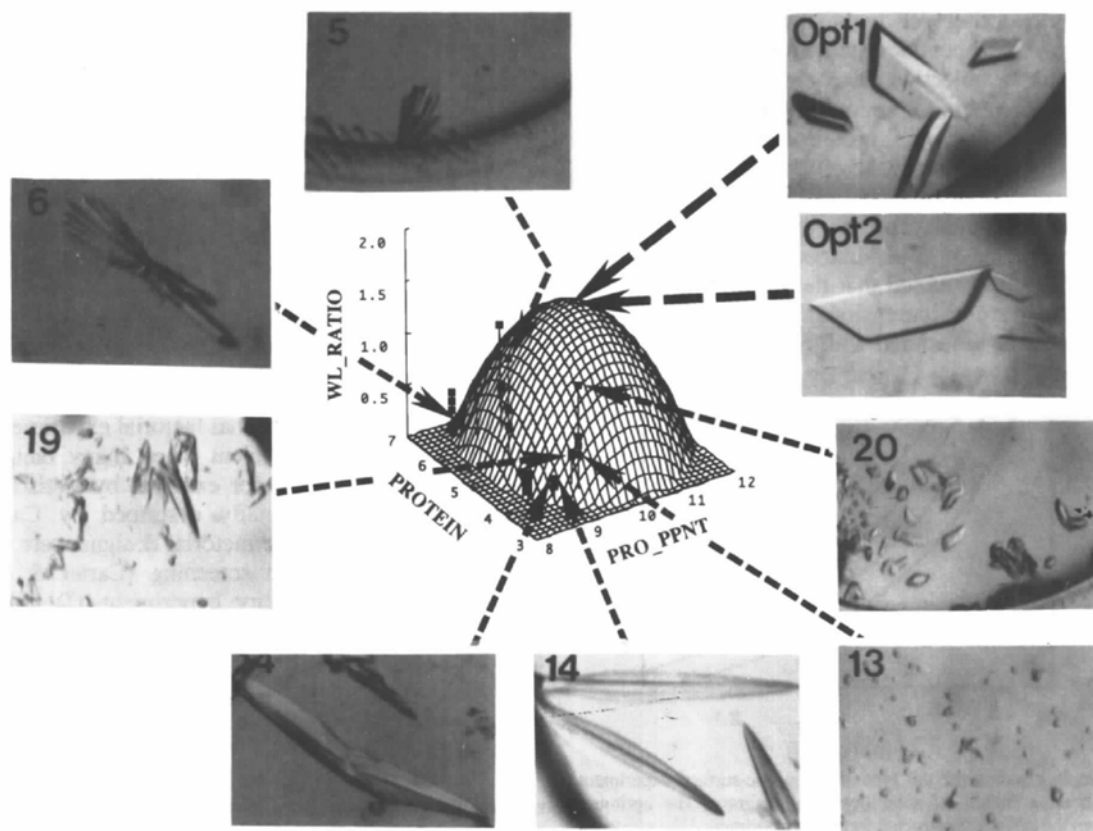


Fig. 7. The type IV crystal response-surface experiment, together with its verification. This figure is constructed in the same manner as Fig. 7. The surface is similar to that shown in Fig. 6, except that the gradient is steeper. Opt1 and Opt2 represent two replicate experiments set up at the optimum point of the surface.

behavior. Response surfaces for the different crystal forms (Figs. 5-8) suggest that these large individual sample variances are indeed associated with conditions far from stationary points, where there may be large partial derivatives with respect to the crystal growth variables. Experiments 15 and 16 are close to the local optimum in the type IV response surface, whereas experiments showing large variation between replicates lie far from such optima (Figs. 7 and 8). Thus, distance of a particular experimental specification from a stationary point in the response surface may place a more important limitation on the reproducibility of a score than does the subjectivity of the experimenter.

A related observation of potentially general significance is the fact that the stationary points for the three different scoring methods in Table 7 are very close to each other. In other words, when we optimized for the asymmetry ratio we also simultaneously found that we got bigger and more uniform crystals. Particularly gratifying is the fact that the multiple growth that had plagued our reseeded experiments does not seem to be a problem with crystals grown at the stationary point for the most isometric crystals (Fig. 7).

5.3. Multidimensionality, search directions and sampling efficiency

It is worth noting yet again how important it is to sample the experimental space appropriately, in pursuit of a given objective. The results described here were all obtained from a small number of experiments, using modest amounts of protein, because the design matrices were constructed carefully. The efficiency of factorial experiments, paradoxically, seems to increase with the dimensionality of the problem. In fact, this increase in efficiency is understood by mathematicians. Bricogne (1993) has noted that the underlying structure of sampling in multi-dimensional spaces is closely related to problems associated with error-correcting codes.

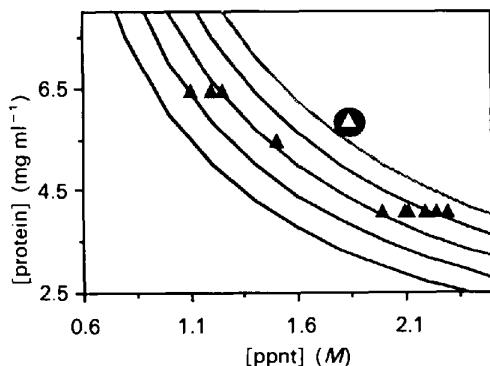


Fig. 8. Sampling points for the type II* response-surface experiment displayed on a traditional solubility-curve diagram. The optimum point from the surface constructed from the experimental data (reverse contrast) lies well outside the range of values tested in the design, and in a region we had never sampled previously because it involved a significantly higher [protein] \times [precipitant] product.

Table 7. Quadratic models for type II* TrpRS crystal growth response surfaces

Models were selected by the same procedure as that used for the type I and type IV models in Table 5. Three different scores were fitted to three different models. The standard error of the estimate for the volume has units of mm^3 ; for the uniformity score it is based on a unitless number between 0 and 10.

Type II* model 1, W/L ratio

$$Q_{W/L \text{ ratio}} = \beta_0 + \beta_{\text{PEG}}\text{PEG} + \beta_C C + \beta_{\text{Prod}}\text{Prod} + \beta_T T + \beta_{\text{PEG}^2}\text{PEG}^2 + \beta_{\text{PEG} \times \text{Prod}}\text{PEG} \times \text{Prod} + \beta_{\text{PEG} \times T}\text{PEG} \times T + \beta_{C^2} C^2 + \beta_{C \times T} C \times T + \beta_{\text{Prod}^2}\text{Prod}^2 + \beta_{\text{Prod} \times T}\text{Prod} \times T$$

Regression statistics		Analysis of variance			
R	R^2	Std error of estimate	Degrees of freedom	F ratio	p
0.975	0.951	0.155	11	38.58	0.8×10^{-11}

Type II* model 2, volume

$$Q_{\text{volume}} = \beta_0 + \beta_{\text{PEG}}\text{PEG} + \beta_{\text{Prod}}\text{Prod} + \beta_T T + \beta_{\text{PEG}^2}\text{PEG}^2 + \beta_{\text{PEG} \times C}\text{PEG} \times C + \beta_{\text{PEG} \times \text{Prod}}\text{PEG} \times \text{Prod} + \beta_{\text{PEG} \times T}\text{PEG} \times T + \beta_{C^2} C^2 + \beta_{C \times \text{Prod}} C \times \text{Prod} + \beta_{C \times T} C \times T + \beta_{\text{Prod}^2}\text{Prod}^2 + \beta_{\text{Prod} \times T}\text{Prod} \times T + \beta_T T^2$$

Regression statistics		Analysis of variance			
R	R^2	Std error of estimate	Degrees of freedom	F ratio	p
0.992	0.984	$7.01 \times 10^{-6} (\text{mm}^3)$	13	65.8	0.3×10^{-9}

Type II* model 3, uniformity

$$Q_{\text{uniform}} = \beta_0 + \beta_{\text{PEG}}\text{PEG} + \beta_C C + \beta_{\text{Prod}}\text{Prod} + \beta_T T + \beta_{\text{PEG}^2}\text{PEG}^2 + \beta_{\text{PEG} \times C}\text{PEG} \times C + \beta_{\text{PEG} \times \text{Prod}}\text{PEG} \times \text{Prod} + \beta_{\text{PEG} \times T}\text{PEG} \times T + \beta_{C^2} C^2 + \beta_{C \times \text{Prod}} C \times \text{Prod} + \beta_{C \times T} C \times T + \beta_{\text{Prod}^2}\text{Prod}^2 + \beta_T T^2$$

Regression statistics		Analysis of variance			
R	R^2	Std error of estimate	Degrees of freedom	F ratio	p
0.973	0.947	0.763	13	18.04	0.3×10^{-5}

Sampling of multi-dimensional factorial experiments can indeed become quite efficient when many factors are explored simultaneously, for example by statistical designs of the type originally described by Carter & Carter (1979). Incomplete factorial designs were crucial to us for crystal growth screening (Carter & Carter, 1979) and phase-permutation experiments (Doublé, Xiang, Bricogne, Gilmore & Carter, 1994). This lends a pleasing symmetry to the use of designed factorial experiments in the initial and final stages of crystal structure determination, both of which are quintessential factorial problems.

A related point concerns how to span the multidimensional spaces encountered in crystal growth projects. Here, the level surfaces in Fig. 5 were very helpful, suggesting the initial supersaturation as a search direction. Using the initial supersaturation in this way forced us to

examine regions we had never tried before, and which pointed the way to a substantial improvement in type II* crystals (Fig. 8). It also improves the chances that every experiment will produce some kind of crystal, thereby reducing 'wasted' experiments.

Finally, it should be noted that 'efficiency' cannot always be achieved simply by finding the smallest number of experiments necessary to do the job. The critical problem in implementing quantitative methods is to determine the residual experimental variance. This has lead us increasingly to perform replicate experiments.

5.4. Other response-surface methods

Most generally applicable procedures for improving inferior crystals involve some kind of systematic search for better growth and/or stabilization conditions and are, generically, response-surface experiments. An appealing feature of the matrix in Table 3 and similar ones that could be generated for different numbers of variables (Hardin & Sloane, 1993) is that it is general enough to be adapted to other, four-variable situations. We have adapted the same design for optimizing *in vitro* transcription yields by assigning different independent variables to the last four columns.

In our three examples the experimental points were sensitive enough to the curvature of the surface to produce sensible quadratic models. Other procedures may be helpful in cases where current conditions are far from a stationary point, and where a design like that in Table 3 may prove wasteful. Such cases can often be treated effectively by choosing a small region around the current conditions and fitting a plane to the sampled points. The direction of steepest ascent in such a fitted plane will often point close enough toward the optimum that a line search of 3-5 points along that direction will provide a fair indication of its location. Then, a design like that in Table 3 can be centered on that new point. These procedures are described in more detail in a very useful discussion of response-surface methods in Box, Hunter & Hunter (1978).

5.5. Real improvements in difficult cases

Finally, it is worth pointing out that these studies, with the possible exception of the type IV response surface, were all carried out with real and difficult problem situations in which the answers were unknown beforehand, and in which further crystallographic work depended heavily on finding better crystals. They were not 'test cases' in which the answers were known beforehand. The reproducibility and efficacy of the Hardin-Sloane response-surface design in revealing where better crystals could be grown represents, therefore, a solid advance in our structural studies research. Response-surface technology has become our routine for making things work better.

This work has been supported, at various times, by the NIH (GM26203), the American Cancer Society (BE54) and the UNC Biochemistry Department. We owe special thanks to Mary Ellen Jones and Mike Douglas for their support, and to Neil Sloane for his design, his advice, for a critical reading of the manuscript and for his enthusiasm. Ongoing experimentation by and discussions with Lee Kuyper clarified various aspects of the use of response-surface experiments. CWC Jr thanks C. W. Carter for his advice and instruction.

References

- ARAKAWA, T. & TIMASHEFF, S. N. (1986). *Methods Enzymol.* **114**, 49-77.
- ATAKA, M. & TANAKA, S. (1986). *Biopolymers*, **25**, 337-350.
- ATKINSON, T., BANKS, G. T., BRUTON, C. J., COMER, J. J., JAKES, R., KAMALAGHARAN, T., WHITAKER, A. R. & WINTER, G. P. (1979). *J. Appl. Biochem.* **1**, 247-258.
- BALDWIN, E. T., CRUMLEY, K. V. & CARTER, C. W. JR (1986). *Biophys. J.* **49**, 47-48.
- BARSTOW, D. A., SHARMAN, A. F., ATKINSON, T. & MINTON, N. P. (1984). *Gene*, **46**, 37-45.
- BOISTELLE, R. & ASTIER, J. P. (1988). *J. Cryst. Growth*, **90**, 14-30.
- BONADIO, A. (1990). *Theorist*. San Francisco: Prescience Corporation.
- BOX, G. E. P., HUNTER, W. G. & HUNTER, J. S. (1978). *Statistics for Experimenters*. New York: Wiley/Interscience.
- BRICOGNE, G. (1993). *Acta Cryst.* **D49**, 37-60.
- CARTER, C. W. JR (1988). *J. Cryst. Growth*, **90**, 168-179.
- 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). *Crystallization of Proteins and Nucleic Acids - a Practical Approach*, edited by A. DUCRUIX & R. GIEGE, pp. 47-71. Oxford: IRL Press.
- CARTER, C. W. JR (1993). *Annu. Rev. Biochem.* **62**, 715-745.
- 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.
- CARTER, C. W. JR, CRUMLEY, K. V., COLEMAN, D. C., HAGE, F. & BRICOGNE, G. (1990). *Acta Cryst.* **A46**, 57-68.
- CARTER, C. W. JR, DOUBLIÉ, S. & COLEMAN, D. E. (1994). *J. Mol. Biol.* In the press.
- CARTER, C. W. JR & GREEN, D. C. (1982). *Anal. Biochem.* **124**, 327-332.
- COLEMAN, D. E. (1988). PhD thesis, Univ. of North Carolina at Chapel Hill, USA.
- COLEMAN, D. E. & CARTER, C. W. JR (1983). *Biochemistry*, **23**, 281-385.
- CONWAY, J. & SLOANE, N. J. A. (1987). *Sphere Packings, Lattices and Groups*. Berlin: Springer-Verlag.
- DOUBLIÉ, S. & CARTER, C. W. JR (1994). *J. Biol. Chem.* Submitted.
- DOUBLIÉ, S., XIANG, S., BRICOGNE, G., GILMORE, C. J. & CARTER, C. W. JR (1994). *Acta Cryst.* **A50**, 164-182.
- FEHER, G. & KAM, Z. (1985). *Methods Enzymol.* **114**, 77-111.
- GEORGE, M. & WILSON, W. W. (1994). *Acta Cryst.* **D50**, 361-365.
- HARDIN, R. H. & SLOANE, N. J. A. (1993). *J. Stat. Plan. Inference*, **37**, 339-369.
- HÉLÈNE, C., BRUN, F. & YANIV, M. (1971). *J. Mol. Biol.* **58**, 349-365.
- KADIMA, W., MCPHERSON, A., DUNN, M. F. & JURNAK, F. A. (1990). *Biophys. J.* **57**, 125-132.
- KAM, Z., SHORE, H. B. & FEHER, G. (1978). *J. Mol. Biol.* **132**, 539-555.
- LAM, S. S. M. & SCHIMMEL, P. R. (1975). *Biochemistry*, **14**, 2775-2780.
- MIKOL, V. & GIEGE, R. (1992). *Crystallization of Proteins and Nucleic Acids - a Practical Approach*, edited by A. DUCRUIX & R. GIEGE, pp. 219-239. Oxford: IRL Press.
- MIKOL, V., HIRSCH, E. & GIEGE, R. (1989). *FEBS Lett.* **258**, 63-66.
- MIKOL, V., HIRSCH, E. & GIEGE, R. (1990). *J. Mol. Biol.* **213**, 187-195.
- NETER, J. & WASSERMAN, W. (1974). *Applied Linear Statistical Models*, ch. 11. Homewood, IL: Richard D. Irwin.

- RIÈS-KAUTT, M. & DUCRUIX, A. (1992). *Crystallization of Nucleic Acids and Proteins - a Practical Approach*, edited by A. DUCRUIX & R. GIEGÉ, pp. 195-218. Oxford: IRL Press.
- ROSENBERGER, F. (1986). *J. Cryst. Growth*, **76**, 618-636.
- ROSENBERGER, F. & MEEHAN, E. J. (1988). *J. Cryst. Growth*, **90**, 74-78.
- SCHUTT, C. E., LINDBERG, U., MYSLIK, J. & STRAUSS, N. (1989). *J. Mol. Biol.* **209**, 735-746.
- STURA, E. & WILSON, I. (1990). In *Methods: a Companion to Methods in Enzymology*, Vol. 1. New York: Academic Press.
- THALLER, C., EICHELE, G., WEAVER, L. H., WILSON, E., KARLSSON, R. & JANSONIUS, J. N. (1985). *Methods Enzymol.* **114**, 132-135.
- THALLER, C., WEAVER, L. H., EICHELE, G., WILSON, E., KARLSSON, R. & JANSONIUS, J. N. (1981). *J. Mol. Biol.* **147**, 465-469.
- WEBER, P. C. (1991). *Adv. Protein Chem.* **41**, 1-36.
- WILKINSON, L. (1987). Editor. *SYSTAT. The System for Statistics*. Evanston, IL: SYSTAT Inc.
- WILSON, W. (1990). In *Methods: a Companion to Methods in Enzymology*, Vol. 1, edited by C. W. CARTER JR, pp. 110-117. New York: Academic Press.