CMLS, Cell. Mol. Life Sci. 61 (2004) 525–536
1420-682X/04/050525-12
DOI 10.1007/s00018-003-3260-z

CMLS Cellular and Molecular Life Sciences

Review

Which strategy for a protein crystallization project?

C. E. Kundrot

Microgravity Science and Applications Department, Science Planning and Program Management Group, Mail Code SD41, Marshall Space Flight Center, Alabama 35812 (USA), Fax: +1 256 544 9305, e-mail: craig.e.kundrot@nasa.gov

Received 1 July 2003; received after revision 6 August 2003; accepted 22 August 2003

Abstract. The three-dimensional, atomic-resolution protein structures produced by X-ray crystallography over the past 50+ years have led to tremendous chemical understanding of fundamental biochemical processes. The pace of discovery in protein crystallography has increased greatly with advances in molecular biology, crystallization techniques, cryocrystallography, area detectors, synchrotrons and computing. While the methods used to produce single, well-ordered crystals have also evolved over the years in response to increased understanding and advancing technology, crystallization strategies continue to be rooted in trial-and-error approaches. This review summarizes the current approaches in protein crystallization and surveys the first results to emerge from the structural genomics efforts.

Key words. Protein crystallization; structural genomics; protein structure; X-ray crystallography; X-ray diffraction.

Introduction

The three-dimensional, atomic-resolution protein structures produced by X-ray crystallography over the past 50+ years have led to tremendous insights into how fundamental biochemical processes are conducted at the chemical level. These insights have been critical for practical applications such as structure-based drug design and the development of enzymes for industrial processes. X-ray crystallography has benefited enormously from advances in molecular biology, cryocrystallography, area detectors, synchrotrons and computing. These advances have lowered the requirements for crystal size, but single, well-ordered crystals are still required. The methods used to produce crystals have also evolved over the years as a result of increased understanding and advancing technology, but crystallization strategies continue to be rooted in trial-and-error approaches refined over the past decades.

The fundamental task in protein crystallization has remained the same over the years. The aim is to create a solution supersaturated with protein that will produce single, well-ordered crystals. More often than not, supersaturated solutions produce precipitate or phase separation instead. There is no a priori theory for discerning which solutions will produce crystals and which will produce precipitate. So the crystallization process is separated into two stages: screening and optimizing. Each stage is conducted with a largely empirical approach. The screening process discovers lead crystallization conditions, or 'hits', which typically produce microcrystals, thin rods or thin plates. Of the three phases of crystal growth, (nucleation, growth and cessation of growth [1]), screening is largely concerned with finding conditions that will support nucleation and some crystal growth. The crystallization conditions are then optimized to produce a crystal suitable for x-ray diffraction, i.e. large and well-ordered. Before the widespread use of synchrotron radiation, 'large' meant crystals that were many hundreds of microns on edge. Today, however, 'large' may mean just tens of microns on edge.

The specific strategies used to crystallize proteins have evolved over the years. Many excellent accounts of the

strategies and methods used in crystallization have been written [2–8]. In the 1960s and 1970s, investigations were largely confined to targets that were naturally abundant. A standard approach to screening for crystallization conditions was to methodically examine several different precipitants as a function of pH [2]. In the 1980s, advances in molecular biology allowed the overexpression of many proteins with low natural abundance. The ability to make site-directed mutants provided a versatile way to make alternative constructs for crystallization. In the 1990s, the Sparse Matrix approach for screening for crystallization conditions was introduced [9]. With the advent of the new millennium, the field of protein crystal growth is about to undergo another large change. Structural genomics is fueling the development of new high-throughput methods for crystallization, and it is generating large amounts of systematic crystallization data that may lead to more predication-based crystallization methods [10–16].

The best crystallization strategy for a project depends heavily on the crystallization throughput of the laboratory. It is useful to describe crystallization throughput in terms of three levels. The low throughput laboratory conducts almost all aspects of the project manually without automated equipment. In particular, crystallization trials are set up by hand. Historically, all laboratories were low throughput. At the other extreme are high-throughput structural genomics (HTGS) efforts. High throughput is achieved by specialized equipment that automates most aspects of the crystallization project. In between these two extremes is the medium-throughput laboratory. It has much less automation than HTGS, but it at least has a commercially available system for setting up crystallization trials. The same basic techniques are available to all three levels of throughput, but the strategies change as a function of throughput. A general theme in protein crystallization is that there is no lack of approaches to try; the trick is to judiciously select the approaches that are most efficient for the project at hand. Greater throughput allows more exhaustive exploration of the possibilities at each step, whereas lower throughput requires more limited exploration.

This review summarizes the current approaches in protein crystallization and surveys the first results to emerge from the structural genomics efforts. It is written primarily from the perspective of a medium-throughput laboratory where crystallization trials can be set up by automated equipment. Low-throughput laboratories may be best served by implementing a subset of the suggestions, whereas high-throughput laboratories can exhaustively pursue all of the strategies described here along with further variations.

Target Selection and Constructs

There is usually considerable flexibility in choosing the exact crystallization target. That flexibility should be used to increase the odds for success. The ideal outcome of the project may be to determine the structure of a protein from a particular organism complexed with a particular small molecule ligand or macromolecule. Many of the biological questions motivating the crystallization project, however, can usually be answered by determining the structure of a modified form of the protein, of a protein from another species or of a different ligation state. Each change to the sequence of the protein or its ligation state generates a new crystallization target for screening. Four different ways to modify a protein to generate a new target for crystallization are to (i) add an N- or C-terminal tag or fusion protein to the protein (ii) truncate the molecule (iii) alter the sequence and (iv) chemically modify specific residues.

The affinity tags commonly engineered onto a target for ease of purification provide one useful means of modifying the crystallization target. Tags are used primarily to aid purification, but they can also improve the level of expression and the solubility. Stevens lists 30 different tags useful for purification [17]. They may be removed from the target or left on for crystallization. It has generally been believed that tags are either neutral or detrimental to crystallization [17, 18].

Recent studies, however, suggest that the presence of tags and the types of tags used should both be considered useful variables in the crystallization. One study showed that the 21-residue amino terminal tag from the Gateway system was compatible with crystallization [19]: 13 of 25 (55%) proteins were crystallized with the tag. A second study tested the effect of five short tags [Arg_5, His_6, FLAG , Strep tag II and biotin acceptor peptide (BAP)] attached to the C-terminus of maltodextrin-binding protein from *Pyrococcus furiosus* [20]. Under the wild-type crystallization conditions, the His_6 and FLAG tags produced crystals of similar diffraction quality as wild type, whereas the BAP tag produced slightly poorer diffraction and the Arg_5 tag produced much worse diffraction. After screening with Wizard I and II, and Hampton I and II, however, the BAP variant produced a crystal form that diffracted better than wild type. No crystals were obtained of the Strep II variant in any condition. This study suggests that the choice of affinity tag is as important as the decision to retain the tag. A third study tested 49 proteins with and without His_6 tags: 11 proteins crystallized both with and without the tag, 9 crystallized only without a tag and 7 crystallized only with a tag [21]. In conclusion, many small affinity tags are compatible with crystallization, and they sometimes enable crystallization.

In contrast to small affinity tags, fusion proteins appear to be more problematic in crystallization. Fusion proteins

consisting of the target protein joined through a linker region to a fusion protein such as glutathione *S*-transferase (GST) or maltose binding protein (MBP) have been crystallized [22–26], but there are relatively few reports of successful crystallization of fusion proteins [27]. The high incidence of unpublished negative results suggests that the crystallization of fusion proteins is difficult.

On the other hand, the removal of flexible regions of the protein or whole domains is a very common approach. Flexible regions can be identified by several methods. Limited proteolysis by a battery of different proteases is common. Low-resolution electron density maps [28] and nuclear magnetic resonance (NMR) [29] have also been used.

More precise changes can be made by site-directed mutagenesis to reproduce the packing found in the crystal structure of homologous proteins, to alter solubility, to alter the surface properties of the target or to stabilize the target.

One can attempt to engineer crystal contacts when the crystal structure of a homologous protein or a low-resolution structure of the target is available. To crystallize human H-chain ferritin, a point mutation was introduced to reproduce the Cd^{2+}-mediated crystal packing contact found in the L-chain ferritins from horse spleen and rat liver [30]. A mutation that altered an intermolecular salt bridge in canavalin trimers produced better diffracting crystals [31]. Crystal packing through the formation of intermolecular beta-sheets was promoted by altering the sequence of solvent accessible β strands in model antibodies [32, 33]. Crystal-packing mutations and their affect on crystallizability were systematically examined for *Thermus thermophilus* aspartyl-transfer RNA (tRNA) synthetase [34]. When attempting to crystallize a complex, crystal contact engineering can be attempted when at least one of the members of the complex has a known structure and there is a reasonable model of what surfaces are buried in the complex, e.g. [35].

Several other types of mutations at surface sites have enabled crystallization. Candidate surface sites may be identified through homology modeling, structure prediction or just the tendency of certain amino acids to be located on the surface. Point mutations may be made to alter function, e.g. the elimination of autoproteolysis [36]. More generally, mutations can be made to alter surface properties. Initial attempts to crystallize wild-type human RhoGDI protein failed. To reduce the conformational entropy of surface side chains, 13 lysine-to-alanine constructs were made, and the structure of four of the variants determined [37]. Hydrophobic surface residues can be mutated to polar residues to increase protein solubility. The F185K mutation of the central core domain of human immunodeficiency virus (HIV) integrase raised the solubility from 1 mg/ml to over 40 mg/ml and enabled crystallization [38]. Directed evolution has been used to generate more soluble constructs [39]. Solvent accessible cysteine residues can be mutated to alanine to reduce the tendency toward oxidation and aggregation [40]. On the other hand, the introduction of an intramolecular disulfide bridge in T4 lysozyme sped nucleation and crystal growth [41]. Even broader 'shotgun' approaches to mutating surface residues have been successful. For example, most mutations of 11 residues on the surface of human thymidylate synthase made the protein easier to crystallize [42]. Many other examples of mutations used to crystallized proteins have been reported [27, 36, 43, 44].

Relatively few examples of chemical modification have been reported for crystallization. One of the most noteworthy examples of a classical chemical modification for crystallization is the S1 myosin head that required acetylated lysines for the structure determination [45]. Deglycosylation may also be regarded as a type of chemical modification. Glycoproteins are commonly deglycosylated in preparation for crystallization [46].

Another way to vary the sequence is to crystallize orthologs – homologous proteins from different organisms. This approach was first used in the crystallization of myoglobin [47] and was key to the structure determination of the TATA-binding protein [48]. One HTGS project found that the species of origin made a large difference in the ability to crystallize particular proteins. The crystallization success rate for 62 different proteins from *Escherichia coli* and *Thermotoga maritime* were 26 and 32%, respectively [21]. But 90% of the proteins crystallized were only crystallized from one of the two organisms. The species of origin, then, should be considered a productive variable.

If the protein forms a complex with another macromolecule, one can attempt to crystallize the protein complexed with different constructs of its cognate partner. The early work on λ repressor demonstrated the usefulness of exploring different DNA oligomers for protein-DNA crystallization [49]. Antibodies or Fab fragments have been used to complex with target proteins [50, 51], as have other high-affinity binding proteins [52]. The same variations that can be made to the protein target can also be made to the cognate partner.

Target Preparation

Sample preparation is critical for obtaining crystals, but it is not known a priori just how stringent the preparation must be. The chemical and conformational purity of the preparation is a critical factor affecting the ability to grow crystals. The amount of resources (time, effort, funding) devoted to sample preparation, however, must be balanced against the resources used for crystallization screening. As Rupp points out, '…the cheapest, fastest and by far the most conclusive measure for crystallization

still remains – crystallization screening' [53]. An effective strategy is to conduct crystallization screens as soon as possible. Analytical techniques such as gel electrophoresis and dynamic light scattering can guide subsequent efforts to (i) improve the quality of the protein preparation used in screening and (ii) optimize the crystallization. These analytical techniques identify problems known to interfere with crystallization and are used to focus extra efforts on specific problems as needed.

Several techniques are useful for assessing the chemical purity of the target. SDS-polyacrylamide gel electrophoresis (PAGE) is the easiest to perform, and it gives a good indication of purity of the target compared with heterogeneous contaminants. Many investigators will not attempt to crystallize a protein unless the preparation is at least 95% pure. SDS-PAGE is not, however, a good method for identifying microheterogeneity. Isoelectric focusing is a better indicator of the homogeneity of charged side chains. Mass spectrometry is a powerful method for assessing both heterogeneous contaminants and microheterogeneity.

Conformational heterogeneity can inhibit crystallization but it is rarely examined in detail because of the lack of convenient, powerful methods. Native PAGE is one indicator of conformational homogeneity, but its interpretation is not straightforward. In practice, conformational heterogeneity tends to be assessed by other means such as limited proteolytic digestion or NMR.

Aggregation of the target molecules usually indicates that crystallization will be very difficult. Electron microscopy has been used to measure aggregation prior to crystallization [54]. But the most common technique is to use light scattering. Dynamic light scattering measures the tendency of the target molecules to be monodisperse, complexed with a well-defined stoichiometry, or aggregated with an ill-defined stoichiometry. One study found that only 1 out of 12 (8%) proteins with a multimodal size distribution crystallized [55]. In contrast, 60 and 77% of those proteins with a broad or a narrow unimodal distribution, respectively, crystallized. A separate study of 25 target proteins found that none of the 13 that were crystallized were aggregated [19]. Yet another estimate is that about 30% of polydisperse or impure solutions will still produce crystals [53]. While one cannot definitively say how much aggregation is required to interfere with crystallization, the trends clearly indicate that the more monodisperse the preparation is, the more likely the molecule is to crystallize [56, 57].

When the target has a known function, functional assays should also be used to characterize the protein. Properties such as specific activity are valuable ways to assess purity. Crystallization is usually possible after a modest reduction in activity, but sometimes the inability to crystallize the protein precedes any detectable reduction in activity.

If the protein target is to be complexed with a small molecule ligand or another macromolecule, there is an additional level of preparation involved. Complexes often have very different solubility properties than the uncomplexed target. Each type of complex with the target protein should be regarded as a separate crystallization target that must be prepared, purified and screened independently.

Lastly, adequate storage conditions are important for ensuring reproducibility. Different storage conditions should be tested and validated as early as possible in the project. Initially, the choice of storage conditions will be guided by the results of the analytical methods described above. Once crystallization conditions are found, however, the reproducibility of the crystallization should be used to validate the method of storage. Storage temperatures of –80 or –20 °C are most common. Rapid freezing in liquid nitrogen with subsequent storage at –80 °C is probably the robust method in general.

Screening

There are a huge number of variables that can be explored in a search for crystallization conditions. McPherson lists 36 factors as affecting crystallization [8]. Among the physical factors are temperature, rate of equilibration, vibration, and homogeneous or heterogeneous nucleants. Chemical factors include precipitant type and concentration, pH, protein concentration, detergents and impurities. Biochemical factors include purity, ligands, aggregation state, posttranslational modification, proteolysis/hydrolysis, chemical modification, genetic modification and history. The challenge is to decide how much variation to explore within each factor: much, little, or none. Fortunately, the experience of crystal growers over the last five decades has produced effective heuristics.

It can be helpful to think of the screening strategy in terms of a hierarchy of factors. The lowest level set may be regarded as a specific set of commercial sparse matrix screens at both 4 ° and 22 °C. These screens will be performed for all the target variants generated by the higher-level factors. The second level may be regarded as the ligation state of the protein. For example, an ATPase may be crystallized without a ligand, with Mg^{++}-ATP, or with Mg^{++}-ADP. The third level of factors may include the nature of the protein construct and its preparation, e. g. plus His tag and minus His tag. One starts by fixing as many high-level factors as possible and conducting the screens. If no hits are obtained, one can systematically vary factors working from lower levels to higher levels.

Foremost among the factors usually fixed is the technique used to achieve supersaturation. Many techniques have been used to grow crystals (table 1). In practice, vapor diffusion using hanging drops or sitting drops and micro-

Table 1. Crystallization Methods [8].

Bulk crystallization
Batch method in vials
Evaporation
Bulk dialysis
Concentration dialysis
Microdialysis
Liquid bridge
Free interface diffusion
Vapor diffusion (sitting drops)
Vapor diffusion (hanging drops)
Sequential extraction
pH-induced
Temperature induced
Crystallization by effector addition

batch under oil are the most popular screening techniques. Three commercially available systems for automatically dispensing crystallization experiments are the Oryx 6/IMPAX 1-5 system (Douglas Instruments; www.douglas.co.uk), the Gilson/Cyberlab workstations (Cyberlab, www.gilson.com) and the Decode Biostructures RoboHTC (www.decode.com). The Gilson/Cyberlab workstations set up vapor diffusion crystallizations. The Oryx6/IMPAX 1-5 system uses the microbatch under oil technique. The vapor diffusion and microbatch under oil techniques do not produce all of the same lead crystallization conditions, but they produce similar numbers of hits [58–60]. Each technique has its own variations, such as equilibration rate [61, 62], and these factors are usually fixed for screening.

Another factor that is usually held constant during screening is the size of the drop. The size of the drop limits the size of the crystal that can be grown. For a 10 mg ml^{-1} protein solution, protein-specific volume of 0.73 cm^3 g^{-1}, and a crystal that is 50% protein, the concentration of protein in the crystal is about 70 times higher than the concentration of protein in solution. A cubic crystal 100 µm on a side, then, requires a drop size of 70 nl to provide the protein needed for growth (assuming no protein remains in solution). In practice, a significant amount of protein may remain in solution, and more than one crystal usually forms. Drop sizes of 1–4 µl, therefore, are common.

Several factors are sampled at a few different values during screening. Temperature is almost always sampled near room temperature, e.g. 22 °C, and it is typically also sampled at 4 °C, i.e. in a cold room. If the protein binds ligands, it will be tested in the presence and the absence of the ligands. If it complexes with another macromolecule, it may be tested with and without its cognate partners. If different constructs of the protein are available, each will be tested separately.

Screening is usually conducted at one or just a few protein concentrations. Most commonly one screens with a protein concentration about 10 mg ml^{-1}. One may more rationally choose a protein concentration by choosing a low concentration that can still be precipitated by reasonable concentrations of representative precipitants such as ammonium sulfate, PEG 4000 and ethanol [53]. Solubility screens [51] are commercially available (Hampton Research, Molecular Dimensions) to provide an initial impression of solubility, but to our knowledge the correlations between results from these solubility screens and the results from the commercially available Sparse Matrix screens have not been published. Regardless of the protein concentrations chosen, the screen is still sparsely sampled with respect to protein concentration. Many of the drops will remain clear or have large amounts of precipitate because the protein supersaturation was either too low or too high. Such drops may give the false impression that the chemical conditions used will not support crystal growth. In practice, however, the logistics of testing many different protein concentrations for each trial in the screen prevents substantial follow-up on clear or heavily precipitated drops.

The only factor that is extensively sampled is the chemical composition of the crystallization trial. The chemical composition space for crystallization is enormous. 383 different compounds have been used in crystallizations reported in the Biological Macromolecular Crystallization Database [63, 64]. There are four general categories of precipitants: (i) salts, (ii) volatile organic solvents, (iii) polymers, and (iv) nonvolatile organic alcohols. Examples include ammonium sulfate, ethanol, polyethylene glycol and MPD, respectively. So while the chemical composition is varied extensively, only a very small subset of the possible conditions is sampled.

How many conditions should be screened? One study concluded that a reasonable lower limit estimate of the crystallization success rate of most proteins is about 2% [65]. Five independent screens of 48 random conditions (for a total of 240 conditions) would provide a 99.2% ($1-0.98^{288}$) chance of obtaining a hit. For a protein with a success rate of only 0.2%, the probability of obtaining a hit would still be about a 40%. So the low- or medium-throughput laboratory can expect a reasonable amount of success by screening with five commercially available or custom random screens. The high-throughput laboratory, however, can easily perform more trials, and these are sometimes crucial. The experience of one HTGS laboratory is that their screen of 1536 conditions often produces 50 or more lead conditions for a target. But there are also several cases (out of 1300 proteins and nucleic acids) where only one condition produced a crystal [66].

While more trials afford more opportunities for success, the observation and analysis of the screens can be deceptively time consuming. For example, the screening facility operated by the Hauptman-Woodward Institute accepts crystallization samples from the scientific community [66]. This facility uses a crystallization plate

containing 1536 different crystallization experiments, and it records seven images of each experiment: just prior to adding target, just after adding target; 1 day, 3 days, 1 week, 2 weeks and 1 month after setup. At 5 s per image, it requires 15 hours of uninterrupted viewing time to examine all 10,752 images. Obviously, the viewing cannot be done continuously, and the viewing must be spread out over many days.

This chemical composition space can be sampled using several commercially available screens (table 2). One study of the relative efficacy of commercially available Sparse Matrix type screens used a test set composed of 25 proteins [19]. The Stura Footprint Screen crystallized 8 proteins, the Crystal Screen plus Crystal Screen 2 crystallized 6 proteins and the Wizard Screen (I and II) 3. Another study using 19 proteins compared ostensibly identical screens from Hampton Research and Molecular Dimensions [67]. 54 crystallizations were produced only by the Hampton Research formulation, and 73 crystallizations produced only by the Molecular Dimensions formulation. Only 38 of 165 crystallizations were produced by both formulations. The formulations are, in fact, not identical since Hampton Research uses HCl and NaOH to adjust the pH of its solutions, whereas Molecular Dimensions uses glacial acetic acid. Since acetic acid is volatile, it is possible for the pH of the Molecular Dimensions' solutions to increase with time. The same study also examined the efficacy of the Wizard I and II screens and concluded that all three pairs of screens were effective and that all three produced distinctly different results.

Alternatives to commercially available Sparse Matrix screens are available. Crystool is a tool for creating one's own screen [65]. Alternative experimental designs for sampling the crystallization space in a statistical manner have also been described [68–70].

More methodical searches can be used instead of or in addition to Sparse Matrix approaches. The method used most commonly prior to the introduction of the Sparse Matrix methods was to examine the effect of pH and precipitant concentration systematically at a fixed temperature. This approach is the idea behind some commercially available 'grid screens.' It is very methodical, but it is also slow and requires much material. The notion is that as the precipitant concentration increases, trials will progress from clear to crystal to precipitate. The operational question is whether or not the crystalline phase is produced between the clear and the precipitate phases. Successively finer sampling is done in the pH-precipitant plane to essentially trace out a quasi-solubility curve (fig. 1). There are also screens intermediate between the methodical grid screens and the Sparse Matrix. Such screens typically vary some variables systematically (e.g. pH) and others somewhat randomly (e.g. a different choice of salt type for each pH) [71].

Another alternative to Sparse Matrices is called Reverse Screening [72]. In Reverse Screening one uses a priori considerations to limit the choice of precipitants in the search for conditions that produce large increases in supersaturation for small changes in precipitant concentration. Streak seeding or cross seeding is recommended to separate the problem of crystal growth from the problem of crystal nucleation [51]. Vapor diffusion is recommended because of its concentrating effect.

Once a screen has been conducted, there are several 'last-resort' manipulations that can be done to obtain a crystal or more information about the solubility of the target. Among these are moving the tray to another temperature (e.g. from 22 to 4 °C), adding additives (e.g. β-octylglucoside) (this is more appropriate for vapor diffusion where the drop is reconcentrated after the additive is added), adding a volatile acid (acetic acid) or base (ammonium hydroxide) to run a pH gradient [71], streak seeding and adding silicone oil to paraffin oil to prompt dehydration of the microbatch under oil [62].

Optimization

Optimizing a lead crystallization condition obtained from a screen, i.e. a 'hit', is more straightforward than making

Table 2. Commercially available screens.

Type	Name	No. of conditions	Company	URL	Ref.
Sparse Matrix	Crystal Screen	50	Hampton Research	http://www.hamptonresearch.com/)	[9]
Sparse Matrix	Crystal Screen 2	48	Hampton Research	http://www.hamptonresearch.com/)	[89]
Sparse Matrix	Structure Screen 1	50	Molecular Dimensions	http://www.moleculardimensions.com/	
Sparse Matrix	Structure Screen 2	50	Molecular Dimensions	http://www.moleculardimensions.com/	
Sparse Matrix	Wizard I	48	Emerald BioStructures	www.decode.com/emeraldbiostructures	
Sparse Matrix	Wizard II	48	Emerald BioStructures	www.decode.com/emeraldbiostructures	
Semisystematic	Clear Strategy Screen I	24	Molecular Dimensions	http://www.moleculardimensions.com/	[90]
Semisystematic	Clear Strategy Screen II	24	Molecular Dimensions	http://www.moleculardimensions.com/	[90]
Semisystematic	JBScreen	240	Jena Bioscience	http://www.jenabioscience.com	
Semisystematic	PEG/Ion Screen	48	Hampton Research	http://www.hamptonresearch.com/	
Semisystematic	Stura Footprint	48	Molecular Dimensions	http://www.moleculardimensions.com/	[51]

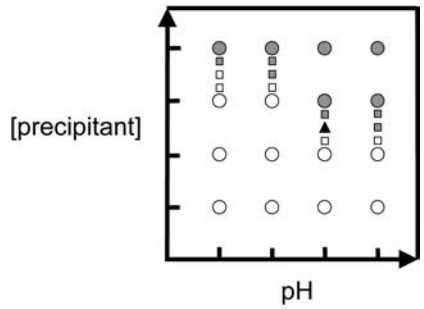

Figure 1. The classic pH-precipitant screening strategy. The first round of conditions comprise a course grid (circles) that produce either clear drops (open symbols), crystals (triangle) or precipitate (filled symbols). The interval between pHs is typically 0.5 or 1.0 pH units. The interval between precipitant concentrations depends on the precipitant but is typically 0.5–1.0 M for salts and 5–10% for PEGs and organics. The second round (squares) more finely samples the precipitant concentrations between the first round clear and precipitate conditions. Additional rounds with finer precipitant concentration sampling may be done. In this example, crystallization conditions were located during the second round.

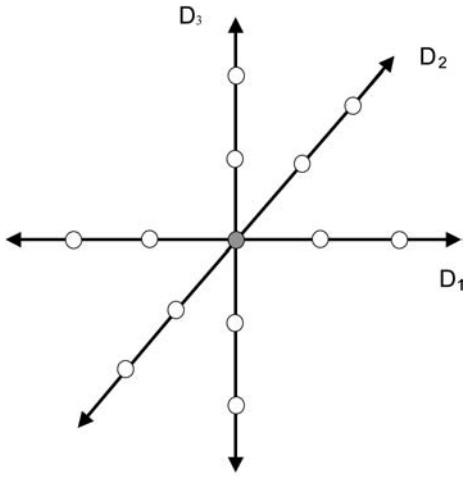

Figure 2. A star search in three dimensions. Conditions to be tested (open circles) are based on the origin condition (filled circle) and change only one parameter at a time. Five conditions, including the origin, are examined along each dimension.

the wide range of decisions before the screening, but there are still many different approaches used by different investigators. Most, if not all, techniques have their origins in mathematical optimization methods that involve quantifiable functions. Some crystallization optimization methods accordingly quantify the results of the crystallization experiments in order to generate the next round of conditions to test, e.g. [73]. Other approaches, however, evaluate the results qualitatively and only use mathematical formalisms to generate new conditions in a methodical manner, e.g. [74]. In either case, one can view the optimization as an iterative procedure. At each step there is the current optimum and a new set of conditions to test. Oftentimes, one will branch out and pursue several distinct optima, each as a separate optimization problem. As progress is made, one switches from judging the quality based on visual properties (e.g. size, lack of epitaxial growth) to X-ray diffraction quality (e.g. resolution, mosaicity).

One simple approach for generating the next round of conditions to test is to systematically vary each parameter while keeping the others constant. We call this a star search in our laboratory because trying to sketch many one-dimensional searches through a common origin on paper results in a starlike figure (fig. 2). For example, imagine that the crystallization conditions in table 3 produced small crystals when 10 mg ml^{-1} protein solution was added 1:1 to the reservoir in a sitting drop experiment incubated at 22 °C. This lead condition is the origin for 6 one-dimensional searches. Each one-dimensional search consists of five conditions: the origin, two conditions below the origin, and two conditions above. In our experience, having five points along a dimension is more useful for assessing trends than three points, especially at the beginning of the optimization.

The first star search provides several types of information. First, it may produce an improved crystallization condition. Second, it shows how sensitive the crystallization in each dimension. Those dimensions that had little or no effect may be omitted in the next several star searches. Third, the first star search may suggest that the step sizes in a particular dimension be increased or decreased.

Assuming that there is at least one improved crystallization condition in the first star search, there are two ways to select the new origin for the second star search. The conservative approach is to simply use the best condition obtained from the first star search. The aggressive approach is to extrapolate from the improvements seen in two or more dimensions to a condition not previously tested. For example in table 3, if the first star search produced larger crystals in both the 140 mM sodium citrate and 25% MPD conditions, the new origin could be taken to be 140 mM sodium citrate and 25% MPD. When opting for this aggressive approach, it is prudent to add one condition from the first star search to serve as a positive control. (With the conservative approach, the origin of the second star search is a positive control because it repeats a condition from the first round.) The generation of subsequent star searches can proceed in the same manner as generating the second from the first.

When the star search approach above no longer generates improved conditions, there are two modifications that can be made. One is to vary two parameters simultaneously in a ‘diagonal star search’. In the example, step increases in the sodium citrate concentration may be accompanied by step decreases in MPD concentration. The diagonal star searches increases the search pattern within a set of dimensions.

Table 3. Crystallization example.

Dimension	Lead condition	First star search	Alternative dimensions
Protein	10 mg/ml	6, 8, 12, 14 mg/ml	± His_6 tag
Sodium citrate	100 mM	60, 80, 120, 140 mM	potassium citrate, sodium chloride
Sodium HEPES (pK_a 7.47)	100 mM	60, 80, 120, 140 mM	PIPES (pK_a 6.76), BICINE (pK_a 8.26)
pH	7.5	7.1, 7.3, 7.7, 7.9	–
2-methyl-2,4-pentanediol	30%	20, 25, 35, 40%	PEG 400
Temperature	22 °C	18, 20, 24, 26 °C	–

The other modification is to test new dimensions. The number of dimensions can be increased by considering other compounds as alternatives to the species used in the Sparse Matrix condition (table 3). The number of dimensions can also be increased by changing other factors in the crystallization such as the rate of equilibration, changing to hanging drop or dialysis, crystallization in microgravity [75], and including additives like DTT or β-octylglucoside. The speed of crystal growth is another dimension. The conventional wisdom has been that it is better to grow crystals slowly (over the course of several days) than fast (overnight). The results of nanoliter crystallizations, however, suggest that it is sometimes preferable to crystallize very rapidly (in hours) [76]. One advantage of speed is that oxidation and other degradative processes will be minimized.

An alternative to the star search approach is the successive grid search approach [77]. In this approach, a grid of conditions on a plane are tested simultaneously (fig. 3). The grid size around the best conditions is decreased in each successive round. The successive grid search is more thorough than the star search, but it requires more material.

Other alternatives to the star search are based on factorial methods. The Central Composite method probes along the axis of each dimension and at each vertex of an N-dimensional hypercube (fig. 4). The Box-Behnken method probes the N-dimensional space by testing all combinations of changing two parameters from the origin value (fig. 5). These two designs have been implemented in concert with crystallization hardware [74].

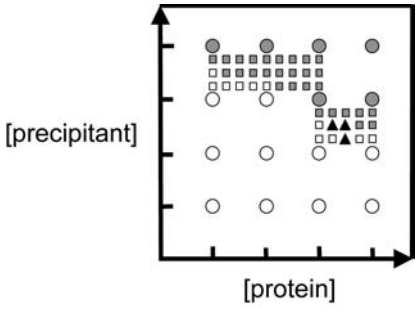

Figure 3. The successive grid search. Two dimensional grids are used in each round rather than the one-dimensional line search of the classic pH-precipitant search. Symbols have the same meaning as in figure 1.

If several hits are obtained from the initial screening of several Sparse Matrices, the Sparse Matrix philosophy can be used to generate a new, secondary Sparse Matrix [78]. The conditions of the hits from the first round of Sparse Matrices can be entered in a spreadsheet with separate fields for precipitant concentration, precipitant type, additive concentration, additive type, pH and buffer. Additional rows are added that expand the concentration ranges sampled for each precipitant, additive and pH. Rows are added for similar compounds. For example, if crystallization occurred in ammonium sulfate, add sodium sulfate and ammonium chloride. If PEG 8000 was the precipitant, add PEG 6000 and PEG 10000. The columns are then shuffled using the sort function on an auxiliary column that contains random numbers. The result is a Sparse Matrix over the input sample space.

An entirely different dimension available to optimization is the use of seeding [79, 80]. Seeding decouples nucle-

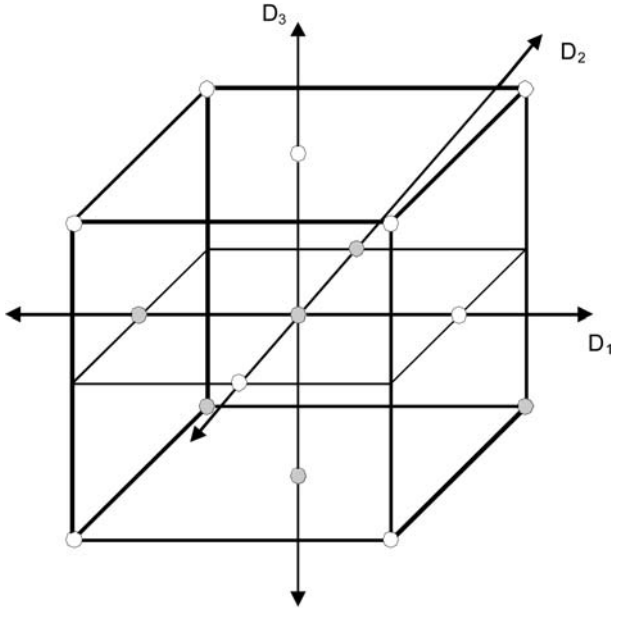

Figure 4. The central composite design. This design contains three types of points based on a step size for each dimension: (i) the origin, (ii) positive and negative single steps along each dimension, and (iii) all combinations of taking one step (positive or negative) in each dimension. In terms of a hypercube, these points correspond to the center, the face centers and the vertices, respectively.

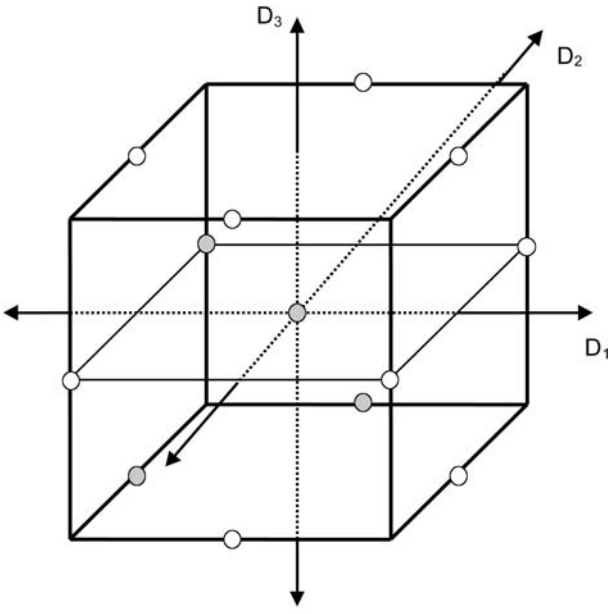

Figure 5. The Box-Behnken design. This design uses a step size for each dimension and contains two types of points: the origin and all combinations of steps (positive or negative) taken in two separate dimensions.

ation from crystal growth. Crystal growth can occur under conditions of much lower supersaturations, and superior crystals are often formed. Three types of seeding are macroseeding, microseeding and streak seeding. Macroseeding uses crystals that can be identified in a microscope and are typically 5–50 μm in size. Only one seed is introduced into the new crystallization drop. Microseeds are too small to be seen in a microscope. They may be generated in many ways [81]. One method is to crush a crystal, centrifuge the solution to eliminate large fragments and then serially dilute the supernatant in a stabilizing solution. Small volumes of the diluted seeds are added to new crystallization drops. A method similar to microseeding has been developed by using vapor diffusion to equilibrate a drop against a high concentration reservoir for a limited time to induce nucleation before switching to a lower concentration reservoir for crystal growth [82]. Streak seeding [79] using an animal hair is the fastest seeding method, but is less reproducible. The hair is touched or stroked over a crystal in its mother liquor and then drawn through a fresh drop containing protein. The hair picks up crystallization nuclei from the crystal and deposits them in the new solution.

Streak seeding can also be used to obtain additional 'hits' and to identify crystalline precipitate. Some investigators obtain new hits by using crystals obtained from initial screens to streak-seed all the drops that remained clear, e.g. [80]. Streak seeding can also be used to distinguish amorphous precipitate from microcrystalline precipitate [83]. Microcrystalline precipitate will produce small crystals when used to streak-seed a fresh drop, whereas amorphous precipitate will only produce more precipitate.

Future developments

Crystallization strategies are expected to undergo substantial refinement as the results of HTGS projects and their technologies become available. Many of the current strategies have benefited from the Biological Macromolecule Crystallization Database (http://wwwbmcd.nist.gov:8080/bmcd/bmcd.html) [64, 84]. As of June 2003, the BMCD included 3547 crystal entries from 2526 biological macromolecules for which diffraction-quality crystals have been obtained. These include proteins, protein-protein complexes nucleic acid, nucleic acid-nucleic acid complexes, protein-nucleic acid complexes and viruses. The HTGS crystallization databases contain much more than recipes for specific crystals. They contain negative as well as positive results, and basic properties of the crystallization targets. The HTGS targets, however, are only a subset of the targets of interest to structural biologists. The HTGS projects typically exclude complexes and membrane proteins. For example, the Protein Structure Factory in Berlin restricts its list of target proteins according to the criteria [18]: small (500 amino acids or fewer), water-soluble (no predicted transmembrane helices, coiled-coil or 'extended low-complexity sequences') proteins that are monomeric or homo-oligomeric are chosen. So while the HTGS results will be very informative, their range of applicability may be limited.

General trends are beginning to emerge from the HTGS efforts. It is difficult to make exacting comparisons between the results obtained by different consortia, because each consortium uses different techniques. For example, the overall efficiency of obtaining lead crystallization conditions in HTSG efforts ranges from 3 to 30% depending on the consortium [27, 59]. Nonetheless, some trends hold across the different consortia.

One consistent observation is that the individual conditions contained in the commercially available screens differ greatly in their effectiveness. In one study, 432 *T. hermotoga maritime* proteins were crystallized in a 480-condition screen. 94% of those proteins could have been crystallized by just using 192 of the best conditions [85]. Another study evaluated Hampton Research Crystal Screen 1 against 755 different proteins from six different bacterial organisms [86]. 45% of the proteins produced crystals. Only 24 conditions from Screen 1 were necessary to produce 94% of the 'hits' and the use of just six conditions from Screen 1 would still produce 60% of the hits. A third study reported the effectiveness of 11 sparse matrix and grid screens available through Hampton Research and Emerald Biosystems [87]. Three Hampton PEG screens (PEG Ion, PEG/LiCl and PEG 6K) pro-

duced 20–100% more hits than individual sparse matrix screens. The conclusion, though, was that here is no magic bullet screen available and that one must try several screens. A fourth study has listed the frequency with which reagents have been used in 203 successful random crystallization experiments, and the frequency of the reagents varies widely [60]. One expects these results to be the basis of a next generation of screens.

Many other issues should be addressed by the future analysis of the HTGS data. Can crystallization conditions be predicted by some combination of physical properties? How does the organism used for overexpression affect the choice of crystallization conditions screened? How does crystal quality or the ability to get crystals depend on purity as judged by SDS-PAGE, native-PAGE, isoelectric focusing PAGE, mass spectrophotometry, dynamic light scattering and so on? Are the identities of the cations and anions important? Are there crystallization conditions that repeatedly produce poorly diffracting crystals or, conversely, that usually produce highly diffracting crystals?

One of the most sought after developments from the HTGS data would be the ability to predict crystallization conditions based on the target's physical properties. An approach to rationally predicting the crystallization conditions or the types of crystallization conditions conducive for a particular (type of) protein have not yet emerged [66]. But one high-throughput crystallization effort has reported some success in using a neural network or Chernov algorithm to predict crystallization conditions based on a small sampling of crystallization space [88]. Such predictors may provide the foundation to extending predictive crystallization to other types of proteins, to nucleic acids and to viruses.

Conclusion

There is no shortage of tools available for crystallization of a new protein. The challenge is to judiciously choose the approaches appropriate for the target and feasible with the resources at hand. The medium-throughput laboratory may be best served by preparing several different targets by making different genetic constructs and/or by preparing different ligation states. Each target should be screened through at least five 48-condition Sparse Matrix screens at 4 and 22 °C. If crystals are not obtained or if optimization of the lead crystallization conditions does not produce diffraction-quality crystals, analytical techniques such as mass spectrometry and dynamic light scattering should be used to identify likely problem areas. Resolution of the problems may require creating new constructs or modifying the preparation of the target. Results emerging from structural genomics may soon help move crystallization from less of a trial-and-error approach to a more directed method.

Acknowledgements. Grateful thanks to Dana Roeber and Linda Martin for assistance in preparing the manuscript.

1 Kam Z., Shore H. B. and Feher G. (1978) On the crystallization of proteins. J. Mol. Biol. **123:** 539–555
2 Blundell T. L. and Johnson L. N. (1976) Crystallization of proteins. In: Protein Crystallography. Academic Press, London pp. 59–82
3 Wyckoff H. W., Hirs C. H. W. and Timasheff S. N. (eds) (1985) Diffraction Methods for Biological Macromolecules. Academic Press, New York
4 Carter C. W. Jr (ed.) (1990) Protein and Nucleic Acid Crystallization. Academic Press, New York
5 Ducruix A. and Giege R. (1992) Methods of crystallization. In: Crystallization of Nucleic Acids and Proteins: A Practical Approach, Giege R. and Ducruix A. (eds.), IRL Press, New York pp. 73–98
6 Carter C. W. Jr and Sweet R. M. (eds) (1997) Macromolecular Crystallography, Part A. Academic Press, New York
7 Bergfors T. M. (ed.) (1999) Protein Crystallization: Techniques, Strategies and Tips. International University Line, La Jolla, CA
8 McPherson A. (1999) Crystallization of Biological Macromolecules, Cold Spring Harbor Laboratory Press, Cold Spring Harbor, NY
9 Jancarik J. and Kim S. H. (1991) Sparse matrix sampling: a screening method for crystallization of proteins. J. Appl. Crystallogr. **24:** 409–411
10 Norvell J. C. and Machalek A. Z. (2000) Structural genomics programs at the US National Institute of General Medical Sciences. Nat. Struct. Biol. **7 Suppl.:** 931
11 Burley S. K. (2000) An overview of structural genomics. Nat. Struct. Biol. **7 Suppl.:** 932–934
12 Terwilliger T. C. (2000) Structural genomics in North America. Nat. Struct. Biol. **7 Suppl.:** 935–939
13 Heinemann U. (2000) Structural genomics in Europe: slow start, strong finish? Nat. Struct. Biol. **7 Suppl**:940-942
14 Yokoyama S., Hirota H., Kigawa T., Yabuki T., Shirouzu M., Terada T. et al. (2000) Structural genomics projects in Japan. Nat. Struct. Biol. **7 Suppl.:** 943–945
15 Dry S., McCarthy S. and Harris T. (2000) Structural genomics in the biotechnology sector. Nat. Struct. Biol. **7 Suppl.:** 946–949
16 Abola E., Kuhn P., Earnest T. and Stevens R. C. (2000) Automation of X-ray crystallography. Nat. Struct. Biol. **7 Suppl.:** 973–977
17 Stevens R. C. (2000) Design of high-throughput methods of protein production for structural biology. Struct. Fold Des. **8:** R177–R185
18 Heinemann U., Bussow K., Mueller U. and Umbach P. (2003) Facilities and methods for the high-throughput crystal structural analysis of human proteins. Acc. Chem. Res. **36:** 157–163.
19 Sulzenbacher G., Gruez A., Roig-Zamboni V., Spinelli S., Valencia C., Pagot F. et al. (2002) A medium-throughput crystallization approach. Acta Crystallogr. D. Biol. Crystallogr. **58:** 2109–2115
20 Bucher M. H., Evdokimov A. G. and Waugh D. S. (2002) Differential effects of short affinity tags on the crystallization of *Pyrococcus furiosus* maltodextrin-binding protein. Acta Crystallogr. D. Biol. Crystallogr. **58:** 392–397
21 Savchenko A., Yee A., Khachatryan A., Skarina T., Evdokimova E., Pavlova M. et al. (2003) Strategies for structural proteomics of prokaryotes: quantifying the advantages of studying orthologous proteins and of using both NMR and X- ray crystallography approaches. Proteins **50:** 392–399
22 Kuge M., Fujii Y., Shimizu T., Hirose F., Matsukage A. and Hakoshima T. (1997) Use of a fusion protein to obtain crystals suitable for X-ray analysis: crystallization of a GST-fused protein containing the DNA-binding domain of DNA replica-

tion-related element-binding factor, DREF. Protein Sci. **6:** 1783–1786
23 Center R. J., Kobe B., Wilson K. A., Teh T., Howlett G.. J., Kemp B. E. et al. (1998) Crystallization of a trimeric human T cell leukemia virus type 1 gp21 ectodomain fragment as a chimera with maltose-binding protein. Protein Sci. **7:** 1612–1619
24 Byrne B., Abramson J., Jansson M., Holmgren E. and Iwata S. (2000) Fusion protein approach to improve the crystal quality of cytochrome bo(3) ubiquinol oxidase from *Escherichia coli*. Biochim. Biophys. Acta **1459:** 449–455
25 Niemann H. H., Knetsch M. L., Scherer A., Manstein D. J. and Kull F. J. (2001) Crystal structure of a dynamin GTPase domain in both nucleotide-free and GDP-bound forms. EMBO J. **20:** 5813–5821
26 Liu Y., Manna A., Li R., Martin W. E., Murphy R. C., Cheung A. L. et al. (2001) Crystal structure of the SarR protein from *Staphylococcus aureus*. Proc. Natl. Acad. Sci. USA **98:** 6877–6882
27 Dale G. E., Oefner C. and D'Arcy A. (2003) The protein as a variable in protein crystallization. J. Struct. Biol. **142:** 88–97
28 Yeh J. I., Biemann H. P., Prive G. G., Pandit J., Koshland D. E. Jr and Kim S. H. (1996) High-resolution structures of the ligand binding domain of the wild-type bacterial aspartate receptor. J. Mol. Biol. **262:** 186–201
29 Sprangers R., Selenko P., Sattler M., Sinning I. and Groves M. R. (2003) Definition of domain boundaries and crystallization of the SMN Tudor domain. Acta Crystallogr. D. Biol. Crystallogr. **59:** 366–368
30 Lawson D. M., Artymiuk P. J., Yewdall S. J., Smith J. M., Livingstone J. C., Treffry A. et al. (1991) Solving the structure of human H ferritin by genetically engineering intermolecular crystal contacts. Nature **349:** 541–544
31 Green M. E., Kirkland N. and Ng J. D. (2001) Effect of a mutation at arginine 301 on the stability, crystal quality and the preliminary crystallographic analysis of recombinant canavalin from Canavalia ensiformis. J. of Crystal Growth **232:** 387–398
32 Edmundson A. B. and Borrebaeck C. A. (1998) Progress in programming antibody fragments to crystallize. Immunotechnology **3:** 309–317
33 Wingren C., Edmundson A. B. and Borrebaeck C. A. (2003) Designing proteins to crystallize through beta-strand pairing. Protein Eng. **16:** 255–264
34 Charron C., Kern D. and Giege R. (2002) Crystal contacts engineering of aspartyl-tRNA synthetase from Thermus thermophilus: effects on crystallizability. Acta Crystallogr. D. Biol. Crystallogr. **58:** 1729–1733
35 Oubridge C., Ito N., Teo C. H., Fearnley I. and Nagai K. (1995) Crystallisation of RNA-protein complexes. II. The application of protein engineering for crystallisation of the U1A protein-RNA complex. J. Mol. Biol. **249:** 409–423
36 Chen P., Tsuge H., Almassy R. J., Gribskov C. L., Katoh S., Vanderpool D. L. et al. (1996) Structure of the human cytomegalovirus protease catalytic domain reveals a novel serine protease fold and catalytic triad. Cell **86:** 835–843
37 Longenecker K. L., Garrard S. M., Sheffield P. J. and Derewenda Z. S. (2001) Protein crystallization by rational mutagenesis of surface residues: Lys to Ala mutations promote crystallization of RhoGDI. Acta Crystallogr. D. Biol. Crystallogr. **57:** 679–688
38 Dyda F., Hickman A. B., Jenkins T. M., Engelman A., Craigie R. and Davies D. R. (1994) Crystal structure of the catalytic domain of HIV-1 integrase: similarity to other polynucleotidyl transferases. Science **266:** 1981–1986
39 Pedelacq J. D., Piltch E., Liong E. C., Berendzen J., Kim C. Y., Rho B. S. et al. (2002) Engineering soluble proteins for structural genomics. Nat. Biotechnol. **20:** 927–932
40 Schwede T. F., Badeker M., Langer M., Retey J. and Schulz G. E. (1999) Homogenization and crystallization of histidine ammonia-lyase by exchange of a surface cysteine residue. Protein Eng. **12**:151–153
41 Heinz D. W. and Matthews B. W. (1994) Rapid crystallization of T4 lysozyme by intermolecular disulfide cross-linking. Protein Eng. **7:** 301–307
42 McElroy H. E., Sisson G. W., Schoettlin W. E., Aust R. M. and Villafranca J. E. (1992) Studies on engineering crystallizability by mutation of surface residues of human thymidylate synthase. J. Crystal Growth **122:** 265–272
43 Braig K., Otwinowski Z., Hegde R., Boisvert D. C., Joachimiak A., Horwich A. L. et al. (1994) The crystal structure of the bacterial chaperonin GroEL at 2.8 Å. Nature **371:** 578–586
44 Zhang F., Basinski M. B., Beals J. M., Briggs S. L., Churgay L. M., Clawson D. K. et al. (1997) Crystal structure of the obese protein leptin-E100. Nature **387:** 206–209
45 Rayment I., Rypniewski W. R., Schmidt-Base K., Smith R., Tomchick D. R., Benning M. M. et al. (1993) Three-dimensional structure of myosin subfragment-1: a molecular motor. Science **261:** 50–58
46 Kwong P. D., Wyatt R., Desjardins E., Robinson J., Culp J. S., Hellmig B. D. et al. (1999) Probability analysis of variational crystallization and its application to gp120, the exterior envelope glycoprotein of type 1 human immunodeficiency virus (HIV-1). J. Biol. Chem. **274:** 4115–4123
47 Kendrew J. C. and Parrish R. G. (1956) The crystal structure of myoglobin. III. Sperm-whole myoglobin. Proceedings of the Royal Society of London **A 238:** 305–324
48 Nikolov D. B., Hu S. H., Lin J., Gasch A., Hoffmann A., Horikoshi M. et al. (1992) Crystal structure of TFIID TATA-box binding protein. Nature **360:** 40–46
49 Jordan S. R., Whitcombe T. V., Berg J. M. and Pabo C. O. (1985) Systematic variation in DNA length yields highly ordered repressor-operator cocrystals. Science **230:** 1383–1385
50 Kovari L. C., Momany C. and Rossmann M. G. (1995) The use of antibody fragments for crystallization and structure determinations. Structure **3:** 1291–1293
51 Stura E. A., Nemerow G. R. and Wilson I. A. (1992) Strategies in the crystallization of glycoproteins and protein complexes. J. Crystal Growth **122:** 273–285
52 Kabsch W., Mannherz H. G., Suck D., Pai E. F. and Holmes K. C. (1990) Atomic structure of the actin:DNase I complex. Nature **347:** 37–44
53 Rupp B. (2003) High-throughput crystallography at an affordable cost: the TB structural genomics consortium crystallization facility. Acc. Chem. Res. **36:** 173–181
54 Horsefield R., Yankovskaya V., Tornroth S., Luna-Chavez C., Stambouli E., Barber J. et al. (2003) Using rational screening and electron microscopy to optimize the crystallization of succinate:ubiquinone oxidoreductase from Escherichia coli. Acta Crystallogr. D. Biol. Crystallogr. **59:** 600–602
55 D'Arcy A. (1994) Crystallizing proteins – a rational approach? Acta Cryst. D. Biol. Crystallogr. **D50:** 469–471
56 Zulauf M. and D'Arcy A. (1992) Light scattering of proteins as a criterion for crystallization. J. Crystal Growth **122:** 102–106
57 Ferre-D'Amare A. R. and Burley S. K. (1994) Use of dynamic light scattering to assess crystallizability of macromolecules and macromolecular assemblies. Structure **2:** 357–359
58 Baldock P., Mills V. and Stewart P. S. (1996) A comparison of microbatch and vapour diffusion for initial screening of crystallization conditions. J. Crystal Growth **168:** 170–174
59 Hui R. and Edwards A. (2003) High-throughput protein crystallization. J. Struct. Biol. **142:** 154–161
60 Rupp B. (2003) Maximum-likelihood crystallization. J. Struct. Biol. **142:** 162–169
61 Mikol V., Rodeau J.-L. and Giege R. (1990) Experimental determination of water equilibration rates in the hanging drop method of protein crystallization. Anal. Biochem. **186:** 332–339
62 D'Arcy A., Mac Sweeney A., Stihle M. and Haber A. (2003) The advantages of using a modified microbatch method for

rapid screening of protein crystallization conditions. Acta Crystallogr. D. Biol. Crystallogr. **59**: 396–399
63 Gilliland G. L., Tung M., Blakeslee D. M. and Ladner J. (1994) The Biological Macromolecule Crystallization Database, Version 3.0: new features, data and the NASA Archive for Protein Crystal Growth Data. Acta Crystallogr. D. Biol. Crystallogr. **D50:** 408–413
64 Gilliland G. L., Tung M. and Ladner J. E. (2002) The Biological Macromolecule Crystallization Database: crystallization procedures and strategies. Acta Crystallogr. D. Biol. Crystallogr. **58:** 916–920
65 Segelke B. W. (2001) Efficiency analysis of sampling protocols used in protein crystallization screening. J. Crystal. Growth **232:** 553–562
66 Luft J. R., Collins R. J., Fehrman N. A., Lauricella A. M., Veatch C. K. and DeTitta G. T. (2003) A deliberate approach to screening for initial crystallization conditions of biological macromolecules. J. Struct. Biol. **142:** 170–179
67 Wooh J. W., Kidd R. D., Martin J. L. and Kobe B. (2003) Comparison of three commercial sparse-matrix crystallization screens. Acta Crystallogr. D. Biol. Crystallogr. **59:** 769–772
68 Carter C. J. and Carter C. W. (1979) Protein crystallization using incomplete factorial experiments. J. Biol. Chem. **254:** 12219–12223
69 Carter C. W. Jr (1990) Efficient Factorial Designs and the Analysis of Macromolecular Crystal Growth Conditions. Methods: A Companion to Methods in Enzymology **1:** 12–24
70 Kingston R. L., Baker H. M. and Baker E. N. (1994) Search designs for protein crystallization based on orthogonal Arrays. Acta Cryst. D. Biological Crystallogr. **D50:** 429–440
71 McPherson A. (1992) Two approaches to the rapid screening of crystallization conditions. J. Crystal Growth **122:** 161–167
72 Stura E. A., Satterthwait A. C., Calvo J. C., Kaslow D. C. and Wilson I. A. (1994) Reverse screening. Acta Cryst. D. Biol. Crystallogr. **D50:** 448–455
73 Carter C. W. and Yin Y. (1994) Quantitative analysis in the characterization and optimization of protein crystal growth. Acta Crystallogr. **D50:** 572–590
74 Shaw Stewart P. and Baldock P. F. M. (1999) Practical experimental design techniques for automatic and manual protein crystallization. J. Crystal Growth **196:** 665–673
75 Kundrot C. E., Judge R. A., Pusey M. L. and Snell E. H. (2001) Microgravity and macromolecular crystallography. Crystal Growth Design **1:** 87–99
76 Santarsiero B. D., Yegian D. T., Lee C. C., Spraggon G., Gu J., Scheibe D. et al. (2002) An approach to rapid protein crystallization using nanodroplets. J. Appl. Crystallogr. **35:** 278–281
77 Cox M. J. and Weber P. C. (1987) Experiments with automated protein crystallization. J. Appl. Crystallogr. **20:** 366–373
78 Doudna J. A., Grosshans C., Gooding A. and Kundrot C. E. (1993) Crystallization of ribozymes and small RNA motifs by a sparse matrix approach. Proc. Natl. Acad. Sci. USA **90:** 7829–7833
79 Stura E. A. and Wilson I. A. (1990) Analytical and production seeding techniques. Methods: A Companion to Methods in Enzymology **1:** 38–49
80 Bergfors T. (2003) Seeds to crystals. J. Struct. Biol. **142:** 66–76
81 Luft JR and Detitta GT (1999) A method to produce microseed stock for use in the crystalliztion of biological macromolecules. Biol. Crystallogr. **55:** 988–993
82 Saridakis E. and Chayen N. E. (2000) Improving protein crystal quality by decoupling nucleation and growth in vapor diffusion. Protein Sci. **9:** 755–757
83 Stura E. A. and Wilson I. A. (1991) Applications of the streak seeding technique in protein crystallization. J. of Crystal Growth **110:** 70–282
84 Gilliland G. L. (1988) A biological macromolecule crystallization database: a basis for a crystallization strategy. J. Crystal Growth **90**: 51–59
85 Lesley S. A., Kuhn P., Godzik A., Deacon A. M., Mathews I., Kreusch A. et al. (2002) Structural genomics of the Thermotoga maritima proteome implemented in a high-throughput structure determination pipeline. Proc. Natl. Acad. Sci. USA **99:** 11664–11669
86 Kimber M. S., Vallee F., Houston S., Necakov A., Skarina T., Evdokimova E. et al. (2003) Data mining crystallization databases: knowledge-based approaches to optimize protein crystal screens. Proteins **51:** 562–568
87 Hosfield D., Palan J., Hilgers M., Scheibe D., McRee D. E. and Stevens R. C. (2003) A fully integrated protein crystallization platform for small-molecule drug discovery. J. Struct. Biol. **142:** 207–217
88 DeLucas L. J., Bray T. L., Nagy L., McCombs D., Chernov N., Hamrick D. et al. (2003) Efficient protein crystallization. J. Struct. Biol. **142:** 188–206
89 Cudney B., Patel S., Weisgraber K., Newhouse Y. and McPherson A. (1994) Screening and optimization strategies for macromolecular crystal growth. Acta Crystallogr. D **50:** 414–423
90 Brzozowski A. M. and Walton J. (2001) Clear strategy screens for macromolecular crystallization. J. Appl. Crystallogr. **34:** 97–101