Acta Crystallographica Section F
Structural Biology Communications
ISSN 2053-230X

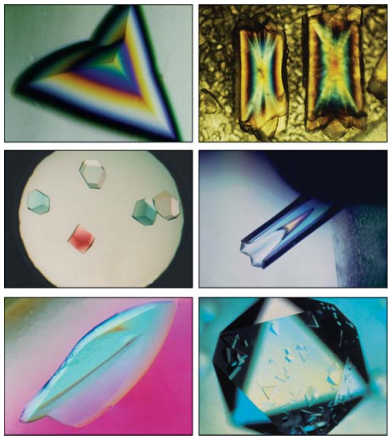
Introduction to protein crystallization

Alexander McPherson[a]* and Jose A. Gavira[b]

[a]Department of Molecular Biology and Biochemistry, University of California, Irvine, 560 Steinhaus Hall, Irvine, CA 92697-3900, USA, and [b]Laboratorio de Estudios Cristalográficos, IACT (CSIC–UGR), Avenida de las Palmeras 4, 18100 Armilla, Granada, Spain

Correspondence e-mail: amcphers@uci.edu

Received 15 November 2013
Accepted 7 December 2013



Protein crystallization was discovered by chance about 150 years ago and was developed in the late 19th century as a powerful purification tool and as a demonstration of chemical purity. The crystallization of proteins, nucleic acids and large biological complexes, such as viruses, depends on the creation of a solution that is supersaturated in the macromolecule but exhibits conditions that do not significantly perturb its natural state. Supersaturation is produced through the addition of mild precipitating agents such as neutral salts or polymers, and by the manipulation of various parameters that include temperature, ionic strength and pH. Also important in the crystallization process are factors that can affect the structural state of the macromolecule, such as metal ions, inhibitors, cofactors or other conventional small molecules. A variety of approaches have been developed that combine the spectrum of factors that effect and promote crystallization, and among the most widely used are vapor diffusion, dialysis, batch and liquid–liquid diffusion. Successes in macromolecular crystallization have multiplied rapidly in recent years owing to the advent of practical, easy-to-use screening kits and the application of laboratory robotics. A brief review will be given here of the most popular methods, some guiding principles and an overview of current technologies.

1. Some history

The first protein crystals, of hemoglobin from worms and fishes (Hunte *et al.*, 2003; Funke, 1851; Reichert & Brown, 1909; Hunefeld, 1840), were observed more than 150 years ago by German biologists (see McPherson, 1991, 1999, for reviews of the history of protein crystal growth). They remained a laboratory curiosity for many years until the 1880s, when Ritthausen (1880, 1881) and Osborne (1891, 1892, 1894, 1899) crystallized, for the purpose of purification, a series of plant seed proteins. Purification and demonstration of purity were the primary reasons that techniques were developed for the crystallization of naturally occurring proteins in the laboratory. A notable success was the crystallization of hen-egg albumin, or ovalbumin (Hofmeister, 1890; Hopkins & Pincus, 1898). Protein crystallization was marked by major successes throughout the 1920s and 1930s, with the crystallization of insulin (Abel *et al.*, 1927) and the demonstration by Sumner (1926) that enzymes could be obtained as crystalline proteins. In the 1930s Northrop and coworkers purified a number of important enzymes by crystallization, most notably from the pancreas of pigs and cows (reviewed in Northrop *et al.*, 1948). A cascade of successes with other enzymes quickly followed, leading to the award of Nobel Prizes to Sumner and Northrop.

Late in the 1930s, crystals of proteins began to assume a new and important role as a consequence of the advent of X-ray crystallography as applied to biological macromolecules. The early work of Bernal, Fankuchen, Crowfoot and Perutz (Dickerson, 2005) made protein crystals important for the three-dimensional structural information that they could potentially yield. The demand for protein crystals expanded rapidly in the 1960s and 1970s as protein crystallography came of age and highly motivated young scientists entered the field.

For 15 years, from about 1965 until 1980, X-ray crystallographers depended very much on the successes of earlier protein chemists, and on their somewhat limited procedures and technologies, to provide

suitable samples for diffraction. Ultimately, however, those sources diminished and the methodologies became inadequate. As a result, the 1970s and 1980s saw a great interest develop in devising new approaches to protein crystallization and in discovering and applying new ways to obtain purified samples of novel and biologically important proteins for crystallization (McPherson, 1982).

This endeavor received its greatest boost from an unexpected source: genetics. With the explosion in genetic engineering and molecular biological research in the 1980s and 1990s came an attendant flood of biologically interesting proteins previously unobtainable because of their low abundance in natural systems. The integration of recombinant DNA technology with X-ray crystallography subsequently produced a revolution in structural biology that has, in turn, totally transformed the field of molecular biology. The two disciplines working in tandem, and in many cases tightly coupled, have spawned the structural genomics enterprise, and ultimately promises to allow the detailed visualization of all biological structures at atomic resolution.

This article contains a brief review of the methods and procedures that have emerged from the last 150 years of protein crystal-growth experience. It contains descriptions of the techniques in common use today. It should, however, not be considered to be entirely comprehensive or exhaustive. In particular, it should in no way set boundaries on the imagination and ingenuity of the reader. There are undoubtedly many contributions yet to be made to this still young, still largely empirical field.

Presently, and in the foreseeable future, the only techniques that can yield atomic level structural images of biological macromolecules are X-ray and neutron diffraction as applied to single crystals. While other methods may produce important structural and dynamic data, for highly precise atomic coordinates only X-ray crystallography is adequate. As its name suggests, application of X-ray crystallography is absolutely dependent on crystals of the macromolecule, and not simply crystals but crystals of sufficient size and quality to permit the collection of accurate diffraction intensities. The quality of the final structural image is directly determined by the quality of diffraction, that is, the size and physical properties of the crystalline specimen; hence, the crystal becomes the linchpin of the entire process and the ultimate determinant of its success (McPherson, 1989).

2. General approach

Macromolecular crystallization, which includes the crystallization of proteins, nucleic acids and larger macromolecular assemblies such as viruses and ribosomes, is based on a rather diverse set of principles, experiences and ideas. There is no comprehensive theory, or even a very good base of fundamental data, to guide our efforts, although they are being accumulated at this time. As a consequence, macromolecular crystal growth is largely empirical in nature, and demands patience, perseverance and intuition.

Complicating the entire process, in addition to our limited understanding of the phenomena involved, is the astonishing complexity and range of the macromolecules before us. Even in the case of rather small proteins, such as cytochrome *c* or myoglobin for example, there are roughly a thousand atoms with thousands of bonds and thousands of degrees of freedom. For viruses or enzyme complexes having molecular weights measured in the millions of daltons, the possibilities for conformation, interaction and mobility are almost uncountable.

Only now are we beginning to develop rational approaches to macromolecular crystallization based on an understanding of the fundamental properties of the systems. We are only now using, in a serious and systematic manner, the classical methods of physical chemistry to determine the characteristics of those mechanisms responsible for the self-organization of large biological molecules into crystal lattices. As an alternative to the precise and reasoned strategies that we commonly apply to scientific problems, we continue to rely, for the time being at least, on what is fundamentally a trial-and-error approach. Macromolecular crystallization is generally a matter of searching, as systematically as possible, the ranges of the individual parameters that influence crystal formation, finding a set, or multiple sets of factors that yield some kind of crystals, and then optimizing the individual variables to obtain the best possible crystals. This is usually achieved by carrying out an extensive series, or establishing a vast matrix, of crystallization trials, evaluating the results and using the information that is obtained to improve conditions in successive rounds of trials. Because the number of variables is so large, and because the ranges are so broad, experience and insight in designing and evaluating the individual and collective trials becomes an important consideration.

3. The nature of protein crystals

X-ray analysis is a singular event confined to the research laboratory and the final product is basic scientific knowledge. The crystals themselves, with some exceptions, have no medicinal or pharmaceutical value, but simply serve as intermediaries in the crystallographic process. The crystals provide the X-ray diffraction patterns that in turn serve as the raw data which allow the direct visualization of the macromolecules or their complexes that the crystals are composed of. Some examples of crystals of proteins and viruses grown in one of the author's laboratories (AM) are shown in Fig. 1.

When crystallizing proteins for X-ray diffraction analysis, one is usually dealing with homogenous, often exceptionally pure macromolecules, and the objective may be to grow only a few large, high-quality, high-performance crystals. It is important to emphasize that while the number of crystals needed may be few, often the amount of protein available may be severely limited. This in turn places grave constraints on the approaches and strategies that may be used to obtain those crystals. While new methodologies such as synchrotron radiation (Helliwell, 1992) and cryocrystallography (Garman & Schneider, 1997) have driven the necessary size and number of specimen crystals consistently downwards, they have not alleviated the need for crystal quality and stability.

Macromolecular crystals are composed of approximately 50% solvent on average, although this may vary from 25 to 90% depending on the particular macromolecule. Protein or nucleic acid occupies the remaining volume so that the entire crystal is in many ways an ordered gel permeated by extensive interstitial spaces through which solvent and other small molecules freely diffuse.

In proportion to its molecular mass, the number of bonds (salt bridges, hydrogen bonds, hydrophobic interactions) that a conventional molecule forms to its neighbors in a crystal far exceeds the very few exhibited by crystalline macromolecules. Since these contacts provide the lattice interactions essential for crystal maintenance, this largely explains the difference in properties between crystals of salts or small molecules and of macromolecules (Chernov, 2003; Vekilov & Chernov, 2002).

Living systems are based almost exclusively on aqueous chemistry within narrow ranges of temperature and pH. Macromolecules have thus evolved an appropriate compatibility, and serious deviations or perturbations are rarely tolerated. As a consequence, all protein and

nucleic acid crystals must be grown from aqueous solutions to which they are tolerant, and these solutions are called mother liquors. Macromolecular crystals have so far only been grown from such media.

Although comparable in their morphologies and appearance, there are important practical differences between crystals of low-molecular-mass compounds and crystals of proteins and nucleic acids. Crystals of conventional molecules are characterized by firm lattice forces, are relatively highly ordered, are generally physically hard and brittle, are easy to manipulate, can usually be exposed to air, have strong optical properties and diffract X-rays intensely. Macromolecular crystals are, by comparison, usually more limited in size, are very soft and crush easily, disintegrate if allowed to dehydrate, exhibit weak optical properties and diffract X-rays poorly. Macromolecular crystals are temperature sensitive and undergo extensive damage after prolonged exposure to radiation. Frequently, several or even many crystals must be analyzed for a structure determination to be successful, although the advent of cryocrystallography (Pflugrath, 1992), CCD area detectors of very high photon-counting efficiency and dynamic range (Gruner *et al.*, 2001), high-intensity synchrotron X-ray sources (Pflugrath, 1992; Helliwell, 1992) and new phasing methods (Rossmann & Arnold, 2001) has greatly lessened this constraint.

The extent of the diffraction pattern from a crystal is directly correlated with its degree of internal order. The more vast the pattern, or the higher the resolution to which it extends, the more structurally uniform are the molecules in the crystal and the more precise is their periodic arrangement. The level of detail to which atomic positions can be determined by crystal structure analysis corresponds closely with this degree of crystalline order. While conventional crystals often diffract to their theoretical limit of resolution, protein crystals, by comparison, produce diffraction patterns of more limited extent.

The liquid channels and solvent-filled cavities that permeate macromolecular crystals are primarily responsible for the limited resolution of the diffraction patterns. Because of the relatively large spaces between adjacent molecules and the consequent weak lattice forces, all molecules in the crystal may not occupy exactly equivalent orientations and positions, but may vary slightly within or between unit cells. Furthermore, because of their structural complexity and their potential for conformational dynamics, protein molecules in a particular crystal may exhibit slight variations in the course of their

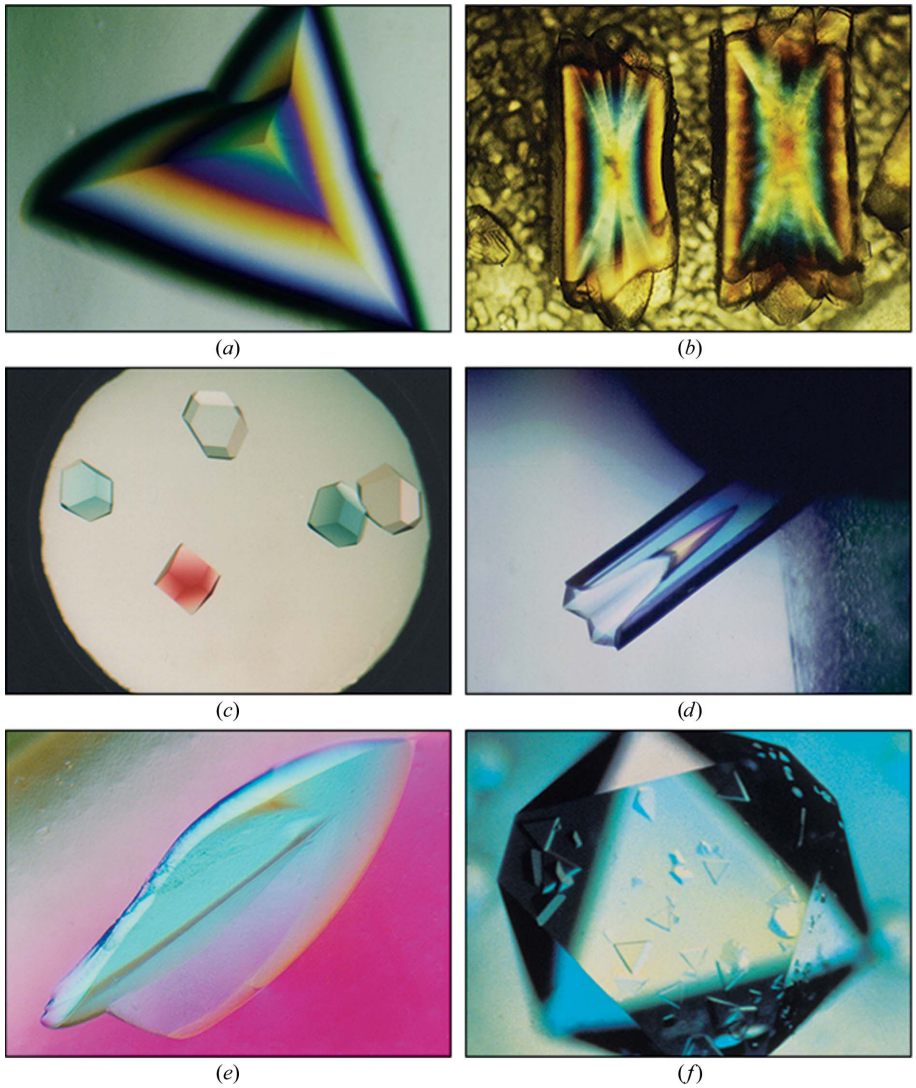


Figure 1
Microphotographs of protein and virus crystals grown in the laboratory of AM showing the variety of habits common to macromolecular crystals. (*a*, *b*, *f*) *Satellite tobacco mosaic virus*, (*c*) *Desmodium yellow mottle virus*, (*d*) hexagonal canavalin and (*e*) intact anti-canine lymphoma antibody.

polypeptide chains or in the dispositions of side groups from one to another.

Although the presence of extensive solvent regions is a major contributor to the generally modest diffraction quality of protein crystals, it is also responsible for their value to biochemists. Because of the high solvent content, the individual macromolecules in protein crystals are surrounded by layers of water that maintain their structure virtually unchanged from that found in solution. As a consequence, ligand binding, enzymatic, spectroscopic characteristics and most other biochemical features are essentially the same as for the fully solvated molecule. Conventional chemical compounds, which may be ions, ligands, substrates, coenzymes, inhibitors, drugs or other effector molecules, may be freely diffused into and out of the crystals. Crystalline enzymes, although immobilized, are completely accessible for experimentation simply through alteration of the surrounding mother liquor.

Polymorphism, as is evident in Fig. 1, is a common phenomenon for protein, nucleic acid and virus crystals. Presumably this is a consequence of their conformational dynamic range and the sensitivity of the lattice contacts involved. Thus, different habits and different unit cells may arise from what, by most standards, would be called identical conditions. In fact, multiple crystal forms are sometimes seen coexisting in the same sample of mother liquor.

There are further differences which complicate the crystallization of macromolecules compared with conventional small molecules (Feigelson, 1988; Feher, 1986; Durbin & Feher, 1996; McPherson, 1982, 1999; McPherson *et al.*, 1995). Firstly, macromolecules may assume several distinctive solid states that include amorphous precipitates, oils or gels as well as crystals, and most of these are kinetically favored. Secondly, macromolecular crystals nucleate, or initiate development, only at very high levels of supersaturation, often two to three orders of magnitude greater than that required to sustain growth. Finally, the kinetics of macromolecular crystal nucleation and growth are generally two to three orders of magnitude slower than for conventional molecules (Kuznetsov *et al.*, 1995; Malkin *et al.*, 1996, 1997). This latter difference arises from their considerably larger size, lowered diffusivity and weaker association tendencies compared with small molecules or ions, as well as the lower probability of incorporation of an incoming macromolecule into a growth step (Chernov, 2003; Vekilov & Chernov, 2002).

4. Screening and optimization of crystallization conditions

There are really two phases in the pursuit of protein crystals for an X-ray diffraction investigation, and these are (i) the identification of

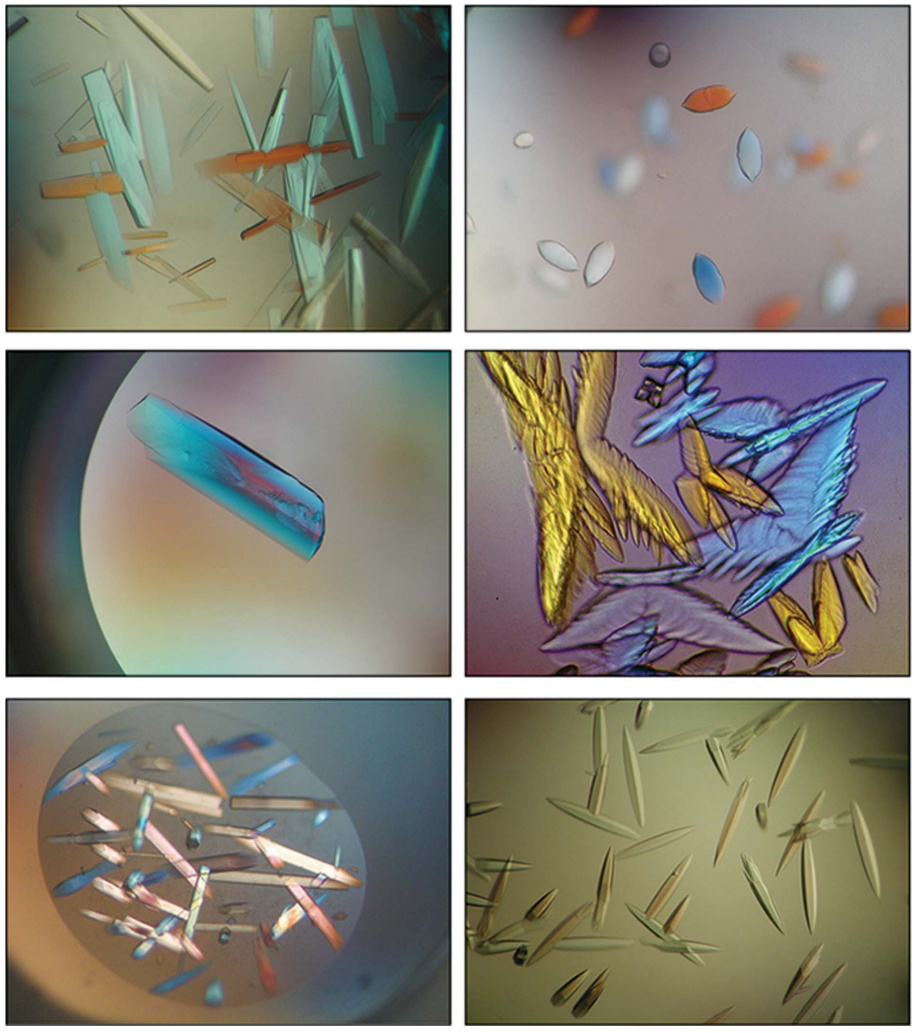


Figure 2
Shown here are a variety of protein crystals that were obtained directly from commercial screening matrices but, as is evident, some are inadequate for X-ray data collection because of morphology or size, implying that the crystallization conditions require optimization.

chemical, biochemical and physical conditions that yield some crystalline material, although that may be entirely inadequate, as illustrated by some of the crystals in Fig. 2, and (ii) the systematic alteration of these initial conditions by incremental amounts to obtain optimal samples for diffraction analysis. The first of these is fraught with the greater risk, as some proteins simply refuse to form crystals and any clues as to why are elusive or absent. The latter, however, often proves to be more demanding, time-consuming and frustrating.

There are basically two approaches to screening for crystallization conditions. The first is a systematic variation of what are believed to be the most important variables: precipitant type and concentration, pH, temperature *etc.* The second is what we might term a shotgun approach, but a shotgun aimed with intelligence, experience and accumulated wisdom. While far more thorough in scope and more congenial to the scientific mind, the first method usually requires a significantly greater amount of protein. In those cases where the quantity of material is limiting, it may simply be impractical. The second technique provides much more opportunity for useful conditions to escape discovery, but in general requires less precious material.

The second approach also has, presently at least, one other major advantage, and that is convenience. Currently, a wide variety of crystallization screening kits from numerous companies are available on the commercial market. The availability and ease of use of these relatively modestly priced kits, which may be used in conjunction with a variety of crystallization methods (hanging- and sitting-drop vapor diffusion, dialysis *etc.*; see below) make them the first tool of choice in attacking a new crystallization problem. With these kits, nothing more is required than combining a series of potential crystallization solutions with one's protein of interest using a micropipette, sealing the samples and waiting for success to smile. Often it does, but sometimes not, and this is when the crystal grower must begin to use his own intelligence to diagnose the problem and devise a remedy.

Once some crystals, even if only microcrystals, are observed and shown to be of protein origin (and one ardently hopes for this event) then optimization begins. Every component in the solution yielding crystals must be noted and considered (buffer, salt, ions *etc.*), along with pH, temperature and whatever other factors (see below) might have an impact on the quality of the results. Each of these parameters or factors is then carefully incremented in additional trial matrices encompassing a range spanning the condition which gave the 'hit'. Because the problem is nonlinear, and one variable may be coupled to another, this process is often more complex and difficult than one might expect (McPherson, 1982, 1999; Bergfors, 1999; Ducruix & Giége, 1992). It is here that the amount of protein and the limits of the investigator's patience may prove to be a formidable constraint.

5. Supersaturation, nucleation and growth of crystals

Crystallization of any molecule, or collection of some chemical species, including proteins, proceeds in two rather distinct but inseparable steps: nucleation and growth. Nucleation is the most difficult problem to address theoretically and experimentally because it represents a first-order phase transition by which molecules pass from a wholly disordered state to an ordered one. Presumably this occurs through the formation of partially ordered or paracrystalline intermediates, in this case protein aggregates having short-range order, and ultimately yields small, completely ordered assemblies which we refer to as critical nuclei.

Critical nuclei must be considered in terms of the molecular dimensions, the supersaturation and the surface free energy of molecular addition. Currently, the critical nuclear size has only been described for a few systems, and for several cases these were only investigated in terms of two-dimensional nuclei developing on the surfaces of already existent crystals (Malkin *et al.*, 1996, 1997). Recently, a theory has emerged which attempts to explain the nucleation phenomenon in terms of statistical fluctuations in solution properties (Ten Wolde & Frenkel, 1997; Haas & Drenth, 1999; Piazza, 1999; Kuznetsov *et al.*, 1998). This idea holds that a distinctive 'liquid protein phase' forms in concentrated protein solutions and that this 'phase' ultimately gives rise to critical nuclei with comprehensive order. This idea is now under study using a variety of experimental techniques in numerous laboratories.

The growth of macromolecular crystals is a better characterized process than nucleation, and its mechanisms are reasonably well understood. Protein crystals grow principally by the classical mechanisms of dislocation growth and growth by two-dimensional nucleation, along with two other less common mechanisms known as normal growth and three-dimensional nucleation (Malkin *et al.*, 1995; McPherson & Malkin, 2000). A common feature of nucleation and growth is that both are critically dependent on what is termed the supersaturation of the mother liquor giving rise to the crystals. Supersaturation is the variable that drives both processes and determines their occurrence and extent and the kinetics that govern them.

Crystallization of a macromolecule absolutely requires the creation of a supersaturated state. This is illustrated by the phase diagram

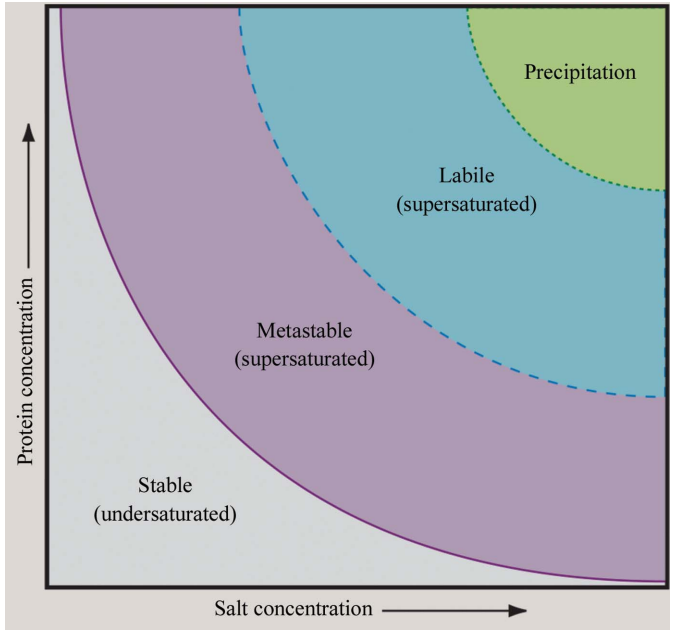


Figure 3
The phase diagram for the crystallization of macromolecules. The solubility diagram is divided sharply into a region of undersaturation and a region of supersaturation by the line denoting maximum solubility at specific concentrations of a precipitant, which may be salt or a polymer. The line represents the equilibrium between the existence of the solid phase and the free-molecule phase. The region of supersaturation is further divided in a more uncertain way into the metastable and labile regions. In the metastable region nuclei will develop into crystals, but no nucleation will occur. In the labile region both might be expected to occur. The final region, at very high supersaturation, is denoted the precipitation region, where this result might be most probable. Crystals can only be grown from a supersaturated solution, and creating such a solution supersaturated in the protein of interest is the immediate objective in growing protein crystals.

for crystal growth presented in Fig. 3. Supersaturation is a non-equilibrium condition in which some quantity of the macromolecule in excess of the solubility limit, under specific chemical and physical conditions, is nonetheless present in solution. Equilibrium is re-established by the formation and development of a solid state, such as crystals, as the saturation limit is attained. To produce the supersaturated solution, the properties of an undersaturated solution must be modified to reduce the ability of the medium to solubilize the macromolecule (*i.e.* reduce its chemical activity), or some property of the macromolecules must be altered to reduce their solubility and/or to increase the attraction of one macromolecule for another. In all cases, the relationships between solvent and solute, or between the macromolecules in solution, are perturbed so as to promote formation of the solid state.

If no crystals or other solid is present as conditions are changed, then solute will not immediately partition into two phases and the solution will remain in the supersaturated state. The solid state does not necessarily develop spontaneously as the saturation limit is exceeded because energy, analogous to the activation energy of a chemical reaction, is required to create the second phase: the stable nucleus of a crystal, or a precipitate. Thus, a kinetic or energy (or probability) barrier allows conditions to proceed further from equilibrium and further into the zone of supersaturation. Once a stable nucleus appears in a supersaturated solution, however, it will proceed to grow until the system regains equilibrium. As long as non-equilibrium forces prevail and some degree of supersaturation exists to drive events, a crystal will grow or a precipitate will form.

6. Promoting supersaturation

In practice, one begins (with the exception of the batch method; see below) with a solution, a potential mother liquor, which contains some concentration of the protein below its solubility limit or alternatively at its solubility maximum. The objective is then to alter matters so that the solubility of the protein in the sample is significantly reduced, thereby rendering the solution supersaturated. This may be performed through several approaches: (i) by altering the protein itself (*e.g.* by a change of pH, which alters the ionization state of surface amino-acid residues), (ii) by altering the chemical activity of the water (*e.g.* by the addition of salt), (iii) by altering the degree of attraction of one protein molecule for another (*e.g.* change of pH or the addition of bridging ions) or (iv) by altering the nature of the interactions between the protein molecules and the solvent (*e.g.* the addition of polymers or ions).

Table 1
Methods for creating supersaturation.

1. Direct mixing of protein and precipitant solutions to immediately create a supersaturated condition (batch method)
2. Alter the temperature
3. Add salt (increase ionic strength), salting out
4. Remove salt (decrease ionic strength), salting in
5. Alter pH through liquid or vapor phase
6. Add a ligand that changes the solubility of the macromolecule
7. Alteration of the dielectric of the medium (by addition of organic solvents)
8. Evaporation
9. Addition of a polymer that produces volume exclusion
10. Addition of a cross-bridging agent that promotes lattice interactions
11. Concentration of the macromolecule by removal of water through a membrane
12. Removal of a solubilizing agent (chaotrope)

Table 1 is a compilation of the methods upon which one might develop strategies for crystallizing a protein for the first time. Indeed there may be others; the limit is only a function of the imagination and cunning of the investigator. The details of these various approaches have been described numerous times elsewhere (McPherson, 1982, 1999; Ducruix & Giége, 1992; Bergfors, 1999; McPherson *et al.*, 2003) and need receive no more attention here. It is probably sufficient to say that if a protein has any propensity to crystallize readily, it can probably be accomplished by variation of precipitant type, precipitant concentration, pH and, to a lesser extent,

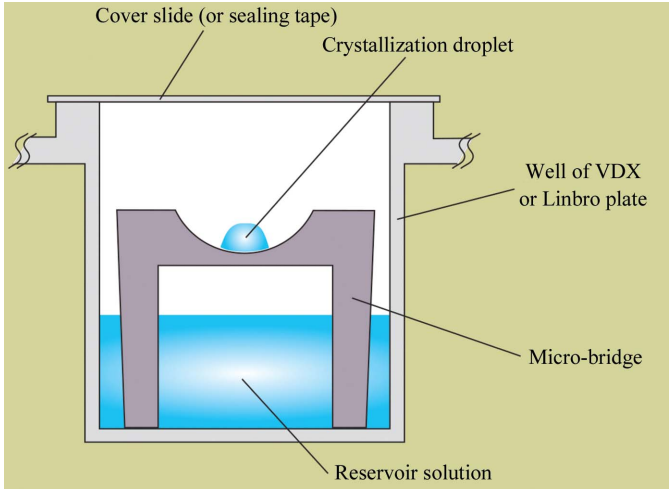


Figure 4
The sitting-drop vapor-diffusion method is illustrated in this schematic diagram. The drop on the elevated platform, which is commonly 2–10 µl, consists of half stock protein solution and half reservoir solution which contains some concentration of a salt or polymer precipitant. About 0.5 ml of the reservoir solution is added to the bottom of the cell before sealing. By water equilibration through the vapor phase the drop ultimately approaches the reservoir in osmolarity, both raising the concentration of the precipitant in the drop and increasing the protein concentration there.

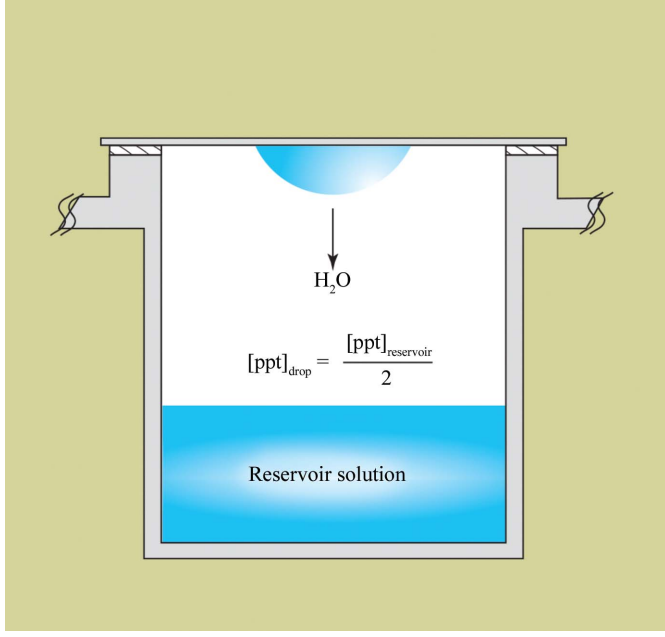


Figure 5
The hanging-drop vapor-diffusion method is illustrated schematically. The components of the drop and reservoir, and the physical equilibration process, are the same here as for the sitting drop. The exception is that the protein drop is suspended from a cover slip over the reservoir rather than resting on a surface. Plasticware for carrying out both sitting- and hanging-drop vapor diffusion are widely and commercially available in numerous formats.

Table 2
Methods for promoting a solubility minimum.

1. Bulk crystallization
2. Batch method in vials
3. Microbatch under oil
4. Controlled evaporation
5. Bulk dialysis
6. Concentration dialysis
7. Microdialysis
8. Free-interface diffusion†
9. Counter-diffusion in capillaries†
10. Liquid bridge†
11. Vapor diffusion on plates (sitting drop)
12. Vapor diffusion in hanging drops
13. Sequential extraction
14. pH-induced crystallization
15. Temperature-induced crystallization
16. Crystallization by effector addition

† 8, 9 and 10 are variations on liquid–liquid diffusion.

temperature, but with all due consideration to the biochemical properties and eccentricities of the protein under investigation. Finally, we are all advised that with real estate there are three important factors, and they are location, location and location. With protein crystallization there are similarly three, and they are purity, purity and homogeneity.

7. Techniques for achieving supersaturation

The growth of protein crystals must be carried out in some physical apparatus that allows the investigator to alter the solubility of the protein, the properties of the mother liquor, using one of the strategies in Table 1. Currently, these almost exclusively use microtechniques. Thus, crystallization 'trials' with a particular matrix of conditions may be carried out with volumes of only a few microlitres or less. Increasingly, these employ plastic multichambered trays for sitting drops (Fig. 4) and hanging drops (Fig. 5), and plexiglass buttons for dialysis or microdrops under oil (Chayen, 1997). Other approaches can be found in Table 2.

Again, all of these devices and their methodologies have been described in detail elsewhere (and are also elaborated upon in other articles in this series). It is unnecessary to comment on each of them again. In addition, detailed instructions are frequently provided by the manufacturers of the crystallization kits, supplies and plasticware, along with much helpful advice. Suffice it to say that currently the hanging-drop and sitting-drop procedures for vapor diffusion, and the batch method using microdrops under oil are most in favor, and are recommended for most investigations. In those cases where the mother-liquor components cannot be transported through the vapor phase (*e.g.* metal ions and detergents) then microdialysis (Fig. 6) may be the only recourse. An important point, however, is that the best method for screening conditions and obtaining an initial set of crystallization parameters may not be the best means for optimization. Thus, one may start with one technique but ultimately find that another gives larger crystals of higher quality.

An interesting, and relatively new, technique for crystallization has been developed (García-Ruiz, 2003; Ng *et al.*, 2003) that is essentially an extension of the old liquid–liquid free-interface diffusion method (Fig. 7), and this has been termed 'counter-diffusion' (Otálora *et al.*, 2009; Gavira *et al.*, 2002). With this technique (Fig. 8) a gel, into which one end of a capillary containing the protein solution is pressed, is then impregnated with a precipitating solution. With time, the precipitant diffuses up the capillary so that a gradient is ultimately established. Thus, the protein is exposed to a continuum of precipitant concentration. Because of the interplay of precipitant diffusion and crystallization, the dynamics of the process in the capillary is more complex than might be thought, but this only enhances the probability of nucleation and more ordered growth. When successful, microcrystals may be observed where the precipitant concentration is highest, near the surface of the gel, and large crystals near the distal end of the capillary. The method has now been used to obtain crystals for X-ray diffraction for many proteins at both room and cryogenic temperatures (García-Ruiz, 2003; Ng *et al.*, 2003), and has been employed to grow crystals in a number of experiments carried out in microgravity.

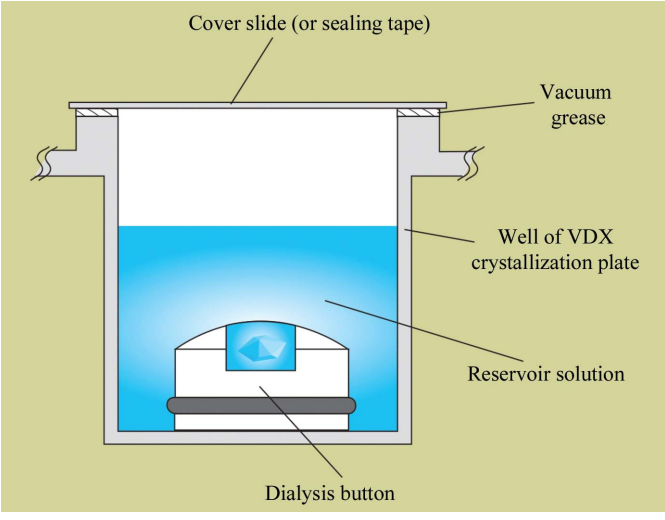


Figure 6
The use of microdialysis buttons to dialyze small volumes of protein solution against a precipitating solution is illustrated. The protein solution volumes may be from 10 to 50 µl. The buttons are commercially available.

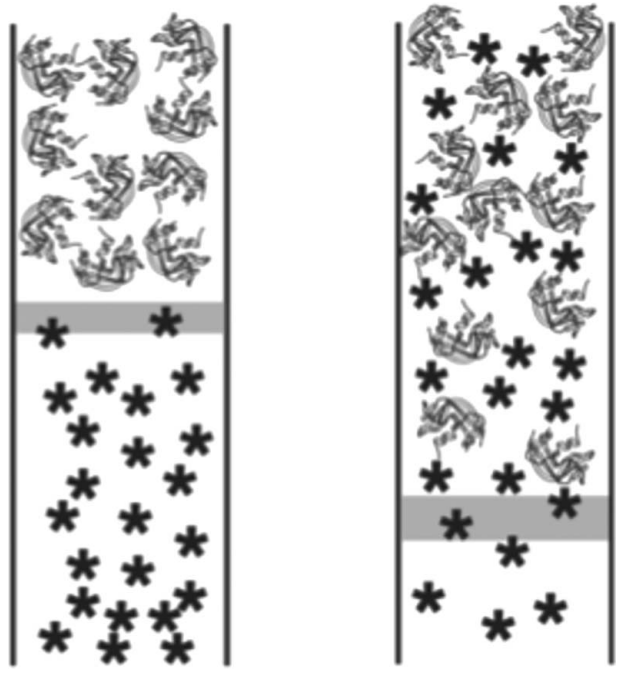


Figure 7
The process of free-interface diffusion to effect crystallization is illustrated. A protein solution is layered atop a precipitant solution in a narrow-bore tube or capillary. Diffusion across the interface, principally of the precipitant, induces nucleation and growth.

Table 3
Precipitants used in macromolecular crystallization.

Salts	Volatile organic solvents	Polymers	Nonvolatile organic solvents
Ammonium sulfate	Ethanol	Polyethylene glycol 1000, 3350, 6000, 8000, 20000	2-Methyl-2,4-pentanediol (MPD)
Ammonium phosphate	Propanol	Jeffamine T, Jeffamine M	2,5-Hexanediol
Lithium sulfate	Isopropanol	Polyethylene glycol monomethyl ester	1,3-Propanediol
Lithium chloride	Dioxane	Polyethylene glycol monostearate	Polyethylene glycol 400
Sodium citrate	Acetone	Polyeneamine	Jeffamine 400
Ammonium citrate	Isobutanol		
Sodium phosphate	*n*-Butanol		
Sodium chloride	*tert*-Butanol		
Potassium chloride	Acetonitrile		
Sodium acetate	Dimethyl sulfoxide		
Ammonium acetate	1,3-Butyrolactone		
Magnesium sulfate			
Magnesium chloride			
Calcium chloride			
Sodium formate			
Sodium tartrate			
Cadmium sulfate			
Sodium succinate			
Sodium malonate			

As described briefly below, and more thoroughly in future articles, screening for crystallization conditions, and even optimization in some cases, has been consigned in high-throughput laboratories to robotic devices. This is particularly true in those of large pharmaceutical companies where many proteins may be under simultaneous investigation. Robotic systems have the advantage of exceptional sample-record maintenance, most can deploy submicrolitre amounts of mother liquor and they can be used to screen vast matrices of conditions that might otherwise be impossible in a practical sense for a lone investigator. Robotic systems are, in addition, now being used to examine and evaluate the results of crystallization trials using optical subsystems and image-processing techniques (Hosfield *et al.*, 2003; DeLucas *et al.*, 2003; Luft *et al.*, 2003). Evaluation of trial arrays of conditions, however, continues to be problematic because of the continuing difficulty in devising meaningful scoring criteria in the absence of actual crystals. That is, the sole presence of various kinds of precipitates or other phases in an individual crystallization trial gives only very ambiguous indications of how near the conditions were to being a successful mother liquor.

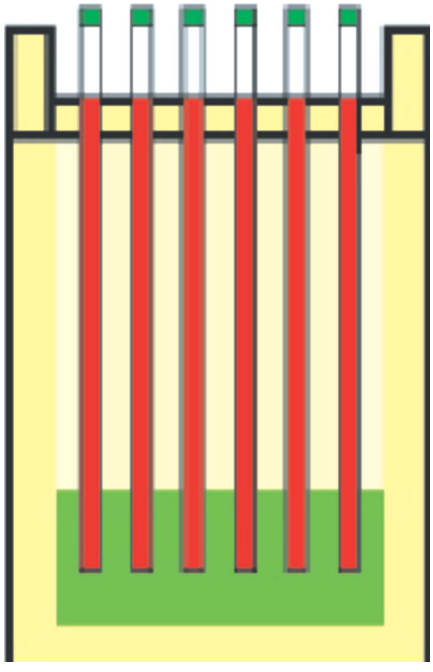


Figure 8
Diagram illustrating the counter-diffusion method for growing protein crystals. Here, the protein solution is shown in red and the gel saturated with the precipitant solution is shown in green. The capillaries are sealed at their distal end but are open where they enter the gel. By diffusion of precipitant up the length of the capillary, a concentration gradient is formed that explores a wide range of precipitant conditions.

8. Crystallization agents and precipitants

If one were to examine the reagents utilized in any of the commercial crystallization screens which are based on shotgun approaches, or to examine the crystallization databases which have been compiled (see below), then it would become immediately apparent that a very wide range of precipitating (crystallizing) agents are used. Indeed, many agents have been employed, and some, such as ammonium sulfate or polyethylene glycol (PEG), for a great number of successes. It is often necessary, however, to explore many precipitants, and it is difficult to

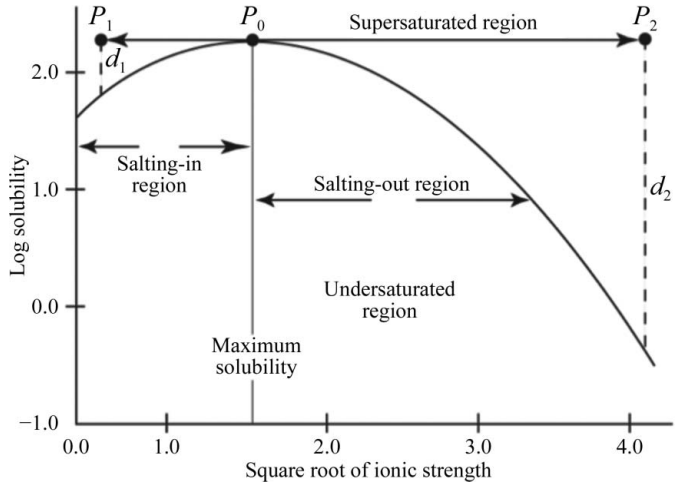


Figure 9
The curve shown here represents a typical solubility curve for a protein and divides, as in Fig. 2, the region of undersaturation from that of supersaturation. It also illustrates the existence of the classical 'salting-in' and 'salting-out' regions for the protein. By taking advantage of these latter effects, supersaturation may be achieved by equilibrating a system from a point of maximum solubility (P_0) to one of reduced solubility (P_1 or P_2) by adjusting the precipitant concentration.

know in advance which might offer the greatest likelihood of obtaining crystals.

Individual precipitants and their properties have also been reviewed (McPherson, 1999) and will not be extensively discussed here. To simplify, however, it is possible to group the precipitants into categories based on their mechanisms for promoting crystallization, and this is shown in Table 3. Precipitants of macromolecules fall into four broad categories: (i) salts, (ii) organic solvents, (iii) long-chain polymers and (iv) low-molecular-weight polymers and nonvolatile organic compounds. The first two classes are typified by ammonium sulfate and ethyl alcohol, respectively, and higher polymers such as polyethylene glycol 4000 are characteristic of the third catergory. In the fourth category we might place compounds such as methylpentanediol and polyethylene glycols of molecular weight lower than about 1000.

The solubility of macromolecules in concentrated salt solutions is a complicated phenomonon, but it can be viewed naively as a competition between salt ions, principally the anions, and the macromolecules for the binding of water molecules, which are essential for the maintenance of solubility (Hofmeister, 1890; Herriott, 1942; Cohn & Ferry, 1943; Cohn & Edsall, 1943). At sufficiently high salt concentrations the macromolecules become so uncomfortably deprived of solvent that they seek association with one another in order to satisfy their electrostatic requirements. In this environment ordered crystals, as well as disordered amorphous precipitate, may form. Some salt ions, chiefly cations, are also necessary to ensure macromolecular solubility. At very low ionic strengths, cation availability is insufficient to maintain macromolecule solubility, and under these conditions crystals may also form. The behavior of typical proteins over the entire range of salt concentrations, including both the 'salting-in' and 'salting-out' regions, is illustrated in Fig. 9.

As described above, salts exert their effect principally by dehydrating proteins through competition for water molecules. A measure of their efficiency in this is the ionic strength, whose value is the sum of the products, one for each ion in solution, of the molarity of that ion with the square of its charge. Thus, multivalent ions, particularly anions, are the most efficient precipitants. Sulfates, phosphates and citrates have, for example, traditionally been employed.

One might anticipate little variation among different salts as long as the valences of their ions are the same. Thus, there should be little expected variation between two different sulfates such as lithium sulfate and ammonium sulfate if only ionic strength were involved. However, this is often observed not to be the case. In addition to salting out, which is a general dehydration effect, or reduction of the chemical activity of water, there are also specific protein–ion interactions that may have other consequences. This is perhaps not unexpected given the unique polyvalent character of individual proteins, their structural complexity and the intimate dependence of their physical properties on their surroundings. It is inadequate, therefore, when attempting to crystallize a protein to examine only one or two salts and ignore the broader range. Alternative salts can sometimes produce crystals of varied quality, morphology and, in some cases, diffraction properties.

It is usually not possible to predict the degree of saturation or molarity of a precipitating agent required for the crystallization of a particular protein or nucleic acid without some prior knowledge of its behavior. In general, however, it is a concentration just a few percent less than that which yields an amorphous precipitate (Sumner & Somers, 1943), and this can be determined for a macromolecule under a given set of conditions using only minute amounts of material (McPherson, 1982). To determine the approximate insolubility points with a particular precipitant, a 10 µl droplet of a 5–15 mg ml^{-1} protein solution can be placed in the well of a depression slide and

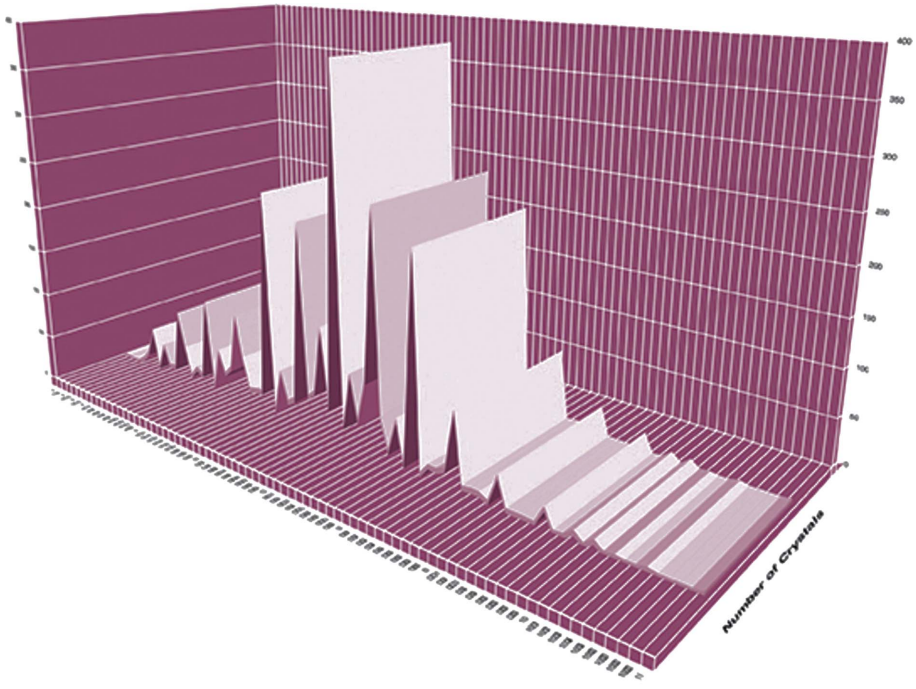


Figure 10
This diagram, based on an analysis of nearly 2800 examples, shows the distribution of the number of protein crystals grown as a function of pH. As one might expect, the great majority have been grown near neutrality, reflecting the desire of investigators to crystallize their protein near physiological conditions. The spread, however, illustrates that protein crystals might reasonably be expected over a very large pH range and that this entire range deserves attention (figure courtesy of Hampton Research, Aliso Viejo, CA, USA).

observed under a low-power light microscope as increasing amounts of saturated salt solution or organic solvent (in 1 or 2 µl increments) are added. If the well is sealed between additions with a cover slip, the increases can be made over a period of many hours.

Along with ionic strength, pH is one of the most important variables influencing the solubility of proteins, and indeed, as illustrated by the diagram in Fig. 10, protein crystals have been obtained over the entire range of pH. This variable provides another powerful approach to creating supersaturated solutions, and hence effecting crystallization. Its manipulation at various ionic strengths and in the presence of diverse precipitants is a fundamental idea in formulating screening matrices and discovering successful crystallization conditions. An example of the effect of pH on a typical protein is illustrated in Fig. 11.

Organic solvents reduce the dielectric of the medium, hence screening the electric fields that mediate macromolecular interactions in solution. As the concentration of organic solvent is increased, attraction between macromolecules increases, the solvent becomes less effective (the activity coefficient of water is reduced) and the solid state is favored (Cohn *et al.*, 1974; Englard & Seifter, 1990). Organic solvents should be used at a low temperature, at or below 0°C, and they should be added very slowly with good mixing (McPherson, 1999). Since they are usually volatile, vapor-diffusion techniques are equally applicable for either bulk or micro amounts. Ionic strength should, in general, be maintained low and whatever means are otherwise available should be pursued to protect against denaturation.

Some polymers, with polyethylene glycols being the most popular (McPherson, 1976*a*; Patel *et al.*, 1995), produce volume-exclusion effects that also induce separation of macromolecules from solution (Ingham, 1990; McPherson, 1976*a*). The polymeric precipitants, which unlike proteins have no consistent conformation, writhe and twist randomly in solution and occupy far more space than they otherwise deserve. This results in less solvent available space for the other macromolecules, which then segregate, aggregate and ultimately form a solid state, often crystals.

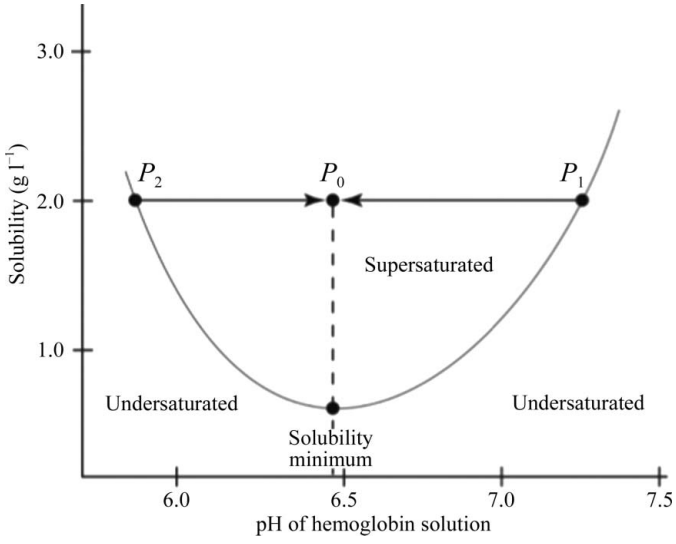


Figure 11
As shown here, most proteins have certain solubility minima as a function of pH. One can take advantage of this property to produce supersaturation by altering a system from a pH permitting high solubility (P_1 or P_2) to a point of low solubility (P_0). This is a powerful approach to promoting crystallization of macromolecules.

Many protein structures have now been solved using crystals grown using polyethylene glycol. These confirm that the protein molecules are in as native a condition in this medium as in any other. This is reasonable because the larger molecular weight polyethylene glycols probably do not even enter the crystals and therefore do not directly contact the interior molecules. In addition, it appears that crystals of many proteins, when grown from polyethylene glycol, are essentially isomorphous with, and exhibit the same unit-cell symmetry and dimensions as, those grown by other means.

PEGs with molecular weights from 400 to 20 000 have successfully provided protein crystals, but the most useful are those in the range 2000–8000. A number of cases have appeared, however, in which a protein could not easily be crystallized using this range but yielded crystals in the presence of PEG 400 or 20 000. The molecular weights are generally not completely interchangeable for a given protein, even within the mid-range. Some produce the best-formed and largest crystals only at, say, a molecular weight of 4000 and lesser quality examples at other weights. This is a parameter which is best optimized by empirical means along with concentration and temperature. The very low molecular weight PEGs such as PEG 200 and 400 are rather similar in character to MPD and hexanediol. There does not appear to be any correlation between the molecular weight of a protein and that of the PEG best used for its crystallization. The higher molecular weight PEGs do, however, have a proportionally greater capacity to force proteins from solution.

A distinct advantage of polyethylene glycol over other precipitating agents is that most proteins crystallize within a fairly narrow range of PEG concentrations: from about 4 to 18% (although there are numerous examples where either higher or lower concentrations were necessary). In addition, the exact PEG concentration at which crystals form is rather insensitive. If one is within a few percent of the optimal value, some success is likely to be achieved. With most crystallizations from high ionic strength solutions or from organic solvents, one must be within 1 or 2% of an optimum lying anywhere between 15 and 85% saturation. The great advantage of PEG is that when conducting a series of initial trials to determine what conditions will give crystals, one can use a fairly coarse selection of concentrations and over a rather narrow total range.

Since PEG solutions are not volatile, PEG must be used like salt or MPD and equilibrated with the protein by dialysis, slow mixing or vapor equilibration. When the reservoir concentration is in the range 5–12%, the protein solution to be equilibrated should be at an initial concentration of about half, conveniently obtained by mixing equal volumes of the reservoir and protein solution. When the final PEG concentration to be attained is much higher than 12%, it is probably advisable to initiate the mother liquor at no more than 4–5% below the final value.

9. How does the novice choose with what screening kits to begin?

The first questions posed by a molecular biologist or biochemist when he begins to think of himself as a potential crystallographer are as follows. (i) Which crystallization kit do I start with? (ii) What do I do if I don't get any crystals? (iii) What happens if I do get crystals? The answer to the first question is in fact the simplest. Start with a screen that does not unreasonably tax your supply of protein but which explores the widest volume of crystallization space, *i.e.* samples the largest number of precipitants and precipitant concentrations over the largest range of pH. There are many kits on the market using 96 sample trials that accomplish this well. If protein is severely limited,

Table 4
Factors affecting crystallization.

Physical	Chemical	Biochemical
1. Temperature/temperature variation	1. pH	1. Purity of the macromolecule/nature of impurities
2. Surfaces/heterogeneous nucleants	2. Precipitant type	2. Ligands, inhibitors, effectors
3. Methodology/approach to equilibrium	3. Final precipitant concentration	3. Aggregation state of the macromolecule
4. Mother-liquor volume	4. Ionic strength	4. Post-translational modifications
5. Geometry of chamber or capillary	5. Cation type and concentration	5. Source of macromolecule
6. Gravity	6. Anion type and concentration	6. Proteolysis/hydrolysis
7. Pressure	7. Degree of supersaturation	7. Chemical modifications
8. Time	8. Reductive/oxidative environment	8. Genetic modifications
9. Vibrations/sound/mechanical perturbations	9. Concentration of the macromolecule	9. Inherent symmetry of the macromolecule
10. Electrostatic/magnetic fields	10. Metal ions	10. Degree of denaturation
11. Dielectric properties of the medium	11. Initial precipitant concentration	11. Isoelectric point
12. Viscosity of the medium	12. Cross-linkers/polyions	12. Unstructured regions
13. Rate of equilibration	13. Detergents/surfactants/amphophiles	13. His tags, purification tags
	14. Non-macromolecular impurities	14. α-Helix content
	15. Chaotropes	15. Conformational states
		16. Thermal stability
		17. Allowable pH range
		18. History of the sample

then there are kits employing 48 trial conditions that do the job adequately. If you experience no success, examine the resulting pattern of precipitates, phase separations, clear drops and assorted odd accumulations, and try to divine what might be the best option to try next. Ask how other proteins similar to your own have been crystallized. Try complexes of your protein with its physiological ligands and effectors. Pursue the abundant advice of 'experts' in the field.

If, on the other hand, you are fortunate and the first kit does indeed yield crystals, but perhaps of insufficient size or quality, or of troubling morphology, then you must optimize the crystals, *i.e.* obtain better ones. Optimization means varying the chemical and physical parameters around those of the reagent mixture that yielded your crystals and searching crystallization parameter space by small increments away from your starting point. There is still a component of art and mystery in science, and this is one instance where its appreciation is paramount. Optimization of conditions will be reviewed in more detail in a later article in this series.

10. Factors affecting crystallization and the inclusion of additives

There are many factors that affect the crystallization of macromolecules (McPherson, 1982, 1999) and many of these are summarized in Table 4. These may affect the probability of its occurring at all, the nucleation probability and rate, the crystal-growth rate, and/or the ultimate sizes and quality of the products. As noted above, pH and salt, or the concentrations of other precipitants, are of great importance. The concentration of the macromolecule, which may vary from as low as 2 mg ml^{-1} to as high as 100 mg ml^{-1}, is an additional significant variable.

Other parameters may be less important in general, but can play crucial roles in specific cases. The presence or absence of ligands or inhibitors, the variety of salts or buffers, the equilibration technique used, the temperature and the presence of detergents are all pertinent considerations. Parameters of somewhat lesser significance include gravity, electric and magnetic fields, and viscosity. It can, in general, not be predicted which of these many variables may be of importance for a particular macromolecule, and the suspected influence of any one must, in general, be evaluated by empirical trials.

The most intriguing problem, or opportunity depending on one's perspective, is what additional components or compounds should comprise the mother liquor in addition to solvent, protein and precipitating agent. The most probable effectors are those which maintain the protein in a single, homogeneous and invarient state. Certain chemical compounds or small molecules may have dramatic effects on the success with which individual proteins crystallize. Additives, as they are often called (McPherson, 1976*b*, 1982, 1999; McPherson & Cudney, 2006), can be decisive in macromolecular crystallization. The most commonly used type of additives, and the only class for which we have any rational basis, are those which may, for physiological reasons, be bound by the protein with consequent favorable changes in its physical–chemical properties or conformation.

Numerous cases have, however, been reported in which small, and sometimes large, molecules were observed to make crucial interactions between macromolecules in the crystal that either helped to guide or secure formation of the lattice (Larson *et al.*, 2008; McPherson & Cudney, 2006). Such small molecules sometimes had a physiological basis for their unexpected presence, but frequently did not. They simply provided essential or at least helpful cross-links within the crystal. Additives that are used in protein crystallization or that might be appropriate for use in crystallization may be classified into eight categories.

(i) Physiologically or biochemically relevant small molecules such as coenzymes, substrate analogues, inhibitors, metal ions, prosthetic groups *etc.* These bind at the active sites of enzymes, or at specific sites elsewhere in protein molecules, and may promote more stable, homogeneous conformations, or they may induce conformational changes into alternate states. In any case, the ultimate protein–ligand complex may exhibit a more monodisperse, less dynamic character. The pertinent molecules here are specific to the individual protein under study, and their selection for inclusion in mother liquors is amenable to rational analysis informed by the enzymology and biochemistry of the protein under study. That is, one considers all of the possible ligands of the protein and includes them in the screen of potential crystallization conditions.

(ii) Chemical protectants. These include reductants such as BME and DTT, heavy-metal ion scavengers such as EDTA and EGTA, and compounds intended to prevent microbial infection such as sodium azide, phenol or chlorobutanol. These too are generally included for well understood reasons; their effects are predictable, and their

impact on the crystallization process is usually, but not always, of marginal significance.

(iii) Solubilizing agents and detergents. These include quaternary ammonium salts (Mirzabekov *et al.*, 1972), sulfobetains (Goldberg *et al.*, 1996), chaotropes such as urea (Bolen, 2004) and a range of surfactant and detergent molecules (Neugebauer, 1990; Wiener, 2004; Zulauf, 1990). Because of the interest in membrane proteins, this class of additives has received extensive study and has been broadly applied to many proteins. Remarkably, there is still no consensus on which are most useful, and which should be included in screening conditions.

(iv) Poisons, as they have traditionally been called (McPherson, 1982, 1999), were originally employed to reduce twinning. These are generally low concentrations, 1–5%(*w*/*v*), of common organic solvents. They include compounds such as ethanol, DMSO, acetone, dioxane, butanol or MPD. Their role in the crystallization process, even after 50 years of use, remains obscure. They are likely to enhance the solubility of the proteins and slightly reduce the degree of supersaturation in the mother liquor, as well as lower the dielectric constant of the medium, but they may have other effects as well.

(v) Osmolytes, co-solvents and cosmotropes are compounds that exert their effects at relatively high concentrations, 1 *M* or more, and include a wide range of molecules such as sucrose, trehalose and other sugars, proline, TMAO, glycine, betaine, taurine, sarcosine and a host of others (Bolen, 2004; Collins, 2004; Collins & Washabaugh, 1985). The effect of their inclusion in the mother liquor is to stabilize (or destabilize) the native conformation of the protein by altering the interaction of the surface of the protein with water, or by altering the hydration layer and possibly the structured waters.

(vi) It has been proposed that the conformations of proteins might be stabilized, and their dynamic character reduced, by providing the proteins with small molecules that could reversibly cross-link charged groups (carboxyl and amino groups) on the surface of the protein, or form intramolecular hydrogen-bonding networks using surface polar groups (Maclean *et al.*, 2002). The molecules that have been explored are usually multivalent molecules such as diamino-containing or dicarboxylic acid-containing molecules, or aliphatic moieties of various lengths carrying some combination of charged groups. It is not known whether the stabilization of proteins by this means is significant enough to affect their crystallization or not. This potential mechanism of altering crystallization behavior, however, may indeed be pertinent.

(vii) Classes of compounds useful for stabilizing proteins through noncovalent intramolecular bonds, as described above, may also help to create and stabilize protein crystals by interposing themselves between protein molecules and forming intermolecular cross-links (McPherson, 1999; McPherson & Cudney, 2006). These cross-bridges may involve purely electrostatic interactions or they may rely on hydrogen-bonding arrangements as well. The compounds that are most favorable for forming such 'lattice interactions' are, again, likely to be multivalent charged compounds, but we might expect that their length, or 'reach', would need to be greater, since they would have to extend from one protein molecule to another.

(viii) The final class of additives are those materials or compounds that somehow serve to enhance nucleation, including unique surfaces. These may include low concentrations of PEG (Ray & Puvathingal, 1986), or other polymeric substances such as Jeffamine, emulsified in solutions of high salt concentration (Kuznetsov *et al.*, 2000, 2001). The micro droplets of the polymeric phase serve to concentrate the protein locally and provide an interface for nucleation to occur. This category should probably also include things like the gel used in cubic lipidic phase crystallization (Caffrey, 2003; Nollert, 2004) and surfaces which promote epitaxy and heterogeneous nucleation (Chayen *et al.*, 2006; McPherson & Shlichta, 1988).

11. Seeding

Often it is desirable to reproduce previously grown crystals of a protein where either the formation of nuclei is limiting or spontaneous nucleation occurs at such a profound level of supersaturation that poor growth patterns result. In such cases it is desirable to induce growth in a directed fashion at low levels of supersaturation. This can sometimes be accomplished by seeding a metastable, supersaturated protein solution with crystals from earlier trials. The seeding techniques (Bergfors, 2003; Stura & Wilson, 1991) fall into two categories: those employing microcrystals as seeds and those using larger macro seeds. In both methods, the fresh solution to be seeded should be only slightly supersaturated so that controlled, slow growth will occur. The two approaches have been described elsewhere in some detail (Fitzgerald & Madsen, 1987; Thaller *et al.*, 1985). Seeding, including the recent development of matrix seeding (D'Arcy *et al.*, 2007), and its attendant complications will be detailed in a future article in this series.

In the method of seeding with microcrystals, the danger is that too many nuclei will be introduced into the fresh supersaturated solution and masses of crystals will result, none of which are suitable for diffraction analysis. To overcome this, a stock solution of microcrystals is serially diluted over a very broad range. Some dilution sample in the series will, on average, have no more than one microseed per microlitre. Others will have severalfold more, or none at all. 1 µl of each sample in the series is then added to fresh protein-crystallization trials under what are perceived to be optimal conditions for growth to occur. This empirical test should, ideally, identify the correct sample to use for seeding by yielding only one or a small number of single crystals when growth is completed. Solutions containing too many seeds will yield microcrystals, and solutions containing too low a concentration of seeds will produce nothing at all. The optimal seeding concentration as determined by the test can then be used to seed many additional samples.

The second approach to seeding involves crystals large enough to be manipulated and transferred under a microscope. Again the most important consideration is to eliminate spurious nucleation by the transfer of too many seeds. Even if a single large crystal is employed, microcrystals adhering to its surface may be carried across to the fresh solution. To avoid this, it is recommended that the macro-seed be thoroughly washed by passing it through a series of intermediate transfer solutions. In so doing, not only are microcrystals removed but, if the wash solutions are chosen properly, some limited dissolution of the seed may take place. This has the effect of freshening the seed crystal surfaces and promoting new growth once it is introduced into the protein solution. Again, the new solution must be supersaturated with respect to protein, but not extremely so, in order to ensure slow and ordered growth.

Seeding is frequently a useful technique for initiating the growth of crystals or inducing nucleation and growth at a lower level of supersaturation than might otherwise spontaneously occur. This can only be performed, however, where crystals, even poor crystals, of the protein under investigation have previously been obtained and can be manipulated to serve as seeds (Gavira *et al.*, 2011). A common problem in macromolecular crystallization is inducing crystals to grow that have never previously been observed. This reflects, of course, the salient fact that the formation of stable nuclei of protein crystals is most often the single major obstacle to obtaining any

crystals at all. In those cases where the immediate problem is simply growing crystals, any crystals, then attention must be focused on the nucleation problem, and any approach that might help to promote nucleation should be considered.

One such technique, which is borrowed in part from classical small-molecule crystal-growth methodology, is the use of heterogeneous or epitaxial nucleants. An unexpected example is presented in Fig. 12. In principle, this means the induction of the growth of crystals of one substance on crystal faces of another. The classical example is gallium arsenide crystals that nucleate and grow from the faces of crystals of silicon. Because protein molecules possess chemical groups, both charged and neutral, that often readily interact with small molecules, membranes or other surfaces, the possibility presents itself that the faces of natural and synthetic minerals might help to order protein molecules at their surfaces and thereby induce the formation of ordered two-dimensional arrays of macromolecules. This ordering might occur by mechanical means owing to steps and dislocations on the crystal faces, or by chemical means derived from complementarity between groups on the mineral and the protein. Such cooperation between mineral faces and nascent protein crystals would be particularly favored when the lattice dimensions of the protein unit cell are integral multiples of the natural spacings in the mineral crystal. McPherson & Shlichta (1988) showed that both heterogeneous nucleation and epitaxial growth of protein crystals from mineral faces do indeed occur. Heterogeneous nucleation has also been observed on other types of surfaces including fibers (Fig. 12), animal hairs, epoxide coatings, and nanoscale etched surfaces of graphite and silicon (McPherson, 1999).

A second approach to enhancing the formation of crystal nuclei has also been described (Ray & Bracker, 1986) that does not use solid surfaces. Here, microdroplets of various concentrations of polyethylene glycol were introduced into protein solutions that were also sufficiently high in salt concentration to support crystal growth once stable nuclei were formed. It was shown that protein left the salt-dominated phase of the mixture and concentrated itself in the polyethylene glycol-rich microdroplets, sometimes reaching effective concentrations in these droplets of several hundred milligrams per millilitre. Using light microscopy techniques, it was shown that crystal nuclei appeared first on the surface of the droplets and then proceeded to grow into the supersaturated salt solution that surrounded them, finally reaching a terminal size appropriate for X-ray analysis.

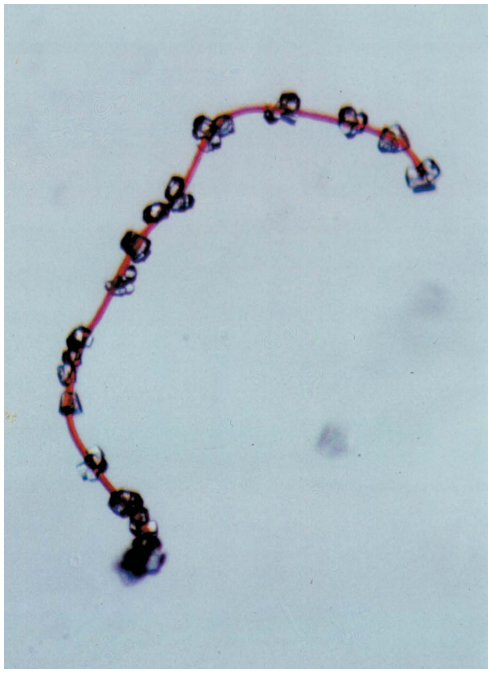


Figure 12
An observation of inadvertent heterogeneous nucleation of protein crystals that is not uncommon is that of crystals growing along the length of a cotton fiber present in the mother liquor. These are crystals of a Fab fragment from an IgG.

12. Robotics and automation of crystallization setups and their analysis

Advances in robotics and microfluidics have made it possible to screen two or more orders of magnitude more crystallization conditions with the same amount of sample as was possible 25 years ago. As a consequence, only a few milligrams of protein may be necessary to explore hundreds of conditions. It has only been the development and application of laboratory robotics that has made possible what we know as 'high-throughput' crystallography and has advanced crystallography into the structural genomics age.

Robotics has had a significant impact on the strategies that we currently use to search for successful conditions. What largely explains the early, rapid acceptance of 'sparse-matrix' and 'shotgun' approaches, and their continued popularity, is that they seemingly allowed the investigator to explore a wide parameter space with a limited amount of material and in a reasonable amount of time. The major weakness of these approaches, however, was that the sampling of parameter space (pH, precipitant concentration, additives, temperature *etc.*) left an uncountable number of gaps and holes. It was therefore entirely possible to miss promising conditions. In addition, optimization of conditions required considerable effort and patience, as well as skill, and often the investigator might lack the necessary commitment and simply accept the first condition that yielded crystals rather than refine it further.

With current tools and use of smaller crystals, the value of the 'shotgun' approach is reduced. Indeed, in many cases it is no longer necessary. Literally thousands of trials can be set up and evaluated robotically with a small amount of material focusing on single parameter matrices (the Slice pH kit from Hampton Research, Aliso Viejo, California, USA is a good example) that sample each variable at fine intervals. Gaps in the parameter screen are thereby eliminated, or minimized, and in addition the need for optimization is reduced. In practical terms, the initial success is much closer to what would be the refined conditions. Thus, the 'systematic search' that was originally the conventional approach appears to have made a comeback.

When an investigation is focused on a single crystallographic objective, the structure solution of a specific macromolecule or macromolecular complex, then it is wise to put as much time into headwork (thinking) as into handiwork (setting up crystallization samples). Examining and evaluating results, extracting insights and divining new directions are certainly as important as actually dispensing samples into crystallization plates. There are, however, instances where it is more efficient to automate the process by which crystallization conditions are identified and optimized. This is true when there is not one but many real or potential targets.

In recent years, robots and other automated instruments, and entire integrated systems, have been developed to accelerate the crystallization process (Bard *et al.*, 2004; DeLucas *et al.*, 2003; Hosfield *et al.*, 2003; Luft *et al.*, 2003). They have the capacity to screen thousands of crystallization conditions, and they do so precisely and reliably, with fewer errors and better record keeping

than most humans. In many large laboratories, these have become essential pieces of equipment.

Robotic systems are efficient, tireless and accurate, but in addition, they offer another important feature. They can carry out experiments using drop samples of very small volume: drops of a microlitre in most cases and of nanolitres in some. This however produces requirements for automated, micro-photographic visualization instruments, and complex storage and handling systems. On the other hand, a great advantage emerges in that they can perform enormous numbers of crystallization trials using remarkably little biological sample. This, in turn, relieves the investigator of a significant burden in terms of preparing and purifying macromolecules.

Many of the robotic systems are based on reproducing procedures currently used for manual experiments, such as sitting and hanging drops, and micro drops under oil. They are simply carried out on a much smaller scale than was possible when drops were made manually. Nanolitre drops of mother liquor are currently the rule in large-scale operations where hundreds of proteins may be in play. Robotic approaches to deploying multiple crystallization condition matrices has required the design and fabrication of new plasticware and plates, and new containers for solutions, but these needs have been adequately met by commercial suppliers.

Recently, even more miniaturized devices have come on the market. These use what is now commonly called nanotechnology to manipulate small amounts of liquids and fluid streams. A more comprehensive review of robotics and its applications will be presented in a future article; thus, touching upon every recent development in this rapidly advancing field is unwarranted here. It should, however, be noted that the use of robotic approaches has perhaps reached its lower limit of scale with the invention of 'microfluidic' plates and chips.

With microfluidic devices, constructed of specialized plastics and membranes, and composed of hundreds of intersecting channels and microcavities, solutions need only be injected at appropriate ports. In so doing, a vast array of crystallization trials is immediately formed by the physical forces acting on the constricted fluids. The 'microfluidic' chips are of such a size that efforts are currently under way to adapt them to the goniometers at synchrotron beamlines (Pinker *et al.*, 2013). If this proves to be successful and practical, then the chips would make it possible to carry out *in situ* X-ray data collection on protein crystals, as suggested some years ago (McPherson, 2000; Bingel-Erlenmeyer *et al.*, 2011).

The application of robotics, the development of an increasing number of crystallization screens and the ability to conduct more and more trials with less and less material has, in due course, expanded the number of samples deployed in experiments into the thousands for a given protein. Thus, automation has needed to be extended beyond simply setting up trays of samples to managing and analyzing the individual samples. This has involved the use of bar codes, plate-handling systems, environmental control and associated computer software systems.

The increased number of samples to be reviewed has further necessitated the development of visualization techniques to determine which samples have successfully produced crystals. While success may be obvious if the crystals are large, well formed and disposed in an otherwise clear drop, it is challenging when the crystals are of micro size, poorly formed or accompanied by precipitate.

'High-throughput' facilities currently use magnification optics and cameras to robotically and periodically examine individual samples and present images in some form to reviewers. While the eye and the mind are often the best judges of success, they are sometimes slow, become tired, and are occasionally subject to error. As a consequence, other, more automated and hopefully more reliable, methods have been developed for visualizing crystals in trials.

The most effective of these are based on fluorescence (Gill, 2010; Pusey, 2011), but there are other alternatives, such as ultraviolet light (Desbois *et al.*, 2013; Dierks *et al.*, 2010), that have been explored. Fluorescence methods may be based on the intrinsic fluorescence of proteins containing aromatic amino acids and nucleic acids composed of nucleotides. Alternately, the methods may require that some small fraction of the macromolecule be labeled with a fluorescent dye (Forsythe *et al.*, 2006). Somewhat surprisingly, the labels appear not to interfere with crystallization, and this methods holds considerable promise.

13. Statistics and data mining

Upon entering the field of macromolecular crystallography one is struck by the extraordinary range of molecules and their properties that one must contend with, and the extensive variety of techniques and conditions that must be tested in order to grow crystals suitable for X-ray diffraction analysis. It would indeed be useful if some comprehensive database existed that at least contained experiences accumulated over the years. Ideally, such a knowledge base would be combined with a system to search for and sift all kinds of relevant information regarding protein crystal growth. A notable example is a crystallization database (Gilliland *et al.*, 1994; Gilliland, 1998) that is distributed through the National Institute of Standards and Technology (http://xpdb.nist.gov:8060/). This database provides a valuable tool for the novice as well as the experienced crystallographer.

With the number of protein structures solved by X-ray crystallography now approaching 100 000, substantial databases of successful crystallization conditions and procedures could, in principle, be developed. From these, it might be expected that some predictive insight would be obtained regarding the most probable crystallization conditions for proteins in general, and for specific families of proteins sharing common physical, chemical or functional properties. The problem of succeeding in this is that each protein remains an individual endowed with its own eccentricities, even within a family, and often these dramatically alter its crystallization behavior. In addition, proteins in general may be exquisitely sensitive to only minor modifications to their properties, further complicating their rational classification in terms of crystal growth.

Nonetheless, statistical analyses are now being widely applied to the expanding databases and some results, both interesting and useful, are beginning to emerge. For example, reduced sets of most favorable crystallization conditions have been proposed by several groups of investigators based on past successes. These may be useful because they reduce the number of crystallization trials in cases where the amount of protein is limited or where a very large number of genetic constructs are to be screened. Favored reagents and, particularly, useful additives have been identified by such analyses. Correlations have been sought between the physical or chemical properties of specific proteins and their manner of crystallization, such as between pI and crystallization pH (Kantardjieff & Rupp, 2004), but this has had only limited success. If such correlations could be identified, however, this would prove a very powerful addition to the available approaches.

14. Difficult and emerging problems in crystallization

The crystallization of membrane proteins (DeLucas, 2009*a*) will be dealt with in greater detail in later articles. They use primarily the

same fundamental techniques for producing supersaturation, and borrow upon the same precipitants and other crystallization components discussed here. Proteins that are naturally membrane-associated or otherwise unusually hydrophobic or lipophilic in nature, however, invariably present unusual problems. Such proteins are, in general, only sparingly soluble in normal aqueous media, and some are virtually insoluble. This, in turn, makes the application of conventional protein crystallization techniques problematic. They are difficult but not intractable. To address these difficulties the use of detergents, particularly non-ionic detergents, has been developed. The essential difficulty with the inclusion of a solubilization agent, such as a detergent, is that it adds an additional dimension to the matrix of conditions that must otherwise be evaluated. For example, if one is content to use a standard 48-well screen of conditions, at least initially, then the additional search for a useful detergent means that the 48-sample screen must then be multiplied by the number of detergent candidates. There are a lot of potentially useful detergents. Hampton Research, a major source of screening reagents, offers three different detergent kits of 24 samples each. Were one to simply apply the basic 48-well screen with each detergent, then this would require a total of 3456 individual trials. While this may actually be possible with highly automated systems, and where a substantial amount of material is available, it is impractical for most laboratories.

The basic crystal screens, whether they are systematic screens or shotgun screens, cannot be abandoned, however. Thus, it becomes essential to reduce, at least in initial screens, the number of detergents to be considered. If, for example, a set of six highly promising detergents could be identified, then less than 300 trials would be called for initially, an undertaking that is well within the capabilities of most laboratories. No one, however, has yet definitively reduced the set to a favored few, everyone has their own opinions as to which detergents should constitute it, and no consensus set has yet emerged from databases or analyses of experiments and successful structure determinations.

To make matters in this area even worse, it appears that some, and perhaps many, detergents function best when accompanied by small amphiphilic molecules such as LDAO. This would of course add yet another dimension to the screening problem, and would seem to convert it into a hopeless exercise. Again, we can only hope that experience and the careful recording of data will provide us with a reduced set of the most promising amphiphiles.

While not as valuable as naming actual candidate detergents, the author can point to a number of useful reviews and discussions that illustrate the properties and virtues of various detergents for membrane crystallization, and also call attention to the chapter by Nollert (2004). Michel (1990) provides a good review of work up until that time: more recently, there are fine discourses and volumes by Loll (2003), Caffrey (2003), Garavito & Ferguson-Miller (2001), Hunte *et al.* (2003), Wiener (2001) and DeLucas (2009*b*).

There are other kinds of samples in addition to membrane proteins, however, that may also require additional consideration. Glycoproteins, for example, are among these. Perhaps of more immediate importance in terms of crystallization are two other categories: large protein complexes and assemblies, and nucleic acids, particularly RNA (Giegé *et al.*, 2012).

In many laboratories, interest in an individual protein independent of its complexes is minimal. How a protein interacts with its partners and contributes to the biological activity of the ensemble is the question of interest. Thus, efforts to crystallize clusters, ordered assemblies and large active particles (ribosomes, exosomes, viruses, chaperonins *etc.*) consumes much effort. It is, however, these studies that will extend our understanding of biological structure upwards on the size scale from macromolecules to organelles and ultimately to the organism level. Complexes too require special attention, more than we can give them here. Dissociation of components, proper assembly and uniformity are all challenges, principally biochemical, that have to be addressed, and all under conditions intended, at least hopefully, to produce crystals.

RNA, and to a lesser extent DNA oligomers also demand a somewhat different perspective. RNA in particular is now assuming many roles in biochemistry and molecular biology previously assigned to proteins, or that were altogether unknown. RNAs are, arguably, the most interesting molecules with which we currently deal, and the molecules about whose structural characteristics we know the least. RNAs do not exhibit the same chemical and physical features as proteins and this is commonly reflected in the differences between those conditions that yield crystals of nucleic acids and those that are successful for proteins. The optimal precipitants are frequently quite different, as are the temperature sensitivity and optimal ranges as well as the most useful additives. Thus, a profitable future field for crystallization science will undoubtedly be the nucleic acids.

Crystals of RNA share an unfortunate feature with membrane proteins in that both tend to give crystals with very high solvent content that diffract to only low, or at best modest, resolution even when using the most powerful radiation sources. A central, persistent problem in crystal growth is how to improve upon these crystals and how to obtain greater diffracting power. Structural biology is being aided in this matter somewhat by the realisation of crystallographers, including computational crystallographers, that more structural information can be extracted from low-resolution electron-density maps. Increasingly, attention is being focused on resolution limits of 3.5 Å and lower, and this is beginning to bear fruit. At the same time, however, it is clear that major advances in our understanding of what are among the most interesting macromolecules would better and more easily be achieved simply by growing better crystals.

15. The protein as a variable

Crystallization, including protein crystallization, was a classical means of purifying a chemical compound. Of equal importance, crystallization was the definitive demonstration that a compound had been obtained in a completely pure form. We accept this assessment, with some reservations, today, but for more than a hundred years it was unquestioned. Thus, it is not surprising that protein crystallization shares many features with protein purification, and in fact uses many of the same approaches. This is particularly evident in that when one considers the importance of all of the components and factors that compose a crystallization experiment, the role of the macromolecule is by far the greatest. As a consequence, the smallest improvement in a purification procedure, or the slightest change in the chemical structure or properties of the macromolecule, may have a profound effect that far exceeds any that might be achieved by variation of the crystallization conditions.

At the risk of belaboring a point, a factor of particular importance is the purity of the macromolecule (Giegé *et al.*, 1994) and this deserves special emphasis. Some proteins, it is true, may crystallize even from very heterogeneous mixtures, and as noted, crystallization has long been used as a useful purification tool. In general, however, the likelihood of success in crystal growth is greatly advanced by increased homogeneity of the sample. Investment in further purification is always warranted, and usually profitable. When every effort to crystallize a macromolecule fails, the best recourse is to further purify.

It is now generally appreciated that the protein itself is indeed the most important and influential variable in the crystallization endeavor (Dale *et al.*, 2003). When all approaches to crystallization have been exhausted, then there are two remaining options: further purifying the protein or modifying the protein. The latter may be accomplished by genetic means if the protein has been created by recombinant DNA techniques; that is, single or multiple point mutations may be introduced or truncated forms of the polypeptide may be generated. Alternately, traditional chemical reactions may be used to modify existing amino acids, or exposure to modifying enzymes may be employed; for example, the production of truncations by limited proteolysis. There are many examples of truncated proteins being successfully crystallized when the full-length polypeptide could not be. With the predictive capability of modern amino-acid sequence analysis and mass spectrometry to identify domains within proteins, designed constructs of predetermined lengths are becoming increasingly used. It has also been suggested that some surface amino acids, such as lysine and glutamic acid, inflict entropic costs when a protein crystallizes (Derewenda & Vekilov, 2005). Their substitution by other amino acids appears to be a useful approach, in some cases, to enhanced crystallizability.

Another strategy that has been used for recalcitrant proteins is to combine them in some manner with a second protein so that the complex of the two provides an additional chance for success. This idea has been used particularly with membrane proteins, where the partner protein was a Fab antibody fragment or the Fv variable light-chain domain of an antibody against the target protein. In these cases the antibody fragment enhanced the solubility of the otherwise insoluble protein and provided additional lattice contacts in the resultant crystals. There is, in principle, no reason why such 'crystallization chaperones' could not also be used with soluble proteins.

Histidine tags and proteins conjugated with a second protein, such as maltose-binding protein, are frequently produced as a basis for purification procedures for recombinant proteins. There is no reason why these 'tagged' or conjugated proteins cannot be crystallized and, indeed, many of them have been. On the other hand, the removal of the tag or conjugate and subsequent crystallization trials provides a further opportunity to obtain crystals of a protein and should also be explored.

Through truncations, mutations, chimeric conjugates and many other protein-engineering contrivances, the probability of crystallization may be significantly enhanced. If we can learn how to go about this in a rational and systematic manner then advances may occur in the succeeding years that match the progress of the past. Even so, the mother liquor must still be made and the optimal conditions identified in order to achieve success.

Stability has long been recognized as an important feature in the propensity of a macromolecule to crystallize (McPherson, 1982). Most of the earliest proteins to be crystallized were unusually stable molecules, and the value of structural stability has more recently been driven home by the vast number of proteins isolated from extremophiles, including ribosomes, studied by X-ray crystallography. A quantitative evaluation of stability, at least thermal stability, has emerged in the 'thermophore' technologies. These, in effect, use fluorescent dyes to measure the exposure of hydrophobic moieties, normally sequestered within a protein's core, as the temperature is increased. It has been observed that certain buffers, ions or other conventional small molecules sometimes positively affect the degree of stability. The results from the measurements can thus serve as a guide to the design of conditions that maintain, or even increase, the stability of a protein and, in some cases, promote its crystallization. Thermophore technologies, like the use of light scattering to detect aggregation, are gradually being integrated into crystallization strategies.

Table 5
Some important principles.

1. **Homogeneity**: begin with as pure and uniform a population of a molecular species as possible; purify, purify, purify
2. **Solubility**: dissolve the macromolecule to a high concentration without the formation of aggregates, precipitate or other phases
3. **Stability**: do whatever is necessary to maintain the macromolecules as stable and unchanging as possible.
4. **Supersaturation**: alter the properties of the solution to obtain a system which is appropriately supersaturated with respect to the macromolecule
5. **Association**: try to promote the orderly association of the macromolecules while avoiding precipitate, nonspecific aggregation or phase separation
6. **Nucleation**: try to promote the formation of a few critical nuclei in a controlled manner.
7. **Variety**: explore as many possibilities and opportunities as possible in terms of biochemical, chemical and physical parameters.
8. **Control**: maintain the system at an optimal state, without fluctuations or perturbations, during the course of crystallization
9. **Impurities**: discourage the presence of impurities in the mother liquor, and the incorporation of impurities and foreign materials into the lattice
10. **Preservation**: once the crystals are grown, protect them from shock and disruption, maintain their stability

16. Important principles

Although the approaches to macromolecular crystallization remain largely empirical, much progress has been made, particularly over the past 40 years. We have now identified useful reagents, devised a host of physical-chemical techniques for studying the crystallization process and gained a better understanding of the unique features of proteins, nucleic acids and macromolecular assemblies that affect their capacity to crystallize. Some principles now stand out regarding the crystallization problem, and these are summarized in Table 5. It remains to the individual investigator to find practical means to institute these ideas and determine for a specific problem which are of critical importance and which will have greatest influence on the likelihood of success.

17. The future of protein crystal growth

A major change in how the problem of protein crystallization is addressed that has occurred over the last 20 years, again owing principally to technological advances, is the size and the number of crystals that are needed to obtain useful X-ray diffraction data. In 1968, when one of the authors (AM) was a graduate student in Michael Rossmann's laboratory, the crystals of lactate dehydrogenase that were required for data collection (on precession cameras or a diffractometer) were roughly 0.25–1.0 mm in dimensions. Data are now frequently being obtained at synchrotrons, with microfocus beams and advanced detectors, from crystals with dimensions as small as 10 µm. Obtaining a complete data set in 1968 might have required 50 to 60 large, good-quality crystals. Now, with cryocrystallography, one crystal might suffice.

Optimistic predictions are that if the potential of free-electron lasers (FELs) for crystallography is realised then only nanocrystals will be needed (Schlichting & Miao, 2012; Yefanov & Vartanyants, 2013). This may, in the end, create a curious problem that we have previously not encountered: how does one avoid large crystals and grow nanocrystals of the appropriate size. If fully successful, the application of FELs to these nanoscale crystals will revolutionize macromolecule structure determination and may eliminate the need for what we now aspire to achieve.

However, perhaps not. Just as one technology is driving the field to smaller and smaller sizes, another is driving it in the opposite direction, and that is the design and construction of extraordinary detector systems for use in neutron diffraction studies (Myles, 2006). Neutron diffraction provides information of a different sort that will become of increasing importance in understanding enzyme mechanisms and filling in the gaps of structural detail that escape conventional X-ray crystallography. Such systems require, optimally, macromolecule crystals of one or more millimetres on their edges, with volumes of several cubic millimetres. Although this limit is also being gradually lowered by technological improvements, for the foreseeable future large crystals will remain an essential requirement.

In addition to the emergence of neutron diffraction as a practical approach to structure analysis, evidence is accumulating that X-ray diffraction data collected from unfrozen crystals at room temperature may be superior in terms of information content, as well as better reflecting physiological conditions (Fraser *et al.*, 2011). More investigators, therefore, are returning to, or at least considering, the old ways of data collection from crystals enclosed in capillaries or membranes. Because crystals at room temperature are far more sensitive to radiation decay than cryocooled crystals, larger crystals and more crystals are again in demand. The crystal shown in Fig. 13 may serve as an inspiring example.

Thus, technology may, in the end, not allow us to escape exclusively to smaller and fewer crystals. In addition, there are some crystals for which acceptable cryoconditions simply cannot be found, and room-temperature data collection is the only alternative. Some virus crystals fall into this category, but so do some protein crystals for reasons that we do not understand. Again, we will be obligated to return to traditional techniques and the persistent problems of today.

The fact that high-quality X-ray diffraction data can frequently be obtained from a single crystal of dimensions in the range of 20–50 µm has changed the objectives considerably from 40 years ago, when many crystals in the millimetre size range were required for a structure analysis. A consequence of this is that attention is turning increasingly from the systematic growth of large protein crystals (Bailey, 1942) to the nucleation and growth of any crystal. This direction has been further promoted by the development in the last 20 years of methods to use even the data from twinned or disordered crystals. While smaller, fewer crystals are now the rule (with the exception of those required for neutron diffraction), this has not eliminated the need for optimizing crystal quality, nor the requirement that at least some sort of crystal be obtained. Thus, attention is now focusing increasingly on nucleation, perhaps always the most problematic step in the crystallization process, and on enhancing crystal perfection. These continue to pose formidable problems.

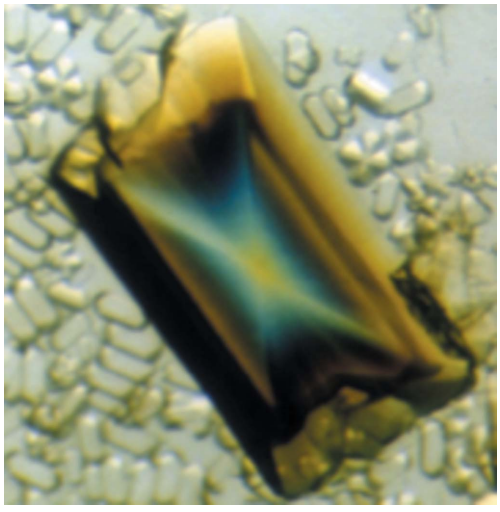


Figure 13
Shown here is a single large crystal of *Satellite tobacco mosaic virus* that is approximately 1.5 mm in the longest dimension and which shows a high degree of birefringence under polarized light. This crystal was grown in microgravity aboard the US Space Shuttle in 1991.

Furthermore, the objectives of crystallization, the entities to be crystallized will, as noted above, continue to become more challenging. In addition to membrane proteins that present difficult problems owing to their solubility (dealt with in another article), interest has increasingly turned toward the solutions of the structures of RNA, glycoproteins, lipoproteins and larger proteins, or protein–nucleic acid complexes and assemblies. It is unlikely that crystals with unit cells much above 1200 Å can be solved even with X-ray technologies currently under development. Assemblies such as large icosahedral viruses, that do yield crystals amenable to analysis are remarkably fragile in a mechanical sense, and large unit-cell sizes require crystals to greatly exceed the small sizes of conventional protein crystals in order to yield diffraction intensities that can be measured accurately. Additional problems will arise from proteins conjugated with other entities of significant size such as lipids and oligosaccharides, which are often disordered, and with proteins that are unstructured, in whole or in part.

Finally, we have come to believe that the structure of a protein in the crystal is the same as the structure of the protein in solution. However, when the protein has a spectrum of conformations in solution, often as a consequence of its function, then to visualize it in full one needs to see it in multiple crystal forms. Thus, it will be increasingly necessary to grow crystals not simply of the apoprotein, but of its important ligand complexes and possibly of several polymorphs. By studying the protein in a variety of crystal forms, its conformational variety may be appreciated and its dynamic range delineated.

Although it is difficult to quantitate its importance, and difficult to assign to it any specific successes, the field has benefitted by an explosion in the understanding at the molecular level of the physics and chemistry of protein crystallization (Chernov, 2003; Vekilov & Chernov, 2002; Chernov & Komatsu, 1995; McPherson, 1999). The application of interferometry, time-lapse video, atomic force microscopy, static and dynamic light scattering and a host of other physical methods has brought forth a wealth of information on thermodynamics and kinetics, the phenomenon of nucleation and the solution factors that affect it, the mechanisms of crystal growth, the factors responsible for growth termination and impurity incorporation, the nature of the defects and their extent, and many other features of crystallization. Thus, we now often know better what we are trying to achieve, even though the way to achieving it may remain cloudy. We better understand our failures and the means of correction. Other articles in this series will address these contributions in greater detail in ways that may profoundly affect the way we think of crystallization.

References

Abel, J. J., Geiling, E. M. K., Roultier, O. A., Bell, F. M. & Wintersteiner, O. (1927). *Crystalline insulin. J. Pharmacol. Exp. Ther.* **31**, 65–85.
Bailey, K. (1942). *The growth of large single crystals of proteins. Faraday Soc. Trans.* **38**, 186–192.
Bard, J., Ercolani, K., Svenson, K., Olland, A. & Somers, W. (2004). *Automated systems for protein crystallization. Methods*, **34**, 329–347.

Bergfors, T. M. (1999). *Protein Crystallization: Techniques, Strategies, and Tips*. La Jolla: International University Line.

Bergfors, T. (2003). *Seeds to crystals. J. Struct. Biol.* **142**, 66–76.

Bingel-Erlenmeyer, R., Olieric, V., Grimshaw, J. P. A., Gabadinho, J., Wang, X., Ebner, S. G., Isenegger, A., Schneider, R., Schneider, J., Glettig, W., Pradervand, C., Panepucci, E. H., Tomizaki, T., Wang, M. & Schulze-Briese, C. (2011). *SLS Crystallization Platform at Beamline X06DA – A Fully Automated Pipeline Enabling in Situ X-ray Diffraction Screening. Cryst. Growth Des.* **11**, 916–923.

Bolen, D. W. (2004). *Effects of naturally occurring osmolytes on protein stability and solubility: issues important in protein crystallization. Methods*, **34**, 312–322.

Caffrey, M. (2003). *Membrane protein crystallization. J. Struct. Biol.* **142**, 108–132.

Chayen, N. E. (1997). *A novel technique to control the rate of vapour diffusion, giving larger protein crystals. J. Appl. Cryst.* **30**, 198–202.

Chayen, N. E., Saridakis, E. & Sear, R. P. (2006). *Experiment and theory for heterogeneous nucleation of protein crystals in a porous medium. Proc. Natl Acad. Sci. USA*, **103**, 597–601.

Chernov, A. A. (2003). *Protein crystals and their growth. J. Struct. Biol.* **142**, 3–21.

Chernov, A. A. & Komatsu, H. (1995). *Principles of crystal growth in protein crystallization*, in *Science and Technology of Crystal Growth*, edited by J. P. van der Eerden & O. S. Bruinsma. Doderecht: Kluwer.

Cohn, E. J. & Edsall, J. T. (1943). Editors. *Proteins, Amino Acids and Peptides as Ions and Dipolarions*. Princeton: Van Nostrand-Reinhold.

Cohn, E. J. & Ferry, J. D. (1943). *The interactions of proteins with ions and dipolar ions. Proteins, Amino Acids and Peptides as Ions and Dipolarions*, edited by E. J. Cohn & J. T. Edsall, pp. 586–622. Princeton: Van Nostrand-Rheinhold.

Cohn, E. J., Hughes, W. L. Jr & Weare, J. H. (1974). *Preparation and Properties of Serum and Plasma Proteins. XIII. Crystallization of Serum Albumins from Ethanol-Water Mixtures. J. Am. Chem. Soc.* **69**, 1753–1761.

Collins, K. D. (2004). *Ions from the Hofmeister series and osmolytes: effects on proteins in solution and in the crystallization process. Methods*, **34**, 300–311.

Collins, K. D. & Washabaugh, M. W. (1985). *The Hofmeister effect and the behaviour of water at interfaces. Q. Rev. Biophys.* **18**, 323–422.

Dale, G. E., Oefner, C. & D'Arcy, A. (2003). *The protein as a variable in protein crystallization. J. Struct. Biol.* **142**, 88–97.

D'Arcy A., Villarda, F. & Marsh, M. (2007). *An automated microseed matrix-screening method for protein crystallization. Acta. Cryst.* D**63**, 550–554.

DeLucas, L. (2009*a*). *Membrane Protein Crystallization. J. Struct. Biol.* **142**, 108–132.

DeLucas, L. (2009*b*). Editor. *Membrane Protein Crystallization.* Amsterdam: Elsevier.

DeLucas, L. J., Bray, T. L., Nagy, L., McCombs, D., Chernov, N., Hamrick, D., Cosenza, L., Belgovskiy, A., Stoops, B. & Chait, A. (2003). *Efficient protein crystallization. J. Struct. Biol.* **142**, 188–206.

Derewenda, Z. S. & Vekilov, P. G. (2005). *Entropy and surface engineering in protein crystallization. Acta. Cryst.* D**62**, 116–124.

Desbois, S., Seabrook, S. A. & Newman, J. (2013). *Some practical guidelines for UV imaging in the protein crystallization laboratory. Acta Cryst.* F**69**, 201–208.

Dickerson, R. E. (2005). *Present at the flood: How structural molecular biology came about. FASEB J.* **20**, 809–810.

Dierks, K., Meyer, A., Oberthür, D., Rapp, G., Einspahr, H. & Betzel, C. (2010). *Efficient UV detection of protein crystals enabled by fluorescence excitation at wavelengths longer than 300 nm. Acta Cryst.* F**66**, 478–484.

Ducruix, A. & Giége, R. (1992). *Crystallization of Nucleic Acids and Proteins: A Practical Approach.* Oxford: IRL Press.

Durbin, S. D. & Feher, G. (1996). *Protein crystallization. Annu. Rev. Phys. Chem.* **47**, 171–204.

Englard, S. & Seifter, S. (1990). *Precipitation techniques. Methods Enzymol.* **182**, 285–300.

Feher, G. (1986). *Mechanisms of nucleation and growth of protein crystals. J. Cryst. Growth*, **76**, 545–546.

Feigelson, R. S. (1988). *The relevance of small molecule crystal growth theories and techniques to the growth of biological macromolecules. J. Cryst. Growth*, **90**, 1–13.

Fitzgerald, P. M. D. & Madsen, N. B. J. (1987). *Improvement of limit of diffraction and uesful X-ray lifetime of crystals of glycogen debranching enzyme. J. Cryst. Growth*, **76**, 600–606.

Forsythe, E., Achari, A. & Pusey, M. L. (2006). *Trace fluorescent labeling for high-throughput crystallography. Acta Cryst.* D**62**, 339–346.

Fraser, J. S., van den Bedem, H., Samelson, A. J., Lang, P. T., Holton, J. M., Echols, N. & Alber, T. (2011). *Accessing protein conformational ensembles using room-temperature X-ray crystallography. Proc. Natl Acad. Sci. USA*, **108**, 16247–16252.

Funke, O. (1851). *Über das milzvenenblut. Z. Rat. Med.* **1**, 172–179.

Garavito, R. M. & Ferguson-Miller, S. (2001). *Detergents as tools in membrane biochemistry. J. Biol. Chem.* **276**, 32403–32406.

García-Ruiz, J. M. (2003). *Counterdiffusion methods for macromolecular crystallization. Methods Enzymol.* **368**, 130–154.

Garman, E. F. & Schneider, T. R. (1997). *Macromolecular Cryocrystallography. J. Appl. Cryst.* **30**, 211–237.

Gavira, J. A., Hernandez-Hernandez, M. A., Gonzalez-Ramirez, L. A., Briggs, R. A., Kolek, S. A. & Shaw Stewart, P. D. (2011). *Combining Counter-Diffusion and Microseeding to Increase the Success Rate in Protein Crystallization. Cryst. Growth. Des.* **11**, 2122–2126.

Gavira, J. A., Toh, D., Lopéz-Jaramillo, J., García-Ruiz, J. M. & Ng, J. D. (2002). *Ab initio crystallographic structure determination of insulin from protein to electron density without crystal handling. Acta Cryst.* D**58**, 1147–1154.

Giegé, R., Jühling, F., Pütz, J., Stadler, P., Sauter, C. & Florentz, C. (2012). *Structure of transfer RNAs: similarity and variability. Wiley Interdiscip. Rev. RNA*, **3**, 37–61.

Giegé, R., Lorber, B. & Théobald-Dietrich, A. (1994). *Crystallogenesis of biological macromolecules: facts and perspectives. Acta Cryst.* D**50**, 339–350.

Gill, H. S. (2010). *Evaluating the efficacy of tryptophan fluorescence and absorbance as a selection tool for identifying protein crystals. Acta Cryst.* F**66**, 364–372.

Gilliland, G. L. (1998). *A biological macromolecule crystallization database: a basis for a crystallization strategy. J. Cryst. Growth*, **90**, 51–59.

Gilliland, G. L., Tung, M., Blakeslee, D. M. & Ladner, J. E. (1994). *Biological macromolecule crystallization database, version 3.0: New features, data and the NASA Archive for protein crystal growth data. Acta Cryst.* D**50**, 51–59.

Goldberg, M. E., Expert-Bezançon, N., Vuillard, L. & Rabilloud, T. (1996). *Non-detergent sulphobetaines: a new class of molecules that facilitate in vitro protein renaturation. Fold. Des.* **1**, 21–27.

Gruner, S. M., Eikenberry, E. F. & Tate, M. W. (2001). *International Tables for Crystallography*, 1st ed., edited by M. G. Rossmann & E. Arnold, pp. 143–153. Dordrecht: Kluwer Academic Publishers.

Haas, C. & Drenth, J. (1999). *Understanding protein crystallization on the basis of the phase diagram. J. Cryst. Growth*, **196**, 388–394.

Helliwell, J. R. (1992). *Macromolecular Crystallography with Synchrotron Radiation.* Cambridge University Press.

Herriott, R. M. (1942). *Solubility Methods in the Study of Proteins. Chem. Rev.* **30**, 413–421.

Hofmeister, T. (1890). *Über die Darstellung von krystallisirtem Eiralbumin und die Krystallisirbarkeit colloider Stoffe. Z. Physiol. Chem.* **14**, 165.

Hopkins, F. G. & Pincus, S. N. (1898). *Animal proteins. J. Physiol.* **23**, 130.

Hosfield, D., Palan, J., Hilgers, M., Scheibe, D., McRee, D. E. & Stevens, R. C. (2003). *A fully integrated protein crystallization platform for small-molecule drug discovery. J. Struct. Biol.* **142**, 207–217.

Hunefeld, F. L. (1840). *Der Chemismus in der tierescher Organization*, p. 160. Leipzig University, Germany.

Hunte, C., von Jagow, G. & Schägger, H. (2003). *Membrane Protein Purification and Crystallization: A Practical Guide*. San Diego: Academic Press.

Ingham, K. C. (1990). *Precipitation of proteins with polyethylene glycol. Methods Enzymol.* **182**, 301–306.

Kantardjieff, K. & Rupp, B. (2004). *Protein pI as a predictor for increased crystallization screening efficiency. Bioinformatics*, **20**, 2162–2168.

Kuznetsov, Y. G., Day, J., Newman, R. & McPherson, A. (2000). *Chimeric Human–Simian Anti-CD4 Antibodies Form Crystalline High Symmetry Particles. J. Struct. Biol.* **131**, 108–115.

Kuznetsov, Y. G., Malkin, A., Greenwood, A. & McPherson, A. (1995). *Interferometric Studies of Growth Kinetics and Surface Morphology in Macromolecular Crystal Growth: Canavalin, Thaumatin, and Turnip Yellow Mosaic Virus. J. Struct. Biol.* **114**, 184–196.

Kuznetsov, Y. G., Malkin, A. & McPherson, A. (1998). *Atomic force microscopy studies of phase separations in macromolecular systems. Phys. Rev. B*, **58**, 6097–6103.

Kuznetsov, Y., Malkin, A. & McPherson, A. (2001). *The influence of precipitant concentration on macromolecular crystal growth mechanisms. J. Cryst. Growth*, **232**, 114–118.

Larson, S. B., Day, J. S., Nguyen, C., Cudney, R. & McPherson, A. (2008). *Progress in the Development of an Alternative Approach to Macromolecular Crystallization. Cryst. Growth Des.* **8**, 3038–3052.

Loll, P. J. (2003). *Membrane protein structural biology: the high throughput challenge. J. Struct. Biol.* **142**, 144–153.
Luft, J. R., Collins, R. J., Fehrman, N. A., Lauricella, A. M., Veatch, C. K. & DeTitta, G. T. (2003). *A deliberate approach to screening for initial crystallization conditions of biological macromolecules. J. Struct. Biol.* **142**, 170–179.
Maclean, D. S., Qian, Q. & Middaugh, C. R. (2002). *Stabilization of proteins by low molecular weight multi-ions. J. Pharm. Sci.* **91**, 2220–2229.
Malkin, A. J., Kuznetsov, Y. G., Glantz, W. & McPherson, A. (1996). *Atomic Force Microscopy Studies of Surface Morphology and Growth Kinetics in Thaumatin Crystallization. J. Phys. Chem.* **100**, 11736–11743.
Malkin, A. J., Kuznetsov, Y. G., Land, T., DeYoreo, J. & McPherson, A. (1995). *Mechanisms of growth for protein and virus crystals. Nature Struct. Biol.* **2**, 956–959.
Malkin, A. J., Kuznetsov, Y. G. & McPherson, A. (1997). *An in situ AFM investigation of catalase crystallization. Surf. Sci.* **393**, 95–107.
McPherson, A. (1976*a*). *Crystallization of proteins from polyethylene glycol. J. Biol. Chem.* **251**, 3600–6303.
McPherson, A. (1976*b*). *Methods of Biochemical Analysis*, edited by D. Glick, pp. 249–345. New York: Academic Press.
McPherson, A. (1982). *Preparation and Analysis of Protein Crystals.* New York: John Wiley & Sons.
McPherson, A. (1989). *Macromolecular crystals. Sci. Am.* **260**, 62–69.
McPherson, A. (1991). *A brief history of protein crystal growth. J. Cryst. Growth*, **110**, 1–10.
McPherson, A. (1999). *Crystallization of Biological Macromolecules.* Cold Spring Harbor: Cold Spring Harbor Laboratory Press.
McPherson, A. (2000). *In situ X-ray crystallography. J. Appl. Cryst.* **33**, 397–400.
McPherson, A. & Cudney, B. (2006). *Searching for silver bullets: an alternative strategy for crystallizing macromolecules. J. Struct. Biol.* **156**, 387–406.
McPherson, A., Cudney, R. & Patel, S. (2003). *The Crystallization of Proteins, Nucleic Acids and Viruses for X-ray Diffraction Analysis. Biopolymers*, Vol. 8, edited by S. R. Fahnestock & A. Steinbüchel, pp. 427–468. Berlin: Wiley-VCH.
McPherson, A., Malkin, A. J. & Kuznetsov, Y. G. (1995). *The science of macromolecular crystallization. Structure*, **3**, 759–768.
McPherson, A., Malkin, A. J. & Kuznetsov Y. G. (2000). *Atomic force microscopy in the study of macromolecular crystal growth. Annu. Rev. Biophys. Biomol. Struct.* **29**, 361–410.
McPherson, A. & Shlichta, P. (1988). *Heterogeneous and epitaxial nucleation of protein crystals on mineral surfaces. Science*, **239**, 385–387.
Michel, H. (1990). *Membrane protein crystallization. J. Struct. Biol.* **142**, 108–132.
Mirzabekov, A. D., Rhodes, D., Finch, J. T., Klug, A. & Clark, B. F. (1972). *Crystallization of tRNAs as cetyltrimethylammonium salts. Nature New Biol.* **237**, 27–28.
Myles, D. (2006). *Neutron protein crystallography: current status and a brighter future. Curr. Opin. Struct. Biol.* **16**, 630–637.
Neugebauer, J. M. (1990). *Detergents: an overview. Methods Enzymol.* **182**, 239–253.
Ng, J. D., Gavira, J. A. & García-Ruiz, J. M. (2003). *Protein crystallization by capillary counterdiffusion for applied crystallographic structure determination. J. Struct. Biol.* **142**, 218–231.
Nollert, P. (2004). *Lipidic cubic phases as matrices for membrane protein crystallization. Methods*, **34**, 348–353.
Northrop, M., Kunitz, M. & Herriott, R. M. (1948). *Crystalline enzymes*. New York: Columbia University Press.
Osborne T. (1891). *The proteids of the Brazil nut. Am. Chem. J.* **13**, 212–218.
Osborne T. (1892). *Crystallized vegetable proteins. Am. Chem. J.* **14**, 215–220.
Osborne T. (1894). *The proteids of the kidney bean. J. Am. Chem. Soc.* **16**, 633–640.
Osborne T. (1899). *On some definite compounds of protein bodies. J. Am. Chem. Soc.* **21**, 486–492.
Otálora, F., Gavira, J. A., Ng, J. D. & García-Ruiz, J. M. (2009). *Counterdiffusion methods applied to protein crystallization. Prog. Biophys. Mol. Biol.* **101**, 26–37.
Patel, S., Cudney, B. & McPherson, A. (1995). *Polymeric precipitants for the crystallization of macromolecules. Biochem. Biophys. Res. Commun.* **207**, 819–828.
Pflugrath, J. W. (1992). *Developments in X-ray detectors. Curr. Opin. Struct. Biol.* **2**, 811–815.
Piazza, R. (1999). *Interactions in protein solutions near crystallisation: a colloid physics approach. J. Cryst. Growth*, **196**, 415–423.
Pinker, F., Brun, M., Morin, P., Deman, A., Chateaux, J., Oliéric, V., Stirnimann, C., Lorber, B., Terrier, N., Ferrigno, R. & Sauter, C. (2013). *ChipX: A Novel Microfluidic Chip for Counter-Diffusion Crystallization of Biomolecules and in Situ Crystal Analysis at Room Temperature. Cryst. Growth Des.* **13**, 3333–3340.
Pusey, M. L. (2011). *Developing a fluorescence-based approach to screening for macromolecule crystallization conditions. Cryst. Growth Des.* **11**, 1135–1142.
Ray, W. J. & Bracker, C. E. (1986). *Polyethylene glycol: Catalytic effect on the crystallization of phosphoglucomutase at high salt concentration. J. Cryst. Growth*, **76**, 562–576.
Ray, W. J. & Puvathingal, J. M. (1986). *The effect of polyethylene glycol on the growth and dissolution rates of a crystalline protein at high salt concentration. Phosphoglucomutase. J. Biol. Chem.* **261**, 11544–11549.
Reichert, E. T. & Brown, A. P. (1909). *The Differentiation and Specificity of Corresponding Proteins and other Vital Substances in Relation to Biological Classification and Organic Evolution: The Crystallography of Hemoglobins*. Washington: Carnegie Institute.
Ritthausen, H. (1880). *Über die Eiweisskorper verschiedenen Oelsamen. Pfluegers Arch.* **21**, 81–104.
Ritthausen, H. (1881). *Krystallinisch Eiweisskorper aus verschiedenen Oelsman. J. Prakt. Chem.* **23**, 481.
Rossmann, M. G. & Arnold, E. (2001). *Crystallography of Biological Macromolecules.*, 1st ed. Dordrecht: Kluwer Academic Publishers.
Schlichting, I. & Miao, J. (2012). *Emerging opportunities in structural biology with X-ray free-electron lasers. Curr. Opin. Struct. Biol.* **22**, 613–626.
Stura, E. A. & Wilson, I. A. (1991). *Applications of the streak seeding technique in protein crystallization. J. Cryst. Growth*, **110**, 270–282.
Sumner, J. B. (1926). *The isolation and crystallization of the enzyme urease: preliminary paper. J. Biol. Chem.* **69**, 435.
Sumner, J. B. & Somers, G. F. (1943). *The Enzymes*. New York: Academic Press.
Thaller, C., Eichele, G., Weaver, L. H., Wilson, E., Karlsson, R. & Jansonius, J. N. (1985). *Diffraction methods for biological macromolecules. Seed enlargement and repeated seeding. Methods Enzymol.* **114**, 132–135.
Vekilov, P. G. & Chernov, A. A. (2002). *The physics of protein crystallization. Solid State Phys.* **57**, 2–147.
Wiener, M. C. (2001). *Existing and emergent roles for surfactants in the three-dimensional crystallization of integral membrane proteins. Curr. Opin. Colloid Interface Sci.* **6**, 412–419.
Wiener, M. C. (2004). *A pedestrian guide to membrane protein crystallization. Methods*, **34**, 364–372.
Wolde, P. R. ten & Frenkel, D. (1997). *Enhancement of protein crystal nucleation by critical density fluctuations. Science*, **277**, 1975–1978.
Yefanov, O. M. & Vartanyants, I. A. (2013). *X-ray Diffraction: Modern Expeerimental Techniques*, edited by O. H. Seech & B. M. Murphy. Singapore: Pan Stanford.
Zulauf, M. (1990). *Crystallization of Membrane Proteins*, edited by H. Michel. Boca Raton: CRC Press.