

# Chapter 2

## Protein Crystallization

Alexander McPherson

### Abstract

Protein crystallization was discovered by chance nearly 200 years ago and was developed in the late nineteenth century as a powerful purification tool, and 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 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 due to the advent of practical, easy-to-use screening kits, and the application of laboratory robotics.

**Key words** Crystals, Supersaturation, Growth mechanisms, Homogeneity, X-ray diffraction, Precipitants, Crystallization methods, Vapor diffusion, Dialysis, Mother liquor

---

## 1 Introduction

Although the technologies of nuclear magnetic resonance and, more recently, cryogenic electron microscopy, have made significant inroads, presently the only technique that can yield atomic level structural images of biological macromolecules is X-ray diffraction analysis as applied to single crystals. While other methods may produce important structural and dynamic data only X-ray crystallography is adequate to precisely define atomic coordinates. The 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 accurate data collection. The quality of the final structural image is directly determined by the perfection and physical properties of the crystalline specimen. The crystals, therefore, become the keystone element of the entire process, and the ultimate determinant of its success. The crystals

themselves have no medicinal or pharmaceutical value, but provide the X-ray diffraction patterns that serve as the fundamental data, which through Fourier synthesis, allow the direct visualization of the macromolecules or their complexes composing the crystals.

When crystallizing proteins for X-ray diffraction analysis, one is usually dealing with homogenous, often exceptionally pure macromolecules, and the objective is to grow only a few large, perfect crystals. The proteins themselves may be purified from natural sources, microbes or tissues of plants and animals, or it may be produced by recombinant DNA techniques. The number of crystals needed for recording data may be few, but often the amount of protein available is severely limited. This in turn places constraints on the approaches and strategies that can be used to obtain those crystals. While new methodologies such as synchrotron radiation [1, 2] and cryocrystallography [3–5] have driven the necessary size of specimen crystals consistently downward, they have not eliminated the need for crystal perfection.

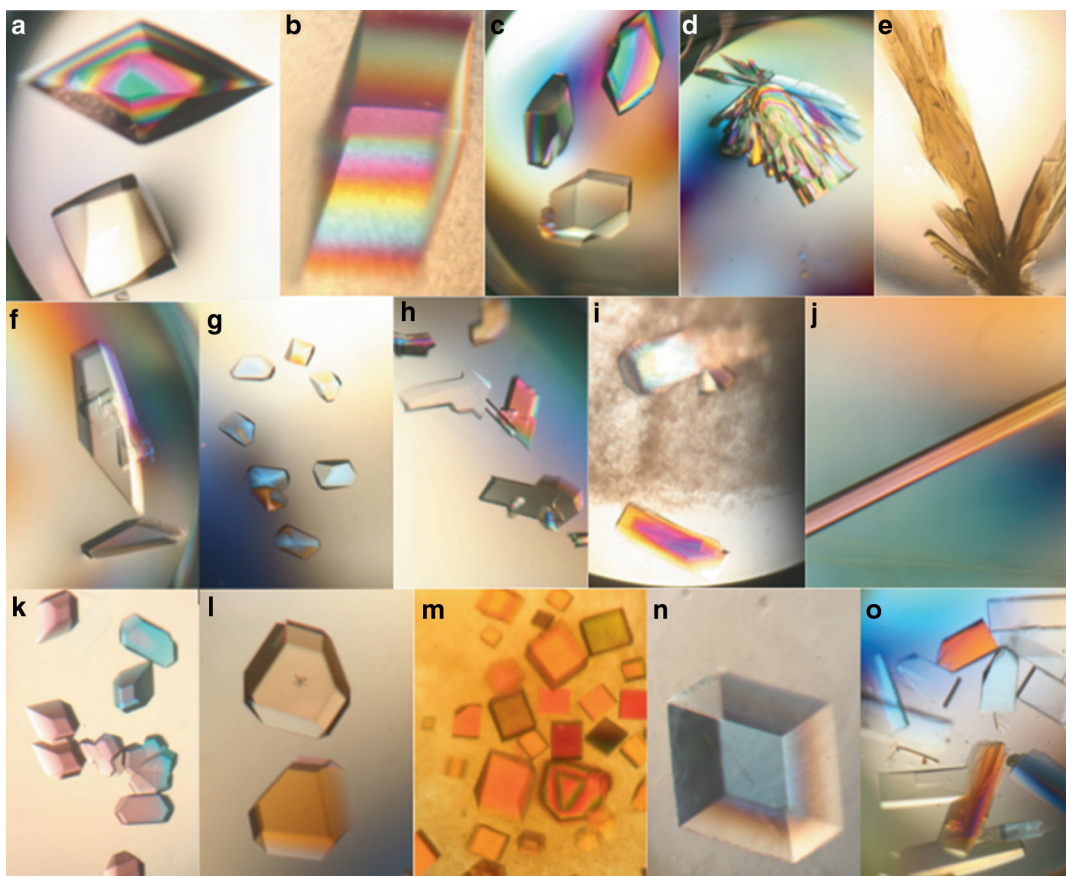
---

## 2 The Nature of Protein Crystals

Protein crystals are composed of approximately 50% solvent on average. Those seen in Fig. 1 vary from 33% solvent for monoclinic lysozyme up to 61% for concanavalin B. At the extremes one finds insulin at about 25% and tropomyosin at 90%. Protein occupies the remaining volume so that the entire crystal may be thought of as an ordered gel permeated by extensive networks of channels and interstitial spaces filled with solvent, through which small molecules can diffuse. There does not appear to be a direct correlation between the solvent volume of a protein crystal and its diffraction properties. It has, however, been noted that transitions of a crystallographic unit cell to smaller volume, with concomitant reduction of included solvent, has frequently produced an improvement in diffraction resolution [6, 7].

In proportion to molecular mass, the number of contacts (salt bridges, hydrogen bonds, hydrophobic interactions) that a conventional organic molecule forms with its neighbors in a crystal far exceeds the very few exhibited by crystalline macromolecules. Since these contacts provide the lattice interactions essential for crystal integrity, this largely explains the differences in properties between crystals of salts or small molecules and macromolecules. It may also explain why the introduction of a few additional contacts, or even one uniquely strong interaction, can profoundly affect the diffraction resolution of a protein crystal.

Living systems are based almost exclusively on aqueous chemistry within narrow ranges of temperature and pH. Macromolecules, thus, have evolved an appropriate compatibility and dependency. Serious deviations or perturbations are rarely tolerated. As a consequence, all protein crystals are grown from aqueous media, ones to



**Fig. 1** An array of protein crystals showing the range of habits they may assume: in (a) thaumatin, (b) bovine trypsin, (c) tetragonal lysozyme, (d) monoclinic lysozyme, (e) beef liver catalase, (f–h) three different crystal forms of bovine RNase S, (i) beta-lactoglobulin, (j) concanavalin B, (k) satellite tobacco mosaic virus, (l) glucose isomerase, (m) concanavalin A, (n) rhombohedral canavalin, and (o) orthorhombic canavalin

which they are tolerant, and these solutions are called mother liquors. As described below, crystals can be made to grow from these mother liquors when the mother liquors are made supersaturated in protein.

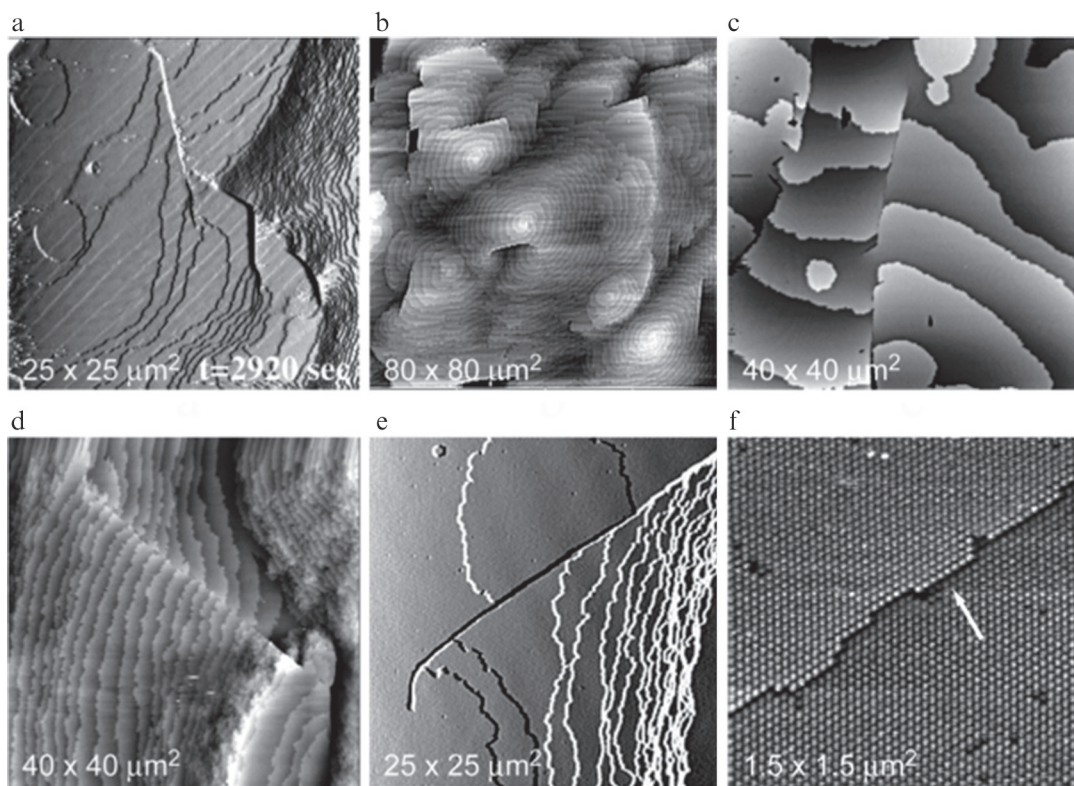
There are important physical and chemical differences between ionic crystals, or those of most low-molecular-mass compounds, and crystals of proteins. For example, protein crystals generally have fairly simple morphologies, or habits as they are called, while conventional crystals often display very complex polyhedral or prismatic appearances. This is mainly due to the absence of centers of symmetry, mirror planes, and glide planes in protein crystals. Proteins exist in only one enantiomeric form and, therefore, cannot have such symmetry elements in their space groups. As a further consequence, protein crystals can fall into only 65 space groups rather than the 230 space groups allowed mixtures of enantiomers, and these 65 tend to have rather simple point group symmetries that are reflected in the habits.

Conventional crystals are characterized by firm lattice interactions, are usually well ordered, physically hard and brittle in general, relatively easy to manipulate, usually can 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. Protein crystals are temperature sensitive and undergo extensive damage after prolonged exposure to radiation. Frequently, several crystals must be analyzed for a structure determination to be successful although the advent of cryocrystallography [3–5, 8] pixel area detectors of very high photon counting efficiency [9], high intensity synchrotron X-ray sources [1, 8], and new phasing methods [10] have greatly lessened this constraint. Those same advancements have also reduced the size (volume) of crystals useful for X-ray diffraction analysis. Until the 1990s, crystals in the range of dimensions 0.25–1.0 mm were commonly required. Currently, structures can be determined from crystals in the range of 20–50  $\mu\text{m}$ .

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 in turn corresponds closely with that 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. Protein crystals, all crystals in fact, are not uniform, flawless solids, but exhibit many defects and dislocations that produce a mosaic pattern of slightly misaligned sectors, or domains. Domain boundaries, often referred to as stacking faults or grain boundaries in conventional crystals, are far more numerous in protein crystals than conventional crystals, probably by several orders of magnitude [11]. These features contribute further to the limitation of diffraction quality. Some defects seen in protein and virus crystals by atomic force microscopy (AFM) are presented in Fig. 2.

The liquid channels and solvent filled cavities that permeate macromolecular crystals and the lack of order they engender are primarily responsible for the limited resolution of the diffraction patterns. Because of the relatively large solvent 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 the aqueous environment of a crystal may exhibit slight variations in the course of their polypeptide chains or the dispositions of side groups from one to another.

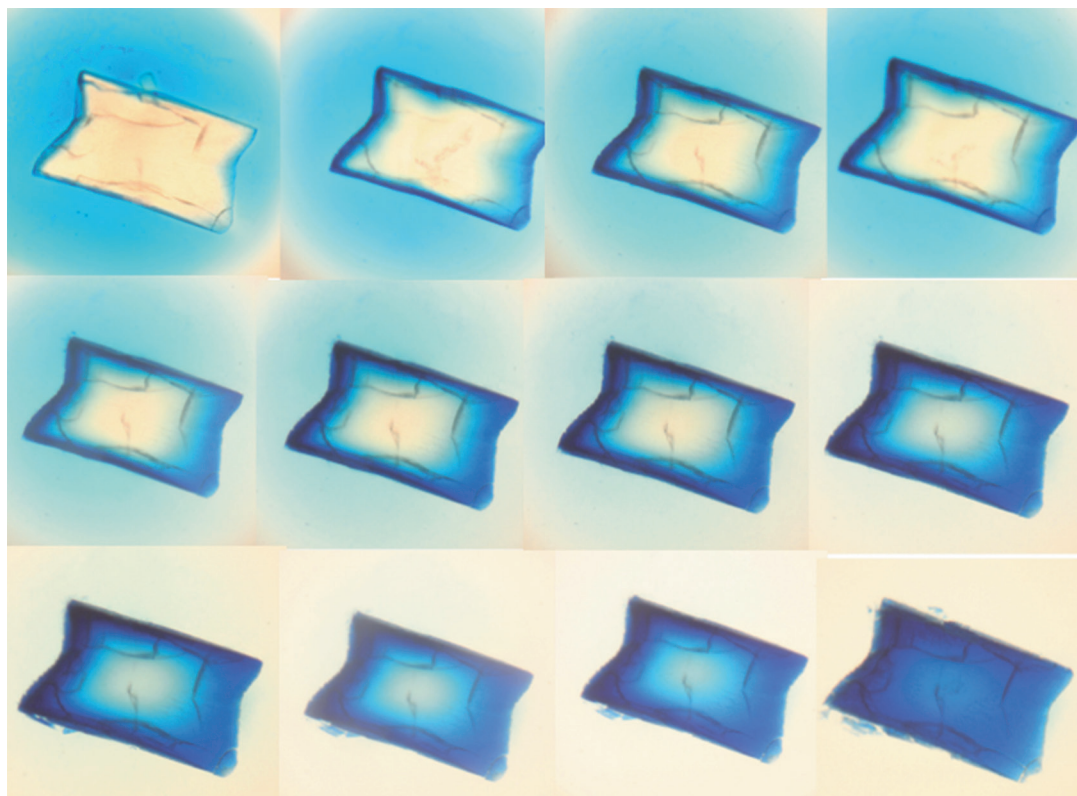




**Fig. 2** Planar defects (stacking faults) in crystals of proteins and viruses. (a), (c), (d), and (e) are surfaces of satellite tobacco mosaic virus crystals, (b) canavalin and (f) a crystal of cucumber mosaic virus. The planar defects, homologous to grain boundaries in conventional crystals, divide the crystal into domains, which in turn are responsible for the mosaicity of the crystals

Although the presence of extensive solvent regions is a major contributor to the generally modest diffraction quality of protein crystals, it is also largely responsible for their value to biochemists as platforms for experimentation. 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 activity, 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, though immobilized, are frequently accessible for experimentation simply through alteration of the surrounding mother liquor.

Figure 3 shows, in a representative manner, how small organic molecules diffuse into a protein crystal through its network of solvent channels. The blue dye xylene cyanol, a molecule in the

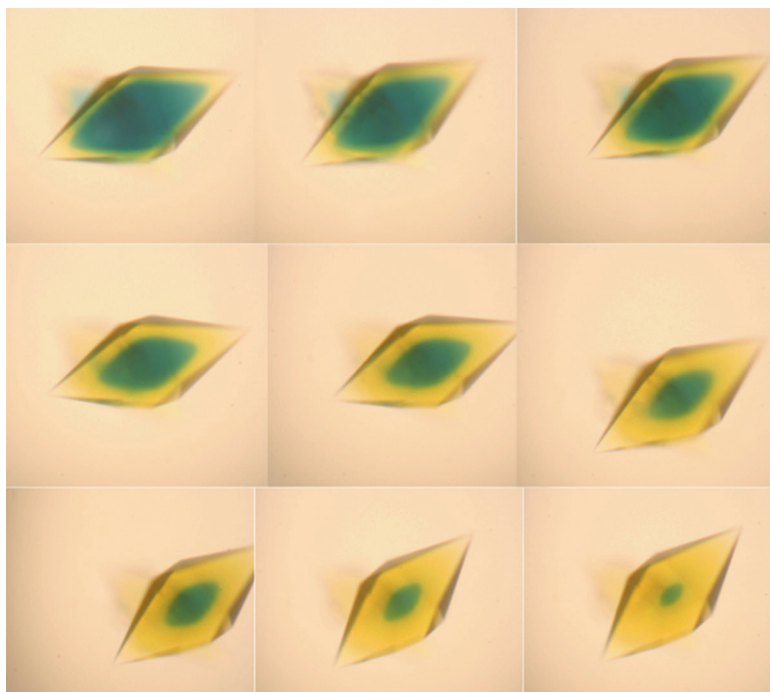


**Fig. 3** A large crystal (about 1.5 mm in length) of the protein canavalin has had its mother liquor replaced with an equivalent mother liquor containing the blue dye xylene cyanol. This series of photographs taken over about 8 h shows the diffusion of the dye molecules into the protein crystal

molecular weight range of a biological coenzyme or a possible drug, was added directly to the mother liquor of a large canavalin crystal. The dye front, as it diffuses into the crystal is clearly evident, and its progress could be recorded and measured. From this it could be estimated that the dye, the small molecule, diffused through the crystal lattice at a rate of about  $60 \mu\text{m}/\text{h}$ .

Figure 4 illustrates another experiment where a large thaumatin crystal was saturated with the pH sensitive dye m-cresol purple at high pH (pH 8) giving it a blue color. The mother liquor was then replaced with an equivalent one but at low pH (pH 6). As  $\text{H}_3\text{O}^+$  ions diffused into the crystal, the dye internal to the crystal changed to a yellow color. Again, the dye transition front and its movement through the crystal was photographically recorded and measured. From this experiment it could be estimated that when a gradient of  $\text{H}_3\text{O}^+$  of  $10^{-8} > 10^{-6}$  exists between the interior of the crystal and its mother liquor,  $\text{H}_3\text{O}^+$  ions diffuse to the center of the crystal with an average rate of about  $1000 \mu\text{m}/\text{h}$ .

A diversity of crystallographic unit cells and habits that we refer to as polymorphism are common phenomena with macromolecular



**Fig. 4** A crystal of thaumatin was saturated with the pH sensitive dye m-bromocresol purple at pH 8 where the dye is *blue*. The mother liquor was then replaced with an equivalent mother liquor at pH 6. As the H<sub>3</sub>O<sup>+</sup> ions diffused into the crystals, the internal m-bromocresol purple dye molecules changed color to *yellow*. The experiment, photographed over a period of about 35 min, shows the progress of the diffusion of the H<sub>3</sub>O<sup>+</sup> ions into the protein crystal

crystals. In Fig. 1 alone we see crystals with three different unit cells of RNase S, two of both lysozyme and canavalin. Glucose isomerase, catalase, and trypsin can also be induced to crystallize in additional unit cells. Presumably this is a consequence of the protein's conformational dynamic range and the sensitivity of the lattice contacts involved. Thus, different unit cells and different symmetries may arise under 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 as in Fig. 5.

---

### 3 Energetics, Kinetics, and Mechanisms of Protein Crystallization

There are further differences that complicate the crystallization of proteins as compared with conventional, small molecules [12–18]. First, proteins may coalesce to form several solid or dense liquid states that include amorphous precipitates, oils, or gels, as well as crystals, and most of these other forms are kinetically favored as supersaturation rises. Second, unlike most conventional crystals,



**Fig. 5** In this vapor diffusion droplet containing sodium nitrate and a trace amount of sodium chloride, the protein lysozyme has crystallized in two distinctly different crystal forms. The large cluster of thin laths is the monoclinic crystal form, while the many smaller, darker crystals are of tetragonal symmetry

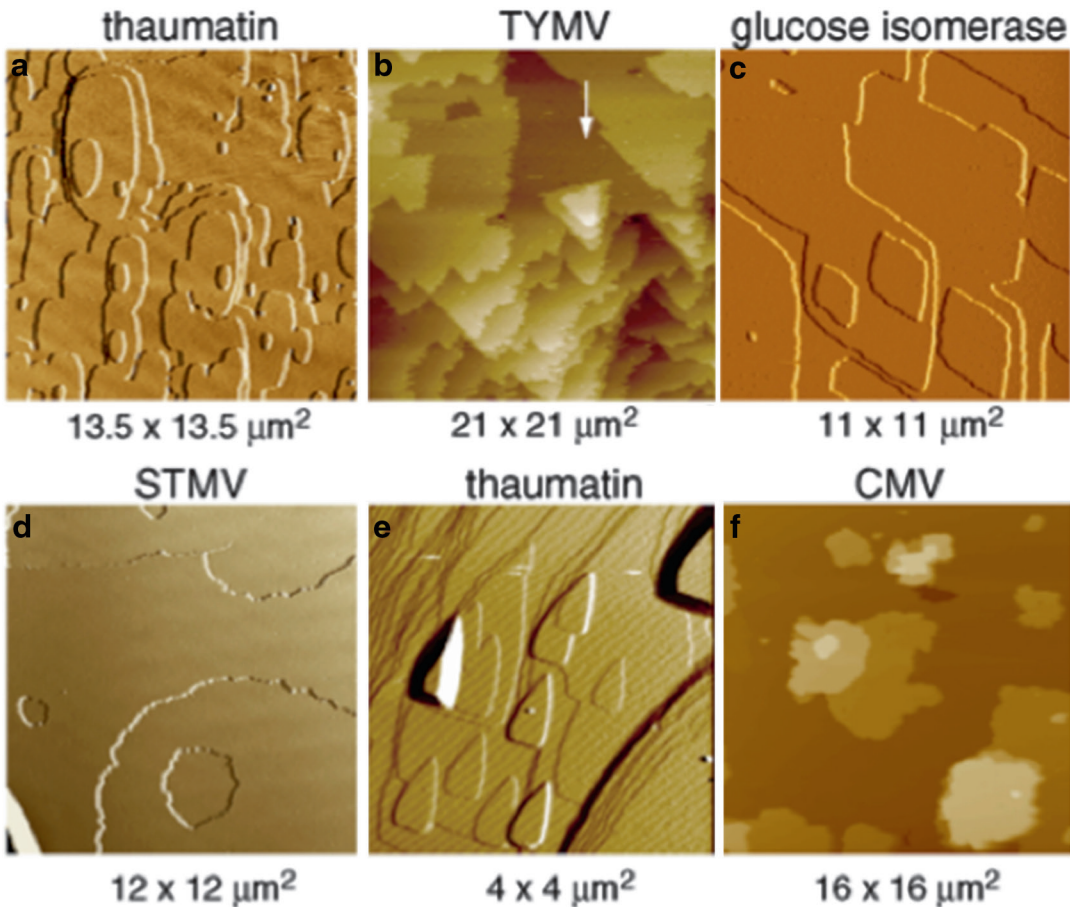
protein crystals nucleate, or initiate development, only at very high levels of supersaturation, often two to three orders of magnitude more than required to sustain crystal growth. This often leads to massive showers of microcrystals, or even more often, precipitate. Further, the kinetics of macromolecular crystal nucleation and growth are generally two to three orders of magnitude slower than for conventional molecules [19–23]. The latter difference arises from the considerably larger size, lowered diffusivity, and weaker association tendencies compared with small molecules or ions, as well as a lower overall probability of incorporation of an incoming protein molecule into a growth step [24].

Crystals, including protein crystals, grow by successive layer addition [18, 22, 23, 25]. The rate limiting step in crystal growth is not, however, the completion of an active layer by recruitment of molecules from solution into growth steps and kinks at the edges of expanding layers (referred to as tangential growth), as this is energetically favorable and rapid [25, 26]. The rate limiting step in crystal growth is the initiation of new, superior growth layers. This is more demanding in terms of self organization, less probable, and by far the slower process. It is referred to as growth in the normal direction.

There are four mechanisms that have been described for protein crystals to provide growth in the normal direction. These were deduced by application of AFM to actively growing crystals [27].



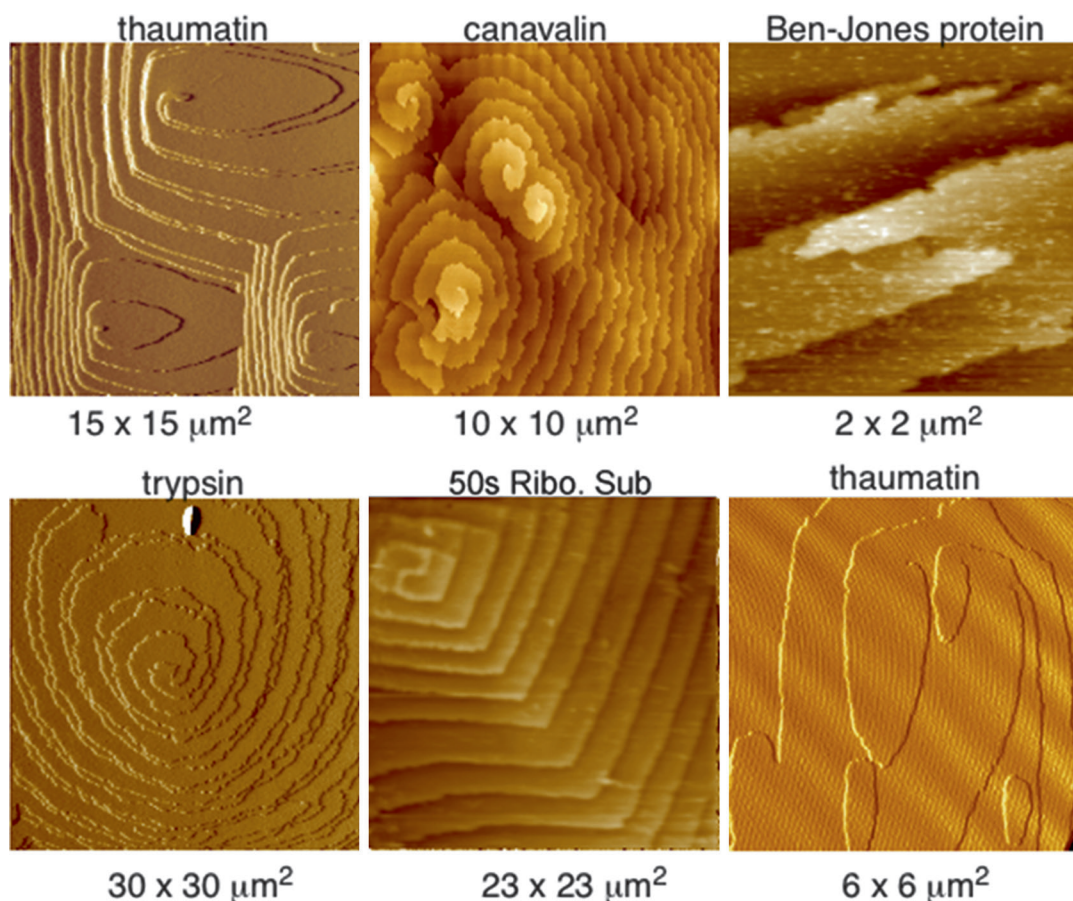
The various mechanisms have been treated in detail elsewhere [23, 27–30], but two of the four predominate. At higher levels of supersaturation the principal mechanism, illustrated by the examples in Fig. 6, is layer initiation by two dimensional nucleation on existing crystal surfaces. Like the formation of a three dimensional critical nucleus (*see* below), this requires molecules otherwise free in solution to self organize on a surface to form a small, crystallographically ordered array that may then expand by tangential growth. The major difference between the formation of three and two dimensional nuclei is that in the latter case the molecules are confined to a surface. This restriction of their freedom encourages their association. In addition, their self organization is guided in an epitaxial manner by the molecular lattice of the underlying, existing layer.



**Fig. 6** A major source of growth steps and layers on the surfaces of growing macromolecular crystals, particularly at medium to high levels of supersaturation, are two dimensional nuclei that exceed critical nuclear size and subsequently develop into two dimensional islands. Shown here are two-dimensional islands on a variety of protein and virus crystals. This is the dominant mechanism for face normal growth for most macromolecular crystals. In (b) the *arrow* denotes a triangular nucleus that reflects the symmetry of the crystal face

The other important mechanism is nucleation of new layers in a continuous manner through the occurrence and activity of screw, or spiral dislocations. Examples of such dislocations on a variety of protein crystals are shown in Fig. 7. They appear as both left and right handed spirals, as simple and compound screws, and they exhibit a variety of appearances dependent on the symmetry of the crystal and various kinetic factors. Because they do not require the improbable ordering of free molecules from solution, screw dislocations produce new layers even at low supersaturation. Together the two mechanisms of two dimensional nucleation and screw dislocation growth account for virtually all protein crystal growth.

Relevant to the practice of crystallization, the specific operable growth mechanism is principally determined by the crystallization conditions and the degree of supersaturation they produce. Often one mechanism may supersede another as supersaturation changes, and occasionally multiple mechanisms may operate simultaneously



**Fig. 7** A major source of growth steps on growing crystals, particularly at lower supersaturation, are screw, or spiral dislocations. Shown here are a variety of screw dislocations on the surfaces of macromolecular crystals that illustrates their diverse character

[16, 29]. The mechanisms of growth are further important because they may determine the amount and distribution of impurities incorporated in the crystal, the crystal defect structure, the ultimate size, and even certain diffraction properties such as mosaicity and resolution limit. It is important in practice to be aware that physical perturbations, such as vibrations, jarring, or temperature fluctuations can disrupt a growth mechanism or produce a shift from one mechanism to another.

---

## 4 Supersaturation, Nucleation, and Growth

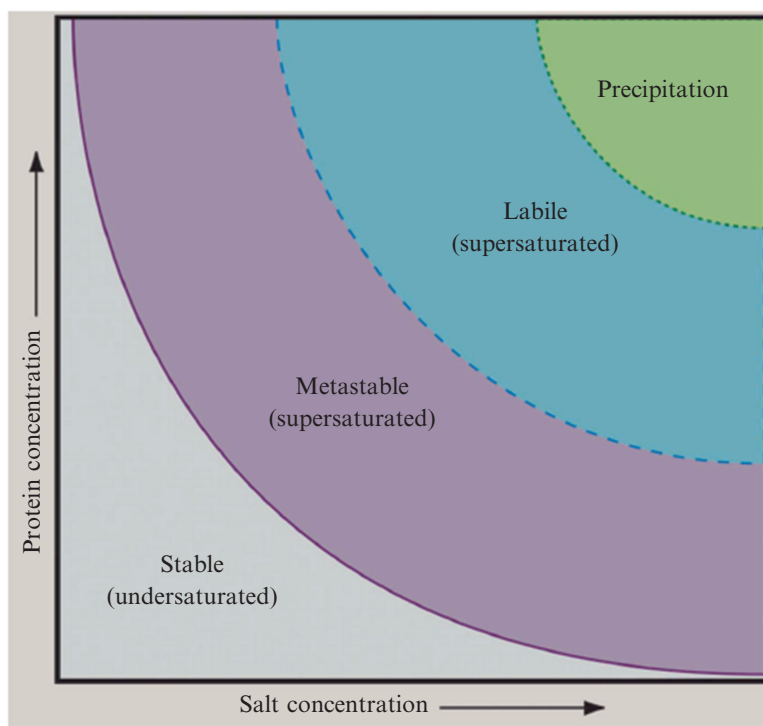
Crystallization of any molecule, or any chemical species including proteins, proceeds in two distinct but inseparable steps, nucleation and growth. Nucleation is the most difficult problem to address both theoretically and experimentally because it represents a mysterious first order phase transition by which molecules pass from a wholly disordered state to an ordered one. We believe that this likely occurs through the initial formation of partially ordered, or paracrystalline intermediates, protein aggregates having only short-range order, that through internal rearrangement ultimately yield small, crystallographically ordered assemblies that we refer to as critical nuclei [16, 29, 31]. A critical nucleus is an ordered cluster of molecules that is of sufficient size (has a surface to volume ratio) such that it acquires new molecules at a rate greater than that of losing molecules.

The size, or number of molecules making up a critical nucleus is dependent on the molecular dimensions, the extent of supersaturation, and the surface free energy of molecular addition. Currently the critical nuclear size has only been described for a few proteins, and for some cases, these were only investigated in terms of two-dimensional nuclei developing on the surfaces of already existent crystals [20, 21]. Recently, a theory has emerged which attempts to explain the nucleation phenomenon in terms of statistical fluctuations in solution properties [32–34]. 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. The hypothesis is supported by observations, using both atomic force microscopy and quasi-elastic light scattering, of a third mode of crystal growth in the normal direction termed growth by three dimensional nucleation [27, 31, 35].

Growth of macromolecular crystals is a better-characterized process than nucleation, and its mechanisms are reasonably well understood. As described above, 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 [27, 29, 31]. A common feature of nucleation and

growth is that both are critically dependent on the supersaturation of the mother liquor giving rise to the crystals. Supersaturation is the variable that drives both processes and determines their occurrence, extent, and the kinetics that govern them.

Crystallization of any molecules, including proteins, absolutely requires the creation of a supersaturated state. This is best illustrated by the phase diagram for crystallization shown in Fig. 8. Supersaturation is a nonequilibrium 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 reestablished by formation and development of a solid state, such as crystals, as the system returns to the saturation limit. To produce a supersaturated state, the properties of an undersaturated, or saturated solution must be



**Fig. 8** 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 existence of solid phase and 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 that result might be most probable



modified to reduce the ability of the medium to sustain the solubility of the protein (i.e., reduce the chemical activity of the solvent). Alternatively, some property of the protein molecules must be altered to reduce protein solubility and/or increase the attraction of one protein molecule for another, thereby inducing association. In any case, relationships between solvent and solute, or between the molecules in solution, are perturbed so as to promote formation of the solid state.

If no crystals or other condensed phase is present as conditions are changed, then solute will not immediately produce a new phase, and the solution will enter and remain in the supersaturated state. The solid state, hopefully a crystal nucleus, but other incipient states as well, does not develop spontaneously as the saturation limit is exceeded. Energy, analogous to the activation energy of a chemical reaction, is required to initiate the second phase, the stable critical nucleus of a crystal, or perhaps an unfortunate precipitate. Thus, a kinetic, or energy (or probability) barrier allows conditions with time to proceed more distant from equilibrium and further into the zone of supersaturation. Once a critical nucleus does appear in a supersaturated solution, however, it will proceed to accumulate molecules from solution and grow until the system regains equilibrium at saturation. So long as nonequilibrium forces prevail and some degree of supersaturation exists to drive events, a crystal will grow or precipitate continue to form.

---

## 5 General Approach

Protein crystallization is based on a diverse set of principles, unique experiences and evolving ideas. There is no comprehensive theory, or even an organized, extensive base of fundamental data to guide an investigator, though that is an effort in progress. As a consequence, protein crystal growth is largely empirical in nature, and demands patience, perseverance and intuition.

What complicates the crystallization process, in addition to our limited understanding of the phenomena involved, is the intimidating 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 hundreds of bonds and thousands of degrees of freedom. For viruses of molecular weights measured in the millions of Daltons, and for large multi-protein complexes, the possibilities for conformation, interaction, and dynamics are almost unlimited.

We are, however, beginning to develop rational approaches to protein crystallization based on an understanding of the fundamental properties of the systems. We are now increasingly using, in a systematic manner, the classical methods of physical chemistry to determine the energetic and kinetic characteristics of the

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, nevertheless, still rely primarily on what is fundamentally a trial and error approach. Protein crystallization is generally a matter of searching, as systematically and intelligently 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, even of poor quality, 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 what information is obtained to improve conditions in successive rounds of trials. Because the number of variables is so large, and the ranges so broad, experience and insight in designing and evaluating the individual and collective trials becomes an important consideration.

---

## 6 Screening for Initial Crystallization Conditions

As noted above, there are usually two phases in the creation of protein crystals for an X-ray diffraction investigation, and these are (a) the identification of chemical, biochemical, and physical conditions that yield a crystalline material, though it may initially be inadequate for X-ray analysis, and (b) the systematic alteration of those initial conditions by incremental amounts to obtain optimal samples for diffraction. The first of these is fraught with the greater risk, as some proteins simply refuse to form crystals, and clues as to why are elusive or absent. Optimization, however, often proves the more demanding of effort, more time consuming, and frustrating.

There are two fundamental approaches to searching for crystallization conditions. The first is a systematic variation of what are believed to be the most important variables, i.e., precipitant type (salt, polymer, organic liquid) and its concentration, pH, temperature, protein concentration, and potential ligands. 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 more effort and a greater amount of protein. In those cases where the quantity of material is limiting, it may simply be impractical. The second technique, however, provides more opportunity for useful conditions to escape discovery. In general, though, it requires less precious material.

The second approach also has, presently at least, one other major advantage, and that is availability and convenience. There is currently on the commercial market, from numerous companies, a wide variety of crystallization screening kits. The availability and ease of use of these relatively inexpensive 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 most popular approach for attacking, at least initially, 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 good fortune to smile. Occasionally it does, but sometimes not, and that is when the crystal grower must begin using his own intelligence to diagnose problems and devise remedies.

Once some crystals, even if only microcrystals, are observed and shown to be of protein origin, 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 might have an impact on the quality of the results (*see* below). Each of these parameters or factors is then carefully incremented in additional trial matrices that encompass ranges spanning the conditions which gave the "hit." Because the problem is nonlinear, that is, one variable may be coupled to another, this process is often more complex and difficult than one might anticipate [15, 16, 36–39]. It is here that the amount of protein and the limits of the investigator's patience may prove a formidable constraint.

---

## 7 Creating Supersaturation in Practice

In practice, one begins with a solution, a potential mother liquor, which contains some concentration of the protein below its solubility limit, or alternatively at its solubility maximum (an exception being the batch method, *see* below). The objective is then to gradually alter conditions so that the solubility of the protein in the sample is significantly reduced, thereby rendering the solution supersaturated. This may be done through a variety of approaches. Principally, these depend upon (a) altering the protein itself (e.g., by change of pH, which alters the ionization state of surface amino acid residues, by binding a ligand, or by introducing mutations), (b) altering the chemical activity of the water (e.g., by addition of salt or organic solvent), (c) altering the degree of attraction of one protein molecule for another (e.g., addition of bridging ions or molecules), or (d) altering the nature of the interactions between the protein molecules and the solvent (e.g., addition of polymers such as PEG), which also reduces the chemical activity of water.

Table 1 is a compilation of approaches upon which one might develop strategies for crystallizing a protein for the first time. Indeed, there are doubtless others that hopefully emerge from the imagination and cunningness of the investigator. The details of the various approaches have been described elsewhere [15, 36, 38, 39, 41] and need receive no extensive treatment here. It is probably sufficient to

**Table 1**  
**Approaches to creating supersaturation**

1. <i>Direct Mixing of Protein and Precipitant:</i> A protein and precipitant solution are thoroughly mixed so that the final solution is immediately supersaturated in protein. This relies on the energy, or probability barrier to critical nucleus formation to restrain the system and limit the number of nuclei and the time of their appearance. The most common application is in microdrops under oil
2. <i>Temperature Alteration:</i> Refers to a raising or lowering of the temperature of a protein–precipitant solution that is very near supersaturation. The temperature change is made in a direction that reduces the solubility of the protein. Most proteins in high salt solutions are more soluble at cold temperature, while protein–PEG combinations and low ionic strength solutions of protein are generally more soluble at warm temperatures
3. <i>Alteration of Ionic Strength:</i> Salt is added to a protein solution to high concentration so that competition for water lowers the solubility of the protein, referred to as “salting out.” Salt ions can also be removed by dialysis to create a low ionic strength state where, because of deprivation of cations, the protein is less soluble, referred to as “salting in.” See the phase diagram in Fig. 8 and the illustration in Fig. 10
4. <i>pH Alteration:</i> As the pH of a protein solution is changed, certain amino acid side chains on the protein molecules’ surfaces alter their ionization, and therefore their charge state. As a consequence the electrostatic surface of the protein molecules change. If this produces charge complementary surfaces and additional favorable interactions, or removes unfavorable interactions, then the molecules will be encouraged to associate and the solubility of the protein will be reduced. See Fig. 11
5. <i>Ligand Binding:</i> The solubility of most proteins, due to both long range and local conformational changes, may be altered as a consequence of ligand binding. The ligands may be coenzymes or other prosthetic groups, inhibitors, or ions. The last of these, particularly divalent cations, can also form bridges between otherwise repulsive groups on protein molecules and transform unfavorable interactions into geometry specific, favorable interactions
6. <i>Alteration of the Dielectric Constant of the Medium:</i> This is usually effected by the direct addition, or addition by dialysis, of an organic liquid of low dielectric constant into the protein solution. This encourages electrostatic and hydrogen bonding interactions between macromolecules
7. <i>Direct Removal of Water:</i> This can be brought about by simple evaporation or by concentration of the protein that reduces the water molecules available for solvation of the protein. Any method that produces dehydration of the protein falls in this category
8. <i>Addition of a Polymer:</i> PEG is most commonly used as a polymeric precipitant and it is hypothesized to act principally through the mechanism of “volume exclusion.” Because of its very large hydrodynamic radius, the disordered polymer restricts the volume of solvent that the protein can access, essentially depriving it of solvating water molecules. It effectively concentrates and dehydrates the protein and thereby reduces its solubility. There is a possibility that PEG may also act as an adhesive intermediary between protein molecules to enhance their association
9. <i>Removal of a Solubilizing Agent:</i> Some proteins can be concentrated to an enhanced degree by inclusion in the solution of some agent that increases its solubility such as a chaotrope or osmolyte [40]. Removal of the agent after concentrating then leaves the protein at reduced solubility and perhaps at supersaturation. The agent may be removed by dialysis
10. <i>Addition of Non-volatile Alcohols and Low Molecular Weight Polymers:</i> Liquid compounds such as MPD, hexanediol, PEGs 200 Da, 400 Da, and low molecular weight Jeffamines reduce the solubility of proteins, probably by competing for water and altering dielectric constants, but their true mechanism remains obscure. They may also incorporate into crystals and favorably alter the solvent structure inside the crystals



say that if a protein has a propensity to crystallize, it can probably be accomplished by variation of precipitant type, precipitant concentration, pH, to a lesser extent 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.

---

## 8 Methodology

The growth of protein crystals must be carried out in some physical apparatus that allows the investigator to reduce the solubility of the protein by altering the properties of the mother liquor, using, perhaps, one of the strategies in Table 1. Currently, these involve, almost exclusively, microtechniques. Crystallization “trials” with a matrix of 48 or 96 conditions may be carried out with volumes of only a fraction of a milliliter if done manually, a few microliters or less with some robotic or microfluidic systems. These employ plastic, multichambered trays for hanging and sitting drops, plexiglass buttons for dialysis, or microdrops under oil. Other methods are found in Table 2.

Crystallization devices and the associated methodologies have also been described in detail elsewhere [14, 36, 39, 45]. Detailed instructions and web sites are frequently provided by the manufacturers of the crystallization kits, supplies, and plasticware, along with many helpful illustrations. The hanging drop and sitting drop procedures for vapor diffusion, and the batch method using microdrops under oil are now most in favor, and are recommended for most investigations. In those cases where mother liquor components cannot be transported through the vapor phase (e.g., metal ions, detergents) then microdialysis 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.

Vapor diffusion, in either the “sitting drop” or “hanging drop” arrangements is the most popular approach. This is illustrated for both arrangements in Fig. 9a, b. Vapor diffusion relies on the equilibration of a small droplet, 1–10  $\mu$ l in volume usually, against a larger liquid reservoir of 0.5–1.0 ml. The droplet is initially a mixture, most commonly 1:1, of a stock protein solution at 10–30 mg/ml, with the reservoir that may contain, for example, a buffered, concentrated salt or PEG solution. Loss of water from the droplet to the reservoir and equilibration of the two over time, hours to days, restores the droplet (almost) to the concentration of

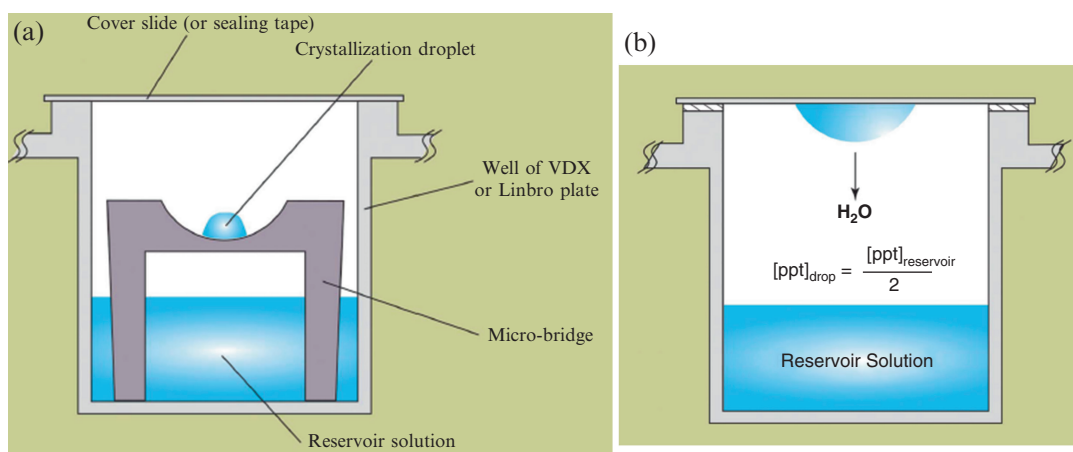
**Table 2**  
**Physical and chemical procedures for producing solubility minima**

1. <i>Bulk (visual) Crystallization:</i> This is a term used by old biochemists to describe the direct addition of salt to a protein solution while watching intently for the appearance of an opalescent sheen (Tyndall effect) that indicates incipient crystallization. This was a skill acquired by researchers for about a century when crystallization was principally used as a powerful purification tool, and ultimately as a demonstration of protein purity
2. <i>Batch Method in Vials:</i> Once a precipitant concentration that produces crystals is identified for a particular protein, then a protein solution is simply mixed with the precipitant at 1–3% less. The protein–salt solution is then dispersed into small screw cap vials and allowed to stand. Very slow evaporation around the caps causes the precipitant concentration to gradually increase in the vials until supersaturation is achieved. This was the most widely used technique until about 1970
3. <i>Evaporation:</i> Probably the oldest method in existence, it was used to produce the first reported protein crystals, those of earthworm hemoglobin [42]. Very slow and controlled evaporation may still be used and effected through the use of narrow capillaries or wicks
4. <i>Dialysis and Microdialysis:</i> This procedure can be performed in bulk using dialysis tubing, or on a microscale using capillaries or dialysis buttons enclosed by a dialysis membrane. Dialysis has the great advantage that it allows investigation of a continuum of precipitant concentrations or a range of pH. It permits the introduction or removal of ligands, coenzymes, inhibitors or ions. It can be used with the same protein sample to carry out multiple, independent experiments simply by changing the exterior solution
5. <i>Concentration Dialysis:</i> Dialysis as described above in (4) but dissolving in the outside solution a high concentration of high molecular weight PEG (PEG 20,000 for example). Water is withdrawn as dialysis proceeds and the protein becomes increasingly concentrated as conditions are altered. Apparatus was once available for performing the concentration function by drawing a vacuum on the system simultaneous with dialysis, in which case no PEG was required
6. <i>Liquid Bridge:</i> A kind of direct, but slow mixing of a protein solution with a precipitant solution. Drops of each are placed in a sealed container maintained at 100% humidity, and a needle is used to draw out a thin connecting liquid bridge between the two drops. Protein diffuses very slowly, thus it is precipitant that gradually diffuses into the protein drop and promotes supersaturation
7. <i>Free Interface Diffusion:</i> In a tube or capillary, a lighter protein solution is gently layered upon a heavier precipitant solution (or vice versa depending on relative densities). Slow diffusion along with some convection across the interface produces a diversity of local concentration gradients that may promote crystal nucleation and support crystal growth. This technique has been particularly useful in microgravity where pure diffusive transport prevails and convection is absent
8. <i>Vapor Diffusion:</i> This refers to any arrangement, microdroplets “sitting” on a plastic support, for example, or “hanging” from a glass cover slip, equilibrating through the vapor phase with a larger volume reservoir solution containing a higher concentration of precipitant. Over time the osmolarity of the drop asymptotically approaches that of the reservoir because of water loss from the protein–precipitant drop. This approach in one manifestation or another is currently in widest and most popular use.
9. <i>Sequential Extraction:</i> This method [43] is primarily used to produce microcrystals to be later used for seeding. It depends on the sequential extraction of a salt induced, protein precipitate (centrifuge pellet) by solutions of decreasing precipitant concentration at 4 °C. The drops of extract are subsequently placed at 25 °C where the protein (in salt solution) is less soluble and supersaturation is achieved

(continued)

**Table 2**  
(continued)

10. <i>pH Induced Crystallization</i> : This is a powerful approach and can be accomplished through direct addition of acid or base, by dialysis, or by vapor diffusion. This takes advantage of the frequent strong dependence of protein solubility as a function of $\text{H}_3\text{O}^+$ concentration.
11. <i>Temperature Induced Crystallization</i> : This method takes advantage of the difference in solubility of some proteins as a function of temperature within the range of 0–37 °C. For most, but not all proteins, this dependence is rather weak so that the technique is used infrequently with proteins. It is extremely important in the crystallization of conventional molecules
12. <i>Effector Addition</i> : This depends on the difference in solubility of a protein when it has a coenzyme, inhibitor, or other ligand bound to it. Dialysis or direct addition can be used to introduce or remove a ligand and thereby affect protein solubility
13. <i>Crystallization in Gels</i> : Diffusion through gels, such as silica or agarose gels, of a mobile precipitant into a protein containing gel, essentially free interface diffusion in a gel, can be used to produce the supersaturated state. Currently in popular use for the crystallization of membrane and lipophilic proteins, the “lipidic cubic phase” [44] for crystallization takes advantage of the complex structure of the gel (mesophase) itself to induce nucleation and allow controlled growth



**Fig. 9** The sitting drop vapor diffusion method is illustrated in this schematic diagram (a). The drop on the elevated platform, which is commonly 2–10  $\mu\text{l}$ , consists of half stock protein solution and half the 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. The hanging drop vapor diffusion method is illustrated schematically in (b). 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

the protein in the stock solution, and increases the precipitant concentration to that of the reservoir. Hopefully this process produces a droplet supersaturated in protein.

Though proportions and volumes are different, dialysis has the same objective, to gradually raise the salt or PEG concentration of a protein solution to the point where it becomes saturated in protein. Drops under oil [46], or the batch method as it is called, uses no equilibration. A protein solution and a potential crystallization promoting solution (salt, PEG, buffer) are simply mixed in some reasonable proportion and droplets dispersed under oil. This method relies on the nucleation energy barrier to allow immediate establishment of supersaturated drops. Other techniques, such as free interface diffusion [47], slow diffusion and mixing of protein and precipitant solutions across a liquid–liquid interface, are also in use, but see less application than batch or vapor diffusion.

In high throughput laboratories, screening for crystallization conditions, and even optimization in some cases, has generally been consigned to robotic devices [48–50]. This is particularly true in those of large pharmaceutical companies where many proteins may be under simultaneous structural investigation. Automated systems have the advantages of exceptional record maintenance, most can deploy sub microliter 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 using manual techniques. Robotic systems are, in addition, now being used to examine and evaluate the results of crystallization trials using optical subsystems and image processing techniques [51–53]. Evaluation of trial arrays of conditions, however, continues to be problematic because of the continuing difficulty in devising meaningful criteria for progress 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 murky indications of how near the conditions were to a successful mother liquor.

---

## 9 Precipitants

One of the most important components in a mother liquor intended to crystallize a protein is sometimes called the precipitant, or other times the crystallization agent. It is generally, but not always, the chemical component that reduces the solubility of the protein or reduces the chemical activity of water. Salts, such as ammonium sulfate or potassium phosphate, or polymers such as PEG are classic examples. The mechanisms by which they act (*see* below) may be different, but their essential role is to deprive the protein of solvating water and to promote protein association.

If one were to examine the reagents utilized in any of the commercial crystallization screens that are based on shotgun approaches,



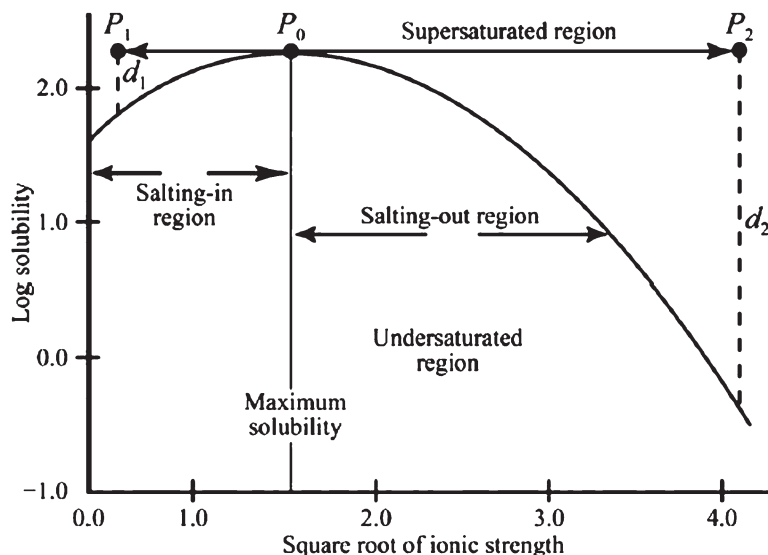
or examined the crystallization conditions that have been compiled into data bases [54, 55], then it would become apparent that a wide variety of precipitating (crystallizing) agents have been used. Indeed many agents have been employed, and some, such as ammonium sulfate and polyethylene glycol have produced a great number of successes. It is often necessary, however, to explore many precipitants, and it is difficult to know initially which might offer the greatest likelihood of obtaining crystals.

Individual precipitants and their properties have also been reviewed in some detail [16] and are not extensively discussed here. To summarize, however, it is possible to group the precipitants into categories based on their mechanisms for promoting crystallization. The majority of precipitants of proteins fall into four broad categories (1) salts, (2) organic solvents, (3) long chain polymers, and (4) low molecular weight polymers and non-volatile alcohols. 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. In the fourth category we might place compounds such as methylpentanediol (MPD) and polyethylene glycols of molecular weight less than about 1000.

The solubility of proteins in concentrated salt solutions is complicated, but it can be viewed naively as a competition between salt ions, principally the anions, and the protein for the binding of water molecules, which are essential for the maintenance of solubility [56–60]. At sufficiently high salt concentrations the protein molecules become so uncomfortably deprived of solvent that they seek association with one another in order to satisfy their electrostatic and amphipathic requirements. In this environment large, semi ordered aggregates that could lead to critical crystal nuclei, as well as disordered amorphous precipitate may form. Other salt ions, chiefly cations, also may be necessary to insure protein solubility. At low ionic strengths, cation availability may be insufficient to maintain protein solubility, and under those conditions too, crystals may form. The behavior of typical proteins over the entire range of salt concentrations, including both the “salting in” and “salting out” regions is illustrated by Fig. 10.

Salts exert their effect principally by dehydrating proteins through competition for water molecules, and a measure of their efficiency in this is the ionic strength, whose value is the product of the molarity of each ion in solution with the square of their valences. Thus, multivalent ions are the most efficient precipitants. Sulfates, phosphates, citrates, and more recently malonates [61] and mixtures of the salts of dicarboxylic acids have traditionally been employed.

One might anticipate little variation among different salts so long as the valences of their ions were the same. Thus there should be little expected variation between two different sulfates such as  $\text{Li}_2\text{SO}_4$  and  $(\text{NH}_4)_2\text{SO}_4$  if only ionic strength were involved. This is often observed not to be the case. In addition to salting out, which



**Fig. 10** The curve shown here represents a typical solubility curve for a protein and divides the region of undersaturation from that of supersaturation. It also illustrates the existence of the classical “salting in” and a “salting out” region for the protein. By taking advantage of the 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

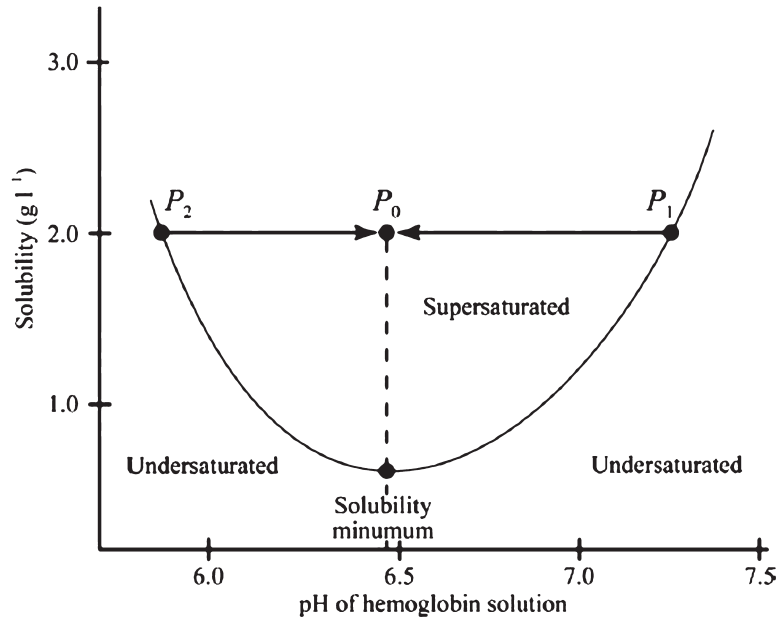
is a general dehydration effect not really much different than evaporation or concentration (except that water is not physically removed) there are also specific protein–ion interactions that may have further consequences [58, 60]. This is perhaps not unexpected given the varied hydration properties of different ions and the unique polyvalent character of individual proteins, protein structural and dynamic 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 without some prior knowledge of its solubility behavior. In general, however, it is a concentration of the precipitant just a few percent less than that which yields an amorphous precipitate [62], and this can be determined for a macromolecule under a given set of conditions using only minute amounts of material [15]. To determine the approximate insolubility points with a particular precipitant a 10  $\mu$ l droplet of a 5–15 mg/ml protein solution can be placed in the well of a depression slide and observed under a low-power light microscope as increasing

amounts of saturated salt solution or organic solvent (in 1- or 2- $\mu$ l increments) are added. If the well is sealed between additions with a coverslip, the increases can be made over a period of many hours.

Along with precipitant type and concentration, pH is usually the most important variable influencing the solubility of proteins. As such, it provides a 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 foundational concept in formulating screening matrices and discovering successful crystallization conditions. An example of how pH might be used to effect crystallization of a protein is illustrated in Fig. 11.

Organic solvents reduce the dielectric constant of the medium, hence the screening of the electric fields that mediate macromolecular interactions in solution. A danger, however, is that they also tend to destabilize protein structure. As the concentration of organic solvent is increased, interaction between protein molecules increases, solvent becomes less effective (the activity coefficient of water is reduced) and the solid state becomes more favored [63, 64]. Organic solvents should be used at low temperature, at or below 0 °C, and they should be added very slowly with good mixing [16]. Since they are usually volatile, vapor diffusion techniques are equally applicable. Ionic strength should, in general, be maintained



**Fig. 11** As shown here, most proteins have specific 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

low and whatever means are otherwise available should be pursued to protect against denaturation.

Some polymers, among which polyethylene glycols (PEG) are most popular [65, 66], produce volume exclusion effects that induce separation of proteins from solution [65, 67]. Polymeric precipitants, unlike proteins, have no consistent, fixed conformation. They writhe and twist randomly in solution and, as a consequence, occupy far more space than their molecular weights would suggest. This effect, referred to as volume exclusion, results in less solvent available space for the protein molecules that then segregate, aggregate, and ultimately form a solid state, in favorable cases crystals. PEG is also extremely hydrophilic and binds water molecules avidly (about 2.3 water molecules per monomeric unit [68, 69]). It, like salts, competes for solvent, thereby dehydrating the protein molecules.

Evidence has emerged recently that suggests PEG, and some related polymeric precipitants, may not exert their effects on protein crystallization exclusively by the mechanism of volume exclusion and dehydration. Some observations indicate that PEG, at least fragments and lower molecular weight components (all PEG preparations exhibit a distribution of lengths about a mean) may actually co-crystallize with the protein due to positive, associative interactions [68, 69], and thus occupy interstitial spaces and channels otherwise filled by solvent alone. Inside the crystal, PEG likely remains disordered, or at best partially ordered, and thereby escapes detection by X-ray analysis. If this PEG incorporation is valid, then it has important ramifications, as PEG could well influence protein association and crystallization by both altering interstitial water structure, and possibly by providing a soft superstructure that helps guide crystal growth. More remains to be done to test this intriguing idea.

A large number of protein structures have now been solved using crystals grown from polyethylene glycol. These confirm that the protein molecules are in as native 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.

PEG sizes from  $M_r = 200$  to 20,000 Da have successfully provided protein crystals, but the most useful seem to be those in the range 2000–8000 Da. A large number of reports have appeared, however, in which a protein could not easily be crystallized using this range but yielded in the presence of PEG 200 Da or 20,000 Da. The molecular weight sizes may not be completely interchangeable for a given protein even within the mid range. Some produce the best-formed and largest crystals only at, say,  $M_r = 4000$  Da, and



less perfect examples at other weights. This is a parameter that is best optimized by empirical means along with concentration. The very low molecular weight PEGs such as 200 and 400 Da are somewhat 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.

An advantage of PEG over most other precipitating agents is that proteins crystallize within a fairly narrow range of PEG concentration; this being from about 4% to 20% (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 (in some cases even more), 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 advantage of PEG, then, 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, free interface diffusion, or vapor equilibration. When the reservoir concentration of PEG is in the range of 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 5–10% below the desired final value.

---

## 10 Factors Affecting Crystallization

There are many factors that can affect the crystallization of proteins and these, too, have been reviewed elsewhere [16, 36, 38, 39]. They fall into categories of physical factors such as temperature, chemical influences such as pH or ionic strength, and biochemical factors that include, among many others, purity and monodispersity. Any one, or any combination may affect the likelihood of crystallization occurring at all, the nucleation probability and rate, crystal growth rate and mechanism, and the ultimate sizes and quality of the products. As noted above, pH and the concentrations of salt and other precipitants are virtually always of importance. The concentration of the protein, which may vary from as low as 2 mg/ml for viruses and large complexes [70], to as much as a hundred mg/ml for some highly soluble proteins, is an additional, significant variable.

Other parameters may be unimportant in some cases but play a crucial role in others. In particular the presence or absence of ligands, coenzymes, or inhibitors, the variety of salt or buffer, the equilibration technique used, temperature fluctuations, and the presence of detergents and chaotropes [40] are all pertinent factors. Parameters of somewhat lesser and largely obscure significance are things like gravity, electric and magnetic fields, or viscosity. It can not be predicted which of these many variables may be of importance for a particular macromolecule, and the influence of any one must, in general, be investigated through empirical trials.

An intriguing problem, or opportunity depending on one's perspective, is what additional components or compounds should be included in the mother liquor in addition to solvent, buffer, protein, and precipitating agent [16, 36, 40, 71, 72]. The most desirable effectors, it would seem, are those which maintain the protein in a single, homogeneous, and invariant state. Reducing agents such as glutathione,  $\beta$ -mercaptoethanol, and dithiothreitol are useful to preserve sulfhydryl groups and prevent oxidation. EDTA and EGTA are effective if one wishes to protect the protein from heavy or transition metal ions. Inclusion of these components may be particularly important when crystallization requires a long period of time to reach completion. When crystallization is carried out at room temperature in polyethylene glycol or low ionic strength solutions, then attention must be given to preventing the growth of microbes. These generally secrete proteolytic enzymes that may have serious effects on the integrity of the protein under study. Inclusion of sodium azide, thymol or chlorobutanol at low levels may be necessary to suppress invasive bacteria and fungi.

Substrates, coenzymes and inhibitors often serve to maintain an enzyme in a more compact and stable form. Thus a greater degree of structural homogeneity may be imposed on a population of protein molecules and a reduced level of statistical variation achieved by complexing the protein with a natural ligand before attempting its crystallization. In some cases an apoprotein and its ligand complexes may be significantly different in their physical behavior and can, in terms of crystallization, be treated as almost entirely separate problems. Complexes may provide additional opportunities for growing crystals if the native apoprotein is refractory. It is worthwhile, therefore, when searching for crystallization conditions, to explore complexes of the macromolecule with substrates, coenzymes, and inhibitors at an early stage. Such complexes are, in addition, often inherently more interesting in a biochemical sense than the apoprotein.

Various metal ions have occasionally been observed to promote the crystallization of proteins. Bacterial glucose isomerase, for example, crystallizes readily from PEG solutions in the presence of  $Mg^{++}$ , but only with difficulty in its absence. Cadmium and some other divalent cations induce immediate crystallization of the iron storage protein ferritin [73]. In some instances ions are essential for activity. It is, therefore, reasonable to expect that they might

aid in maintaining certain structural features of the molecule. There are other examples, however, where metal ions, particularly divalent metal ions of the transition series such as  $\text{Ca}^{++}$ , were found to encourage crystal growth but played no recognized role in the protein's activity or structure. They likely serve as bridging agents between molecules in the crystal lattice.

---

## 11 The Protein as a Variable

A factor of particular importance to crystallization is the homogeneity and monodispersity of the protein [74, 75] and this deserves special emphasis. Some proteins may crystallize even from very heterogeneous mixtures (egg albumin, lysozyme, canavalin,  $\alpha$ -amylase, for example), and indeed, crystallization has long been used as a powerful purification tool. It is the reason, in fact, why it originated as a technique and has been held in such high regard. In general, however, the likelihood of success in crystal growth is greatly advanced by increased homogeneity of the protein sample. Investment in further purification is always warranted, and usually profitable. When every effort to crystallize a protein fails, the best recourse is to further purify.

Recombinant DNA technology provided an enormous impetus to crystal growth research and X-ray crystallography 35 years ago, as it provided crystallographers access to proteins found in very low abundance that nevertheless played important roles in cells. Indeed it may be on the verge of providing another advance at this very time. Arguably, the most important parameter in protein crystallization is the protein itself. Until recently we have had little or no direct control over most of the important features of that parameter. Modification at the genetic level, however, provides us that opportunity, and its possibilities are only now being realized [76–78].

Through truncations, mutations, chimeric conjugates, and many other protein engineering contrivances, the probability of crystallization has been significantly enhanced. If we can learn how to go about this in a rational and systematic manner then advances may occur in future years that match the progress of the past. Approaches to application of mutation will be addressed and elaborated by others elsewhere in this volume.

---

## 12 Optimization of Crystallization Conditions

Optimization means adjusting the parameters of crystallization conditions, initially estimated from screening matrices [16, 37, 39], with the objective of discovering improved conditions that ultimately yield the best crystals for diffraction data collection. Optimization is in a sense refinement, but it is complicated somewhat because the parameters almost certainly are not independent of one another. They may

be linked or correlated. Furthermore, solubility diagrams, which would have many dimensions, do not exist for specific proteins. Every protein has a unique length and amino acid sequence, and a unique three-dimensional conformation. Every protein is an individual with its own eccentricities and peculiarities. A further complication is that there can be an “embarrassment of riches” where many “hits” are obtained initially and the question arises as to which deserve the effort required for further improvement.

Optimization, as it is often practiced is in principle relatively straightforward. The parameters that define the initial conditions are first identified (pH, precipitant type, precipitant concentration, temperature, ion concentration, etc.). Following this, solutions are made that incrementally and systematically vary the parameters about the initial values. That is, if the pH of the initial hit was 7.0, then the same mother liquor might be composed but at pH 6, 6.2, 6.4, 6.6, etc. up to pH 8.0. This does not guarantee that one will arrive at optimal conditions, parameters may be correlated, but it is the best approach that we have.

While simple in principle, optimization becomes demanding in the laboratory. First of all, the number of parameters or effecting conditions may be large [15, 16, 36, 37]. It may not be clear which parameters are actually important, or what the range for exploration should be. Thus we have as an initial goal of optimization to deduce what variables are relevant and how to prioritize each relative to another so that adjustments can be made, all the while minimizing or neglecting the least or irrelevant factors.

Optimization may require a substantial amount of protein sample, and this may be severely limited. Thus, efficiency and economy become essential, and the use of very small volume trials [48, 50, 52] will be tempting. Small volumes, however, should be treated with caution. One seldom obtains large crystals from nanoliter volumes of mother liquor, and when promising results from very small drops are scaled up to larger volumes to grow larger crystals (which larger volumes tend to yield) the increase in scale fails to materialize.

The greatest obstacle to success in optimization is most frequently an absence of sufficient commitment, or a lack of effort on the part of the investigator. Screening for new crystallization conditions can be made almost, but not quite, painless. Commercial kits can be purchased that contain precisely prepared solutions. Robotics are now employed to dispense samples into plates, further robotic devices categorize and store the plates, and automated photographic systems present images of the many drops for viewing [49, 50, 52].

Automated systems, however, cannot make optimization effortless, and that is because optimization requires composition of a vast number of solutions that must be formulated or purchased, and the use of robotics in optimization presents as many problems as it solves, at least at this point in time. Making up a myriad of solutions, adjusting their pHs to exact values, and so on is tedious.

In other words, doing a lot of basic laboratory chemistry demands a lot of hard labor. Many investigators would rather struggle with marginal, or even miserable crystals obtained from the first hit than undertake the optimization effort.

---

## 13 Membrane Proteins

Proteins that are naturally membrane associated, or that are otherwise unusually hydrophobic or lipophilic in nature present unique problems. Such proteins are, in general, only sparingly soluble in normal aqueous media, some virtually insoluble, others lose their active conformations, and this in turn makes the application of conventional protein crystallization techniques problematic. Problems are difficult but not intractable. To address these difficulties the use of detergents, particularly non-ionic detergents, has been developed [79–83]. No attempt will be made here to describe the various techniques or the combinations of detergents and accessory molecules that have been used, as that involves a number of complexities and considerations that are covered in another chapter.

An essential difficulty associated with 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 drop screen of conditions, at least initially, then the additional search for a useful detergent means that the 48 trial screen must be multiplied by the number of detergent candidates. A further problem is that there are a great number of potentially useful detergents. Hampton Research (Aliso Viejo, CA), a major source of screening reagents, offers three different detergent kits of 24 compounds each. Were one to simply apply a basic 48-well screen with each detergent, then that would require a total of 3456 individual trials. While this may actually be possible with highly automated, nanoscale systems, and where a substantial amount of material is available, it is impractical for most laboratories.

Basic crystal screens, whether they are systematic screens or shotgun screens, should not, however, be abandoned. It becomes essential though to reduce, at least initially, 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 well within the capabilities of most labs. No one, however, has yet reduced the set to a favored few. Everyone has an opinion as to which detergents should be favored, and no consensus has yet emerged from data bases and analyses of experiments. To make matters even more challenging, it appears that some, 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.



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. Reference 83 is a good review of workup until that time, and more recently, there are fine discourses by Loll [80], Caffrey [44], Garavito and Ferguson-Miller [84], Hunte, et al. [85], and Wiener [79, 86], as well as a chapter in this book.

## 14 Some Important Concepts

Although approaches to protein crystallization remain largely empirical, substantial progress has been made. 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 and their complex assemblies that affect their capacity to crystallize. Some principles now stand out regarding the crystallization problem, and these are summarized in Table 3.

**Table 3**  
**Some important concepts in protein crystallization**

1. <i>Protein Purity</i> —Crystallization occurs because a population of structurally and chemically homogeneous molecules are made amenable to the formation of periodic bonding arrangements. Molecular misfits create disruptions of order and inhibit critical nucleus formation and crystal growth. Efforts to make the most pure and uniform protein sample as possible are never wasted
2. <i>Solubility and Monodispersity</i> —High protein concentration generally means more reliable crystallization and a greater overall chance of success in initial screens, and this depends on solubility of the protein. Solubility, however, also implies protein monodispersity and the absence of arbitrary oligomers and aggregates in the sample that are little more than contaminants
3. <i>Stability</i> —A foundational concept in crystallization is the unchanging nature of the molecules with regard to conformation and physical–chemical properties. It is now a given that the more stable a protein, the more likely it is to crystallize. The investigator must do whatever possible to insure that the protein molecules remain in their native state
4. <i>Supersaturation</i> —This is the crucial, controlling factor in determining nucleation probability, and both the mechanisms and kinetics of crystal growth. It can be achieved in many ways, and the path by which it is reached is as important as the ultimate value. A solution supersaturated in protein is a physical necessity for crystallization
5. <i>Association</i> —Supersaturation can be reached in many cases by enhancing attractive, specific interactions between protein molecules and thereby reducing their solubility. Additives, ions, protein modifications are traditional approaches. Reducing the chemical activity of the solvent abets this process and is the mechanism by which most precipitating agents operate
6. <i>Nucleation</i> —This is essential to start the crystallization process, and it is largely dependent on probability. That in turn depends on the degree of supersaturation and the path (through the phase diagram) by which supersaturation is reached. Competition from other condensed phases, such as precipitate, is the primary adversary. Enough supersaturation is necessary; too much supersaturation is a damper.

(continued)

**Table 3**  
(continued)

7. <i>Variety</i> —Because of the stochastic elements involved in crystallization, chance is an important factor. The more chances one has, the more likely is success. Explore as many possibilities and opportunities as possible in terms of sample source, sample conformation, physical, chemical, and biochemical parameters
8. <i>Constancy</i> —Physical and/or chemical perturbations can inject energy into a dynamic, crystallizing system and cause deviations of otherwise ordered growth mechanisms. Disruptions from mechanical jarring, evaporation, or from temperature fluctuations can be devastating. Maintain the crystallizing samples at an optimal state during the full course
9. <i>Impurities</i> —The incorporation of impurities, not only molecules present in the protein sample, but in the reagents, apparatus, or from the environment can seriously contribute to unwanted nucleation and growth termination
10. <i>Preservation</i> —Crystals vary in their long term stability once they have reached terminal size. It may be necessary to take “post crystallization” measures to insure that the crystals maintain their quality until X-ray data collection can begin. These may include lowering temperature, increasing the precipitant concentration, prevention of evaporation through plastics, addition of stabilizers, cryo-vitrification, or mounting in sealed capillaries. Shock and handling must be avoided

## References

- Helliwell JR (1992) Macromolecular crystallography with synchrotron radiation. Cambridge University Press, Cambridge, UK
- Bingel-Erlenmeyer R, Olieric V, Grimshaw J et al (2011) SLS crystallization platform at Beamline X06DA-A fully automated pipeline enabling in situ X-ray diffraction screening. *Crystr Growth Des* 11:916–923
- Garman EF, Schneider TR (1997) Macromolecular cryocrystallography. *J Appl Crystallogr* 30:211–237
- Pflugrath JW (2004) Macromolecular cryocrystallography—methods for cooling and mounting protein crystals at cryogenic temperatures. *Methods* 34:415–423
- Pflugrath JW (2015) Practical macromolecular cryocrystallography. *Acta Crystallogr F Struct Biol Commun* 71:622–642
- Heras B, Edeling MA, Byriel KA et al (2003) Dehydration converts DsbG crystal diffraction from low to high resolution. *Structure* 11:139–145
- Kiefersauer R, Than ME (2000) A novel free-mounting system for protein crystals: transformation and improvement of diffraction power by accurately controlled humidity changes. *J Appl Crystallogr* 33:1223–1230
- Pflugrath JW (1992) Developments in X-ray detectors. *Curr Opin Struct Biol* 2:811–815
- Gruner SM, Eikenberry EF, Tate MW (2001) Comparison of X-ray detectors. In: Rossmann MG, Arnold E (eds) *International tables for crystallography*, vol F. Kluwer Academic Publishers, Dordrecht, pp 143–153
- Rossmann MG, Arnold E (2001) *Crystallography of biological macromolecules*, vol F. International tables for crystallography. Dordrecht, Kluwer Academic Publishers
- Malkin AJ, Kuznetsov YG, McPherson A (1996) Defect structure of macromolecular crystals. *J Struct Biol* 117:124–137
- Feigelson RS (1988) The relevance of small molecule crystal growth theories and techniques to the growth of biological macromolecules. *J Cryst Growth* 90:1–13
- Feher G (1986) Mechanisms of nucleation and growth of protein crystals. *J Cryst Growth* 76:545–546
- Durbin SD, Feher G (1996) Protein crystallization. *Annu Rev Phys Chem* 47:171–204
- McPherson A (1982) *The preparation and analysis of protein crystals*. Wiley, New York
- McPherson A (1999) *Crystallization of biological macromolecules*. Cold Spring Harbor Laboratory Press, Cold Spring Harbor, NY
- McPherson A, Malkin AJ, Kuznetsov YG (1995) The science of macromolecular crystallization. *Structure* 3(8):759–768

18. Rosenberger F (1986) Inorganic and protein crystal growth—similarities and differences. *J Cryst Growth* 76:618
19. Kuznetsov YG, Malkin AJ, Greenwood A et al (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
20. Malkin AJ, Kuznetsov YG, Glantz W et al (1996) Atomic force microscopy studies of surface morphology and growth kinetics in thaumatin crystallization. *J Phys Chem* 100:11736–11743
21. Malkin AJ, Kuznetsov YG, McPherson A (1997) An in situ investigation of catalase crystallization. *Surf Sci* 393:95–107
22. Chernov AA, Komatsu H (1995) Principles of crystal growth in protein crystallization. In: Bruinsma JPEOSL (ed) *Science and technology of crystal growth*. Kluwer, Dordrecht, The Netherlands, p 67
23. Vekilov PG, Chernov AA (2002) The physics of protein crystallization. *Solid State Phys* 57:2–147
24. Chernov AA (2003) Protein crystals and their growth. *J Struct Biol* 142:3–21
25. Rosenberger A (1979) *Fundamentals of crystal growth*. Springer-Verlag, Berlin
26. Chernov AA (1984) *Modern crystallography III*. Crystal growth. Springer-Verlag, Berlin.
27. Malkin AJ, Kuznetsov YG, Land TA et al (1995) Mechanisms of growth for protein and virus crystals. *Nat Struct Biol* 2:956–959
28. McPherson A, Malkin AJ, Kuznetsov YG et al (2001) Atomic force microscopy applications in macromolecular crystallography. *Acta Crystallogr D Biol Crystallogr* 57:1053–1060
29. McPherson A, Malkin AJ, Kuznetsov YG (2000) Atomic force microscopy in the study of macromolecular crystal growth. *Annu Rev Biophys Biomol Struct* 29:361–410
30. McPherson A, Kuznetsov YG (2014) Mechanisms, kinetics, impurities and defects: consequences in macromolecular crystallization. *Acta Crystallogr F Struct Biol Commun* 70:384–403
31. Kuznetsov Yu G, Malkin A, McPherson A (1998) Atomic force microscopy studies of phase separations in macromolecular systems. *Phys Rev B* 58:6097–6103
32. Ten Wolde R, Frenkel D (1997) Enhancement of protein crystal nucleation by critical density fluctuations. *Science* 277:1975–1978
33. Haas C, Drenth J (1999) Understanding protein crystallization on the basis of the phase diagram. *J Cryst Growth* 196:388–394
34. Piazza R (1999) Interactions in protein solutions near crystallization: a colloid physics approach. *J Cryst Growth* 196:415–423
35. Malkin A, McPherson A (1994) Light scattering investigations of nucleation processes and kinetics of crystallization in macromolecular systems. *Acta Crystallogr D Biol Crystallogr* 50:385–395
36. McPherson A, Gavira JA (2014) Introduction to protein crystallization. *Acta Crystallogr F Struct Biol Commun* 70:2–20
37. McPherson A, Cudney B (2014) Optimization of crystallization conditions for biological macromolecules. *Acta Crystallogr F Struct Biol Commun* 70:1445–1467
38. Ducruix A, Giége R (1992) *Crystallization of nucleic acids and proteins, a practical approach*. IRL Press, Oxford
39. Bergfors TM (1999) *Protein crystallization: techniques, strategies and tips*. International University Line, La Jolla, CA
40. Bolen DW (2004) Effects of naturally occurring osmolytes on protein stability and solubility: issues important in protein crystallization. *Methods* 34:312–322
41. McPherson A, Cudney R, Patel S (2003) The crystallization of proteins, nucleic acids, and viruses for X-ray diffraction analysis. In: Fahnestock SR, Steinbuchel A (eds) *Biopolymers*, vol 18(16), pp 427–468
42. Hünefeld FL (1840) *Der Chemismus in der thierischen organisation*. FA Brockhaus, Leipzig
43. Jacoby WB (1968) A technique for the crystallization of proteins. *Anal Biochem* 26:295
44. Caffrey M (2003) Membrane protein crystallization. *J Struct Biol* 142:108–132
45. McPherson A (1976) The growth and preliminary investigation of protein and nucleic acid crystals for X-ray diffraction analysis. *Methods Biochem Anal* 23:249–345
46. Chayen NE, Shaw-Stuart PD, Blow DM (1992) Microbatch crystallization under oil: a new technique allowing many small volume crystallization trials. *J Cryst Growth* 122:176–180
47. Salemme FR (1972) A free interface diffusion technique for the crystallization of proteins for X-ray crystallography. *Arch Biochem Biophys* 151:533–539
48. Bard J, Ercolani K, Svenson K, Olland A, Somers W (2004) Automated systems for protein crystallization. *Methods* 34:329–347
49. Hui R, Edwards A (2003) High-throughput protein crystallization. *J Struct Biol* 142:154–161
50. Santarsiero BD, Yegian DT, Lee CC et al (2002) An approach to rapid protein crystallization.

- zation using nanodroplets. *J Appl Crystallogr* 35:278–281
51. Hosfield D, Palan J, Hilger M et al (2003) A fully integrated protein crystallization platform for small-molecule drug discovery. *J Struct Biol* 142:207–217
  52. DeLucas LJ, Bray TL, 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
  53. Luft JR, Collins RJ, Fehrman NA et al (2003) A deliberate approach to screening for initial crystallization conditions of biological macromolecules. *J Struct Biol* 142:170–179
  54. Gilliland GL, Tung M, Blakeslee DM et al (1994) Biological macromolecules crystallization database, version 3.0: new features, data and the NASA archive for protein crystal growth data. *Acta Crystallogr D Biol Crystallogr* 50:408–413
  55. Gilliland GL (1988) A biological macromolecule crystallization database: a basis for a crystallization strategy. *J Cryst Growth* 90:51–60
  56. Cohn EJ, Ferry JD (1943) The interactions of proteins with ions and dipolar ions. In: Cohn EJ, Edsall JT (eds) *Proteins, amino acids and peptides as ions and dipolar ions*. Reinhold, New York, pp 586–622
  57. Cohn EJ, Edsall JT (eds) (1943) *Proteins, amino acids and peptides as ions and dipolar ions*. Van Nostrand-Reinhold, Princeton, NJ
  58. Collins KD (2004) Ions from the Hofmeister series and osmolytes: effects on proteins in solution and in the crystallization process. *Methods* 34:300–311
  59. Herriott RM (1942) Solubility methods in the study of proteins. *Chem Rev* 30:413
  60. Hofmeister F (1888) *Zur Lehre von der Wirkung der Salze*. Nauyn—Schneidebergs Arch Exp Pathol Pharmacol 24:247
  61. McPherson A (2001) A comparison of salts for the crystallization of macromolecules. *Protein Sci* 10:418–422
  62. Sumner JB, Somers GF (1943) *The enzymes*. Academic Press, New York
  63. Englard S, Seifter S (1990) Precipitation techniques. *Methods Enzymol* 182:301–306
  64. Cohn EJ, Hughes WL, Weare JH (1974) Crystallization of serum albumin from ethanol water mixtures. *J Am Chem Soc* 69:1753–1761
  65. McPherson A (1976) Crystallization of proteins from polyethylene glycol. *J Biol Chem* 251:6300–6303
  66. Patel S, Cudney R, McPherson A (1995) Polymeric precipitants for the crystallization of macromolecules. *Biochem Biophys Res Commun* 207:819–828
  67. Ingham KC (1990) Precipitation of proteins with polyethylene glycol. *Methods Enzymol* 182:301–306
  68. Israelachvili J (1997) The different faces of poly(ethylene glycol). *Proc Natl Acad Sci U S A* 94:8378–8379
  69. Sheth SR, Leckband D (1997) Measurements of attractive forces between proteins and end-grafted poly(ethylene glycol) chains. *Proc Natl Acad Sci U S A* 94:8399–8404
  70. McPherson A, Larson SB (2015) A guide to the crystallographic analysis of icosahedral viruses. *Crystallogr Rev* 21:4–55
  71. Timasheff SN, Arakawa T (1988) Mechanism of protein precipitation and stabilization by cosolvents. *J Cryst Growth* 90:39–46
  72. McPherson A, Cudney B (2006) Searching for silver bullets: an alternative strategy for crystallizing macromolecules. *J Struct Biol* 156:387–406
  73. Granick S (1942) Ferritin: I. Physical and chemical properties of horse spleen ferritin. *J Biol Chem* 146:451–461
  74. Giege R, Lorber B, Theobald-Dietrich A (1994) Crystallogenes of biological macromolecules: facts and perspectives. *Acta Crystallogr D Biol Crystallogr* 50:339–350
  75. McPherson A, Malkin A, Kuznetsov YG et al (1996) Incorporation of impurities into macromolecular crystals. *J Cryst Growth* 168:74–92
  76. Dale GE, Oefner C, D'Arcy A (2003) The protein as a variable in protein crystallization. *J Struct Biol* 142:88–97
  77. Derewenda ZS (2004) The use of recombinant methods and molecular engineering in protein crystallization. *Methods* 34:354–363
  78. Derewenda ZS, Vekilov PG (2006) Entropy and surface engineering in protein crystallization. *Acta Crystallogr D Biol Crystallogr* 62:116–124
  79. DeLucas L (ed) (2009) *Membrane protein crystallization: current topics in membranes*, vol 63. Elsevier, Amsterdam
  80. Loll PJ (2003) Membrane protein structural biology: the high throughput challenge. *J Struct Biol* 142:144–153
  81. Michel H (ed) (1990) General and practical aspects of membrane protein crystallization. *Crystallization of membrane proteins*. CRC Press, Boca Raton, FL
  82. Wiener MC (2004) A pedestrian guide to membrane protein crystallization. *Methods* 34:364–372

83. Zulauf M (1990) Detergent phenomena in membrane protein crystallization. In: Michel H (ed) *Crystallization of membrane proteins*. CRC Press, Boca Raton, FL
84. Garavito RM, Ferguson-Miller S (2001) Detergents as tools in membrane biochemistry. *J Biol Chem* 276:32403–32406
85. Hunte C, Jagow G, Schagger H (2003) *Membrane protein purification and crystallization: a practical guide*. Academic Press, San Diego
86. Weiner MC (2001) Existing and emergent roles for surfactants in the three-dimensional crystallization of integral membrane proteins. *Curr Opin Struct Biol* 6:412–419



Protein Crystallography

Methods and Protocols

Wlodawer, A.; Dauter, Z.; Jaskolski, M. (Eds.)

2017, XIII, 672 p. 173 illus., 145 illus. in color.,

Hardcover

ISBN: 978-1-4939-6998-2

A product of Humana Press