UC Irvine

Title

Introduction to protein crystallization

Permalink

Journal

ISSN

Author

McPherson, Alexander

Publication Date

DOI

Copyright Information

Peer reviewed

Introduction to protein crystallization

Alexander McPherson*

Department of Molecular Biology and Biochemistry, University of Califonia, Irvine 560 Steinhaus Hall, Irvine, CA 92697-3900, USA

Accepted 24 March 2004

Abstract

Biological macromolecules can be crystallized by a variety of techniques, and using a wide range of reagents which produce supersaturated mother liquors. These may, in turn, be applied under different physical conditions such as temperature. The fundamental approaches to devising successful crystallization conditions and the factors that influence them are summarized here. For the novice, it is hoped that this brief review might serve as a useful introduction and a stepping-stone to a successful X-ray strucutre determination. In addition, it may provide a framework in which to place the articles that follow.
© 2004 Elsevier Inc. All rights reserved.

1. Some history

Protein crystallization developed in the latter half of the 19th century for three reasons, (a) it provided a means for the purification of specific proteins from an otherwise impure mixture at a time when few other means existed, (b) it served as a demonstration that a protein had been purified (which even now is taken as a pretty good measure), and (c) it was an interesting laboratory curiosity. Initially, the crystallization of hemoglobin from a variety of sources was really nothing more than that [36], though a thoughtful attempt to relate hemoglobin crystal form to evolution was made around 1900 [48]. The first two reasons, however, were dominant in the last quarter of the 19th century and biochemists such as Osborne used it extensively to isolate and characterize proteins, particularly those from seeds. The meaning of this early research for us today is to show first of all that specific proteins can often be isolated from quite impure preparations, but more important, these pioneers deduced many of the approaches for the growth of protein crystals that are still in use.

Between 1900 and 1940, the emphasis was on enzymes, and again, crystallization in the hands of Summer, Northrup, Kunitz, Herriott, and their colleagues [50–52,43] proved an important tool in establishing the properties and nature of catalytic macromolecules. In the late 1930s, however, a new application for protein crystallization appeared as a result of the studies, using X-ray diffraction analysis, of Bernal and Crowfoot [2], Perutz [45], and others. Today, though crystallization is still an admired and respected procedure by enzymologists and protein chemists, X-ray crystallography, and the structure determination of macromolecules and their complexes stand as the principal objective of those involved in crystallization.

A fundamental change in protein crystallization, its investigation and its application, occurred in the 1980s. This was due to the development of recombinant DNA technology which permitted researchers, for the first time, to prepare ample amounts of otherwise rare and elusive proteins. Currently, the majority of proteins addressed by X-ray diffraction are derived from recombinant sources. Among the other consequences is that today we are generally working with more homogeneous protein samples which exhibit greater reproducibility than ever before. Needless to say, this has both accelerated the progress of X-ray crystallography, and greatly expanded both its applicability and its appeal to biochemists and molecular biologists [37,38].

Ultimately, structural biologists would like to describe all living systems, and the materials they produce, in molecular and even atomic terms. This first requires a precise knowledge of the building block molecules, the

---
* Fax: 1-949-824-5992.
E-mail address: amcphers@uci.edu.

proteins, nucleic acids, lipids, and polysaccharides. It further requires information of a somewhat different nature, rules or guidelines to specify how the building block molecules are joined, organized, and assembled into higher order structures. Those greater structures include macromolecular complexes, assemblies, organelles, cell walls, membranes, cytoskeletons, etc. From these assemblies we can finally reach a molecular and atomic level description of the living cell, and from that understand, in terms of classical chemistry and physics, the architecture and mechanics of living matter.

The dynamics within the cell, the mechanisms responsible for the dynamics, and the cell's interactions with exterior influences are equally important. To understand those, however, we first need to delineate how the building block molecules respond to chemical and physical forces, how the responses are regulated, and how the responses are transmitted through the hierarchy of assemblies and higher structures. This in turn means visualizing the building block molecules, not in a single state, but in all of those available as a consequence of their molecular interactions.

The salient elements of this more detailed and comprehensive understanding of life's design and processes are the structures of the building block molecules, and the principles of how they assemble and interact. To the precision required, these properties can only be addressed by X-ray crystallography. The atomic structures of the building block molecules, the proteins and other macromolecules, must be elucidated. This includes not only those easily solubilized and crystallized, but also those that resist current techniques.

Progress in molecular biology and its application to human medicine, agriculture, and industrial processes have for the past two decades been crucially dependent on a detailed knowledge of macromolecular structure at the atomic level. This has included proteins, nucleic acids, viruses, and other large macromolecular complexes and assemblies. Redundancies in structural elements emerging from the now constant flow of newly determined molecular structures suggest that the number of naturally occurring structural motifs and substructures (domains) may be finite. Ultimately then, all macromolecular structures may be classified and catalogued according to polypeptide folds. Once all, or most of the folds which are utilized by nature, are known, then this will provide predictive insight, based primarily on amino acid sequence, into the structures and functions of unknown proteins. The sequences of most proteins, it is important to note, are currently being elucidated by a broad array of sequencing efforts, such as the human genome project, carried out both by government and the private sector. Extension of these genome projects to the three-dimensional structural level appears the next logical step, and this effort, under the broad rubric of structural genomics, is now in the initial stages.

In addition to the dramatic impact that knowledge of three-dimensional structures of proteins has had on fundamental research in biochemistry and biology, macromolecular structure, is of formidable value in biotechnology as well. Here, it provides the essential knowledge required to apply the technique of rational drug design in the creation and discovery of new drugs and pharmaceutical products [37,38]. It serves as the basis of powerful approaches now being applied in emerging biotechnology enterprises, as well as major pharmaceutical companies, to identify lead compounds to treat a host of human ailments, veterinary problems, and crop diseases in agriculture. The underlying hypothesis is that if the structure of the active site of a salient enzyme in a metabolic or regulatory pathway is known, then chemical compounds, such as drugs, can be rationally designed to inhibit or otherwise affect the behavior of that enzyme.

A second approach, of equal importance to biotechnology that also requires knowledge of three-dimensional macromolecular structure is the genetic engineering of proteins. Although recombinant DNA techniques provide the essential synthetic role that permits modification of proteins, structure determination by X-ray crystallography provides the analytical function. It serves as a structural guide for rational and purposeful changes, in place of random and chance amino acid substitutions. Direct visualization of the structural alterations that are introduced by mutation offers new directions for chemical and physical enhancements.

Presently, and in the foreseeable future, 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, for the purposes described above, only X-ray crystallography is adequate. As its name suggests, application of X-ray crystallography is absolutely dependent on crystals of the macromolecule, and not simply crystals, but crystals of sufficient size and quality to permit accurate data collection. The quality of the final structural image is directly determined by the perfection, size, and physical properties of the crystalline specimen, hence the crystal becomes the keystone element of the entire process, and the ultimate determinant of its success [35].

When crystallizing proteins for X-ray diffraction analysis, one is usually dealing with homogeneous, often exceptionally pure, macromolecules, and the objective may be to grow only a few large, perfect crystals. It is important to emphasize that while the number of crystals needed may be few, often the amount of protein available may be severely limited. This in turn places grave constraints on the approaches and strategies that may be used to obtain those crystals. While new methodologies such as synchrotron radiation [21] and cryocrystallography [15] have driven the necessary size of

specimen crystals consistently downward, they have not alleviated the need for crystal perfection.

It is also well to remember that X-ray analysis is a singular event confined to the research laboratory and the final product is basic scientific knowledge. The crystals themselves have no medicinal or pharmaceutical value, but simply serve as intermediaries in the crystallography process. The crystals provide the X-ray diffraction patterns that in turn serve as the raw data which allow the direct visualization of the macromolecules or their complexes composing the crystals.

2. General approach

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

Complicating the entire process, in addition to our limited understanding of the phenomena involved, is the astonishing complexity and the range of 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 weights measured in the millions of daltons, the possibilities for conformation, interaction, and mobility are almost unimaginable.

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

into designing and evaluating the individual and collective trials becomes an important consideration.

3. The nature of protein crystals

Macromolecular crystals like those seen in Fig. 1 are composed of approximately 50% solvent on average, though this may vary from 25 to 90% depending on the particular macromolecule. Protein or nucleic acid occupies the remaining volume so that the entire crystal is in many ways an ordered gel permeated by extensive interstitial spaces through which solvent and other small molecules freely diffuse.

In proportion to molecular mass, the number of bonds (salt bridges, hydrogen bonds, and hydrophobic interactions) that a conventional 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 maintenance, this largely explains the differences in properties between crystals of salts or small molecules and macromolecules.

Living systems are based almost exclusively on aqueous chemistry within narrow ranges of temperature and pH. Macromolecules have, thus, evolved an appropriate compatibility. Serious deviations or perturbations are rarely tolerated. As a consequence, all protein and nucleic acid crystals must be grown from aqueous solutions, ones to which they are tolerant, and these solutions are called mother liquors. Macromolecular crystals have not yet been grown except from such medium.

Although comparable in their morphologies and appearance, there are important practical differences between crystals of low-molecular-mass compounds and crystals of proteins and nucleic acids. Crystals of conventional molecules are characterized by firm lattice forces, are relatively highly ordered, generally physically hard and brittle, 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. Macromolecular 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 [46], CCD area detectors of very high photon counting efficiency [19], high intensity synchrotron X-ray sources [21,46], and new phasing methods (Chapters 12–16) [49] have greatly lessened this constraint.

The extent of the diffraction pattern from a crystal is directly correlated with its degree of internal order. The more vast the pattern, or the higher the resolution to which it extends, the more structurally uniform are the
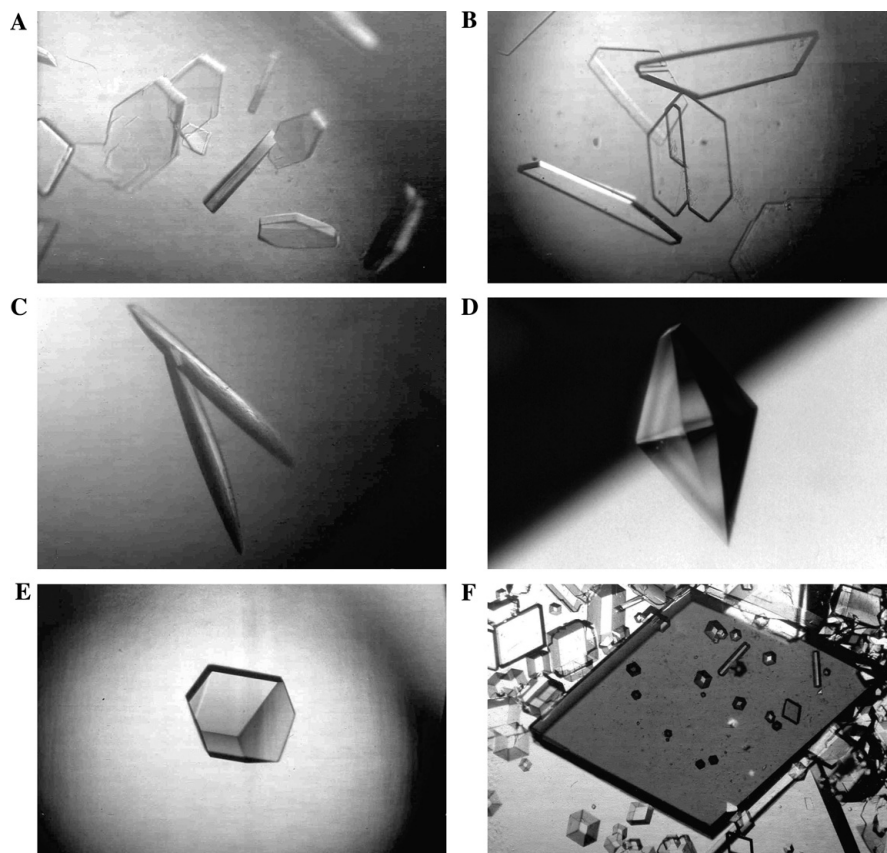
Fig. 1. An array of protein and virus crystals grown for X-ray diffraction analysis under a variety of conditions. In (A) the β-subunit of leutinizing hormone, (B) satellite tobacco mosaic virus, (C) a CH_2 domain deleted human antibody, (D) the sweet protein thaumatin (E) the $T = 1$ particle of brome mosaic virus, and (F) the seed storage protein canavalin.

molecules in the crystal and the more precise is their periodic arrangement. The level of detail to which atomic positions can be determined by crystal structure analysis corresponds closely with 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.

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

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

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

There are further differences which complicate the crystallization of macromolecules as compared with conventional, small molecules [10,12,13,34,39,41]. First, macromolecules may assume multiple distinctive states that include amorphous precipitates, oils, or gels as well as crystals, and most of these are kinetically favored. Second, macromolecular crystals nucleate, or initiate development

only at very high levels of supersaturation, often two to three orders of magnitude greater than required to sustain growth. Finally, the kinetics of macromolecular crystal nucleation and growth are generally two to three orders of magnitude slower than for conventional molecules [27,30,32]. This 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 probability of incorporation of an incoming macromolecule into a growth step [4].

4. Screening and optimization

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

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

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

Once some crystals, even if only microcrystals are observed and shown to be of protein origin (and one ardently hopes for this event) then optimization begins. Every component in the solution yielding crystals must be noted and considered (buffer, salt, ions, etc.), along with pH, temperature, and whatever other factors (see below) might have an impact on the quality of the results. Each of these parameters or factors is then carefully incremented in additional trial matricies encompassing a range spanning the conditions which gave the "hit." Because the problem is non linear, and one variable may be coupled to another, this process is often more complex and difficult than one might expect [1,9,34,39]. It is here that the amount of protein and the limits of the investigator's patience may prove a formidable constraint.

5. Supersaturation, nucleation, and growth

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

Critical nuclei must be considered in terms of the molecular dimensions, the supersaturation, and the surface free energy of molecular addition. Currently the critical nuclear size has only been described for a few systems, and for several cases these were only investigated in terms of two-dimensional nuclei developing on the surfaces of already existent crystals [30,32]. Recently, a theory has emerged which attempts to explain the nucleation phenomenon in terms of statistical fluctuations in solution properties [20,47,53]. This idea holds that a distinctive "liquid protein phase" forms in concentrated protein solutions, and that this "phase" ultimately gives rise to critical nuclei with comprehensive order. This idea is now under study by a variety of experimental techniques in numerous laboratories.

Growth of macromolecular crystals is a better-characterized process than nucleation, and its mechanisms are reasonably well understood. Protein crystals grow principally by the classical mechanisms of dislocation growth, and growth by two-dimensional nucleation, along with two other less common mechanisms known as normal growth and three-dimensional nucleation [31,42]. A common feature of nucleation and growth is that both are critically dependent on what is termed the supersaturation of the mother liquor giving rise to the crystals. Supersaturation is the variable that drives both

processes and determines their occurrence, extent, and the kinetics that govern them.

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

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

6. Creating a state of supersaturation

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

Table 1 is a compilation of the methods upon which one might develop strategies for crystallizing a protein

Table 1
Methods for creating supersaturation

1	Direct mixing to immediately create a supersaturated condition (Batch Method)
2	Alter temperature
3	Alter salt concentration (salting in or out)
4	Alter pH
5	Add a ligand that changes the solubility of the macromolecule
6	Alteration of the dielectric of the medium
7	Direct removal of water (evaporation)
8	Addition of a polymer that produces volume exclusion
9	Addition of a cross bridging agent
10	Concentration of the macromolecule
11	Removal of a solubilizing agent

for the first time. Indeed there may be others, the limit is only a function of the imagination and cunning of the investigator. The details of these various approaches have been described in detail numerous times elsewhere [1,9,34,39,40] and need receive no more attention here. It is probably sufficient to say that if a protein has any propensity to crystallize readily, it can probably be accomplished by variation of precipitant type, precipitant concentration, pH, 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.

7. Methodology

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

Again, all of these devices and their methodologies have been described in detail elsewhere (and also elaborated upon in other chapters of this volume). It is unnecessary to comment on each of them again. In addition, detailed instructions are frequently provided by the manufacturers of the crystallization kits, supplies, and plasticware along with much helpful material. Suffice it to say that currently the hanging drop and sitting drop procedures for vapor diffusion, and the batch method using microdrops under oil are most in favor, and are recommended for most investigations. In those cases where mother liquor components cannot be transported through the

Table 2
Methods for promoting a solubility minimum

1	Bulk crystallization
2	Batch method in vials
3	Evaporation
4	Bulk dialysis
5	Concentration dialysis
6	Microdialysis
7	Liquid bridge
8	Free interface diffusion
9	Vapor diffusion on plates (sitting drop)
10	Vapor diffusion in hanging drops
11	Sequential extraction
12	pH-induced crystallization
13	Temperature-induced crystallization
14	Crystallization by effector addition

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 optimal means. Thus, one may start with one technique but ultimately find that another gives larger crystals of higher quality.

As illustrated particularly in the article by Sommers, et al. in this volume, screening for crystallization conditions, and even optimization in some cases, has been consigned in high throughput laboratories to robotic devices. This is particularly true in those of large pharmaceutical companies where many proteins may be under simultaneous investigation. Robotic systems have the advantages of exceptional sample record maintenance, most of them 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. Robotic systems are, in addition, now being used to examine and evaluate the results of crystallization trials using optical subsystems and image processing techniques [8,24,29]. Evaluation of trial arrays of conditions, however, continues to be

problematic because of the continuing difficulty in devising meaningful criteria in the absence of actual crystals. That is, the sole presence of various kinds of precipitates or other phases in an individual crystallization trial gives only very murky indications of how near the conditions were to a successful mother liquor.

8. Precipitants

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

Individual precipitants and their properties have also been reviewed [39] and will not be extensively discussed here. To simplify, however, it is possible to group the precipitants into categories based on their mechanisms for promoting crystallization, and this is done in Table 3. Precipitants of macromolecules fall into four broad categories (1) salts, (2) organic solvents, (3) long chain polymers, and (4) low molecular weight polymers and non-volatile organic compounds. The first two classes are typified by ammonium sulfate and ethyl alcohol, respectively, and higher polymers such as polyethylene glycol 4000 are characteristic of the third. In the fourth category, we might place compounds such as methylpentanediol and polyethylene glycols of molecular weight less than about 1000.

The solubility of macromolecules in concentrated salt solutions is complicated, but it can be viewed naively as a competition between salt ions, principally the anions,

Table 3
Precipitants used in macromolecular crystallization

Salts	Volatile organic solvents	Polymers	Non-Volatile organic solvents
Ammonium phosphate sulfate	Ethanol	Poly(ethylene glycol) 1000, 3350, 6000, 8000, 20,000	2-Methyl-2, 4-pentanediol
Lithium sulfate	Propanol and isopropanol	Jeffamine T, Jeffamine M	
Sodium or ammonium citrate	1,3-Propanediol	Poly(ethylene glycol) monomethyl ester	2,5-Hexandediol
Sodium or potassium phosphate	2-Methyl-2, 4-pentanediol	Poly(ethylene glycol) monostearate	Ethylene glycol 400
Sodium or potassium or ammonium chloride	Dioxane	Polyamine	
Sodium or ammonium acetate	Acetone		
Magnesium or calcium sulfate	Butanol		
Cetyltriethyl ammonium salts	Acetonitrile		
Calcium chloride	Dimethyl sulfoxide		
Ammonium or sodium nitrate			
Sodium or magnesium formate	Methanol		
Sodium or potassium tartrate	1,3-Butyrolactone		
Cadmium sulfate	Ethylene glycol 400		

and the macromolecules for the binding of water molecules which are essential for the maintenance of solubility [5,6,22,23]. At sufficiently high salt concentrations the macromolecules become so uncomfortably deprived of solvent that they seek association with one another in order to satisfy their electrostatic requirements. In this environment, ordered crystals as well as disordered amorphous precipitate may form. Some salt ions, chiefly cations, are also necessary to insure macromolecular solubility. At very low ionic strengths, cation availability is insufficient to maintain macromolecule 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. 2.

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

One might anticipate little variation among different salts so long as the valences of their ions were the same. Thus, there should be little expected variation between two different salts such as $(NH_4)_2PO_4$ and $(NH_4)_2SO_4$ if only ionic strength were involved. This, however, is often observed not to be the case. In addition to salting out, which is a general dehydration effect, or reduction of the chemical activity of water, there are also specific protein–ion interactions that may have other consequences. This is perhaps not unexpected given the unique polyvalent character of individual proteins, their structural complexity, and the intimate dependence of their physical properties on their surroundings. It is inadequate, therefore, when attempting to crystallize a protein to examine only one or two salts and ignore the broader range. Alternative salts can sometimes produce crystals of varied quality, morphology, and in some cases diffraction properties.

It is usually not possible to predict the degree of saturation or molarity of a precipitating agent required for the crystallization of a particular protein or nucleic acid without some prior knowledge of its behavior. In general, however, it is a concentration just a few percent less than that which yields an amorphous precipitate [52], and this can be determined for a macromolecule under a given set of conditions using only minute amounts of material [34]. To determine the approximate insolubility points with a particular precipitant, a 10 µl 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-µl 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 ionic strength, pH is one of the most important variables influencing the solubility of proteins. As such, it provides another powerful approach to creating supersaturated solutions, and hence effecting crystallization. Its manipulation at various ionic strengths and in the presence of diverse precipitants is a fundamental idea in formulating screening matrices and discovering successful crystallization conditions. An example of the effect of pH on two different proteins is illustrated in Fig. 3.

Organic solvents reduce the dielectric of the medium, hence the screening of the electric fields that mediate macromolecular interactions in solution. As the concentration of organic solvent is increased, attraction between macromolecules increases, solvent becomes less
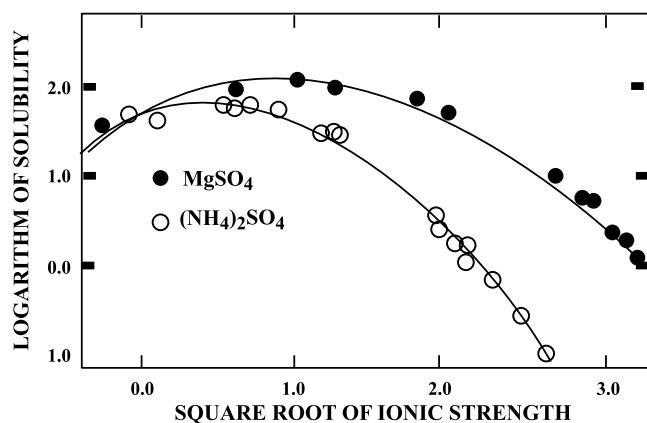

Fig. 2. The solubility of a typical protein, enolase, is shown here as a function of ionic strength produced by two different, widely used salts. The regions of the end points of the curves where solubility decreases are called, at low ionic strength, the "salting in" region, and at high ionic strength, the "salting out" region. Both provide opportunities for the creation of supersaturated macromolecular solutions and crystal growth.
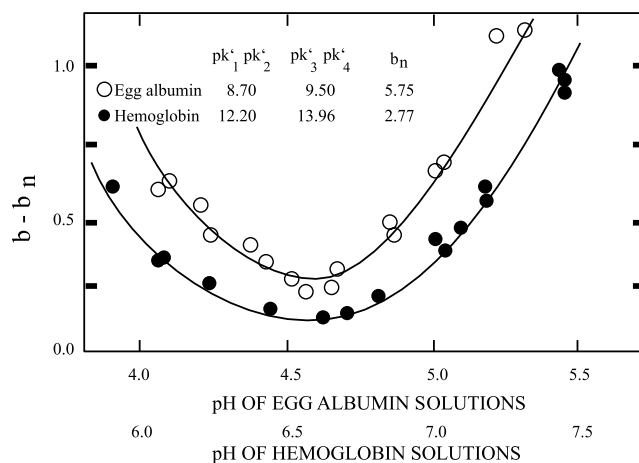

Fig. 3. Solubility of two typical proteins, hen egg albumin and hemoglobin, as a function of pH. All parameters are otherwise constant. Both proteins show dramatic decreases in their solubilities at characteristic pH values, a feature that can be used to advantage in creating supersaturated solution of the proteins.

effective (the activity coefficient of water is reduced), and the solid state is favored [7,11]. Organic solvents should be used at a low temperature, at or below 0 °C, and they should be added very slowly with good mixing [39]. Since they are usually volatile, vapor diffusion techniques are equally applicable for either bulk or micro amounts. Ionic strength should, in general, be maintained low and whatever means are otherwise available should be pursued to protect against denaturation.

Some polymers, polyethylene glycols are most popular [33,44], produce volume exclusion effects that also induce separation of macromolecules from solution [26,33]. The polymeric precipitants, unlike proteins, have no consistent conformation, writhe and twist randomly in solution, and occupy far more space than they otherwise deserve. This results in less solvent available space for the other macromolecules which then segregate, aggregate, and ultimately form a solid state, often crystals.

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

PEG sizes from $M_r = 400$ to 20,000 have successfully provided protein crystals, but the most useful are those in the range 2000–8000. A number of cases have appeared, however, in which a protein could not easily be crystallized using this range but yielded in the presence of PEG 400 or 20,000. The molecular weight sizes are generally not completely interchangeable for a given protein even within the mid range. Some produce the best-formed and largest crystals only at, say, $M_r = 3350$ and less perfect examples at other weights. This is a parameter which is best optimized by empirical means along with concentration and temperature. The very low

molecular weight PEGs such as 200 and 400 are rather similar in character to MPD and hexanediol. There does not appear to be any correlation between the molecular weight of a protein and that of the PEG best used for its crystallization. The higher molecular weight PEGs do, however, have a proportionally greater capacity to force proteins from solution.

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

Since PEG solutions are not volatile, PEG must be used like salt or MPD and equilibrated with the protein by dialysis, slow mixing, or vapor equilibration. When the reservoir concentration is in the range 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 4–5% below the final value.

9. Factors affecting crystallization

There are many factors that affect the crystallization of macromolecules [34,39] and many of these are summarized in Table 4. These may affect the probability of

Table 4
Factors effecting crystallization

	Physical		Chemical		Biochemical
1	Temperature/temperature variation	1	pH	1	Purity of the macromolecule/impurities
2	Surfaces	2	Precipitant type	2	Ligands, inhibitors, effectors
3	Methodology/approach to equilibrium	3	Precipitant concentration	3	Aggregation state of the macromolecule
4	Gravity	4	Ionic strength	4	Post-translational modifications
5	Pressure	5	Specific ions	5	Source of macromolecule
6	Time	6	Degree of supersaturation	6	Proteolysis/hydrolysis
7	Vibrations/sound/mechanical perturbations	7	Reductive/oxidative environment	7	Chemical modifications
8	Electrostatic/magnetic fields	8	Concentration of the macromolecules	8	Genetic modifications
9	Dielectric properties of the medium	9	Metal ions	9	Inherent symmetry of the macromolecule
10	Viscosity of the medium	10	Crosslinkers/polyions	10	Stability of the macromolecule
11	Rate of equilibration	11	Detergents/surfactants/amphophiles	11	Isoelectric point
12	Homogeneous or heterogeneous nucleants	12	Non-macromolecular impurities	12	History of the sample

its occurring at all, the nucleation probability and rate, crystal growth rate, and the ultimate sizes and quality of the products. As noted above, pH and salt, or the concentrations of other precipitants are of great importance. The concentration of the macromolecule, which may vary from as low as 2 mg/ml to as much as 100 mg/ml, is an additional, significant variable.

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

The most intriguing problem, or opportunity depending on one's perspective, is what additional components or compounds should comprise the mother liquor in addition to solvent, protein, and precipitating agent. The most probable effectors are those which maintain the protein in a single, homogeneous, and invariant state. Reducing agents such as glutathione or β-mercaptoethanol 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 desirable 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 or 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 macromolecules, 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 refractile. It is worthwhile, therefore, when searching for crystallization conditions, to explore complexes of the macromolecule with substrates, coenzymes, analogues, and inhibitors at an early stage. Such complexes are, in addition, inherently more interesting in a biochemical sense than the apoprotein.

Various metal ions have occasionally been observed to promote the crystallization of proteins and nucleic acids. In some instances, these ions were essential for activity. It was, therefore, reasonable to expect that they might aid in maintaining certain structural features of the molecule. In other cases, however, metal ions, particularly divalent metal ions of the transition series, were found that encouraged crystal growth but played no known role in the macromolecule's activity. Likely, they serve as bridging agents between molecules in the crystal lattice.

9.1. Membrane proteins

Proteins that are naturally membrane associated or otherwise unusually hydrophobic or lipophilic in nature invariable present unusual problems. Such proteins are, in general, only sparingly soluble in normal aqueous media, some virtually insoluble, and this in turn makes the application of conventional protein crystallization techniques problematic. Such cases are difficult but not intractable. To address these difficulties the use of detergents, particularly non-ionic detergents, has been developed. 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 inappropriate here.

The essential difficulty with the necessity of including 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 in using a standard 48-well screen of conditions, at least initially, then the additional search for a useful detergent means that the 48 sample screen must then be multiplied by the number of detergent candidates. The problem is that there are a lot of potentially useful detergents. Hampton Research, a major source of screening reagents, offers three different detergent kits of 24 samples each. Were one to simply apply the basic 48-well screen with each detergent, then that would require a total of 3456 individual trials. While this may actually be possible with highly automated systems, and where a substantial amount of material is available, it is impractical for most laboratories.

The basic crystal screens, whether they are systematic screens or shotgun screens, cannot be abandoned, however. Thus, it becomes essential to reduce, at least in initial screens, the number of detergents to be considered. If, for example, a set of 6 highly promising detergents could be identified, then less than 300 trials would be called for initially, an undertaking well within the capabilities of most laboratories. No one, however, has yet reduced the set to a favored few, everyone has their own opinion as to which detergents should constitute it, and certainly no consensus set has yet emerged from databases

or from analyses of experiments and the successful structure determinations that have been carried out. Hopefully, such a reduction in the detergent variable will be among the first important products of the structural genomics enterprise. This will be true, however, only if membrane and lipophilic proteins are addressed with the same enthusiasm and intensity as are the soluble macromolecules.

To make matters in this area even worse, 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 and seem to convert it into a hopeless exercise. Again, we can only hope that experience and the careful recording of data will provide us with a reduced set of most promising amphiphiles.

While not as valuable as naming actual candidate detergents, the author can point to a number of useful reviews and discussions that illustrate the properties and virtues of various detergents for membrane crystallization, and also call attention to the chapter by Nollert in this volume. Michel (1990) [55] is a good review of work up until that time, more recently, there are fine discourses by Loll [28], Caffrey [3], Garavito and Ferguson-Miller [14], Hunte et al. [25], and Wiener [54].

9.2. The protein as a variable

At the risk of belaboring a point, a factor of particular importance is the purity of the macromolecule [16] and this deserves special emphasis. Some proteins, it is true, may crystallize even from very heterogeneous mixtures, and indeed, crystallization has long been used as a powerful purification tool. In general, however, the likelihood of success in crystal growth is greatly advanced by increased homogeneity of the sample. Investment in further purification is always warranted and usually profitable. When every effort to crystallize a macromolecule fails, the best recourse is to further purify.

Upon entering the field of macromolecular crystallography one is struck by the extraordinary range of molecules and their properties that one must contend with, and the extensive variety of techniques and conditions that must be tested in order to grow crystals suitable for X-ray diffraction analysis. It would indeed be useful if some comprehensive database existed that at least contained the experiences accumulated over the years. Indeed, such a knowledge base, combined with a system to search and sift for all kinds of relevant information regarding protein crystal growth, has been compiled and is readily available. This is the crystallization database devised by Gilliland [17,18] and distributed through the National Institute of Standards and Technology (www.bmcd@NIST.gov). This database provides a valuable tool for the novice as well as the experienced crystallographer. It includes virtually all of those conditions used

to grow crystals of individual proteins, and it provides innumerable ideas regarding procedures and techniques.

Recombinant DNA technology provided an enormous impetus to crystal growth research and X-ray crystallography 25 years ago, and it may be on the verge of providing another at this very time. Arguably, but hardly so, 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, now provides us that opportunity, and its possibilities are only now beginning to be realized.

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

9.3. Important principles

Although the approaches to macromolecular crystallization remain largely empirical, much progress has been made, particularly over the past 25 years. We have now identified useful reagents, devised a host of physical-chemical techniques for studying the crystallization process, and gained a better understanding of the unique features of proteins, nucleic acids, and macromolecular

Table 5
Some important principles

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

assemblies that affect their capacity to crystallize. Some principles now stand out regarding the crystallization problem, and these are summarized in Table 5. It remains to the individual investigator to find practical means to institute these ideas and determine for a specific problem which are of critical importance, and which will have greatest influence on the likelihood of success.

References

[1] T.M. Bergfors, Protein Crystallization: Techniques, Strategies and Tips, International University Line, La Jolla, CA, 1999.

[2] J.D. Bernal, D. Crowfoot, Nature 133 (1934) 794.

[3] M. Caffrey, J. Struc. Biol. 142 (2003) 108–132.

[4] A.A. Chernov, J. Struc. Biol. 142 (2003) 3–21.

[5] E.J. Cohn, J.T. Edsall (Eds.), Proteins, Amino Acids and Peptides as Ions and Dipolar Ions, Van Nostrand-Reinhold, Princeton, NJ, 1943.

[6] E.J. Cohn, J.D. Ferry (Eds.), The Interactions of Proteins with Ions and Dipolar Ions. Proteins, Amino Acids and Peptides, Van Nostrand-Rheinhold, New Jersey, 1943.

[7] E.J. Cohn, W.L. Hughes, et al., J. Am. Chem. Soc. 69 (1974) 1753–1761.

[8] L.J. DeLucas, T.L. Bray, et al., J. Struc. Biol. 142 (2003) 188–206.

[9] A. Ducruix, R. Giége, Crystallization of Nucleic Acids and Proteins, A Practical Approach, IRL Press, Oxford, 1992.

[10] S.D. Durbin, G. Feher, Annu. Rev. Phys. Chem. 47 (1996) 171–204.

[11] S. Englard, S. Seifter, Precipitation Techniques, Methods Enzymol. 182 (1990) 301–306.

[12] G. Feher, J. Cryst. Growth 76 (1986) 545–546.

[13] R.S. Feigelson, J. Cryst. Growth 90 (1988) 1–13.

[14] R.M. Garavito, S. Ferguson-Miller, J. Biol. Chem. 276 (2001) 32403–32406.

[15] E.F. Garmen, T.R. Schneider, J. Appl. Cryst. 30 (1997) 211–237.

[16] R. Giege, B. Lorber, et al., Acta Cryst. D 50 (1994) 339–350.

[17] G.L. Gilliland, J. Cryst. Growth 90 (1998) 51–59.

[18] G.L. Gilliland, M. Tung, et al., Acta Cryst. D 50 (1994) 408–413.

[19] S.M. Gruner, E.F. Eikenberry, et al., Comparison of X-ray Detectors, in: M.G. Rossmann, E. Arnold (Eds.), International Tables for Crystallography, F, Kluwer Academic Publishers, Dordrecht/Boston/London, 2001, pp. 143–153.

[20] C. Haas, J. Drenth, J. Cryst. Growth 196 (1999) 388–394.

[21] J.R. Helliwell, Macromolecular Crystallography with Synchrotron Radiation, Cambridge University Press, Cambridge, UK, 1992.

[22] R.M. Herriott, Chem. Rev. 30 (1942) 413.

[23] T. Hofmeister, Z. Physiol. Chem. 14 (1890) 165.

[24] D. Hosfield, J. Palan, et al., J. Struc. Biol. 142 (2003) 207–217.

[25] C. Hunte, V. Jagow, et al., Membrane Protein Purification and Crystallization: A Practical Guide, Academic Press, San Diego, 2003.

[26] K.G. Ingham, Methods Enzymol. 182 (1990) 301–306.

[27] Y.G. Kuznetsov, A.J. Malkin, et al., J. Struc. Biol. 114 (3) (1995) 184–196.

[28] P.J. Loll, J. Struct. Biol. 142 (1) (2003) 144–153.

[29] J.R. Luft, R.J. Collins, et al., J. Struc. Biol. 142 (2003) 170–179.

[30] A.J. Malkin, Y.G. Kuznetsov, et al., J. Phys. Chem 100 (1996) 11736–11743.

[31] A.J. Malkin, Y.G. Kuznetsov, et al., Nat. Struc. Biol 2 (11) (1995) 956–959.

[32] A.J. Malkin, Y.G. Kuznetsov, et al., Surface Sci 393 (1997) 95–107.

[33] A. McPherson, J. Biol. Chem. 251 (1976) 3600–6303.

[34] A. McPherson, The Preparation and Analysis of Protein Crystals, Wiley, New York, 1982.

[35] A. McPherson, Sci. Am. 260 (1989) 62.

[36] A. McPherson, J. Cryst. Growth 110 (1991) 1–10.

[37] A. McPherson, J Cryst. Growth 122 (1992) 161–167.

[38] A. McPherson, The Role of X-ray crystallography in structure based rational drug design, in: Chemical and Structural Approaches to Rational Drug Design, CRC Press, Boca Raton, FL, 1994.

[39] A. McPherson, Crystallization of Biological Macromolecules, Cold Spring Harbor Laboratory Press, Cold Spring Harbor, NY, 1999.

[40] A. McPherson, R. Cudney, et al., The crystallization of proteins, nucleic acids, and viruses for X-ray diffraction analysis, in: Biopolymers, vol. 18 (2003) 427–468.

[41] A. McPherson, A.J. Malkin, et al., Structure 3 (8) (1995) 759–768.

[42] A. McPherson, A.J. Malkin, et al., Ann. Rev. Biophys. Biomol. Struc. 29 (2000) 361–410.

[43] J.H. Northrop, M. Kunitz, et al., Crystalline Enzymes, Columbia University Press, New York, 1948.

[44] S. Patel, R. Cudney, et al., Biochemistry and Biophysics Research Communications 207 (1995) 819–828.

[45] M.F. Perutz, Sci. Am. 211 (1964) 64.

[46] J.W. Pflugrath, Curr. Opin. Struct. Biol. 2 (1992) 811–815.

[47] R. Piazza, J. Cryst. Growth 196 (1999) 415–423.

[48] E.T. Reichert, A.P. Brown, The Differentiation and Specificity of Corresponding Proteins and Other Vital Substances in Relation To Biological Classification and Organic Evolution: The Crystallography of Hemoglobins, Carnegie Institute, Washington, DC, 1909.

[49] M.G. Rossmann, E. Arnold, Crystallography of Biological Macromolecules, Kluwer Academic Publishers, Dordrecht/Boston/London, 2001.

[50] J.B. Sumner, J. Biol. Chem. 69 (1926) 435.

[51] J.B. Sumner, A.L. Dounce, J. Biol. Chem. 1231 (1937) 417.

[52] J.B. Sumner, G.F. Somers, The Enzymes, Academic Press, New York, 1943.

[53] R. Ten Wolde, D. Fraenkel, Science 277 (1997) 1975–1978.

[54] M.C. Weiner, Curr. Opin. Struct. Biol. 6 (2001) 412–419.

[55] H. Michel (Ed.), General and practical aspects of membrane protein crystallization. Crystallization of membrane proteins. CRC Press, Boca Raton, FL, 1990.