defects, disorder, impurity incorporation and internal stress. Such imperfections can be visualized using laser confocal microscopy combined with differential interference contrast microscopy (Iimura *et al.*, 2005; Sazaki *et al.*, 2005). They can also be evaluated by the resolution limit of diffraction, which expresses the average disorder of the molecules in the crystal lattice. Wilson plots provide good illustrations of diffraction quality for protein crystals. Other sources of disorder, such as the mosaic structure of the crystal, may strongly influence the quality of the diffraction data. They are responsible for increases in the diffuse background scatter and a broadening of diffraction intensities. These defects are difficult to monitor with precision, and dedicated techniques and instruments are required for accurate analysis.

Mosaicity can be defined experimentally by X-ray rocking width measurements. An overall diagnostic of crystal quality can be obtained by X-ray diffraction topography. Both techniques have been refined with lysozyme as a test case and were used for comparative analysis of crystals grown under different conditions, both on earth and in microgravity. For lysozyme and thaumatin, improvement of the mosaicity, as revealed by decreased rocking widths measured with synchrotron radiation, was observed for the microgravity-grown crystals (Snell *et al.*, 1995; Ng, Lorber *et al.*, 1997; Lorber, Sauter, Robert *et al.*, 1999).

An illustration of mosaic block character in a lysozyme crystal was provided by X-ray topography (Fourme *et al.*, 1995). Comparison of earth- and microgravity-grown lysozyme crystals showed a high density of defects in the earth control crystals, while in the microgravity case, several discrete regions were visible (Stojanoff *et al.*, 1996). X-ray topographs have also been used to compare crystal polymorphs (Izumi *et al.*, 1996), to monitor temperature-controlled growth of tetragonal lysozyme crystals (Stojanoff *et al.*, 1997), to study the effects of solution variations during growth on crystal perfection (Dobrianov *et al.*, 1998; Otálora *et al.*, 1999), to compare crystals grown in solution and in agarose gel (Lorber, Sauter, Ng *et al.*, 1999), and to map defects in the bulk of protein crystals (Hu *et al.*, 2001).

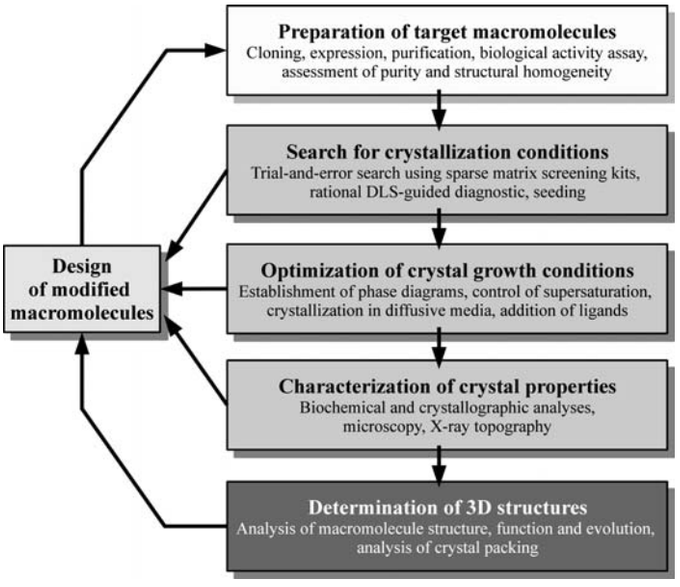

Figure 4.1.6.1
From the target molecule to its three-dimensional structure: a flowchart for a structure detemination. The figure illustrates the different steps the crystal grower will go through to access three-dimensional information. Ideally, the trajectory will be linear, crystallization remaining the major bottleneck in the whole process. If one gets stuck at any stage, the difficulties can often be circumvented by modifying the target molecule, either genetically or (bio)chemically, by improving its purity/homogeneity, or by adding ligands to stabilize its structure.

4.1.6. How to crystallize a new macromolecule: the structural biology view

Fig. 4.1.6.1 schematizes the progress of structural biology projects and highlights the pivotal role of the crystallization step. Despite tremendous progress in macromolecule and crystal preparation, crystallization remains the most serious bottleneck. Various strategies have been proposed to overcome the difficulty (*e.g.* D'Arcy, 1994; Ducruix & Giegé, 1999; Kundrot, 2004; McPherson, 2004; DeLucas *et al.*, 2005; Pusey *et al.*, 2005; Hughes & Ng, 2007; Li *et al.*, 2007; Chayen & Saridakis, 2008). The choice is largely dependent on the type of project [functional biology with well defined protein(s) or structural genomics with large ensembles of known and unknown proteins] and thus on non-exclusive rational, incomplete factorial screening, or empirical approaches.

4.1.6.1. How to start and how to choose what screening kits to start with

In practice, the very first questions posed when one begins to think about crystallization assays are: (i) which crystallization kit to start with, (ii) what to do if no crystals grow and (iii) what happens when one gets crystals. The answer to the first question is in fact simplest. One should start with a screen that does not unreasonably tax the supply of protein but which explores the widest volume of crystallization space, *i.e.* samples the largest number of crystallants and crystallant concentrations over the largest range of pH. There are many kits on the market using 96-sample trials that accomplish this well. If protein is severely limited, then there are 48-sample kits that do the job satisfactorily. In case of no success one should examine the resulting pattern of precipitates, phase separations, clear drops and assorted odd accumulations, and try to divine what might be the best option to try next. Good advice is to check how others have crystallized similar proteins and to try complexes of the protein with its physiological ligands and effectors.

If, on the other hand, one is fortunate and the first kit does indeed yield crystals, but perhaps of insufficient size or quality, or of troubling morphology, then crystallization must be optimized to get better ones. One should also check whether it is really the macromolecule that crystallized and not a ligand or any component of the crystallization mixture. Optimization means varying the chemical and physical parameters of the reagent mix that yielded the crystals and searching the crystallization parameter space by small increments away from the starting point. There is still a component of art and mystery in science, and this is one instance where its appreciation is paramount.

As a practical example, one can consider the 'Silver bullet' concept, which employs a limited set of fundamental crystallization conditions combined with a broad screen of potentially useful small-molecule additives (McPherson & Cudney, 2006), with the hypothesis that some of them might establish intermolecular noncovalent cross-links (Larson *et al.*, 2008). In the 'Uppsala' concept, a crystallization project entails two distinct steps: screening and optimization. The aim of the initial phase is to screen the parameters affecting crystallization as broadly as possible. If promising conditions are found, these are optimized with other protocols including grid and additive screens, seeding, and manipulation of drop kinetics (Bergfors, 2007).

4.1.6.2. Rules and general principles

Rules for sample preparation. The first concern is to obtain a macromolecular sample of the highest quality, second to collate all biochemical and biophysical information characterizing the macromolecule in order to design the best crystallization strategy, and finally, to establish precise protocols that ensure reproducibility of experiments. Solubility screening is an important issue and can be automated (Listwan *et al.*, 2009). Important as well is to clean and sterilize by filtration (over 0.22 µm porosity membranes, or even 0.10 µm) all solutions in contact with pure proteins, to remove dust and other solid particles, and to avoid contamination by microbes (Chayen, 2009). Inclusion of sodium azide in crystallizing solutions may discourage invasive bacteria and fungi. In vapour-diffusion assays, placing a small grain of thymol in the reservoir can prevent contamination. Thymol, however, can occasionally have specific effects on crystal growth (Chayen *et al.*, 1989) and thus may serve as an additive in screenings as well.

The preparation of solutions for crystallization experiments should follow some common rules. Stocks should be prepared with chemicals of the purest grade dissolved in double-distilled water, filtered through 0.22 µm membranes and ultracentrifuged (1 h at 100 000*g*) to remove aggregates. The chemical nature of the buffer is an important parameter, and the pH of buffers, which must be strictly controlled, is often temperature dependent, especially that of Tris buffers. Commercial PEG may contain ionic contaminants (Jurnak, 1986), or contaminants derived from peroxidation, and thus repurification is recommended (Ray & Puvathingal, 1985).

Mother liquors are defined as the solutions that contain all compounds (buffer, crystallant *etc.*) at the final concentration for crystallization, except the protein (to avoid dissolution of the crystals when transferred to a fresh mother liquor, soluble protein should be present). Samples of proteins often contain quantities of salt of unknown composition, and it is therefore wise to dialyse new batches against well characterized buffers. Whatever the crystallization method used, it almost always requires a high concentration of protein. This may imply concentration steps using devices operating under nitrogen pressure, by centrifugation or by lyophilization (note that lyophilization may denature proteins and that non-volatile salts will accumulate). Dialysis against high-molecular-weight PEG may also be used. During concentration, the pH and ionic strength may vary and, if not kept at the appropriate values, denaturation of samples may occur.

How to use phase diagrams and favour reproducibility. Crystallization requires bringing the protein to a supersaturated state that favours nucleation. Use of phase diagrams may be important for that purpose (Haas & Drenth, 1999; Sauter, Lorber *et al.*, 1999; Asherie, 2004; Lorber & Witz, 2008). If solubilities or phase diagrams are unavailable, it is nevertheless important to understand the correlation between solubility and the way supersaturation is reached in the different crystallization methods (Fig. 4.1.3.1). In dialysis, the protein concentration remains constant during equilibration. The initial concentration of the crystallant in the exterior solution leaves the macromolecule in an undersaturated state. With increasing concentration of this agent in the exterior solution, a state of supersaturation can be attained, leading to crystallization or precipitation. In a vapour-diffusion experiment, where the concentration of crystallant in the reservoir exceeds that in the drop, the protein will begin to concentrate from an undersaturated to a supersaturated state, with both protein and crystallant concentrations increasing. Crystals appear in the labile region. For crystals that appear first, the trajectory of equilibration is complex and the remaining concentration of protein in solution will converge towards a point located on the solubility curve. In batch crystallization using a closed vessel, three situations can occur: if the concentration of the protein is undersaturated, crystallization never occurs (unless another parameter such as temperature is varied); if it belongs to the supersaturated region between solubility and precipitation curves, crystals can grow until the remaining concentration of the protein in solution equals its solubility; if supersaturation is too high, the protein precipitates immediately, although in some cases crystals can grow from precipitates by Ostwald ripening (Ng *et al.*, 1996; Lorber & Witz, 2008). Altogether, the crystal habits (*e.g.* Sauter, Lorber *et al.*, 1999; Zhu *et al.*, 2001) and the underlying growth mechanisms (Fig. 4.1.5.1) are dependent on the region in the phase diagram where nucleation occurred and on the equilibration trajectories (Figs. 4.1.1.1 and 4.1.3.1).

Reproducibility should be a strategic concern (Newman *et al.*, 2007), but is often not reached despite the efforts of investigators to work with protein and crystallant solutions of identical chemical composition. Besides the many reasons that could explain this situation, the effect of the mixing protocol of the different components constituting a crystallization drop on the output of the experiment is often overlooked. Since mixing is not instantaneous, it will generate local supersaturation maxima and minima that can affect the crystallization process. Thus, when seeking reproducibility, active mixing of the crystallization drops is recommended, which is easy to perform when working manually but not trivial in automated procedures (Howard *et al.*, 2009).

4.1.6.3. Database mining and statistics

With the number of protein structures solved by X-ray crystallography now approaching 100 000, substantial databases of successful crystallization conditions and procedures have been developed (*e.g.* Peat *et al.*, 2005; Li *et al.*, 2007; Tung & Gallagher, 2009). From these it might be expected that some predictive insight would be obtained regarding the most probable crystallization conditions for proteins in general (*e.g.* Kimber *et al.*, 2003), and for specific families of proteins sharing common physical, chemical or functional properties such as aminoacyl-tRNA synthetases (Giegé *et al.*, 2008). However, the problem is that each protein remains an individual endowed with its own eccentricities, even within a family, and often these dramatically alter its crystallization behaviour. In addition, proteins in general may be exquisitely sensitive to only minor modifications to their properties, further complicating their rational classification in terms of crystal growth.

Nonetheless, statistical analyses are now being widely applied to the expanding databases and some results, both interesting and useful, are beginning to emerge. For example, reduced sets of the most favourable crystallization conditions have been proposed by several groups of investigators based on past successes. These may be useful because they reduce the number of crystallization trials in cases where either the amount of protein is limited or a very large number of constructs are to be screened. Favoured reagents and, particularly, useful additives have been identified by such analyses. Correlations have been sought between the physical or chemical properties of specific proteins and their manner of crystallization, such as between pI and crystallization pH, but this has had only limited success. A promising method

generates relatively accurate predictions from protein composition and collocation of amino acids, pI and hydrophobicity deduced from sequence data, and number of solved structures of similar sequences (Mizianty & Kurgan, 2009). Interestingly, data mining of biophysical properties that control protein crystallization led to the conclusion that crystallization propensity depends primarily on the prevalence of well ordered surface epitopes capable of mediating interprotein interactions rather than on the overall thermodynamic stability of the proteins (Price *et al.*, 2009).

4.1.6.4. Strategic concerns: a summary

Homogeneity and stability. Perhaps the most important property of a system to be crystallized is its purity. Crystallization presupposes that identical units are available for incorporation into a periodic lattice. If crystallization fails, reconsidering purification protocols often helps achieve success. No homogeneous molecular population can remain so if its members alter their form, folding or association state. Hence, it is crucial that proteins in solution not be allowed to denature, aggregate or undergo conformational changes.

Solubility. Before a molecule can be crystallized, it must be solubilized. This means creation of monodisperse solutions free of aggregates and molecular clusters. Solubility and crystallizability strongly depend on substances (organic solvents and PEGs) that reduce the ionic strength of the solution (Papanikolau & Kokkinidis, 1997).

Supersaturation and nucleation. Crystals grow from systems displaced from equilibrium so that restoration requires formation of the solid state. Thus, the first task is to find ways to alter properties of the crystallizing solutions, such as by pH or temperature change, and to create supersaturated states. The number, size and quality of crystals depend on the mechanisms and rates of nuclei formation. For diffraction work in crystallization, one must seek to induce limited nucleation by adjustment of the physical and chemical properties of the system.

Association and variety. In forming crystals, molecules organize themselves through self-association to produce periodically repeating, three-dimensional arrays. Thus, it is necessary to facilitate positive molecular interactions while avoiding formation of precipitate, unspecific aggregates or phase separation. Proteins may crystallize under a wide spectrum of conditions and form many polymorphs. Thus, one should explore as many opportunities for crystallization as possible and explore the widest spectrum of biochemical, chemical and physical parameters.

Control and perfection. The ultimate value of any crystal is dependent on its perfection. Perturbations of the mother liquor are, in general, deleterious. Thus, crystallizing systems have to be maintained at an optimal state, without fluctuations or shock, until the crystals have matured. Crystallization conditions should be such as to favour crystal perfection, to minimize defects and high mosaicity of the growing crystals, and to minimize internal stress and incorporation of impurities. Predictions from crystal-growth theories may help to define such conditions (Chernov, 2003).

Impurities and preservation. Impurities can contribute to a failure to nucleate or to grow quality crystals. Thus, one must discourage their presence in the mother liquor and their incorporation into the lattice. Protein crystals may degrade and lose diffraction quality upon ageing. Thus, once grown, crystals may be stabilized by temperature change, addition of more crystallant or by some other suitable alteration in the mother liquor.

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

4.1.7. The future of protein crystal growth

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

Furthermore, the objectives of crystallization, the entities to be crystallized, will continue to become more challenging. In addition to membrane proteins that pose difficult problems because of their solubility (see Chapter 4.2), interest has increasingly turned towards the solutions of the structures of RNA, glycoproteins (Chang *et al.*, 2007), lipoproteins, and larger protein or protein–nucleic acid complexes and assemblies. It is unlikely that crystals with unit cells much above 1200 Å can be solved with current X-ray technologies, but even those assemblies, such as large icosahedral viruses, that do yield crystals amenable to analysis are remarkably fragile in a mechanical sense, and the large unit-cell size requires that the crystals greatly exceed the small sizes of conventional protein crystals. Additional problems will arise from proteins conjugated with other entities of significant size such as lipids and oligosaccharides, which are often disordered, and with proteins that are unstructured, in whole or in part.

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