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.