4.3. Application of protein engineering to improve crystal properties

BY D. R. DAVIES AND A. BURGESS HICKMAN

4.3.1. Introduction

There is accelerating use of protein engineering by protein crystallographers. Site-directed mutations are being used for a variety of purposes, including solubilizing the protein, developing new crystal forms, providing sites for heavy-atom derivatives, constructing proteolysis-resistant mutants and enhancing the rate of crystallization. Traditionally, if the chosen protein failed to crystallize, a good strategy was to examine a homologous protein from a related species. Now, the crystallographer has a variety of tools for directly modifying the protein according to his or her choice. This is owing to the development of techniques that make it easy to produce a large number of mutant proteins in a timely manner (see Chapter 3.1).

The relevance to macromolecular crystallography of these mutational procedures rests on the assumption that the mutations do not produce conformation changes in the protein. It is often possible to measure the activity of the protein *in vitro* and, therefore, test directly whether mutation has affected the protein's properties. Several observations suggest that changes of a small number of surface residues can be tolerated without changing the three-dimensional structure of a protein. The work on haemoglobins demonstrated that mutant proteins generally have similar topologies to the wild type (Fermi & Perutz, 1981). The systematic study of T4 phage lysozyme mutants by the Matthews group (Matthews, 1993; Zhang *et al.*, 1995) has confirmed and significantly extended these studies and has provided a basis for mutant design. This work revealed that, for monomeric proteins, 'Substitutions of solvent-exposed amino acids on the surfaces of proteins are seen to have little if any effect on protein stability or structure, leading to the view that it is the rigid parts of proteins that are critical for folding and stability' (Matthews, 1993). It was also concluded that point mutants do not interfere with crystallization unless they affect crystal contacts. The corollary from this is that if the topology of the protein is known from sequence homology with a known structure, the residues that are likely to be located on the surface can be defined and will provide suitable targets for mutation. Fortunately, even in the absence of such information, it is usually possible to make an informed prediction of which residues (generally charged or polar) will, with reasonable probability, be found on the surface.

Here, we shall outline some of the procedures that have been used successfully in protein crystallography. We have tried to provide representative examples of the variety of techniques and creative approaches that have been used, rather than attempting to assemble a comprehensive review of the field. The identification of appropriate references is a somewhat unreliable process, because information regarding these attempts is usually buried in texts; we apologize in advance for any significant omissions.

There have been several reviews on the general topic of the application of protein engineering to crystallography. An overview of the subject is provided by D'Arcy (1994), while Price & Nagai (1995) 'focus on strategies either to obtain crystals with good diffraction properties or to improve existing crystals through protein engineering'. In addition to attempts at a rational approach to protein engineering, it is worth emphasizing the role of serendipity in achieving the goal of diffraction-quality crystals. One example is given by the structure of GroEL (Braig *et al.*, 1994), where better crystals were obtained by the accidental introduction of a double mutation, which arose from a polymerase error during the cloning process. The second example is provided by the search for crystals of the complex between the U1A spliceosomal protein and its RNA hairpin substrate (Oubridge *et al.*, 1995). Initially, only poorly diffracting crystals (7–8 Å) could be obtained, which were similar in morphology to those of the protein alone. A series of mutations were made, designed to improve the crystal contacts, but the end result was a new crystal form that diffracted to 1.7 Å.

Dasgupta *et al.* (1997), in an informative review, have compared the contacts formed between molecules in crystal lattices and in protein oligomerization. They found that there are more polar interactions in crystal contacts, while oligomer contacts favour aromatic residues and methionine. Arginine is the only residue prominent in both, and for a protein that is difficult to crystallize, they recommend replacing lysine with arginine or glutamine. Carugo & Argos (1997) also examined crystal-packing contacts between protein molecules and compared these with contacts formed in oligomers. They observed that the area of the crystal contacts is generally smaller, but that the amino-acid composition of the contacts is indistinguishable from that of the solvent-accessible surface of the protein and is dramatically different from that observed in oligomer interfaces.

4.3.2. Improving solubility

Frequently, a protein is so insoluble that there is only a small probability of direct crystallization. Not only does the limited amount of protein hinder crystallization, but the departure from optimal solubility conditions by the addition of almost any crystallization medium frequently results in rapid precipitation of the protein from solution. When this happens, it is sometimes possible to find surface mutations that enhance solubility. Two strategies have been successfully applied, depending on whether or not the overall topology is known.

An early investigation of the effects of surface mutations (McElroy *et al.*, 1992) involved the crystallization of human thymidylate synthase, where the *Escherichia coli* enzyme structure was known, but the human enzyme could only be crystallized in an apo form unsuitable for studying inhibitors owing to disorder in the active site. The effect of surface mutations was systematically explored by making 12 mutations in 11 positions, and it was found that some of the mutations dramatically changed the protein solubility. Some of the mutant proteins were easier to crystallize than the wild type, and, furthermore, three crystal forms were obtained that differed from that of the wild type.

A second example of the rational design of surface mutations based on prior knowledge of the structure of a related protein is demonstrated by the studies of the trimethoprim-resistant type S1 hydrofolate reductase (Dale *et al.*, 1994). This protein was rather insoluble and precipitated at concentrations greater than 2 mg ml^{-1}. The authors changed four neutral, amide-containing side chains to carboxylates and examined the expressed proteins for improved solubility. Three of the four mutant proteins were more soluble than the wild-type protein, and a double mutant, Asn48 → Glu and Asn130 → Asp, was particularly soluble; this mutant protein crystallized in thick plates, ultimately enabling the structure to be determined.

In the absence of any knowledge of the structure, more heroic procedures are required, as illustrated by the crystallization of the HIV-1 integrase catalytic domain (residues 50–212). This domain had been a focus of intensive crystallization attempts, which were hindered by the low solubility of the protein. The strategy used was to replace all the single hydrophobic residues with lysine and to replace groups of adjacent hydrophobic amino acids with alanines (Jenkins *et al.*, 1995). A simple assay for improved solubility based on the overexpression of the protein was employed, which did not require isolating the purified protein; cell lysis followed by

references