International Tables for Crystallography (2006). Vol. F, Section 4.3.2, pp. 100–101.

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

centrifugation and SDS–PAGE analysis were used to determine which mutant proteins were sufficiently soluble to appear in the supernatant. The initial application of this method to 30 mutants resulted in one, Phe185 → Lys, which was soluble and which was subsequently crystallized and its structure determined (Dyda *et al.*, 1994). The protein formed a dimer, and the mutated residue was observed at the periphery of the dimer interface where the introduced lysine formed a hydrogen bond with a backbone atom of the second subunit, an interaction not possible for the unmutated protein. The position of the mutation was remote from the active site, and the physiological relevance of the observed dimer interaction was later confirmed by studies on an avian retroviral integrase (Bujacz *et al.*, 1995).

In further mutational work, it was observed that the HIV-1 integrase core-domain mutant suffered from an inability to bind to Mg^{2+} in the crystal, despite the evidence that Mg^{2+} or Mn^{2+} is needed for activity. The original crystallization took place using cacodylate as a buffer and also had dithiothreitol present in the crystallization medium. Under these conditions, cacodylate can react with –SH groups, and there were two cysteines in the structure that were clearly bonded to arsenic atoms. To avoid this problem, attempts were made to crystallize in the absence of cacodylate. These were successful only when a second mutation, designed to improve solubility, was introduced, Trp131 → Glu (Jenkins *et al.*, 1995; Goldgur *et al.*, 1998). The use of this mutant led to crystals that had the desired property of binding to Mg^{2+} and, in addition, revealed the conformation of a flexible loop that had not been previously defined.

4.3.3. Use of fusion proteins

Fusion proteins have been frequently used in a variety of applications (reviewed by Nilsson *et al.*, 1992), such as preventing proteolysis, changing solubility and increasing stability. They have also been used – although less frequently – for crystallization. The disadvantage in the context of crystallography is that the length and flexibility of the linker chain often introduce mobility of one protein domain relative to the other, which can impede, rather than enhance, crystallization.

Donahue *et al.* (1994) were able to determine the three-dimensional structure of the 14 residues representing the platelet integrin recognition segment of the fibrinogen γ chain by constructing a fusion protein with lysozyme, which was then crystallized from ammonium sulfate. Kuge *et al.* (1997) successfully obtained crystals of a fusion protein consisting of glutathione S-transferase (GST) and the DNA-binding domain (residues 16–115) of the DNA replication-related element-binding factor, DREF, under crystallization conditions similar to those used for GST alone.

In many cases, a fusion protein is made to aid in the isolation and purification of the target protein, and the intervening linker is engineered to contain a proteolytically susceptible sequence. However, subsequent cleavage to separate the two proteins can introduce the possibility of accidental proteolysis elsewhere in the protein. This was observed with a fusion protein between thioredoxin and VanH, a D-lactate dehydrogenase, where attempts to remove the carrier resulted in non-specific proteolysis and VanH inactivation (Stoll *et al.*, 1998). Fortunately, cleavage was unnecessary, and conditions were identified under which the authors were able to crystallize the intact fusion protein.

A novel approach to crystallizing membrane proteins is provided by the fusion protein in which cytochrome b_{562} was inserted into a central cytoplasmic loop of the lactose permease from *Escherichia coli* (Privé *et al.*, 1994). Although crystals have not yet been reported, the cytochrome attachment provides increased solubility together with the ability to use the red colour to assay the progress of crystallization trials.

4.3.4. Mutations to accelerate crystallization

A common problem encountered in crystallization is that certain crystals appear late and grow slowly. Sometimes, the slow appearance of crystals is the result of proteolytic processing, but often the reasons are not apparent. There are several examples where protein engineering has resulted in an increase in the rate of crystallization.

Heinz & Matthews (1994) explored the crystallization of T4 phage lysozyme using a strategy based on their understanding of the structure of the enzyme and its crystallization properties. The crystallization of the wild-type protein required the presence of β-mercaptoethanol (BME), an additive which could not be replaced with dithiothreitol. It had also been observed that the oxidized form of BME, hydroxyethyl disulfide, was trapped in the dimer interface between two lysozyme molecules (Bell *et al.*, 1991). It was hypothesized that dimer formation might be the rate-limiting step in crystallization, so dimerization was enhanced by cross-linking two monomers by disulfide-bridge formation. Applying rules developed for constructing S–S bridges, they selected Asn68 → Cys and Ala93 → Cys. In the presence of oxidized BME, the rate of crystallization of these mutant proteins was substantially increased, with crystals reaching full size in two days, in contrast to two weeks for the unmutated protein. Furthermore, they were able to crystallize a previously uncrystallizable mutant. Unexpectedly, however, the dimer formed in this way was lacking in activity, despite the selection of mutation sites on the opposite side of the molecule to the active site.

Mittl *et al.* (1994) wanted to improve the resolution of their crystals of glutathione reductase. From the 3 Å map, they could see a hole in the crystal packing where two molecules within 6 Å of each other just missed forming a crystal contact; they filled this hole by mutating Ala90 → Tyr and Ala86 → His. This designed double mutant did not improve the resolution, but did increase the rate of crystallization 40-fold, *i.e.*, initial crystals were observed within 1.5 h *versus* 60 h for the wild-type enzyme.

4.3.5. Mutations to improve diffraction quality

Another commonly encountered situation is that crystals can be obtained, but they diffract poorly. There are many examples where investigators have applied protein engineering in an effort to overcome this problem.

Proteolytic trimming is one possible approach to improving diffraction quality. For example, Zhang *et al.* (1997) attempted to crystallize a homodimer of the C2 domain of adenylyl cyclase. The initial crystals diffracted poorly (to 3.8 Å), so the effects of limited proteolysis with chymotrypsin, trypsin, GluC and LysC were investigated. A stable cleavage product was observed with GluC, approximately 4 kDa smaller than the full-length protein, but in order to avoid minor products formed during GluC proteolysis, the cleavage site was re-engineered as a thrombin site. Since there was already an atypical thrombin site seven residues from this site, proteolysis resulted in a smaller protein than expected; nevertheless, this modified protein crystallized readily and diffracted to 2.2 Å.

The importance of applying a variety of strategies to improve crystal quality is exemplified by the work of Oubridge *et al.* (1995), in which initial attempts to crystallize wild-type U1A complexed with RNA hairpins resulted in cubic crystals diffracting to 7–8 Å. By mutating surface residues, changing the N-terminal sequence to reduce heterogeneity and varying the sequence of the RNA hairpin, a new crystal form which diffracted to 1.7 Å was ultimately crystallized. However, in order to achieve this result, many variants were constructed and examined. For the protein, mutations were introduced which it was believed (incorrectly) would affect the crystal packing, and which were selected based on the observed similarity of space group and cell dimensions between crystals of