the complex and those of the protein alone. One of these mutations, together with an additional mutation resulting from a polymerase chain reaction (PCR) artefact, yielded crystals that diffracted to 3.5 Å. Additional variation of the length and composition of the RNA hairpin led to a new crystal form of this double mutant in the presence of a 21-base RNA that diffracted to 1.7 Å. A further mutation, Ser29 → Cys, was made to allow mercury binding (see Section 4.3.8), also resulting in crystals that diffracted to 1.7 Å. The authors commented that 'If any principle emerges from this study, it is that the key to success is not in concentrating on exhausting any one approach, but in the diversity of approaches used.'

The relevance of this comment is illustrated by the attempts of Scott *et al.* (1998) to obtain diffraction-quality crystals of the I-A^d class II major histocompatibility complex (MHC) protein. This complex exists *in vivo* as a heterodimer, but expression in recombinant form did not lead to satisfactory dimer formation. A leucine zipper peptide was therefore added to each chain to enhance dimerization. Attempts to crystallize this heterodimer after removal of the leucine zippers and in the presence of bound peptides led to poorly diffracting crystals. To enhance the affinity of an ovalbumin peptide for the MHC dimer, the peptide was then attached through a six-residue linker to the N-terminus of the chain, tethering it in the vicinity of the binding site. This construct, in conjunction with removal of the leucine zippers from the heterodimer, resulted in crystals that diffracted to 2.6 Å.

4.3.6. Avoiding protein heterogeneity

Protein heterogeneity can arise from many sources, including proteolysis, oxidation and post-translational modifications, and can have a severe effect on crystal quality or can prevent crystallization altogether. Limited proteolysis has frequently been used to modify proteins for crystallization, in order to avoid heterogeneity from proteolysis occurring during expression and to remove relatively unstructured regions that might hinder crystallization. Some examples are given below.

Windsor *et al.* (1996) crystallized a complex of interferon γ with the extracellular domain of the interferon γ cell surface receptor. To obtain satisfactory crystals, it was necessary to re-engineer the receptor with an eight-amino-acid residue deletion at the N-terminus to avoid the observed heterogeneity owing to proteolysis, since 2–10% of the purified protein was cleaved during expression.

Crucial to the structure determination of the complex of transducin-α bound to GTPγS (Noel *et al.*, 1993) was the systematic examination of proteolysis of the intact protein (Mazzoni *et al.*, 1991). This work revealed a cluster of protease-sensitive sites near residues Lys17–Lys25. Homogeneous material consisting of residues 26–350 of activated rod transducin, $G_{t\alpha}$, was obtained by proteolysis of the full-length protein with endo-proteinase LysC; the truncated protein was subsequently used to solve the structure.

Hickman *et al.* (1997) identified a site near the C-terminus of HIV-1 integrase that was susceptible to proteolytic cleavage during protein expression, resulting in severe protein heterogeneity in which up to 30% of the purified protein was cleaved. The proteolysis site was identified by mass spectrometry analysis, and several point mutations on either side of this site were made and evaluated for their effect on proteolysis. Substitution of either Gly or Lys for Arg284 eliminated the protease sensitivity, yielding homogeneous material.

Some proteins have surface cysteines that are susceptible to oxidation and can be adventitiously cross-linked *via* a disulfide bridge that does not exist in the native protein. If there are relatively few cysteines, this problem may be circumvented by mutating the individual cysteines to determine which ones are responsible. Conversely, cysteines can be introduced into proteins to enhance the binding of interacting molecules (see also Section 4.3.8). An elegant example of the latter case is provided by the recent structure of HIV-1 reverse transcriptase (Huang *et al.*, 1998), which was mutated to introduce a cysteine in a position near the known binding side of the double-stranded DNA substrate. Using an oligonucleotide with a modified base that contained a free thiol group, cross-links were specifically introduced between the protein and the DNA; this covalently linked complex was used to obtain crystals that contained the incoming nucleoside triphosphate, a crystallographic problem that had defied other solutions.

Post-translationally modified proteins, such as glycoproteins, present some of the most difficult problems in X-ray crystallography, since the carbohydrate side chains are usually flexible and often heterogeneous. In some cases, enzymes can be used to trim the carbohydrate and produce a protein suitable for crystallization. Alternatively, the protein sequence can be altered so that unwanted glycosylation does not occur. A combination of approaches was used by Kwong *et al.* (1998) to determine the structure of the HIV-1 envelope glycoprotein, gp120, a protein which is extensively modified *in vivo*. The N- and C-termini were truncated, 90% of the carbohydrate was removed by deglycosylation and two large, flexible loops of the protein were replaced by tripeptides. The resulting simplified version of the glycoprotein retained its ability to bind the CD4 receptor, and crystals were ultimately obtained of a ternary complex of the envelope glycoprotein, a two-domain fragment of CD4 and an antibody Fab.

Occasionally, an mRNA sequence will fortuitously result in a false initiation of translation, resulting in a truncated form co-purifying with the intended protein. In attempting to crystallize a trimethoprim-resistant form of dihydrofolate reductase, Dale *et al.* (1994) observed that a fragment of the protein was being expressed through false initiation of translation, beginning at Ala43. They also found most of the protein in inclusion bodies and recovery was poor. They noticed that there was a putative Shine–Dalgarno sequence ten nucleotides up from the AUG codon of Met42, which could result in the expression of a smaller protein. They replaced the middle base of the Shine–Dalgarno sequence, GGGAA, with GGCAA and removed unusual codons from the first 18 amino acids. These two changes resulted in a 20-fold increase in expression level, together with removal of the contaminating fragment. Similar heterogeneity problems owing to translation initiation at an internal Shine–Dalgarno sequence upstream of Met50 were observed during expression of full-length recombinant HIV-1 integrase and were also resolved by altering the DNA to eliminate the Shine–Dalgarno sequence without changing the sequence of encoded amino acids (Hizi & Hughes, 1988).

4.3.7. Engineering crystal contacts to enhance crystallization in a particular crystal form

It is often the case that the structure of some related form of a protein is known, but the protein of interest crystallizes in a different space group. There have been attempts to use this knowledge to obtain crystals in a form that could be readily analysed. However, it may not be necessary to resort to molecular engineering approaches, since molecular replacement methods can often be successfully applied to determine the protein structure.

In one of the first applications of protein engineering to obtain crystals, Lawson *et al.* (1991) reported the crystal structure of ferritin H. Ferritin has two types of chains, H and L; the structure of rat L ferritin was known. Despite high sequence identity to L ferritin, human recombinant H ferritin did not crystallize satisfactorily. To obtain the structure of a human H ferritin homopolymer, the sequence in the subunit interface was modified to give crystals that were isomorphous with the rat L ferritin. The

mutation Lys86 → Gln was introduced, which enabled metal bridge contacts to form, resulting in crystals that diffracted to 1.9 Å. Although the mutant was designed to crystallize from $CdSO_4$, it did not do so. Rather, $CaCl_2$ gave large crystals which were isomorphous with rat and horse L ferritin crystals. In these latter crystals, Ca^{2+} is coordinated between Asp84 and Gln86, providing the rationale for the mutation.

4.3.8. Engineering heavy-atom sites

Another application of protein engineering to crystallography involves the mutation of wild-type residues to cysteines, thus creating potential heavy-atom binding sites (reviewed by Price & Nagai, 1995). This was first systematically investigated by Sun *et al.* (1987), who made five cysteine mutants of T4 phage lysozyme. They demonstrated that modification of the protein usually, but not always, introduced differences in isomorphism with the wild type. When the lack of isomorphism was not large, its effects could be reduced by comparing the mutant crystal with and without heavy atoms. The authors suggested that serine would be an attractive site for substitution, since it is structurally similar to cysteine and has a high probability of being on the protein surface.

However, in the absence of a known structure, the choice of a successful cysteine substitution site involves some luck. A general sense of the success rate of this approach can be gauged from three studies. Martinez *et al.* (1992) prepared 14 mutant forms of *Fusarium solani* cutinase in which each serine was replaced by cysteine. Four of these gave isomorphous crystals and led to useful derivatives with mercuric acetate. Nagai *et al.* (1990), as part of an attempt to crystallize a domain of the U1 small RNA-binding protein, engineered ten mutants to give cysteine replacements for polar side chains; of these, four yielded mercury derivatives that were isomorphous with the native protein. Finally, in a study of the ribosomal protein L9 (Hoffman *et al.*, 1994), eight cysteine mutants were prepared, but only one crystallized well and was isomorphous with wild-type crystals. In addition, two methionine mutants were engineered, and both crystallized isomorphously to the wild type (discussed below).

When the protein being examined belongs to a homologous superfamily, the sequences can be analysed to provide likely sites. For example, in a structure determination of ribosomal protein L6 (Golden *et al.*, 1993), a heavy-atom binding site was constructed with the mutant Val124 → Cys. This site was chosen because it is a cysteine residue in other L6 proteins. The mutant protein crystallized with the same space group and cell dimensions as the wild type. It was reacted with parachloromercuribenzoate to provide a heavy-atom derivative. However, at high resolution, the crystals were not isomorphous with the wild type, so a derivative was prepared by replacing the two methionines with selenomethionine, illustrating a second approach to engineering heavy-atom sites, discussed below. The structure was ultimately solved with a combination of multiple isomorphous replacement, anomalous scattering and solvent flattening. The same approach was used for OmpR (Martínez-Hackert *et al.*, 1996), in which cysteine residues were similarly engineered into positions determined by comparison with other proteins of the superfamily.

The increasing popularity of multiwavelength anomalous dispersion (Karle, 1980; Hendrickson, 1991; Hendrickson & Ogata, 1997) for phase determination, using selenomethionine (Se-Met) in place of methionine, has led to the engineering of proteins to create selenomethionine sites. The original substitution of Se-Met for Met was described by Cowie & Cohen (1957). The potential for crystallography was demonstrated for thioredoxin (Hendrickson *et al.*, 1990) and was used to solve the structure of ribonuclease H (Yang, Hendrickson, Crouch & Satow, 1990; Yang, Hendrickson, Kalman & Crouch, 1990). Methods for preparing Se-Met-substituted proteins are reviewed by Doublié (1997). Budisa *et al.* (1995) have also reported successful incorporation of telluromethionine into a protein, although this approach is not yet routine.

Since the frequency of methionines in proteins is about 1 in 60 (Dayhoff, 1978; Hendrickson *et al.*, 1990), it is not unusual for the protein being studied to contain no methionine residues. A number of investigators have introduced methionine into a protein sequence so that it can subsequently be replaced by Se-Met. These include Leahy *et al.* (1994), who crystallized domains FN7–10 of human fibronectin. Attempts to obtain mercury-soaked diffraction-quality crystals of FN7–10P, a double mutant that resulted from a (yet another!) PCR error, were unsuccessful, as were attempts to solve the structure by molecular replacement. They therefore prepared a mutant in which three residues (two leucines and one isoleucine) were substituted with methionine. Diffraction-quality data were subsequently obtained from the Se-Met derivatives.

Sometimes the protein cannot be crystallized satisfactorily in the Se-Met form, and further modification is required. The Se-Met derivative of UmuD′, an *Escherichia coli* SOS response protein, did not crystallize under conditions that gave native crystals (Peat *et al.*, 1996). Comparison with homologous proteins indicated that two of the Met sites were either conserved or replaced by hydrophobic residues. The third site, Met138, was variable and often replaced by a polar residue. The authors hypothesized that this methionine might, therefore, be surface-exposed, rendering the Se-Met version highly susceptible to oxidation and heterogeneity. When this penultimate Met was mutated to Met138 → Val or Met138 → Thr, these mutant proteins yielded crystals both with and without introduction of Se-Met.

As a final note, it is worth returning to the study involving the crystallization of the U1A/RNA complex (Oubridge *et al.*, 1995), in which the authors comment: ‘In retrospect it is clear that too much was assumed about interactions within crystals, and that the “design” of good crystals *per se* was not feasible . . .. It may be that almost anything can be crystallized to give well ordered crystals as long as enough constructs are tried; however, one only knows the right condition when the crystals are obtained.’

Acknowledgements

We gratefully acknowledge the kind assistance of B. Cudney, A. D’Arcy, J. Ladner and A. McPherson in the early stages of researching this article, and A. McPherson and K. Nagai for reviewing the manuscript. We also thank A. Stock and S. Hughes for their help in coordinating references and for sharing their manuscript prior to publication.