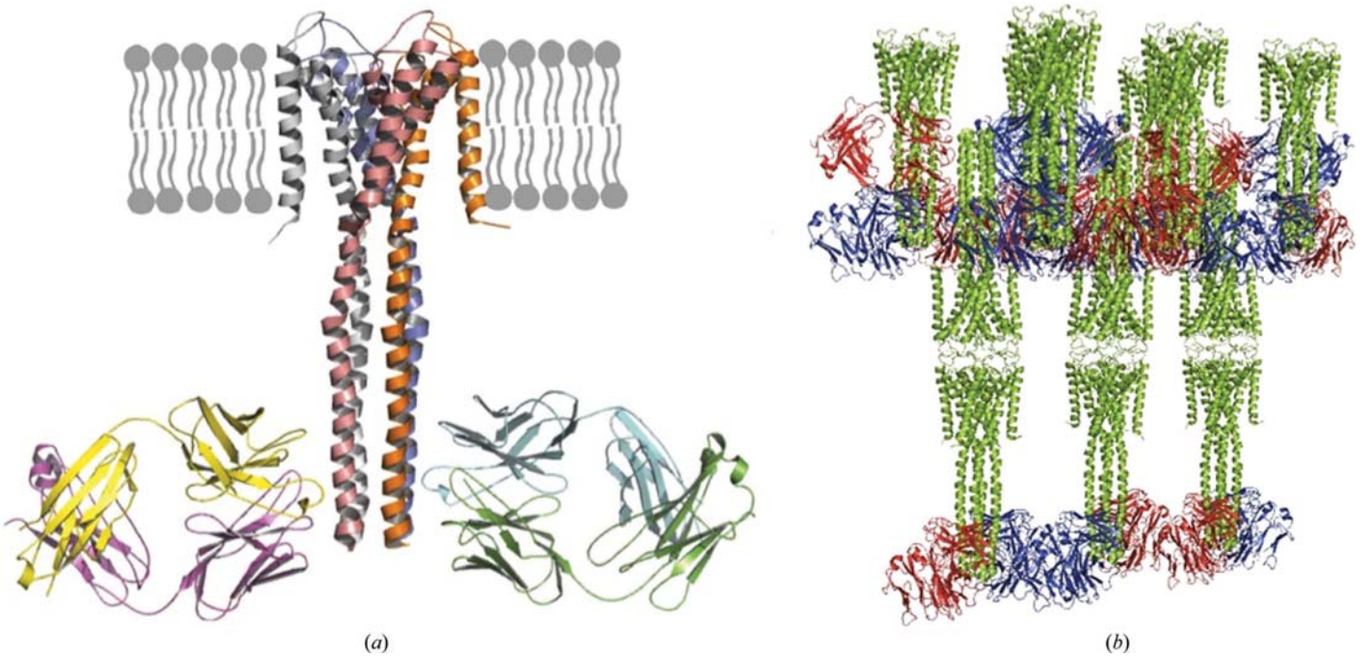

Figure 4.3.6.1
Phage-display-generated Fab fragments as crystallization chaperones: the structure of the KcsA channel in the closed conformation in complex with a synthetic Fab (PDB code 3eff; Uysal *et al.*, 2009). (*a*) A diagram showing how the Fab binds to the cytosolic portion in reference to the transmembrane domain. (*b*) The crystal structure of the complex showing how the synthetic Fab molecules mediate the major crystal contacts. (Figure courtesy of Dr Anthony Kossiakoff, University of Chicago.)

and expensive. In principle, a more efficient approach is to carry out *in vitro* selection of Fab fragments using phage display (Lee *et al.*, 2004) or ribosome display (Lipovsek & Pluckthun, 2004). However, since a typical antibody–antigen interface involves ∼30 amino acids, the total number of possible sequences of a given template Fab significantly exceeds the available combinatorial libraries. Consequently, traditional phage-display libraries greatly diminish diversity at the mutated sites, which explains why synthetic antibodies were initially weaker binders than natural ones (Hawkins *et al.*, 1992; Koide, 2009). This problem was successfully overcome using a different type of phage-display library based on a 'reduced genetic code' and comprised of only a few amino acids, *e.g.* four, which produces high-affinity binders based on a single Fab scaffold (Fellouse *et al.*, 2004; Lee *et al.*, 2004). In contrast to natural antibodies, such synthetic Fab fragments can be generated against unique conformations, complexes or weak antigens such as RNA. Among recent examples are the crystallization and structure determination of the closed form of the full-length KcsA potassium channel with its cognate synthetic Fab (Uysal *et al.*, 2009; Fig. 4.3.6.1) and the crystallographic study of the ΔC209 P4-P6 domain of the *Tetrahymena* group I intron, a structured RNA molecule (Ye *et al.*, 2008).

The *in vitro* display methods also allow the engineering of non-antibody scaffolds as alternative protein binders and crystallization chaperones (Koide, 2009). For example, a fibronectin type III domain (FN3) scaffold was successfully used to generate binders with a reduced genetic code phage-display library (Koide, Gilbreth *et al.*, 2007; Gilbreth *et al.*, 2008). A similar approach was used for DARPins, *i.e.* designed ankyrin-repeat proteins (Sennhauser & Grütter, 2008), based on ribosome-display selection (Lipovsek & Pluckthun, 2004; Sennhauser & Grütter, 2008). Several new protein structures have been solved as complexes with DARPin chaperones, including polo-like kinase 1 (Bandeiras *et al.*, 2008), the trimeric integral membrane multidrug transporter AcrB (Sennhauser *et al.*, 2007) and the receptor-binding protein (RBP, the BppL trimer) of the baseplate complex of the lactococcal phage TP901-1 (Veesler *et al.*, 2009).

4.3.7. Removal of post-translational modifications

A number of proteins undergo post-translational modifications which can adversely affect crystallization. By far the most ubiquitous is N- and O-glycosylation, primarily of membrane-associated, secreted and lysosomal proteins. In a number of cases successful crystallization of glycoproteins purified from natural sources has been reported and carbohydrate groups have often been found to be ordered and occasionally sequestered between the protein molecules, thus even contributing in a positive way to crystallization (Mark *et al.*, 2003; Aleshin *et al.*, 1994). In general terms, however, the flexible and heterogeneous carbohydrate moieties, particularly the oligosaccharides linked by N-glycosylation, can account for a significant fraction of the surface area of the protein and can therefore be detrimental to crystallization. The preparation of recombinant proteins in *E. coli* eliminates these post-translational modifications and may sometimes solve the problem (Mohanty *et al.*, 2009), but N-glycosylation is often required for appropriate folding and solubility, so this approach is not always possible. However, if a eukaryotic expression system is a necessity, the problem can often be resolved by mutating the asparagines within the relevant glycosylation motifs (Asn-*X*-Thr/Ser), *e.g.* to aspartates, as was performed in the case of the extracellular domain of the metabotropic glutamate receptor expressed in insect cells (Muto *et al.*, 2009), or to glutamines, as was performed for the human testis angiotensin-converting enzyme (Gordon *et al.*, 2003). Alternatively, glycosylation at these sites can be eliminated by mutation of the Thr/Ser residues in the glycosylation motif to alanine or other amino acids, as described for rat cathepsin B (Lee *et al.*, 1990), or valine, as was the case with the Ebola virus glycoprotein (Lee *et al.*, 2008, 2009). Similarly, potentially glycosylated threonines or serines in O-

glycosylated glycoproteins can be mutated to other amino acids (Horan *et al.*, 1998) to avoid or reduce glycosylation.

Other post-translational modifications occur less frequently. Prenylation and N-myristoylation can occur at the C- and N-termini, respectively. Expression in *E. coli*, often using truncated versions of target proteins, is a common remedy (Pai *et al.*, 1990).

4.3.8. Stabilization of protein targets

There is currently no clear consensus regarding a possible correlation between the thermostability of a protein and its propensity to form crystals. It is often assumed on the basis of somewhat anecdotal evidence that thermostable proteins are more readily crystallizable and therefore if a specific target protein is recalcitrant to crystallization then a homologue from a thermophilic organism should instead be used. In some cases, low thermostability may correlate with the presence of unstructured loops or termini and consequently the construct-optimization strategies, as described above, are likely to yield a more crystallizable variant with a concomitantly increased stability. For example, a study of MAPKAP kinase 2 showed that truncated variants with increased thermal stability also showed higher crystallization propensity (Malawski *et al.*, 2006). However, it is uncertain whether the relationship is causal or serendipitous. A recent analysis of large-scale data from a structural genomics project showed that when partly or fully unfolded proteins and hyperstable proteins (with melting temperatures T_m of greater than 90 °C) are excluded from comparisons, thermostability *per se* does not correlate with propensity for crystallization (Price *et al.*, 2009). Consequently, it appears that in a general case of a well behaving protein, attempts to increase thermostability by site-directed mutagenesis may not necessarily yield variants with enhanced crystallization properties, although when the prospective crystallization target is inherently unstable, engineering more stable variants may be helpful. In fact, this strategy has been successfully used for membrane proteins, which are often unstable in detergent environments. The first structure of a recombinant G-protein-coupled receptor (GPCR), *i.e.* bovine rhodopsin in complex with 11-*cis* retinal, was obtained using a thermostable variant with an engineered disulfide bond (Standfuss *et al.*, 2007). In the more recent case of the turkey adrenergic β2 receptor, 318 variants were screened and six mutations were identified that increased thermostability. A variant containing all six mutations had an apparent T_m that was 21 °C higher than that of the native protein in dodecylmaltoside (DDM), was more stable in short-chain detergents and was successfully crystallized (Warne *et al.*, 2008, 2009). The effort required for such vast screening is substantial, but it appears that within protein families (such as GPCRs) the pattern of mutations enhancing thermostability is preserved, thus making it possible to transfer the mutations from one family member to another (Serrano-Vega & Tate, 2009).

Finally, it should be noted that protein thermostability is strongly dependent on solution parameters such as the ionic strength. In the case of ribonuclease SA, the melting temperature (T_m) increased from ∼40 to ∼60 °C on transfer from pure 50 m*M* diglycine buffer to 0.9 *M* ammonium sulfate (Trevino *et al.*, 2007). Similarly, binding small molecules, either physiological or non-physiological ligands, typically promote stability (Matulis *et al.*, 2005). High-throughput screening methods have been developed to aid in screening for conditions and ligands that enhance stability (Vedadi *et al.*, 2006; Mezzasalma *et al.*, 2007) and in a general case this strategy appears to hold better promise than attempts to engineer higher stability through mutagenesis.

4.3.9. Surface-entropy reduction (SER)

For well ordered intrinsically stable proteins that show none of the problems addressed above, the propensity of the molecules to associate together and form crystals mediated by weak but specific interactions is ultimately defined by the physical chemistry and topology of the molecular surface. As already pointed out, large flexible amino acids on the surface, such as Lys, Glu and Gln, constitute an impediment to intermolecular interactions and consequently to protein crystallization (Cieślik & Derewenda, 2009; Price *et al.*, 2009). In fact, it has been suggested that these residues and the 'entropy shield' that they form play a role in protein evolution which, given the high average concentration of proteins in cells, disfavours protein–protein interactions unless they are biologically functional (Doye, 2004). Thus, an intuitively obvious way to generate crystallizable variants is to replace selected large and surface-exposed residues with smaller residues such as alanine. This crystal-engineering strategy based on the surface-entropy reduction (SER) concept was extensively tested using as a model system the globular domain of the human Rho-specific guanine nucleotide-dissociation inhibitor (RhoGDI), which is recalcitrant to crystallization in its wild-type form owing to a high content of Lys and Glu residues, which constitute more than 20% of the sequence (Longenecker, Garrard *et al.*, 2001; Mateja *et al.*, 2002; Derewenda, 2004; Cooper *et al.*, 2007). These experiments established that in order to be most effective the SER strategy requires simultaneous mutations of clusters of two to three solvent-exposed high-entropy amino acids, typically Lys, Glu or Gln, located in close sequence proximity. These amino acids are replaced with alanine, although threonine and tyrosine, which is known to make a positive contribution at protein–protein interfaces such as antibody–antigen complexes (Fellouse *et al.*, 2006), can also be used (Cooper *et al.*, 2007). Engineered low-surface-entropy variants of RhoGDI produced new and unique crystal forms, many with superior diffraction quality when compared with the wild-type protein. Importantly, in the vast majority of these crystals the mutated surface patches mediated crystal contacts, suggesting that SER engineering directly drives crystallization in a rational fashion by creating suitable crystal-contact-forming interfaces. The general utility of the method was further established by the crystallization of several novel protein targets found to be recalcitrant to crystallization in their wild-type form (Longenecker, Lewis *et al.*, 2001; Derewenda *et al.*, 2004; Devedjiev *et al.*, 2004; Janda *et al.*, 2004).

SER is quickly becoming a method of choice for engineering crystallizable variants of both individual proteins and protein–protein complexes. To date (December 2009), there have already been more than 100 depositions made to the Protein Data Bank (Berman *et al.*, 2007) based on diffraction studies of crystals generated by SER and corresponding to 47 novel structures, seven novel protein–protein complexes, several studies of proteins in complexes with drug leads aimed at rational drug development and two membrane proteins. The current list of crystal structures obtained using SER crystals includes a number of cases of exceptionally high biological interest. For example, the EscJ protein from enteropathogenic *E. coli*, the oligomerization of which initiates assembly of the type III bacterial secretion system, was crystallized with three entropy-reducing mutations (E62A, K63A and E64A) forming a key contact in the crystal structure (Yip *et al.*, 2005). Likewise, the HIV CcmK4 capsid