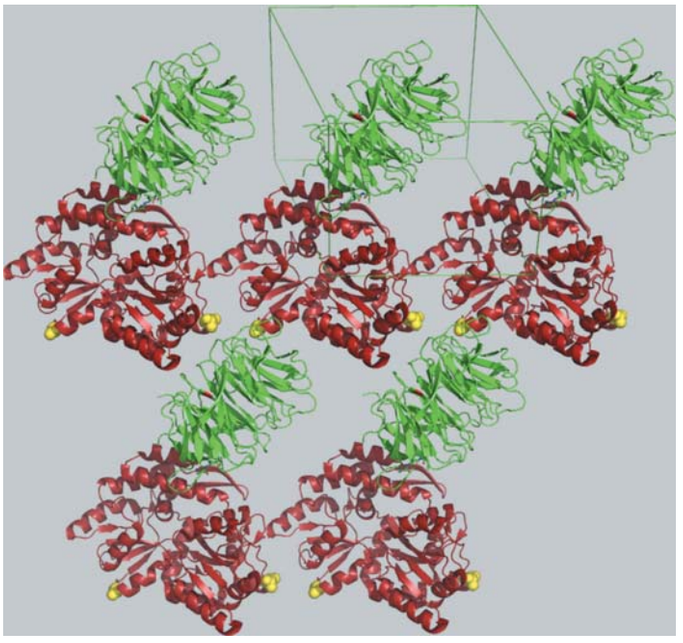

Figure 4.3.5.1
An example of the use of a fused carrier protein in crystallization: the crystal structure of the RACK1 protein (green) crystallized in fusion with an engineered variant of the maltose-binding protein (MBP; red); the major crystal contacts are mediated by MBP (PDB code 3dm0; Ullah *et al.*, 2008). The yellow spheres show alanines introduced by site-directed mutagenesis (see text for further details). Figs. 4.3.5.1 and 4.3.9.1 were generated using *PyMOL* (http://www.pymol.org).

downstream or upstream of the target protein and are often separated from it by a protease-sensitive linker sequence. They are cleaved proteolytically following expression and partial purification of the fusion protein and removed, leaving the isolated target ready for crystallization. However, in some cases the target protein may not be adequately soluble after cleavage or may resist crystallization. One of the possible solutions is to use the intact fusion protein in the crystallization screens in the hope that the carrier protein will both confer solubility on the construct and mediate crystal contacts. Not surprisingly, the canonical carrier proteins, all of which crystallize fairly easily on their own, constitute the obvious first choice. Using this strategy, the DNA-binding domain of DNA replication-related element-binding factor (DREF) was crystallized in fusion with *Escherichia coli* GST (Kuge *et al.*, 1997) and the U2AF homology motif (UHM) domain of splicing factor Puf60 was crystallized as a fusion with thioredoxin (Corsini *et al.*, 2008). A key problem limiting the utility of this technique is the inherent flexibility of a two-domain fusion protein, which is detrimental to its crystallizability. A possible solution to this problem is shortening the linker between the two proteins until a relatively rigid construct is identified (Smyth *et al.*, 2003). This approach was successfully pioneered for maltose-binding protein (MBP), which was used as a fusion chaperone to crystallize the human T-cell leukemia virus type 1 gp21 ectodomain fragment (Center *et al.*, 1998). The same strategy was employed in the crystallization of the ZP-N domain of ZP3 (Monne *et al.*, 2008), the islet amyloid polypeptide (IAPP; Wiltzius *et al.*, 2009) and the MATα1 homeodomain (Ke & Wolberger, 2003). Recently, a genetically modified version of MBP (see below) was used as an N-terminal fusion chaperone to crystallize the signal transduction regulator RACK1 from *Arabidopsis thaliana* (Ullah *et al.*, 2008; Fig. 4.3.5.1). Thus, MBP remains the most successful fusion chaperone for protein crystallization, even though the absolute number of proteins crystallized in this way is still limited.

In addition to the canonical fusion chaperones, which were originally designed as affinity tags, other carrier proteins can be used to assist crystallization. For example, a module made up of two sterile α motif (SAM) domains has been engineered to polymerize in response to a pH drop and was shown to drive the crystallization of 11 target proteins in a pilot study (Nauli *et al.*, 2007). In another example, barnase, a secreted ribonuclease from *Bacillus amyloliquefaciens*, was recently used as a carrier protein for crystallization of the disulfide-rich protein McoEeTI (Niemann *et al.*, 2006).

An alternative to N- or C-terminal fusions is an insertion fusion, in which a carrier protein is inserted into a loop in the sequence of a poorly soluble target. To date, this approach has exclusively been used in membrane-protein crystallization and was initially pioneered for the *E. coli* lactose permease, in which cytochrome b_{562}, flavodoxin and T4 lysozyme were tested as carrier proteins inserted into one of the loops (Privé *et al.*, 1994; Engel *et al.*, 2002). In this specific case none of these variants actually yielded useful crystals and the structure of lactose permease was eventually solved using crystals obtained using a variant containing the C154G mutation which stabilized a single conformation in complex with a lactose analogue (Abramson *et al.*, 2003). In contrast, a similar insertion fusion with T4 lysozyme replacing the third intracellular loop of the β2-adrenergic receptor was highly successful and yielded good-quality crystals that allowed structure determination at 2.4 Å resolution (Cherezov *et al.*, 2007; Rosenbaum *et al.*, 2007). This spectacular result attests to the potential of insertion-fusion proteins, but the method is not trivial as the constructs must be carefully evaluated for both structural and functional consequences of the insertion and a number of variants may have to be screened before a suitable one is identified.

4.3.6. Noncovalent crystallization chaperones

Noncovalent crystallization chaperones, *i.e.* engineered binding proteins that produce noncovalent complexes with target macromolecules, constitute an exciting alternative to fusion carrier proteins. Complexes with such chaperones often exhibit enhanced solubility and/or crystallizability in comparison to the isolated targets. The Fab and Fv fragments of antibodies are most commonly used for this purpose (Kovari *et al.*, 1995; Hunte & Michel, 2002; Prongay *et al.*, 1990; Ostermeier *et al.*, 1995; Jiang *et al.*, 2003; Dutzler *et al.*, 2003; Lee *et al.*, 2005). In its canonical version, the technique requires animal immunization with subsequent purification of hybridoma-derived antibodies and their proteolytic digestion to obtain pure homogeneous Fab fragments (Karpusas *et al.*, 2001; Kovari *et al.*, 1995). Alternatively, the Fab fragment can be directly sequenced and a synthetic gene can be used for *E. coli* expression, although this is not trivial owing to the presence of disulfides and two separate polypeptide chains in an Fab molecule. To overcome this bottleneck, a more efficient method of recombinant production of antibody fragments using mammalian HEK 293T has recently been proposed (Nettleship *et al.*, 2008). Another possibility is the use of so-called nanobodies, *i.e.* single-chain fragments derived from camelid antibodies (Koide, Tereshko *et al.*, 2007; Lam *et al.*, 2009; Korotkov *et al.*, 2009). However, this strategy requires immunization of camels or llamas, which is not technically easy.

Regardless of the specific strategy, the use of hybridoma technology and animal immunization is always time-consuming

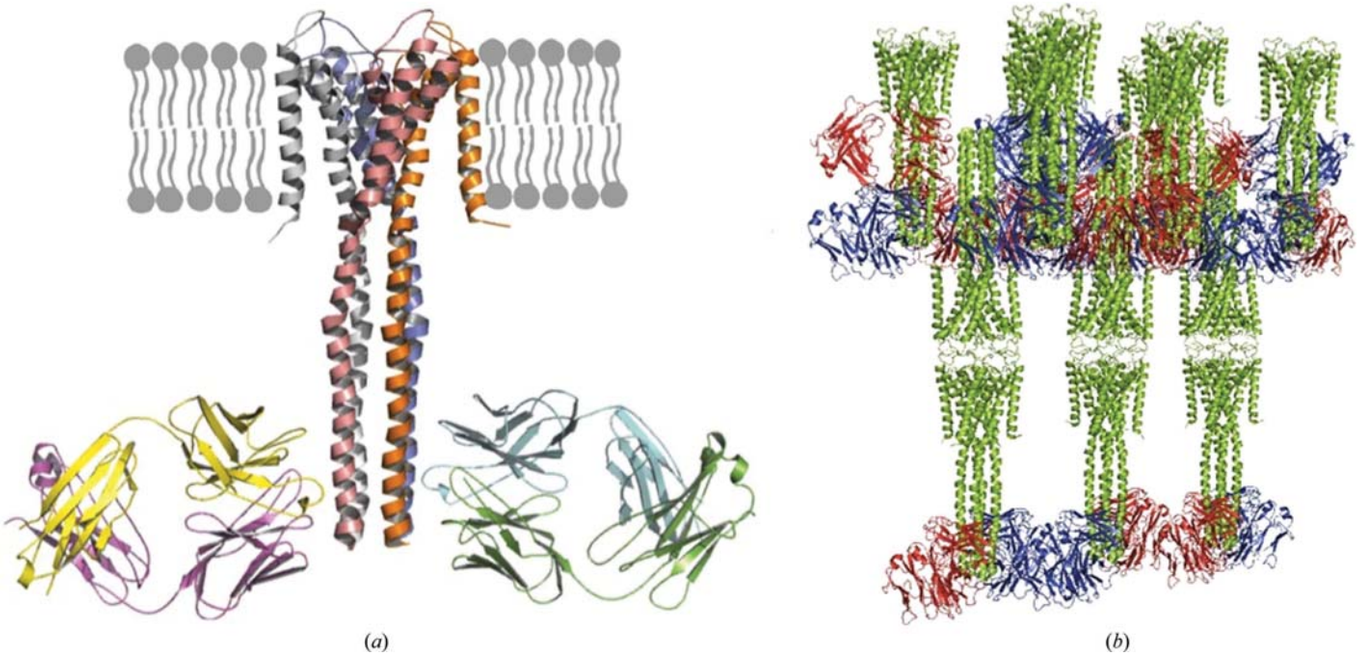

Figure 4.3.6.1
Phage-display-generated Fab fragments as crystallization chaperones: the structure of the KcsA channel in the closed conformation in complex with a synthetic Fab (PDB code 3eff; Uysal *et al.*, 2009). (*a*) A diagram showing how the Fab binds to the cytosolic portion in reference to the 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-