Crystal structures of fusion proteins with large-affinity tags

DOUGLAS R. SMYTH,1 MAREK K. MROZKIEWICZ,1 WILLIAM J. MCGRATH,1,3 PAWEL LISTWAN,1,2 AND BOSTJAN KOBE1,2

1Department of Biochemistry and Molecular Biology, Institute for Molecular Bioscience, and Special Research Centre for Functional and Applied Genomics and 2Cooperative Research Centre for Chronic Inflammatory Disease, University of Queensland, St Lucia, Queensland 4072, Australia
3Biology Department, Brookhaven National Laboratory, Upton, New York 11973, USA

(RECEIVED December 20, 2002; FINAL REVISION March 28, 2003; ACCEPTED April 17, 2003)

Abstract

The fusion of a protein of interest to a large-affinity tag, such as the maltose-binding protein (MBP), thioredoxin (TRX), or glutathione-S-transferase (GST), can be advantageous in terms of increased expression, enhanced solubility, protection from proteolysis, improved folding, and protein purification via affinity chromatography. Unfortunately, crystal growth is hindered by the conformational heterogeneity induced by the fusion tag, requiring that the tag is removed by a potentially problematic cleavage step. The first three crystal structures of fusion proteins with large-affinity tags have been reported recently. All three structures used a novel strategy to rigidly fuse the protein of interest to MBP via a short three- to five-amino acid spacer. This strategy has the potential to aid structure determination of proteins that present particular experimental challenges and are not conducive to more conventional crystallization strategies (e.g., membrane proteins). Structural genomics initiatives may also benefit from this approach as a way to crystallize problematic proteins of significant interest.

Keywords: Chimera; fusion protein; protein crystallization; protein expression; membrane proteins; molecular replacement; structural genomics; X-ray crystallography

Fusion (or chimeric) proteins are utilized in the forefront of protein science research for applications as diverse as biochemical purification, immunodetection, protein therapies, vaccine development, functional genomics, analysis of protein trafficking, and analyses of protein–protein or protein–nucleic acid interactions (Beckwith 2000). In structural biology, where milligram quantities of homogeneous protein sample are usually required, the most common utility of chimeras involves the separation of the fusion protein from the cell lysate using affinity chromatography. The most common affinity tags include the hexa-histidine (His-tag; Bornhorst and Falke 2000), Escherichia coli maltose-binding protein (MBP; Sachdev and Chirgwin 2000), Schistosoma japonicum glutathione-S-transferase (GST; Smith 2000), E. coli thioredoxin (TRX; LaVallie et al. 2000), and avidin/streptavidin Strep tags (Skerra and Schmidt 2000). Several other tags have also been developed (Stevens 2000).

To grow crystals of a protein of interest for X-ray diffraction studies, large-affinity tags, such as MBP or GST, are usually removed using site-specific proteolysis in the engineered linker region, followed by purification to separate the protein of interest from the affinity tag moiety and the protease. However, particular problems may be encountered during the cleavage step, including low yield, precipitation of the target protein, tedious optimization of cleavage conditions, high cost of proteases (e.g., factor Xa and enterokinase), or failure to recover active or structurally intact...
protein (Baneyx 1999). The alternative is to circumvent the cleavage and repurification steps and leave the affinity tag in place for crystallization trials. Unfortunately, this brings about a new challenge, as multidomain proteins are usually (1) less conducive to forming well-ordered, diffracting crystals, presumably due to the conformational heterogeneity allowed by the flexible linker region; and (2) too large for NMR studies. These problems explain why small affinity tags, such as the His-tag, are the tags of choice in structural biology, especially for high throughput/structural genomics approaches; they do not increase the size of the protein substantially, and cleavage of small tags is often not required to grow suitable crystals (Bucher et al. 2002).

Despite the considerations mentioned above, the first three-dimensional (3D) structures of fusion protein containing large-affinity tags have recently been reported (Kobe et al. 1999; Liu et al. 2001; Ke and Wolberger 2003). The aim of this review is to compare the use of small and large-affinity tags, focusing on their use in structural biology. In particular, we highlight the factors that have contributed to the successful crystallizations of MBP fusion proteins, and to present ideas for the potential utility of large-affinity tags for problematic structural targets. The advantages of large-affinity tags in facilitating the structural studies of small peptides have been discussed previously (Zhan et al. 2001).

**Expression of fusion proteins using small and large-affinity tags**

Recent estimates indicate that perhaps one-third to one-half of all prokaryotic proteins cannot be overexpressed in bacteria in soluble form using a His-tag (Edwards et al. 2000; Stevens 2000). Three recent high-throughput studies have indicated that this number is higher for eukaryotic proteins (Braun et al. 2002; Hammarstrom et al. 2002; Shih et al. 2002), particularly larger multidomain proteins. If the problem of insoluble expression of the His-tagged protein in *E. coli* is encountered, one or more of the following options are typically explored: altering culture growth conditions, co-expressing chaperones, changing cell lines, or switching to a different affinity tag such as MBP, GST, TRX, or NusA (Stevens 2000). Single domains may be targeted after accurate mapping of the domain boundaries using limited proteolysis and fragment analysis and/or emerging bioinformatics tools (Marsden et al. 2002; Miyazaki et al. 2002). Alternatively, eukaryotic expression systems may be used. In structural genomics, the protein may initially be left behind in the pursuit of the “low-hanging fruit”, to maximize the output and leave the problematic cases to be revisited when methodology improves (Edwards et al. 2000).

Apart from affinity purification, the aforementioned large-affinity tags offer several advantages. In a recent report, TRX and MBP enhanced the solubility and expression of a test set of 32 small (<20 kD) human proteins in *E. coli*, compared to the His-tag expression (Hammarstrom et al. 2002). For the test sets of 32 larger human proteins (17–158 kD; Braun et al. 2002) and 40 eukaryotic proteins (9–100 kD; Shih et al. 2002), the large-affinity tags MBP (40 kD), NusA (54 kD), and GST (26 kD) were demonstrated to be helpful in improving the yield of soluble protein, whereas thioredoxin (12 kD) did not provide significant improvement in solubility compared to His-tag. In an earlier study, thioredoxin and GST provided only minor or no improvement in solubility of six notoriously insoluble proteins, whereas MBP greatly enhanced the solubility of five of the proteins (Kapust and Waugh 1999). Furthermore, chaperone-like qualities have been attributed to MBP when fused at the N-terminus, assisting in correct protein folding and acquiring active proteins (Baneyx 1999; Kapust and Waugh 1999; Sachdev and Chirgwin 2000).

**The use of large-affinity tags in structural biology**

The considerations discussed above suggest that large-affinity tags may offer several advantages for structural biology applications. The 3D structures of *E. coli* MBP in apo- (Sharff et al. 1992) and maltose-bound (Spurlino et al. 1991; Quiocho et al. 1997) forms, *S. japonicum* GST (McTigue et al. 1995), and oxidized and reduced forms of *E. coli* TRX (Katti et al. 1990; Jeng et al. 1994) have been determined. These structures can be used as search models to solve the crystallographic phase problem by molecular replacement (MR) methods. Another possible advantage may be that the crystal contacts and the crystallization conditions successful in crystallizing the affinity tag may also be explored for crystallizing the fusion protein (Carter et al. 1994). However, this particular benefit has only been demonstrated for small peptides of 5–42 residues in length fused to *S. japonicum* GST (Zhan et al. 2001) or *Pyrococcus furiosis* MBP (Bucher et al. 2002), where the short peptides occupy the void near the location of the fusion, present among neighboring GST or MBP molecules in the crystal. Larger polypeptides would not fit into the available space, reducing the described advantages. The largest detriment to successful crystallization of fusion proteins with large-affinity tags is considered to be the conformational heterogeneity introduced by the flexible linker between the affinity tag and the protein of interest.

Recently, the first crystal structures of fusion proteins containing large-affinity tags have been reported (Kobe et al. 1999; Liu et al. 2001; Ke and Wolberger 2003); all three structures contain *E. coli* MBP as the affinity tag. Preliminary X-ray diffraction results have also been reported for crystals of GST (Kuge et al. 1997), TRX (Stoll et al. 1998), and MBP fusion proteins (Kukimoto et al. 2000; Table 1).
Crystallization of proteins fused to large-affinity tags

**Maltose binding protein (MBP).**

In the first crystallization report of an MBP-fusion protein, two fragments of the ectodomain of the human T cell leukaemia virus type 1 (HTLV-1) envelope protein gp21 were crystallized (Center et al. 1998). Crystallization of MBP fusion proteins was pursued because of the low solubility of the gp21 fragments on their own. Crystallization trials of the longer MBP/gp21 construct (residues 335–445) containing the unmodified linker between MBP and gp21 yielded no crystals. However, after extensive modification of the junction between the MBP and gp21 domains (Table 2), thin plate and needle crystals were obtained, albeit not suitable for structure determination. Expressing a shorter fragment (residues 338–425) with the same modifications at the fusion junction yielded 3D crystals that diffracted to 2.5 Å resolution and allowed the structure to be determined (Kobe et al. 1999). The sequence modifications in the MBP-gp21 junction included the substitution of the 25-amino acid linker with only three alanine residues, and the mutation of charged residues near the C-terminus of MBP to alanines (Table 2; Center et al. 1998).

The extracellular domain of the cell surface antigen CD38 in complex with ganglioside G T1b, a heptasaccharide containing lipid, is another example of a protein crystallized as a fusion to MBP (Kukimoto et al. 2000). Two crystal forms were found, one of which diffracted to 2.4 Å resolution. Attempts to solve the structure using MR with MBP or a CD38 homolog as the search model have failed, and no structure has yet been reported.

The Staphylococcus accessory regulator R (SarR) from *Streptococcus aureus* was also crystallized as an MBP fusion protein (Liu et al. 2001), and the analysis of the sequence of the crystallized construct shows that similar modifications to those used for the successful crystallization of MBP/gp21 were employed at the fusion junction (Table 2). Some charged residues at the C-terminus of MBP were mutated to alanine, and the linker length was shortened to the five-residue sequence AAAEF. The crystals diffracted to 2.3-Å resolution and the structure was determined using MR methods analogous to those reported for MBP/gp21.

**Table 1. Summary of crystallization data for protein fusions to large affinity tags**

<table>
<thead>
<tr>
<th>Protein</th>
<th>Fused/total amino acids (kDa)</th>
<th>Cloning vector</th>
<th>Linker</th>
<th>Protein concentration</th>
<th>Well solution</th>
<th>Resolution (Å)</th>
</tr>
</thead>
<tbody>
<tr>
<td>MBP/gp21</td>
<td>88/459 (9.9/50.4)</td>
<td>pMAL-c2</td>
<td>AAA</td>
<td>18 mg/mL</td>
<td>22% PEG 4000</td>
<td>2.5</td>
</tr>
<tr>
<td>(338–425)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>0.1 M NaOAc, pH 4.7</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>0.2 M (NH₄)₂SO₄</td>
<td></td>
</tr>
<tr>
<td>MBP/gp21b</td>
<td>108/479 (9.9/52.7)</td>
<td>pMAL-c2</td>
<td>AAA</td>
<td>18 mg/mL</td>
<td>1) 20% PEG 10,000</td>
<td>—</td>
</tr>
<tr>
<td>(338–445)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>0.1 M HEPES, pH 7.5</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>2) 18% PEG 8000</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>0.1 M cacodylate, pH 6.5</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>0.2 M Zn(OAc)₂</td>
<td></td>
</tr>
<tr>
<td>MBP/SarR</td>
<td>115/488 (13.7/54.5)</td>
<td>pMAL-c2</td>
<td>AAAEF</td>
<td>15 mg/mL</td>
<td>18–22% PEG-MME 2000</td>
<td>2.3</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>0.1 M NaOAc, pH 4.6</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>0.1 M NaCl</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>5 mM β-mercaptoethanol</td>
<td></td>
</tr>
<tr>
<td>MBP/MATa1</td>
<td>50/422 (6.0/46.5)</td>
<td>pMAL-c2</td>
<td>AAAAA</td>
<td>15 mg/mL</td>
<td>1) 2.4 M (NH₄)₂SO₄</td>
<td>1) 2.1</td>
</tr>
<tr>
<td>(77–126)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>0.1 M MES, pH 5.0</td>
<td>2) 2.3</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>1) 20% PEG 6000</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>0.1 M MES pH 6.0</td>
<td></td>
</tr>
<tr>
<td>MBP/CD38c</td>
<td>256/–635 (29.6/71.9)</td>
<td>pMAL-cR1</td>
<td>NA</td>
<td>10 mg/mL (+1.5 mg/mL G41k)</td>
<td>10–20% PEG 20,000</td>
<td>2.4</td>
</tr>
<tr>
<td>(45–300)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>0.1 M HEPES, pH 7.5</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>+ 0.1 M NaI (Form I)</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>+ 0.1 M glycine (Form II)</td>
<td></td>
</tr>
<tr>
<td>TRX/VanHd</td>
<td>322/–440 (35.8/–48)</td>
<td>pTrxFus</td>
<td>NA</td>
<td>4 mg/mL</td>
<td>0.8 M NaH₂PO₄</td>
<td>3.0</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>0.4 M K₂HPO₄</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>0.1 M HEPES, pH 7</td>
<td></td>
</tr>
<tr>
<td>GST/DREF</td>
<td>100/326</td>
<td>pGEX-2T</td>
<td>SDLVPRGS</td>
<td>15 mg/mL</td>
<td>5% PEG 3350</td>
<td>2.5</td>
</tr>
<tr>
<td>(16–115)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>50 mM KH₂PO₄, pH 5.2</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>10% ethylene glycol</td>
<td></td>
</tr>
</tbody>
</table>

| a | pMal-c2 was modified to introduce truncations and mutations (see Table 2). |
| b | Two crystal habits were grown but were unsuitable for X-ray diffraction. |
| c | Two polymorphs. |
| d | Two polymorphs grown in the same drop. |
| e | NA, information not available. |
The most recent example of a crystallized MBP fusion protein involves a fragment of the MATa1 protein (residues 77–126) from Saccharomyces cerevisiae (Ke and Wohberger 2003). The fusion junction of MBP/MATa1 was also modified to closely resemble that of MBP/gp21 (Table 2), with the linker sequence truncated to a pentaa Alanine. Two crystal forms were obtained, diffracting to 2.1 Å and 2.3 Å resolution, respectively. The structure was determined using MR, with the maltotetraose-bound MBP structure (Quiocco et al. 1997) as a search model.

The inspection of the crystallization conditions of MBP-fusion proteins (Table 1) shows that polyethylene glycols (PEG) and related molecules are the most successful precipitants, and the pH is generally low (Table 1). These observations may be of some use in devising specific crystallization screens for MBP fusion proteins (e.g., focusing on PEGs and acidic pH); however, no clear conclusions can be drawn from the small sample size presently available, and the two MBP/MATa1 crystal forms used very different conditions.

Glutathione-S-transferase (GST).

Successful crystallizations of target proteins fused to GST have been reported for the DNA-binding domain (residues 16–115) of the Drosophila DNA replication-related element-binding factor (GST/DREF; Kuge et al. 1997) and for the mouse estrogen receptor hormone binding domain (residues 281–599; GST/ERHBD; Lally et al. 1998). These reports have been discussed in the context of carrier-driven GST-peptide crystallization (Zhan et al. 2001). Of the two proteins, only GST/DREF yielded X-ray diffraction quality crystals; however, no subsequent structure has been reported. Although the thin crystals of GST/ERHBD were unsuitable for X-ray analysis, gel electrophoresis and electron microscopy confirmed the presence of the intact fusion protein.

Thioredoxin (TRX).

Vancomycin resistance protein (VanH), a D-lactate dehydrogenase from Enterococcus faecium, has been crystallized fused to thioredoxin (Stoll et al. 1998). Crystallization of the VanH-TRX fusion protein was attempted only after enterokinase cleavage failed to yield structurally intact VanH. Two crystal forms grew from the same crystallization conditions, one of which diffracted to 3.0 Å resolution (Table 1). No structure has yet been reported.

Crystal structures of MBP-fusion proteins

MBP/gp21.

The envelope protein gp21 is involved in the fusion of the viral and the host cell membranes during HTLV-1–mediated infection. The chimeric protein containing MBP and the HTLV-1 gp21 ectodomain has a trimeric mushroom-like structure (Fig. 1A) (Kobe et al. 1999). The ~70 Å long “stalk” is comprised of the 88 residues of each gp21 monomer assembled around a threefold crystallographic symmetry axis to form a parallel coiled coil structure. The three MBP units comprise the “cap,” and importantly do not hinder the formation of the trimeric complex despite the short linker between the two domains. The trimeric state is biologically relevant in vitro and in vivo (Center et al. 1998).

The tri-alanine linker connects the C-terminal helix of MBP with the N-terminal region of the fused protein (Protein). Residues constituting and the N-terminal helix of the fused protein are underlined, with the effective linker region between these secondary structural elements highlighted in bold. Mutations are highlighted in italic. The linker sequence typical for pMAL-c2 is shown in the example of the sequence corresponding to the fusion protein of MBP/HTLV gp21 (335–445).

<table>
<thead>
<tr>
<th>Fusion protein</th>
<th>MBP</th>
<th>Linker</th>
<th>Protein</th>
</tr>
</thead>
<tbody>
<tr>
<td>MBP/gp21 (338–445)</td>
<td>TVDEALKDAQTN</td>
<td>S,N,O,L,GIEGRIEFGS</td>
<td>TGMSLAS</td>
</tr>
<tr>
<td>MBP/gp21 (338–425)</td>
<td>TVDAALAAATQTN</td>
<td>AAA</td>
<td>MSLASGKS</td>
</tr>
<tr>
<td>MBP/SarR</td>
<td>TVDEALAAATQTN</td>
<td>AAEF</td>
<td>MSKIND</td>
</tr>
<tr>
<td>MBP/MATa1 (77–126)</td>
<td>TVDAALAAATQ</td>
<td>AAAA</td>
<td>ISPOARAF</td>
</tr>
</tbody>
</table>

Three sections are shown, the C-terminal helix of MBP (MBP), the linker region (Linker), and the N-terminal region of the fused protein (Protein). Residues constituting and the N-terminal helix of the fused protein are underlined, with the effective linker region between these secondary structural elements highlighted in bold. Mutations are highlighted in italic. The linker sequence typical for pMAL-c2 is shown in the example of the sequence corresponding to the fusion protein of MBP/HTLV gp21 (335–445).
double helix. The SarR monomers contain the typical helix-turn-helix DNA-binding domain. The MBP domains do not participate in, or hinder the dimer formation.

The first residue of the AAAEF linker represents the last residue of the C-terminal helix of MBP. The remaining four residues form part of a 10-residue loop joining the C-terminal helix of MBP to the first helix in SarR (Table 2; Fig. 1B).

**MBP/MATα1.**

MATα1 and MATα2 from *S. cerevisiae* bind DNA cooperatively to repress the transcription of haploid-specific genes. Crystallization as MBP-chimera was pursued because no crystals of free MATα1 could be obtained; the MBP/MATα1 (residues 77–126) chimera, on the other hand, produced crystals readily (one-fifth of conditions in a commercial crystallization screen produced crystals; Ke and Wolberger 2003). The structure reveals a typical homeodomain structure for the MATα1 fragment. The functional regions of MATα1 (the DNA-recognition helix and the MATα2-binding tail) are not obstructed by MBP, and the DNA-binding behavior of the chimera is the same as for the free protein.

The five-alanine linker between MBP and MATα1 adopts a turn conformation. The exact same disposition of the two moieties in the chimera is found in an alternative crystal form of the same protein. There are two residues between the poly-alanine linker and the first helix in MATα1 (Table 2; Fig. 1C).

**Comparative analysis.**

The most significant similarity between the three MBP fusion structures involves the short linker fusing the target proteins to MBP. The long flexible linker containing the
protease cleavage site was substituted with AAA for gp21, AAAAAE for SarR and AAAAAA for MATa1. Each linker sequence terminates the C-terminal MBP α-helix with a turn motif immediately preceding an N-terminal α-helix from the protein of interest. The original rationale behind using a three-alanine linker in the case of MBP/gp21 was to attempt to form a rigid connection through constructing a continuous helix between the C-terminal helix of MBP and the N-terminal helix of gp21. The MBP fusion protein structures now show that instead there may be structural reasons for the formation of a 90° turn at the end of the C-terminal helix of MBP.

Significantly, no obstruction of the biologically relevant quaternary states of gp21 and SarR is induced by the close proximity of the large MBP moiety in MBP/gp21 and MBP/SarR. Only limited interactions are formed between MBP and the fused proteins, making MBP an appealing affinity tag. The structures reveal physiologically relevant multimeric states and give insight into the mechanism of biologic function.

MBP adopts distinct conformational states depending on the absence or presence of the bound maltose (Spurlino et al. 1991; Sharff et al. 1992; Boos and Shuman 1998). The MBP/gp21 and MBP/SarR structures feature MBP in the “closed” conformation, retaining a bound maltose molecule (Kd = 35 µM; Quiocio et al. 1997) from the purification stage (no maltose was added to the crystallization solution). By contrast, the two crystal forms of MBP/MATa1 have no maltose in the active site of MBP. MBP/MATa1 was purified using strong cation exchange resin, which may have facilitated the release of maltose. Importantly, the structure could still be solved by MR using a ligand-bound structure as a search model. However, the MATa1 case emphasizes that care must be taken to avoid a partial occupancy of maltose, which would result in mixed conformational states and may inhibit well-ordered crystals from forming. Extra maltose is not required to ensure conformational homogeneity of the MBP using purification and crystallization conditions similar to those employed for MBP/gp21 and MBP/SarR. However, there may be some conditions that facilitate partial maltose release producing a mixture of the two MBP conformations, in which case maltose would have to be added to the protein solution for crystallization trials.

The three protein targets in MBP/gp21, MBP/SarR, and MBP/MATa1 are small in comparison to the 368 residues of E. coli MBP to which they are fused (Table 1). Until further structures are determined, it is difficult to determine whether the MBP/protein ratio is significant to the success of the crystallization process. One obvious advantage of a greater MBP to protein ratio is that it facilitates the structure determination by molecular replacement methods using MBP as a search model. A further implication of having small proteins fused to MBP is the dominance of MBP in crystal lattice formation. No direct crystal contacts (other than within the oligomer) are observed between gp21, SarR, or MATa1 molecules. Instead, the crystals are assembled by the combination of MBP/MBP and MBP/protein contacts. An analysis of the crystal packing arrangements in the presently available MBP and MBP fusion protein crystal structures does not suggest any clear parallels, except that the loops protruding out furthest from the globular structure of MBP (around residues 83, 141, and 173) are most frequently involved in crystal contacts. More correlations may emerge as new structures become available, and these may be exploited to design focused crystallization protocols. Although structures of larger fusion proteins should be possible, small proteins may be more conducive to this technique by allowing the affinity tag to direct the construction of the crystal lattice.

It is possible that some of the reported crystals of fusion proteins never led to successful structure determinations because no interpretable electron density for the protein of interest could be found, as a result of the mobility of this portion of the protein in the crystals. The chance of such an outcome is minimized, in parallel with increasing the likelihood of crystallization, by a rigid connection between the affinity tag and the protein of interest.

It is also possible that a soluble fusion protein is produced, but the protein of interest is not completely folded and exists as an ensemble of conformers (Sachdev and Chirgwin 1998; Nomine et al. 2001). As such heterogeneity would be expected to impede crystallization, it is advisable to characterize the fusion protein using biophysical tools or activity assays.

Potential applications of large-affinity tags in structural biology

The use of rigid fusions of proteins to large-affinity tags may be a viable alternative to the use of small tags in structural biology, when it overcomes the problem of obtaining an active, soluble sample in sufficient quantity and concentration (particularly when there are additional problems associated with the removal of the tag). This strategy is likely to find a niche role in the determination of challenging target proteins that would otherwise prove fruitless using other more conventional strategies.

Membrane proteins

Membrane proteins represent a specific class of proteins that continue to be challenging for structural studies. Difficulties pose themselves both during recombinant expression in heterologous systems and crystallization steps. Overexpression of membrane proteins in E. coli (Grisshammer and Tate 1995; Tate 2001; Quick and Wright 2002) fused to large-affinity tags has been reported using MBP (Grisshammer et al. 1993, 1994; Chen and Gouaux 1996; Su et al. 1996; Kanamori et al. 1999; Stanasila et al. 1999; Weiss and Grisshammer 2002), GST (Panayotova-Heiermann et al.
Crystals of membrane proteins are categorized into three general architectures: 2D, type I 3D, and type II 3D crystals (Michel 1991; Abramson and Iwata 1999). A 2D crystal is constructed when membrane protein molecules are ordered side by side in a lipid bilayer representative of the native membrane structure (Fig. 2A). Type I 3D crystals are layers of 2D crystals built up in the third dimension. In type II crystals, detergent-solubilized protein molecules are held together by crystal contacts mediated by the hydrophilic portions of the protein (protruding from the detergent-covered, hydrophobic transmembrane regions). Medium- to low-resolution 3D structural information can be obtained from ultrathin 2D crystals (by cryoelectron microscopy; Saibil 2000), although type I and type II 3D crystals may be suitable for X-ray diffraction.

One method of growing 2D crystals utilizes affinity of the protein for the polar head groups of natural or synthetic lipids, to allow the self-assembly of protein molecules on a planar lipid film (Fig. 2B; Brisson et al. 1999; Levy et al. 1999, 2001). Using this technique, the ligand lipid is diluted with a second lipid and deposited onto an aqueous solution of the protein so that the lipids spread into a monolayer at the water–air interface. Fusion proteins containing large-affinity tags may bind lipids that incorporate the ligand for the affinity tag; for example, lipids carrying a maltose polar head group may be used to bind MBP fusion proteins. Assembly into 2D crystals occurs through substitution of the detergent micelles with a lipid bilayer and the formation of specific intermolecular contacts between neighboring hydrophilic portions of molecules.

Increasing the size of the hydrophilic domain of membrane proteins through the use of antibody fragments has been demonstrated to aid crystallization (Hunte and Michel 2002). A hydrophilic domain, such as an affinity tag, may be used to play a similar role (Hunte and Michel 2002). Rigid fusion can be achieved through the short linker approach described above for soluble proteins, or by insertion of the fusion protein into one of the extramembrane loops. This latter approach has been successful in preparing 2D crystals of the fusion of cytochrome \(b_{562}\) to lactose permease (Prive and Kaback 1996; Zhuang et al. 1999).

Structural genomics.

In the early stages of structural genomics, challenging targets such as those that express poorly or yield insoluble protein will be put aside in the pursuit of the “lower hanging

Figure 2. (A) Three architectures of membrane protein crystals. (B) A method for growing 2D membrane protein crystals using MBP fusion proteins.
fruit” (Edwards et al. 2000). The “low-hanging fruit” usually corresponds to proteins that express as soluble His-tag fusion proteins in bacteria, but they may not always be the most desired targets (Vitkup et al. 2001). One alternative is expression in *E. coli* as a fusion to a large-affinity tag such as MBP, NusA, GST, thioredoxin, and others (Braun et al. 2002; Hammarstrom et al. 2002; Shih et al. 2002). Incorporation of an additional small affinity tag (such as the His-tag) greatly increases the efficiency of purification as well as minimizes the losses due to poor binding to the affinity column, a problem often encountered for MBP fusions (Baneyx 1999; Braun et al. 2002; Routzahn and Waugh 2002). The proteins expressed and purified in this way need to progress through cleavage and repurification steps before entering crystallization trials or NMR studies.

Expression of proteins rigidly fused to the MBP tag could be pursued in parallel as an alternative approach. Potentially, proteins expressed this way would retain the solubility and purification advantages observed for the longer construct (Routzahn and Waugh 2002) yet could proceed straight from purification to crystallization trials.

High-throughput approaches require methods for cloning and expression that can be applied simultaneously to a large number of target proteins. One efficient system with such properties is the directional TOPO cloning technology (Invitrogen Life Technologies) that requires no ligase or restriction enzymes. The technology could be adapted to the large-affinity tag-short linker system; however, the linker would have to contain three amino acids (the result of topoisomerase I action and an integral part of TOPO directional cloning), in addition to the three-alanine or similar spacer of choice (Fig. 3A).

Another possibility would be to use a modification of a classic bacterial plasmid expression vector, encoding MBP followed by the sequence GCCGCTGCGCA. This sequence encodes three alanines, with the underlined nucleotides representing the recognition site for the blunt end-producing restriction enzyme FspI. The sequence could be modified to incorporate more alanine residues, if required. After a restriction digest of the vector (resulting in linearized blunt-ended DNA), 3’-terminal thymidines would be added to both ends (Fig. 3B), and the resulting overhangs used for an efficient ligation of the target protein PCR product (generated by thermostable polymerases producing A-overhangs). The use of this vector system would avoid the incorporation of any unwanted amino acids (as in the case of the TOPO-based strategy), as well as the use of a multiple cloning site (used in most available expression systems). However, the proposed method is a nondirectional cloning method, requiring an additional step confirming the correct orientation of cloned sequence.

An advantage of the “rigid fusion” technology is that X-ray diffraction data could potentially be collected at the home laboratory and solved using MR methods, or known binding sites for heavy atoms can be utilized for phasing (Spurlino et al. 1991; Rubin et al. 2002), as alternatives to Se-Met incorporation routinely used in structural genomics (MBP contains 7 Met residues).

**Conclusions**

It is clear that large-affinity tags offer an advantage over a small tag when expressing recombinant proteins, and the

---

**Figure 3.** (A) Schematic diagram of the proposed TOPO expression vector, incorporating the MBP affinity tag, three alanines, and a TOPO directional cloning site. T7, T7 promoter; lacO, lac operon; RBS, ribosome binding site; ATG, translation initiation codon; 6xHis, His-tag; MBP, maltose binding protein; T7 term, T7 termination region; TOPO, topoisomerase I. (B) Schematic diagram of the proposed nondirectional expression vector, incorporating the MBP affinity tag and three alanines, labeled as in (A). 3’ T overhangs are indicated.
resulting fusion proteins have found extensive use in diverse functional studies. Although only a few successful crystallizations of proteins fused to large-affinity tags have been reported, they set an important precedent for also using the fusion proteins for structural studies. The analysis of these structures demonstrates that deliberate modification of the fusion junction will usually be required to generate high quality crystals for X-ray diffraction. A rigid connection must be established between the two domains to remove the impediment of conformational heterogeneity that exists when long flexible linkers are present. This technique may occupy a niche role in the structural biologist’s toolbox for obtaining 3D information of challenging target proteins.

Acknowledgments

We thank Ben Hankamer and Trazel Teh for useful discussions. B.K. is an NHMRC Senior Research Fellow.

Note added in proof

Chao et al. reported the crystal structure of the MBP–Saccharomyces cerevisiae ribosomal protein L30 fusion protein. The protein was crystallized in the presence of maltose, and the crystal structure was solved by molecular replacement using the structure of Saccharomyces cerevisiae. The protein produced as fusion proteins in Escherichia coli when long flexible linkers are present. This technique may provide an important precedent for also using the fusion proteins for structural studies. The analysis of these structures demonstrates that deliberate modification of the fusion junction will usually be required to generate high quality crystals for X-ray diffraction. A rigid connection must be established between the two domains to remove the impediment of conformational heterogeneity that exists when long flexible linkers are present. This technique may occupy a niche role in the structural biologist’s toolbox for obtaining 3D information of challenging target proteins.

References


Kapust, R.B. and Waugh, D.S. 1999. Escherichia coli maltose-binding protein is uncommonly effective at promoting the solubility of polypeptides to which it is fused. Protein Sci. 8: 1668–1174.


