Journal of Structural and Functional Genomics **2:** 83–92, 2002.

CABM Symposium

Differential effects of supplementary affinity tags on the solubility of MBP fusion proteins

Karen M. Routzahn and David S. Waugh *
*Protein Engineering Section, Macromolecular Crystallography Laboratory, Center for Cancer Research, National Cancer Institute at Frederick, P.O. Box B, Frederick, Maryland 21702, USA; *Author for correspondence (tel: 301-846-1842; fax: 301-846-7148; e-mail: waughd@ncifcrf.gov)*

Received 29 December 2001; Accepted in revised form 25 March 2002

Key words: affinity tags, fusion protein, maltose binding protein, solubility, structural genomics

Abstract

It is difficult to imagine any strategy for high-throughput protein expression and purification that does not involve genetically engineered affinity tags. Because of its ability to enhance the solubility and promote the proper folding of its fusion partners, *Escherichia coli* maltose-binding protein (MBP) is a particularly useful affinity tag. However, not all MBP fusion proteins bind efficiently to amylose resin, and even when they do it is usually not possible to obtain a sample of adequate purity after a single affinity step. To address this problem, we endeavored to incorporate supplemental affinity tags within the framework of an MBP fusion protein. We show that both the nature of the supplemental tags and their location can influence the ability of MBP to promote the solubility of its fusion partners. The most promising configurations for high-throughput protein expression and purification appear to be a fusion protein with a biotin acceptor peptide (BAP) on the N-terminus of MBP and/or a hexahistidine tag (His-tag) on the C-terminus of the passenger protein.

Abbreviations: BAP, biotin acceptor peptide; EDTA, ethelenediaminetetraacetic acid; IPTG, isopropyl-β-D-thiogalactopyranoside; MBP, *E. coli* maltose-binding protein; GFP; green fluorescent protein; Ni-NTA, nickel-nitrilotriacetic acid; ORF, open reading frame; PCR; polymerase chain reaction; R5, polyarginine tag; SDS-PAGE, sodium dodecyl sulfate-polyacrylamide gel electrophoresis; TEV, tobacco etch virus; WT, wild-type

Introduction

High-throughput expression and purification of recombinant proteins for large-scale structural biology initiatives (i.e., structural genomics) will require a general approach that yields predictable results. However, proteins are a chemically and structurally diverse group of macromolecules that do not readily lend themselves to generic purification strategies, and the yield and solubility of recombinant proteins in heterologous environments can vary widely. In principal, both of these problems can be mitigated to some degree by the use of affinity tags. Not only can they be exploited to devise generic protocols for protein purification, but affinity tags have also been observed to improve the yield of recombinant proteins, protect them from intracellular proteolysis, and even enhance their solubility [1].

Among the affinity tags that have been characterized to date [2], only *E. coli* MBP is also a potent solubility enhancer [3]. In many cases, polypeptides that normally accumulate as inclusion bodies in *E. coli* can be recovered in a properly folded, biologically active form if they are fused to the C-terminus of MBP (e.g., [3–5]). Because insolubility appears to be the major obstacle to high-throughput protein expres-

sion and purification [6], this chaperone-like quality of MBP makes it a particularly valuable fusion partner.

One shortcoming of MBP, however, is that MBP fusion proteins do not always bind efficiently to amylose resin [4]. Besides, it is usually not possible to obtain protein of sufficient purity for structural studies after a single affinity step. One way around this problem would be to incorporate additional affinity tags within the framework of an MBP fusion protein. These supplemental tags could be used in concert with MBP to achieve greater purity, and would also afford more flexibility by enabling alternative strategies to be employed if unanticipated problems arise. From a practical standpoint, the challenge is to identify locations in which these tags can be placed so that they will function as intended without also interfering with the ability of MBP to act as a solubility enhancer. In the present report, we test various locations for supplemental tags and evaluate their impact on the ability of MBP to promote the solubility of its fusion partners. On the basis of these results, we propose two configurations involving multiple affinity tags that appear to be well suited for high-throughput protein expression and purification.

Materials and methods

Materials

Restriction endonucleases, T4 DNA ligase, Deep Vent DNA polymerase, pMal-C2, and pMal-C2x were purchased from New England Biolabs (Beverly, MA). *Pfu* DNA polymerase was obtained from Stratagene (La Jolla, CA). Synthetic oligodeoxyribonucleotides and Gateway cloning reagents were obtained from Life Technologies (Gaithersburg, MD), a subsidiary of Invitrogen (Carlsbad, CA). IPTG was obtained from Research Products International (Mount Prospect, IL). GelCode Blue was from Pierce (Rockford, IL). Novex 10–20% Tris–Glycine Gels were purchased from Invitrogen.

MBP expression vectors with supplementary N-terminal tags

pKM1074, which produces wild-type MBP with an N-terminal BAP (Figure 2a), was constructed by replacing the DNA between the unique *Bsi*WI and *Bam*HI sites in pDW363 [7] with the corresponding fragment of pDW533 [8]. pKM1084, the vector that produces MBP with an N-terminal Arg-tag (Figure 2a), was constructed in two steps. First, a portion of the MBP open reading frame (ORF) extending from the N-terminus to just beyond the unique *Bgl*II site in pMal-C2x was amplified by polymerase chain reaction (PCR), using an oligonucleotide primer with an unpaired 5′ extension that encoded the N-terminal Arg-tag preceded by an *Nde*I site. This PCR amplicon was cleaved with *Nde*I and *Bgl*II, and then ligated with the *Nde*I/*Bgl*II vector backbone of pMal-C2x to create pKM771. In the second step, the *Bsi*WI/*Bam*HI fragment of pKM771 was replaced by the corresponding fragment from pDW533 as described above.

MBP expression vectors with supplementary internal tags

The starting point for the construction of MBP expression vectors with internal tags between residues 133 and 143 was pDW483, a vector that produces MBP with a stuffer fragment in this position that is flanked by *Bbs*I sites. To construct the vector with an internal BAP tag (Figure 2b), first a pair of synthetic oligodeoxyribonucleotides (5′-CTTGCCACCGTG CCATTCGATTTTCTGAGCTTCGAAAATATCGTT CAGGCCTCCCGG-3′ and 5′-GATTCCGGGAGG CCTGAACGATATTTTCGAAGCTCAGAAAAT CG AATGGCACGGTGG-3′) was annealed and the resulting double-stranded fragment was inserted between the two *Bbs*I sites in pDW483 to create pKM606. Next, the *Bgl*II/*Nco*I fragment of pKM606 that encompasses this region was ligated with the *Bgl*II/*Nco*I vector backbone of pDW533 to generate the final expression vector (pKM1143). The vector with an Arg-tag in this position (pKM1144) was constructed in the same manner, starting with a different pair of oligodeoxyribonucleotides (5′-GATTCCGCG TCGCCGGCGTCGTGG-3′ and 5′-CTTGCCACG ACGCCGGCGACGCGG-3′). The vector that produces an MBP with a BAP tag between residues 250 and 254 (Figure 2b) was constructed by overlap-extension PCR [9] using the following pair of overlapping mutagenic primers (5′-GACATCTTCGAGGCC CAGAAAATCGAGTGGCATGGTGGTCCATCCAA ACCGTTCGTTGGCGTGCT- 3′ and 5′-GATTTTC TGGGCCTCGAAGATGTCATTCAGACCACCGAA GGTCGGCAGTACCGTTACACCATAATTC-3′) in concert with a pair of primers that flank the *Bgl*II/*Nco*I fragment within the MBP ORF. The final PCR amplicon was cleaved with *Bgl*II and *Nco*I and then

ligated with the *Bgl*II/*Nco*I vector fragment of pDW533 to generate the expression vector (pKM1145).

MBP expression vectors with supplementary C-terminal tags

To construct pKM648, which produces MBP with the C-terminal R5#1 Arg-tag (Figure 2c), the Gateway Cloning Cassette in pKM596 [10] was amplified by PCR, using a pair of primers with unpaired 5′ extensions. The 5′ extension of the N-terminal primer encoded an Arg-tag preceded by a *Sac*I site. The 5′ extension of the C-terminal primer included a *Hin*dIII site. The amplicon was cleaved with *Sac*I and *Hin*dIII, and then ligated with the *Sac*I/*Hin*dIII vector fragment of pMal-C2. To make the vector that produces MBP with the C-terminal R5#2 tag (Figure 2c), a pair of PCR primers (5′-AAGCTTGGCACTGGCCGT CGTTTTAC-3′ and 5′-GCGACGGCGACG ACGAGTCTGCGCGTCTTTCAG-3′) was used to amplify nearly the entire length of pMal-C2 after the DNA was linearized with *Eco*RI. One of the primers included an unpaired 5′ extension that added an in-frame Arg-tag to the C-terminus of the MBP ORF. The blunt-ended amplicon was subsequently ligated with the Gateway Cloning Cassette (RfA) to generate the expression vector (pKM681).

MBP fusion vectors with supplemental His-tags on the N- or C-termini of the passenger proteins

A two-step PCR procedure was used to construct Gateway entry clones of the passenger proteins (p16, E6, and GFP) with no supplemental His-tags. First, each ORF was amplified with a pair of gene-specific primers. The N-terminal primers included unpaired 5′ extensions that encoded a cleavage site (ENLYFQG) for tobacco etch virus (TEV) protease [11], and the C-terminal primers included an attB2 recombination site as an unpaired 5′ extension. The initial PCR amplicons were then used as templates for a second PCR with the same gene-specific C-terminal primers and a generic N-terminal primer that anneals to the TEV site and includes an attB1 recombination site as an unpaired 5′ extension. The amplicons generated after the second round of PCR were recombined via the BxP reaction into pDONR201 (Invitrogen) to yield the entry clones pKM1122 (GFP), pKM991 (E6), and pKM992 (p16).

The same approach was used to construct Gateway entry clones of the passenger proteins with supplemental N-terminal His-tags. However, in this instance the N-terminal gene-specific primers included unpaired 5′ extensions that encoded a His-tag preceded by a TEV protease cleavage site. The amplicons generated after the second round of PCR were recombined via the BxP reaction into pDONR201 to yield entry clones pKM1128 (GFP), pKM1127 (E6), and pKM1125 (p16).

Entry clones of the passenger proteins with supplementary C-terminal His-tags were constructed as follows. First, each ORF was amplified by PCR using a gene-specific N-terminal primer that included a TEV protease site as an unpaired 5′ extension and a gene-specific C-terminal primer that added six histidine codons to the C-terminus of the ORF. The initial amplicons were then used as templates for a second round of PCR with a generic N-terminal primer that anneals to the TEV protease cleavage site and includes the attB1 recombination site as an unpaired 5′ extension and a generic C-terminal primer that anneals to the His-tag and includes the attB2 recombination site as an unpaired 5′ extension. The final PCR amplicons were recombined into pDONR201 via the Gateway BxP reaction to yield entry clones pKM617 (GFP), pKM1126 (E6), and pKM1124 (p16).

The MBP destination vector pKM596 has been described elsewhere [10]. The BAP-MBP destination vector pKM1107 was constructed by inserting the Gateway cloning cassette (RfA) between the unique *Sac*I and *Bam*HI sites in pDW342 [7] after blunting the ends with T4 DNA polymerase and deoxynucleotide triphosphates. The Arg-MBP destination vector pKM1072 was created by ligating the *Bst*EII/*Bsi*WI fragment of pKM771 that includes the N-terminus of the Arg-MBP ORF with the *Bst*EII/*Bsi*WI vector fragment of pKM596.

Expression vectors for the production of MBP fusion proteins with or without supplementary His-tags on the N- or C-termini of the passenger proteins (Figure 5) were constructed by recombining the passenger proteins from the corresponding entry vectors (above) into the destination vector pKM596 via the Gateway LxR reaction, using the standard protocol (Invitrogen). Expression vectors for the production of MBP fusion proteins with a supplementary BAP or Arg-tag on the N-terminus of MBP were created by recombining the passenger proteins from the entry vectors with no His-tags into the destination vector pKM1107 (BAP-MBP) or pKM1072 (Arg-MBP) via

the Gateway LxR reaction, using the standard protocol. Expression vectors for the production of MBP fusion proteins with Arg-tags in the linker region between MBP and the passenger proteins were constructed in a similar manner, using the destination vectors pKM648 (R5#1) or pKM681 (R5#2).

The nucleotide sequences of the relevant ORFs encoded by all expression vectors used in this study were verified experimentally.

Protein expression, SDS-PAGE analysis, and densitometry

Protein expression, preparation of total and soluble intracellular protein samples, SDS-PAGE, and quantitative densitometry were performed essentially as described [3] using *E. coli* BL21(DE3) cells containing the tRNA accessory plasmid pRIL (Stratagene). The only exception was the experiment involving pKM596, pKM648 and pKM681 (Figure 3c). These three vectors had to be maintained in *E. coli* DB5, a host strain with a *gyrA* mutation that renders the cells immune to the action of a DNA gyrase poison (CcdB) produced by the plasmids. Due to the presence of an in-frame translation termination codon immediately after the attB1 site, these destination vectors could be used to produce the fusion proteins depicted in Figure 2c. All experiments were performed in duplicate.

Results

Supplemental affinity tags

Because MBP is already a relatively large protein (ca. 42 kDa), we considered only small peptides as supplemental affinity tags. Among the smallest tags that have been described are the polyarginine tag (Arg-tag; 5 residues), the hexahistidine tag (His-tag; 6 residues), and the biotin acceptor peptide (BAP; 13–24 residues). The principal behind the Arg-tag is that it should endow even an acidic protein with the ability to bind to a cation exchange resin at alkaline pH (>8.0), conditions under which the vast majority of endogenous proteins will fail to adhere [12]. Elution of the Arg-tagged protein is effected with a salt gradient. His-tagged proteins will bind tightly to immobilized divalent metal ions such as nickel or cobalt [13]. Bound proteins can be released from Ni-NTA resin with an imidazole gradient, by reducing the pH to $\sim$4.5, or by stripping the bound metal ions from the column with a chelating agent like EDTA. The BAP tag is a substrate for site-specific enzymatic biotinylation *in vivo* by *E. coli* biotin holoenzyme synthetase (BirA) [14]. Once biotinylated, a BAP-tagged protein can be captured on immobilized avidin or streptavidin resin. However, because the interaction between biotin and avidin or streptavidin is incredibly strong ($K_d \sim 10^{-15}$ mol/L), modified forms of these proteins with reduced affinity for biotin are usually employed for affinity chromatography [15, 16].

Potential locations for supplemental tags

The most obvious locations to place supplemental affinity tags within the framework of an MBP fusion protein are at its N- or C-termini or in the linker region between MBP and the passenger protein (Figure 1). However, we considered that it might also be possible to insert tags within the folded structure of MBP itself. One potential location was revealed by a study in which a genetic approach was employed to identify permissive sites in MBP that could accommodate small insertions or deletions without abolishing its ability to bind and transport maltose [17]. It was subsequently demonstrated that many different foreign peptides and even the entire TEM β-lactamase protein could be tolerated in one of these locations, between residues 133 and 143 of MBP [18, 19]. Another possible location for an accessory tag within MBP was suggested by comparing the crystal structure of *E. coli* MBP [20] with that of its ortholog from the hyperthermophilic archaeon *Pyrococcus furiosus* [21]. The solvent-exposed loop corresponding to res-

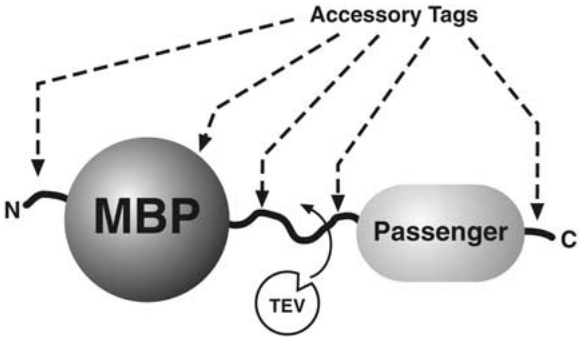

Figure 1. Potential locations for supplementary affinity tags within the framework of an MBP fusion protein. Accessory tags on the N-terminal side of the TEV protease processing site remain associated with MBP after digestion of the fusion protein whereas accessory tags on the C-terminal side of the TEV protease site remain associated with the passenger protein.

idues 251–253 in *E. coli* MBP is elongated in *P. furiosus* MBP, suggesting that it might be able to accommodate an insertion.

Depending on its location, a supplementary tag may or may not be retained on the passenger protein after the MBP moiety is removed by proteolytic digestion of the fusion protein (Figure 1). For those tags that remain associated with the passenger protein, one must be cognizant of their potential impact on its propensity to crystallize. A recent study demonstrated that the addition of a His-tag to the C-terminus of a model protein had no effect on the formation of crystals or their ability to diffract X-rays, whereas a BAP or Arg-tag in the same location was clearly detrimental [22]. We also note that more than one hundred crystal structures of proteins with a His-tag on one of their termini have been deposited in the RCSB Protein Data Bank to date [23]. Thus, the available evidence suggests that, among the supplemental affinity tags under consideration, the His-tag is least likely to interfere with the crystallization of the passenger protein. Accordingly, only the His-tag was tested in locations where it could not be separated from the passenger. Conversely, the BAP and Arg-tag were only employed in positions where they would remain associated with MBP after proteolytic processing of the fusion protein.

The influence of supplemental tags on the solubility of MBP in its unfused state

For those locations in which the supplemental tags would remain associated with MBP after proteolytic digestion of the fusion protein, we first investigated whether their presence would affect the solubility of MBP in its unfused state. Two N-terminal tags were tested: a 24- residue BAP and an Arg-tag (Figure 2a). Although the BAP employed in these experiments is longer than the minimal substrate for *in vivo* biotinylation [24], we previously demonstrated that this BAP-MBP fusion protein can be expressed at a high level and biotinylated efficiently in *E. coli* [7]. The N-terminal Arg-tag was preceded by an initiator methionine and a lysine residue. The lysine was added in an effort to improve the yield of Arg-MBP, which was very low in its absence (data not shown).

The two N-terminally tagged MBPs were overproduced in *E. coli* and their yield and solubility was compared with that of wild-type MBP under the same conditions (Figure 3a). The results demonstrate that neither N-terminal tag adversely affects the yield or solubility of MBP. Using a streptavidin–horse radish peroxidase conjugate as a probe in a Western blot experiment, we also confirmed that the N-terminal BAP tag was biotinylated in *E. coli*, in agreement with previous results [7]. The other protein that reacts with the probe is biotin carboxyl carrier protein (BCCP), the only endogenous biotinyl protein in *E. coli* [25].

In a second series of experiments, residues 133–143 of MBP were deleted and replaced with either a BAP or an Arg-tag (Figure 2b). In contrast to the N-terminally tagged proteins, the solubility of these two internally tagged MBPs was severely compromised (Figure 3b). A similar result was obtained when residues 251–253 were deleted and replaced by a BAP (Figures 2b and 3b). Nevertheless, a Western blot revealed that both of the MBPs with internal BAPs were biotinylated *in vivo* (Figure 3b). To the best of our knowledge, this is the first demonstration that the BAP can function as a substrate for BirA when both of its ends are constrained. The fact that much of the biotinylated MBP is present in the insoluble fraction of the cell lysate implies that biotinylation occurs very rapidly *in vivo* (i.e. prior to the formation of inclusion bodies).

In a third series of experiments, we examined the impact of adding an Arg-tag to the C-terminus of MBP, which would place it in the linker between MBP and the passenger protein in the context of a fusion protein (Figure 1). Only the Arg-tag was tested in this position because we were concerned that more drastic alterations in the length of the linker might interfere with the ability of MBP to promote the solubility of its fusion partners. However, a previous study established that a His-tag can also be used to good effect in this location, although its impact on the solubilizing properties of MBP was not directly assessed [4] Two different constructs were tested (Figure 2c): one with an Arg-tag located very close to the C-terminus of MBP (R5#2) and another with an Arg-tag in a more central location within the linker (R5#1). The other residues that comprise the linker sequences are derived from the attB recombination site in the Gateway destination vectors (see Materials and methods). Neither of these C-terminally Arg-tagged MBPs was any less soluble than the corresponding protein with no Arg-tag (Figure 3c).

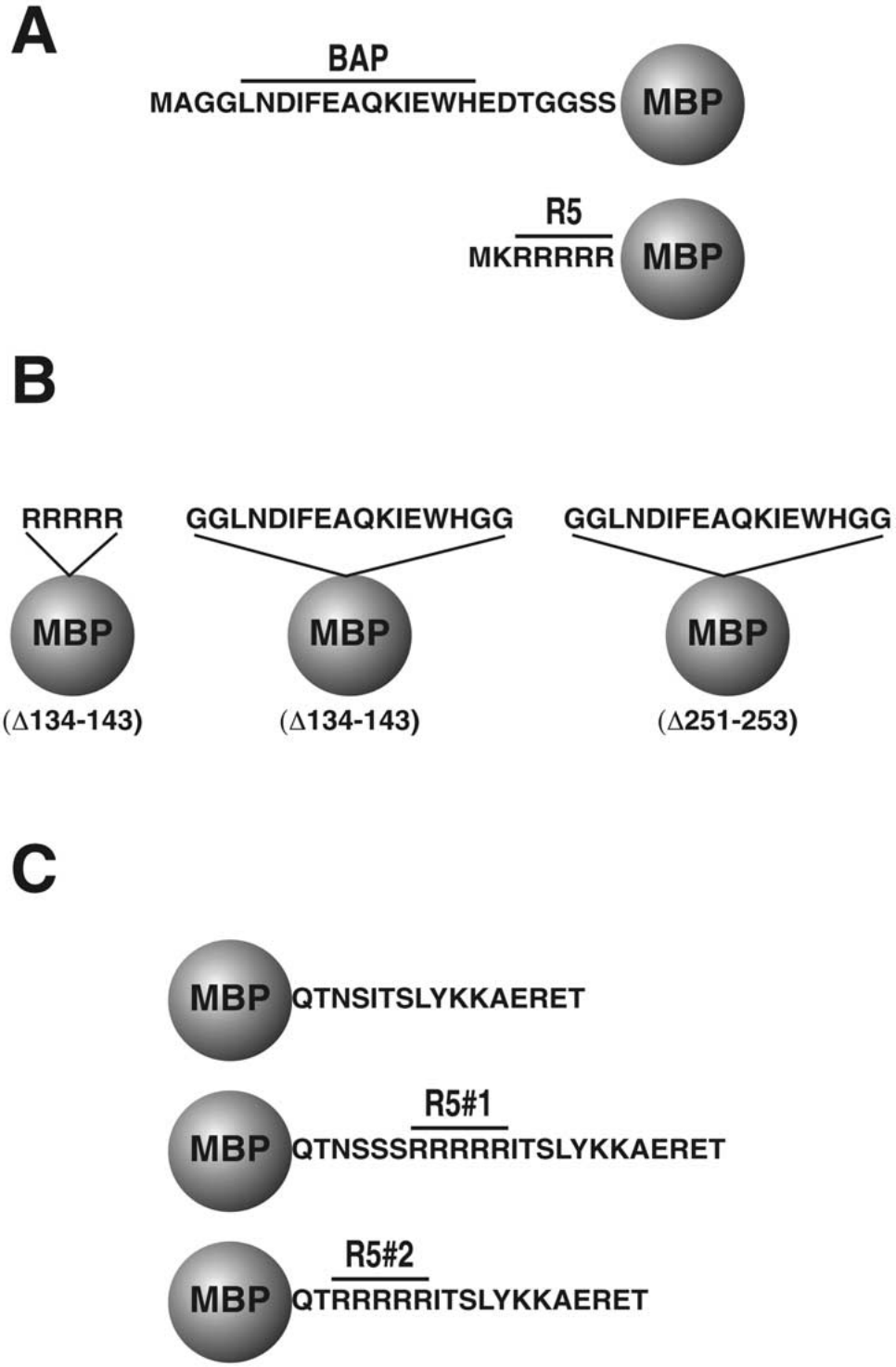

Figure 2. Schematic representations of MBPs with supplementary affinity tags. (**A**) N-terminal tags. (**B**) Internal tags. (**C**) C-terminal tags.

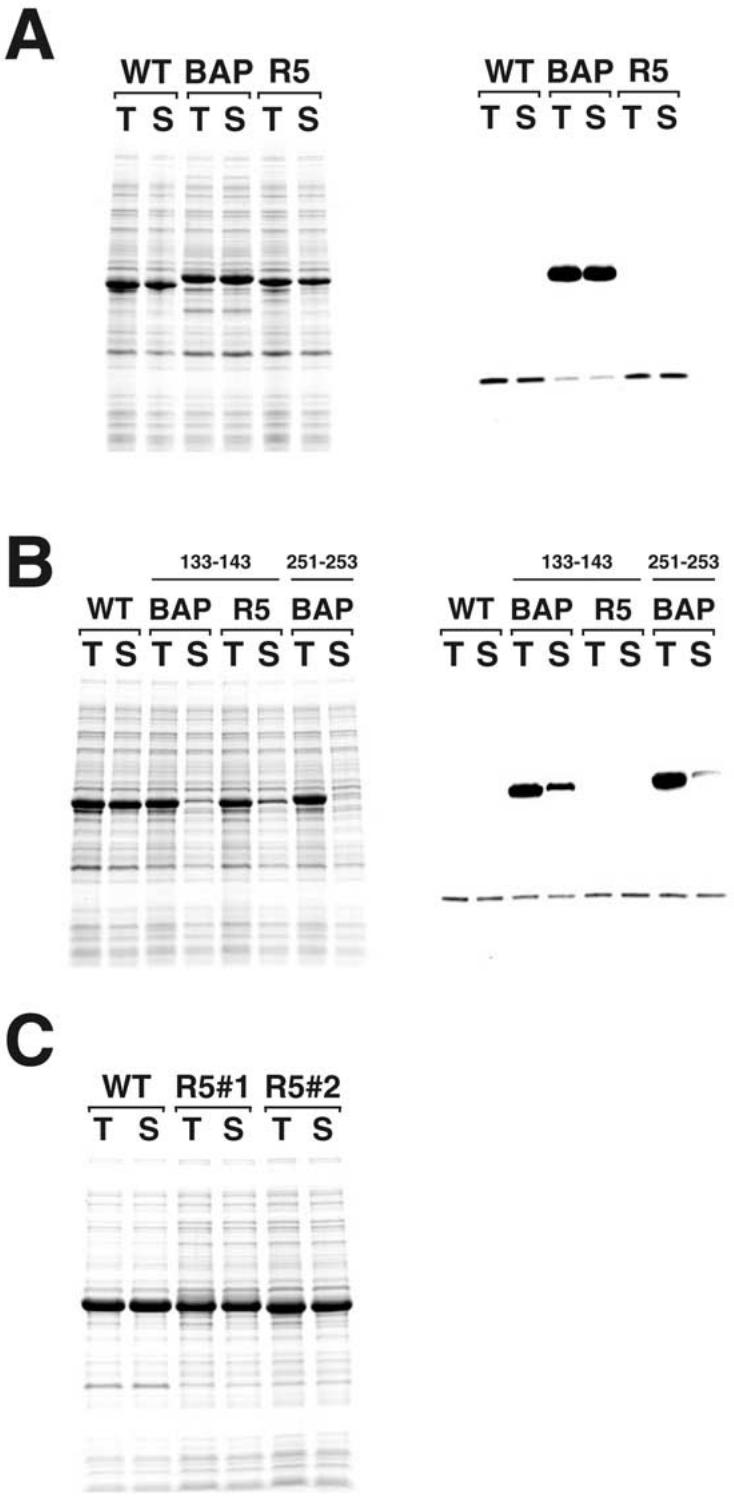

Figure 3. Yield and solubility of MBPs with supplementary affinity tags. (**A**) N-terminal tags. (**B**) Internal tags. (**C**) C-terminal tags. Left: Tris-glycine SDS polyacrylamide gel (10-20% gradient) stained with GelCode Blue. Right: Western blot probed with a streptavidin-horse radish peroxidase conjugate. Abbreviations: WT, wild-type (untagged) MBP; BAP, biotin acceptor peptide; R5, Arg-tag; T, total intracellular protein; S, soluble intracellular protein.

The influence of supplemental tags on the solubility of MBP fusion proteins

Those supplemental tags that did not reduce the solubility of MBP in its unfused state were subsequently tested in the context of several different fusion proteins to ascertain whether they would interfere with MBPs ability to promote the solubility of its fusion partners. The three passenger proteins selected for these experiments were human $p16^{INK4}$, human papilloma virus oncoprotein E6, and wild-type *Aquorea victoria* green fluorescent protein (GFP). These proteins accumulate predominantly or exclusively in an insoluble form when they are overproduced in *E. coli*, but the corresponding MBP fusion proteins are highly soluble [3, 8]. The yield of the six MBP fusion proteins with supplementary N-terminal affinity tags was very similar (data not shown). However, whereas the solubility of the MBP fusion proteins with N-terminal BAPs was about the same as the solubility of the corresponding fusion proteins with no supplemental tags, the solubility of all three fusion proteins with N-terminal Arg-tags was markedly reduced (Figure 4a). The presence of an Arg-tag in the linker region between MBP and the passenger protein also reduced the solubility of the fusion proteins, particularly when the Arg-tag was located closer to the C-terminus of MBP (Figure 4b).

C-terminal His tags reduce the solubility of MBP fusion proteins

A series of His-tagged MBP fusion proteins was constructed to determine what would happen if supplementary tags were added to either end of the passenger protein, where they will remain after proteolytic removal of the MBP moiety (Figure 5). The same three passenger proteins (p16, E6, and GFP) were used to assess the impact of the supplementary His-tags on the solubility of MBP fusion proteins. Two of the three fusion proteins with non-removable His-tags on their C-termini (MBP-p16-His and MBP-E6-His) were less soluble than their counterparts with no supplementary His-tags (Figure 4c). On the other hand, the presence of His-tags on the N-termini of the passengers had no significant effect on the solubility of the MBP fusion proteins (Figure 4c).

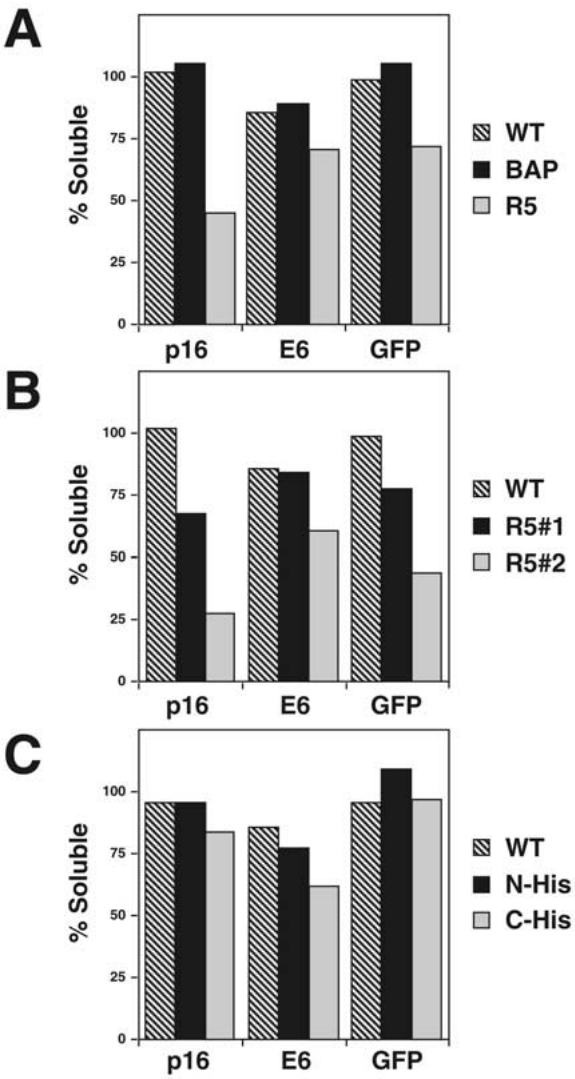

Figure 4. Solubility of MBP fusion proteins with supplementary affinity tags. (**A**) Fusion proteins with supplementary tags fused to the N-terminus of MBP. (**B**) Fusion proteins with supplementary Arg-tags in the linker between MBP and the passenger proteins. (**C**) Fusion proteins with supplementary His-tags attached to the N- or C-termini of the passenger proteins. Abbreviations: WT, wild-type fusion proteins (with no supplementary tags); BAP, biotin acceptor peptide; R5, Arg-tag; N-His and C-His, fusion proteins with nonremovable His-tags added to the N-termini or C-termini of the passenger proteins, respectively.

Discussion

MBP has several attributes that contribute to its utility as a fusion partner for the production of recombinant proteins. First, its natural affinity for maltodextrins can be exploited for affinity purification. Second, MBP fusion proteins are almost always produced in

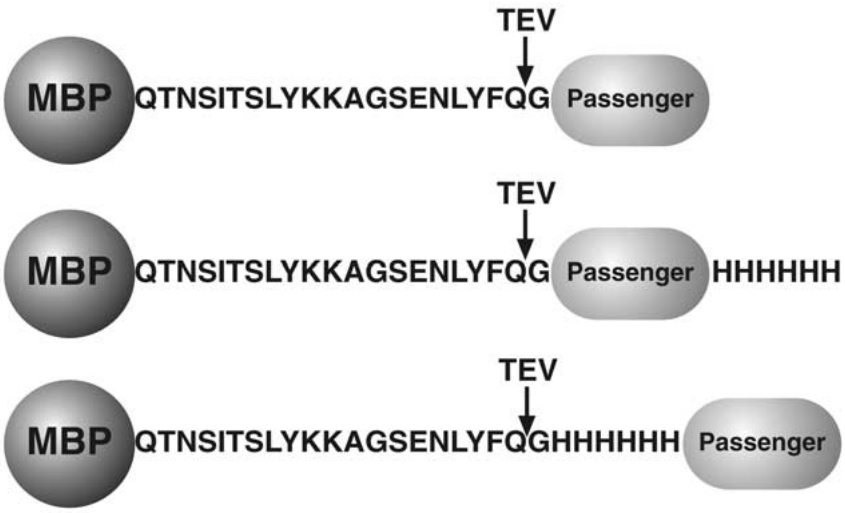

Figure 5. Schematic representations of MBP fusion proteins with no supplementary affinity tags and with supplementary His-tags on the N- or C-termini of the passenger proteins. The processing sites for tobacco etch virus (TEV) protease are indicated.

great abundance in *E. coli*. Third, MBP can afford some degree of protection from intracellular proteolysis. Finally, MBP has a remarkable ability to enhance the solubility and promote the proper folding of its fusion partners. Despite these powerful advantages, MBP also has its weaknesses. For instance, a significant fraction (ca. 20%) of MBP fusion proteins fails to adhere to amylose resin, and even when amylose affinity chromatography works as well as can be expected, most of the time the fusion protein will still contain an unacceptably high level of contaminants.

In an effort to overcome these problems, we endeavored to incorporate supplementary affinity tags within the framework of an MBP fusion protein. A central concern was that these supplementary tags should not interfere with the ability of MBP to promote the solubility of its fusion partners. All attempts to insert additional affinity tags within MBP itself were unsuccessful, yielding predominantly insoluble protein even in the unfused state. By contrast, the addition of supplementary affinity tags to the N- or C-terminus of MBP did not reduce its solubility in the unfused state. However, differential effects on solubility were observed when the same N- and C-terminal tags were examined in the context of several different MBP fusion proteins. Whereas the MBP fusion proteins with an N-terminal BAP tag were just as soluble as their counterparts with no BAP tag, the fusion proteins with an N-terminal polyarginine tag were clearly less soluble. This result indicates that the nature of the supplementary tag, as well as its location, can influence MBP's ability to function as a solubilizing agent. Nearly all of the fusion proteins with an Arg-tag in the linker region between MBP and the passenger were less soluble than their counterparts with no polyarginine tag, but the configuration in which the Arg-tag is closer to the C-terminus of MBP was consistently more deleterious. Thus, of all the options investigated, the N-terminal BAP appears to be the most promising. It neither affects the yield of fusion protein nor interferes with MBP's ability to promote the solubility of its fusion partners. Moreover, as shown previously, it is also a very efficient substrate for enzymatic biotinylation *in vivo*, and biotinylated BAP-MBP fusion proteins bind reversibly to monomeric avidin resin [7].

We also investigated the effect of adding a supplementary His-tag to either the N- or C-terminus of the passenger protein on the solubility of several MBP fusion proteins. The C-terminus of the passenger is an attractive location for an additional tag because it creates an 'affinity sandwich' configuration, thereby enabling full-length fusion proteins to be separated from truncated products after successive affinity steps. However, a C-terminal tag would remain associated with the passenger protein after the N-terminal MBP moiety is removed by proteolytic digestion of the fusion protein. In view of this, we elected to use the His-tag for these experiments because the available evidence suggests that it is less likely to interfere with the crystallization of proteins than other small tags [22]. Although the presence of a C-terminal His-tag

reduced the solubility of the MBP-p16 and MBP-E6 fusion proteins, the effect was rather modest. Therefore, we believe that the advantage of the affinity sandwich would offset a slight reduction in the yield of soluble fusion protein that might occur under these circumstances. When His-tags were added to the N-termini of the passenger proteins, immediately after the TEV protease cleavage site, they seemed to have even less impact on the solubility of the MBP fusion proteins. One potential advantage of adding a His-tag to the N-terminus of the passenger protein, as opposed to its C-terminus, is that this may make the TEV protease cleavage site more accessible to the enzyme.

On the basis of the results presented here, we suggest that the most useful configuration for high-throughput protein expression and purification is likely to be an MBP fusion protein with a BAP on the N-terminus of MBP and/or a His-tag on the C-terminus of the passenger protein (i.e. BAP-MBP-passenger-His). Either MBP or the BAP could be used in conjunction with the C-terminal His-tag for affinity sandwich purification, after which the BAP-MBP moiety could be removed by digesting the fusion protein with TEV protease or a suitable alternative. Because it would be retained by the passenger protein, the C-terminal His-tag could be used to facilitate the separation of the passenger protein from the other byproducts of the TEV protease digest. In the event that the passenger protein failed to crystallize or if the fusion protein could not be cleaved by TEV protease because of steric occlusion, then the His-tag could be relocated to the N-terminus of the passenger protein instead. Although the advantage of the affinity sandwich would be lost, the purification protocol probably would not have to be altered because the passenger protein would still retain a His-tag after the fusion protein is cleaved by TEV protease.

By capitalizing on successive affinity interactions, it should be possible to develop a generic protocol that can be used to purify many different proteins. Moreover, because purification protocols that rely entirely on affinity interactions never require a protein to interact directly with a chromatographic matrix, even marginally stable proteins that would be difficult or impossible to purify by conventional methods (e.g. by ion exchange chromatography) may yield to these approaches. The development of MBP fusion proteins with supplemental affinity tags represents a useful step in this direction.

Acknowledgements

We wish to thank Dr. Jeffrey Fox and Paragon Bioservices, Inc. (Baltimore, MD) for constructing some of the plasmid expression vectors used in this work, and Dr. D. Eric Anderson for drawing our attention to the Arg-tag.

References

1. Baneyx, F. (1999) *Curr. Opin. Biotechnol.* **10**, 411–421.
2. Stevens, R.C. (2000) *Structure* **8**, R177–R185.
3. Kapust, R.B., and Waugh, D.S. (1999) *Prot. Sci.* **8**, 1668–74.
4. Pryor K.D., and Leiting, B. (1997) *Protein Expr. Purif.* **10**, 309–319.
5. Bach, H., Mazor, Y., Shaky, S., Shoham-Lev, A., Berdichevsky, Y., Gutnick, D.L., and Benhar, I. (2001) *J. Mol. Biol.* **312**, 79–93.
6. Edwards, A.M., Arrowsmith, C.H., Christendat, D., Dharamsi, A., Friesen, J.D., Greenblatt, J.F., and Vedadi, M. (2000) *Nat. Struct. Biol.* **7** Suppl., 970–972.
7. Tsao, K.-L., DeBarbieri, B., Michel, H., and Waugh, D.S. (1996) *Gene* **169**, 59–64.
8. Fox, J.D., Kapust, R.B., and Waugh, D.S. (2001) *Prot. Sci.* **10**, 622–630.
9. Ho, S.N, Hunt, H.D., Horton, R.M., Pullen, J.K., and Pease, L.R. (1989) *Gene* **77**, 51–59.
10. Fox, J.D., and Waugh, D.S. (2002) *Meth. Mol. Biol. in press*.
11. Parks, T.D., Leuther, K.K., Howard, E.D., Johnston, S.A., and Dougherty, W.G. (1994) *Anal. Biochem.* **216**, 413–417.
12. Sassenfeld, H.M., and Brewer, S.J. (1984) *Bio/Technology* **2**, 76–81.
13. Hochuli, E., Bannwarth, W., Dabeli, H., Gentz, R., and Stuber, D. (1988) *Bio/Technology* **6**, 1321–1325.
14. Schatz, P.J. (1993) *Bio/Technology* **11**, 1138–1143.
15. Kohanski, R.A., and Lane, M.D. (1990) *Meth. Enzymol.* **184**, 194–200.
16. Morag, E., Bayer, E.A., and Wilchek, M. (1996) *Anal. Biochem.* **243**, 257–263.
17. Betton, J.-M., Jacob, J.P., Hofnung, M., and Broome-Smith, J.K. (1997) *Bio/Technology* **15**, 1276–1279.
18. Martineau, P., Guillet, J.-G., Leclerc, C., and Hofnung, M. (1992) *Gene* **113**, 35–46.
19. Betton, J.-M., Martineau, P., Saurin, W., and Hofnung, M. (1993) *FEBS Letts.* **325**, 34–38.
20. Spurlino, J.C., Lu, G.-Y., and Quiocho, F.A. (1991) *J. Biol. Chem.* **266**, 5202–5219.
21. Evdokimov, A.G., Anderson, D.E., Routzahn, K.M., and Waugh, D.S. (2001) *J. Mol. Biol.* **305**, 891–904.
22. Bucher, M.H., Evdokimov, A.G., and Waugh, D.S. (2002) *Acta Crystallogr. D***58**, 392–397.
23. Berman, H.M., Westbrook, J., Feng, Z., Gilliland, G., Bhat, T.N., Weissig, H., Shindyalov, I.N., and Bourne, P.E. (2000) *Nucleic Acids Res.* **28**, 235–242.
24. Beckett, D., Kovaleva, E., and Schatz, P.J. (1999) *Prot. Sci.* **8**, 921–929.
25. Fall, R.R. (1979) *Meth. Enzymol.* **62**, 390–398.