Protein Science (1999), *8*:1668–1674. Cambridge University Press. Printed in the USA.

Escherichia coli maltose-binding protein is uncommonly effective at promoting the solubility of polypeptides to which it is fused

RACHEL B. KAPUST AND DAVID S. WAUGH

ABL-Basic Research Program, NCI-Frederick Cancer Research and Development Center, P.O. Box B, Frederick, Maryland 21702-1201

(RECEIVED March 15, 1999; ACCEPTED April 19, 1999)

Abstract

Although it is usually possible to achieve a favorable yield of a recombinant protein in *Escherichia coli*, obtaining the protein in a soluble, biologically active form continues to be a major challenge. Sometimes this problem can be overcome by fusing an aggregation-prone polypeptide to a highly soluble partner. To study this phenomenon in greater detail, we compared the ability of three soluble fusion partners—maltose-binding protein (MBP), glutathione S-transferase (GST), and thioredoxin (TRX)—to inhibit the aggregation of six diverse proteins that normally accumulate in an insoluble form. Remarkably, we found that MBP is a far more effective solubilizing agent than the other two fusion partners. Moreover, we demonstrated that in some cases fusion to MBP can promote the proper folding of the attached protein into its biologically active conformation. Thus, MBP seems to be capable of functioning as a general molecular chaperone in the context of a fusion protein. A model is proposed to explain how MBP promotes the solubility and influences the folding of its fusion partners.

Keywords: aggregation; fusion protein; glutathione S-transferase; inclusion bodies; maltose-binding protein; protein folding; solubility; thioredoxin

Fusion proteins have become a cornerstone of modern biological research, with an ever widening range of applications (reviewed by Uhlen et al., 1992). Unrelated proteins originally were fused together (at the genetic level) to facilitate the detection and/or purification of one partner (Uhlen et al., 1983). Thereafter, it became apparent that the yield of a recombinant protein often could be improved by producing it in the form of a fusion (Butt et al., 1989). Yet another unexpected attribute of fusion proteins has begun to emerge in recent years: Sometimes the solubility of a recombinant protein can be improved by fusing it to a highly soluble partner. This is an exciting development because it may offer a means of circumventing the "inclusion body problem," which is one of the greatest technical obstacles to the production of biologically active recombinant proteins in heterologous systems (Schein, 1989). Examples of fusion partners that have been touted as solubilizing agents include thioredoxin (TRX) (LaVallie et al., 1993), glutathione S-transferase (GST) (Nygren et al., 1994), maltose-binding protein (MBP) (Pryor & Leiting, 1997), Protein A (Samuelsson et al., 1994), ubiquitin (Power et al., 1990), and DsbA (Zhang et al., 1998).

Although widely recognized and potentially of great importance, this solubilizing effect remains poorly understood. It is not clear, for example, what characteristics besides intrinsically high solubility epitomize an effective solubilizing agent. Are all soluble fusion partners equally proficient at this task, or are some consistently more effective than others? Similarly, it is not known whether the solubility of many different polypeptides can be improved by fusing them to a highly soluble partner or whether this approach is only effective in a small fraction of cases. To address these fundamental questions, we compared the solubility of 18 different fusion proteins in *Escherichia coli*, collectively representing every combination of three highly soluble proteins and six very insoluble ones. Our results indicate that many aggregation-prone polypeptides can be rendered soluble by fusing them to an appropriate partner, but that some fusion partners are much better solubilizing agents than others. These findings have mechanistic implications, enabling us to propose a model that may explain why proteins that normally accumulate in an insoluble form often manage to avoid this fate when they are fused to an appropriate partner.

Reprint requests to: David S. Waugh, ABL-Basic Research Program, NCI-Frederick Cancer Research and Development Center, P.O. Box B, Frederick, Maryland 21702-1201; e-mail: waughd@ncifcrf.gov.

Results

Design and construction of fusion proteins

It seems reasonable to assume that insoluble proteins would be poor solubilizing agents, but are all highly soluble proteins equally effective? To answer this question, we selected three soluble fusion

partners for direct comparisons: *E. coli* His-patch thioredoxin (TRX), *Schistosoma japonicum* glutathione S-transferase (GST), and the mature form of *E. coli* maltose-binding protein (MBP). Not only are all of these proteins highly soluble in the *E. coli* cytoplasm, but they each exhibit an appreciable affinity for a specific ligand, which can be exploited to facilitate their purification. Accordingly, we refer to them as affinity domains. The affinity of His-patch thioredoxin for immobilized metal ions is an engineered trait that does not affect its performance as a solubilizing agent (Lu et al., 1996).

To evaluate and compare the efficacy of these three affinity domains as solubilizing agents, we fused each of them to six different polypeptides that normally accumulate predominantly or exclusively in an insoluble form in *E. coli* (Fig. 1). These "passenger proteins" are structurally, functionally, chemically, and evolutionarily diverse (Table 1). One of them is not a wild-type protein: CATΔ9 lacks nine amino acids that normally comprise the C-terminus of chloramphenicol acetyl transferase (CAT). However, this deletion, which has the impact of transforming an extraordinarily soluble protein into a completely insoluble one, is thought to exert its influence on the folding pathway rather than the native state (Robben et al., 1993). In all cases, the passenger proteins were fused to the C-termini of the affinity domains. Care was taken not to vary the sequence of the peptide linkers between the domains of the fusion proteins so that this factor could not influence the results.

Table 1. *Properties of passenger proteins* [a]

Protein	Activity	MW	pI	S-S
TIMP	MMP inhibitor	14 kDa	7.2	3
p16	Cdk4 inhibitor	16 kDa	6.0	0
E6	Oncoprotein	19 kDa	8.6	0
CATΔ9	Enzyme	25 kDa	6.6	0
GFP	Bioluminescence	27 kDa	6.2	0
TEV	Protease	29 kDa	8.5	0

[a] Abbreviations used: MW, molecular weight; pI, isoelectric point; S-S, number of disulfide bonds; TIMP, N-terminal inhibitory domain (residues 1-127) of human tissue inhibitor of metalloproteinases-2 (Williamson et al., 1994); MMP, matrix metalloproteinase; p16, human cyclin-dependent kinase 4 (Cdk4) inhibitor (Serrano et al., 1993); E6, oncoprotein encoded by human papillomavirus 18 (Vousden, 1993); CATΔ9, a form of chloramphenicol acetyl transferase in which nine residues have been deleted from the C-terminus (Robben et al., 1993); GFP, *Aequorea victoria* green fluorescent protein (Prasher et al., 1992); TEV, catalytic domain of the nuclear inclusion protease from tobacco etch virus (Parks et al., 1995).

Yield and solubility of fusion proteins

All of the fusion proteins were expressed at a high level in *E. coli* BL21/DE3 cells (Studier et al., 1991), with yields ranging between 15 and 35% of the total intracellular protein (Fig. 2A). The solubility of each fusion protein was estimated on the basis of data obtained by laser scanning densitometry of Coomassie-stained gels (Fig. 2B), as described (see Materials and methods).

The most striking result to emerge from these experiments is that the MBP fusion proteins invariably proved to be more soluble than their GST and TRX counterparts. The difference was considerable in nearly every instance. The only MBP fusion protein that was not mostly soluble at 37 °C is MBP-CATΔ9. However, CATΔ9 is not a wild-type protein and may not be capable of adopting a stably folded conformation (Robben et al., 1993). Notwithstanding this, in stark contrast to the TRX-CATΔ9 and GST-CATΔ9 fusion proteins, the majority of the MBP-CATΔ9 fusion protein accumulated in soluble form at 25 °C (data not shown). In fact, we observed that lower temperature improved the solubility of several fusion proteins, a phenomenon previously reported for unfused proteins expressed in *E. coli* (Schein & Noteborn, 1988). Nonetheless, the MBP fusion proteins were always markedly more soluble than their GST and TRX counterparts, irrespective of the temperature at which the experiment was performed.

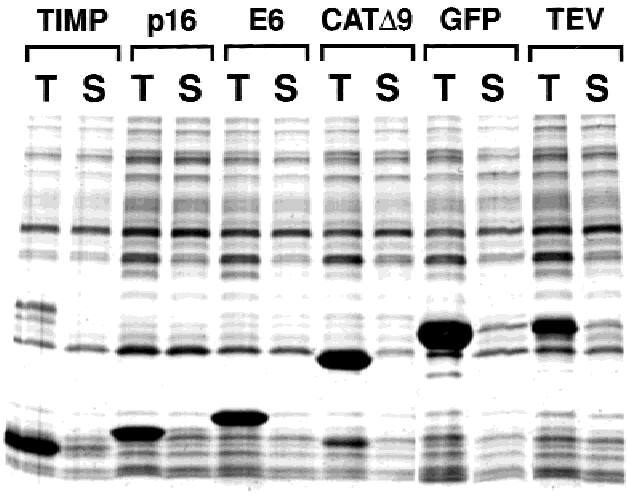

Fig. 1. Insolubility of passenger proteins produced in an unfused form in *E. coli*. Samples of the total (T) and soluble (S) intracellular protein fractions from cells producing each passenger protein were prepared as described (Materials and methods), and the results were analyzed by SDS-PAGE. Abbreviations are defined in Table 1.

The data also substantiate a second, more subtle revelation. The (smaller) TRX fusion proteins were marginally but consistently more soluble in *E. coli* than were their GST counterparts, and TRX was able to improve the solubility of some proteins that are larger than it is (e.g., TIMP and TEV). Taken together, these results indicate that the solubility of a fusion protein is not determined solely by the relative sizes of the soluble and insoluble fusion partners.

Fusion to MBP can promote the proper folding of passenger proteins

Whether or not a passenger protein can attain its biologically active conformation once it is rendered soluble by fusion to MBP is a question of considerable practical importance. We could demonstrate that cells expressing the MBP-CATΔ9 fusion protein confer a modest degree of resistance to chloramphenicol (ca. 30 μg/mL at 37 °C and 70 μg/mL at 25 °C), preparations of MBP-GFP fusion protein are fluorescent, and the MBP-E6 fusion protein can promote the ubiquitin-dependent degradation of p53 (data not shown). Therefore, some fraction of these passenger proteins must be properly folded.

A more powerful demonstration of the benefit that can be achieved by fusing proteins to MBP was possible in the case of TEV protease (Fig. 3). In cells expressing polyhistidine-tagged TEV pro-

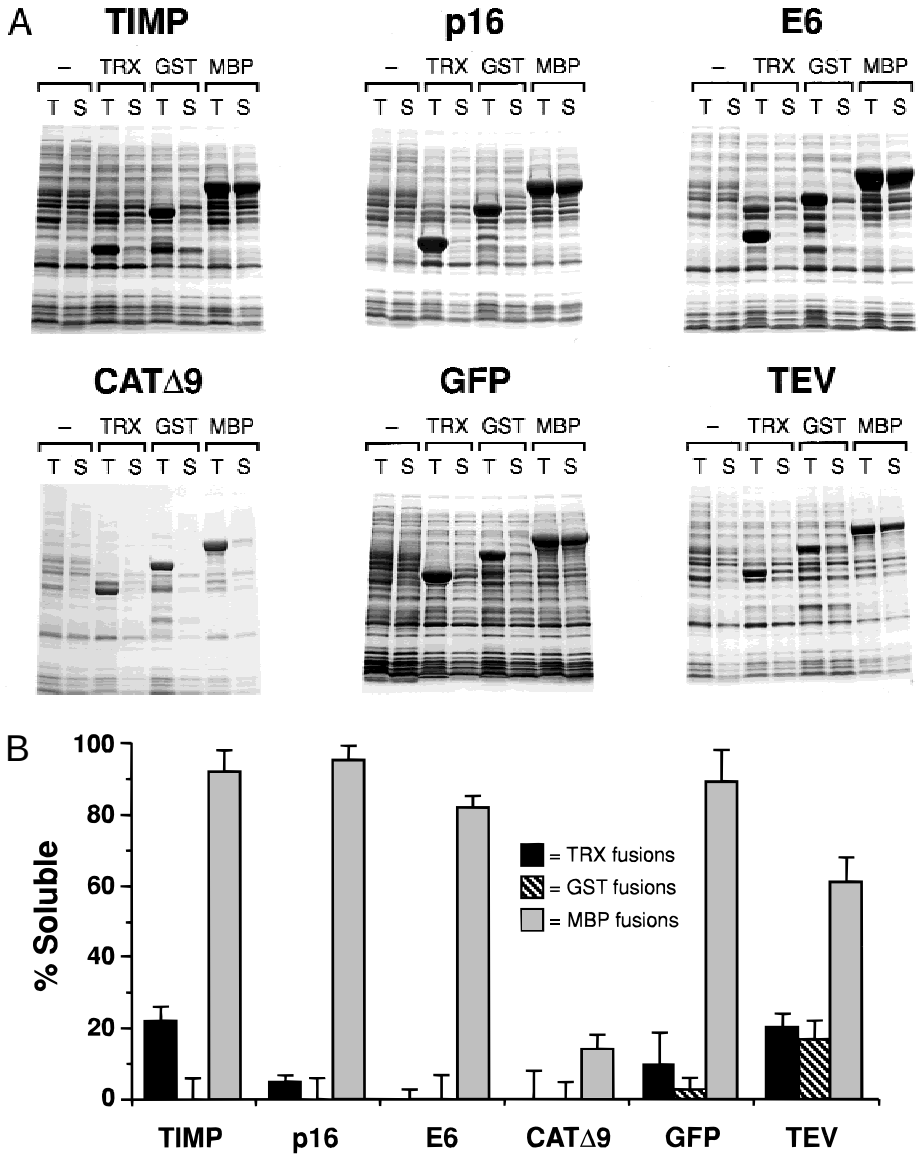

Fig. 2. Solubility of TRX, GST, and MBP fusion proteins. **A:** Samples of the total and soluble intracellular protein fractions from cells producing various fusion proteins were prepared as described (see Materials and methods) and analyzed by SDS-PAGE. **B:** The solubility of each fusion protein (% soluble) was estimated from data obtained by laser scanning densitometry of Coomassie-stained gels, as described (see Materials and methods). Values represent the mean of 3–7 independent experiments. Errors are reported in the form of standard deviations. In a few cases, the mean was slightly less than (but not more than 1 standard deviation from) 0%; these numbers were rounded to 0%. Abbreviations: MBP, maltose-binding protein (mature form); GST, glutathione S-transferase; TRX, His-patch thioredoxin; T, total intracellular protein; S, soluble intracellular protein; —, samples from cells containing no protein expression vector. Other abbreviations are defined in Table 1.

tease (His-TEV) directly (i.e., in an unfused form), a substantial amount of protein accumulated in response to isopropyl-1-thio-β-D-galactopyranoside (IPTG) induction but nearly all of this material was insoluble and inactive (Fig. 3B, Lanes 3 and 4). On the other hand, when TEV was produced in the form of an MBP fusion protein, the majority of the fusion protein was soluble in the crude cell extract (Fig. 3B, Lanes 5 and 6). The most interesting result was obtained when an MBP-TEV fusion protein containing a canonical TEV recognition site in the linker between the two domains was produced under the same conditions (Fig. 3B, Lanes 7 and 8). In this instance, all of the fusion protein was processed intracellularly to yield separate MBP and His-TEV domains. Strikingly, in contrast to the result obtained when His-TEV was produced in an unfused form, nearly all of the processed His-TEV was soluble in the crude cell extract. Moreover, its specific activity was indistinguishable from that of the soluble His-TEV expressed in an unfused form or purchased from a commercial vendor (data not shown). It is noteworthy that there was about as much soluble His-TEV in these cells as there was insoluble His-TEV in cells producing the protein in an unfused form. This indicates that what limits the solubility of His-TEV in *E. coli* is not aggregation of the native state of the protein; rather, it must be a folding intermediate

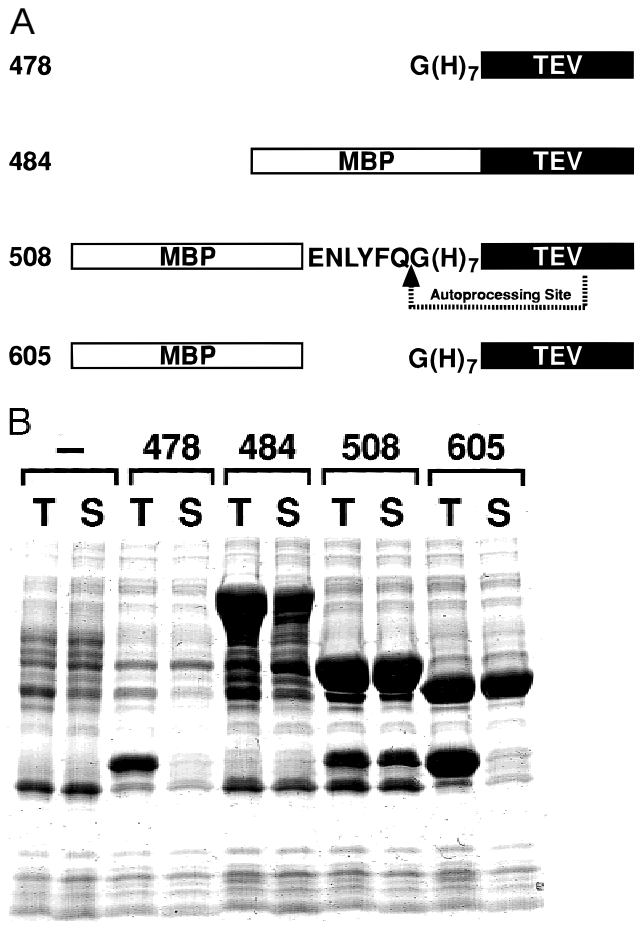

Fig. 3. Proper folding of TEV protease is promoted by fusion to MBP. **A:** Schematic illustration of the proteins produced by four different TEV expression vectors (not to scale): 478, His-tagged TEV protease; 484, MBP-TEV fusion protein; 508, self-processing MBP-TEV fusion protein; 605, separate MBP and His-tagged TEV protease domains synthesized from a single, dicistronic mRNA. **B:** Samples of the total and soluble intracellular protein fractions from cells containing the TEV expression vectors in **A** were prepared as described (see Materials and methods) and analyzed by SDS-PAGE.

that is prone to form insoluble aggregates. This experiment also demonstrates that His-TEV is not just temporarily immune from aggregation while it is associated with MBP. On the contrary, most of the His-TEV remains soluble after cleavage in vivo, an event that occurs very rapidly (data not shown). However, covalent fusion of the two domains is required, at least transiently, to promote the solubility (i.e., proper folding) of His-TEV. When similar amounts of the separate MBP and His-TEV domains were produced simultaneously from a dicistronic mRNA, no improvement in the solubility of His-TEV was observed (Fig. 3B, Lanes 9 and 10).

Discussion

Our results indicate that MBP is a much better solubilizing agent than either GST or TRX. Because not all soluble proteins perform this task equally well, we have to modify the simple idea that fusing an aggregation-prone protein to any highly soluble partner will routinely give rise to a soluble fusion protein. Rather, it appears that only a few soluble proteins, such as MBP, have the ability to act in this capacity. A larger number of soluble fusion partners will have to be tested to ascertain just how unique MBP is in this regard. On the other hand, the solubility of all six passenger proteins that we tested was markedly improved by fusing them to MBP. This result demonstrates that a single solubilizing agent can be effective in conjunction with a wide variety of insoluble partners.

How do proteins that normally accumulate in an insoluble form in *E. coli* avoid this fate when they are fused to MBP? The answer could depend, in part, on whether aggregation occurs before or after a protein adopts its native conformation, since these two modes of aggregation may be affected differently or to varying degrees by fusion to MBP and other soluble partners. Thus, it will be difficult to draw any firm conclusions about the mechanism(s) of solubilization unless we understand why the passenger proteins form insoluble aggregates in the first place. We have presented compelling evidence that aggregation occurs during, rather than after, the folding of TEV protease in *E. coli* (Fig. 3), and ample proof exists that the same is true of GFP (Cormack et al., 1996; Crameri et al., 1996; Siemering et al., 1996; Yang et al., 1996; Kimata et al., 1997). On the other hand, comparatively little is known about what causes the aggregation of other passenger proteins we have studied. Thus, although we can conclude that fusion to MBP (but not GST or TRX) has a beneficial impact on the folding of some proteins in *E. coli*, we cannot say with certainty what effect it may have on proteins that form insoluble aggregates in their native state. This question is clearly something that merits further investigation.

To explain how MBP influences the folding of its fusion partners, we propose that it can function as a general molecular chaperone in the context of a fusion protein (Fig. 4). According to this model, nascent fusion proteins initially adopt a form (Folding Intermediate) in which the affinity domain is properly folded but the passenger protein is not. If the passenger protein subsequently attains its native conformation, then this gives rise to a soluble fusion protein (Native Structure). Alternatively, incompletely folded passenger proteins can self-associate to form insoluble fusion protein aggregates (Insoluble Aggregates). The fate of the Folding Intermediate depends, in large measure, on its concentration inside the cell. A high concentration, such as typically occurs in our experiments (Fig. 2A), tends to favor intermolecular association (aggregation), whereas the unimolecular folding reaction is more prevalent at lower concentrations. However, in the case of the MBP fusion proteins, we propose that this intermediate can rearrange into a form in which a physical interaction exists between MBP and the incompletely folded passenger protein (Sequestered Intermediate), effectively occluding its self-association. Although weak and reversible, this nonspecific association is promoted by the close proximity of the interacting partners in the context of a fusion protein. Consequently, the concentration of the aggregation-prone intermediate in the cell at any given time is low, so the formation of Insoluble Aggregates is avoided.

In accord with this proposal, MBP has been shown to interact preferentially with unfolded proteins and to promote their folding in vitro (Richarme & Caldas, 1997), albeit less efficiently than we have observed in the context of a fusion protein. Additionally, others have reported that fusion to MBP can greatly improve the yield of soluble protein obtained after refolding (Sachdev & Chirgwin, 1998a) and evidence of tertiary noncovalent interactions

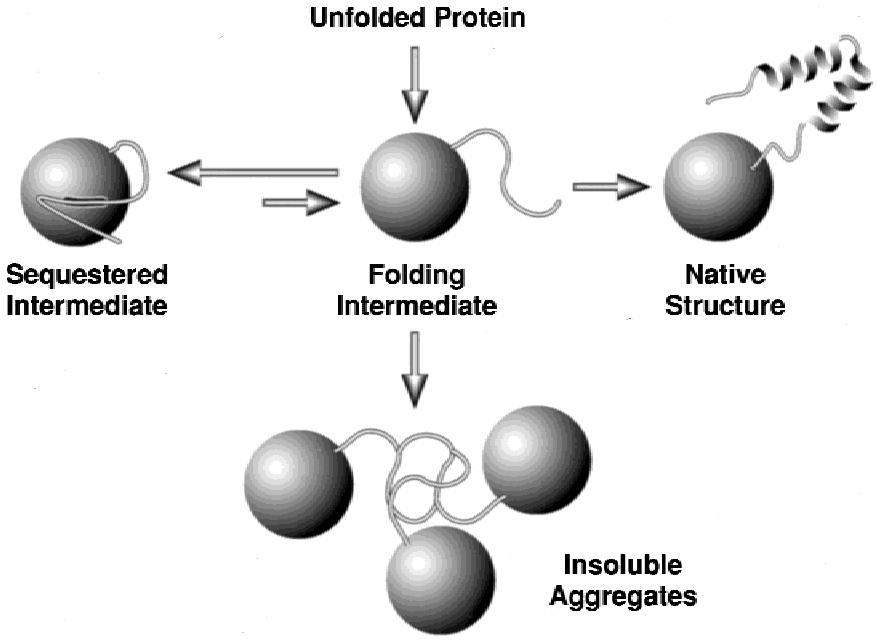

Fig. 4. A model to explain how MBP can improve the solubility and promote the proper folding of its fusion partners. See text for details. The sphere represents native (properly folded) MBP; the attached strings and helix represent the incompletely folded and native states of a passenger protein, respectively.

between MBP and its fusion partners has been uncovered serendipitously in at least two instances (Blondel et al., 1996; Lorenzo et al., 1997). Finally, another study has demonstrated that in *E. coli* the solubility of MBP fusion proteins depends on the order in which the two domains are synthesized; soluble fusion proteins are obtained only when the passenger proteins are fused to the C-terminus of MBP (Sachdev & Chirgwin, 1998b). This observation is consistent with the involvement of a folding intermediate of the sort that we have proposed (Fig. 4).

We suggest that MBP plays a passive role in protein folding, merely serving to prevent the off-pathway aggregation of intermediates in the folding process. The iterative interaction of MBP with its fusion partners might eventually steer them toward their native conformations, but only if they are able to fold spontaneously in the *E. coli* cytoplasm. The results reported here, together with diverse examples from the literature (Derbyshire & Grindley, 1992; Mottershead et al., 1996; Rao & Bodley, 1996; Thomas et al., 1996; Perez-Martin et al., 1997; Pryor & Leiting, 1997), indicate that this can occur in some instances. However, other examples of proteins that can be solubilized by fusion to MBP but evidently do not attain their native, biologically active conformations have been reported (Louis et al., 1991; Saavedra-Alanis et al., 1994; Sachdev & Chirgwin, 1998b) and will surely be encountered in the future. Such fusion proteins may exist principally in the form of soluble, sequestered folding intermediates (Fig. 4). If so, then perhaps they can be exploited to study nonnative folding states of proteins that would otherwise be inaccessible due to aggregation and insolubility. Moreover, it might be possible to extend the utility of this approach even further, on a case by case basis, by deliberately manipulating conditions to promote folding. For example, a protein that contains disulfide bonds may not adopt a stable fold in the unfavorable redox environment of the *E. coli* cytoplasm but might be able to do so if the fusion protein is exposed to the appropriate redox conditions for a suitable period of time. At present, we do not fully understand how often and efficiently passenger proteins attain their native conformations once they are rendered soluble by fusion to MBP. However, the available evidence suggests that proper folding of passenger proteins is not a rare occurrence.

How might MBP interact with its fusion partners? If they are bound in a nonspecific fashion and in an incompletely folded form, then hydrophobic interactions could play an important role. There are several clusters of hydrophobic residues on the surface of MBP (Spurlino et al., 1991), some of which are used to contact other proteins in the maltose transport apparatus (Martineau et al., 1990). One feature that distinguishes MBP from the other soluble fusion partners examined in this study is its deep hydrophobic cleft, which serves as the maltose-binding site. We note that the *E. coli* chaperonin GroEL utilizes a hydrophobic cleft to interact with its targets (Fenton et al., 1994; Buckle et al., 1997). Any or all of these hydrophobic zones on the surface of MBP could serve as binding sites for incompletely folded passenger proteins.

In summary, our results indicate that MBP is a much more effective solubilizing agent than the other (equally soluble) affinity domains that we tested, and that a wide variety of aggregation-prone polypeptides can be recovered in soluble form as MBP fusion proteins. Moreover, we have demonstrated that sometimes MBP can promote the proper folding of its fusion partners. These chaperone-like qualities serve to distinguish MBP from other affinity domains and enhance its value as a fusion partner. Admittedly, however, not every protein can be efficiently solubilized by fusion to MBP (Chen & Gouaux, 1996; Hering et al., 1996; Aoki et al., 1998; Reddy et al., 1998). Further characterization of these exceptions to the rule may provide important clues about the mechanism of solubilization and yield further insight into the pathways of protein aggregation in vivo.

Materials and methods

Plasmid expression vectors

To construct vectors for the production of passenger proteins in an unfused form (Fig. 1), the appropriate open reading frames (ORFs) were amplified by the polymerase chain reaction (PCR), using either natural or synthetic gene sequences as templates. The PCR products were cleaved with *Nco*I and *Bam*HI (or *Nco*I and *Bgl*II in the case of E6), and then ligated with the *Nco*I/*Bam*HI backbone of pET3d or pET11d (Novagen, Madison, Wisconsin). The N-terminus of the TEV catalytic domain ORF was extended to encode the amino acid sequence MGHHHHHHH during the construction of the His-TEV vector (pDW478). The derivative of pET11d that produces GFP was a gift from Gottfried Palm.

The immediate precursor of the GST fusion vectors (pDW418) was constructed by annealing two synthetic oligodeoxyribonucleotides (5′-GATCGCGAGCTCGGCCATGGTACGTAGGCCTAAG CTTGGATCCTCGAG-3′ and 5′-AATTCTCGAGGATCCAAGCT TAGGCCTACGTACCATGGCCGAGCTCGC-3′) and ligating the double-stranded cassette with the *Bam*HI/*Eco*RI vector backbone of pGEX-3X (Pharmacia, Piscataway, New Jersey). The immediate precursor of the TRX fusion vectors (pDW480) was constructed by annealing two different synthetic oligodeoxyribonucleotides (5′-GATCGAGCTCCATGGTACGTAGGCCTGGATCCTCGAG AATT-3′ and 5′-CTAGAATTCTCGAGGATCCAGGCCTACGTA CCATGGAGCTC-3′) and ligating the double-stranded cassette with the *Bam*HI/*Xba*I vector fragment of pThioHis A (Invitrogen, San Diego, California). The nucleotide sequences of both inserts were confirmed experimentally.

The MBP fusion vectors were constructed first; ORFs encoding the six passenger proteins (Table 1) were amplified by PCR and inserted between the unique *Xmn*I and *Bam*HI sites in pMal-C2 (New England Biolabs, Beverly, Massachusetts). Once the nucleotide sequences of these inserts were confirmed, ORFs encoding the passenger proteins, together with the interdomain linker peptide encoded by the pMal-C2 vector, were excised with *Sac*I and *Bam*HI. The MBP-E6 fusion vector, however, had to be cleaved with *Sac*I and *Sal*I instead, due to the presence of a *Bam*HI site within the E6 gene. These restriction fragments were then ligated with the *Sac*I/*Bam*HI (or *Sac*I/*Xho*I) backbones of pDW418 and pDW480 to construct the corresponding GST and TRX fusion vectors, respectively.

A two-step PCR strategy was used to construct the self-cleaving MBP-TEV fusion vector (pRK508). First, two separate PCR reactions were performed using the MBP-TEV fusion vector (pDW484) as the template: one with PE-29 (5′-GATGAAGCCCT GAAAGACGCGCAG-3′) and PE-78 (5′-ATCACCTTGAAAATAA AGATTTTCTCCCCTTCCCTCGATCCCGAGGTTGTTG-3′) as the primers, and the other with PE-45 (5′-GAAAATCTTTATTTT CAAGGTCATCATCATCATCATCATGGAGAAAGCTTGT TTAAGGGGCCGCGTGATTACA-3′) and PE-30 (5′-GCAAGG CGATTAAGTTGGGTAACGC-3′) as the primers. These two overlapping PCR products then were combined and used as the template for another PCR amplification, this time using only the outer primers (PE-29 and PE-30). This approach created a PCR product that included the entire TEV coding sequence preceded by a polyhistidine tract and a TEV cleavage site and flanked by sequences that were present in the original template. This PCR fragment was cleaved with *Sac*I and *Bam*HI, and then ligated with the *Sac*I/*Bam*HI vector backbone of pMal-C2. The nucleotide sequence of pRK508 was confirmed experimentally. The vector used for the simultaneous production of MBP and TEV from a dicistronic mRNA (pRK605) was constructed by ligating an *Xba*I/*Sca*I fragment of the vector used to produce TEV protease in an unfused form with the *Xba*I/*Sca*I backbone of pDW533, a plasmid that produces the mature form of MBP under *tac* promoter control but is otherwise identical to pMal-C2.

Protein expression and SDS-PAGE analysis

The TEV protease coding sequence contains a number of arginine codons that are not commonly used in *E. coli* (AGG and AGA). To obtain a high yield of recombinant protein, TEV vectors must be used in conjunction with a compatible plasmid that constitutively overproduces the cognate tRNA (*argU*), such as pDC952 (Calderone et al., 1996). The presence of this plasmid has no impact on the yield or solubility of the other passenger proteins or fusion proteins studied here. However, all cells contained pDC952 so that the experimental conditions would be uniform.

Cells from single, drug resistant colonies of *E. coli* BL21/DE3 containing one of the protein expression vectors and pDC952 were grown to saturation in LB broth (Miller, 1972) supplemented with 100 μg/mL ampicillin and 30 μg/mL chloramphenicol at 37 °C. The saturated cultures were diluted 50-fold in the same medium and grown in shake-flasks to mid-log phase (A_{600} = 0.5–0.7), at which time IPTG was added to a final concentration of 1 mM. After 3 h, the cells were recovered by centrifugation. The cell pellets were resuspended in 0.1 culture volumes of lysis buffer (50 mM Tris-HCl (pH 8.0), 150 mM NaCl, 1 mM EDTA), and disrupted by sonication. A total protein sample was collected from the cell suspension after sonication, and a soluble protein sample was collected from the supernatant after the insoluble debris was pelleted by centrifugation (20,000 × *g*). These samples were subjected to SDS-PAGE and proteins were visualized by staining with Coomassie Brilliant Blue.

Densitometry

At least three independent experiments were performed to obtain numerical estimates of the solubility of each fusion protein in *E. coli*. Coomassie-stained gels were scanned with a Molecular Dynamics Personal Densitometer and the pixel densities of the bands corresponding to the fusion proteins were obtained directly by volumetric integration. In each lane, the collective density of all *E. coli* proteins that are larger than the largest fusion protein was also determined by volumetric integration and used to normalize the values in each lane relative to the others. The percent solubility of each fusion protein (Fig. 2B) was calculated by dividing the amount of soluble fusion protein by the total amount of fusion protein in the cells, after first subtracting the normalized background values obtained from negative control lanes (cells containing no expression vector). Descriptive statistical data (e.g., the mean and standard deviation) were generated by Microsoft Excel.

Acknowledgments

We thank Kwei-Lan Tsao and Siobain Duffy for expert technical assistance. We are also grateful to Michael Kubbutat and Karen Vousden for assaying the bioactivity of the MBP-E6 fusion protein. Anne Arthur and Richard Frederickson assisted with the preparation of this manuscript. This research was sponsored by the National Cancer Institute, DHHS, under contract with ABL.

References

Aoki T, Tomaki E, Satoh M, Takashiro M, Onagi H, Itoh M, Teramoto T, Morikawa J, Wantabe H. 1998. Purification of recombinant human pepsinogens and their application as immunoassay standards. *Biochem Mol Biol Int 45*:289–301.

Blondel A, Nageotte R, Bedoulle H. 1996. Destabilizing interactions between the partners of a bifunctional fusion protein. *Protein Eng 9*:231–238.

Buckle AM, Zahn R, Fersht AR. 1997. A structural model for GroEL-polypeptide recognition. *Proc Natl Acad Sci USA 94*:3571–3575.

Butt TR, Jonnalagadda S, Monia BP, Sternberg EJ, Marsh JA, Stadel JM, Ecker DJ, Crooke ST. 1989. Ubiquitin fusion augments the yield of cloned gene products in *Escherichia coli. Proc Natl Acad Sci USA 86*:2540–2544.

Calderone TL, Stevens RD, Oas TG. 1996. High-level misincorporation of lysine for arginine at AGA codons in a fusion protein expressed in *Escherichia coli. J Mol Biol 262*:407–412.

Chen GQ, Gouaux JE. 1996. Overexpression of bacterio-opsin in *Escherichia coli* as a water-soluble fusion to maltose-binding protein: Efficient regeneration of the fusion protein and selective cleavage with trypsin. *Protein Sci 5*:456–467.

Cormack BP, Valdivia RH, Falkow S. 1996. FACS-optimized mutants of the green fluorescent protein (GFP). *Gene 173*:33–38.

Crameri A, Witehorn EA, Tate E, Stemmer WPC. 1996. Improved green fluorescent protein by molecular evolution using DNA shuffling. *Nat Biotechnol 14*:315–319.

Derbyshire KM, Grindley NDF. 1992. Binding of the IS903 transposase to its inverted repeat in vitro. *EMBO J 11*:3449–3455.

Fenton WA, Kashi Y, Furtak K, Horwich AL. 1994. Residues in chaperonin GroEL required for polypeptide binding and release. *Nature 371*:614–619.

Hering TM, Kollar J, Huynh TD, Varelas JB. 1996. Purification and characterization of decoring core protein expressed in *Escherichia coli* as a maltose-binding protein fusion. *Anal Biochem 240*:98–108.

Kimata Y, Iwaki M, Lim CR, Kohno K. 1997. A novel mutation which enhances the fluorescence of green fluorescent protein at high temperatures. *Biophys Biochem Res Comm 232*:69–73.

LaVallie ER, DiBlasio EA, Kovacic S, Grant KL, Schendel PF, McCoy JM. 1993. A thioredoxin gene fusion expression system that circumvents inclusion body formation in the *E. coli* cytoplasm. *BioTechnology 11*:187–193.

Lorenzo HK, Farber D, Germain V, Acuto O, Alzari PM. 1997. The MBP fusion protein restores the activity of the first phosphatase domain of CD45. *FEBS Lett 411*:231–235.

Louis JM, McDonald RA, Nashed NT, Wondrak EM, Jerina DM, Oroszlan S, Mora PT. 1991. Autoprocessing of the HIV-1 protease using purified wild-type and mutated fusion proteins expressed at high levels in *Escherichia coli. Eur J Biochem 199*:361–369.

Lu Z, DiBlasio-Smith EA, Grant KL, Warne NW, LaVallie ER, Collins-Racie LA, Follettie MT, Williamson MJ, McCoy JM. 1996. Histidine patch thioredoxins: Mutant forms of thioredoxin with metal chelating affinity that provide for convenient purification of thioredoxin fusion proteins. *J Biol Chem 271*:5059–5065.

Martineau P, Saurin W, Hofnung M, Spurlino JC, Quiocho FA. 1990. Progress in the identification of interaction sites on the periplasmic maltose-binding protein from *E. coli*. *Biochimie 72*:397–402.

Miller JH. 1972. *Experiments in molecular genetics.* Cold Spring Harbor, New York: Cold Spring Harbor Press. p 433.

Mottershead DG, Polly P, Lyons RJ, Sutherland RL, Watts CK. 1996. High activity, soluble, bacterially expressed human vitamin D receptor and its ligand binding domain. *J Cell Biochem 61*:325–337.

Nygren P-A, Stahl S, Uhlen M. 1994. Engineering proteins to facilitate bioprocessing. *Trends Biotechnol 12*:184–188.

Parks TD, Howard ED, Wolpert TJ, Arp DJ, Dougherty WG. 1995. Expression and purification of a recombinant tobacco etch virus NIa proteinase: Biochemical analyses of the full-length and a naturally occurring truncated proteinase form. *Virology 210*:194–201.

Perez-Martin J, Cases I, de Lorenzo V. 1997. Design of a solubilization pathway for recombinant polypeptides in vivo through processing of a bi-protein with a viral protease. *Protein Eng 10*:725–730.

Power RF, Conneely OM, McDonnell DP, Clark JH, Butt TR, Schrader WT, O'Malley BW. 1990. High level expression of a truncated chicken progesterone receptor in *Escherichia coli. J Biol Chem 265*:1419–1424.

Prasher DC, Eckenrode VK, Ward WW, Prendergast FG, Cormier MJ. 1992. Primary structure of the *Aequorea victoria* green-fluorescent protein. *Gene 111*:229–233.

Pryor KD, Leiting B. 1997. High-level expression of soluble protein in *Escherichia coli* using a His_6-tag and maltose-binding-protein double-affinity fusion system. *Protein Exp Purif 10*:309–319.

Rao S, Bodley JW. 1996. Expression, purification, and characterization ofthe G domain of *Saccharomyces cerevisiae* elongation factor 2. *Protein Exp Purif 8*:91–96.

Reddy A, Grimwood BG, Plummer TH, Tarentino AL. 1998. High-level expression of the Endo-beta-N-acetylglucosaminidase F2 gene in *E. coli*: One step purification to homogeneity. *Glycobiology 8*:633–636.

Richarme G, Caldas TD. 1997. Chaperone properties of the bacterial periplasmic substrate-binding proteins. *J Biol Chem 272*:15607–15612.

Robben J, Van der Schueren J, Volckaert G. 1993. Carboxyl terminus is essential for intracellular folding of chloramphenicol acetyltransferase. *J Biol Chem 268*:24555–24558.

Saavedra-Alanis VM, Rysavy P, Rosenberg LE, Kalousek F. 1994. Rat liver mitochondrial processing peptidase: Both alpha- and beta-subunits are required for activity. *J Biol Chem 269*:9284–9288.

Sachdev D, Chirgwin JM. 1998a. Solubility of proteins isolated from inclusion bodies is enhanced by fusion to maltose-binding protein or thioredoxin. *Protein Exp Purif 12*:122–132.

Sachdev D, Chirgwin JM. 1998b. Order of fusions between bacterial and mammalian proteins can determine solubility in *Escherichia coli*. *Biophys Biochem Res Comm 244*:933–937.

Samuelsson E, Moks T, Nilsson B, Uhlen M. 1994. Enhanced in vitro refolding of insulin-like growth factor I using a solubilizing fusion partner. *Biochemistry 33*:4207–4211.

Schein CH. 1989. Production of soluble recombinant proteins in bacteria. *BioTechnology 7*:1141–1148.

Schein CH, Noteborn MHM. 1988. Formation of soluble recombinant proteins in *Escherichia coli* is favored by lower growth temperature. *BioTechnology 6*:291–294.

Serrano M, Hannon GJ, Beach D. 1993. A new regulatory motif in cell-cycle control causing specific inhibition of cyclin D/CDK4. *Nature 366*:704–707.

Siemering KR, Golbik R, Sever R, Haseloff J. 1996. Mutations that suppress the thermosensitivity of green fluorescent protein. *Curr Biol 6*:1653–1663.

Spurlino JC, Lu G-Y, Quiocho FA. 1991. The 2.3 Å resolution structure of the maltose- or maltodextrin-binding protein, a primary receptor of bacterial active transport and chemotaxis. *J Biol Chem 266*:5202–5219.

Studier FW, Rosenberg AH, Dunn JJ, Dubendorff JW. 1991. Use of T7 RNA polymerase to direct expression of cloned genes. *Methods Enzymol 185*: 60–89.

Thomas S, Soriano S, d'Santos C, Banting G. 1996. Expression of recombinant rat myo-inositol 1,4,5-triphosphate 3-kinase B suggests a regulatory role for its N-terminus. *Biochem J 319*:713–716.

Uhlen M, Forsberg G, Moks T, Hartmanis M, Nilsson B. 1992. Fusion proteins in biotechnology. *Curr Opin Biotechnol 3*:363–369.

Uhlen M, Nilsson B, Guss B, Lindberg M, Gatenbeck S, Philipson L. 1983. Gene fusion vectors based on the gene for staphylococcal protein A. *Gene 23*:369–378.

Vousden K. 1993. Interactions of human papillomavirus transforming proteins with the products of tumor suppressor genes. *FASEB J 7*:872–879.

Williamson RA, Martorell G, Carr MD, Murphy G, Docherty AJP, Freedman RB, Feeney J. 1994. Solution structure of the active domain of tissue inhibitor of metalloproteinases-2. A new member of the OB fold protein family. *Biochemistry 33*:11745–11759.

Yang T-T, Cheng L, Kain SR. 1996. Optimized codon usage and chromophore mutations provide enhanced sensitivity with green fluorescent protein. *Nucleic Acids Res 24*:4592–4593.

Zhang Y, Olsen DR, Nguyen KB, Olson PS, Rhodes ET, Mascarenhas D. 1998. Expression of eukaryotic proteins in soluble form in *Escherichia coli*. *Protein Exp Purif 12*:159–165.