Maltodextrin-binding proteins from diverse bacteria and archaea are potent solubility enhancers

Jeffrey D. Fox, Karen M. Routzahn, Matthew H. Bucher, David S. Waugh*

Protein Engineering Section, Macromolecular Crystallography Laboratory, National Cancer Institute at Frederick, P.O. Box B, Frederick, MD 21702-1201, USA

Received 15 October 2002; revised 17 January 2003; accepted 17 January 2003

First published online 29 January 2003

Edited by Takashi Gojobori

Abstract ***Escherichia coli*** **maltose-binding protein (MBP) is frequently used as an affinity tag to facilitate the purification of recombinant proteins. An important additional attribute of MBP is its remarkable ability to enhance the solubility of its fusion partners. MBPs are present in a wide variety of microorganisms including both mesophilic and thermophilic bacteria and archaea. In the present study, we compared the ability of MBPs from six diverse microorganisms (*** ***E. coli*** **, *** ***Pyrococcus furiosus*** **, *** ***Thermococcus litoralis*** **, *** ***Vibrio cholerae*** **, *** ***Thermotoga maritima*** **, and *** ***Yersinia pestis*** **) to promote the solubility of eight different aggregation-prone proteins in *** ***E. coli*** **. In contrast to glutathione S-transferase (GST), all of these MBPs proved to be effective solubility enhancers and some of them were even more potent solubilizing agents than *** ***E. coli*** **MBP.**

Key words: Maltose-binding protein; MBP; Inclusion body; Fusion protein; Solubility enhancer

1. Introduction

The expression of recombinant proteins in *Escherichia coli* is frequently hampered by poor solubility. Sometimes insoluble proteins that are deposited in the form of inclusion bodies can be denatured and refolded, but this is an uncertain and time-consuming undertaking. One way to circumvent the problem of inclusion body formation is to exploit the innate ability that certain proteins have to enhance the solubility of their fusion partners. Originally it was presumed that virtually any highly soluble protein could function as a general solubilizing agent, but this has not turned out to be the case. In a systematic comparison, *E. coli* maltose-binding protein (MBP) proved to be a much more effective solubility enhancer than glutathione S-transferase or thioredoxin despite the fact that all three of these proteins are highly soluble [1]. It appears, therefore, that the ability of one fusion partner to promote the solubility of another is a relatively rare trait.

We reasoned that one way to gain some insight into the properties of *Eco* MBP that make it such an unusually effective solubilizing agent would be to compare the ability of orthologous proteins to promote solubility; perhaps some patterns would emerge that would reveal clues about the mechanism of the solubilizing effect. Moreover, because there is no a priori reason to believe that nature optimized *Eco* MBP for this task, we wondered if MBPs from other organisms might be even better solubilizing agents. At the same time, we were also curious to know if there is any correlation between the thermostability of a protein and its ability to function as an effective solubilizing agent. Accordingly, we selected five orthologs exhibiting varying degrees of amino acid sequence identity with *Eco* MBP (*Yersinia pestis* (*Ype*), 85%; *Vibrio cholerae* (*Vch*), 68%; *Thermotoga maritima* (*Tma*), 35%; *Thermococcus litoralis* (*Tli*), 30%; and *Pyrococcus furiosus* (*Pfu*), 27%) and compared their ability to promote the solubility of eight different aggregation-prone proteins in *E. coli*. The *Tma*, *Tli*, and *Pfu* MBPs were from hyperthermophiles, whereas the *Eco*, *Ype*, and *Vch* MBPs were of mesophilic origin. An alignment of the six MBP sequences is included in the web supplement (doi: 10.1016/S0014-5793(03)00070-X).

2. Materials and methods

2.1. Native expression vectors

Native *E. coli* (*Eco*) MBP was expressed from pDW533 [2]. The open reading frames (ORFs) encoding the mature MBPs (without their N-terminal signal peptides) from *Y. pestis* (*Ype*), *T. maritima* (*Tma*), *T. litoralis* (*Tli*), *P. furiosus* (*Pfu*) and *V. cholerae* (*Vch*) were amplified by polymerase chain reaction (PCR) from the corresponding genomic DNAs. The PCR products were then digested with the appropriate restriction enzymes and ligated into either pET11c or pET11d (Novagen) to yield the native expression vectors. The *Ype* MBP gene was amplified with primers PE-819 (5′-CCT CCC ATA TGA AAA TTG AAG AAG GTA AAC TGG TTA TC-3′) and PE-820 (5′-CAG CCT GGA TCC TTA GGC CTT CGT GAT ACG GGT TGC CGC ATC-3′), cut with *Nde*I and *Bam*HI, and inserted into pET11c to yield pJF1105. The *Tma* MBP gene was amplified with PE-668 (5′-CCT CCC ATA TGA AAA TTG AAC AAA CAA AGC TCA CCA TCT GGT CTT CCG AAA AGC AGG-3′) and PE-669 (5′-CAG CCT AGA TCT TAG GCC TTT TCT ATC TGT GCC TTG ATT TTG TCC AC-3′), cut with *Nde*I and *Bgl*II, and inserted into pET11c to yield pKM980. The *Tli* MBP gene was amplified with PE-666 (5′-CCT CCC ATA TGA AAA TTG AAG AAG GAA AGA TAG TAT TTG CTG TAG GAG G-3′) and PE-667 (5′-CAG CCT AGA TCT TAG GCC TTG CTG TAT TGT TTA ACT AAT TCC TCT G-3′), cut with *Nde*I and *Bgl*II, and inserted into pET11c to yield pKM979. The *Pfu* MBP gene was amplified with PE-472 (5′-CCT CCC ATA TGA AAA TCG AAG AAG GAA AAG TTG TTA TTT GGC ATG CAA TG-3′) and PE-473 (5′-CAG CCT GGA TCC ATT ATC CTT GCA TGT TGT TAA GGA TTT CTT G-3′), cut with *Nde*I and *Bam*HI, and inserted into pET11c to yield pKM820. The *Vch* MBP gene was amplified with PE-874 (5′-CCT GCT CAT GAA AAT TGA AGA AGG ACA ACT CAC TAT TTG G-3′) and PE-875 (5′-CAG CCT GGA TCC TTA CCC GGG TTT CGT CAT CTG CTT TTC AGC ATC-3′), cut with *Bsp*HI and *Bam*HI, and inserted into pET11d to yield pKM1136.

*Corresponding author. Fax: (1)-301-846 7148.
E-mail address: waughd@ncifcrf.gov (D.S. Waugh).

0014-5793/03/$22.00 © 2003 Published by Elsevier Science B.V. on behalf of the Federation of European Biochemical Societies.
doi:10.1016/S0014-5793(03)00070-X

2.2. Gateway destination vectors

The Gateway™ recombinational cloning system (Invitrogen) was used to facilitate the construction of fusion protein expression vectors. The *Eco* MBP destination vector, pKM596, was described previously [3]. The *Ype* MBP destination vector, pJF1106, was constructed by inserting the RfC Gateway cloning cassette into *Stu*I-digested pJF1105. The *Tma* MBP destination vector was constructed by inserting the RfC cassette into *Stu*I-digested pKM980. The *Tli* MBP destination vector was constructed by inserting the RfC cassette into *Stu*I-digested pKM979. The *Pfu* MBP destination vector was constructed by inserting the RfA cassette between the unique *Sac*I and *Bam*HI sites in a precursor of pKM820, after the sticky ends were blunted with T4 DNA polymerase and dNTPs. The *Vch* MBP destination vector was constructed by inserting the RfA cassette into *Sma*I-digested pKM1136.

2.3. Gateway entry clones

The genes encoding p16, GFP, CATΔ9, and E6 were described previously [1]. The cloned bovine rhodanese gene was obtained from Dr. Paul M. Horowitz (University of Texas Health Science Center at San Antonio, TX, USA). The luciferase gene originated from the plasmid pZA31luc [4]. The G3PDH and DHFR ORFs were obtained from the Invitrogen line of GeneStorm™ clones (catalog numbers M-M32599M and H-J00140M, respectively). To construct the Gateway entry clones, each passenger protein ORF was amplified by PCR, using a pair of gene-specific primers with 5′ extensions that added an in-frame TEV protease recognition site and a hexahistidine tag to their N- and C-termini, respectively. Next, these PCR amplicons were used as the templates for another PCR with primers PE-277 and PE-278 [5], which are designed to anneal to the sequences encoding the TEV protease recognition site and the His-tag, respectively, and add *att*B1 and *att*B2 recombination sites to the ends of the amplicon. The final PCR amplicons were inserted by recombinational cloning into the entry vector pDONR201 to create the entry clones pKM992 (E6), pKM617 (GFP), pKM1038 (CATΔ9), pKM991 (E6), pJF849 (rhodanese), pJF853 (luciferase), pJF929 (G3PDH), and pJF930 (DHFR). The nucleotide sequences of all eight ORFs were verified experimentally.

2.4. Fusion protein expression vectors

48 MBP fusion protein expression vectors were constructed by recombining each passenger protein ORF (p16, GFP, CATΔ9, E6, rhodanese, luciferase, G3PDH, and DHFR) into each MBP destination vector (*Eco*, *Ype*, *Tma*, *Tli*, *Vch*, *Pfu*), using the standard LxR protocol (Invitrogen). The GST fusion protein expression vectors were constructed in a similar fashion, using the destination vector pGST-DV3 (Invitrogen).

2.5. Protein expression, sodium dodecyl sulfate–polyacrylamide gel electrophoresis (SDS–PAGE) analysis, and densitometry

Protein expression experiments were performed as described previously [1] except that the tRNA accessory plasmid pRIL (Stratagene) was included in all cases. Preparation of samples, SDS–PAGE, and quantitative densitometry of the stained gels were all carried out essentially as described [1]. Between three and five experiments were performed to obtain the average solubility (and standard error) for each fusion protein.

3. Results

3.1. Overproduction of unfused MBPs in E. coli

MBPs and other periplasmic solute-binding proteins are present in a wide variety of microorganisms [6]. However, thus far only *E. coli* MBP has been exploited to facilitate the purification and enhance the solubility of recombinant proteins [1,3,7–9]. The objective of this study was to determine whether MBPs from other microorganisms can also function as solubility enhancers in the context of a fusion protein. To this end, five MBPs from diverse microbial sources were selected for comparative analysis.

The mature domain of every MBP is preceded by a hydrophobic N-terminal leader sequence that either serves as a secretion signal or a membrane anchor. In some cases (*Vch*, *Tma*, *Tli* and *Pfu*), it was uncertain exactly where the leader peptide ended and the mature domain began. For this reason, and also to improve the odds of achieving efficient translation initiation in *E. coli*, the N-termini of these proteins were modified to resemble that of the mature *Eco* MBP. Additionally, to allow for the insertion of a Gateway™ cloning cassette in the proper reading frame, one or two non-native residues were added to the C-termini of some MBPs to create unique restriction sites. The non-native residues are colored red in the sequence alignment (see web supplement).

To begin with, each MBP was overproduced in *E. coli* to assess its yield and solubility in the unfused state (Fig. 1). The yields of the *Ype*, *Vch*, *Tma*, *Tli* and *Pfu* MBPs were very similar to that of *Eco* MBP, easily comprising the majority of the intracellular protein. Moreover, although a statistical model [10] predicted that four of the six MBPs were likely to be insoluble in *E. coli*, we found that all of them were highly soluble at 37°C. *Tli* MBP does not bind to amylose resin [11], but all of the other MBPs were quantitatively retained on an amylose column, indicating that they were properly folded (data not shown).

3.2. Insolubility of GST fusion proteins in E. coli

Having established that all six MBPs are highly soluble in an unfused state, next we compared their ability to promote the solubility of eight different aggregation-prone passenger proteins: human $p16^{INK4}$, *Aquorea victoria* green fluorescent protein (GFP), chloramphenicol acetyltransferase-Δ9 (CATΔ9), human papillomavirus E6 oncoprotein, bovine rhodanese, *Photinus pyralis* luciferase, murine glyceraldehyde-3-phosphate dehydrogenase (G3PDH), and human dihydrofolate reductase (DHFR). These proteins represent a wide range of origins, sizes, and functions. Rhodanese, luciferase, G3PDH and DHFR are commonly used as model substrates for the molecular chaperone GroEL. The latter proteins are also more difficult to solubilize than most of the other passengers used in this study, and so they afford an opportunity for a more rigorous assessment of solubility enhancement.

As demonstrated previously, GST has virtually no ability to enhance the solubility of its fusion partners [1]. Consequently, the solubility of a GST fusion protein is a good indicator of the solubility of its passenger protein in the unfused state. At the same time, uniformly high expression levels are easier to

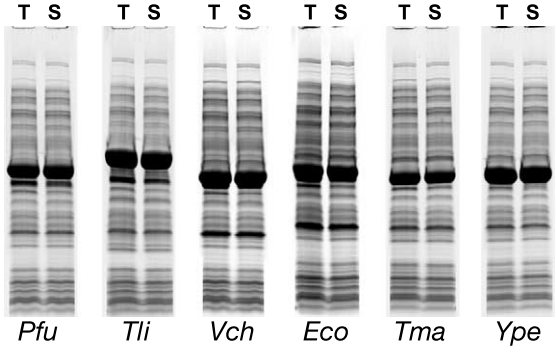

Fig. 1. Overproduction of unfused MBPs in *E. coli*. Samples of the total (T) and soluble (S) intracellular protein fractions are shown after SDS–PAGE. Abbreviations: *Pfu*, *P. furiosus*; *Tli*, *T. litoralis*; *Vch*, *V. cholerae*; *Tma*, *T. maritima*; *Eco*, *E. coli*; *Ype*, *Y. pestis*.

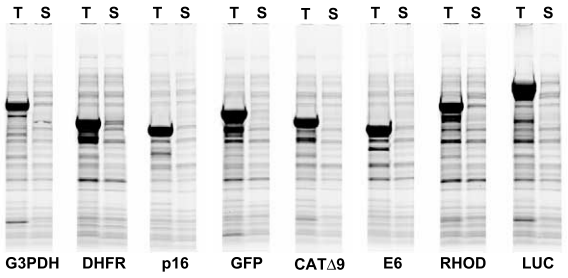

Fig. 2. Insolubility of GST fusion proteins in *E. coli*. Samples of the total (T) and soluble (S) intracellular proteins are shown after SDS–PAGE. Abbreviations: G3PDH, glyceraldehyde-3-phosphate dehydrogenase; DHFR, dihydrofolate reductase; p16, $p16^{INK4}$; GFP, green fluorescent protein; CATΔ9, a mutant of chloramphenicol acetyltransferase lacking the nine C-terminal residues; E6, human papillomavirus oncoprotein E6; RHOD, rhodanese; LUC, luciferase.

obtain with GST fusions than would be the case with unfused passengers. As shown in Fig. 2, all of the passenger proteins used in this study are poorly soluble as GST fusions.

3.3. *Solubility of MBP fusion proteins in* E. coli

Next, 48 different MBP fusion protein expression vectors, comprising all possible combinations of the six MBPs and the eight passenger proteins, were constructed by Gateway™ recombinational cloning. The length and amino acid sequences of the interdomain linkers were nearly identical in all of the fusion proteins. The MBP fusion proteins were expressed in *E. coli* and their solubility was estimated by SDS–PAGE and densitometry. All of the fusion proteins were expressed at a uniformly high level (data not shown). The quantitative results are summarized in Fig. 3.

All of the MBPs were more effective solubilizing agents than GST, but some were consistently better than others. The best solubilizing agent was *Pfu* MBP, the most distant relative of *Eco* MBP, whereas the closest relative of *Eco* MBP, *Ype* MBP, was the least effective overall. For the most part, those passenger proteins that were solubilized most readily by *Ype* MBP (GFP, p16 and E6) tended also to be solubilized most efficiently by the other MBPs, suggesting that the underlying mechanism of the solubilizing effect is likely to be similar for all six MBPs. The two passenger proteins that were consistently most difficult to solubilize, CATΔ9 and luciferase, exhibited a dramatic increase in solubility when they were fused to *Pfu* MBP. When the average solubility of each set of MBP fusion proteins (*Ype*, *Eco*, *Vch*, *Tma*, *Tli*, *Pfu*) is compared, the trend becomes even clearer (Fig. 4). From these data, it can be seen that on average *Pfu* MBP is about 50% more effective than *Eco* MBP and almost twice as effective as *Ype* MBP at promoting the solubility of the eight passenger proteins employed in this study. The average solubility of the corresponding GST fusion proteins is negligible by comparison.

4. Discussion

Although not every highly soluble protein can function as a solubility enhancer, our results indicate that this is a common property of MBPs from diverse microbial sources. In contrast to GST, all six of the MBPs we tested were able to enhance the solubility of aggregation-prone proteins to varying degrees. Unexpectedly, the closest relative of *Eco* MBP, *Ype* MBP, proved to be the least effective solubilizing agent. There are 53 amino acid substitutions in *Ype* MBP relative to *Eco* MBP, most of which are conservative in nature. Important clues about the mechanism of the solubilizing effect might be uncovered by attempting to determine which amino acid substitution(s) make the latter MBP a more effective solubilizing agent than the former.

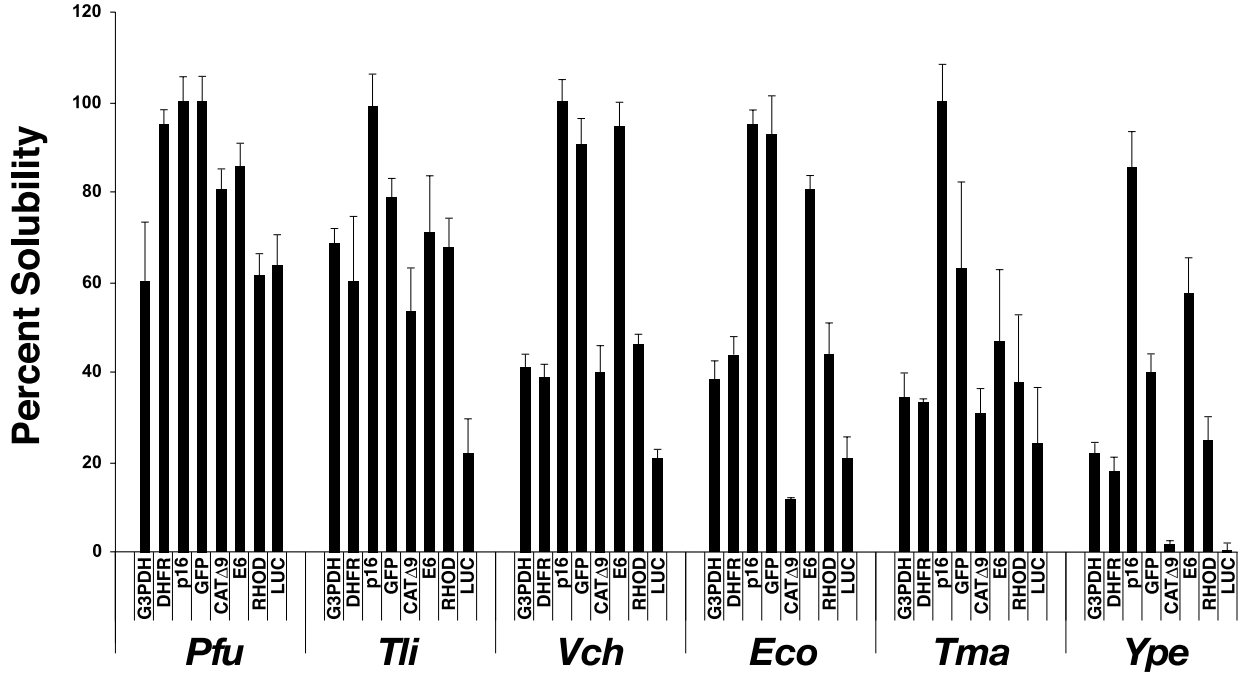

Fig. 3. Solubility of MBP fusion proteins in *E. coli*. Each fusion protein was expressed, analyzed by SDS–PAGE, and its solubility estimated by laser densitometry. Error bars indicate the standard error for the solubility of each fusion protein. Abbreviations are given in the legends to Figs. 1 and 2.

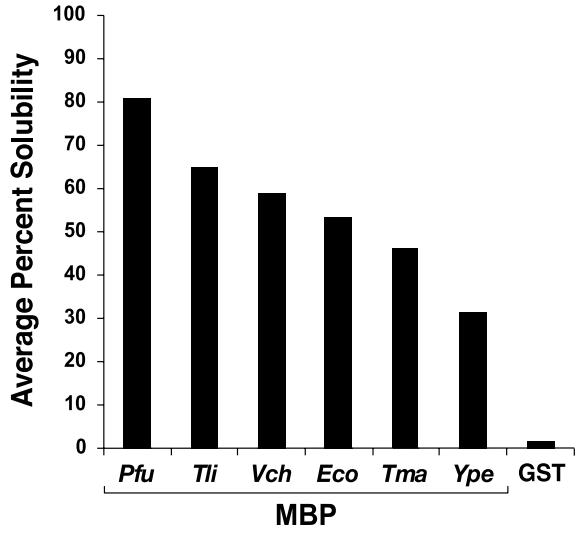

Fig. 4. Average solubility of each type of MBP fusion protein. The solubility of all eight fusions to each MBP was averaged. The average solubility of the corresponding set of GST fusion proteins is also shown for comparison. Abbreviations are given in the legend to Fig. 1.

It is conceivable that, among highly soluble proteins, the larger ones tend to be the most effective solubilizing agents. This might explain why MBP consistently outperformed both GST and thioredoxin in side-by-side comparisons [1]. Although a direct test of this hypothesis would be difficult, because any collection of highly soluble proteins of varying sizes would be heterogeneous with respect to other properties as well, the question can also be approached from a different angle. If, in addition to the quality of being highly soluble in *E. coli*, the ability of a protein to function as a solubility enhancer depended primarily on its size, then soluble proteins of similar size should be equally effective solubilizing agents. Yet, our results appear to contradict this prediction. All of the MBPs tested in this study are highly soluble in *E. coli* and approximately the same size, but they vary widely in their efficacy as solubilizing agents. Thus, if the size of the soluble fusion partner makes any difference, it seems unlikely to be the principal factor.

The two best solubilizing agents identified in this study, *Pfu* MBP and *Tli* MBP, are extremely thermostable proteins [12,13], as is *Tma* MBP [14]. However, the latter protein was a less effective solubilizing agent than *Vch* MBP or *Eco* MBP (Fig. 4), both of which are of mesophilic origin. The thermostability of *Vch* MBP has not been formally investigated, but *Eco* MBP is far less stable than *Tma* MBP [15,16]. Therefore, it appears that thermophilic proteins are not necessarily more effective solubilizing agents than their mesophilic counterparts.

What properties do all of these MBPs have in common that might explain their ability to promote the solubility of their fusion partners? For one thing, they are all very acidic proteins with theoretical isoelectric points ranging between 4.41 (*Pfu* MBP) and 5.39 (*Ype* MBP). *E. coli* NusA, another effective solubility enhancer [17], is also a very acidic protein (p$I \sim 4.35$). In contrast, the isoelectric points of GST (p$I \sim 6.52$) and chloramphenicol acetyltransferase (p$I \sim 6.36$), two highly soluble proteins that do not function as solubility enhancers [1,18], are much closer to neutral. The correlation between low isoelectric point and potency as a solubilizing agent is intriguing and may be significant, but it is not steadfast; the predicted isoelectric point of thioredoxin (~ 5.21) is the same as that of *Tli* MBP, but thioredoxin is a far less effective solubilizing agent than even *Eco* MBP [1]. Further research will be required to ascertain whether or not acidic proteins tend to be the most effective solubilizing agents.

MBPs belong to a family of periplasmic solute-binding proteins that interact with sugars and amino acids [6]. All of them are involved in solute uptake or chemotaxis. The crystal structures of *Eco*, *Tli* and *Pfu* MBP revealed that although their amino acid sequences are quite different, all three proteins adopt a similar tertiary fold [12,13,19]. It therefore seems likely that the other MBPs examined in this study also share the same general architecture. It is possible that, for whatever reason, this tertiary fold is particularly well-suited for solubility enhancement. If so, then even more distant relatives of MBP within the superfamily of periplasmic solute-binding proteins may also possess the ability to promote the solubility of their fusion partners. Moreover, if the ability of a protein to function as a solubility enhancer is correlated with its tertiary structure, then the phylogenetic comparative approach described here could also be used to identify orthologs of other solubility enhancing proteins, like *E. coli* NusA [17], with improved performance characteristics.

A potential practical advantage of the thermostable MBPs (*Tli*, *Tma*, *Pfu*) may be their utility as 'solubility handles' for refolding proteins. Not all passenger proteins that can be rendered soluble by fusing them to MBP are able to fold spontaneously into their native, biologically active conformation. Because the thermostable MBPs do not unfold in the presence of high concentrations of urea or guanidine hydrochloride [12–14], passenger proteins could be denatured and subsequently refolded while still fused to a folded MBP domain. This approach might conceivably result in a greater yield of properly folded protein than could be obtained by refolding the same protein in the unfused state.

In conclusion, although relatively few proteins appear to be generally effective solubilizing agents [1], this seems to be a common property of even distantly related maltodextrin-binding proteins. Some of these MBPs are clearly more effective solubilizing agents than *E. coli* MBP, but whether they will also be more effective at promoting the proper folding of their fusion partners remains to be determined. Although many polypeptides can be produced in a soluble form as MBP fusion proteins, they are frequently unable to fold into their native conformations but exist instead as soluble aggregates [20,21]. Therefore, solubility is not a reliable indicator of structural integrity and one must bear in mind that the most effective solubilizing agent may not necessarily be the most efficient 'foldase'. This issue clearly needs to be addressed in the future, not only for the MBPs described here, but also for other proteins that have been touted as solubilizing agents.

References

[1] Kapust, R.B. and Waugh, D.S. (1999) Protein Sci. 8, 1668–1674.
[2] Fox, J.D., Kapust, R.B. and Waugh, D.S. (2001) Protein Sci. 10, 622–630.
[3] Fox, J.D. and Waugh, D.S. (2003) Methods Mol. Biol. 205, 99–117.

[4] Lutz, R. and Bujard, H. (1997) Nucleic Acids Res. 25, 1203–1210.
[5] Evdokimov, A.G., Tropea, J.E., Routzahn, K.M. and Waugh, D.S. (2002) Acta Crystallogr. D58, 398–406.
[6] Tam, R. and Saier Jr., M.H. (1993) Microbiol. Rev. 57, 320–346.
[7] Riggs, P. (2000) Mol. Biotechnol. 15, 51–63.
[8] Shih, Y.-P., Kung, W.-M., Chen, J.-C., Yeh, C.-H., Wang, A.H.-J. and Wang, T.F. (2002) Protein Sci. 11, 1714–1719.
[9] Hammarstrom, M., Hellgren, N., van Den Berg, S., Berglund, H. and Hard, T. (2002) Protein Sci. 11, 313–321.
[10] Wilkinson, D.L. and Harrison, R.G. (1991) Biotechnology 9, 443–448.
[11] Horlacher, R., Xavier, K.B., Santos, H., DiRuggiero, J., Kossmann, M. and Boos, W. (1998) J. Bacteriol. 180, 680–689.
[12] Evdokimov, A.G., Anderson, D.E., Routzahn, K.M. and Waugh, D.S. (2001) J. Mol. Biol. 305, 891–904.
[13] Diez, J., Diederichs, K., Greller, G., Horlacher, R., Boos, W. and Welte, W. (2001) J. Mol. Biol. 305, 905–915.
[14] Wassenberg, D., Liebl, W. and Jaenicke, R. (2000) J. Mol. Biol. 295, 279–288.
[15] Novokhatny, V. and Ingham, K. (1997) Protein Sci. 6, 141–146.
[16] Ganesh, C., Shah, A.N., Swaminathan, C.P., Surolia, A. and Varadarajan, R. (1997) Biochemistry 36, 5020–5028.
[17] Davis, G.D., Elisee, C., Newham, D.M. and Harrison, R.G. (1999) Biotechnol. Bioeng. 65, 382–388.
[18] Maxwell, K.L., Mittermaier, A.K., Forman-Kay, J.D. and Davidson, A.R. (1999) Protein Sci. 8, 1908–1911.
[19] Spurlino, J.C., Lu, G.Y. and Quiocho, F.A. (1991) J. Biol. Chem. 266, 5202–5219.
[20] Sachdev, D. and Chirgwin, J.M. (1998) Biochem. Biophys. Res. Commun. 244, 933–937.
[21] Nominé, Y., Ristriani, T., Laurent, C., Lefèvre, J.-F., Weiss, É. and Travé, G. (2001) Prot. Expr. Purif. 23, 22–32.