Enhancement of soluble protein expression through the use of fusion tags

Dominic Esposito and Deb K Chatterjee

The soluble expression of heterologous proteins in *Escherichia coli* remains a serious bottleneck in protein production. Although alteration of expression conditions can sometimes solve the problem, the best available tools to date have been fusion tags that enhance the solubility of expressed proteins. However, a systematic analysis of the utility of these solubility fusions has been difficult, and it appears that many proteins react differently to the presence of different solubility tags. The advent of high-throughput structural genomics programs and advances in cloning and expression technology afford us a new way to compare the effectiveness of solubility tags. This data should allow us to better predict the effectiveness of tags currently in use, and might also provide the information needed to identify new fusion tags.

Addresses
Protein Expression Laboratory, Research Technology Program, SAIC-Frederick, Inc., NCI-Frederick, Frederick, MD 21702, USA

Corresponding author: Chatterjee, Deb K (chatterjee@ncifcrf.gov)

Current Opinion in Biotechnology 2006, **17**:353–358

This review comes from a themed issue on
Protein technologies
Edited by Deb K Chatterjee and Joshua LaBaer

Available online 15th June 2006

0958-1669/$ – see front matter

DOI 10.1016/j.copbio.2006.06.003

Introduction

Production of proteins, whether for biochemical analysis, therapeutics or structural studies, requires the success of three individual factors: expression, solubility and purification. Although protein expression is no longer considered a major limiting step and protein purification techniques have improved dramatically in the past decade, the problem of producing soluble proteins for purification has continued to be a major bottleneck in the field. As the number of high-throughput structural genomics projects increases, the reported percentages of soluble heterologous proteins expressed in *Escherichia coli* has continued to decrease; recent reports cite numbers ranging from 13% to 23% [1,2]. In addition, many of the most biochemically interesting families of proteins, including kinases, phosphatases, membrane-associated proteins and many other enzymes, are extremely difficult to produce as soluble proteins in *E. coli*. Although the reason why it is difficult to express soluble mammalian proteins is unknown, two factors that might contribute are the rate of translation and the rate of protein folding, which are almost an order of magnitude faster in *E. coli* as compared with eukaryotic systems [3]. Although eukaryotic expression hosts are sometimes able to overcome these problems, they are not without their own difficulties in terms of ease of use, time, cost and experimental flexibility.

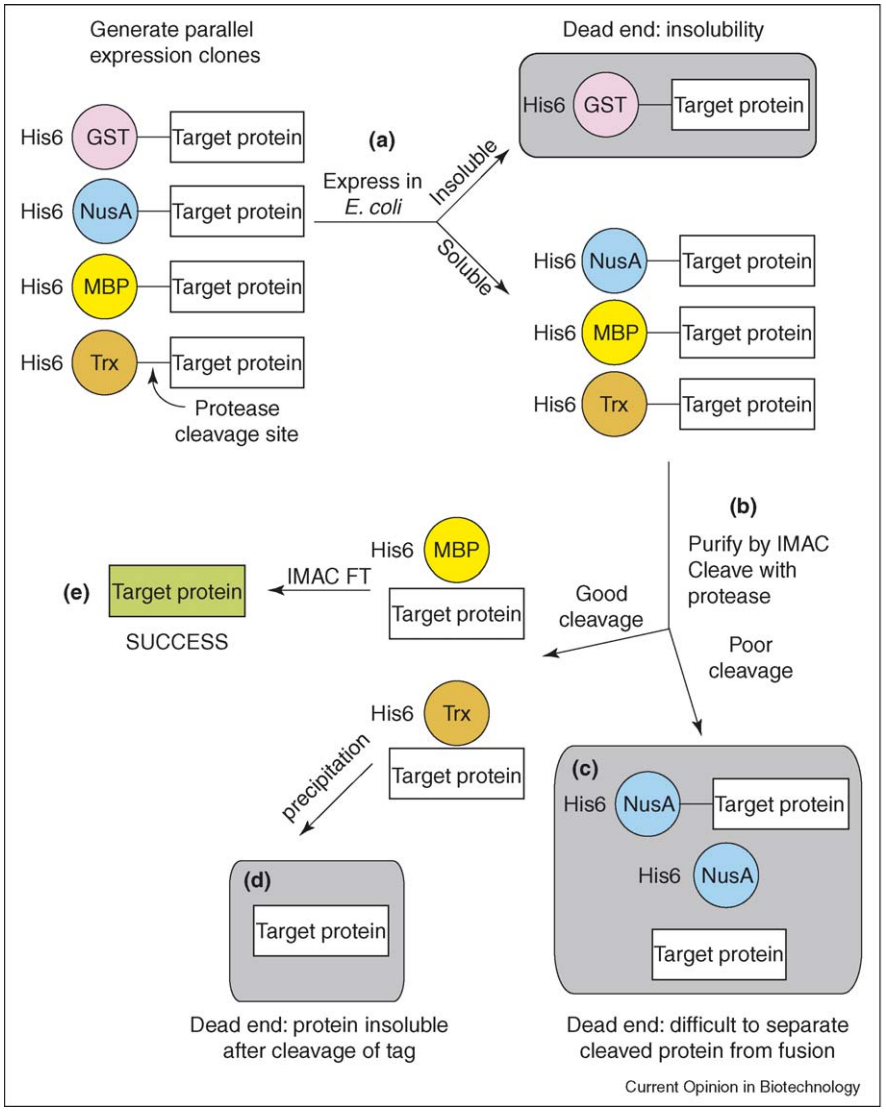
These problems in protein production have led to significant research into ways of enhancing the production of soluble proteins using currently available expression hosts. Although some efforts have been directed towards finding alternate expression conditions that can assist in making proteins soluble, the majority of the work in the field has focused on the discovery, development and refinement of solubility fusion tags. As the name implies, these tags are proteins or peptides that are fused to the protein of interest and, in the best case, help to properly fold their partners leading to enhanced solubility in the protein of interest (Figure 1). The concept of solubility fusions is not a new one, but advances in high-throughput cloning and expression methods have given us more power to test the function of these fusion tags and to attempt to make generalizations about their utility. Unfortunately, the addition of fusion tags brings with it a new set of problems, including issues surrounding the ultimate removal of these tags and the question as to whether the proteins made in this way retain their native structure and activity. In this review the benefits and some of the problems of using fusion tags for the soluble expression of proteins are discussed.

Expression hosts and conditions

Although a number of expression hosts are available for protein production, the standard in the field still remains *E. coli* [4,5]. Considerable effort is currently underway to make alternative hosts more accessible and affordable, and eukaryotic systems including mammalian, yeast and insect cell expression are becoming easier to use and less expensive [6–9]. Cell-free protein synthesis also has great potential for overcoming some of the problems of soluble protein expression, but remains a work in progress for the time being [10,11]. In the end, *E. coli* has significant benefits of cost, ease of use and scale, all of which make it essential to find ways to overcome the difficulty of generating soluble heterologous proteins in *E. coli*.

Improving the solubility of recombinant proteins in *E. coli* commonly involves changing some of the expression conditions. Factors such as reduced temperature [12], changes in the *E. coli* expression strain [13], different promoters or

Figure 1



Schematic representation of the pathway from protein expression to purification using solubility tags. Four arbitrary tagged versions of the protein of interest are generated in *E. coli*. **(a)** After expression, some fusions will remain in the insoluble fraction and be lost from the pathway. **(b)** Soluble fusions are purified by IMAC (immobilized metal affinity chromatography) using the attached His6 tag. A protease is then used to cleave the fusion tag from the partner protein. **(c)** Some fusions will not cleave efficiently, and will leave behind a mixture of cleaved and uncleaved proteins that cannot be easily separated. **(d)** Other fusions will cleave efficiently, but when separated from the solubility tag the partner protein will become insoluble and precipitate. **(e)** However, a well-behaved fusion will remain in solution and can be purified by a second IMAC step to remove the His6-tagged solubility tag and protease, leaving only the target protein in the flow through (IMAC FT).

induction conditions [14], and co-expression of molecular chaperones and folding modulators [15] have all been examined and in some specific cases lead to enhancements of soluble protein production. In particular, lower expression temperatures routinely improve the solubility of *E. coli*-expressed proteins; however, the improvements can be minimal and many proteins will remain insoluble or behave poorly even under these conditions [16••]. In many cases, none of these factors will solve the problem and proteins will be expressed in insoluble inclusion bodies when overproduced in *E. coli* [17].

Solubility-enhancing fusion tags

It was discovered many years ago that some affinity tags were able to enhance the solubility of some of the partner proteins to which they were attached [18,19]. None of these tags worked universally with every partner protein, however, and the hope for a 'magic bullet' to solve the solubility problem continued. Although this entity has yet to be discovered, the repertoire of fusion partners has increased steadily over the years. Coupled with advances in high-throughput cloning and expression, these fusion partners have increased our ability to find conditions for

Table 1

Some commonly used solubility-enhancing fusion partners

Tag	Protein	Source organism	References
MBP	Maltose-binding protein	*Escherichia coli*	[39•,40]
GST	Glutathione-*S*-transferase	*Schistosoma japonicum*	[19]
Trx	Thioredoxin	*Escherichia coli*	[41]
NusA	N-Utilization substance	*Escherichia coli*	[42]
SUMO	Small ubiquitin-modifier	*Homo sapiens*	[2]
SET	Solubility-enhancing tag	Synthetic	[29]
DsbC	Disulfide bond C	*Escherichia coli*	[43]
Skp	Seventeen kilodalton protein	*Escherichia coli*	[28]
T7PK	Phage T7 protein kinase	Bacteriophage T7	[28]
GB1	Protein G B1 domain	Streptococcus sp.	[30]
ZZ	Protein A IgG ZZ repeat domain	*Staphylococcus aureus*	[32]

soluble protein expression. Today, there are a number of common solubility-enhancing fusion tags that are used to express proteins in *E. coli* (Table 1). In some cases, these tags double as affinity tags, not only facilitating soluble expression but also increasing the efficiency of protein purification. In other cases, these solubility tags have been combined with a simple hexahistidine (His6) tag, allowing the fusion partner to maintain its solublizing functionality and also double as an affinity tag. Additional affinity tags that can be combined with many of these solubility-enhancing tags are also available and have been successfully used to produce purified proteins (Table 2).

Recombinational cloning has led to a major advancement in the field of soluble protein expression. In the past, cloning a gene into multiple vectors with different tags was a laborious process involving repeated PCR amplification, ligation and sequencing. Modern cloning technology has made parallel cloning into multiple vectors with different tags a routine matter, with the ability to make and compare 8 or 16 different solubility-tagged protein expression vectors in a single experiment [16••,20••]. Coupled with the ease of generating new expression vectors using these systems [20••,21••], we now have the tools to study solubility tags on a side-by-side basis and in a high-throughput manner, alleviating the problem of focusing on a single protein or small family of proteins and allowing us to draw global conclusions about the utility of given tags.

Although, as Table 1 shows, there are numerous solubility tags reported in the literature, the majority of recent work has continued to focus on a few major players, notably maltose-binding protein (MBP), N-utilization substance A (NusA), thioredoxin (Trx), and glutathione-*S*-transferase (GST). Both MBP and GST have an additional benefit in that they can function as affinity tags; MBP binds strongly to amylose resin [22], whereas GST binds to glutathione resin [23]. In *E. coli*, however, the overwhelming evidence has shown that GST is, at best, a poor solubility enhancer [12,20••]. This has left MBP as one of the most well-studied solubility factors, and a significant body of evidence exists to show that N-terminal MBP fusions can frequently produce soluble proteins when the unfused partners are insoluble [16••,20••,21••,24••]. The *E. coli* NusA protein has also been shown to function at a similar level to MBP in producing soluble partner proteins, although NusA has no independent affinity functionality [25•,26,27•]. *E. coli* thioredoxin has been reported in several studies to be nearly as efficient as MBP at promoting solubility [12,25•], although other studies have shown it to be less effective [2].

Recently, new solubility tags have begun to appear in the literature, as scientists try to identify highly soluble proteins that might be able to promote solubility in their fusion partners. A fragment of the bacteriophage T7 protein kinase gene (T7PK) has recently been shown to function as a solubility enhancer, and also appears to

Table 2

Some commonly used affinity purification tags

Tag	Protein	Affinity matrix	References
His6	Hexahistidine tag	Metal chelates	[44]
GST	Glutathione *S*-transferase	Glutathione	[23]
MBP	Maltose-binding protein	Amylose	[22]
FLAG	FLAG tag peptide	Anti-FLAG antibody	[45]
BAP	Biotin acceptor peptide	Avidin	[46]
Strep-II	Streptavidin-binding peptide	Streptavidin	[47]
CBP	Calmodulin-binding peptide	Calmodulin	[48]

enhance overall levels of expression [28]. Small peptide tags called SET tags, which feature highly acidic amino acid sequences, have also been shown to stimulate solubility in a few partner proteins [29]. If the SET tags are shown to have a more universal effect, they could be highly valued as their small size (<30 amino acids) might lead to less folding interference and make them more amenable for structural studies without the need to remove the tag. Similarly, two other small protein tags, GB1 and ZZ, have been used with some success to enhance the expression and solubility of peptides and small proteins [30–32]. Another potentially promising recent development is the SUMO tag, a ubiquitin-related protein that has been reported to enhance solubility and in some cases appears to be as effective as MBP [2].

Currently, many of these tags suffer from the same problem — they do not function equally well with all partner proteins. Although many studies have demonstrated that a particular tag is better at solublizing a given partner protein or small set of proteins, the large-scale analysis of hundreds of proteins with tens of tags remains to be accomplished. Recent studies have begun to look at larger numbers of target proteins, and the types of target proteins being analyzed are becoming more diverse as well [20••,25•]. But, the contradictory data observed for many of these tags simply highlights both the difficulty in sample size and the true complexity of proteins — every protein is different, and its reaction to various solubility tags could very well also be different. We must look to high-throughput analyses and proteomics technologies for the larger-scale comparisons that are necessary to validate new solubility tags. One such technology is POET, a high-throughput proteomics approach that should be useful for comparing the solubility-enhancing power of multiple tags on large pools of proteins [33]. Data from high-throughput structural genomics centers could also help clarify which tags have a more generic function across a wider variety of targets from different hosts or different classes of proteins [1,16••,20••].

The removal of fusion tags

The ability of solubility fusion tags to produce soluble protein is only the first step in the pathway towards protein production. As many of these tags are large proteins, it is often necessary to remove the tags after they have been used to make soluble fusion partners. Although some groups have shown the possibility of generating crystal structures of the fusion proteins [34], biochemical studies and therapeutic proteins require the removal of the tags and the maintenance of a stable, soluble protein of interest. In order to achieve this, the most common solution is to engineer a protease cleavage site between the solubility tag and the partner protein, permitting an *in vitro* reaction after purification to remove the fusion tag. By placing an affinity tag at the N terminus of the fusion partner, one can purify the protein, cleave the tag, and then re-purify on the same affinity matrix to remove the cleaved tag [35•]. This process is becoming a very common and successful approach to making highly purified proteins. There are numerous choices for the protease used for this purpose (Table 3), but one of the most commonly used is the TEV protease from tobacco etch virus, which shows exquisitely high specificity, is relatively easy to make in large quantities, and cleaves in most cases to leave a native N terminus [36,37].

Unfortunately, difficulties are often encountered when attempts are made to remove the solubility tags from apparently soluble fusion proteins. In some cases, the protease cleavage fails, owing perhaps to steric issues or aggregation. Sometimes, the problem can be solved by introducing short linkers between the protease site and the passenger protein (D Esposito, unpublished). In other cases, the protease cleavage is successful, but the passenger protein does not remain soluble once the fusion partner is removed. The most widely viewed hypothesis for this result is that the seemingly soluble proteins are actually existing as 'soluble aggregates', held in solution by interactions with the solubility partner, but not in their true, native, soluble form [35•,38]. Removal of the solubility partner causes them to revert to their natural insolubility and they precipitate. Solving this problem is a key focus at present, as a significant number of proteins exhibit this behavior independently of the type of solubility tag being employed [39•]. Possible solutions to this problem that are being investigated include the use of additives, such as detergents or chaperones, or

Table 3

Some proteases commonly used to remove fusion tags

Protease	Source	Cleavage site[a,b]	References
TEV	Tobacco etch virus protease	ENLYFQ/X	[37]
3C	Human rhinovirus 3C protein	EVLFQ/GP	[44]
Xa	Factor Xa	IEGR/	[49]
EntK	Enterokinase	DDDDK/	[50]
Thr	Thrombin	LVPR/GS	[49]
Caspase	Caspase-3	DXXD/	[51]

[a] The '/' indicates the site of protease cleavage.
[b] The cleavage site is given in single-letter amino acid code, where X indicates any amino acid.

alterations in expression conditions. So far, however, no one method has proven widely successful.

Conclusions

Currently, the production of soluble proteins in *E. coli* remains a hit-or-miss affair. Although there are clearly some solubility tags that seem to perform, on average, better than others, there is still no *a priori* guarantee that a given tag will work with a protein of interest. It is for this reason that an emphasis on large-scale comparisons of protein solubility with different tags is underway. With the increase in high-throughput cloning and expression projects, it is likely that the next few years will give us a better feel for the true winners among the current set of tags. At the same time, there is a likelihood that many other proteins might exist that can carry out similar functions, and the expectation is that additional solubility tags will be discovered and compared with the current set. Overall, it remains likely that the individuality of proteins will force scientists to keep a toolbox of solubility tags to hand, any one of which might prove the best tool for a given task. The advent of recombinational cloning and high-throughput expression techniques has made this a much easier task, and data produced over the next few years should give us a better idea as to what tools to use in what circumstances. Barring the discovery of the 'magic bullet' tag, this is likely to be the best path forward for soluble protein production in *E. coli*.

Acknowledgements

This project has been funded in whole or in part with federal funds from the National Cancer Institute, National Institutes of Health, under contract N01-CO-12400. The content of this publication does not necessarily reflect the views or policies of the Department of Health and Human Services, nor does mention of trade names, commercial products or organizations imply endorsement by the US Government.

References and recommended reading

Papers of particular interest, published within the annual period of review, have been highlighted as:

- • of special interest
- •• of outstanding interest

1. Chambers SP, Austen DA, Fulghum JR, Kim WM: **High-throughput screening for soluble recombinant expressed kinases in *Escherichia coli* and insect cells**. *Protein Expr Purif* 2004, **36**:40-47.

2. Marblestone JG, Edavettal SC, Lim Y, Lim P, Zuo X, Butt TR: **Comparison of SUMO fusion technology with traditional gene fusion systems: enhanced expression and solubility with SUMO**. *Protein Sci* 2006, **15**:182-189.

3. Widmann M, Christen P: **Comparison of folding rates of homologous prokaryotic and eukaryotic proteins**. *J Biol Chem* 2000, **275**:18619-18622.

4. Baneyx F: **Recombinant protein expression in *Escherichia coli***. *Curr Opin Biotechnol* 1999, **10**:411-421.

5. Goulding CW, Perry LJ: **Protein production in *Escherichia coli* for structural studies by X-ray crystallography**. *J Struct Biol* 2003, **142**:133-143.

6. Hu YC: **Baculovirus as a highly efficient expression vector in insect and mammalian cells**. *Acta Pharmacol Sin* 2005, **26**:405-416.

7. Kost TA, Condreay JP, Jarvis DL: **Baculovirus as versatile vectors for protein expression in insect and mammalian cells**. *Nat Biotechnol* 2005, **23**:567-575.

8. Daly R, Hearn MT: **Expression of heterologous proteins in *Pichia pastoris*: a useful experimental tool in protein engineering and production**. *J Mol Recognit* 2005, **18**:119-138.

9. Greene JJ: **Host cell compatibility in protein expression**. *Methods Mol Biol* 2004, **267**:3-14.

10. Murthy TV, Wu W, Qiu QQ, Shi Z, LaBaer J, Brizuela L: **Bacterial cell-free system for high-throughput protein expression and a comparative analysis of *Escherichia coli* cell-free and whole cell expression systems**. *Protein Expr Purif* 2004, **36**:217-225.

11. Sitaraman K, Esposito D, Klarmann G, Le Grice SF, Hartley JL, Chatterjee DK: **A novel cell-free protein synthesis system**. *J Biotechnol* 2004, **110**:257-263.

12. Hammarstrom M, Hellgren N, van Den Berg S, Berglund H, Hard T: **Rapid screening for improved solubility of small human proteins produced as fusion proteins in *Escherichia coli***. *Protein Sci* 2002, **11**:313-321.

13. Miroux B, Walker JE: **Over-production of proteins in *Escherichia coli*: mutant hosts that allow synthesis of some membrane proteins and globular proteins at high levels**. *J Mol Biol* 1996, **260**:289-298.

14. Qing G, Ma LC, Khorchid A, Swapna GV, Mal TK, Takayama MM, Xia B, Phadtare S, Ke H, Acton T *et al.*: **Cold-shock induced high-yield protein production in *Escherichia coli***. *Nat Biotechnol* 2004, **22**:877-882.

15. de Marco A, De Marco V: **Bacteria co-transformed with recombinant proteins and chaperones cloned in independent plasmids are suitable for expression tuning**. *J Biotechnol* 2004, **109**:45-52.

16. •• Kataeva I, Chang J, Xu H, Luan CH, Zhou J, Uversky VN, Lin D, Horanyi P, Liu ZJ, Ljungdahl LG *et al.*: **Improving solubility of *Shewanella oneidensis* MR-1 and *Clostridium thermocellum* JW-20 proteins expressed into *Escherichia coli***. *J Proteome Res* 2005, **4**:1942-1951.
Results from a high-throughput structural genomics project comparing MBP, GST, NusA and His6 solubility tags. The study of 152 proteins from mesophilic and thermophilic bacteria suggested that MBP was the best tag for making soluble fusions, and also showed the positive effect of lowered expression temperature on solubility. The authors discuss algorithms for predicting the solubility of proteins.

17. Chayen NE: **Turning protein crystallisation from an art into a science**. *Curr Opin Struct Biol* 2004, **14**:577-583.

18. Kapust RB, Waugh DS: ***Escherichia coli* maltose-binding protein is uncommonly effective at promoting the solubility of polypeptides to which it is fused**. *Protein Sci* 1999, **8**:1668-1674.

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

20. •• Dyson MR, Shadbolt SP, Vincent KJ, Perera RL, McCafferty J: **Production of soluble mammalian proteins in *Escherichia coli*: identification of protein features that correlate with successful expression**. *BMC Biotechnol* 2004, **4**:32.
The authors chose 30 diverse human genes and compared expression with six different N-terminal and eight different C-terminal solubility fusions. Results showed that MBP and Trx were the best tags and that MBP also worked as a C-terminal fusion, contrary to previous results. The authors nicely highlight the protein features that seem to correlate with solubility: molecular weight, number of hydrophobic residues, and low complexity regions.

21. •• Busso D, Delagoutte-Busso B, Moras D: **Construction of a set gateway-based destination vectors for high-throughput cloning and expression screening in *Escherichia coli***. *Anal Biochem* 2005, **343**:313-321.
This article describes the construction of 10 recombinational cloning vectors and their use for comparing expression and solubility of a small number of genes from *Bacillus subtilis*. Methods for generating and testing new vectors are clearly discussed and the results of the solubility analysis suggest that MBP and NusA are much better than GST or Trx in the case of these partner proteins.

22. di Guan C, Li P, Riggs PD, Inouye H: **Vectors that facilitate the expression and purification of foreign peptides in *Escherichia coli* by fusion to maltose-binding protein**. *Gene* 1988, **67**:21-30.

23. Smith DB, Johnson KS: **Single-step purification of polypeptides expressed in *Escherichia coli* as fusions with glutathione *S*-transferase**. *Gene* 1988, **67**:31-40.

24. •• Braud S, Moutiez M, Belin P, Abello N, Drevet P, Zinn-Justin S, Courcon M, Masson C, Dassa J, Charbonnier JB *et al.*: **Dual expression system suitable for high-throughput fluorescence-based screening and production of soluble proteins**. *J Proteome Res* 2005, **4**:2137-2147.
The authors describe the construction of vectors not only for generating soluble proteins, but also for screening for soluble proteins using fluorescence in a high-throughput manner. Results of comparisons of seven tags are presented for a small set of proteins, highlighting the fact that different proteins can give very different solubility patterns. Methods for expression and solubility analysis in high-throughput format are explained in great detail.

25. • Dummler A, Lawrence AM, de Marco A: **Simplified screening for the detection of soluble fusion constructs expressed in *E. coli* using a modular set of vectors**. *Microb Cell Fact* 2005, **4**:34.
A series of expression vectors were generated to compare solubility tags using fluorescent proteins as the partner proteins. Findings showed that NusA and Trx were both good promoters of solubility, but that DsbC worked very poorly in these cases.

26. Schrodel A, Volz J, de Marco A: **Fusion tags and chaperone co-expression modulate both the solubility and the inclusion body features of the recombinant CLIPB14 serine protease**. *J Biotechnol* 2005, **120**:2-10.

27. • Turner P, Holst O, Karlsson EN: **Optimized expression of soluble cyclomaltodextrinase of thermophilic origin in *Escherichia coli* by using a soluble fusion-tag and by tuning of inducer concentration**. *Protein Expr Purif* 2005, **39**:54-60.
Although examining only one particular protein of interest, this paper is one of the rare examples where alterations of expression conditions show a strong effect on solubility. Using a special strain of *E. coli* called Tuner, the authors were able to modulate expression using lower concentrations of inducer and observed an increase in protein solubility. NusA fusions were also shown to be successful in enhancing soluble protein production.

28. Chatterjee DK, Esposito D: **Enhanced soluble protein expression using two new fusion tags**. *Protein Expr Purif* 2006, **46**:122-129.

29. Zhang YB, Howitt J, McCorkle S, Lawrence P, Springer K, Freimuth P: **Protein aggregation during overexpression limited by peptide extensions with large net negative charge**. *Protein Expr Purif* 2004, **36**:207-216.

30. Bao WJ, Gao YG, Chang YG, Zhang TY, Lin XJ, Yan XZ, Hu HY: **Highly efficient expression and purification system of small-size protein domains in *Escherichia coli* for biochemical characterization**. *Protein Expr Purif* 2006, **47**:599-606.

31. Cheng Y, Patel DJ: **An efficient system for small protein expression and refolding**. *Biochem Biophys Res Commun* 2004, **317**:401-405.

32. Zhao Y, Benita Y, Lok M, Kuipers B, van der Ley P, Jiskoot W, Hennink WE, Crommelin DJ, Oosting RS: **Multi-antigen immunization using IgG binding domain ZZ as carrier**. *Vaccine* 2005, **23**:5082-5090.

33. Gillette WK, Esposito D, Frank PH, Zhou M, Yu LR, Jozwik C, Zhang X, McGowan B, Jacobowitz DM, Pollard HB *et al.*: **Pooled ORF expression technology (POET): using proteomics to screen pools of open reading frames for protein expression**. *Mol Cell Proteomics* 2005, **4**:1647-1652.

34. Smyth DR, Mrozkiewicz MK, McGrath WJ, Listwan P, Kobe B: **Crystal structures of fusion proteins with large-affinity tags**. *Protein Sci* 2003, **12**:1313-1322.

35. • Waugh DS: **Making the most of affinity tags**. *Trends Biotechnol* 2005, **23**:316-320.
A recent review of affinity purification and the problems associated with it, focusing in large part on the issue of making soluble proteins and the ability to cleave those proteins from the fusion tags. Also contains an analysis of the advantages and disadvantages of the various affinity tags and their combinations with solubility tags.

36. Kapust RB, Tozser J, Copeland TD, Waugh DS: **The P1′ specificity of tobacco etch virus protease**. *Biochem Biophys Res Commun* 2002, **294**:949-955.

37. Kapust RB, Tozser J, Fox JD, Anderson DE, Cherry S, Copeland TD, Waugh DS: **Tobacco etch virus protease: mechanism of autolysis and rational design of stable mutants with wild-type catalytic proficiency**. *Protein Eng* 2001, **14**:993-1000.

38. Nomine Y, Ristriani T, Laurent C, Lefevre JF, Weiss E, Trave G: **Formation of soluble inclusion bodies by HPV e6 oncoprotein fused to maltose-binding protein**. *Protein Expr Purif* 2001, **23**:22-32.

39. • Nallamsetty S, Waugh DS: **Solubility-enhancing proteins MBP and NusA play a passive role in the folding of their fusion partners**. *Protein Expr Purif* 2006, **45**:175-182.
Work focusing on MBP and NusA demonstrated that many partner proteins react similarly no matter which tag is attached to them. This observation suggests that the formation of soluble aggregates could be an intrinsic feature of some proteins, and perhaps means that some proteins will never respond to the solubility tag approach. Discussions are presented about the possible mechanisms of solubility enhancement by fusion tags.

40. Fox JD, Waugh DS: **Maltose-binding protein as a solubility enhancer**. *Methods Mol Biol* 2003, **205**:99-117.

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

42. Davis GD, Elisee C, Newham DM, Harrison RG: **New fusion protein systems designed to give soluble expression in *Escherichia coli***. *Biotechnol Bioeng* 1999, **65**:382-388.

43. Zhang Z, Li ZH, Wang F, Fang M, Yin CC, Zhou ZY, Lin Q, Huang HL: **Overexpression of DsbC and DsbG markedly improves soluble and functional expression of single-chain Fv antibodies in *Escherichia coli***. *Protein Expr Purif* 2002, **26**:218-228.

44. Gaberc-Porekar V, Menart V: **Perspectives of immobilized-metal affinity chromatography**. *J Biochem Biophys Methods* 2001, **49**:335-360.

45. Einhauer A, Jungbauer A: **The FLAG peptide, a versatile fusion tag for the purification of recombinant proteins**. *J Biochem Biophys Methods* 2001, **49**:455-465.

46. Schatz PJ: **Use of peptide libraries to map the substrate specificity of a peptide-modifying enzyme: a 13 residue consensus peptide specifies biotinylation in *Escherichia coli***. *Biotechnology (NY)* 1993, **11**:1138-1143.

47. Voss S, Skerra A: **Mutagenesis of a flexible loop in streptavidin leads to higher affinity for the strep-tag II peptide and improved performance in recombinant protein purification**. *Protein Eng* 1997, **10**:975-982.

48. Vaillancourt P, Zheng CF, Hoang DQ, Breister L: **Affinity purification of recombinant proteins fused to calmodulin or to calmodulin-binding peptides**. *Methods Enzymol* 2000, **326**:340-362.

49. Jenny RJ, Mann KG, Lundblad RL: **A critical review of the methods for cleavage of fusion proteins with thrombin and factor Xa**. *Protein Expr Purif* 2003, **31**:1-11.

50. Kwon SY, Choi YJ, Kang TH, Lee KH, Cha SS, Kim GH, Lee HS, Kim KT, Kim KJ: **Highly efficient protein expression and purification using bacterial hemoglobin fusion vector**. *Plasmid* 2005, **53**:274-282.

51. Feeney B, Soderblom EJ, Goshe MB, Clark AC: **Novel protein purification system utilizing an N-terminal fusion protein and a caspase-3 cleavable linker**. *Protein Expr Purif* 2005, **47**:311-318.