CHAPTER SIXTEEN

Tagging for Protein Expression

Arun Malhotra

Contents

Abstract

Tags are frequently used in the expression of recombinant proteins to improve solubility and for affinity purification. A large number of tags have been developed for protein production and researchers face a profusion of choices when designing expression constructs. Here, we survey common affinity and solubility tags, and offer some guidance on their selection and use.

Department of Biochemistry and Molecular Biology, University of Miami Miller School of Medicine, Miami, Florida, USA

Methods in Enzymology, Volume 463
ISSN 0076-6879, DOI: 10.1016/S0076-6879(09)63016-0

1. Introduction

Most proteins are made using recombinant techniques in expression systems. Additional residues or tags can be engineered on either the N- or C-terminal end of the protein of interest during the cloning step. These tags, which can range in size from just a few residues to full-length proteins or domains, can be used to improve protein production or to confer new properties that can be used for characterization and study of the target protein. The term ''fusion protein'' is also often used instead of the term ''tag,'' and fusion sometimes refers to the simpler end-to-end joining of two proteins while tags are typically shorter and include linker regions; here, we will use these terms interchangeably.

The focus of this chapter is to survey some of the common tags that are used for improving the production of proteins, and highlight the advantages and pitfalls of their use. Given the wild profusion of tags, often in different combinations and with different cleavage sites, this review is by no means exhaustive and the reader is encouraged to look at constructs available from many commercial sources as well as repository sources such as Addgene (http://www.addgene.org) or ATCC (http://www.atcc.org). Detailed protocols related to the use of many individual tags have been described previously in *Methods in Enzymology* (e.g., Volumes 326 and 327), and in specialized journals in this field such as *Protein Expression and Purification* and the newer *Microbial Cell Factories*. Handbooks from suppliers of vectors for most expression systems (Amersham/GE Healthcare Inc., Clontech Inc., Invitrogen Inc., New England Biolabs Inc., Novagen/EMD Biosystems Inc., Roche Inc., Sigma-Aldrich Inc., and others) are another rich source of protocols.

Tags used to improve the production of recombinant proteins can be roughly divided into purification and solubility tags. The former are used along with affinity binding to allow rapid and efficient purification of proteins, while the latter refer to tags that enhance the proper folding and solubility of a protein. Tables 16.1 and 16.2 list some of the common purification and solubility tags that are used for protein expression. Table 16.3 summarizes common endoproteases used to remove tags and recover the target protein of interest.

While such tags are quite useful, other tags can be fused to proteins for a broad range of applications—labeling for imaging and localization studies, protein detection and quantification, protein–protein interaction studies, subcellular localization or transduction, and many others. It is important to keep in mind some of these additional capabilities that can be engineered into a protein as recombinant constructs are being designed, since multiple tags can be added together in different combinations (Fig. 16.1).

Table 16.1 Common affinity tags

Tag	Size	Affinity matrix
His-tag	6–10 His	Immobilized metal ions—Ni, Co, Cu, Zn
GST (glutathione-*S*-transferase)	211 aa	Glutathione resin
FLAG-tag	8 aa (DYKDDDDK) (22 aa for 3xFLAG)	Anti-FLAG mAB
Strep-II-tag	8 aa (WSHPQFEK)	Strep-Tactin (modified streptavidin)
Protein A (staphylococcal Protein A)	280 aa	Immobilized IgG
MBP (maltose-binding protein)	396 aa	Cross-linked amylose
CBP (calmodulin-binding protein)	26 aa	Immobilized calmodulin
CBD (chitin-binding domain)	51 aa	Chitin
HaloTag	~300 aa	Chloroalkane

Table 16.2 Common solubility tags

Tag	Size	Protein
MBP	396 aa	Maltose-binding protein
NusA	495 aa	N-utilization substance
Trx	109 aa	Thioredoxin
SUMO	~100 aa	Small ubiquitin modifier
GB1	56 aa	IgG domain B1 of streptococcus Protein G
SET/SEP	<20 aa	Hydrophilic solubility enhancing peptide sequences
HaloTag	~300 aa	Mutated dehalogenase

2. Some Considerations When Designing a Tagged Protein

2.1. Affinity and/or solubility?

Two challenges in the production of heterologous proteins in *Escherichia coli*, the workhorse of protein expression systems, are poor or low expression, and the misfolding of the expressed protein into insoluble aggregates called

Table 16.3 Common endoproteases used to remove protein tags

Protease	Cleavage site
Enterokinase	DDDDK↑
Factor Xa	IEGR↑
Thrombin	LVPR↑GS
PreScission (human rhinovirus 3C protease)	LEVLFQ↑GP
TEV (tobacco etch virus protease)	ENLYFQ↑G
TVMV (tobacco vein mottling virus protease)	ETVRFQG↑S
SUMO protease (catalytic core of Ulp1)	Recognizes SUMO tertiary structure and cleaves at the C-terminal end of the conserved Gly–Gly sequence in SUMO

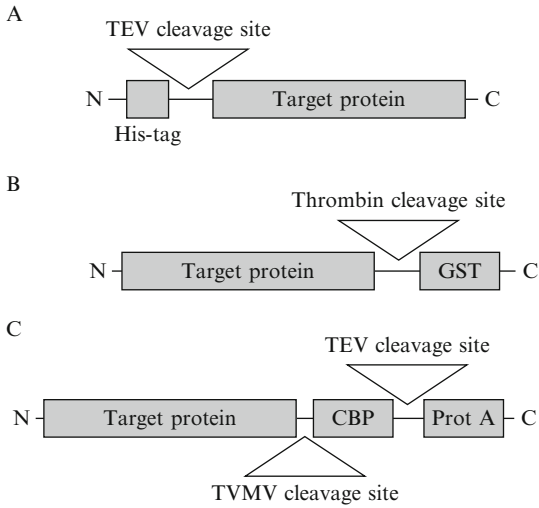

Figure 16.1 Example schematics of single and tandem tagged proteins. (A) N-terminally His-tagged protein; (B) C-terminally GST tagged protein; and (C) target protein with a removable (using TVMV proteases) C-terminal tandem affinity tag (TAP tag).

inclusion bodies. Expression problems caused by weak promoters, poor translation initiation, or the presence of rare codons, can be alleviated by introducing corrective sequences in the gene when assembling overexpression plasmids and the use of *E. coli* strains supplemented with rare tRNAs. Attaching a highly translated native gene as a fusion on the N-terminal end of the heterologous target protein is another approach to improving yield,

and has also the added benefit of increasing the solubility of the target protein; this is the basis of solubility tags (Table 16.2).

Affinity tags, on the other hand, are crucial during protein purification, and allow the use of a variety of strategies to bind the target protein on an affinity matrix (Table 16.1). Some protein tags can function in both affinity and solubility roles—for example, the glutathione-*S*-transferase (GST) tag improves solubility of some proteins, while solubility tags such as maltose-binding protein (MBP) can be used for affinity purification of the target protein.

2.2. Which tag(s) to use?

Protein affinity and solubility tags vary widely in their size, and thus in the metabolic load that they impose on the host cell. For example, an MBP tag (43 kDa) on a 100-residue target protein will require the expression of 5 mg of fusion protein for every mg of the target protein produced. Affinity tags also vary in the cost of purification, since different affinity media have different expenses, both for the resin itself, and for operating costs (ease of regeneration and reuse, elution agent, binding capacity, etc.). Lichty *et al.* (2005) did a comparison of eight affinity tags looking at many different aspects such as purity, yield, and cost, and many similar studies have been carried out. Several recent reviews provide good overviews (Terpe, 2003), and list advantages and disadvantages of affinity tags (Arnau *et al.*, 2006b; Waugh, 2005) and solubility tags (Esposito and Chatterjee, 2006).

Apart from the cost, the choice between different affinity tags often depends on finding purification buffer conditions suitable for the target protein. For proteins susceptible to oxidation or proteolytic damage, the His-tag may not be very suitable since immobilized metal affinity chromatographic (IMAC) media cannot tolerate reducing agents or EDTA. Similarly, care has to be taken when IMAC media is used with target proteins that are sensitive to metal ions. Conversely, for target proteins requiring denaturing conditions or refolding, the His-tag and IMAC purification is an excellent choice.

Expression levels can dictate the choice of tags in some cases—solubility tags such as MBP, Trx, NusA, as well as GST have strong translational initiation signals and can drive expression levels higher which is useful for structural studies. On the other hand, when low expression levels are desirable, such as when studying complexes or physiological interactions, more stringent epitope tags or tandem tagging may be more appropriate.

A major source of comparative studies and innovation in protein tags is coming from ongoing structural genomics efforts, since protein production is a big roadblock for these projects (Pédelacq *et al.*, 2002; Structural Genomics Consortium *et al.*, 2008; Yokoyama, 2003). The reader is encouraged to take advantage of the consensus ''what to try first'' strategies that are emerging from these efforts (Structural Genomics Consortium *et al.*, 2008).

Many high-throughput studies have looked more specifically at the effectiveness of individual solubility tags (e.g., Cabrita *et al.*, 2006; Kataeva *et al.*, 2005), and reporter systems have been developed to monitor the solubility and proper folding of proteins (Listwan *et al.*, 2009; Liu *et al.*, 2006; Waldo *et al.*, 1999). Directed evolution of proteins can also be used for improving protein expression and solubility (Roodveldt *et al.*, 2005; Waldo, 2003). Use of tags for production of transmembrane proteins is another area of active development (Cunningham and Deber, 2007).

2.3. Tandem tags?

Multiple tags can be attached on target proteins, allowing for improved purification, expression, or tracking. The tandem affinity purification (TAP) tag was originally developed to allow for two purification steps—first on Protein A IgG beads, followed by cleavage of the Protein A tag and subsequent purification on calmodulin coated beads (Fig. 16.1; Puig *et al.*, 2001). The use of tandem tags allows TAP-tagged proteins to be detected and purified in native conditions even when expressed at very low levels. TAP tagging has proved very useful for studying protein complexes (Bauer and Kuster, 2003), and many other variations of the TAP tags have been designed (Collins and Choudhary, 2008).

Solubility tags such as Trx or NusA can also be linked with affinity tags such as His-tags for efficient purification of the fusion protein. Other tags such as S-tags or FLAG-tags can be added to allow protein detection when low levels of protein expression are expected and less specific antibodies (such as those against His-tags) may not be adequate. Affinity tags can also be attached at both ends of a target protein (Mueller *et al.*, 2003).

2.4. N- or C-terminal?

Tags can be placed at either the N- or C-terminus of a target protein. One advantage of placing a tag on the N-terminal end is that the construct can take advantage of efficient translation initiation sites on the tag. Solubility tags based on highly expressing proteins such as MBP, Trx, and NusA are also more efficient at solubilizing target proteins when positioned at the N-terminal end (Sachdev and Chirgwin, 1998), though recent high-throughput studies have shown than the MBP tag is still quite effective when positioned at the C-terminal end (Dyson *et al.*, 2004). Another advantage of placing a tag on the N-terminal site is that the tag can be removed more cleanly, since most endoproteases cut at or near the C-terminus of their recognition sites.

While placing a tag, care should be taken to preserve the positioning of any signal sequences or modification sites. Sequences at termini of the fusion protein should be examined for effects on the stability of the final construct,

especially at the N-terminal end, which should be inspected for the host cell's N-end rule degradation signals (Bachmair *et al.*, 1986; Wang *et al.*, 2008). It is also useful to examine the sequence of the tagged protein for any inadvertently created interaction or cleavage sites using motif databases such as PROSITE (Hulo *et al.*, 2008).

2.5. Cleavage sites to remove tags

Tags can interfere with the structure and function of the target protein, and provision must be made to remove tags after the expression and purifications steps. Multiple cleavage sites can be engineered into the expression construct to remove individual tags at different stages of purification (Fig. 16.1). Table 16.3 lists some of the endoproteases used for tag removal, and these are discussed in this chapter.

3. Protein Affinity Tags

3.1. His-tag

The His-tag (also called 6xHis-tag) is one of the simplest and most widely used purification tags, with six or more consecutive histidine residues. These residues readily coordinate with transition metal ions such as Ni^{2+} or Co^{2+} immobilized on beads or a resin for purification. IMAC is the preferred choice as a first step during the purification of His-tagged proteins, though small batch reactions or spin columns with IMAC beads can be used for expression tests or small-scale preparations. Metal ions are immobilized using linkages such as Ni(II)-nitrilotriacetic acid (Ni-NTA) or Co^{2+}-carboxymethyl-aspartate on resins and beads available from many commercial sources. IMAC media typically has high binding capacities (5–40 mg of His-tagged protein/ml of media), is relatively low cost, and can easily be sanitized. For nickel binding media, the metal ion can often be stripped (using buffers with EDTA) and recharged for multiple use cycles. Some cobalt based resins (such as Talon, Clontech Inc.) use proprietary linkages that are more durable and cannot be recharged; such resins can be reused three and four times, but offer the advantage of being more specific for polyhistidine tags and almost no metal leakage during protein elution. IMAC can also be used under denaturing conditions, since the His-tag does not need a specific protein conformation for metal binding; indeed, binding to IMAC resins is stronger under denaturing conditions as the His-tag becomes more exposed.

His-tags bind the immobilized metal via the histidine imidazole ring, and tagged protein bound on IMAC media can be easily eluted using elution buffers with imidazole (100–250 m*M*) or low pH (4.5–6). Since some

endogenous proteins also exhibit weak binding to IMAC media (Bolanos-Garcia and Davies, 2006), a low level of imidazole (5–20 m*M*) should be included in the loading and wash buffers to minimize nonspecific binding.

While the mild elution conditions for IMAC is one of the positive aspects of His-tags, care has to be taken to avoid EDTA (or EGTA) in any of the buffers. Cell extracts loaded on IMAC columns should not contain any EDTA, and only EDTA-free protease inhibitor cocktails should be used for sample preparation. TRIS salts weakly chelate metal ions as well, and the use of TRIS buffers should be minimized (50 m*M* or less). Most IMAC media is also very sensitive to reducing agents such as DTT or DTE, and low levels of β-mercaptoethanol (<10 m*M*) should be used instead.

The small size of the His-tag minimizes interference with the folding and structure of the target protein, and this tag can be positioned at either the N- or C-terminal ends (Fig. 16.1A). While the tag can be removed by introducing a protease cleavage site, there are many examples of proteins that have been crystallized with intact His-tags and little or no impact on the structure of the tagged protein (Carson *et al.*, 2007). In some cases, the His-tag can actually assist in crystal formation (e.g., Smits *et al.*, 2008). The His-tag can also be used with commercially available His-tag-specific antibodies for protein detection.

There are several variations on the standard 6xHis sequence used in His-tags (Terpe, 2003). These tags intersperse multiple histidines among other residues to lower the charge or improve stability, such as the HAT-tag and 6xHN tag which also bind better to the Co^{2+}-Talon resin (Clontech Inc.), and the MAT-tag (Sigma-Aldrich Inc.).

3.2. GST tag

GST is an abundantly expressed 26 kDa eukaryotic protein, and GST cloned from *Schistosoma japonicum* was shown to promote solubility and expression as an N-terminal fusion (Smith and Johnson, 1988). When positioned at the C-terminal end (Fig. 16.1B), GST is less efficient at improving protein solubility but still functions well as an affinity tag. GST binds to resin immobilized glutathione, and this property is used for affinity purification of GST tagged proteins. After the fusion protein is bound to the resin, it can be eluted under rather mild conditions using free reduced glutathione (10–40 m*M*) at neutral pH. Resins used for GST fusion purification, such as Glutathione-Sepharose beads, are relatively cheap, have high binding capacity (5–10 mg of GST/ml of resin), and can be regenerated and reused multiple times.

The GST tag has to be properly folded to bind glutathione, and thus the fusion protein needs to be soluble and in nondenaturing conditions for efficient purification. GST fusion proteins are often expressed at high levels, and thus protein solubility has to be monitored carefully. For some proteins, GST can act as a solubility tag (e.g., Kim and Lee, 2008). GST can be

detected by a colorimetric assay with the substrate 1-chloro-2,4 dinitrobenzene (CDNB) (Habig *et al.*, 1974); this assay can also be used to monitor proper folding and accessibility of GST in the fusion protein when binding to glutathione resin is suboptimal. Commercial anti-GST antibodies are also available for detecting this tag.

The large size of the GST tag increases potential for degradation by proteases, and GST fusion proteins need to be purified quickly to minimize sample loss. Unlike for His-tagged proteins, EDTA can be used in buffers during sample preparation to reduce proteolytic damage. Care should also be taken to use reducing conditions since GST has four solvent exposed cysteines that can be involved in oxidative aggregation (Kaplan *et al.*, 1997). GST forms a homodimer in solution, which makes it a poor choice for tagging oligomeric proteins.

The kinetics of GST binding to glutathione and its elution are relatively slow, and so GST fusion proteins need to be loaded and eluted from GST columns at slow flow rates. The monitoring of protein during elution also has to be done carefully since glutathione absorbs strongly at 280 nm.

3.3. Other purification tags

3.3.1. Epitope tags

A number of short amino acid (aa) sequences, recognized by commercially available antibodies, can be used as tags for detection and purification of proteins. Epitope tagging is widely used in molecular biology as a general method for tracking recombinant proteins (Fritze and Anderson, 2000), and has the advantage of high specificity and the use of short tags that minimizes deleterious effects on the structure and function of the target protein. These tags are mostly placed at the N- or C-termini, but can also be placed within a target protein (in loops or between structural domains in a solvent exposed region). Epitope tags are usually sequences absent in the host cell, making the detection of the target protein straightforward. For purification, however, epitope tag binding media (typically, monoclonal antibodies immobilized on chromatographic resins) is expensive and less suitable for large-scale preparations than other affinity media. Short peptides corresponding to the tag, low pH, or other approaches (e.g., chelation of calcium required for tag binding, or use of salt and polyol) can be used to elute the target protein, but some of these procedures tend to be more harsh than other affinity purification methods (however, see Chapter 28 in this volume).

The FLAG-tag is a short, eight-residue (DYKDDDDK) hydrophilic peptide tag that can be used for detection and purification of target proteins (Einhauer and Jungbauer, 2001; Prickett *et al.*, 1989). Apart from availability of several high specificity anti-FLAG mABs, the FLAG-tag has the advantage of incorporating the enterokinase cleavage site (DDDK) that allows the complete tag to be cleanly removed after purification. Another

variant of the FLAG-tag is the 3xFLAG-tag (Sigma-Aldrich, Inc.) that is made up of three tandem repeats of FLAG-like sequences (Hernan *et al.*, 2000). Other commonly used epitope tags are the HA-tag and the c-Myc tag (Fritze and Anderson, 2000).

3.3.2. The S-tag

S-tag, reviewed by Raines *et al.* (2000), takes advantage of the tight association between the N-terminal S-peptide (residues 1–20) and the S-protein (residues 21–124) in RNase S (Richards and Vithayarhil, 1959). The S-tag system (Novagen Inc.) uses the N-terminal 15 residues of the S-peptide (Potts *et al.*, 1963), with S-protein immobilized on agarose beads for target protein purification. S-tag vectors typically encode a site-specific protease cleavage site, and elution of the target protein requires cleavage of the tag or harsher denaturing conditions that disrupt the S-tag S-protein interaction.

3.3.3. STREP-II-tag

STREP-II-tag (WSHPQFEK) is a tag that takes advantage of the strong and specific interaction between biotin and streptavidin (Schmidt and Skerra, 1994). This peptide tag binds in the biotin pocket of streptavidin; Strep-Tactin, a recombinant form of streptavidin optimized to bind the Strep-II-tag, is used in affinity media. Bound target protein can be eluted with low levels (2.5 mM) of desthiobiotin, a biotin analog that competes for binding to Strep-Tactin in a reversible manner (Skerra and Schmidt, 2000). The elution conditions are gentle, and buffers can include high levels (up to 50 mM) of reducing agents such as DTT or β-mercaptoethanol, as well as chelating agents such as EDTA, which makes this an excellent purification technique for proteins that are very sensitive to oxidation. Strep-Tactin chromatographic resins can regenerated and reused a few times. While there is a low background of naturally biotinylated proteins in most cell extracts, these usually do not interfere in the purification; if necessary, avidin can be added to clear biotinylated host proteins (Schmidt and Skerra, 2007). There has also been some work on using streptavidin/avidin in fusion tags (Sano and Cantor, 2000), but they are difficult to use for affinity purification because the very strong interaction between these proteins is difficult to disrupt.

3.3.4. CBP-tag

The calmodulin-binding peptide (CBP) is a short 26-residue sequence derived from the C-terminus of skeletal muscle myosin light chain kinase that binds specifically to calmodulin (reviewed in Terpe, 2003). Calmodulin (CaM) immobilized on chromatographic media (such as CaM-Sepharose) can be used for affinity purification of target proteins tagged with CBP. Though calmodulin binds very strongly with CBP (nanomolar affinity), this interaction is dependent on calcium and the bound protein can be eluted in a single step using gentle buffers containing a calcium-chelating agent, such

as EGTA. Another application of the CBP-tag is for ^{32}P isotopic labeling of the fusion protein, since the tag includes a protein kinase A target sequence (Vaillancourt *et al.*, 2000). The CBP-tag cannot be used for expression in eukaryotic systems, since many endogenous proteins in eukaryotes also interact with calmodulin.

4. Solubility Tags

Production of well-folded, soluble proteins is a major bottleneck in recombinant protein expression, especially when heterologous eukaryotic proteins are expressed in bacterial cells. Several soluble proteins are used as tags to improve folding of the target protein. These tags should be used in conjunction with other approaches to improve protein folding such as lowering temperature after protein induction or coexpression of chaperones (Baneyx and Mujacic, 2004; de Marco *et al.*, 2007; Sahdev *et al.*, 2008). It is also useful to screen multiple solubility tags (Peleg and Unger, 2008). For recalcitrant proteins, protein purification using denaturing conditions and refolding can be tried (Cabrita and Bottomley, 2004; Jungbauer and Kaar, 2007; Qoronfleh *et al.*, 2007; see Chapter 17 in this volume).

4.1. MBP tag

MBP is a solubility enhancing tag (di Guan *et al.*, 1988) that can also be used for effective affinity purification, since it binds specifically to maltose or amylose. MBP is a large 43 kDa secreted *E. coli* protein that can be expressed at very high levels, and helps keep proteins fused at its C-terminal end soluble (Kapust and Waugh, 1999). More recent surveys have shown than the MBP tag is also effective when placed on the C-terminal end of target proteins (Dyson *et al.*, 2004). The large size of this tag puts a heavy metabolic load on the host cell, but in high-throughput tests, MBP ranks as one of the best tags for making soluble fusions (Dyson *et al.*, 2004; Kataeva *et al.*, 2005). About a quarter of proteins expressed as MBP fusions, however, remain insoluble or are prone to aggregation when the MBP tag is removed. In our work on the expression of human calcitonin gene-related peptide-receptor component protein (CGRP-RCP) in *E. coli*, for example, we observed protein aggregation when the N-terminal MBP tag was removed by enterokinase cleavage even after successful expression and purification (Tolun *et al.*, 2007). Each solubility tag also has different effects, and several tags may need to be tried for recalcitrant proteins. In the work on CGRP-RCP, a thioredoxin tag was not effective and the fusion protein remained insoluble (Tolun *et al.*, 2007). Nallamsetty and Waugh (2006) have suggested that solubility tags such as MBP and NusA act as passive partners in the folding of target proteins, and that solubility of

aggregation-prone target proteins after removal of the tag appears to depend on the characteristics of the target protein rather than the tag used.

Commercial vectors for tagging MBP are available for both cytoplasmic and periplasmic expression of target proteins, with a variety of cleavage sites (New England Biolabs Inc.). Cross-linked amylose resin is used to bind MBP tagged proteins, and the bound fusion protein can be easily eluted by adding 10 m*M* maltose to the wash buffers. This allows for an easy one-step purification of MBP tagged proteins under very gentle conditions; however, amylose affinity purification cannot be carried out with reducing agents or under denaturing conditions. Amylose resins are degraded somewhat by amylase activity in crude extracts, especially from cells grown in rich LB media, but this can be minimized by including glucose in the media (0.2%). Amylose resins can be regenerated and reused several times.

4.2. Trx tag

Thioredoxin (Trx) is a thermostable, 12-kDa intracellular *E. coli* protein that is easily overexpressed and soluble even when overexpressed up to 40% of the total cellular protein (LaVallie *et al.*, 1993), and is very useful as a tag in avoiding inclusion body formation in recombinant protein production (LaVallie *et al.*, 2000). Tests by Dyson *et al.* (2004) indicate that the Trx tag is more effective when placed on the N-terminal end of the target protein.

Thioredoxin accumulates at cytoplasmic membrane adhesion sites (Bayer *et al.*, 1987), which allows Trx fusion proteins to be released by simple osmotic shock or freeze/thaw treatments, providing a simple initial purification step. Additional affinity tags such as His-tags are typically attached to the Trx tag for further purification steps.

4.3. NusA tag

NusA is a large (495 aa) N-utilizing substance A transcription antitermination factor that was chosen as a potential tag as it ranked among the most soluble *E. coli* proteins as a fusion for heterologous protein expression, based on a statistical solubility model (Davis *et al.*, 1999; De Marco *et al.*, 2004). In some large-scale screening tests (Busso *et al.*, 2005; Kataeva *et al.*, 2005), NusA has performed as well as or better than MBP as a solubility tag, but the results vary for individual proteins. NusA tags are used in conjunction with other affinity tags such as His-tags (de Marco, 2006).

4.4. Other solubility tags

A number of smaller solubility enhancement tags (SETs) or solubility enhancement peptide (SEP) tags have been developed, which use highly acidic sequences to enhance solubility of some target proteins (Kato *et al.*, 2007;

Zhang *et al.*, 2004). Other small tags are the GB1-tag (56 residues) which is based on the IgG binding B1 domain of the streptococcal Protein G (Cheng and Patel, 2004; Zhou *et al.*, 2001), as well as the IgG binding domain of Protein A (ZZ domain, 116 residues; Inouye and Sahara, 2009; Rondahl *et al.*, 1992). These small tags are especially useful in protein preparations for NMR studies (Kato *et al.*, 2007), and some progress has been made in the creation of NMR-invisible solubility tags using protein ligation methods (Durst *et al.*, 2008; Kobashigawa *et al.*, 2009).

Small ubiquitin-like modifier (SUMO) protein (~11 kDa) has been shown to significantly improve protein stability and solubility as an N-terminal fusion (Marblestone *et al.*, 2006). The tag can be removed after purification using SUMO protease (catalytic domain of Ulp1) that recognizes the SUMO structure (Lee *et al.*, 2008; Panavas *et al.*, 2009). The SUMO-fusion system has also been adapted for use in insect cells and other eukaryotic expression systems (Liu *et al.*, 2008).

HaloTag is a recently created modular tagging system that uses a 34-kDa modified haloalkane dehalogenase protein that can bind a variety of synthetic ligands (HaloTag ligands; Promega Inc.). These ligands are comprised of a constant reactive linker that binds covalently to the HaloTag, and a variable reporter end that can impart a variety of useful properties to the fusion protein. Thus, a single tag can be used for subcellular imaging within live cells, cell labeling and sorting, affinity purification, or even immobilization on solid supports (Los *et al.*, 2008). Ohana *et al.* (2009) have used a version of this tag (HaloTag7) for affinity purification, using the chloroalkane linker attached to agarose beads. Since the HaloTag binds covalently to this linker in a very specific nonreversible manner, even poorly expressed proteins can be bound efficiently to the chloroalkane resin. The target protein can then be eluted off the resin using tobacco etch virus (TEV) protease that cuts at a cleavage site engineered between the HaloTag and the target protein. Surprisingly, the monomeric compact HaloTag was seen to dramatically improve fusion protein solubility when tested on a panel of difficult-to-express recombinant human proteins being expressed in *E. coli* (Ohana *et al.*, 2009). In these tests, the HaloTag fared significantly better than MBP, and appears to function as a bona fide solubility tag.

5. Removal of Tags

After a protein tag has been used for solubility enhancement or affinity purification, it is often useful to remove it for biological and functional studies since the tag can potentially interfere with the proper functioning of the target protein. This is especially true for large tags such GST or MBP,

though there are some examples where the fusion protein was more amenable for crystallization (Smyth *et al.*, 2003).

Most commercial expression vectors that are used to add tags on target proteins also include cleavage sites with specific sequences that allow the tag to be removed using recombinant endoproteases. After the initial affinity purification step, the sample can be treated with the endoprotease to cleave off the tag, which can subsequently be separated from the target protein by passing the sample back on the affinity column and collecting the flow-through. The recombinant endoprotease usually also comes with an affinity tag, allowing for its easy removal after the cleavage reaction.

Some commonly used endoproteases are listed in Table 16.3. Enterokinase and Factor Xa are useful for removal of N-terminal tags since they cut at the C-terminal end of their recognition sequence, allowing for the complete tag and recognition sequence to be removed. However, both these enzymes are somewhat promiscuous and can cleave at secondary sites, often at other basic residues. Similar secondary cleavage sites have also been seen for thrombin, another protease used to remove tags (Jenny *et al.*, 2003; Liew *et al.*, 2005). One advantage with proteases such as thrombin, especially for large-scale protein production, is that it is cheaper and more efficient than other more specific proteases.

The PreScission protease is a more specific protease with a longer and stricter recognition sequence. This is the protease 3C from human rhinovirus-14 (3Cpro) with a GST tag, which makes it especially useful for removing GST tags. Another very specific and popular protease is the TEV protease (Kapust *et al.*, 2001). The TEV protease prefers a Gly after the cleavage site (Table 16.3), but can tolerate other residues with only a modest decrease in activity, which allows it to cleave N-terminal tags with no additional residues remaining on the target protein in many cases (Kapust *et al.*, 2002). The TEV protease is easy to produce in-house in large quantities and is usually expressed with a His-tag for convenient purification and removal after the cleavage reaction (Tropea *et al.*, 2009). Many forms of TEV proteases, with other affinity tags or enhanced for higher activity and stability, are sold commercially.

A variant of the TEV protease is the tobacco vein mottling virus (TVMV) protease that recognizes a different sequence (Nallamsetty *et al.*, 2004), and can be used to design separate cleavage sites between different tags (Fig. 16.1C).

Exoproteases can also be used to remove N-terminal tags, such as the TAGzyme system (Arnau *et al.*, 2006a, available commercially from Qiagen Inc.). This approach uses dipeptide aminopeptidase I (DAPase) to sequentially chew up the N-terminal tag until a dipeptide ''stop point'' in the sequence is reached. Additional variations of this system have been designed to work with a variety of sequences (Arnau *et al.*, 2006b, 2008).

Complete removal of C-terminal tags is more problematic, since most endoproteases cut toward the C-terminal end of their recognition sequence.

If complete removal of a tag is necessary, more specialized cleavage sites can be designed that take advantage of structure-based recognition (such as with the SUMO protease; Malakhov *et al.*, 2004) or an autocatalytic protein self-splicing element (Inteins; Saleh and Perler, 2006). Both of these cleavage systems have been coupled with affinity purification tags—the SUMO-fusion system (Butt *et al.*, 2005; Lee *et al.*, 2008), and intein-chitin-binding domain (CBD; also commercialized as the IMPACT system by New England Biolabs Inc.; Chong *et al.*, 1997) or the intein-polyhydroxybutyrate-binding (PHB) and similar protein purification systems (Gillies *et al.*, 2008).

In some cases, tags can also be cleaved chemically. Though chemical cleavage uses cheap reagents and can be very efficient, these reactions require harsh solvents and denaturing conditions, and are used typically for preparation of small peptides. Cyanogen bromide (CNBr) cleaves at methionines, and can be used when the fusion protein can be designed to have a unique methionine between the tag and the target peptide (Döbeli *et al.*, 1998; Fairlie *et al.*, 2002). Another chemical cleavage agent, hydroxylamine, cleaves the peptide bond between Asn and Gly (Hu *et al.*, 2008).

Practical considerations: While designing a specific cleavage site between a tag and a target protein is relatively straightforward, efficient cleavage of the tag is not always possible or easily predictable. Each construct has to be experimentally tested both for cleavage efficiency, and for any secondary cleavages that may occur when promiscuous proteases are used. Often, the level of the protease and the duration of incubation have to be optimized. Maximizing the efficiency of cleavage is especially important for oligomeric proteins where tags have to be removed for each monomer for productive yields of the target protein (Kenig *et al.*, 2006). The cleavage sequence also has to be sterically accessible to the protease and relatively unstructured; poor cleavage can sometimes be alleviated by introducing a spacer or linker of a few residues between the recognition site and the target protein, or by using sequences around the cleavage site that are unlikely to form secondary structures. For proteases that lack an affinity tag, on-column cleavage (where the protease is injected onto the affinity column with the fusion protein bound) is often more efficient than batch reactions.

6. Conclusions

A large variety of protein tags are available for facilitating the soluble expression and purification of recombinant proteins. However, even with this wide arsenal of tags, structural genomics protein production facilities are getting soluble purified proteins with success rates of less than 50% (Structural Genomics Consortium *et al.*, 2008). While many good strategies

have emerged from these large-scale studies, given the diversity in how proteins fold and their distinct biochemical characteristics, there is no common set of solubility or affinity tags that works for all proteins. Rather, the choice of tags largely depends on the protein being expressed and the task at hand. Herein, we have surveyed the most effective protein expression tags and issues related to their use.

ACKNOWLEDGMENT

The author is supported in part by a grant (R01-GM69972) from the National Institutes of Health.

REFERENCES

Arnau, J., Lauritzen, C., and Pedersen, J. (2006a). Cloning strategy, production and purification of proteins with exopeptidase-cleavable His-tags. *Nat. Protoc.* **1,** 2326–2333.

Arnau, J., Lauritzen, C., Petersen, G. E., and Pedersen, J. (2006b). Current strategies for the use of affinity tags and tag removal for the purification of recombinant proteins. *Protein Expr. Purif.* **48,** 1–13.

Arnau, J., Lauritzen, C., Petersen, G. E., and Pedersen, J. (2008). The use of TAGZyme for the efficient removal of N-terminal His-tags. *Methods Mol. Biol.* **421,** 229–243.

Bachmair, A., Finley, D., and Varshavsky, A. (1986). *In vivo* half-life of a protein is a function of its amino-terminal residue. *Science* **234,** 179–186.

Baneyx, F., and Mujacic, M. (2004). Recombinant protein folding and misfolding in *Escherichia coli*. *Nat. Biotechnol.* **22,** 1399–1408.

Bauer, A., and Kuster, B. (2003). Affinity purification-mass spectrometry. Powerful tools for the characterization of protein complexes. *Eur. J. Biochem.* **270,** 570–578.

Bayer, M. E., Bayer, M. H., Lunn, C. A., and Pigiet, V. (1987). Association of thioredoxin with the inner membrane and adhesion sites in *Escherichia coli*. *J. Bacteriol.* **169,** 2659–2666.

Bolanos-Garcia, V. M., and Davies, O. R. (2006). Structural analysis and classification of native proteins from *E. coli* commonly co-purified by immobilised metal affinity chromatography. *Biochim. Biophys. Acta* **1760,** 1304–1313.

Busso, D., Delagoutte-Busso, B., and Moras, D. (2005). Construction of a set Gateway-based destination vectors for high-throughput cloning and expression screening in *Escherichia coli*. *Anal. Biochem.* **343,** 313–321.

Butt, T. R., Edavettal, S. C., Hall, J. P., and Mattern, M. R. (2005). SUMO fusion technology for difficult-to-express proteins. *Protein Expr. Purif.* **43,** 1–9.

Cabrita, L. D., and Bottomley, S. P. (2004). Protein expression and refolding—A practical guide to getting the most out of inclusion bodies. *Biotechnol. Annu. Rev.* **10,** 31–50.

Cabrita, L. D., Dai, W., and Bottomley, S. P. (2006). A family of *E. coli* expression vectors for laboratory scale and high throughput soluble protein production. *BMC Biotechnol.* **6,** 12.

Carson, M., Johnson, D. H., McDonald, H., Brouillette, C., and Delucas, L. J. (2007). His-tag impact on structure. *Acta Crystallogr. D Biol. Crystallogr.* **63,** 295–301.

Cheng, Y., and Patel, D. J. (2004). An efficient system for small protein expression and refolding. *Biochem. Biophys. Res. Commun.* **317,** 401–405.

Chong, S., Mersha, F. B., Comb, D. G., Scott, M. E., Landry, D., Vence, L. M., Perler, F. B., Benner, J., Kucera, R. B., Hirvonen, C. A., *et al.* (1997). Single-column purification of free recombinant proteins using a self-cleavable affinity tag derived from a protein splicing element. *Gene* **192,** 271–281.

Collins, M. O., and Choudhary, J. S. (2008). Mapping multiprotein complexes by affinity purification and mass spectrometry. *Curr. Opin. Biotechnol.* **19,** 324–330.

Cunningham, F., and Deber, C. M. (2007). Optimizing synthesis and expression of transmembrane peptides and proteins. *Methods* **41,** 370–380.

Davis, G. D., Elisee, C., Newham, D. M., and Harrison, R. G. (1999). New fusion protein systems designed to give soluble expression in *Escherichia coli*. *Biotechnol. Bioeng.* **65,** 382–388.

de Marco, A. (2006). Two-step metal affinity purification of double-tagged (NusA-His6) fusion proteins. *Nat. Protoc.* **1,** 1538–1543.

De Marco, V., Stier, G., Blandin, S., and de Marco, A. (2004). The solubility and stability of recombinant proteins are increased by their fusion to NusA. *Biochem. Biophys. Res. Commun.* **322,** 766–771.

de Marco, A., Deuerling, E., Mogk, A., Tomoyasu, T., and Bukau, B. (2007). Chaperone-based procedure to increase yields of soluble recombinant proteins produced in *E. coli*. *BMC Biotechnol.* **7,** 32.

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

Döbeli, H., Andres, H., Breyer, N., Draeger, N., Sizmann, D., Zuber, M. T., Weinert, B., and Wipf, B. (1998). Recombinant fusion proteins for the industrial production of disulfide bridge containing peptides: Purification, oxidation without concatemer formation, and selective cleavage. *Protein Expr. Purif.* **12,** 404–414.

Durst, F. G., Ou, H. D., Löhr, F., Dötsch, V., and Straub, W. E. (2008). The better tag remains unseen. *J. Am. Chem. Soc.* **130,** 14932–14933.

Dyson, M. R., Shadbolt, S. P., Vincent, K. J., Perera, R. L., and McCafferty, J. (2004). Production of soluble mammalian proteins in *Escherichia coli*: Identification of protein features that correlate with successful expression. *BMC Biotechnol.* **4,** 32.

Einhauer, A., and Jungbauer, A. (2001). The FLAG peptide, a versatile fusion tag for the purification of recombinant proteins. *J. Biochem. Biophys. Methods* **49,** 455–465.

Esposito, D., and Chatterjee, D. K. (2006). Enhancement of soluble protein expression through the use of fusion tags. *Curr. Opin. Biotechnol.* **17,** 353–358.

Fairlie, W. D., Uboldi, A. D., De Souza, D. P., Hemmings, G. J., Nicola, N. A., and Baca, M. (2002). A fusion protein system for the recombinant production of short disulfide-containing peptides. *Protein Expr. Purif.* **26,** 171–178.

Fritze, C. E., and Anderson, T. R. (2000). Epitope tagging: General method for tracking recombinant proteins. *Methods Enzymol.* **327,** 3–16.

Gillies, A. R., Hsii, J. F., Oak, S., and Wood, D. W. (2008). Rapid cloning and purification of proteins: Gateway vectors for protein purification by self-cleaving tags. *Biotechnol. Bioeng.* **101,** 229–240.

Habig, W. H., Pabst, M. J., and Jakoby, W. B. (1974). Glutathione S-transferases. The first enzymatic step in mercapturic acid formation. *J. Biol. Chem.* **249,** 7130–7139.

Hernan, R., Heuermann, K., and Brizzard, B. (2000). Multiple epitope tagging of expressed proteins for enhanced detection. *Biotechniques* **28,** 789–793.

Hu, J., Qin, H., Sharma, M., Cross, T. A., and Gao, F. P. (2008). Chemical cleavage of fusion proteins for high-level production of transmembrane peptides and protein domains containing conserved methionines. *Biochim. Biophys. Acta* **1778,** 1060–1066.

Hulo, N., Bairoch, A., Bulliard, V., Cerutti, L., Cuche, B. A., de Castro, E., Lachaize, C., Langendijk-Genevaux, P. S., and Sigrist, C. J. (2008). The 20 years of PROSITE. *Nucleic Acids Res.* **36,** D245–D249.

Inouye, S., and Sahara, Y. (2009). Expression and purification of the calcium binding photoprotein mitrocomin using ZZ-domain as a soluble partner in *E. coli* cells. *Protein Expr. Purif.* **66,** 52–57.

Jenny, R. J., Mann, K. G., and Lundblad, R. L. (2003). A critical review of the methods for cleavage of fusion proteins with thrombin and factor Xa. *Protein Expr. Purif.* **31,** 1–11.

Jungbauer, A., and Kaar, W. (2007). Current status of technical protein refolding. *J. Biotechnol.* **128,** 587–596.

Kaplan, W., Hüsler, P., Klump, H., Erhardt, J., Sluis-Cremer, N., and Dirr, H. (1997). Conformational stability of pGEX-expressed *Schistosoma japonicum* glutathione S-transferase: A detoxification enzyme and fusion-protein affinity tag. *Protein Sci.* **6,** 399–406.

Kapust, R. B., and Waugh, D. S. (1999). *Escherichia coli* maltose-binding protein is uncommonly effective at promoting the solubility of polypeptides to which it is fused. *Protein Sci.* **8,** 1668–1674.

Kapust, R. B., Tözsér, J., Fox, J. D., Anderson, D. E., Cherry, S., Copeland, T. D., and Waugh, D. S. (2001). Tobacco etch virus protease: Mechanism of autolysis and rational design of stable mutants with wild-type catalytic proficiency. *Protein Eng.* **14,** 993–1000.

Kapust, R. B., Tözsér, J., Copeland, T. D., and Waugh, D. S. (2002). The P1′ specificity of tobacco etch virus protease. *Biochem. Biophys. Res. Commun.* **294,** 949–955.

Kataeva, I., Chang, J., Xu, H., Luan, C. H., Zhou, J., Uversky, V. N., Lin, D., Horanyi, P., Liu, Z. J., Ljungdahl, L. G., *et al.* (2005). Improving solubility of *Shewanella oneidensis* MR-1 and *Clostridium thermocellum* JW-20 proteins expressed into *Escherichia coli*. *J. Proteome Res.* **4,** 1942–1951.

Kato, A., Maki, K., Ebina, T., Kuwajima, K., Soda, K., and Kuroda, Y. (2007). Mutational analysis of protein solubility enhancement using short peptide tags. *Biopolymers* **85,** 12–18.

Kenig, M., Peternel, S., Gaberc-Porekar, V., and Menart, V. (2006). Influence of the protein oligomericity on final yield after affinity tag removal in purification of recombinant proteins. *J. Chromatogr. A* **1101,** 293–306.

Kim, S., and Lee, S. B. (2008). Soluble expression of archaeal proteins in *Escherichia coli* by using fusion-partners. *Protein Expr. Purif.* **62,** 116–119.

Kobashigawa, Y., Kumeta, H., Ogura, K., and Inagaki, F. (2009). Attachment of an NMR-invisible solubility enhancement tag using a sortase-mediated protein ligation method. *J. Biomol. NMR* **43,** 145–150.

LaVallie, E. R., DiBlasio, E. A., Kovacic, S., Grant, K. L., Schendel, P. F., and McCoy, J. M. (1993). A thioredoxin gene fusion expression system that circumvents inclusion body formation in the *E. coli* cytoplasm. *Biotechnology (NY)* **11,** 187–193.

LaVallie, E. R., Lu, Z., Diblasio-Smith, E. A., Collins-Racie, L. A., and McCoy, J. M. (2000). Thioredoxin as a fusion partner for production of soluble recombinant proteins in *Escherichia coli*. *Methods Enzymol.* **326,** 322–340.

Lee, C. D., Sun, H. C., Hu, S. M., Chiu, C. F., Homhuan, A., Liang, S. M., Leng, C. H., and Wang, T. F. (2008). An improved SUMO fusion protein system for effective production of native proteins. *Protein Sci.* **17,** 1241–1248.

Lichty, J. J., Malecki, J. L., Agnew, H. D., Michelson-Horowitz, D. J., and Tan, S. (2005). Comparison of affinity tags for protein purification. *Protein Expr. Purif.* **41,** 98–105.

Liew, O. W., Ching Chong, J. P., Yandle, T. G., and Brennan, S. O. (2005). Preparation of recombinant thioredoxin fused N-terminal proCNP: Analysis of enterokinase cleavage products reveals new enterokinase cleavage sites. *Protein Expr. Purif.* **41,** 332–340.

Listwan, P., Terwilliger, T. C., and Waldo, G. S. (2009). Automated, high-throughput platform for protein solubility screening using a split-GFP system. *J. Struct. Funct. Genomics* **10,** 47–55.

Liu, J. W., Boucher, Y., Stokes, H. W., and Ollis, D. L. (2006). Improving protein solubility: The use of the *Escherichia coli* dihydrofolate reductase gene as a fusion reporter. *Protein Expr. Purif.* **47,** 258–263.

Liu, L., Spurrier, J., Butt, T. R., and Strickler, J. E. (2008). Enhanced protein expression in the baculovirus/insect cell system using engineered SUMO fusions. *Protein Expr. Purif.* **62,** 21–28.

Los, G. V., Encell, L. P., McDougall, M. G., Hartzell, D. D., Karassina, N., Zimprich, C., Wood, M. G., Learish, R., Ohana, R. F., Urh, M., *et al.* (2008). HaloTag: A novel protein labeling technology for cell imaging and protein analysis. *ACS Chem. Biol.* **3,** 373–382.

Malakhov, M. P., Mattern, M. R., Malakhova, O. A., Drinker, M., Weeks, S. D., and Butt, T. R. (2004). SUMO fusions and SUMO-specific protease for efficient expression and purification of proteins. *J. Struct. Funct. Genomics* **5,** 75–86.

Marblestone, J. G., Edavettal, S. C., Lim, Y., Lim, P., Zuo, X., and Butt, T. R. (2006). Comparison of SUMO fusion technology with traditional gene fusion systems: Enhanced expression and solubility with SUMO. *Protein Sci.* **15,** 182–189.

Mueller, U., Büssow, K., Diehl, A., Bartl, F. J., Niesen, F. H., Nyarsik, L., and Heinemann, U. (2003). Rapid purification and crystal structure analysis of a small protein carrying two terminal affinity tags. *J. Struct. Funct. Genomics* **4,** 217–225.

Nallamsetty, S., and Waugh, D. S. (2006). Solubility-enhancing proteins MBP and NusA play a passive role in the folding of their fusion partners. *Protein Expr. Purif.* **45,** 175–182.

Nallamsetty, S., Kapust, R. B., Tözsér, J., Cherry, S., Tropea, J. E., Copeland, T. D., and Waugh, D. S. (2004). Efficient site-specific processing of fusion proteins by tobacco vein mottling virus protease *in vivo* and *in vitro*. *Protein Expr. Purif.* **38,** 108–115.

Ohana, R. F., Encell, L. P., Zhao, K., Simpson, D., Slater, M. R., Urh, M., and Wood, K. V. (2009). HaloTag7: A genetically engineered tag that enhances bacterial expression of soluble proteins and improves protein purification. *Protein Expr. Purif.* **68,** 110–120.

Panavas, T., Sanders, C., and Butt, T. R. (2009). SUMO fusion technology for enhanced protein production in prokaryotic and eukaryotic expression systems. *Methods Mol. Biol.* **497,** 303–317.

Pédelacq, J. D., Piltch, E., Liong, E. C., Berendzen, J., Kim, C. Y., Rho, B. S., Park, M. S., Terwilliger, T. C., and Waldo, G. S. (2002). Engineering soluble proteins for structural genomics. *Nat. Biotechnol.* **20,** 927–932.

Peleg, Y., and Unger, T. (2008). Application of high-throughput methodologies to the expression of recombinant proteins in *E. coli*. *Methods Mol. Biol.* **426,** 197–208.

Potts, J. T., Young, D. M., and Anfinsen, C. B. (1963). Reconstitution of fully active RNase S by carboxypeptidase-degraded RNase S-peptide. *J. Biol. Chem.* **238,** 2593–2594.

Prickett, K. S., Amberg, D. C., and Hopp, T. P. (1989). A calcium-dependent antibody for identification and purification of recombinant proteins. *Biotechniques* **7,** 580–589.

Puig, O., Caspary, F., Rigaut, G., Rutz, B., Bouveret, E., Bragado-Nilsson, E., Wilm, M., and Séraphin, B. (2001). The tandem affinity purification (TAP) method: A general procedure of protein complex purification. *Methods* **24,** 218–229.

Qoronfleh, M. W., Hesterberg, L. K., and Seefeldt, M. B. (2007). Confronting high-throughput protein refolding using high pressure and solution screens. *Protein Expr. Purif.* **55,** 209–224.

Raines, R. T., McCormick, M., Van Oosbree, T. R., and Mierendorf, R. C. (2000). The S.Tag fusion system for protein purification. *Methods Enzymol.* **326,** 362–376.

Richards, F. M., and Vithayarhil, P. J. (1959). The preparation of subtilisin-modified ribonuclease and the separation of the peptide and protein components. *J. Biol. Chem.* **234,** 1459–1465.

Rondahl, H., Nilsson, B., and Holmgren, E. (1992). Fusions to the 5′ end of a gene encoding a two-domain analogue of staphylococcal protein A. *J. Biotechnol.* **25,** 269–287.

Roodveldt, C., Aharoni, A., and Tawfik, D. S. (2005). Directed evolution of proteins for heterologous expression and stability. *Curr. Opin. Struct. Biol.* **15,** 50–56.

Sachdev, D., and Chirgwin, J. M. (1998). Order of fusions between bacterial and mammalian proteins can determine solubility in *Escherichia coli*. *Biochem. Biophys. Res. Commun.* **244,** 933–937.
Sahdev, S., Khattar, S. K., and Saini, K. S. (2008). Production of active eukaryotic proteins through bacterial expression systems: A review of the existing biotechnology strategies. *Mol. Cell. Biochem.* **307,** 249–264.
Saleh, L., and Perler, F. B. (2006). Protein splicing in cis and in trans. *Chem. Rec.* **6,** 183–193.
Sano, T., and Cantor, C. R. (2000). Streptavidin-containing chimeric proteins: Design and production. *Methods Enzymol.* **326,** 305–311.
Schmidt, T. G., and Skerra, A. (1994). One-step affinity purification of bacterially produced proteins by means of the "Strep tag" and immobilized recombinant core streptavidin. *J. Chromatogr. A* **676,** 337–345.
Schmidt, T. G., and Skerra, A. (2007). The Strep-tag system for one-step purification and high-affinity detection or capturing of proteins. *Nat. Protoc.* **2,** 1528–1535.
Skerra, A., and Schmidt, T. G. (2000). Use of the Strep-tag and streptavidin for detection and purification of recombinant proteins. *Methods Enzymol.* **326,** 271–304.
Smith, D. B., and Johnson, K. S. (1988). Single-step purification of polypeptides expressed in *Escherichia coli* as fusions with glutathione S-transferase. *Gene* **67,** 31–40.
Smits, S. H., Mueller, A., Grieshaber, M. K., and Schmitt, L. (2008). Coenzyme- and His-tag-induced crystallization of octopine dehydrogenase. *Acta Crystallogr. Sect. F Struct. Biol. Cryst. Commun.* **64,** 836–839.
Smyth, D. R., Mrozkiewicz, M. K., McGrath, W. J., Listwan, P., and Kobe, B. (2003). Crystal structures of fusion proteins with large-affinity tags. *Protein Sci.* **12,** 1313–1322.
Structural Genomics Consortium, China Structural Genomics Consortium, Northeast Structural Genomics Consortium, Gräslund, S., Nordlund, P., Weigelt, J., Hallberg, B. M., Bray, J., Gileadi, O., Knapp, S., *et al.* (2008). Protein production and purification. *Nat. Methods* **5,** 135–146.
Terpe, K. (2003). Overview of tag protein fusions: From molecular and biochemical fundamentals to commercial systems. *Appl. Microbiol. Biotechnol.* **60,** 523–533.
Tolun, A. A., Dickerson, I. M., and Malhotra, A. (2007). Overexpression and purification of human calcitonin gene-related peptide-receptor component protein in *Escherichia coli*. *Protein Expr. Purif.* **52,** 167–174.
Tropea, J. E., Cherry, S., and Waugh, D. S. (2009). Expression and purification of soluble His(6)-tagged TEV protease. *Methods Mol. Biol.* **498,** 297–307.
Vaillancourt, P., Zheng, C. F., Hoang, D. Q., and Breister, L. (2000). Affinity purification of recombinant proteins fused to calmodulin or to calmodulin-binding peptides. *Methods Enzymol.* **326,** 340–362.
Waldo, G. S. (2003). Genetic screens and directed evolution for protein solubility. *Curr. Opin. Chem. Biol.* **7,** 33–38.
Waldo, G. S., Standish, B. M., Berendzen, J., and Terwilliger, T. C. (1999). Rapid protein-folding assay using green fluorescent protein. *Nat. Biotechnol.* **17,** 691–695.
Wang, K. H., Oakes, E. S., Sauer, R. T., and Baker, T. A. (2008). Tuning the strength of a bacterial N-end rule degradation signal. *J. Biol. Chem.* **283,** 24600–24607.
Waugh, D. S. (2005). Making the most of affinity tags. *Trends Biotechnol.* **23,** 316–320.
Yokoyama, S. (2003). Protein expression systems for structural genomics and proteomics. *Curr. Opin. Chem. Biol.* **7,** 39–43.
Zhang, Y. B., Howitt, J., McCorkle, S., Lawrence, P., Springer, K., and Freimuth, P. (2004). Protein aggregation during overexpression limited by peptide extensions with large net negative charge. *Protein Expr. Purif.* **36,** 207–216.
Zhou, P., Lugovskoy, A. A., and Wagner, G. (2001). A solubility-enhancement tag (SET) for NMR studies of poorly behaving proteins. *J. Biomol. NMR* **20,** 11–14.