Title: To fuse or not to fuse: What is your purpose?

Authors: Mark R. Bell, Mark J. Engleka, Asim Malik, and James E. Strickler

Affiliation: LifeSensors, Inc., 271 Great Valley Parkway, Malvern, PA 19083

Corresponding author:

James E. Strickler

LifeSensors, Inc.

271 Great Valley Parkway

Malvern, PA 19355

e-mail: strickler@lifesensors.com

phone: 610-644-8845

fax: 610-644-8616

Running title:

Protein fusion technology

Total number of pages: 33

1-With apologies to William Shakespeare
Abstract

Since the dawn of time, or at least the dawn of recombinant DNA technology (which for many of today’s scientists is the same thing), investigators have been cloning and expressing heterologous proteins in a variety of different cells for a variety of different reasons. These range from cell biological studies looking at protein-protein interactions, post-translational modifications, and regulation, to lab-scale production in support of biochemical, biophysical and structural studies, to large scale production of potential biotherapeutics. In parallel, fusion-tag technology has grown-up to facilitate micro-scale purification (pull-downs), protein visualization (epitope tags), enhanced expression and solubility (protein partners, e.g. GST, MBP, TRX, SUMO), and generic purification (e.g. His-tags, streptag, FLAG™-tag). Frequently, these latter two goals are combined in a single fusion partner. In this review, we examine the most commonly used fusion methodologies from the perspective of the ultimate use of the tagged protein. That is, what are the most commonly used fusion partners for pull-downs, for structural studies, for production of active proteins, or for large-scale purification? What are the advantages and limitations of each? This review is not meant to be exhaustive and the approach undoubtedly reflects the experiences and interests of the authors. For the sake of brevity, we have largely ignored epitope tags although they receive wide use in cell biology for immunoprecipitation. Finally, in the spirit of full disclosure, the authors are employed by LifeSensors, Inc. which developed and markets a line of expression vectors based on the use of SUMO as a fusion partner.

Keywords: Fusion tag; Protein expression; GST; MBP;SUMO;TRX; His6
Introduction

Advances in genomics, proteomics, and bioinformatics over the last thirty years have dramatically increased the use of recombinant DNA as a way to study proteins of interest for a variety of applications. Combined with affinity tagging, recombinant DNA techniques allow for the identification, modification, production, isolation, and purification of proteins from a range of host systems, including *E. coli*, yeast, plant, insect, and mammalian cell lines. On the other hand, production of recombinant proteins routinely encounters problems, including the formation of inclusion bodies, incorrect protein conformation, toxicity to the host cell, or low protein yield. These issues are most often addressed by changing expression hosts or through fusion of the protein of interest (POI) to a carrier protein (fusion tag). Located at either the N- or C- terminus of the protein of interest, fusion tags can improve protein solubility, achieve native protein folding, and increase total yield by improving expression and decreasing degradation. Fusion tags may be used in tandem with affinity tags and other markers to improve detection, allow for protein secretion, and achieve greater total yield.

There are several published reviews on both affinity and fusion tags from the past several years (1; 2). While these reviews do an excellent job of describing the many tags and tag removal systems currently available, it can be difficult to determine which tags are the best candidates for specific applications. For example, it is estimated that 20-40% of eukaryotic proteins cannot be expressed in soluble form in prokaryotic hosts(3). Given the variability of protein structures, many tags may have similar issues. This review, therefore, focuses on tags that are utilized in specific protein applications, including protein-protein interaction ‘pull-down’ assays, structure determination, e.g. X-
ray crystallography, control and maintainence of protein functionality, and large scale manufacturing. While this list is by no means exhaustive, we hope to provide insight on the prominent tags used for these applications.

**Protein-protein interaction ‘pull-down’ assay**

Proteins do not work in isolation, but instead interact in complex networks. When studying any one protein, the isolation of other proteins in its complex can have several uses, including the purification of one or more of the binding partners, or the identification of unknown binding partners. Protein complex immunoprecipitation (Co-IP) is a technique in which an antibody is bound to a known target protein, allowing this protein and other proteins that are bound to it to be precipitated, or ‘pulled-down’, out of solution and analyzed.

While effective, one of the problems with this technique is the difficulty in generating specific antibodies to the target protein. A solution is to clone the DNA of the target protein into an expression vector containing a fusion tag at either terminus of the protein. Depending on the tag, either affinity chromatography or an antibody can be used for capture of the complex. The use of affinity chromatography drastically speeds up the process of protein isolation and identification, and allows the same purification process to be used repeatedly. Additionally, this system can be used to increase the expression of the target protein beyond endogenous levels, potentially allowing more complete pull-down of the protein complex and providing greater amounts of specific bound proteins (4; 5).

The most common fusion tag used in pull-down assays is glutathione-S-transferase (GST). A 26 kDa protein from the parasitic helminth *Schistosoma japonicum*,
GST binds with high affinity to glutathione (6). When used as a fusion tag, GST can increase protein yield by allowing efficient initiation of translation (7). GST has been used in a wide range of cell types, including *E. coli* (8-10), yeast (11; 12), plant (13; 14), insect (15; 16), and mammalian cells (17; 18). For purification, the GST-protein fusion is bound to glutathione immobilized to a solid support such as agarose beads or magnetic particles.. The fusion construct is eluted by the addition of 10 mM reduced glutathione. The target and associated proteins can be analyzed by standard methods such as SDS-PAGE or western blotting (19). GST has been used as both an N- and C-terminal tag, and in many commercially available systems, a protease cleavage site is encoded between GST and the target protein, allowing removal of GST after purification.

One of the problems that can occur with GST-based pull down assays is the solubility of the binding protein. Specifically, proteins that are either highly hydrophobic or larger than 100 kDA tend to form insoluble aggregates and inclusion bodies when tagged with GST, rendering them inactive. To correct for this, detergents such as Triton X and CHAPS are often used in the purification process to enhance the solubility of the fusion complex. If the detergents disrupt the biological activity of the binding protein, a high salt buffer can also be used to encourage solubility (20).

Another issue with GST is its propensity to dimerize. Native GST exists as a homodimer, and when fused with a target protein that can also oligomerize, the resulting fusion can form large complexes that are not easily eluted from the bound glutathione resin (21). Strategies to prevent dimerization include modification of salt or pH, or the addition of a strong reducing agent such as dithiothreitol (DTT). Another tactic is to promote elution by the addition of extra free glutathione, or by switching the
elution agent to S-butylglutathione, which has a 25-fold higher affinity for GST than does glutathione (22). As in the case of insolubility above, several of these solutions may disrupt the functionality of the tagged protein, and some researchers have suggested that GST is not suitable for pull down assays of proteins that are known to oligomerize (7; 23).

While the most common, GST is not the only tag that is used for pull-down assays. Technically, any affinity tag that can be fused to the target protein will work, and as the most widely used affinity tag in general, it comes as little surprise that polyhistidine tags (usually hexahistidine or His6) are also popular for pull-down assays. Although a His6 tag does not offer increased expression or solubility levels, it is small (0.84 kDa), immunogenically inactive, and doesn't dimerize. Like GST, the His6 tag may be attached to the target protein at either the N- or C-terminus, and most proteins are functional with the tag attached. Purification of His6-tagged fusions involves immobilized metal-affinity chromatography (IMAC), as the negatively charged histidine binds to the positively charged metal ions, most commonly Ni2+ (24). The fusion construct can be eluted with an imidazole gradient (either stepwise or linear) (25). One disadvantage of imidazole elution is the observation that high imidazole concentrations have been found to remove metal ions from a variety of proteins leaving them inactive and possibly altering the nature of their protein-protein interactions (26). Beyond both GST and His6, a number of additional tags are used for pull-down assays, including Strep-tag (27), and Fc-fusions (28), although these are seen in much lower overall numbers.
Tags for structural studies

Generation of good quality protein for structural studies, such as NMR spectroscopy, X-ray crystallography, and cryo-electron microscopy, places stringent demands on protein production. These approaches require multi-milligram quantities of protein at high purity and production techniques that minimize the use of detergents, chaotropes and reducing agents that might alter the final structure of the protein or prevent the formation of good quality diffracting crystals. While incorporating fusion tags can overcome some of these challenges by increasing yield, enhancing folding and streamlining purification, they can also create new obstacles. Multi-domain fusion proteins joined by a flexible linker may be less likely to form well-ordered, diffracting crystals or be too large for NMR studies. Strategies that require tag removal introduce challenges including optimization of cleavage conditions, added costs of proteases for tag removal, and failure to recover soluble or structurally intact protein after tag removal. On balance however, the advantages of using fusion tags for producing proteins for structural studies outweigh the disadvantages, and fusion tags have been widely adopted as the method of choice for protein production for structural study. Over 75% percent of proteins produced for crystallization are expressed as fusion constructs (29).

By far, His\(_6\) tag is the most common fusion partner, being used in 60% of crystallographic studies (30). The His\(_6\)-tag's small size does not generally interfere with crystallization and its utility in nickel affinity chromatography facilitates simple and cost-effective purification at the multi-milligram scale (31). His\(_6\)-tags do have certain disadvantages, however, as they are variably gluconoylated in \textit{E. coli} (32), which can lead to heterogeneity that is not conducive to crystal formation. While His\(_6\)-tags have
clearly been effective in structural studies, the Structural Genomics Center has estimated that up to 50% of all prokaryotic proteins are insoluble when expressed in *E. coli* with a His-tag (33; 34), and additional studies suggest that this number is higher for eukaryotic proteins (35-37).

Many large fusion tags, such as maltose binding protein (MBP) (38), glutathione S-transferase (GST) (39), thioredoxin (40), and small ubiquitin-like modifier (SUMO) (41), enhance solubility and, either directly or in combination with small affinity tags, simplify purification. Therefore, expression of target proteins as fusions with these tags is often used as a rescue strategy for difficult to express proteins (cf. (42)). In most cases, tags are proteolytically removed prior to crystallization by engineering endoprotease cleavage sites between the fusion tag and the protein of interest. The high homogeneity necessary for crystallization trials requires minimal non-specific or variable cleavage during tag removal. Thrombin, enterokinase (enteropeptidase) and Factor Xa, commonly used for tag removal, have historically shown spurious cleavage at sites distinct from the engineered site (43). In contrast, the viral proteases, tobacco etch virus (TEV) or human rhinovirus 3C protease, have a lower turnover, which results in fewer catalytic events at sub optimal cleavage sites and thus greater specificity (43). However, the lower turnover requires that large quantities of protease are required for tag removal and increases costs of protein production particularly at the multi-milligram scale. While limited to removal of SUMO tags, SUMO proteases provide both high specificity and high efficiency (44). SUMO proteases specifically recognize the tertiary structure of the SUMO tag, a structure not found elsewhere in the proteome, rather than a linear peptide sequence, and cleave precisely at the C-terminus of the tag.
Furthermore, SUMO protease has a $k_{\text{cat}}$ ~25-fold higher than TEV making the SUMO tag/SUMO protease system an efficient and cost-effective option for structural work (45).

A growing number of cases are being documented where crystallization is performed with the intact fusion protein (reviewed in Smyth (3) and Moon (46). Successful crystals have been described for fusions with MBP (47-49), GST (50-53), Thioredoxin (54; 55), Lysozyme (56-58) and SUMO (59). Leaving the fusion tag in place presents a number of advantages in so-called ‘carrier-driven’ crystallization (60-62), particularly when the protein of interest is small enough to occupy available space between the neighboring tag molecules in the crystal lattice. High resolution structures of MBP (63-65), GST (66), TRX (67; 68) and SUMO (69; 70) have been determined. The tags provide surfaces favorable to crystal lattice formation and by extension, the conditions for crystallization of the tag may translate to conditions for successful crystallization of the fusion protein. In addition, the structures of the tags can be used as search models to solve the crystallographic phase problem by molecular replacement methods. Examples of protein structures determined with or without tag removal are shown in Figure 1.

GST has a number of properties that make it a good candidate for carrier-driven crystallization (39). GST’s hydrophilic surface can improve the solubility of the protein of interest, GST fusions can be readily purified via glutathione resins, and finally, the structure of recombinant GST is known (66). Several structures of small peptides and protein regulatory domains have been determined as GST fusions, including gp41 from HIV (60), the C-terminal fibrinogen gamma chain (53), the ankryin-binding domain of α-
Na/K ATPase (52), acute myelogenous leukemia-1 nuclear matrix targeting sequence (AML-1 NMTS) (51), and DNA replication-related element-binding factor (DREF) (50). A further set of proteins has been crystallized but no structures have been reported, presumably because the fused fragment is disordered. The success of carrier-driven GST appears to be limited to date to protein fragments less than 100 amino acids (62). The use of GST presents added drawbacks in that it does not improve solubility in all cases (71) and it forms dimers (72; 73), which can lead to aggregation of certain targets.

MBP has also been used successfully in carrier-driven crystallization (3; 46) Like GST, MBP confers increased solubility to its fusion partner (71) and its crystal structure has been solved (63-65). In addition, the C-terminus of MBP forms a solvent accessible \( \alpha \)-helix, which provides a rigid support for linking the protein of interest (3). Although MBP provides a strategy for affinity purification by amylose affinity chromatography, MBP-fusion proteins often fail to bind to the amylose resin in practice (31; 74; 75). Furthermore, MBP adopts different conformations in the presence or absence of bound maltose (63-65), so partial occupancy of maltose can lead to heterogeneity in the final product, which can be detrimental to crystal formation. The addition of excess maltose may be required under certain purification schemes and crystallization trials. To circumvent these issues, MBP is frequently used in conjunction with a His\(_6\)-tag for purification. To date, MBP has shown the highest degree of success as a carrier protein for protein structure determination with over 25 structures determined as MBP fusions (46). Design of a short, rigid linker between the fusion tag and the protein of interest is critical to the success of carrier-driven crystallization. A variety of linkers and N-terminal
truncations of the target protein should be tried. Linkers too long will result in excessive flexibility, whereas linkers too short may influence the target protein structure.

Additional tags have been incorporated into successful crystals, including thioredoxin (54; 55), lysozyme (56; 58) and SUMO (76), but to a lesser extent than MBP or GST. Like MBP, the C-terminus of thioredoxin ends in a rigid, solvent accessible α-helix (55). Much of thioredoxin's polar surface can form crystal contacts and thioredoxin readily crystallizes. Despite these features, only two crystals and one solved structure have been reported using thioredoxin as a fusion partner (54; 55). Rather than as a traditional solubility or affinity tag, lysozyme was used to replace an unstructured loop in the β2-adrenergic G-protein coupled receptor (56; 58). Lysozyme provided an increased polar surface conducive to lattice formation. Multiple structures of SUMO fusions have been placed in the NCBI structure database. The SUMO fusion tag provides an interesting option for carrier-driven crystallization. As discussed above, SUMO improves yield and solubility across a wide range of targets (44; 45; 77-81) and its structure has been solved to high resolution (69; 70). Most importantly, the SUMOstar system is amenable to expression in eukaryotic hosts, allowing production of proteins for structural study that require expression in a eukaryotic host to produce proper folding or desired eukaryotic post-translational modifications (41; 78; 82).

**Tags for functional activity**

The need to generate functionally active proteins is a necessity of many studies, but is especially important when the protein in question is a potential therapeutic. Conformational characteristics, including proper folding and solubility, are an essential component of functionally active proteins, and these can be improved by the presence
of fusion tags (83). On the other hand, the generation of a native N-terminus is also critical for functional activity, particularly among cytokines, small peptides, and cytotoxic proteins, presenting additional challenges for tag use.

Cytokines, including interleukins (e.g. IL-6, IL-8), interferons (e.g. hIFN-γ), colony stimulating factors (G-CSF, M-CSF and GM-CSF), and hematopoietic factors such as erythropoeitin (EPO) and thromobopoeitin (TPO) are of interest for their immunomodulatory effects and therapeutic potential. A range of tags have been used to express cytokines, with varying levels of success. GST and thioredoxin have performed inconsistently to solubilize these types of proteins (1; 84-86). MBP and NusA are found to display solubility enhancing properties and increases expression levels of the target protein, perhaps due to their size (84; 87). In fact, NusA has been used to successfully express and purify cytokine homologs with the IL8-like fold (85). Removal of these tags requires endoproteases that leave additional amino acid residues on the N-terminus of the processed protein(1; 85), which can affect the structure and/or function of the protein (88). A tag that has seen more consistent success is SUMO. The conformation-specific activity of SUMO protease ensures that cleavage occurs precisely at the N-terminus of the target protein, and SUMO has been shown to produce an array of functionally active cytokines. IL-1β and IL-8 were expressed at high levels using SUMO, and both were biologically active. (88; 89) Recently, IFN-γ was generated in a functionally active form following SUMO expression and purification, although it was expressed as an inclusion body previously. (90) TNF-α was also generated as a mature and active product for use in drug development assays (91). Finally, the activity of chemokines can be drastically altered by truncation or extension at the N-terminus (see
for instance (92; 93)). In an extensive study, Lu et al. used a SUMO-tag to express and purify 15 chemokines in active form (94). In their system, SUMO fusion did not help with solubility, but was essential for producing the proteins with the authentic N-terminus. They used both the standard His$_6$-SUMO-tag and, in an interesting variation on the theme, they also used a tandem fusion; His$_6$-thioredoxin-SUMO, to promote refolding of the chemokines in the absence of a redox buffer.

Antimicrobial peptides are studied for their roles in physiology and potential as therapeutics. The main peptide families of interest, defensins and cathelicidins, are synthesized as precursor-proteins that are proteolytically cleaved to produce mature peptides (95; 96). These precursors protect host cells from the cytotoxic effects of the mature peptides (97; 98), and fusion tags are used to mimic these precursors, preventing cytotoxicity and successfully generating functional peptides.

Similar to expression work in cytokines, tags used for the production of antimicrobial peptides have encountered mixed results (see review by Li (99)). Once again GST performs inconsistently, as in a number of cases the tag failed to protect against proteolytic cleavage of the precursor peptide (100-102). This may be due to the large size of GST (28 kDa), especially in relation to the small antimicrobial peptides. MBP has been used for the production of Human β-defensin 25 (hBD25) and Human β-defensin 28 (hBD28), but both required refolding steps to recover fusion protein from aggregates in purification (98; 103).

Additional tags have been more consistently successful in expressing antimicrobial peptides. Thioredoxin has been used to achieve high-yields of precursor peptides in the cytoplasm, perhaps because of its smaller size (11.8 kDa) (40). SUMO,
also small in size (11.2 kDa), has also been used successfully in generating defensins and cathelicidins. In the production of Human β-Defensin-4, 166 mg per 1 L fermentation was obtained after purification (104). LL-37, regarded as the only cathelicidin-derived antimicrobial peptide found in humans(105), has been produced in conjunction with thioredoxin in a dual-tag expression (104; 106). Other functionally active proteins produced with SUMO include the antibacterial peptide CM4 (ABP-CM4) (107), the PnTx3-4 toxin isolated from spider venom (108), and the antitumor-analgesic peptide (AGAP) purified from scorpion venom (108; 109).

**Self-cleaving affinity tags**

The use of self-splicing tags (inteins) for the purification of recombinant proteins was first described in 1997 (110; 111) by a group working at New England Biolabs. Their work gave rise to NEB’s IMPACT (Intein-mediated purification with an affinity chitin-binding tag) system which remains the most frequently used intein purification methodology. The technology revolves around the use of a yeast-derived protein splicing element coupled to a bacterial chitin-binding affinity purification tag. The expressed protein is captured on a chitin column, washed free of host derived proteins, and then released by induction of cleavage of the intein. Mechanistically, the intein spontaneously undergoes an S-N acyl shift at its N-terminal cysteine residue to form a thioester bond with the protein of interest. This thioester is then readily cleaved by a number of small molecules such as 2-mercaptoethanol, 2-mercaptoethansulfonic acid, or dithiothreitol. The protein of interest is released as a thioester of the small molecule which, depending on the intended use, can be used to carry-out C-terminal modification or simply hydrolyzed to generate a C-terminal carboxylate. While this technology has
worked well for a large number of proteins from small anti-bacterial toxins (112-114) to antibody fragments (115; 116) to human choline acetyltransferase (117) it has not become a widely used method for general protein purification. Rather it has found its widest use to generate proteins that can subsequently be used for C-terminal modification via the thioester generated during cleavage. This includes expressed protein ligation (EPL) or native chemical ligation (NCL). In this technique, two disparate peptide segments are joined together through the ligation of a C-terminal peptide containing an N-terminal cysteine residue to the N-terminal peptide with a C-terminal thioester. Attack of the sulfhydryl of the cysteine on the thioester results in a thioester linkage between the two peptides. In a reverse of the intein reaction, the thioester undergoes an N to S migration generating a standard amide bond. This methodology is frequently used to introduce non-natural amino acids into a protein when one of the two partners is produced synthetically rather than biosynthetically. One interesting use was the joining of two peptides, one of which was expressed as a heavy isotope labeled protein (\(^{13}\)C and \(^{15}\)N) in order to produce a more highly refined NMR structure of two protein domains which identified an extended interaction interface between the two proteins previously thought to behave independently (118). In another iteration of the methodology, the thioester is used to introduce small molecules such as fluorophores at the C-terminus of the protein of interest to generate enzyme substrates. For instance, a small cottage industry has grown up around the C-terminal modification of ubiquitin (Ub) and ubiquitin-like proteins (Ubls) as substrates/inhibitors of Ubl-deconjugating enzymes (deubiquitylases (DUBs), desumoylases, deNEDDylases, etc.) These include inhibitors such as ubiquitin-aldehyde (119) and ubiquitin-vinylsulfone (120) among others (121-
123) and substrates such as ubiquitin-amidomethylcoumarin (AMC) (124), ubiquitin-rhodamine 110 (125), and ubiquitin-amidomethyl-luciferin (AML) (126).

There are two principle disadvantages to intein technology. First, the binding capacity of chitin-agarose is very low. Typically, the capacity of chitin-agarose is 1-2 mg of recombinant protein/mL of matrix. Contrast this to Ni\(^{+2}\)-IMAC or IEX-sepharose columns with capacities of 40-80 mg/mL of resin. As a consequence, one must use relatively large columns to insure complete capture of the expressed protein. While this is generally not a major concern at lab-scale, it decreases the desirability of this methodology at manufacturing scales. Second, the cleavage reaction is very slow. On column cleavage is usually carried out for 16 hr at room temperature (110) or to up to 5 days at 4°C.(114; 127) Long incubation times and elevated temperatures can lead to increased non-specific proteolysis by contaminating proteases or increased risk of protein denaturation. Finally, the use of high levels of reducing agents (30 to 100 mM) can be detrimental to the recovery of disulfide linked proteins leading to the subsequent need for refolding.

To overcome some of these disadvantages, the Wood lab at Princeton has been actively engaged in modification of the technology. Their major focus has been on replacing the chitin-binding domain with other types of purification entities to avoid the low capacity of chitin-agarose as well as replace column chromatography with mechanical methods of primary separation (see review by Banki and Wood (128)). Briefly, these new methods either require a precipitation step mediated by the controlled aggregation of elastin-like polypeptide tags attached to the intein (129-132) or adsorption of phasin-tagged inteins to polyhydroxybutyrate (PHB) nanogranules.
produced in specially engineered strains of *E. coli* (133-136). None of these methods have been tried extensively outside of the originator’s lab. In a separate attempt to improve purification capacity of the intein technology Wang, et al. (137) added either a His$_6$-Ubiquitin-tag or a His$_6$-SUMO-tag at the N-terminus of the intein with a C-terminal target protein. They could then use Ni$^{2+}$-IMAC for purification and then induce intein cleavage to release the protein of interest without the use of a protease. As an added bonus, the SUMO- and ubiquitin-tags enhanced expression levels compared to the intein alone.

None of these approaches address the issue of slow cleavage rate. In order to tackle this problem a number of groups have moved away from inteins to utilize proteases themselves as part of the fusion tag. These include a His$_6$-tagged cysteine protease domain from *Vibrio cholerae* (138) and a His$_6$-tagged catalytic sortase domain from *Staphylococcus aureus* (139). In both cases, enzymatic cleavage is initiated on-column through the addition of a small molecule, inositol hexakisphosphate or Ca$^{2+}$, respectively. Again, these methodologies have not gained wide acceptance outside of the originators’ laboratories. Perhaps the most well-developed of these approaches is that developed by Bryan’s group (140) and marketed by BioRad, Inc. as the Profinity system. In this instance, the affinity tag is the prodomain of the bacterial enzyme subtilisin. An engineered form of subtilisin, which binds with high affinity to this prodomain, is coupled to agarose for affinity purification. As with the two protease tags described above, cleavage is initiated by a small molecule, in this case fluoride. Although less potent, the enzyme can be activated by other halide ions limiting their use in buffers. Another drawback of this self-cleaving system is the requirement for separate
expression, purification, and conjugation of the mutant subtilisin which could make the system cost-prohibitive at large scale. With the other two protease-based systems, the enzyme is expressed as part of the fusion protein insuring that sufficient protease is always present.

Large scale manufacturing

Large scale protein manufacturing is often quite different from small scale research production. While protein production for research and development might typically involve a few liters of culture or less to obtain the desired amount of purified product, commercial production often requires bioreactors capable of holding several thousand liters each. The challenges of high quantity and low cost bioprocessing mean that affinity and fusion tags are not commonly used, due in part to the extra time and cost associated with tag removal and the subsequent purification steps at such large scales. Therefore, when a tag system is used, it must present clear advantages in production or to the therapeutic itself.

Fc-fusion proteins are the best-known example of a fusion tag being used in large scale manufacturing. First created in 1989 as a potential AIDS therapeutic (141), several commercially available drugs are based on Fc-fusions. These include belatacept and alefacept for organ rejection, abatacept and etanercept for rheumatoid arthritis, and afibercept for macular degeneration, among others (142). Fc-fusion proteins consist of a protein of interest linked to a fragment crystallization (Fc) domain of an immunoglobulin, which is the segment of the heavy chain furthest from the antigen binding site. At approximately 250 amino acids in size, the Fc domain can be attached to either terminus of the protein of interest. Currently all commercial therapeutics utilize...
the Fc domain from human IgG1, although other options, such as IgG3, IgA and IgM are currently being explored (143).

As an affinity tag, the Fc domain binds with high affinity to Protein A, a surface protein originally isolated from *Staphylococcus aureus* that is often bound to a stationary phase resin. Elution can be accomplished via a pH gradient or by commercial detergents, and the combination of high affinity and simple elution allows the Fc system to be cost-effective at scales large enough for manufacturing. This purification process is only an ancillary benefit; the addition of an Fc domain to a therapeutic compound provides the fusion with several beneficial pharmacological properties. Most prominently, the addition of an Fc domain increases the serum half-life of a therapeutic, increasing its activity. This is accomplished by the Fc domain binding with the neonatal Fc-receptor (FcRn), which plays a role in protein recycling by preventing lysosomal degradation (144-146). The Fc domain can also interact with Fc-receptors present on certain cells in the immune system, such as B-lymphocytes and natural killer cells, an ability which is important to certain types of oncology therapeutics (147; 148). Finally, the addition of an Fc domain can improve the solubility of the overall fusion due to the fact that it folds independently of its fusion partner (149).

Since Fc-domains are essentially a functional section of the final therapeutic, no cleavage step is necessary. In order for a fusion system that does undergo tag removal to be considered for larger scale protein expression, additional benefits must be present to offset the additional purification costs. One such system is SUMO. As previously mentioned, SUMO can increase both the yield and solubility of its fusion partner and cleavage of the tag is performed by a specific SUMO protease which
leaves no additional amino acids behind that can disrupt therapeutic function. When combined with a His$_6$ or other affinity tag on both the SUMO tag and the protease, a simple purification scheme based on IMAC chromatography can be used, and current SUMO tags can be used in several expression systems, including *E. coli*, *Pichia pastoris*, and mammalian cells such as HEK and CHO.(45) Recent examples include the production of the antiviral protein cyanovirin-N (CVN), a microbicide with anti-HIV effects. The expression of CVN with a His$_6$-SUMO tag lead to a soluble protein level in *E. coli* of over 30% of the total soluble protein using a 30L bioreactor (150). Additionally, SUMO was used in the production of subunit antigens to anthrax and swine flu in a test sponsored by DARPA that required both rapid antigen production and cost-per-dose thresholds. At the 30L scale, the maximum product output was 14 g/L, and the cost-per-dose was well under the requirements.
References


63. Spurlino JC, Lu GY, Quiocho FA (1991) The 2.3-A resolution structure of the maltose- or maltodextrin-binding protein, a primary receptor of bacterial active transport and chemotaxis. J Biol Chem 266:5202-5219. PMID: 2002054 {Medline}


147. Beck A, Reichert JM (2011) Therapeutic Fc-fusion proteins and peptides as successful alternatives to antibodies. MAbs 3:415-416. PMID: 21785279 {Medline}


Figure 1. Structures of two proteins expressed as SUMO-fusions in E. coli. Panel A shows the structure of a peptidyl-prolyl cis-trans isomerase from Burkholderia pseudomallei bound to 8-deethyl-8-[but-3-enyl]-ascomycin (solid surface). The protein was crystallized and the structure determined to 1.9Å resolution with SUMO still attached (labeled). (PDB ID 3UQB, Fox III, Abendroth, Staker, and Stewart, Seattle Structural Genomics Center for Infectious Disease, deposited Nov. 2011). Panel B shows the structure of the human ubiquitin E3 ligase Parkin. SUMO was enzymatically removed prior to crystallization. The structure was determined at 2.0Å resolution (PDB ID 4I1H, Riley et al. [42]).