
Comparison of SUMO fusion technology with traditional gene fusion systems: Enhanced expression and solubility with SUMO

JEFFREY G. MARBLESTONE,¹ SUZANNE C. EDAVETTAL,² YITING LIM,²
PETER LIM,¹ XUN ZUO,² AND TAUSEEF R. BUTT²

¹Progenra Inc., Malvern, Pennsylvania 19355, USA

²LifeSensors Inc., Malvern, Pennsylvania 19355, USA

(RECEIVED August 26, 2005; FINAL REVISION October 12, 2005; ACCEPTED October 12, 2005)

Abstract

Despite the availability of numerous gene fusion systems, recombinant protein expression in *Escherichia coli* remains difficult. Establishing the best fusion partner for difficult-to-express proteins remains empirical. To determine which fusion tags are best suited for difficult-to-express proteins, a comparative analysis of the newly described SUMO fusion system with a variety of commonly used fusion systems was completed. For this study, three model proteins, enhanced green fluorescent protein (eGFP), matrix metalloprotease-13 (MMP13), and myostatin (growth differentiating factor-8, GDF8), were fused to the C termini of maltose-binding protein (MBP), glutathione S-transferase (GST), thioredoxin (TRX), NUS A, ubiquitin (Ub), and SUMO tags. These constructs were expressed in *E. coli* and evaluated for expression and solubility. As expected, the fusion tags varied in their ability to produce tractable quantities of soluble eGFP, MMP13, and GDF8. SUMO and NUS A fusions enhanced expression and solubility of recombinant proteins most dramatically. The ease at which SUMO and NUS A fusion tags were removed from their partner proteins was then determined. SUMO fusions are cleaved by the natural SUMO protease, while an AcTEV protease site had to be engineered between NUS A and its partner protein. A kinetic analysis showed that the SUMO and AcTEV proteases had similar K_M values, but SUMO protease had a 25-fold higher k_{cat} than AcTEV protease, indicating a more catalytically efficient enzyme. Taken together, these results demonstrate that SUMO is superior to commonly used fusion tags in enhancing expression and solubility with the distinction of generating recombinant protein with native sequences.

Keywords: fusion protein; protease; structural genomics; SUMO; ubiquitin-like protein

The lack of efficient methods to express structurally diverse proteins in *Escherichia coli* is a major obstacle in structural genomics. In fact, the Southeast Collaboratory

for Structural Genomics (SECSG) reports that only 22.9% of proteins they have expressed in *E. coli* have been soluble (1463 soluble proteins of 6397 expressed; as published on the SECSG Web site 08/11/2005). Numerous technological advancements have vastly improved recombinant protein expression in *E. coli*, including the development of strong promoters (Studier and Moffatt 1986), coexpression with chaperones and foldases (Ikura et al. 2002), and the use of protein fusions. Protein fusions have been particularly successful at enhancing the expression and solubility of recombinant proteins. Fusion systems are characterized by their ability to enhance protein expression, reduce proteolytic degradation of the recombinant protein, improve protein folding, solubility, and

Reprint requests to: Tauseef R. Butt, LifeSensors Inc., 271 Great Valley Parkway, Malvern, PA 19355, USA; e-mail: butt@lifesensors.com; fax: (610) 644-8616.

Abbreviations: DUB, deubiquitinating enzyme or ubiquitin specific protease/hydrolase; eGFP, enhanced green fluorescent protein; IPTG, isopropyl- β -D-thiogalactopyranoside; MBP, *E. coli* maltose-binding protein; Ni-NTA, nickel-nitrilotriacetic acid; PCR, polymerase chain reaction; SDS-PAGE, sodium dodecyl sulfate-polyacrylamide gel electrophoresis; Ub, ubiquitin; Ubl(s), ubiquitin-like protein(s); ULP, catalytic domain of Ulp1.

Article published online ahead of print. Article and publication date are at <http://www.proteinscience.org/cgi/doi/10.1110/ps.051812706>.

simplify purification and detection. A variety of structures have been used as fusion motifs, including maltose-binding protein (MBP), glutathione S-transferase (GST), thioredoxin (TRX), NUS A, ubiquitin (UB), and SUMO (Table 1; Pryor and Leiting 1997; Wang et al. 1999; De Marco et al. 2004). There is nothing in common among these fusion proteins in terms of molecular weight, structure, or function, with the exception of Ub and SUMO, which share a common structure (Bayer et al. 1998). As such, predicting which fusion tag will enhance the solubility of a difficult-to-express protein remains empirical. Some comparison studies have been completed; however, none have examined the ability of the newly described SUMO fusion system to enhance expression and solubility in comparison to other commonly used tags (Davis et al. 1999; Wang et al. 1999; De Marco et al. 2004).

SUMO (small ubiquitin-related modifier), an ~100-residue protein, modulates protein structure and function by covalent modification of target proteins in eukaryotes (Johnson and Blobel 1999; Melchior 2000; Tatham et al. 2001). The SUMO pathway is highly conserved in eukaryotes and notably absent in prokaryotes. Yeast has a single SUMO gene (*SMT3*), while three genes have been described in vertebrates (*SUMO-1*, *SUMO-2*, and *SUMO-3*) (Kawabe et al. 2000). The three human SUMOs are highly homologous, with human SUMO-1 sharing 50% sequence identity with human SUMO-2 and SUMO-3 (Muller et al. 1998), and human SUMO-2 and SUMO-3 sharing 87% sequence identity with each other (Melchior 2000). SUMO shares 47% sequence identity with human SUMO-1. Although the overall sequence identity between SUMOs and Ub is ~18%, structure determination reveals that they share a common three-dimensional structure that is characterized by a tightly packed globular fold with β -sheets wrapped around one α -helix (Bayer et al. 1998).

The conjugation of SUMO to target proteins is a highly regulated and dynamic process, similar to Ub. The enzymes involved in cleaving SUMO fusions, the SUMO proteases, have been extensively studied. Hochstrasser and

Li demonstrated that the yeast SUMO proteases, Ulp1 and Ulp2, remove SUMO from proteins and play a role in progression through the G₂/M phase and recovery of cells from check point arrest, respectively (Li and Hochstrasser 1999, 2000). Ulp1 and Ulp2 cleave the C-termini of SUMO (-GGATY) to form a mature SUMO (-GG) and also deconjugate it from the side chains of lysines within modified proteins. The sequence similarity of the two enzymes is restricted to a 200-amino-acid sequence called ULP domain that contains the catalytically active region. The three-dimensional structure of the ULP domain from Ulp1 has been determined in a binary complex with SUMO (Mossessova and Lima 2000). It is interesting to note that SUMO proteases are not related to the deubiquinating enzymes (DUBs), but are distantly related to adenoviral processing protease (Li and Hochstrasser 1999, 2000).

Recently, a SUMO fusion system that facilitates efficient expression of recombinant proteins in *E. coli* has been described (Malakhov et al. 2004). Several proteins, including severe acute respiratory syndrome coronavirus (SARS-CoV) 3CL protease, nucleocapsid, and membrane proteins, have been recombinantly expressed using the SUMO fusion system (Zuo et al. 2005b). The SUMO fusion tag has led to enhanced expression and solubility. A hexahistidine SUMO fusion construct has been shown to enhance expression and facilitate purification with Ni-NTA chromatography (Zuo et al. 2005a).

One distinguishing feature of the SUMO fusion system is the ability of its associated SUMO protease to cleave a variety of fusion partners with remarkable fidelity and efficiency (Malakhov et al. 2004). Traditional gene fusion systems require engineered cleavage sites, which are recognized by the proteases and are positioned between the fusion tag and the protein target. Proteases that have been used to cleave fusion tags include tobacco etch virus (TEV) protease (Carrington et al. 1989), factor Xa, or thrombin protease (Jenny et al. 2003). A major drawback to the use of engineered cleavage sites and traditional proteases is the generation of non-native N-terminal amino acids. Many structural and therapeutic proteins require specific N-terminal amino acids for biological activity (e.g., chemokines). Cleavage by traditional proteases results in the retention of several amino acids, which are downstream from the cleavage site and required for protease recognition. For example, thrombin will cleave the sequence LVPRGS at the arginine residue, resulting in an N-terminal extension of the target protein by two amino acids (GS) (Jenny et al. 2003). Those proteins that require a specific N terminus for biological activity, half-life, or structural stability, will not be successfully expressed using gene fusions with traditional proteases. However, direct fusion of the recombinant protein to the C terminus of SUMO results in

Table 1. Sequence and size of tags

Tag	Residues	Cleavage site	Size (kDa)
Poly-His	6	—	0.84
Ubiquitin	76	ENLYFQ'GXX	8
SUMO	100	GG'XXX	11.5 ^a
Maltose Binding protein	396	ENLYFQ'GXX	40
Glutathione S-transferase	211	ENLYFQ'GXX	26
Thioredoxin	109	ENLYFQ'GXX	12
NUS A	495	ENLYFQ'GXX	55

^aSUMO migrates aberrantly at ~20 kDa in an SDS-polyacrylamide gel.

the production of protein with the desired N-terminal amino acid. In addition, when using traditional gene fusion systems, if the target protein or fusion tag contains the cleavage recognition sequence, the target protein will also be cleaved (e.g., erroneous cleavage of the NUS A tag has been observed when using Factor Xa) (Davis et al. 1999). The SUMO protease recognizes the tertiary structure of SUMO, and as such, does not cleave erroneously within the target protein.

Previous experience with the SUMO fusion system suggests that it represents a technological advancement in recombinant protein expression, as this system enhances expression and solubility and utilizes a highly specific protease capable of generating native N-terminal amino acids. The aim of this study is to provide a direct comparison of SUMO with other fusion systems to determine whether it truly represents such advancement. Three candidate proteins, enhanced green fluorescent protein (eGFP), and two previously described difficult-to-express proteins, matrix metalloprotease-13 (MMP13) and myostatin (growth differentiating factor-8, GDF8), were expressed as fusions with maltose-binding protein (MBP), glutathione S-transferase (GST), thioredoxin (TRX), NUS A, ubiquitin (Ub), and SUMO. These constructs were expressed in *E. coli* and evaluated for expression and solubility. In addition, the catalytic efficiency of the SUMO and commonly used AcTEV proteases were evaluated.

Results

Comparison of SUMO with other commonly used fusion tags

The ability of commonly used fusion tags to enhance protein expression and solubility was investigated using three candidate proteins (eGFP, MMP13, and GDF8) and six fusion tags (SUMO, Ub, MBP, GST, TRX, and NUS A). Vectors were transformed into *E. coli* and

cultures were induced for protein expression under the control of the T7 promoter with IPTG. Culture inductions were conducted at 20°C overnight for optimal expression levels and solubility. Cell lysates from uninduced (UI) and induced (I) cultures, plus the soluble (S) and insoluble (IB) fractions from the induced cultures were analyzed by SDS-PAGE (Figs. 1–3). There was a clear difference among the fusion tags with respect to their ability to enhance expression and solubility.

The impact of different gene fusions on the expression and solubility of eGFP following induction at 20°C overnight are shown in Figure 1. The His₆-eGFP construct was observed to have very low expression, while the SUMO-, TRX-, and NUS A-eGFP constructs had the greatest enhancement of soluble expression. The SUMO and NUS A fusions resulted in the best soluble expression of eGFP (~90% and 100% soluble, respectively). While TRX fusion resulted in greatly enhanced expression, only ~50% of the expressed protein was soluble. The Ub, MBP, and GST fusions enhanced eGFP expression equivocally, relative to each other.

The difficult-to-express MMP13 (residues 20–274) was then investigated for enhanced solubility and expression with the various gene fusions (Fig. 2). As expected, the His₆-MMP13 fusion did not generate any MMP13. SUMO, NUS A, and TRX fusions produced the greatest degree of expression enhancement, as was observed for eGFP. Also consistent with the results from the eGFP study, the SUMO and NUS A fusions resulted in enhanced soluble expression (~40% and 80% fusion MMP13 in the soluble fraction, respectively). In contrast to eGFP, the TRX-MMP13 fusion solely generated insoluble protein. The Ub-MMP13 fusion was observed to have a moderate amount of expression; however, the majority of the protein produced was insoluble (~90%). Again, the GST fusions only enhanced expression mildly, and the MBP fusion failed to enhance expression.

GDF8 (mature), also considered to be a difficult-to-express protein, was evaluated for enhanced expression

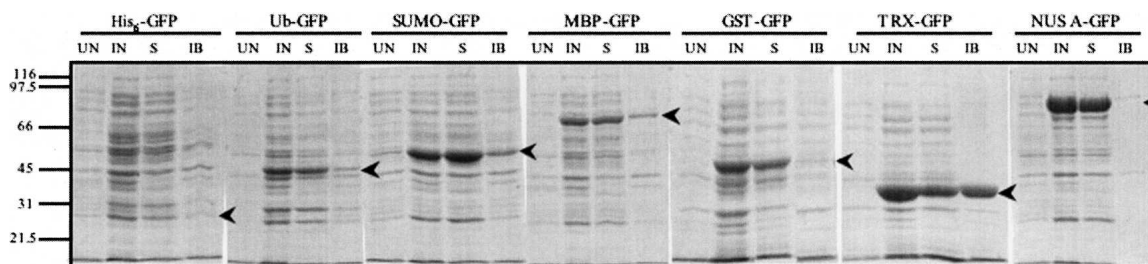


Figure 1. Expression of His₆-GFP and fusions with Ub, SUMO, MBP, GST, TRX, and NUS A proteins. *E. coli* grown in LB to an OD₆₀₀ = 0.6 at 37°C, induced with 1 mM IPTG and incubated at 20°C overnight. Protein fractions were resolved by SDS-PAGE and stained with Coomassie. Arrowheads indicate expected/observed positions of respective protein bands. (UN) Uninduced culture; (IN) induced culture; (S) soluble cellular lysate; (IB) inclusion body.

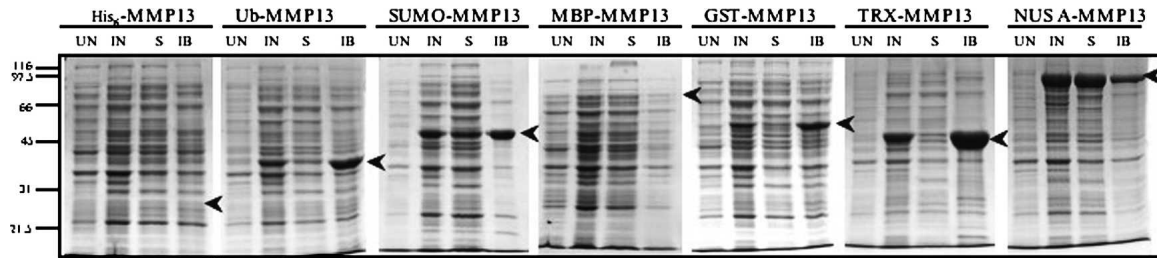


Figure 2. Expression of His₆-MMP13 (amino acids 20–274) and fusions with Ub, SUMO, MBP, GST, TRX, and NUS A proteins. *E. coli* grown in LB to an OD₆₀₀ = 0.6 at 37°C, induced with 1 mM IPTG and incubated at 20°C overnight. Protein fractions were resolved by SDS-PAGE and stained with Coomassie. Arrowheads indicate expected/observed positions of respective protein bands. (UN) Uninduced culture; (IN) induced culture; (S) soluble cellular lysate; (IB) inclusion body.

and solubility with the various gene fusions (Fig. 3). The Ub- and GST-GDF8 fusions seem to have enhanced expression levels, but the majority of the recombinant protein was found in the inclusion body, with little in the soluble fraction. Similarly, the TRX-GDF8 fusion yielded high levels of expression; however, this fusion was exclusively insoluble. The SUMO- and NUS A-GDF8 fusions were both observed to enhance expression at high levels, with the SUMO and NUS A fusions being the only tags that were able to preserve solubility. As was observed for the MMP13 and eGFP constructs, NUS A produced almost entirely soluble protein (~95%) with the SUMO fusion less successful at retaining recombinant protein solubility (~50%). The MBP and His₆ fusions were observed to have very little enhanced expression of GDF8.

Comparison of SUMO protease with TEV protease

The SUMO- and NUS A-tags produced the highest expression levels and solubility among the fusion proteins tested. A comparison of their respective proteases was then conducted to determine the catalytic efficiency of removing the fusion tag (Fig. 4; Table 2). The commonly used AcTEV protease was used to cleave the

NUS A tag, and an AcTEV recognition sequence (ENLYFQ'GXX) was engineered between the NUS A tag and the fusion partner. The kinetic parameters K_M and k_{cat} for AcTEV and SUMO protease were determined using the NUS A-eGFP and SUMO-eGFP substrates, respectively (Table 2). The generated data were fit to the Michaelis-Menton equation using nonlinear regression (R^2 values were equal to 0.97 and 0.99 for the SUMO protease and AcTEV plots, respectively) (Fig. 4). The apparent K_M of AcTEV for the NUS A-eGFP fusion (6.3 mM) is very similar to that of SUMO protease for the SUMO-eGFP fusion (3.3 mM), but its k_{cat} (0.028 sec⁻¹) is ~25-fold less than that of SUMO protease (0.782 sec⁻¹), resulting in an enzyme with ~25 times lower catalytic efficiency. The K_M and k_{cat} values obtained for AcTEV are not similar to what has been previously reported (Nallamsetty et al. 2004). This is most likely due to the choice of substrates, as previous studies have been completed using peptide substrates with canonical recognition sites. The SUMO protease does not recognize a linear sequence like AcTEV, and as such, a full fusion construct had to be utilized for comparison purposes. It is expected that lower K_M and higher k_{cat} values would be obtained for the peptide substrate than for the full fusion construct.

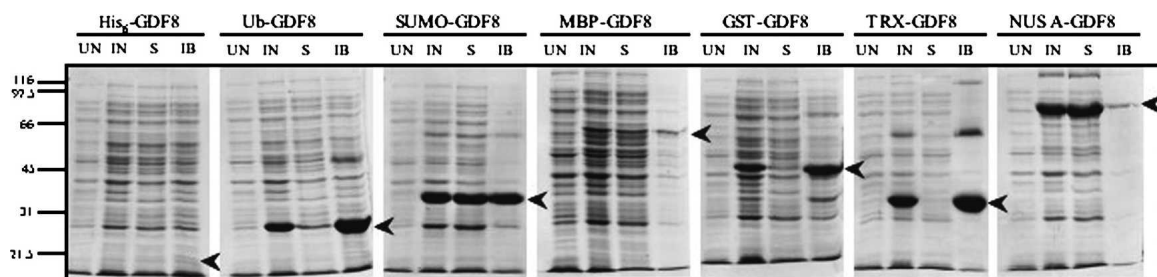


Figure 3. Expression of His₆-GDF8 and fusions with Ub, SUMO, MBP, GST, TRX, and NUS A proteins. *E. coli* was grown in LB to an OD₆₀₀ = 0.6 at 37°C, induced with 1 mM IPTG, and incubated at 20°C overnight. Protein fractions were resolved by SDS-PAGE and stained with Coomassie. Arrowheads indicate expected/observed positions of respective protein bands. (UN) Uninduced culture; (IN) induced culture; (S) soluble cellular lysate; (IB) inclusion body.

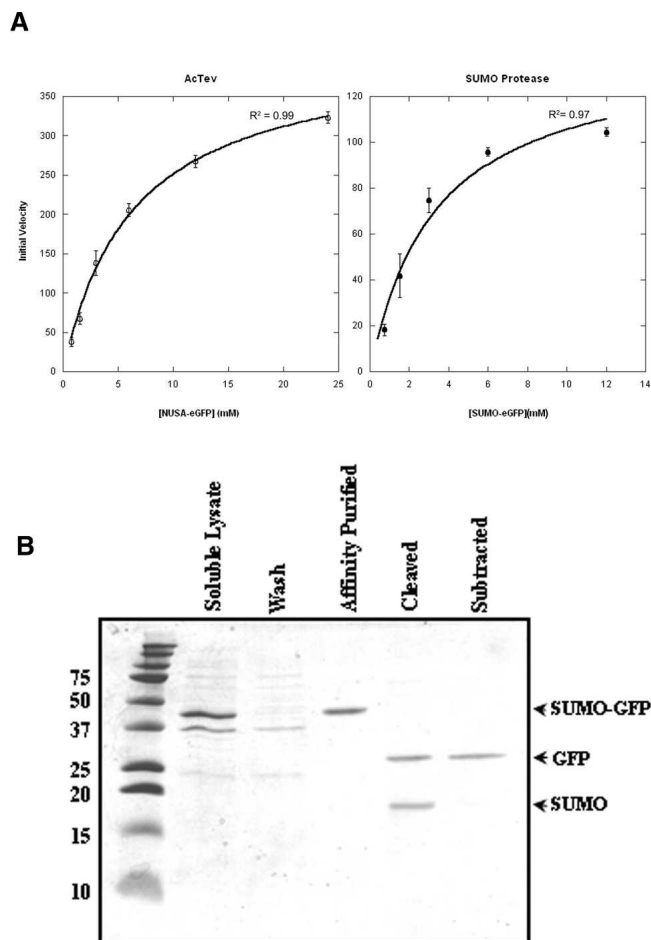


Figure 4. (A) Kinetic data for cleavage of SUMO-eGFP and NUSA-eGFP by SUMO and AcTEV proteases. Data represent the mean of three independent experiments, with error bars representing the range. Curve fitting to the Michaelis-Menten equation was completed using nonlinear regression (R^2 as shown on each graph), and used to calculate K_M and k_{cat} (Table 2). (B) An SDS gel depicting the purification of GFP using the SUMO fusion tag. Soluble cell lysate was passed over a Ni-NTA column, washed with 20 mM imidazole, and eluted with 300 mM imidazole (affinity purified). Cleavage with the SUMO protease was conducted under standard conditions, and the sample was passed over another Ni-NTA column to remove the SUMO protease and tag (subtracted). These various steps of a typical purification of SUMO-GFP were analyzed in a 15% SDS–polyacrylamide gel. Proteins were visualized using Coomassie Blue staining. Arrowheads indicate expected/observed positions of respective protein bands.

Discussion

The preferred host for heterologous protein expression is *E. coli*, as this host system provides a simple and inexpensive means to produce recombinant protein. Many obstacles are encountered, however, when using *E. coli* as a host system. Gene fusion technology has been very successful at overcoming most of these obstacles and has increased the success of heterologous expression in *E. coli*. Ideally, a fusion tag should enhance expression and

solubility of a wide variety of proteins. While several studies have compared the effectiveness of commonly used gene fusions in enhancing the expression and solubility of difficult-to-express proteins, none have included the newly described SUMO fusion system (Davis et al. 1999; Wang et al. 1999; De Marco et al. 2004).

Three candidate proteins were chosen for this study, eGFP, GDF8, and MMP13. eGFP was chosen due to its common use in biochemical laboratories and the ease at which it can be assayed. GDF8 is essential for proper regulation of human skeletal muscle mass (Lee and McPherron 2001). GDF8 was chosen, as recombinant expression in *E. coli* would greatly facilitate research efforts aimed at exploiting this potentially useful therapeutic target. Previous attempts to express GDF8 in *E. coli* have been hampered by poor yields, solubility, and lack of biological activity, and therefore, GDF8 is considered to be a difficult-to-express protein (Thomas et al. 2000; Taylor et al. 2001). Similarly, MMP13 is generally considered a potential therapeutic target and a difficult-to-express protein (Pathak et al. 1998; Hardern et al. 2000). Although only a limited number of proteins were assayed for this study, patterns do emerge as to which tags are best for enhanced expression and solubility, especially for the difficult-to-express proteins used. In general, His₆ tag afforded little to no enhanced expression. MBP was also observed to have little impact on expression, and did not enhance expression at all for the difficult-to-express proteins in this study (MMP13 and GDF8). Both the Ub and GST tags did provide mild enhancement of expression; however, the recombinant protein was mostly contained within the inclusion bodies. SUMO, TRX, and NUS A were the best tags at enhancing expression. The TRX tag did not enhance solubility for the difficult-to-express proteins, while both SUMO and NUS A had pronounced soluble fusion protein expression.

For all polypeptides, there is a competition between aggregation and folding that rely on similar molecular interactions. Fusion partners have been shown to act as solubility enhancers, although the exact mechanism by which they improve solubility has not been described. It is speculated that fusion partners are able to keep the target protein in solution long enough for it to undergo its natural folding process, otherwise the target protein aggregates before it has sufficient opportunity to fold. A soluble fusion partner that slows the aggregation process

Table 2. Kinetic parameters for SUMO protease and AcTEV protease

Enzyme	Substrate	K_M (mM)	k_{cat} (mM ⁻¹ s ⁻¹)
SUMO protease	SUMO-eGFP	3.34 ± 0.9	0.782 ± 0.17
AcTEV protease	NUS A-eGFP	6.33 ± 0.6	0.028 ± 0.002

inadvertently shifts the fusion protein to a folding pathway, increasing the amount of soluble protein. Fusion tags have also been hypothesized to enhance the solubility of the protein target by acting as a nucleus of folding ("molten globule hypothesis") (Creighton 1997; Englander 2000). This theory suggests that a fusion tag acts as a nucleation site for the folding of the target protein. Attachment of SUMO to the N terminus of a partner protein has been observed to promote correct folding and solubility. SUMO and Ub have highly homologous, rapidly folding structures (Khorasanizadeh et al. 1996). It may be that the tight, rapidly folding soluble structure of SUMO provides a nucleation site for the proper folding of C-terminally fused partner proteins. Indeed, expression of alternative configurations of the fusion (i.e., eGFP-SUMO) was not enhanced in *E. coli* (data not shown). The 55-kDa NUS A fusion tag also facilitates solubility of partner proteins. NUS A, a hydrophilic tag, was identified in a screen for *E. coli* proteins that have the highest potential for solubility when overexpressed (Davis et al. 1999). Protein expression levels are dependent on the stability of their mRNA, where degradation plays an important role controlling the levels of unstable transcripts (Arechaga et al. 2003). Both NUS A and SUMO were observed to enhance expression, but the role they play in stabilizing the mRNA transcript is unknown. Whereas NUS A levels of expression have been well documented, the novel SUMO tag has had modest published results. Table 3 represents a variety of fusion proteins that have been successfully expressed using the SUMO system. These SUMO fusion proteins have established a 5–25-fold increase in expression levels compared with their unfused counterparts (T.R. Butt, pers. comm.).

In this study, neither of the commonly used tags, GST or MBP, dramatically enhanced expression or solubility.

Table 3. *Proteins expressed as SUMO fusions in E. coli*

Protein/Protein Family
Liver X Receptor (LXR)
Human IL 21
Human IL 23
Human IL 18
β -Glucuronidase
Human Oxidoreductase
Bovine Pancreatic Trypsin Inhibitor
Human MAP kinase
D-Amino acid oxidase
HIV gp120 peptide
β -Catenin peptide
Wnt peptide
Tcf 4 peptide
TGF- β Related Growth Factor
T4 bacteriophage helicase

While for the purpose of the present study MBP was fused solely to the N terminus of partner proteins, successful expression has also been achieved when MBP is fused to the C terminus of a protein (Sachdev and Chirgwin 2000). In fact, it has been found that N-terminal MBP fusions can reduce the efficiency of translation, which may explain the low yield in these experiments with MBP (Hamilton et al. 2002; Podmore and Reynolds 2002). These studies underscore the need for a comprehensive analysis of the fusion tags for the expression of various protein families. If the present study is confirmed, it would appear that commonly used fusions for enhanced soluble protein expression are not the best fusion tags available.

Removal of the chosen fusion tag can also present a formidable challenge. In the present study, the best tags for enhanced expression and solubility (SUMO and NUS A) were further examined for the ease at which they are removed by their respective proteases (SUMO and AcTEV). Both AcTEV and SUMO proteases had similar affinities to their substrates. However, they differed dramatically in terms of catalytic efficiency (k_{cat}), with SUMO protease being ~25 times more efficient than AcTEV. In addition to being more catalytically efficient, SUMO protease also has the advantage of recognizing the tertiary structure of SUMO and not a linear amino acid sequence like AcTEV. This characteristic prevents the SUMO protease from erroneously cleaving within the target protein, and thus, SUMO protease behaves as a universal protease, suitable for nearly every substrate. In addition, direct fusion of the target protein to SUMO allows for the generation of recombinant protein with native N-terminal sequences. Taken together, these data suggest that the SUMO fusion system may be ideal for the soluble expression of proteins that have been impossible using traditional gene fusions.

Materials and methods

Construction of the fusion tags

All plasmids were constructed by standard methods. *E. coli* strains DH5 α and TOP10 (Invitrogen) were used for plasmid construction and manipulation, and pET24d (Novagen) was used as the backbone. N-terminal or C-terminal hexahistidine-tags were added to all fusions except for GST. In addition, all of the fusion tags except SUMO contained a cleavage site for AcTEV protease to facilitate the proper removal of the fusion tag upon purification. AcTEV was chosen in this study based on its high cleavage activity and it is commonly used in fusion technology. The cloning strategy used NcoI and BamHI sites upstream and downstream, respectively, of the fusion tag sequence, which was ligated pET24d vector. A BsaI restriction site was introduced by PCR directly downstream of the fusion tag sequence, upstream of the BamHI site in the reverse primer to ensure a consistent cloning strategy.

Construction of GFP, MMP13, and GDF8 fusion proteins

The cloning strategy to express the fusion proteins used BsaI in the forward primer directly upstream of *eGFP* (Clontech), *MMP13* (residues 20–274), or *GDF8* (mature) with either a HindIII or EcoRI restriction site downstream of the gene sequence in the reverse primer. These primers allowed the DNA fragment to be cloned in frame with the fusion tags. All plasmids were sequenced routinely (DNA Clinic Inc.).

Expression of model fusion proteins in *E. coli*

In a typical experiment, a single colony of transformed Rosetta (DE3) pLysS (Novagen) was inoculated into 2 mL of LB-broth with chloramphenicol (25 $\mu\text{g}/\text{mL}$) and kanamycin (30 $\mu\text{g}/\text{mL}$) and grown overnight at 37°C with shaking. A total of 1 mL of this overnight culture was inoculated into 100 mL of LB (25 $\mu\text{g}/\text{mL}$ Cm and 30 $\mu\text{g}/\text{mL}$ Kan), and the culture was grown to $\text{OD}_{600} = 0.6$ at 37°C with shaking. The culture was then cooled on ice, induced with 1 mM IPTG, and incubated at 20°C overnight with shaking. Cells were isolated by centrifugation at 10,000g for 20 min and resuspended in 3 mL of PBS (pH 8.0), 300 mM NaCl, 10 mM imidazole. Cells were lysed by mild sonication. Triton x-100 (Sigma) was added to 1% (v/v) and incubated at 4°C for 1 h, with shaking. The culture was centrifuged at 10,000g for 30 min, and the supernatant was decanted and stored at 4°C (soluble protein sample). Inclusion body samples were prepared by resuspending the insoluble material in 3 mL of solubilization buffer (50 mM CAPS at pH 11, 0.3 M NaCl, 0.3% N-lauryl sarcosine, and 1 mM DTT), and allowed to shake for 20 min at room temperature. The IB fraction was then centrifuged 10,000g for 20 min, and the supernatant removed and stored at 4°C (inclusion body protein sample).

Enzyme kinetics

SUMO-eGFP and NUS A-eGFP, the substrates used for kinetic analysis, were purified from soluble fractions prepared as described above. Both fusion constructs contained an N-terminal His₆ tag and were purified by Ni-NTA superflow resin (Qiagen) using the BioLogic Duo-Flow FPLC (Bio-Rad) as described previously (Zuo et al. 2005a). Briefly, the soluble cellular lysate was loaded onto a Ni-NTA (5 mL resin vol) column (20 cm \times 1.6 cm) at 1 mL/min and the resin was washed extensively with 40 mL wash buffer (PBS [pH 8.0], 20 mM imidazole, 150 mM NaCl) at 2 mL/min. Purified protein was then eluted in 15 mL of PBS (pH 8.0), 150 mM NaCl, 300 mM imidazole at 2 mL/min. The purified protein eluted typically as an isolated, symmetrical UV peak. Peak fractions were pooled (10 mL) and dialyzed using 3.5 kDa MWCO dialysis tubing against PBS (pH 7.5) at 4°C for 24 h.

The SUMO protease assays were initiated by mixing 6 μM SUMO protease (LifeSensors) with purified SUMO-eGFP (12, 6, 3, 1.5, 0.75, and 0.375 mM) in PBS (pH 7.5), 1 mM DTT. The reactions were incubated at 30°C for 1 h and stopped by addition of SDS-loading buffer. Samples were then heated at 95°C for 3 min and loaded onto a 15% SDS-polyacrylamide gel for analysis. The AcTEV protease assays were initiated by mixing 0.48 mM AcTEV protease (Invitrogen) with purified NUS A-eGFP (24, 12, 6, 3, 1.5, 0.75 mM) in PBS (pH 7.5), 1 mM DTT. The reactions were incubated at 30°C for 2 h and stopped by addition of SDS-loading buffer. Samples were then heated at 95°C for 3 min and loaded onto a 15% SDS-polyacrylamide gel

for analysis. Time course experiments concluded that 1 and 2 h were well within the initial velocity rates for SUMO protease and AcTEV protease, respectively (that is, the rate of reaction was within in the linear region [data not shown]).

The SDS-polyacrylamide gels were stained with Coomassie Blue and scanned such that each band could be quantified using the software Scion Image version Beta 4.0.2 (Scion Corporation). Densitometry analyses were performed, and when compared with a loading standard (known amount of product) used to determine the amount of product generated per unit time. The initial velocity values used for curve fitting were the mean of three independent experiments. These initial velocity measurements were plotted against the substrate concentration and fit to the Michaelis-Menton equation using KaleidaGraph 3.5 (Synergy Corporation). The k_{cat} values were calculated by assuming 100% activity for the enzyme.

Acknowledgments

We thank Drs. Stefan Masure, Susan Weiss, and Hiep Tran for their help with the SUMO fusion project. We also thank Dr. Tejvir S. Khurana and his colleagues at the University of Pennsylvania Medical School for help in expression and analysis of the GDF8 project. Numerous colleagues in the protein expression field have kindly shared with us their problems and ideas, for which we are most thankful. Research supported in part by grant numbers GM 068404-01, AI51752-01/02, and HL69744 awarded to T.R.B. by the NIH.

References

- Arechaga, I., Miroux, B., Runswick, M.J., and Walker, J.E. 2003. Over-expression of *Escherichia coli* F1F(o)-ATPase subunit a is inhibited by instability of the uncB gene transcript. *FEBS Lett.* **547**: 97–100.
- Bayer, P., Arndt, A., Metzger, S., Mahajan, R., Melchior, F., Jaenicke, R., and Becker, J. 1998. Structure determination of the small ubiquitin-related modifier SUMO-1. *J. Mol. Biol.* **280**: 275–286.
- Carrington, J.C., Cary, S.M., Parks, T.D., and Dougherty, W.G. 1989. A second proteinase encoded by a plant potyvirus genome. *EMBO J.* **8**: 365–370.
- Creighton, T.E. 1997. How important is the molten globule for correct protein folding? *Trends Biochem. Sci.* **22**: 6–10.
- 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, 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.
- Englander, S.W. 2000. Protein folding intermediates and pathways studied by hydrogen exchange. *Annu. Rev. Biophys. Biomol. Struct.* **29**: 213–238.
- Hamilton, S.R., O'Donnell Jr., J.B., Hammet, A., Stapleton, D., Habinowski, S.A., Means, A.R., Kemp, B.E., and Witters, L.A. 2002. AMP-activated protein kinase kinase: Detection with recombinant AMPK α 1 subunit. *Biochem. Biophys. Res. Commun.* **293**: 892–898.
- Hardern, I.M., Knauper, V., Ernil, R.J., Taylor, I.W., Cooper, K.L., and Abbott, W.M. 2000. An analysis of two refolding routes for a C-terminally truncated human collagenase-3 expressed in *Escherichia coli*. *Protein Expr. Purif.* **19**: 246–252.
- Ikura, K., Kokubu, T., Natsuka, S., Ichikawa, A., Adachi, M., Nishihara, K., Yanagi, H., and Utsumi, S. 2002. Co-overexpression of folding modulators improves the solubility of the recombinant guinea pig liver transglutaminase in *Escherichia coli*. *Prep. Biochem. Biotechnol.* **32**: 189–205.
- 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.
- Johnson, E.S. and Blobel, G. 1999. Cell cycle-regulated attachment of the ubiquitin-related protein SUMO to the yeast septins. *J. Cell. Biol.* **147**: 981–994.

- Kawabe, Y., Seki, M., Seki, T., Wang, W.S., Imamura, O., Furuichi, Y., Saitoh, H., and Enomoto, T. 2000. Covalent modification of the Werner's syndrome gene product with the ubiquitin-related protein, SUMO-1. *J. Biol. Chem.* **275**: 20963–20966.
- Khorasanizadeh, S., Peters, I.D., and Roder, H. 1996. Evidence for a three-state model of protein folding from kinetic analysis of ubiquitin variants with altered core residues. *Nat. Struct. Biol.* **3**: 193–205.
- Lee, S.J. and McPherron, A.C. 2001. Regulation of myostatin activity and muscle growth. *Proc. Natl. Acad. Sci.* **98**: 9306–9311.
- Li, S.J. and Hochstrasser, M. 1999. A new protease required for cell-cycle progression in yeast. *Nature* **398**: 246–251.
- . 2000. The yeast ULP2 (SMT4) gene encodes a novel protease specific for the ubiquitin-like Smt3 protein. *Mol. Cell. Biol.* **20**: 2367–2377.
- 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.
- Melchior, F. 2000. SUMO—nonclassical ubiquitin. *Annu. Rev. Cell. Dev. Biol.* **16**: 591–626.
- Mosesso, E. and Lima, C.D. 2000. Ulp1-SUMO crystal structure and genetic analysis reveal conserved interactions and a regulatory element essential for cell growth in yeast. *Mol. Cell* **5**: 865–876.
- Muller, S., Matunis, M.J., and Dejean, A. 1998. Conjugation with the ubiquitin-related modifier SUMO-1 regulates the partitioning of PML within the nucleus. *EMBO J.* **17**: 61–70.
- Nallamsetty, S., Kapust, R.B., Tozser, 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.
- Pathak, N., Hu, S.I., and Koehn, J.A. 1998. The expression, refolding, and purification of the catalytic domain of human collagenase-3 (MMP-13). *Protein Expr. Purif.* **14**: 283–288.
- Podmore, A.H. and Reynolds, P.E. 2002. Purification and characterization of VanXY(C), a D, D-dipeptidase/D, D-carboxypeptidase in vancomycin-resistant *Enterococcus gallinarum* BM4174. *Eur. J. Biochem.* **269**: 2740–2746.
- Pryor, K.D. and Leiting, B. 1997. High-level expression of soluble protein in *Escherichia coli* using a His6-tag and maltose-binding-protein double-affinity fusion system. *Protein Expr. Purif.* **10**: 309–319.
- Sachdev, D. and Chirgwin, J.M. 2000. Fusions to maltose-binding protein: Control of folding and solubility in protein purification. *Methods Enzymol.* **326**: 312–321.
- Studier, F.W. and Moffatt, B.A. 1986. Use of bacteriophage T7 RNA polymerase to direct selective high-level expression of cloned genes. *J. Mol. Biol.* **189**: 113–130.
- Tatham, M.H., Jaffray, E., Vaughan, O.A., Desterro, J.M., Botting, C.H., Naismith, J.H., and Hay, R.T. 2001. Polymeric chains of SUMO-2 and SUMO-3 are conjugated to protein substrates by SAE1/SAE2 and Ubc9. *J. Biol. Chem.* **276**: 35368–35374.
- Taylor, W.E., Bhasin, S., Artaza, J., Byhower, F., Azam, M., Willard Jr., D.H., Kull Jr., F.C., and Gonzalez-Cadavid, N. 2001. Myostatin inhibits cell proliferation and protein synthesis in C2C12 muscle cells. *Am. J. Physiol. Endocrinol. Metab.* **280**: E221–E228.
- Thomas, M., Langley, B., Berry, C., Sharma, M., Kirk, S., Bass, J., and Kambadur, R. 2000. Myostatin, a negative regulator of muscle growth, functions by inhibiting myoblast proliferation. *J. Biol. Chem.* **275**: 40235–40243.
- Wang, C., Castro, A.F., Wilkes, D.M., and Altenberg, G.A. 1999. Expression and purification of the first nucleotide-binding domain and linker region of human multidrug resistance gene product: Comparison of fusions to glutathione S-transferase, thioredoxin and maltose-binding protein. *Biochem. J.* **338**: 77–81.
- Zuo, X., Li, S., Hall, J., Mattern, M.R., Tran, H., Shoo, J., Tan, R., Weiss, S.R., and Butt, T.R. 2005a. Enhanced expression and purification of membrane proteins by SUMO fusion in *E. coli*. *J. Struct. Funct. Genomics* **6**: 103–111.
- Zuo, X., Mattern, M.R., Tan, R., Li, S., Hall, J., Sterner, D.E., Shoo, J., Tran, H., Lim, P., Sarafianos, S.G., et al. 2005b. Expression and purification of SARS Coronavirus proteins using SUMO fusions. *Protein Express Purif.* **42**: 100–110.