Review

SUMO fusion technology for difficult-to-express proteins

Tauseef R. Butt *, Suzanne C. Edavettal, John P. Hall, Michael R. Mattern

LifeSensors, Inc., 271 Great Valley Parkway, Malvern, PA 19355, USA

Abstract

The demands of structural and functional genomics for large quantities of soluble, properly folded proteins in heterologous hosts have been aided by advancements in the field of protein production and purification. *Escherichia coli*, the preferred host for recombinant protein expression, presents many challenges which must be surmounted in order to over-express heterologous proteins. These challenges include the proteolytic degradation of target proteins, protein misfolding, poor solubility, and the necessity for good purification methodologies. Gene fusion technologies have been able to improve heterologous expression by overcoming many of these challenges. The ability of gene fusions to improve expression, solubility, purification, and decrease proteolytic degradation will be discussed in this review. The main disadvantage, cleaving the protein fusion, will also be addressed. Focus will be given to the newly described SUMO fusion system and the improvements that this technology has advanced over traditional gene fusion systems.
© 2005 Elsevier Inc. All rights reserved.

Keywords: Protein expression; Gene fusion; Protein fusion; Ubhiquitin; SUMO; Ubiquitin-like proteins

Efficient recombinant protein expression is a major bottleneck for structural genomics and proteomics. Despite progress in automation, soluble protein expression is frequently the rate-limiting step for many researchers. The preferred host for recombinant protein expression has historically been *Escherichia coli* (*E. coli*) due to the simplicity and low costs associated with using this host. While *E. coli* has proved a successful host for the expression of many heterologous proteins, it is frequently not capable of expressing soluble heterologous proteins. The Southeast Collaboratory for Structural Genomics (SECSG)[1] reports that of the 6386 proteins they have expressed in *E. coli* only 22.7% (1452) have been soluble (as published on SECSG web page 03/04/

2005). Much advancement has been made toward improving recombinant protein expression in *E. coli*, including the development of strong promoters [1], co-expression with chaperones [2], and through the use of protein fusions. No other technology has been as effective at improving the solubility of recombinant proteins as fusion systems, especially for difficult-to-express proteins. A variety of structures have been used as fusion motifs (Table 1). These fusion proteins are frequently employed to enhance protein expression and facilitate purification [3–5]. There is no particular similarity among these proteins in terms of molecular weight, structure, or function, with the exception of ubiquitin (Ub) and SUMO, which share a common structure [6]. This review will discuss the advantages of fusion technologies. These advantages include the manner in which protein expression is enhanced, proteolytic degradation of the target protein is decreased, protein folding and solubility are increased, and purification and detection are simplified. The main disadvantage of fusion technology, cleaving the protein fusion, is also covered. In addition, focus is given to the newly described SUMO fusion

---

* Corresponding author. Fax: +1 610 644 8616.
  E-mail address: butt@lifesensors.com (T.R. Butt).

[1] *Abbreviations used:* SECSG, Southeast Collaboratory for Structural Genomics; SUMO, small ubiquitin modifying protein; Ni–NTA, nickel–nitriloacetic acid; CTHS, C-terminal half of SUMO; NTHS, N-terminal half of SUMO; GST, glutathione *S*-transferase; MBP, maltose binding protein; Trx, thioredoxin; Ub, ubiquitin; Tev, tobacco etch virus; DUBs, deubiquitylating enzymes.

Table 1
Fusion partner proteins, which enhance expression and simplify purification

Fusion partner	Molecular weight (kDa)	Purification method
6× His[a]-SUMO	12.2	Ni–NTA[a]
CTHS	7	NTHS
GST	27.3	Glutathione–Sepharose
6× His-NusA	58.4	Ni–NTA
MBP	45	Amylose resin
TRX	14.3	ThioBond resin
6× HIS-Ub	8.6	Ni–NTA
Flag peptide	1	Anti-Flag antibodies

Abbreviations used: SUMO, small ubiquitin modifying protein; Ni–NTA, nickel–nitriloacetic acid; CTHS, C-terminal half of SUMO; NTHS, N-terminal half of SUMO; GST, glutathione *S*-transferase; MBP, maltose binding protein; Trx, thioredoxin; Ub, ubiquitin.

[a] The N-terminus of SUMO is very flexible and can accommodate a variety of affinity tags (e.g., GST). The SUMO system can therefore be tailored to the researcher's desired affinity tag.

systems. SUMO fusions appear to have all of the advantages of traditional fusion systems, and due to the activity of SUMO proteases do not encounter the same difficulties with cleavage.

Advantages of recombinant protein fusion technology

Protection from degradation

Proteolysis is highly regulated and plays critical roles in maintaining cellular homeostasis, including removing unwanted or incorrectly folded proteins from the cell [7]. Often, recombinant proteins are viewed as unwanted by cells and are subjected to proteolytic degradation, decreasing the level of recombinant protein expression (reviewed in [8]). Several strategies have been developed to protect recombinant proteins from degradation including the use of protease inhibitors [9], secretion into the periplasm [10] or culture medium [11], and generating protective fusions [12]. Fusions between the N-terminus of target proteins and protein tags (Table 1) have been shown to protect the target protein from degradation [13–15]. Furthermore, fusions with C-terminus [16], dual C- and N-terminal fusions [17], and tandem fusions of multiple copies of the target gene [18] have been shown to afford protection from proteolytic degradation.

The compartmentalization hypothesis describes the mechanism by which gene fusions protect against proteolytic degradation [19]. Fusions can promote the translocation of their partner proteins to different cellular compartments, thereby decreasing the concentration of the recombinant protein in the protease-rich cytosol. For example, maltose binding protein (MBP) can translocate to the membrane compartment of the cell [20] or SUMO can translocate from the cytosol to the nucleus [21]. The tags can thereby compartmentalize their part-

ner proteins and decrease the susceptibility to proteolytic degradation.

Enhanced recombinant protein expression

Protein expression is a complex process dependent upon mRNA stability and translational efficiency, as well as transcriptional regulation. Enhanced recombinant protein expression is the result of a high mRNA copy number, efficient translational initiation and elongation, stability of the mRNA, and translation enhancers (reviewed in [22]). Many heterologous genes are not translated efficiently due to codon usage bias. Codon bias has been implicated as one of the main reasons for inefficient translation (e.g., malaria parasite genes [23]). Codon bias has been overcome by engineering new strains or cell lines that contain rare tRNAs or by altering the problematic codons to more common prokaryotic codons [24]. Strong and highly regulated promoters from *E. coli*, yeast, and insect cells are available [14,25,26], and as such transcription of the heterologous gene is usually not a rate-limiting factor. It has been observed that many fusion proteins enhance protein expression [14,15,27], however, the exact mechanism by which fusion proteins achieve this enhanced expression is unknown. It has been speculated that enhanced expression is the result of the highly conserved structures of these proteins. Attachment of a highly evolved translational frame at the N-terminus of an inefficiently translated protein may help to improve the latter's efficiency of expression [28].

Improved protein folding

Although *E. coli* is usually the first choice as a recombinant expression organism, many eukaryotic proteins, especially proteins with disulfide bridges, cannot be expressed as soluble, active, and properly folded proteins in *E. coli* [29]. When a large quantity of protein is expressed, macromolecular crowding (200–300 mg/ml in the cytoplasm) presents an unfavorable environment for protein folding. Frequently, the result of a high concentration of incorrectly folded recombinant protein is the formation of inclusion bodies [30]. In fact, the level of aggregated protein can increase to the extent that inclusion bodies are observable by light microscopy as round bodies surrounding the cytosol of *E. coli* [31]. While the formation of inclusion bodies does afford protection from proteolytic degradation, and re-folding can recover active proteins from inactive inclusion bodies, the initial expression of correctly folded, soluble recombinant protein is ideal [32].

Several strategies have been developed to promote the expression of properly folded recombinant protein, including co-expression of molecular chaperones [2] and foldases [33], expression of secreted proteins [34], and expression of protein fusions [35]. Fusion partners (Table 1) have been shown to act as solubility enhancers,

although the exact mechanism by which they improve solubility has not been described. It has been hypothesized that these fusions may act as chaperones [36]. The fusion of a stable or conserved structure to an insoluble recombinant protein may serve to stabilize and promote proper folding of the recombinant protein. MBP has been shown to function as a general molecular chaperone in the context of a fusion protein by binding to aggregation-prone folding intermediates of passenger proteins and preventing their self-association [36]. Fusion tags have also been hypothesized to enhance the solubility of the protein target by acting as a nucleus of folding ("molten globule hypothesis") [37,38]. This theory suggests that a fusion tag acts as a nucleation site for the folding of the target protein. Ub, which has a highly hydrophobic core, has been shown to be the fastest folding protein known [39]. Ub's ancient structure is highly conserved in SUMO, as well as in other ubiquitin-like proteins used as fusions [6]. When fused to Ub or SUMO, otherwise insoluble proteins have been observed to fold properly and be soluble [14,15].

Simplified purification and detection

The use of protein fusion technology offers the opportunity to simplify and facilitate purification and detection of recombinant proteins. Fusing a tag to the target protein provides a one-step purification procedure by passing cell extracts or supernatants over an appropriate affinity matrix. Numerous examples of affinity purification exist for fusion proteins, including nickel–nitriloacetic acid to isolate hexahistidine-fused proteins [40], biotin to isolate streptavidin-fused proteins [41], and amylose to isolate MBP-fused proteins [42]. The fusion tag can also be used to identify the recombinant product through Western blotting by way of anti-tag antibodies [43].

Disadvantages of recombinant protein fusion technology

Cleaving the fusion protein

Cleavage of protein fusions to generate free protein remains the major disadvantage of protein fusion technologies. Cleavage of the fusion is usually necessary because the fusion interferes with the structural or functional properties of the recombinant protein [44]. A variety of chemical and enzymatic methodologies have been developed to cleave fusions [45,46]. These methods include the use of engineered cleavage sites, which are recognized by the proteases and are positioned between the fusion tag and the protein target. Proteases that have been employed to cleave fusion tags include tobacco etch virus (Tev) protease [47], factor Xa, or thrombin protease (reviewed in [46]). Problems associated with proteolytic cleavage of fusion tags are low yield, precipitation

of the protein of interest, labor-intensive optimization of cleavage conditions, expense of proteases, and failure to recover active, structurally intact protein [48].

Another major problem encountered with the cleavage of fusion proteins is the generation of non-native N-terminal amino acids. Many structural and therapeutic proteins require a specific N-terminus, other than methionine, for biological activity (e.g., chemokines). All nascent proteins have methionine as their N-terminal residue; some (e.g., precursor proteins, which undergo proteolytic maturation) then undergo post-translational processing and modification, leading to a variety of N-terminal amino acids. Experiments performed by Varshavsky and co-workers [49] in the 1980s demonstrated that the expression of ubiquitin-β-galactosidase in yeast leads to rapid processing of ubiquitin in the cells. The processing of Ub-X-fusion, where X is any amino acid, was independent of the identity of residue X [19,49]. It was observed that the in vivo half-life of the resulting protein varied as a function of the N-terminal residue of the protein (residue X). The "N-end rule" proposes that there is a relationship between the identity of a protein's N-terminal residue and its half-life. For an in depth discussion of the "N-end rule," please see appropriate publications of Varshavsky and co-workers [19,49].

Cleavage by the aforementioned 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 LVP**R**GS at the arginine residue, resulting in an N-terminal extension of the target protein by two amino acids (GS) [46]. Since many proteins require a specific N-terminus for biological activity, half-life, or structural stability, this characteristic of protease cleavage can have serious effects on the ability to recombinantly produce active protein. Notable exceptions include the intein, SUMO, and Ub fusion systems. The intein fusion system utilizes the inducible self-cleavage activity of engineered protein splicing elements (termed inteins). In the presence of thiols such as DTT, β-mercaptoethanol or cysteine, the intein undergoes specific self-cleavage which releases the target protein from the tag. When fused directly to the N-terminus of the target protein cleavage results in the production of a target protein without any extra non-native residues attached to its terminus after cleavage [45].

Unlike intein-mediated cleavage, the Ub and SUMO fusion systems do require the use of proteases for removal of the tag. However, the SUMO proteases and the deubiquitylating enzymes (DUBs) are distinct from other proteases that recognize a peptide sequence, as they recognize the tertiary structure of the SUMO or Ub tag. When the target protein is fused directly to the C-terminus of SUMO or Ub, cleavage will not result in extraneous residues at the N-terminus of the target protein and therefore will yield native-like protein [15,49]. The DUBs

have a major drawback in that complete cleavage of the Ub tag requires a large amount of enzyme (1:10 molar ratio of DUB to target) [50]. The SUMO protease has been reported to be much more robust, requiring only a 1:5000 molar ratio of protease to target [15]. The new generation of SUMO proteases has been shown to be even more catalytically active requiring only 1:100,000 molar ratio of protease to target (Butt et al. unpublished results) (see further discussion below).

SUMO fusion technology

SUMO and SUMO pathways

The SUMO family of proteins has been described extensively in the literature (reviewed in [51]). SUMO and the SUMO pathway are highly conserved in all eukaryotes from yeast to humans, but are absent from prokaryotes (Fig. 1) [52–54]. *Saccharomyces cerevisiae* has only a single SUMO gene, *SMT3* which is essential for viability [55,56]. In contrast, vertebrates have three SUMO genes, SUMO-1, SUMO-2, and SUMO-3. The three human SUMOs are highly homologous, with human SUMO-1 sharing 50% sequence identity with human SUMO-2 and SUMO-3 [56], and human SUMO-2 and SUMO-3 sharing 87% sequence identity with each other [57]. SMT3 shares 47% sequence identity with human SUMO-1. Although overall sequence identity between Ub and SUMO is only 18%, structure determination by nuclear magnetic resonance (NMR) reveals that the two proteins possess a common three-dimensional structure that is characterized by a tightly packed globular fold with β-sheets wrapped around one α-helix [6].

The family of SUMO proteins function, like Ub, as covalent modifiers of other proteins, and sumoylation occurs in a similar fashion to the ubiquitination of proteins [57–59] (Fig. 1). SUMO is activated in an ATP-dependent step by the formation of a thioester bond with the SUMO activating enzyme E1. Following activation, SUMO is transferred to SUMO E2, or SUMO-conjugating enzyme. Following conjugation to E2, SUMO is transferred to its target protein through the activity of SUMO E3. These modifications can have a variety of cellular consequences. Sumoylation can antagonize the action of ubiquitination by preventing ubiquitin-meditated proteolysis [60]. SUMO conjugation has also been observed to impact higher-order chromatin structure [61], transcriptional regulation [62], DNA repair pathways [63], nuclear transport [64], and signal transduction pathways [65].

SUMO conjugation to target proteins is a dynamic process, which changes in response to a variety of stimuli [66]. SUMO can be removed from target proteins enzymatically by SUMO C-terminal hydrolases-isopeptidases, several of which are now known in many species including yeast (Ulp1 and Ulp2) [67–69], *Arabidopsis*
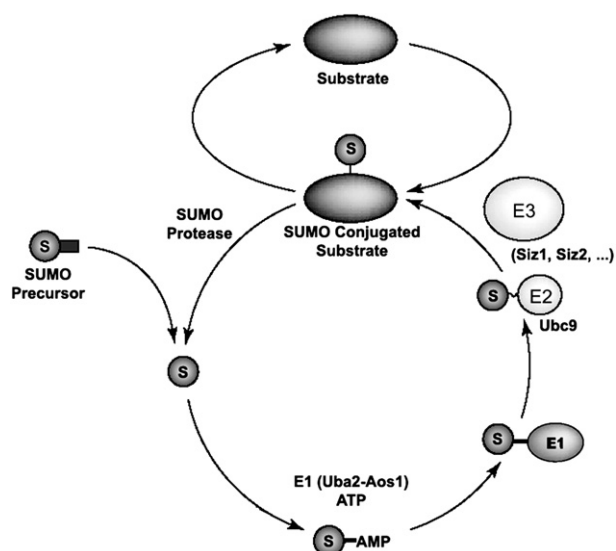

Fig. 1. The SUMO cycle. SUMO is synthesized as a precursor and cleaved by SUMO proteases (yeast Ulp1 or Ulp2). Activating enzyme (E1), conjugating enzyme (E2), and ligase (E3) are named according to yeast where most were discovered.

[70], and humans [59,71–73]. SUMO proteases share a common C-terminal domain (Ulp domain), and have no sequence homology to the DUBs [67–69].

SUMO fusion enhances protein expression, solubility, and purification in prokaryotes

Recently, SUMO has been fused to the N-terminus of several proteins, including matrix metalloprotease (MMP13), green fluorescent protein (GFP), and SARS-CoV 3CL protease, and used as a recombinant expression system [15,74] (www.lifesensors.com). SUMO fusion leads to enhanced expression and solubility (Fig. 2A and unpublished results). For example, when MMP13 is expressed without SUMO fusion in *E. coli*, it is contained primarily in inclusion bodies. However, when MMP13 is expressed as a SUMO fusion, MMP13 was observed primarily in the soluble fraction [15]. The effect that SUMO has on enhancing protein solubility has been explained in part by the structure of SUMO. SUMO has an external hydrophilic surface and inner hydrophobic core, which may exert a detergent-like effect on otherwise insoluble proteins [15]. A hexahistidine SUMO fusion construct has been shown to enhance expression and facilitate purification with Ni–NTA chromatography. Ni–NTA chromatography has been used to purify the fusion from the cellular lysate (Fig. 2B) [15].

SUMO proteases

As described above, the major disadvantage of fusion protein technology is cleavage of the tag. Com-
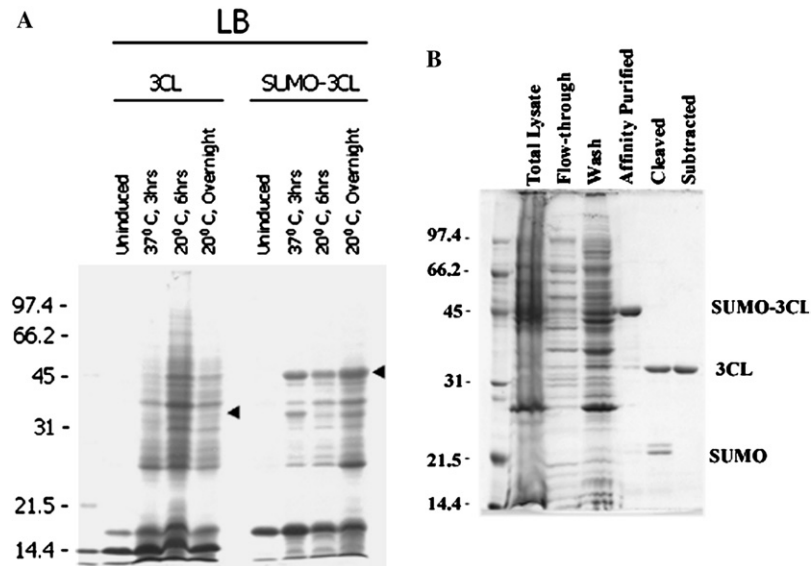
Fig. 2. Enhanced expression and purification using the SUMO fusion system. (A) Enhanced expression of SARS-CoV 3CL protease (3CL) by SUMO fusion in *E. coli*. Cells grown in Luria–Bertani (LB) media were induced at the temperatures and for the lengths of time indicated. Just before expression was induced and after induction was completed the cells from a 1.5 ml aliquot of culture were lysed. Samples of whole cell lysates (∼7.5 μl) from the various expression conditions were resolved in 12% SDS-gels and stained with Coomassie blue. Molecular weights (in kDa) were as indicated, and arrowheads highlight expected/observed positions of respective expressed protein bands [74]. (B) An SDS gel depicting the purification of SARS-CoV 3CL protease. Total cell lysate was passed over a Ni–NTA column, washed with 40 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). Aliquots of the samples (each containing ∼5 μg protein) were separated on a 12% SDS-gel and stained with Coomassie blue. The migration positions of the SUMO fusion and the proteins resulting from the cleavage are as indicated [74].

monly used proteases do not cleave all fusions efficiently, accurately and, moreover, can generate extraneous amino acids at the N-terminus of the target protein [75]. SUMO proteases, which are members of the cysteine protease superfamily, are able to overcome these difficulties. SUMO proteases (LifeSensors) are accurate and efficient in cleavage of the SUMO tag and allow for retention of the desired N-terminus. To date, we have cleaved approximately 100 SUMO fusions and never observed erroneous cleavage within the partner protein. Unlike other proteases that recognize a peptide sequence, SUMO proteases recognize the tertiary structure of the SUMO tag and cleavage does not result in an extended the N-terminus in the partner protein. Owing to its unique structure, SUMO protease can accommodate a variety of SUMO fusion partner proteins (Fig. 4). SUMO proteases have been observed to completely cleave a broad range (6–110 kDa) of partner proteins fused to SUMO [15,74]. In addition, SUMO protease is able to cleave efficiently under a wide range of conditions, including pH, temperature, and ionic strength (Fig. 3). Unfortunately, SUMO protease is not able to cleave target proteins which contain an N-terminal proline. SUMO proteases have a con-
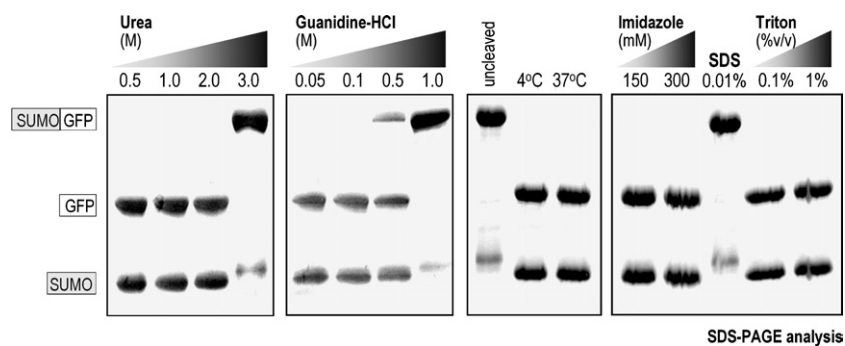

Fig. 3. Effect of various conditions on the activity of SUMO protease. SUMO-green fluorescent protein fusion (SUMO-GFP) (2.5 μg) was incubated with SUMO protease (1:5000 molar ratio of SUMO-GFP to protease) for 1 h under conditions described in the figure: temperatures of 4 or 37 °C, and concentration ranges of imidazole (at 25 °C), sodium dodecyl sulfate (SDS) (at 25 °C), Triton X-100 (at 25 °C), urea (at 25 °C), or guanidine hydrochloride (at 25 °C). The data show that the enzyme is active over a broad temperature range and tolerates highly adverse biochemical conditions [15].
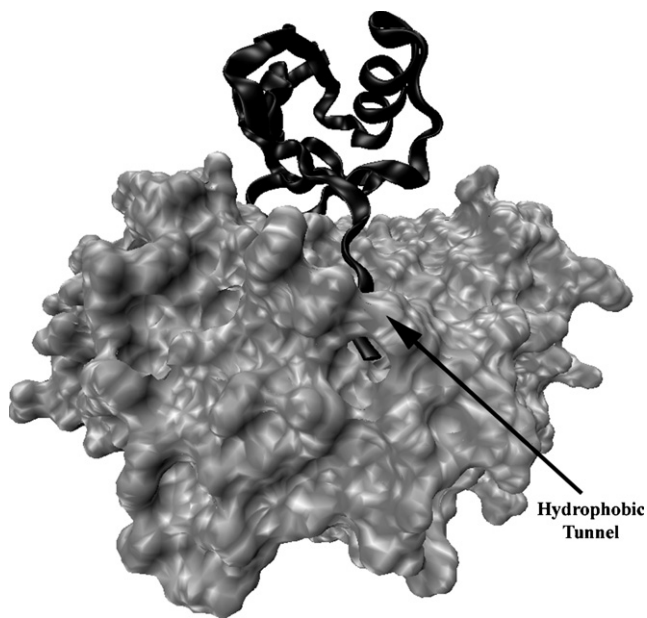
Fig. 4. The X-ray crystal structure of human SUMO protease (Senp2, grey) in complex with human SUMO-1 (black) [76]. SUMO-1 must pass through a constrictive hydrophobic tunnel (arrow) within the active site in order to be cleaved by Senp2.

strictive hydrophobic tunnel within the active site, and substrates must pass through this tunnel during cleavage (Fig. 4) [76]. It has been hypothesized that the constrictive tunnel is unable to accommodate the structural changes induced by prolines near the cleavage site.

SUMO fusion in eukaryotic hosts, the split SUMO solution

This review is primarily focused on the role of SUMO fusion expression systems in prokaryotic cells. However, it is relevant to mention that fusion of Ub to under-expressed genes in eukaryotes enhances the level of protein production, even though the Ub tag is cleaved by endogenous DUBs [77]. We have also observed dramatic enhancement of under-expressed proteins in eukaryotes after fusion with SUMO. Similar to Ub, the SUMO tag is cleaved soon after translation (Edavettal and Butt, unpublished). These studies suggest that protein expression enhancing properties of Ub and like proteins (SUMO) are preserved in eukaryotes.

The lack of an endogenous SUMO protease in _E. coli_ facilitates the use of SUMO as a purification tag in this host organism. However, in eukaryotic systems, endogenous SUMO proteases immediately cleave SUMO fusions, making purification using the SUMO tag impossible. A novel fusion system, split SUMO, has been developed to overcome this problem. In the split SUMO system, SUMO is bifurcated into N- and C-terminal halves (NTHS and CTHS) (Fig. 5A). Fusion of the CTHS to the N-terminus of a target protein allows enhancement of expression in eukaryotes and, most importantly, the fusion is not recognized by endogenous SUMO proteases; therefore it is not cleaved in vivo and can facilitate purification. CTHS has strong hydrophobic interactions with NTHS, and mixture of
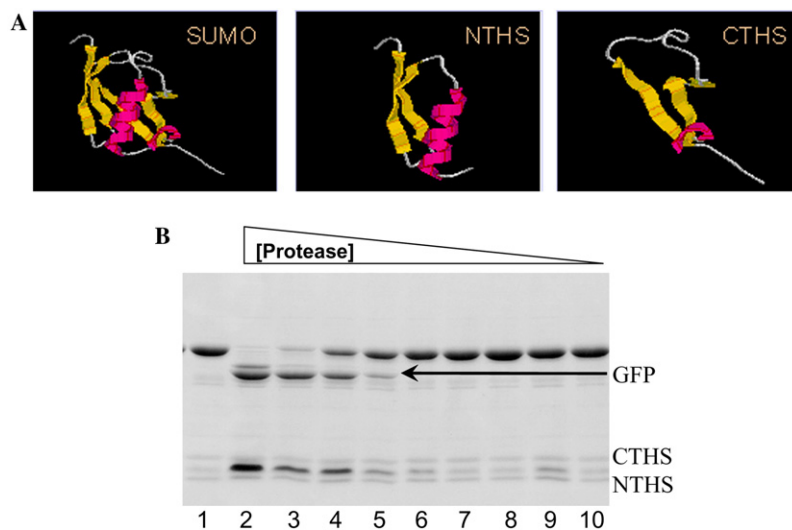

Fig. 5. The split SUMO expression system. (A) The structure of SUMO, and the N- and C-terminal halves (NTHS and CTHS). The target protein is fused to the CTHS, and the full SUMO structure is reconstituted after purification by incubating with NTHS. (B) The reconstitution of cleavable structure on the CTHS fusion in vitro and cleavage by Ulp1. An SDS PAGE of 6× His-CTHS-GFP (8 μg) fusion protein purified from _E. coli_ that was incubated for 30 min at 30 °C with purified 6× His-NTHS (2 μg) and increasing concentrations of SUMO protease (lane 1, CTHS-GFP + NTHS and lanes 2–9, CTHS-GFP + NTHS with decreasing concentrations of SUMO protease (1000, 500, 250, 125, 62.5, 31.3, 15.6, 7.8, and 3.9 ng)). Note the release of free GFP indicating the reconstitution of the full SUMO structure at high protease concentrations.

NTHS with the CTHS fusion allows for reconstitution of SUMO in vitro. Reconstitution of intact SUMO by non-covalent interactions between the two halves in vitro generates a substrate for SUMO protease, permitting cleavage and purification of the partner protein (Fig. 5B). We have also observed that fusion of a secretory signal to the N-terminus of the SUMO fusion permits enhanced expression and prevents cleavage of the SUMO tag (Edavettal et al., unpublished). We believe that the secretory signal affords protection from the endogenous SUMO proteases because the nascent protein is captured by the endoplasmic reticulum, and secreted into the media, therefore bypassing the protease-rich cytosol.

Conclusions and future directions

Rapid, efficient, and cost-effective protein expression and purification strategies are required for high throughput structural genomics and the production of therapeutic proteins. Fusion protein technology represents one strategy to achieve these goals. Fusion protein technology allows for the enhanced expression of recombinant, proteins, which are protected from degradation and have improved solubility and simplified purification and detection. The major drawback to most fusion systems is cleavage of the fusion tag, which can result in the generation of non-native N-termini and erroneous cleavage. The SUMO fusion expression system affords the advantages of other fusion technologies, but also the SUMO protease is efficient, accurate, and does not result in the extraneous residues at the N-terminus of the target protein. We are currently developing the second generation of SUMO tags, which have demonstrated added enhanced expression of under-expressed proteins, and the second generation of SUMO proteases, which are even more robust than Ulp1.

The SUMO and split SUMO fusion technologies allow for efficient recombinant expression in both prokaryotic and eukaryotic hosts. This technology has been used to efficiently express and purify a variety of proteins, including membrane proteins [78]. Another innovative application of the SUMO fusion technology is its use as an immobilization tool for protein arrays. Applications for protein arrays include expression profiling, protein isolation and purification, protein–protein interaction studies, and small molecular drug discovery [79–81]. SUMO has both flexible N- and C-termini, allowing dynamic processes to occur with relative ease, and the complex can be released from the solid support by the action of SUMO protease, which facilitates the identification and characterization of the partner protein(s). SUMO and other fusion technologies will allow for efficient recombinant expression of proteins, aiding in numerous future discoveries.

References

[1] F.W. Studier, A.H. Rosenberg, J.J. Dunn, J.W. Dubendorff, Use of T7 RNA polymerase to direct expression of cloned genes, Methods Enzymol. 185 (1990) 60–89.

[2] K. Ikura, T. Kokubu, S. Natsuka, A. Ichikawa, M. Adachi, K. Nishihara, H. Yanagi, S. Utsumi, Co-overexpression of folding modulators improves the solubility of the recombinant guinea pig liver transglutaminase in *Escherichia coli*, Prep. Biochem. Biotechnol. 32 (2002) 189–205.

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

[4] C. Wang, A.F. Castro, D.M. Wilkes, G.A. Altenberg, 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 (1999) 77–81.

[5] K.D. Pryor, B. Leiting, High-level expression of soluble protein in *Escherichia coli* using a His6-tag and maltose-binding-protein double-affinity fusion system, Protein Express. Purif. 10 (1997) 309–319.

[6] P. Bayer, A. Arndt, S. Metzger, R. Mahajan, F. Melchior, R. Jaenicke, J. Becker, Structure determination of the small ubiquitin-related modifier SUMO-1, J. Mol. Biol. 280 (1998) 275–286.

[7] A.L Goldberg, J.F. Dice, Intracellular protein degradation in mammalian and bacterial cells, Annu. Rev. Biochem. 43 (1974) 835–869.

[8] A. Rozkov, S.-O. Enfors, Analysis and control of proteolysis of recombinant proteins in *Escherichia coli*, Adv. Biochem. Eng./Biotechnol. 89 (2004) 163–195.

[9] W. Prouty, A. Goldberg, Effects of protease inhibitors on protein breakdown in *Escherichia coli*, J. Biol. Chem. 247 (1972) 3341–3352.

[10] K. Talmadge, W. Gilbert, Cellular location affects protein stability in *Escherichia coli*, Proc. Natl. Acad. Sci. USA 79 (1982) 1830–1833.

[11] A. Hogset, O.R. Blingsom, O. Saether, V.T. Gautvik, E. Holmgren, M. Hartmanis, S. Josephson, O.S. Gabrielsen, J.O. Gordeladze, P. Alestrom, Expression and characterization of a recombinant human parathyroid hormone secreted by *Escherichia coli* employing the staphylococcal protein A promoter and signal sequence, J. Biol. Chem. 265 (1990) 7338–7344.

[12] M. Murby, M. Uhlen, S. Stahl, Upstream strategies to minimize proteolytic degradation upon recombinant production in *Escherichia coli*, Protein Express. Purif. 7 (1996) 129–136.

[13] A. Martinez, P.M. Knappskog, S. Olafsdottir, A.P. Doskeland, H.G. Eiken, R.M. Svebak, M. Bozzini, J. Apold, T. Flatmark, Expression of recombinant human phenylalanine hydroxylase as

fusion protein in *Escherichia coli* circumvents proteolytic degradation by host cell proteases. Isolation and characterization of the wild-type enzyme, Biochem. J. 306 (1995) 589–597.

[14] T.R. Butt, S. Jonnalagadda, B.P. Monia, E.J. Sternberg, J.A. Marsh, J.M. Stadel, D.J. Ecker, S.T. Crooke, Ubiquitin fusion augments the yield of cloned gene products in *Escherichia coli*, Proc. Natl. Acad. Sci. USA 86 (1989) 2540–2544.

[15] M.P. Malakhov, O.A. Malakhova, M. Drinker, S. Weeks, T.R. Butt, SUMO fusion and SUMO-specific proteases for efficient expression and purification of proteins, J. Struct. Funct. Genom. 5 (2004) 75–86.

[16] M.H. Koken, H.H. Odijk, M. van Duin, M. Fornerod, J.H. Hoeijmakers, Augmentation of protein production by a combination of the T7 RNA polymerase system and ubiquitin fusion: overproduction of the human DNA repair protein, ERCC1, as an ubiquitin fusion protein in *Escherichia coli*, Biochem. Biophys. Res. Commun. 195 (1993) 643–653.

[17] M. Murby, L. Cedergren, J. Nilsson, P.A. Nygren, B. Hammarberg, B. Nilsson, S.O. Enfors, M. Uhlen, Stabilization of recombinant proteins from proteolytic degradation in *Escherichia coli* using a dual affinity fusion strategy, Biotechnol. Appl. Biochem. 14 (1991) 336–346.

[18] S.H. Shen, Multiple joined genes prevent product degradation in *Escherichia coli*, Proc. Natl. Acad. Sci. USA 81 (1984) 4627–4631.

[19] A. Varshavsky, The N-end rule: functions, mysteries, uses, Proc. Natl. Acad. Sci. USA 93 (1996) 12142–12149.

[20] H. Nikaido, Maltose transport system of *Escherichia coli*: an ABC-type transporter, FEBS Lett. 346 (1994) 55–58.

[21] A. Kishi, T. Nakamura, Y. Nishio, H. Maegawa, A. Kashiwagi, Sumoylation of Pdx1 is associated with its nuclear localization and insulin gene activation, Am. J. Physiol. Endocrinol. Metab. 284 (2003) E830–E840.

[22] A.C. Makrides, Strategies for achieving high-level expression of genes in *Escherichia coli*, Microbiol. Rev. 60 (1996) 512–538.

[23] D.L. Narum, S. Kumar, W.O. Rogers, S.R. Fuhrmann, H. Liang, M. Oakley, A. Taye, B.K. Sim, S.L. Hoffman, Codon optimization of gene fragments encoding *Plasmodium falciparum* merzoite proteins enhances DNA vaccine protein expression and immunogenicity in mice, Infect. Immun. 69 (2001) 7250–7253.

[24] U. Brinkmann, R.E. Mattes, P. Buckel, High-level expression of recombinant genes in *Escherichia coli* is dependent on the availability of the *dna Y* gene product, Gene 85 (1989) 109–114.

[25] J.W. Dubendorff, F.W. Studier, Controlling basal expression in an inducible T7 expression system by blocking the target T7 promoter with the lac repressor, J. Mol. Biol. 219 (1991) 45–59.

[26] S.M. Thiem, L.K. Miller, Differential gene expression mediated by late, very late and hybrid baculovirus promoters, Gene 91 (1990) 87–94.

[27] D. Sachdev, J.M. Chirgwin, Fusions to maltose-binding protein: control of folding and solubility in protein purification, Methods Enzymol. 326 (2000) 312–321.

[28] I. Arechaga, B. Miroux, M.J. Runswick, J.E. Walker, Over-expression of *Escherichia coli* $F_1F(o)$-ATPase subunit α is inhibited by instability of the uncB gene transcript, FEBS Lett. 547 (2003) 97–100.

[29] W.A. Prinz, F. Aslund, A Holmgren, J. Beckwith, The role of the thioredoxin and glutaredoxin pathways in reducing protein disulfide bonds in the *Escherichia coli* cytoplasm, J. Biol. Chem. 272 (1997) 15661–15667.

[30] J. Chen, T.B. Acton, S.K. Basu, G.T. Montelione, M. Inouye, Enhancement of the solubility of proteins overexpressed in *Escherichia coli* by heat shock, J. Mol. Microbiol. Biotechnol. 4 (2002) 519–524.

[31] M.M. Carrio, R. Cubarsi, A. Villaverde, Fine architecture of bacterial inclusion bodies, FEBS Lett. 471 (2000) 7–11.

[32] G. Georgiou, P. Valax, Expression of correctly folded proteins in *Escherichia coli*, Curr. Opin. Biotechnol. 7 (1996) 190–197.

[33] C. Wulfing, A. Pluckthun, Correctly folded T-cell receptor fragments in the periplasm of *Escherichia coli*. Influence of folding catalysts, J. Mol. Biol. 242 (1994) 655–669.

[34] Y. Sugamata, T. Shiba, Improved secretory production of recombinant proteins by random mutagenesis of hlyB, an α-hemolysin transporter from *Escherichia coli*, Appl. Environ. Microbiol. 71 (2005) 656–662.

[35] A.K. Chopra, A.R. Brasier, M. Das, X.J. Xu, J.W. Peterson, Improved synthesis of *Salmonella typhimurium* enterotoxin using gene fusion expression systems, Gene 144 (1994) 81–85.

[36] J.D. Fox, R.B. Kapust, D.S. Waugh, Single amino acid substitutions on the surface of *Escherichia coli* maltose-binding protein can have a profound impact on the solubility of fusion proteins, Protein Sci. 10 (2001) 622–630.

[37] S.W. Englander, Protein folding intermediates and pathways studied by hydrogen exchange, Annu. Rev. Biophys. Biomol. Struct. 29 (2000) 213–238.

[38] T.E. Creighton, How important is the molten globule for correct protein folding?, Trends Biochem. Sci. 22 (1997) 6–10.

[39] S. Khorasanizadeh, I.D. Peters, H. Roder, Evidence for a three-state model of protein folding from kinetic analysis of ubiquitin variants with altered core residues, Nat. Struct. Biol. 3 (1996) 193–205.

[40] S.K. Sharma, D.B. Evans, A.F. Vosters, T.J. McQuade, W.G. Tarpley, Metal affinity chromatography of recombinant HIV-1 reverse transcriptase containing a human renin cleavable metal binding domain, Biotechnol. Appl. Biochem. 14 (1991) 69–81.

[41] C.F. Ford, I. Suominen, C.E. Glatz, Fusion tails for the recovery and purification of recombinant proteins, Protein Express. Purif. 2 (1991) 95–107.

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

[43] Q.M. Sun, L. Cao, L. Fang, C. Chen, J. Dai, L.L. Chen, Z.C. Hua, Expression, purification of human vasostatin120–180 in *Escherichia coli* and its anti-angiogenic characterization, Protein Express. Purif. 39 (2005) 288–295.

[44] P. Balbas, Understanding the art of producing protein and non-protein molecules in *Escherichia coli*, Mol. Biotechnol. 19 (2001) 251–267.

[45] S. Chong, F.B. Mersha, D.G. Comb, M.E. Scott, D. Landry, L.M. Vence, F.B. Perler, J. Benner, R.B. Kucera, C.A. Hirvonen, J.J. Pelletier, H. Paulus, M.Q. Xu, Single-column purification of free recombinant proteins using a self-cleavable affinity tag derived from a protein splicing element, Gene 192 (1997) 277–281.

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

[47] J.C. Carrington, S.M. Cary, T.D. Parks, W.G. Dougherty, A second proteinase encoded by a plant polyvirus genome, Embo J. 8 (1989) 365–370.

[48] F. Baneyx, Recombinant protein expression in *Escherichia coli*, Curr. Opin. Biotechnol. 10 (1999) 411–421.

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

[50] A. Catanzariti, T.A. Soboleva, D.A. Jans, P.G. Board, R.T. Baker, An efficient system for high-level expression and easy purification of authentic recombinant proteins, Protein Sci. 13 (2004) 1331–1339.

[51] E.S. Johnson, Protein modification by SUMO, Annu. Rev. Biochem. 73 (2004) 355–382.

[52] M. Hochstrasser, Evolution and function of ubiquitin-like protein-conjugation systems, Nat. Cell Biol. 2 (2000) E153–E157.

[53] S. Jentsch, G. Pyrowolakis, Ubiquitin and its kin: how close are the family ties?, Trends Cell Biol. 10 (2000) 335–342.

[54] S. Muller, C. Hoege, G. Pyrowolakis, S. Jentsch, SUMO, ubiquitin's mysterious cousin, Nat. Rev. Mol. Cell Biol. 2 (2001) 202–210.

[55] E.S. Johnson, I. Schwienhorst, R.J. Dohmen, G. Blobel, The ubiquitin-like protein Smt3p is activated for conjugation to other proteins by an Aos1p/Uba2p heterodimer, Embo J. 16 (1997) 5509–5519.

[56] S. Muller, M.J. Matunis, A. Dejean, Conjugation with the ubiquitin-related modifier SUMO-1 regulates the partitioning of PML within the nucleus, Embo J. 17 (1998) 61–70.

[57] F. Melchior, SUMO—nonclassical ubiquitin, Annu. Rev. Cell Dev. Biol. 16 (2000) 591–626.

[58] M. Hochstrasser, SP-RING for SUMO: new functions bloom for a ubiquitin-like protein, Cell 107 (2001) 5–8.

[59] E.T. Yeh, L. Gong, T. Kamitani, Ubiquitin-like proteins: new wines in new bottles, Gene 248 (2000) 1–14.

[60] J.S. Steffan, N. Agrawal, J. Pallos, E. Rockabrand, L.C. Trotman, N. Slepko, K. Illes, T. Lukacsovich, Y.Z. Zhu, E. Cattaneo, P.P. Pandolfi, L.M. Thompson, J.L. Marsh, SUMO modification of Huntingtin and Huntington's disease pathology, Science 304 (2004) 100–104.

[61] Y. Shiio, R.N. Eisenman, Histone sumoylation is associated with transcriptional repression, Proc. Natl. Acad. Sci. USA 100 (2003) 13225–13230.

[62] G. Zheng, Y.C. Yang, ZNF76, a novel transcriptional repressor targeting TATA-binding protein, is modulated by sumoylation, J. Biol. Chem. 41 (2004) 42410–42421.

[63] M.J. Matunis, On the road to repair: PCNA encounters SUMO and ubiquitin modifications, Mol. Cell 10 (2002) 441–442.

[64] L.D. Wood, B.J. Irvin, G. Nucifora, K.S. Luce, S.W. Hiebert, Small ubiquitin-like modifier conjugation regulates nuclear export of TEL, a putative tumor suppressor, Proc. Natl. Acad. Sci. USA 100 (2003) 3257–3262.

[65] T.T. Huang, S.M. Wuerzberger-Davis, Z.H. Wu, S. Miyamoto, Sequential modification of NEMO/IKKgamma by SUMO-1 and ubiquitin mediates NF-κB activation by genotoxic stress, Cell 115 (2003) 565–576.

[66] L.L. Manza, S.G. Codreanu, S.L. Stamer, D.L. Smith, K.S. Wells, R.L. Roberts, D.C. Liebler, Global shifts in protein sumoylation in response to electrophile and oxidative stress, Chem. Res. Toxicol. 17 (2004) 1706–1715.

[67] S.J. Li, M. Hochstrasser, A new protease required for cell-cycle progression in yeast, Nature 398 (1999) 246–251.

[68] S.J. Li, M. Hochstrasser, The yeast ULP2 (SMT4) gene encodes a novel protease specific for the ubiquitin-like Smt3 protein, Mol. Cell. Biol. 20 (2000) 2367–2377.

[69] S.J. Li, M. Hochstrasser, The Ulp1 SUMO isopeptidase: distinct domains required for viability, nuclear envelope localization, and substrate specificity, J. Cell Biol. 160 (2003) 1069–1081.

[70] G. Murtas, P.H. Reeves, Y.F. Fu, I. Bancroft, C. Dean, G. Coupland, A nuclear protease required for flowering-time regulation in *Arabidopsis* reduces the abundance of small ubiquitin-related modifier conjugates, Plant Cell 15 (2003) 2308–2319.

[71] J. Cheng, D. Wang, Z. Wang, E.T. Yeh, SENP1 enhances androgen receptor-dependent transcription through desumoylation of histone deacetylase 1, Mol. Cell. Biol. 24 (2004) 6021–6028.

[72] T. Kadoya, H. Yamamoto, T. Suzuki, A. Yukita, A. Fukui, T. Michiue, T. Asahara, K. Tanaka, M. Asashima, A. Kikuchi, Desumoylation activity of Axam, a novel Axin-binding protein, is involved in downregulation of beta-catenin, Mol. Cell. Biol. 22 (2002) 3803–3819.

[73] J. Kurepa, J.M. Walker, J. Smalle, M.M. Gosink, S.J. Davis, T.L. Durham, D.-Y. Sung, R.D. Vierstra, The small ubiquitin-like modifier (SUMO) protein modification system in *Arabidopsis*: accumulation of SUMO-1 and -2 conjugates is increased by stress, J. Biol. Chem. 278 (2003) 6862–6872.

[74] X. Zuo, M.R. Mattern, R. Tan, S. Li, J. Hall, D.E. Sterner, J. Shoo, H. Tran, P. Lim, S.G. Sarafianos, L. Kazi, S. Navas-Martin, S.R. Weiss, T.R. Butt, Expression and purification of SARS Coronavirus proteins using SUMO fusions, Protein Express. Purif. (2005) in press, doi:10.1016/j.pep.2005.02.004.

[75] F. Schaefer, A. Schaefer, K. Steinert, A highly specific system for efficient removal of tags from recombinant proteins, J. Biomol. Techn. 13 (2002) 158–171.

[76] D. Reverter, C.D. Lima, A basis for SUMO protease specificity provided by analysis of human Senp2 and a Senp2–SUMO complex, Structure 12 (2004) 1519–1531.

[77] D.J. Ecker, T.R. Butt, J. Marsh, E. Sternberg, A. Shatzman, J.S. Dixon, P.L. Weber, S.T. Crooke, Ubiquitin function studied by disulfide engineering, J. Biol. Chem. 264 (1989) 1887–1893.

[78] X. Zuo, S. Li, J. Hall, M.R. Mattern, H. Tran, J. Shoo, R. Tan, S.R. Weiss, T.R. Butt, Enhanced expression and purification of membrane proteins by SUMO fusion in *Escherichia coli*, J. Struct. Funct. Genom. (2005) in press.

[79] G. MacBeath, S.L. Schreiber, Printing proteins as microarrays for high-throughput function determination, Science 289 (2000) 1760–1763.

[80] P. Mitchell, A perspective on protein microarrays, Nat. Biotechnol. 20 (2002) 225–229.

[81] H. Zhu, M. Bilgin, R Bangham, D. Hall, A. Casamayor, P. Bertone, N. Lan, R Jansen, S. Bidlingmaier, T. Houfek, T. Mitchell, P. Miller, R.A. Dean, M. Gerstein, M. Snyder, Global analysis of protein activities using proteome chips, Science 293 (2001) 2101–2105.