Published in final edited form as:

Protein Expr Purif. 2007 May ; 53(1): 40–50.

Ligation Independent Cloning Vectors for Expression of SUMO Fusions

Stephen D. Weeks¹, Mark Drinker², and Patrick J. Loll¹

*1*Department of Biochemistry & Molecular Biology, Drexel University College of Medicine, 245 N 15th Street, Mailstop 497, Philadelphia, PA 19102-1192, USA

*2*Division of Infectious Diseases, Children's Hospital of Philadelphia, Philadelphia, PA 19104, USA

Abstract

With demand increasing for the production of many different proteins for biophysical or biochemical analyses, rapid methods are needed for the cloning, expression and purification of native recombinant proteins. In particular, generic methods are required that are independent of the target gene sequence. To address this challenge we have constructed four *E. coli* expression vectors that can be used for ligation independent cloning (LIC) of an amplified target gene sequence. These vectors represent the combinatorial pairing of two different parent vector backbones with two different affinity tags. The target gene is cloned downstream of the sequence coding for an affinity-tagged small ubiquitin related modifier (SUMO). Using enhanced green fluorescent protein (eGFP) as an example we demonstrate that the LIC procedure works with high efficiency for all four of the vectors. We also show that the resultant recombinant SUMO-eGFP fusion can be overexpressed in *E. coli* and readily isolated by standard affinity purification techniques. Importantly, the purified fusion product can be treated with recombinant SUMO hydrolase to yield a mature target protein with any residue except proline at the amino terminus. We demonstrate an application of this by generating recombinant eGFP containing a non-native amino terminal cysteine residue and using it as a substrate for expressed protein ligation (EPL). The reagents and techniques described here represent a generic method for the rapid cloning and production of a target protein, and would be appropriate for a high throughput genomic scale expression project.

Keywords

Ligation independent cloning; LIC; SUMO; subtractive purification; expressed protein ligation; EPL; structural genomics

Introduction

The rapid and efficient production of recombinant proteins demands generally applicable “one size fits all” methodologies. Typically, however, production of a protein varies with its sequence: The nucleotide sequence limits which restriction endonucleases can be used during cloning, while the amino acid sequence determines the protein's physical properties and thereby dictates the purification procedure. Sequence-independent techniques have been developed

Corresponding Author: Patrick J. Loll, Department of Biochemistry & Molecular Biology, Drexel University College of Medicine, 245 N 15th Street, Mailstop 497, Philadelphia, PA 19102-1192, USA. Telephone: 215-762-7706. Fax: 215-762-4452. e-mail: pat.loll@drexel.edu

Publisher's Disclaimer: This is a PDF file of an unedited manuscript that has been accepted for publication. As a service to our customers we are providing this early version of the manuscript. The manuscript will undergo copyediting, typesetting, and review of the resulting proof before it is published in its final citable form. Please note that during the production process errors may be discovered which could affect the content, and all legal disclaimers that apply to the journal pertain.

that address cloning or purification individually, but examples of their successful combination are limited.

Traditionally a target gene is amplified by PCR using primers containing unique restriction sites. The PCR product is digested and then ligated into a vector containing compatible cohesive ends [1]. When a desired restriction site is already present within the target sequence, Type IIS restriction endonucleases can be used to produce the sticky ends [2], so long as their recognition sites are not present within the target sequence. Alternatively, a PCR-based strategy can be used that incorporates methylated bases in order to eliminate specific cutting sites within an amplified product [3]. Although these procedures allow for seamless directional cloning, the cloning steps are quite slow and the restriction endonucleases cleave the ends of PCR products with variable efficiency [4].

Flanking sequences other than restriction sites can also be used to facilitate cloning. In TA-cloning, single adenosines are added to the 3' ends of the amplified target and used to ligate the product into a vector containing the complementary 3' thymidine extensions [5]. While this method is not directional, correctly oriented clones can be selected using restriction endonuclease digestion [6]. However, the target sequences must not contain the selection restriction site.

Longer flanking extensions can be used for recombinogenic cloning [7-9], in which recombinases transfer the target into an acceptor plasmid. The target gene must be flanked by long regions homologous to the acceptor plasmid, and the PCR primers therefore contain extensions ranging from 29 to 50 bases in length. The efficiency of recombinogenic cloning is high, but the requirement for large homologous flanking regions can lead to the incorporation of non-native amino acids at one or both ends of the recombinant target protein [10,11].

Ligation independent cloning (LIC) adds flanking sequence extensions to the target gene that are longer than those used in TA-cloning, but shorter than those used in recombinogenic methods. These extensions are used to generate complementary single stranded overhangs ranging from 10 to 15 bases in length on the vector and insert. The two components are mixed and anneal to form a multiply-nicked but stable plasmid that can be directly transformed into *E. coli* [12,13]. Preparation of the overhangs requires that continuous stretches of sequence be found at the ends of both the insert and linearised plasmid that lack one of the four nucleotides. Treatment with a polymerase having 3-5' exonuclease activity in the presence of the dNTP corresponding to the missing base results in the formation of the long single stranded ends. This process depends only upon the sequence of the regions flanking the insert, and is otherwise independent of the sequences of the target gene and the vector. Hence, LIC can be used to clone any target gene into any site of a vector as long as the sequence requirements in the flanking regions are met.

Generic purification procedures that are applicable to wide range of different proteins typically rely on the use of affinity fusion partners such as His₆ tags, GST, or MBP [14-16]. Fusion partners can reduce the activity or hinder the crystallization of the target protein [17,18], and so it is prudent to introduce a protease cleavage site into the recombinant product that allows the partner's removal. Such cleavage sites include short linear epitopes such as those recognized by thrombin, enterokinase, factor Xa, and TEV NIa protease [15,16,19], as well as the sequences of entire proteins such as ubiquitin [20] or the related SUMO [21], which are recognized by specific hydrolases [22]. By using a particular affinity partner it is possible to generate (with variable efficiency) many different recombinant proteins using a single standardized purification protocol. Depending upon the protease recognition sequence used, however, these recombinant proteins may or may not contain the precise native sequence of the target.

Ideally a generic approach that standardizes both the cloning and purification steps of target gene expression should not compromise the sequence of the final product. In this work we have constructed LIC-adapted vectors for the generation of SUMO fusions that fulfill this requirement. SUMO fusions increase solubility of target proteins in *E. coli* [23] and can be easily matured using a recombinant protease [21]. This protease displays robust activity over a range of temperatures and in the presence of numerous additives. For these reasons, SUMO fusions are attractive vehicles for protein expression in structural genomics efforts [24]. We show that by using LIC it is possible to efficiently generate a SUMO fusion, independent of the target gene sequence. The relative ease of maturation of SUMO fusions allows the target protein to be isolated using a standardized purification procedure, and the properties of the SUMO hydrolase allow the target protein to be produced with a completely native sequence (so long as the N-terminal residue is not a proline; 21). In addition we demonstrate an application of this technology in which a protein with a non-native N-terminal residue is generated and used as a substrate for expressed protein ligation.

Materials and Methods

Enzymes, chemicals, media and strains

Enzymes required for the cloning steps were purchased from New England Biolabs (NEB, Ipswich, MA) unless otherwise stated. All reagents for media were bought from Fisher Scientific (Pittsburgh, PA). All chemicals were purchased from Sigma-Aldrich (St. Louis, MO) unless otherwise stated. PCR primers were purchased from MWG-biotech Inc. (High Point, NC). Six thymidine residues were included at the 5' ends of all primers to ensure efficient cleavage by restriction endonucleases; for the sake of clarity, these have been omitted from the sequences described below.

LB media was used for plasmid production and protein expression using the pASK-derived plasmids. For expression of pET-based constructs the auto-inducing media ZYP-5052 was used [25]. For plate selection ampicillin was used at 100 µg/ml and kanamycin at 30 µg/ml. For initial vector construction the bacterial strain DH5α (Invitrogen, Carlsbad, CA) was used. For cloning and expression of pASHSUL and pASS2SUL LIC products Mach1-T1^r (Invitrogen) was used. For the pETHSUL and pETS2SUL LIC products the strain Mach1-T1^r (DE3) was used. The latter strain was generated by treatment of Mach1-T1^r cells with the DE3 prophage using the DE3 lysogenisation kit (EMD Biosciences, Madison, WI). Both strains were made chemically competent for transformation using the method of Inoue *et al.* [26].

Construction of SUMO Fusion Expression Vectors

To generate the SUMO fusion LIC vectors two acceptor plasmids were constructed based on the pET and pASK architectures that have the same restriction sites bounding their multiple cloning regions. For the pET-based constructs a C-terminal His₆ tag vector, called pETCH, was created by amplification of a 427 bp fragment from pET21d (EMD biosciences) using the primers 5'-

GTCGACCCCGGGCACCACCACCACCACCTAATAAGCTTCTGCTAACAAAGC-3' and 5'-CGATGGCCCACTACGTGAAC-5'. The PCR product was digested with SalI and DraIII and cloned into pET21Δ (an in house vector containing a silent point mutation in the β-lactamase gene that eliminates the BsaI restriction site) that had been similarly digested. The vector pAST, containing the *tetA* promoter/operator region, was constructed by the insertion of the pET21d multiple cloning site between the XbaI and HindIII restriction sites, inclusive, into pASK-IBA2 (IBA, St. Louis, MO).

Tagged variants of the SUMO sequence were generated by a number of cloning steps. A His₆SUMO fusion sequence was generated by amplification of the *S. cerevisiae* strain RSY335

genomic DNA with 5'-CGTCTCTAGGTTCTGACTCCGAAGTCAATCAAGAAGCTAAG-3' and 5'-GGATCCGGTCTCACATGCCACCAATCTGTTTCGCGGT-3'. The PCR product was digested with BsmBI and BamHI and cloned into pET24His (in house vector containing a 5' His₆ sequence in the multiple cloning site) that had been digested with BsaI and BamHI, generating pET24HSU. The newly constructed plasmid was digested with XbaI and HindIII and the resultant 397 bp fragment cloned into pET21Δ to yield pETHSU.

To create a Strep-II tagged version of SUMO the primers 5'-TTCGAAAAATCTGACTCCGAAGTCAATCAAGAAGCTAAG-3' and 5'-GGATCCGGTCTCACATGCCACCAATCTGTTTCGCGGT-3' were used to amplify the SUMO sequence from *S. cerevisiae* RSY335 genomic DNA. The PCR product was digested with BstBI and BamHI and ligated into the equivalently digested pASK-IBA6 (IBA) to generate pASKIBA6SU. This vector was digested with NheI and BamHI and the 331 bp fragment, containing the tagged SUMO sequence, was subcloned into the same restriction sites of pET21Δ generating the vector pETS2SU.

The final LIC vectors were constructed by PCR using pETHSU and pETS2SU as templates with the primers 5'-ACAAATCCCCTCTAGAAATAATTTT-3' and 5'-AAGCTTCTCGAGGAGAGTTTAGACGCCACCAATCTGTTTCGCGGT-3'. The forward primer anneals to the region between the T7 promoter and the initiation codon of the T7lac plasmids. The reverse primer contains three restriction sites: HindIII, XhoI and BseRI. The latter enzyme is necessary for generation of the LIC compatible ends. The PCR products were digested with XbaI and HindIII and ligated into equivalently digested pETCH and pAST, to yield the four vectors: pETHSUL, pETS2SUL, pASHSUL and pASS2SUL. The integrity of the fusion sequence and multiple cloning site were ensured by sequencing.

Construction of dtUD1 expression vector and purification

A dual tagged clone of the catalytic domain (dtUD1) of the *S. cerevisiae* SUMO hydrolase was constructed as follows. The gene fragment, corresponding to residues 403 to 621 of the full-length protein [27], was amplified with the primers 5'-CCATGGGACTTGTTCCTGAATTAAATGAAAAAG-3' and 5'-TTTTTCCCGGGTTTAAAGCGTCGGTTAAATCAAA-3' using the *S. cerevisiae* RSY335 genomic DNA as a template. The PCR product was digested with NcoI and XmaI and ligated into the similarly digested pETCH, generating pSUPCH. pSUPCH was digested with NcoI and DraIII, and the resulting 1065 bp fragment was ligated into pETS2T (in house vector containing the sequence encoding for an amino terminal Strep-II tag followed by a TEV protease cleavage site) which had been digested with BsaI and HindIII yielding pSUPDT. The NcoI/DraIII fragment of pSUPDT, containing the double tagged hydrolase sequence, was then cloned into pETSU (in house vector containing an untagged yeast SUMO sequence) that had been digested with BsaI and HindIII creating pSUPER. The final vector, pSUPER, encodes a SUMO fusion of the dtUD1 that self cleaves when expressed in *E. coli*, resulting in two over-expressed bands when analysed by SDS-PAGE (data not shown).

The *E. coli* strain BL21-CodonPlus(DE3)-RIL (Stratagene, La Jolla, CA) was used for expression of dtUD1 from pSUPER. 5 ml of LB, containing 100 µg/ml ampicillin and 34 µg/ml chloramphenicol, was inoculated with a single transformed colony and allowed to grow at 37°C for 9 hours. 500 µl of the preculture was used to inoculate 1 litre of the auto-inducing media ZYP-5052 containing 100 µg/ml of ampicillin and 0.01 % V/V antifoam SE-15 in a 4 litre baffled flask. The culture was grown at 24°C for 20 hours. Cells were harvested by centrifugation at 1500 g and washed once with cold sterile water. The resultant pellet was resuspended in 2 times w/v LE buffer (50 mM sodium phosphate, 25 mM imidazole, 250 mM

NaCl, 10% w/v glycerol, pH 8.0). The slurry was stored at -80°C until ready to proceed with the next step.

The resuspended pellet was rapidly thawed and diluted a further 4 fold in LE buffer containing 5 mM β -mercaptoethanol (BME), 10 mM MgCl_2 and 1 $\mu\text{g}/\text{ml}$ each of DNase and RNase. Cells were lysed by three passes through an iced Emulsiflex-C5 homogenizer (Avestin, Ontario, Canada) at approximately 173 MPa backpressure. The lysate was clarified by centrifugation at 39,000 g for 30 minutes and then by ultracentrifugation at 257,000 g for 60 minutes, both at 4°C . The resultant supernatant was passed through a 0.45 micron MCE syringe filter (Fisher Scientific).

The dtUD1 was column purified using a 5 ml bed volume of nickel charged His-Bind agarose (EMD Biosciences) per litre of original culture. The supernatant was loaded at 38 cm/hr onto the column. The bound protein was washed with 20 column volumes (CV) of LE buffer plus 5 mM BME and then eluted in elution buffer (50 mM sodium phosphate, 300 mM imidazole, 250 mM NaCl, 10% w/v glycerol pH 8.0). The concentration of pooled fractions was determined and the protein was diluted to 4 mg/ml in dialysis buffer (20 mM Tris-HCl, 150 mM NaCl, 20 % w/v glycerol, 1 mM EDTA and 5 mM DL-dithiothreitol (DTT), pH 8.0). The protease was dialysed against two changes of dialysis buffer at 4°C before diluting to 0.5 mg/ml in storage buffer (20 mM Tris-HCl, 150 mM NaCl, 50 % glycerol v/v, 1 mM EDTA and 5 mM DTT, pH 8.0). Protein could be stored at either -20°C or -80°C for 6 months with no apparent loss of activity.

Ligation Independent Cloning

For the preparation of the plasmid two alternative methods were performed:

- i. Direct digestion: 2 μg of each vector was digested with 4 units of BseRI in a 50 μl reaction for 1.5 hours at 37°C . Five units of calf intestinal alkaline phosphatase (CIAP) were then added to each reaction and incubated for a further 30 minutes at 37°C . The digested DNA was purified by agarose gel electrophoresis and isolated using the Qiaquick gel extraction kit (Qiagen, Valencia, CA). The eluted DNA was incubated for 30 minutes at room temperature with 3 units of T4 DNA polymerase in the presence of 1 mM dCTP (Promega, Madison, WI). The reaction was stopped by incubating at 75°C for 20 minutes. The LIC-ready vectors were then stored at -20°C until ready for use.
- ii. PCR amplification: 2 μg of each vector was digested and CIAP treated as described above. The digested DNA was ethanol precipitated and resuspended in TE buffer (10 mM Tris-HCl, 1 mM EDTA pH 8.0). 100 ng of each digested vector was amplified in a 40 μl pfu turbo (Stratagene) PCR reaction using 5'-TCTAAACTCTCCTCGAGAAGCTTG-3' and 5'-CCACCAATCTGTTCGCGGTGA-3' as primers and reaction conditions suggested by the manufacturer. Following amplification the reaction was ethanol precipitated and the TE resuspended product treated with 5 units of DpnI for 30 minutes to digest the template DNA. The target product was purified by agarose gel electrophoresis and isolated using the Qiaquick gel extraction kit. The resulting product was treated with T4 DNA polymerase as described above.

To test the LIC vectors the eGFP sequence was amplified from pETCS2-eGFP (in house vector) with pfu turbo using the primers 5'-AGATTGGTGGCTGCGTGAGCAAGGGCGAGGAGC-3' and 5'-GAGGAGAGTTTAGACTATTATTTTTCGAAGTGC GG GTG-3'. The PCR product was purified by agarose gel electrophoresis and isolated using the Qiaquick gel extraction kit. The

resulting product was treated with T4 DNA polymerase as described above substituting the dCTP with 1 mM dGTP.

LIC was carried out as described for commercial LIC vectors (EMD Biosciences). Briefly, 0.001 pmol of the plasmid were mixed with 0.005 pmol of the eGFP insert. Following 5 minutes of incubation at room temperature, EDTA was added to each reaction at a final concentration of 5 mM. LIC reactions were transformed by the protocol of Inoue *et al.* [26], using Mach1-T1^r(DE3) for the pET-based vectors and Mach1-T1^r cells for the pAS-based vectors.

Efficiency of Ligation Independent Cloning Strategy

To determine the efficiency of the cloning strategy fluorescence was measured from liquid cultures expressing eGFP fusions. The protocol followed was dependent on the vectors used. Individual colonies from a plate of Mach1-T1^r(DE3) transformed with the pETHSUL and pETS2SUL LIC reactions were used to inoculate 1 ml of LB media containing 0.5% w/v glucose in 96 square deep well plates. For each plate 8 wells were inoculated with colonies transformed with the empty vector to act as negative controls. The plates were covered with Airpore tape (Qiagen) and placed in an orbital shaker at 30°C and 325 rpm for approximately 16 hours. Following overnight growth 20 µl of each well was used to inoculate 1 ml of ZYP-5052 media, in a fresh 96 square deep well plate. The new plate, again covered with Airpore tape, was placed in the orbital shaker at 30°C and 325 rpm and cultures were allowed to grow for 8 hours. Following growth 100 µl of each cell culture was transferred to a white 96 MicroWell plate (Nalge Nunc, Rochester, NY). The fluorescence properties of the individual liquid cultures were measured using the FluoroMax-3 spectrofluorometer with the Micromax microplate reader attachment (Jobin Yvon, Edison, NJ). An excitation and emission wavelength of 488 and 509 nm, respectively, were used with a 2nm bandpass. The signal was integrated for 1 second for each well.

For the Mach1-T1^r cells transformed with the pASHSUL and pASS2SUL LIC reactions, colonies were picked from a plate and grown overnight to saturation in LB in 96 square deep well plates covered with Airpore tape. 22 µl of each overnight culture was used to inoculate 1 ml of fresh LB media in a new 96 square deep well plate. The plate was covered and the cultures grown to an OD_{600nm} of ~0.5 at 30°C and 325 rpm (approximately 3 hours) at which point cells were induced with 200 µg/l anhydrotetracycline. The deep well plates were returned to the orbital shaker and, following 3 hours of induction, the fluorescence properties of the liquid cultures were measured in the same manner as for the pET based transformants.

Cloning, expression and purification of AtxQ30

The sequence encoding human ataxin-3 containing an internal stretch of 30 glutamines (AtxQ30) was amplified from the vector pETHM3c-AtxQ30 (in house vector) using the primers 5'-AGATTGGTGGCGGTATGGAGTCCATCTTCCACGAG-3' and 5'-GAGGAGAGTTTAGACTATTATGTCAGATAAAGTGTGAAGGT-3'. The PCR product was prepared for LIC as described above and cloned into pETS2SUL generating pETS2SUL-AtxQ30. The latter vector was transformed into BL21-CodonPlus(DE3)-RIL for protein expression. Culture and harvesting of cells was performed as described for dtUD1. Pelleted cells were stored at -80°C until ready for further workup.

A cell pellet, corresponding to 250 ml of the liquid culture, was rapidly thawed and diluted 10 fold w/v in sactinL buffer (50 mM Tris-HCl, 300 mM NaCl, 1 mM EDTA, 5 mM DTT pH 8.0) containing 10 mM MgSO₄ and 2 µg/ml each of DNase and RNase. Cells were lysed by two passes through a chilled Emulsiflex-C5 homogenizer at approximately 173 MPA. The lysate was clarified by centrifugation at 12,000 g and then ultracentrifugation at 311,000 g. The resultant supernatant was passed through a 5 micron PVDF syringe filter (Fisher Scientific).

The fusion protein was purified by loading the lysate, at 17 cm/h, onto a column containing 10 ml of Strep-Tactin Superflow resin (IBA, St. Louis, MO). Following loading the resin was washed with 2 column volumes of sactinL buffer at 17 cm/h. The bound protein was eluted at 34 cm/h using sactinE buffer (50 mM Tris-HCl, 300 mM NaCl, 1 mM EDTA, 5 mM DTT, 2.5 mM desthiobiotin pH 8.0).

The SUMO-AtxQ30 fusion protein was cleaved directly in the column elution buffer by the addition of 25 µg of dtUD1 to the column eluate. The reaction was incubated for 1 hour at 18°C resulting in complete cleavage of the fusion. The eluate was then dialysed against two changes of sactinD buffer (25 mM Tris-HCl, 300 mM NaCl, 1 mM EDTA, 2 mM DTT pH 8.0) at 4°C. The dialysed protein was reloaded at 17 cm/h onto the same 10 ml Strep-Tactin Superflow resin that had been regenerated using sactinR buffer (50 mM Tris-HCl, 300 mM NaCl, 1 mM EDTA, 5 mM DTT 4-hydroxyazobenzene-2-carboxylic acid pH 8.0) and then equilibrated in sactinD buffer. The flowthrough, containing the target AtxQ30 protein, was kept.

Purification of eGFP and MBPthioester

The constructed pETHSUL-eGFPS2 was transformed into BL21-CodonPlus(DE3)-RIL. Auto-inducing cultures were generated as described for dtUD1 expression. The pelleted cells were resuspended in two volumes of IMAC8 (25 mM TrisHCl, 500 mM NaCl, 8 mM imidazole 10% w/v glycerol, 10 mM BME pH 7.4) and stored at -80°C until ready for further workup.

A cell slurry, corresponding to a 500 mL culture, was thawed rapidly and diluted 4 fold further in IMAC8 containing 10 mM MgCl₂ and 1 µg/ml each of DNase and RNase. Cells were lysed by three passages through an iced Emulsiflex-C5 homogenizer at approximately 173 MPa. The lysate was clarified by centrifugation at 39,000 g followed by ultracentrifugation at 257,000 g for 60 minutes, both at 4°C. The resultant supernatant was then passed through a 0.45 micron MCE syringe filter.

The His₆ tagged SUMO-eGFP fusion was column purified using 5 ml of Chelating Sepharose 6 FF (GE Healthcare, Piscataway, NJ) charged with nickel. The supernatant was loaded at 38 cm/h. The resin was washed with 20 CV of IMAC8 and the bound fusion protein was eluted from the column at 76 cm/h using IMAC250 (25 mM TrisHCl, 500 mM NaCl, 250 mM imidazole, 10% w/v glycerol, 10 mM BME, pH 7.4).

EDTA was added to the eluate to a final concentration of 1mM. 50 µg of dtUD1 was then added and the cleavage reaction was allowed to proceed at 4°C for approximately 6 hours. The cleavage reaction was then dialysed against two changes of IMAC8 at 4°C. The dialysed protein was then reloaded onto the 5 ml column of Chelating Sepharose 6 FF at 38 cm/h. The flowthrough, containing eGFP, was kept.

The subtractively purified protein was concentrated using a YM10 Centriplus device (Millipore, Billerica, MA) and then further purified by gel filtration using an S-100 Sephacryl column (GE Healthcare) equilibrated in 20 mM MES, 500 mM NaCl pH 6.5. The protein was reconcentrated using a YM10 centriplus device to 45 mg/ml and then flash frozen in liquid nitrogen and stored at -80°C until required.

The MBP thioester was generated using the vector pMYB5 (NEB) containing the gene encoding for MBP in frame with the *S. cerevisiae* VMA intein. The vector was transformed into BL21(DE3) and a single colony was used in auto-inducing cultures as described above. Following harvest the pelleted cells were resuspended in two volumes of CLAW buffer (25 mM HEPES, 500 mM NaCl pH 7.4) and stored at -80°C until further workup.

A cell slurry, corresponding to a 1 litre culture, was rapidly thawed and diluted 4 fold in CLAW buffer containing 10 mM MgCl₂ and 1 µg/ml each of DNase and RNase. Cells were lysed and clarified exactly as described for the eGFP purification. The filtered supernatant was loaded onto a 20 ml chitin agarose column (NEB) at 21 cm/h and 4°C. Upon loading the column was washed with 10 CV of CLAW buffer followed by 2 CV of CE buffer (20 mM MES, 500 mM NaCl pH 6.0) and then 2 CV of CE buffer containing 50 mM sodium 2-mercaptoethane sulfonate (MESNA). The column was transferred to room temperature and left to stand for approximately 17 hours. The MBP-MESNA thioester product was eluted from the column using a further 2 CV of CE buffer. The eluate was concentrated to approximately 28 mg/ml using a YM30 Centrplus device and then flash frozen in liquid nitrogen and stored at -80°C until required.

Expressed Protein Ligation

The previously purified eGFP and MBP-thioester were thawed, mixed in a buffer containing 100 mM MESNA, 500mM NaCl, 100 mM Tris pH 8.0, and incubated at room temperature; final protein concentrations were 0.625 mM (17.5 mg/ml) for eGFP and 0.25 mM (10.7 mg/ml) for the MBP-thioester. Samples were removed from the reaction at different times and the reaction was stopped by the addition of 6X SDS buffer followed by freezing at -20°C.

Results

Vectors

Four LIC vectors have been constructed, derived from the combinatorial pairing of two different parent vector backbones with two different affinity tags (Table 1). The two parent vectors (pET and pASK) use different promoters to drive expression of the cloned fusion protein (Figure 1A). The two affinity tags are both small peptides and are incorporated upstream of the *S. cerevisiae* SUMO sequence (Figure 1B).

The pET vectors allow for the LacI-repressed but strongly inducible production of transcript by the highly processive T7 RNA polymerase [28], and must be used either with a lysogenic DE3 *E. coli* strain or exogenous polymerase provided by a recombinant phage [29]. The pASK vectors employ the tightly regulated *tetA* promoter [30], and are just over half the size of the pET-based constructs (Table 1). These vectors encode all the regulatory elements required for transcriptional control of the host RNA polymerase and therefore can be used in any strain of *E. coli*.

Ligation independent cloning

The LIC process is outlined in Figure 2. The vector is linearized and treated with T4 DNA polymerase in the presence of only dCTP to generate 10- and 14-base single strand overhangs (Figure 2A). To generate a LIC-compatible insert, a target gene is amplified using the following primers: 5'-AGATTGGTGGCN_x-3' (forward) and 5'-GAGGAGAGTTTAGACN_x-3' (reverse), where N_x denotes the target nucleotide sequence required for the specific annealing of the primers to the template. The PCR product is treated with T4 DNA polymerase in the presence of dGTP to generate the complimentary single stranded overhangs (Figure 2B). Note that the reverse primer should include a translation termination sequence such as the optimal TAAT sequence [31]; if the termination sequence is omitted, the resultant expressed protein will contain an additional valine residue at its carboxy end (Figure 2B). The treated vector and insert are simply mixed and transformed into *E. coli*, where the host naturally ligates the two fragments together.

As all four vectors constructed in this work contain the same LIC site (Figure 1B), they can all be prepared in an identical fashion. We compared two methods of vector preparation to

determine which yielded fewer false positive clones (Figure 2A). In both methods each vector was treated with the type IIS restriction endonuclease BseRI, which cuts downstream of the SUMO sequence (Figure 1B) to produce a two base pair 3' overhang (Figure 2A); the vector was then dephosphorylated using calf intestine alkaline phosphatase. In the first method, the digested plasmids were gel purified and treated with T4 DNA polymerase in the presence of dCTP, and then used directly for LIC. For the second method, we used PCR to amplify the linearized, dephosphorylated vectors as previously reported [12,13]. The resultant blunt-ended products were treated with DpnI to digest the template DNA, gel purified, and treated with T4 DNA polymerase in the presence of dCTP to generate the LIC-compliant overhangs.

To compare the efficiencies of the different methods of vector preparation, we developed a reporter assay using enhanced green fluorescent protein (eGFP). The 750 bp PCR product encoding eGFP was prepared for LIC, mixed with the treated plasmids, and transformed into competent cells. On average transformation resulted in 3.5×10^4 and 1×10^5 colonies per microgram of pET- and pASK-based vector DNA, respectively, with little difference between the two methods of vector preparation. Individual colonies were used to inoculate liquid cultures in a 96 deep well format. Expression of the SUMO-eGFP fusion was achieved using auto-inducing media and anhydrotetracycline induction for the pET- and pASK-based vectors, respectively. Background-corrected fluorescence levels were then determined for each culture. For selected cultures plasmid DNA was isolated and subjected to restriction analysis and protein expression was assessed by SDS-PAGE. High levels of fluorescence above background correlated well with both the presence of a restriction fragment corresponding to the eGFP insert and expression of the SUMO-eGFP fusion protein (data not shown).

The results from the reporter assay are summarized in Table 1, and show that for all vectors the LIC procedure is highly efficient, giving on average over 90% correct clones. In our hands similar values are obtained with different batches of prepared vectors and with different inserts, ranging in size from 180 to 2000 bp (data not shown). Little difference is observed in the cloning efficiency for the two methods of vector preparation. For the small vectors, pASHSUL and pASS2SUL, the additional PCR preparation step increased the number of positive clones by 6 to 8 %, presumably by minimizing contamination by the original uncut vector. Interestingly, for both pETHSUL and pETS2SUL, the additional PCR step actually reduced cloning efficiency. It is possible that PCR amplification of these large plasmids leads to incomplete extension, yielding products with single stranded ends that may self-anneal and then circularize upon transformation into the host strain.

Subtractive purification of SUMO fusions

The presence of either a small His₆ or Strep-II tag at the amino terminus of the SUMO sequence (Figure 1B) allows for the simple capture and purification of the recombinant fusion protein by affinity chromatography [32,33]. Sample purifications using each tag are shown in Figures 3 and 4.

To demonstrate affinity purification using StrepTactin resin the human polyglutamine protein ataxin-3 was used. Polyglutamine proteins are prone to aggregation [34], and prokaryotic expression of active ataxin-3 requires care [35]. A version of the protein containing a 30 residue polyglutamine stretch (AtxQ30) was cloned by LIC into pETS2SUL. Upon induction the fusion construct accumulates to high levels, with the majority of the protein being found in the soluble fraction (Figure 3, Lanes 1 and 2). The SUMO-AtxQ30 fusion was captured directly from the clarified lysate using a StrepTactin column, and could be readily eluted by desthiobiotin (Figure 3, Lanes 3 and 4). In initial purification trials it was found that fusion constructs containing Strep-II tags slowly leached from the column during prolonged wash steps (data not shown); therefore, to maximize yield, after loading only two column volumes of wash buffer were applied before elution. Even with this low volume wash the eluted Strep-II SUMO-AtxQ30 is

greater than 90 % pure (Figure 3, Lane 4) and concentrated approximately 10 fold. A similar result was observed for the IMAC purification of His₆ tagged SUMO-eGFP expressed from a pETHSUL vector and is shown in Figure 4A (Lanes 1 to 4). Following binding to a nickel resin and extensive washing, the fusion protein eluted by imidazole is about 90% pure (Figure 3A, Lane 4) and typically concentrated 100 fold.

In the process of making the LIC vectors we also constructed a vector for the overexpression of the UD1 domain of the *S. cerevisiae* Ulp1 protein, using published domain boundaries [27]. This domain retains the full SUMO-specific proteolytic activity. The recombinant protease has an amino terminal Strep-II tag and a carboxy terminal His₆ tag, and is dubbed dtUD1 (doubly tagged UD1). These tags allow for easy purification of the recombinant protease; they also allow the protease to be removed after cleavage of the fusion protein, using subtractive affinity chromatography with either an IMAC or an immobilized StrepTactin resin. Examples of subtractive purification are shown for both the Strep-II tagged SUMO-AtxQ30 and His₆ tagged SUMO-eGFP fusions (Figures 3, Lanes 6-9 and 4A, Lanes 5-7). Addition of dtUD1 to either of the purified SUMO fusion constructs results in almost complete cleavage of the chimeras following incubation at 4° C for 16 hours (Figures 3, Lane 6 and 4A, Lane 5). After removal of desthiobiotin or imidazole by dialysis the cleavage reaction was reapplied to the original affinity resin. In both cases the uncleaved fusion protein, tagged SUMO, and dtUD1 remain bound to the resin along with other protein impurities carried over from the initial affinity purification (Figures 3, Lane 9 and 4A, Lane 6), and the flow-through fractions contain the purified target AtxQ30 and eGFP proteins (Figures 3, Lane 8 and 4A, Lane 7). Comparable results have been observed for other fusions using each of the two affinity tags (data not shown).

Expressed protein ligation

Using the LIC strategy and the PCR primers described above, the first codon of the amplified target is placed directly after the terminal glycine codon of the SUMO sequence and can encode any amino acid (Figure 2B). dtUD1 cleaves the fusion protein directly after this glycine (Figure 1B), and is capable of cleaving any Gly-X bond except where X = Pro [21]. Therefore, the recombinant protein product can be engineered to have any residue present at its amino terminus except proline. We exploited this capability to generate a version of eGFP containing a non-native cysteine at the amino terminus (cys-eGFP) and used it as a substrate for expressed protein ligation (EPL; [36]). Following expression and subtractive affinity chromatography the cys-eGFP was further purified by gel filtration to yield a product suitable for EPL (Figure 4B, Lane 2). A 2.5 molar excess of the purified cys-eGFP was mixed with purified MBP-thioester and incubated at room temperature. Within seconds of mixing the MBP-eGFP fusion protein appeared (compare control lanes 1 and 2 with lane 3 of Figure 4B). After a 24 hour incubation at room temperature, approximately 66% of the MBP-thioester was converted to the MBP-eGFP fusion (Figure 4B, Lane 4).

Discussion

We have designed a generic approach to recombinant protein production that combines the simplicity of LIC [12] with the strength of the SUMO fusion system [21]. The procedure allows for the systematic cloning of target genes, independent of their sequence, and provides a platform for the standardized purification of recombinant proteins having either native or engineered amino terminal residues. The procedures described here are equally applicable in small laboratory settings focusing on a few protein targets and in environments requiring high-throughput protein production.

We chose to create two sets of expression vectors derived from two different parent vectors (Figure 1A). One set uses the popular T7 promoter to drive high level protein expression in DE3 lysogenic strains of *E. coli* [29]. The second set use the promoter/operator region of the

Tn10 tetracycline resistance gene *tetA* to drive expression [30]. Although the latter vectors may yield lower levels of expression than those using the T7 promoter, they can be used in any strain of *E. coli*. This may prove useful if the effect of different *E. coli* genotypes on the yield or solubility of product is being studied [37,38], or if auxotrophic strains are being used to label proteins with isotopically enriched residues for NMR studies [39].

Cloning into the LIC-based vectors described here is totally independent of the target sequence, therefore it offers an advantage over a previously reported method for generating SUMO fusions that uses type IIS restriction enzymes to generate compatible sticky ends on both the vector and insert [21]. The LIC method requires relatively modest primer extensions that are comparable in size to those used in RF cloning [40] and considerably smaller than the extensions required for cloning into Gateway vectors [11]

Preparation of vector and amplified insert for LIC is a simple process requiring a small number of readily available enzymes. In our hands the LIC procedure results in approximately 90% correct clones (Table 1), a value that compares favorably with success rates recently reported for LIC vectors containing the TEV protease cleavage sequence [41,42]. We could not match the 100% cloning efficiency of earlier LIC reports [12,13], even when amplifying the linearised plasmid by PCR as suggested. However, 90% efficiency should be ample even in a structural genomics setting, since duplicate cloning trials would give greater than 98% coverage of the chosen targets.

The combination of the affinity tags with the SUMO fusion partner allows the target protein to be isolated using subtractive affinity chromatography, a procedure commonly used in structural genomics initiatives [41,42]. The SUMO protease used to cleave the fusion is active over a range of temperatures and in the presence of numerous additives [21] and so can be used in cases where a target protein is prone to aggregation.

The LIC SUMO fusion vectors described here allow recombinant proteins to be produced with any desired sequence at the amino terminus, so long as the initial residue is not proline. In contrast, previously described LIC and Gateway fusion vectors contribute extra residues at the amino terminus that cannot be avoided without re-designing the vector [41-43]. The freedom to dictate the N-terminal sequence offered by the LIC SUMO vectors is likely to prove useful in cases where the amino terminus of the protein is critical for activity [44,45], or when the N-terminal sequence is to be engineered for a specific application like EPL. To date, protein substrates containing N-terminal cysteines for use in EPL have been generated either by using inteins or by proteolytic cleavage of fusions [36]. Premature cleavage can occur with intein fusions [46], which leads to the loss of the affinity tag used for purification; and commonly employed proteases, such as Factor Xa, show variable efficiency and can sometimes cleave at secondary sites [19]. As the SUMO protease cleaves the fusion efficiently and with no unwanted proteolytic degradation, SUMO fusions seem ideal for producing EPL substrates having N-terminal cysteines.

It is common to compare different fusion protein partners to determine which provides best expression and solubility for a given target [11,41]. This option is not available with the LIC procedure described here, which works specifically with the SUMO sequence. SUMO has been reported to increase solubility of at least some fused partners [23], but it is unclear whether it will prove as useful in this respect as other proteins such as MBP or NusA. However, since SUMO is only 98 amino acids long it is easy to envision fusing these other proteins upstream of SUMO to determine if they improve solubility. Similarly, a signal peptide sequence could be added to the amino terminus of the protein to direct secretion of the SUMO fusion into the periplasm and permit disulphide bond formation.

Acknowledgements

The authors would like to thank Kimberly Grasty for technical support. This work was supported in part by NIH grants GM64676 & MH73060 (P. J. L.) and by the Fanny Ripple Foundation.

Abbreviations and symbols

BME, 2-mercaptoethanol; DTT, DL-Dithiothreitol; eGFP, enhanced green fluorescent protein; EPL, expressed protein ligation; LIC, ligation independent cloning; MESNA, sodium 2-mercaptoethane sulfonate; PCR, polymerase chain reaction; SUMO, small ubiquitin-related modifier.

References

1. Sambrook, J.; Russell, DW. *Molecular Cloning: A Laboratory Manual*. Cold Spring Harbor Laboratory Press; Cold Spring Harbor, NY: 2001.
2. Szybalski W, Kim SC, Hasan N, Podhajski AJ. Class-II restriction enzymes--a review. *Gene* 1991;100:13–26. [PubMed: 2055464]
3. Padgett KA, Sorge JA. Creating seamless junctions independent of restriction sites in PCR cloning. *Gene* 1996;168:31–35. [PubMed: 8626061]
4. Zimmermann K, Schögl D, Mannhalter JW. Digestion of terminal restriction endonuclease recognition sites on PCR products. *Biotechniques* 1998;24:582–584. [PubMed: 9564528]
5. Marchuk D, Drumm M, Saulino A, Collins FS. Construction of T-vectors, a rapid and general system for direct cloning of unmodified PCR products. *Nucleic Acids Res* 1991;19:1154. [PubMed: 2020552]
6. Goda N, Tenno T, Takasu H, Hiroaki H, Shirakawa M. The PRESAT-vector: asymmetric T-vector for high-throughput screening of soluble protein domains for structural proteomics. *Protein Sci* 2004;13:652–658. [PubMed: 14978305]
7. Hartley JL, Temple GF, Brasch MA. DNA cloning using in vitro site-specific recombination. *Genome Res* 2000;10:1788–1795. [PubMed: 11076863]
8. Liu Q, Li MZ, Leibham D, Cortez D, Elledge SJ. The univector plasmid-fusion system, a method for rapid construction of recombinant DNA without restriction enzymes. *Curr. Biol* 1998;8:1300–1309. [PubMed: 9843682]
9. Muirers JP, Zhang Y, Stewart AF. Techniques: Recombinogenic engineering-new options for cloning and manipulating DNA. *Trends Biochem. Sci* 2001;26:325–331. [PubMed: 11343926]
10. Brasch MA, Hartley JL, Vidal M. ORFeome cloning and systems biology: standardized mass production of the parts from the parts-list. *Genome Res* 2004;14:2001–2009. [PubMed: 15489318]
11. Busso D, Delagoutte-Busso B, Moras D. Construction of a set Gateway-based destination vectors for high-throughput cloning and expression screening in *Escherichia coli*. *Anal. Biochem* 2005;343:313–321. [PubMed: 15993367]
12. Aslanidis C, de Jong PJ. Ligation-independent cloning of PCR products (LIC-PCR). *Nucleic Acids Res* 1990;18:6069–6074. [PubMed: 2235490]
13. Aslanidis C, de Jong PJ, Schmitz G. Minimal length requirement of the single-stranded tails for ligation-independent cloning (LIC) of PCR products. *PCR Methods Appl* 1994;4:172–177. [PubMed: 7580902]
14. Nilsson J, Ståhl S, Lundberg J, Uhlén M, Nygren P-Å. Affinity fusion strategies for detection, purification, and immobilization of recombinant proteins. *Protein Expr. Purif* 1997;11:1–16. [PubMed: 9325133]
15. Terpe K. Overview of tag protein fusions: from molecular and biochemical fundamentals to commercial systems. *Appl. Microbiol. Biotechnol* 2003;60:523–533. [PubMed: 12536251]
16. Waugh DS. Making the most of affinity tags. *Trends Biotechnol* 2005;23:316–320. [PubMed: 15922084]
17. Bucher MH, Evdokimov AG, Waugh DS. Differential effects of short affinity tags on the crystallization of *Pyrococcus furiosus* maltodextrin-binding protein. *Acta Crystallogr. D. Biol. Crystallogr* 2002;58:392–397. [PubMed: 11856823]

18. Jenkins JL, Tanner JJ. High-resolution structure of human D-glyceraldehyde-3-phosphate dehydrogenase. *Acta Crystallogr. D. Biol. Crystallogr* 2006;62:290–301. [PubMed: 16510976]
19. Jenny RJ, Mann KG, Lundblad RL. A critical review of the methods for cleavage of fusion proteins with thrombin and factor Xa. *Protein Expr. Purif* 2003;31:1–11. [PubMed: 12963335]
20. Catanzariti AM, Soboleva TA, Jans DA, Board PG, Baker RT. An efficient system for high-level expression and easy purification of authentic recombinant proteins. *Protein Sci* 2004;13:1331–1339. [PubMed: 15096636]
21. Malakhov MP, Mattern MR, Malakhova OA, Drinker M, Weeks SD, Butt TR. SUMO fusions and SUMO-specific protease for efficient expression and purification of proteins. *J. Struct. Funct. Genomics* 2004;5:75–86. [PubMed: 15263846]
22. Amerik AY, Hochstrasser M. Mechanism and function of deubiquitinating enzymes. *Biochim. Biophys. Acta* 2004;1695:189–207. [PubMed: 15571815]
23. Marblestone JG, Edavettal SC, Lim Y, Lim P, Zuo X, Butt TR. Comparison of SUMO fusion technology with traditional gene fusion systems: enhanced expression and solubility with SUMO. *Protein Sci* 2006;15:182–189. [PubMed: 16322573]
24. Chance MR, Bresnick AR, Burley SK, Jiang JS, Lima CD, Sali A, Almo SC, Bonanno JB, Buglino JA, Boulton S, Chen H, Eswar N, He G, Huang R, Ilyin V, McMahan L, Pieper U, Ray S, Vidal M, Wang LK. Structural genomics: a pipeline for providing structures for the biologist. *Protein Sci* 2002;11:723–738. [PubMed: 11910018]
25. Studier FW. Protein production by auto-induction in high density shaking cultures. *Protein Expr. Purif* 2005;41:207–234. [PubMed: 15915565]
26. Inoue H, Nojima H, Okayama H. High efficiency transformation of *Escherichia coli* with plasmids. *Gene* 1990;96:23–28. [PubMed: 2265755]
27. Mossesso E, Lima CD. Ulp1-SUMO crystal structure and genetic analysis reveal conserved interactions and a regulatory element essential for cell growth in yeast. *Mol. Cell* 2000;5:865–876. [PubMed: 10882122]
28. Studier FW, Rosenberg AH, Dunn JJ, Dubendorff JW. Use of T7 RNA polymerase to direct expression of cloned genes. *Methods Enzymol* 1990;185:60–89. [PubMed: 2199796]
29. Studier FW, Moffatt BA. Use of bacteriophage T7 RNA polymerase to direct selective high-level expression of cloned genes. *J. Mol. Biol* 1986;189:113–130. [PubMed: 3537305]
30. Skerra A. Use of the tetracycline promoter for the tightly regulated production of a murine antibody fragment in *Escherichia coli*. *Gene* 1994;151:131–135. [PubMed: 7828861]
31. Poole ES, Brown CM, Tate WP. The identity of the base following the stop codon determines the efficiency of in vivo translational termination in *Escherichia coli*. *EMBO J* 1995;14:151–158. [PubMed: 7828587]
32. Porath J, Carlsson J, Olsson I, Belfrage G. Metal chelate affinity chromatography, a new approach to protein fractionation. *Nature* 1975;258:598–599. [PubMed: 1678]
33. Voss S, Skerra A. Mutagenesis of a flexible loop in streptavidin leads to higher affinity for the Strep-tag II peptide and improved performance in recombinant protein purification. *Protein Eng* 1997;10:975–982. [PubMed: 9415448]
34. Chow MK, Ellisdson AM, Cabrita LD, Bottomley SP. Purification of polyglutamine proteins. *Methods Enzymol* 2006;413:1–19. [PubMed: 17046388]
35. Bevivino AE, Loll PJ. An expanded glutamine repeat destabilizes native ataxin-3 structure and mediates formation of parallel beta -fibrils. *Proc. Natl. Acad. Sci. U. S. A* 2001;98:11955–11960. [PubMed: 11572942]
36. Muir TW. Semisynthesis of proteins by expressed protein ligation. *Annu. Rev. Biochem* 2003;72:249–289. [PubMed: 12626339]
37. Makrides SC. Strategies for achieving high-level expression of genes in *Escherichia coli*. *Microbiol. Rev* 1996;60:512–538. [PubMed: 8840785]
38. Winstone TL, Duncalf KA, Turner RJ. Optimization of expression and the purification by organic extraction of the integral membrane protein EmrE. *Protein Expr. Purif* 2002;26:111–121. [PubMed: 12356478]

39. Muchmore DC, McIntosh LP, Russell CB, Anderson DE, Dahlquist FW. Expression and nitrogen-15 labeling of proteins for proton and nitrogen-15 nuclear magnetic resonance. *Methods Enzymol* 1989;177:44–73. [PubMed: 2691846]
40. Geiser M, Cebe R, Drewello D, Schmitz R. Integration of PCR fragments at any specific site within cloning vectors without the use of restriction enzymes and DNA ligase. *Biotechniques* 2001;31:88–90. [PubMed: 11464525]
41. Cabrita LD, Dai W, Bottomley SP. A family of *E. coli* expression vectors for laboratory scale and high throughput soluble protein production. *BMC Biotechnol* 2006;6:12. [PubMed: 16509985]
42. Stols L, Gu M, Dieckman L, Raffin R, Collart FR, Donnelly MI. A new vector for high-throughput, ligation-independent cloning encoding a tobacco etch virus protease cleavage site. *Protein Expr. Purif* 2002;25:8–15. [PubMed: 12071693]
43. Haun RS, Moss J. Ligation-independent cloning of glutathione S-transferase fusion genes for expression in *Escherichia coli*. *Gene* 1992;112:37–43. [PubMed: 1339364]
44. Gohara DW, Ha CS, Kumar S, Ghosh B, Arnold JJ, Wisniewski TJ, Cameron CE. Production of “authentic” poliovirus RNA-dependent RNA polymerase (3D^{pol}) by ubiquitin-protease-mediated cleavage in *Escherichia coli*. *Protein Expr. Purif* 1999;17:128–138. [PubMed: 10497078]
45. Kirkpatrick RB, McDevitt PJ, Matico RE, Nwagwu S, Trulli SH, Mao J, Moore DD, Yorke AF, McLaughlin MM, Knecht KA, Elefante LC, Calamari AS, Fornwald JA, Trill JJ, Jonak ZL, Kane J, Patel PS, Sathe GM, Shatzman AR, Tapley PM, Johanson KO. A bicistronic expression system for bacterial production of authentic human interleukin-18. *Protein Expr. Purif* 2003;27:279–292. [PubMed: 12597888]
46. David R, Richter MP, Beck-Sickinger AG. Expressed protein ligation. Method and applications. *Eur. J. Biochem* 2004;271:663–677. [PubMed: 14764082]

Figure 1.

Map and details of the cloning sites for the LIC-SUMO plasmids. (A) Map of pET and pASK based plasmids showing the topology of the encoded sequences (B) The sequence of the common cloning site found in the pET- and pASK-based vectors. Two alternative affinity tags are available for each plasmid backbone, either a hexa-histidine (His₆) or a Strep-II sequence (boxed region). The tags are encoded in the same reading frame with the amino terminus of the yeast SUMO sequence. Cleavage of fusions with the Sumo-specific protease (dtUDI) occurs on the carboxyl side of the terminal glycine of the Sumo sequence, at the position marked by the arrow.

Figure 2.

Flowchart describing the method used to generate long single stranded overhangs for LIC. (A) Preparation of the plasmid. Each plasmid is digested with BseRI and then treated with calf intestine alkaline phosphatase (CIAP) to minimize self ligation. Following restriction digestion the resulting product can be directly treated with T4 DNA polymerase in the presence of dCTP to generate the single stranded overhangs. Alternatively the digested product can first be amplified by PCR and then treated with T4 DNA polymerase; the PCR will yield a blunt-ended product lacking the underlined bases. (B) Preparation of the insert. The target sequence is amplified by PCR using primers containing the underlined 5' extensions. The resultant product is treated with T4 DNA polymerase in the presence of dGTP to generate single stranded ends

complementary to those in the plasmid. As the forward primer extension encodes the C-terminal residues of SUMO (shown above sequence), the final base of which terminates the T4 DNA polymerase exonuclease activity in the presence of dGTP, the first codon (nnn) of the amplified target can encode any residue.

Figure 3. 12% SDS-PAGE Coomassie stained gels showing subtractive purification of AtxQ30. Lane 1, Cell lysate; Lane 2, Supernatant following clarification by centrifugation; Lane 3, StrepTactin column flowthrough; Lane 4, desthiobiotin elution; Lane 5, dtUD1 catalysed cleavage reaction; Lane 6, Dialysed cleavage reaction; Lane 7, Subtracted flowthrough from StrepTactin column; Lane 8, Desthiobiotin elution following second pass through the StrepTactin column. Lanes 4 to 8 are loaded at twice the relative concentration of the cell lysate sample.

Figure 4. 10-20% SDS-PAGE Coomassie stained gels showing subtractive purification and expressed protein ligation (EPL) of eGFP. Position of molecular weight markers are shown (kDa units). (A) Subtractive purification of His₆SUMO fusion of eGFP. Lane 1, Cleared cell lysate; Lane 2, Nickel column flowthrough; Lane 3, 7.5mM imidazole wash; Lane 4, 250 mM imidazole elution; Lane 5, dtUD1 catalysed cleavage reaction; Lane 6, 250 mM Imidazole elution following second nickel column pass; Lane 7, Subtracted flowthrough from nickel column. (B) Ligation of eGFP containing an amino terminal free cysteine to MBP. Lane 1, Purified MBP-thioester; Lane 2, Purified eGFP; Lane 3, EPL reaction 0 hrs; Lane 4, EPL reaction 24 hrs.

Table 1
Vector properties and comparison of LIC efficiency

Vector ^a	Size (bp)	Preparation ^b	Positive clones (%) ^c
pETHSUL	5677	PCR	92.0
		Restriction Digestion	93.2
pETS2SUL	5680	PCR	86.4
		Restriction Digestion	92.0
pASHSUL	3473	PCR	94.3
		Restriction Digestion	88.6
pASS2SUL	3476	PCR	94.3
		Restriction Digestion	85.2

^aVector naming convention: The first two letters are derived from the parent pET or pASK vector names. They are followed by a code identifying the affinity tag attached to SUMO's N-terminus, where "H" represents a hexa-histidine tag and "S2" a Strep-II tag. The letters "SUL" stand for SUMO LIC vector.

^bMethod used to prepare linearised vectors for LIC cloning: Either restriction digestion followed by PCR amplification, or restriction digestion alone.

^cDetermined from measuring the fluorescence properties of individual cultures for 88 clones exciting at 488 nm. A positive clone was identified as such if the emission value at 509 nm, in counts per second, was greater than 50% of the value of the difference between the average of 8 negative control cultures and median value of the test cultures.