(12) INTERNATIONAL APPLICATION PUBLISHED UNDER THE PATENT COOPERATION TREATY (PCT)

(19) World Intellectual Property Organization
International Bureau

(43) International Publication Date
17 July 2003 (17.07.2003)

PCT

(10) International Publication Number
WO 03/057174 A2

(51) International Patent Classification[7]: A61K

(21) International Application Number: PCT/US03/00436

(22) International Filing Date: 7 January 2003 (07.01.2003)

(25) Filing Language: English

(26) Publication Language: English

(30) Priority Data:
60/346,449 7 January 2002 (07.01.2002) US

(71) Applicant *(for all designated States except US)*: **LIFE-SENSORS, INC.** [US/US]; 271 Great Valley Parkway, Malvern, PA 19356 (US).

(72) Inventors; and

(75) Inventors/Applicants *(for US only)*: **BUTT, Tauseef, R.** [US/US]; 1020 Sanderling Circle, Audubon, PA 19403 (US). **WEEKS, Steven, D.** [GB/US]; 1828 Pine Street, Apt. 3F, Philadelphia, PA 19103 (US). **TRAN, Hiep, T.** [VN/US]; 1180 Queen Lane, Apt. 03, West Chester, PA 19382 (US). **MALAKHOV, Michael, P.** [RU/US]; 199 Canterburg Court, West Chester, PA 19380 (US). **MALAKHOVA, Oxana, A.** [RU/US]; 199 Canterburg Court, West Chester, PA 19380 (US).

(74) Agents: **RIGAUT, Kathleen, D.** et al.; Dann, Dorfman, Herrell and Skillman, Suite 720, 1601 Market Street, Philadelphia, PA 19103 (US).

(81) Designated States *(national)*: AE, AG, AL, AM, AT, AU, AZ, BA, BB, BG, BR, BY, BZ, CA, CH, CN, CO, CR, CU, CZ, DE, DK, DM, DZ, EC, EE, ES, FI, GB, GD, GE, GH, GM, HR, HU, ID, IL, IN, IS, JP, KE, KG, KP, KR, KZ, LC, LK, LR, LS, LT, LU, LV, MA, MD, MG, MK, MN, MW, MX, MZ, NO, NZ, OM, PH, PL, PT, RO, RU, SC, SD, SE, SG, SK, SL, TJ, TM, TN, TR, TT, TZ, UA, UG, US, UZ, VC, VN, YU, ZA, ZM, ZW.

(84) Designated States *(regional)*: ARIPO patent (GH, GM, KE, LS, MW, MZ, SD, SL, SZ, TZ, UG, ZM, ZW), Eurasian patent (AM, AZ, BY, KG, KZ, MD, RU, TJ, TM), European patent (AT, BE, BG, CH, CY, CZ, DE, DK, EE, ES, FI, FR, GB, GR, HU, IE, IT, LU, MC, NL, PT, SE, SI, SK, TR), OAPI patent (BF, BJ, CF, CG, CI, CM, GA, GN, GQ, GW, ML, MR, NE, SN, TD, TG).

Published:
— *without international search report and to be republished upon receipt of that report*

For two-letter codes and other abbreviations, refer to the "Guidance Notes on Codes and Abbreviations" appearing at the beginning of each regular issue of the PCT Gazette.

WO 03/057174 A2

(54) Title: METHODS AND COMPOSITIONS FOR PROTEIN EXPRESSION AND PURIFICATION

(57) Abstract: Methods for enhancing expression levels and secretion of heterologous fusion proteins in a host cell are disclosed.

METHODS AND COMPOSITIONS FOR PROTEIN EXPRESSION AND PURIFICATION

By Tauseef R. Butt
Stephen D. Weeks
Hiep T. Tran
Oxana A. Malakhova
Michael P. Malakhov

CROSS REFERENCE TO RELATED APPLICATION

This application claims priority to US Provisional Application 60/346,449 entitled "Methods for Protein Expression and Purification" filed January 7, 2002, the entire disclosure of which is incorporated by reference herein.

FIELD OF THE INVENTION

The present invention relates to the field of recombinant gene expression and purification of expressed proteins. More specifically, the invention provides materials and methods which facilitate purification of heterologous proteins from a variety of different host species.

BACKGROUND OF THE INVENTION

Several publications and patent documents are cited throughout the specification in order to describe the state of the art to which this invention pertains. Full citations for these references can be found at the end of the specification. Each of these citations is incorporated herein as though set forth in full.

Functional genomic studies have been hampered by the inability to uniformly express and purify biologically active proteins in heterologous expression systems. Despite the use of identical transcriptional and translational signals in a given expression vector, expressed protein levels have been observed to vary

dramatically (5, 7). For this reason, several strategies have been developed to express heterologous proteins in bacteria, yeast, mammalian and insect cells as gene-fusions.

The expression of heterologous genes in bacteria is by far the simplest and most inexpensive means available for research or commercial purposes. However, some heterologous gene products fail to attain their correct three-dimensional conformation in *E. coli* while others become sequestered in large insoluble aggregates or "inclusion bodies" when overproduced. Major denaturant-induced solubilization methods followed by removal of the denaturant under conditions that favor refolding are often required to produce a reasonable yield of the recombinant protein. Selection of ORFs for structural genomics projects has also shown that only about 20% of the genes expressed in *E.coli* render proteins that were soluble or correctly folded (36, 38). These numbers are startlingly disappointing especially given that most scientists rely on *E.coli* for initial attempts to express gene products. Several gene fusion systems such as NUS A, maltose binding protein (MBP), glutathione S transferase (GST), and thioredoxin (TRX) have been developed (17). All of these systems have certain drawbacks, ranging from inefficient expression to inconsistent cleavage from desired structure. Comprehensive data showing that a particular fusion is best for a certain family of proteins is not available.

Ubiquitin and ubiquitin like proteins (UBLs) have been described in the literature. The SUMO system has also been characterized. SUMO (small ubiquitin related modifier) is also known as Sentrin, SMT3, PIC1, GMP1 and UBL1. SUMO and the SUMO pathway are present throughout the eukaryotic kingdom and the proteins are highly conserved from yeast to humans (12, 15, 28). SUMO homologues have also been identified in C.elegans and plants. SUMO has 18 % sequence identity with ubiquitin (28, 39). Yeast has only a single SUMO gene, which has also been termed *SMT3* (23, 16). The yeast Smt3 gene is essential for viability (29). In contrast to yeast, three members of SUMO have been described in vertebrates: SUMO-1 and close homologues SUMO-2 and SUMO-3. Human SUMO-1, a 101 amino-acid polypeptide, shares 50% sequence identity with human SUMO-1/SUMO-2 (29). Yeast SUMO (SMT3) shares 47 % sequence identity with

mammalian SUMO-1. Although overall sequence homology between ubiquitin and SUMO is only 18%, structure determination by nuclear magnetic resonance (NMR) reveals that the two proteins share a common three dimensional structure that is characterized by a tightly packed globular fold with β-sheets wrapped around one α-helix(4). Examination of the chaperoning properties of SUMO reveals that attachment of a tightly packed globular structure to N-termini of proteins can act as nucleus for folding and protect the labile protein. All SUMO genes encode precursor proteins with a short C-terminal sequence that extends from the conserved C-terminal Gly-Gly motif. The extension sequence, 2-12 amino acids in length, is different in all cases. Cells contain potent SUMO proteases that remove the C-terminal extensions. The C-terminus of SUMO is conjugated to ε amino groups of lysine residues of target proteins. The similarity of the enzymes of the sumoylation pathway to ubiquitin pathway enzymes is remarkable, given the different effects of these two protein modification pathways. Sumoylation of cellular proteins has been proposed to regulate nuclear transport, signal transduction, stress response, and cell cycle progression (29). It is very likely that SUMO chaperones translocation of proteins among various cell compartments, however, the precise mechanistic details of this function of SUMO are not known.

Other fusions promote solubility of partner proteins presumably due to their large size (e.g., NUS A). Fusion of proteins with glutathione S-transferase (GST) or maltose binding protein (MBP) has been proposed to enhance expression and yield of fusion partners. However, enhanced expression is not always observed when GST is used as GST forms dimers and can retard protein solubility. Another problem with GST or other fusion systems is that the desired protein may have to be removed from the fusion. To circumvent this problem, protease sites, such as factor X, thrombin or Tev protease sites are often engineered downstream of the fusion partner. However, incomplete cleavage and inappropriate cleavage within the fusion protein is often observed. The present invention circumvents these problems.

SUMMARY OF THE INVENTION

In accordance with the present invention compositions and methods for enhancing expression levels of a protein of interest in a host cell are provided. An exemplary method comprises i) operably linking a nucleic acid sequence encoding molecule selected from the group consisting of SUMO, RUB, HUB, APG8, APG12, URM1, and ISG15 to a nucleic acid sequence encoding said protein of interest thereby generating a construct encoding a fusion protein, ii) introducing said nucleic acid into said host cell, whereby the presence of said molecule in said fusion protein increases the expression level of said protein of interest in said host cell. In a preferred embodiment the molecule is SUMO encoded by a nucleic acid of SEQ ID NO: 2. The method optionally entails cleavage of said fusion protein and isolation of the protein of interest.

In yet another embodiment of the invention, an exemplary method for generating a protein of interest having an altered amino terminus is provided. Such a method comprises i) providing a nucleic acid sequence encoding the protein of interest; ii) altering the N-terminal amino acid coding sequence in the nucleic acid; iii) operably linking a SUMO molecule to the nucleic acid sequence; and iv) expressing the nucleic acid in a eukaryotic cell, thereby producing the protein of interest in the cell, wherein the eukaryotic cell expresses endogenous SUMO cleaving enzymes, which effect cleavage of SUMO from the sequence encoding the protein of interest, thereby producing a protein of interest having an altered amino terminus. All amino acids with the exception of proline may be added to the amino terminus using this method.

The invention also provides a method for producing a sumolated protein for tracking protein localization within a host cell. An exemplary method comprises i) providing a nucleic acid sequence encoding said protein; ii) substituting the N-terminal amino acid coding sequence in the nucleic acid for a codon which encodes proline; iii) operably linking a SUMO molecule to said nucleic acid sequence; and iv) expressing said SUMO linked protein in said host cell.

In yet another aspect of the invention, a method for enhancing secretion levels of a protein of interest from a host cell is provided. Such a method comprises i) operably linking a nucleic acid sequence encoding molecule selected from the group consisting of SUMO, RUB, HUB, URM1, and ISG15 to a nucleic acid sequence encoding said protein of interest thereby generating a construct encoding a fusion protein, ii) introducing said nucleic acid into said host cell, whereby the presence of said molecule in said fusion protein increases the secretion of said protein of interest from said host cell.

BRIEF DESCRIPTION OF THE DRAWINGS

Figure 1 is a schematic drawing illustrating the conjugation pathways for ubiquitin and ubiquitin-like proteins (UBLs). An arrow in the "C-terminal hydrolase" column indicates the cleavage of the precursor proteins. Only enzymes previously described are provided. The failure to list a particular enzyme in a particular pathway does not preclude the existence of that enzyme.

Figure 2 is a schematic representation of the cloning strategy used to express SUMO fusion proteins. In this cloning strategy, a Bsa I site is introduced directly downstream of a SUMO sequence within a desired vector. The nucleic acid sequence encoding the protein to be expressed as a fusion with SUMO is amplified by PCR with primers that introduce a Bsa I site at the 5' end. The vector and the PCR product are cleaved by Bsa I and an appropriate restriction enzyme (represented by Xxx) that allows for insertion of the cleaved PCR product into the vector.

Figure 3 is a circular map of pSUMO, an E.coli SUMO expression vector. The nucleic acid sequence provided (SEQ ID NO: 37) encompasses the SUMO encoding region and the multiple cloning site. The amino acid sequence provided (SEQ ID NO: 38) is 6xHis tagged SUMO. Restriction enzymes are indicated above their

recognition sequence. The pSUMO expression vector has been constructed in the backbone of the pET-24d expression vector (Novagen).

Figures 4A and 4B show Coomassie stained gels and graphic data that demonstrate that the attachment of the carboxy-terminus of UBLs to the amino-terminus of target proteins increases expression and/or enhances solubility of the protein in E.coli. Green fluorescence protein (GFP) and UBL-GFP fusions encoded in pET24d E.coli expression vectors were expressed in the E.coli Rosetta pLysS strain (Novagen). Expression was induced either at 37°C with 1 mM IPTG for four hours either in LB medium (Fig 4A) or in minimal media with 1 mM IPTG at 26°C overnight (Fig 4B). Left panels are Coomassie stained SDS-polyacrylamide gels of total cellular protein (top) and soluble proteins (bottom). The first lanes of each gel are molecular weight markers. Dark arrow indicates observed GFP species and light arrow indicates size of expected GFP species. Right panel is quantitative representation in Arbitrary Units (AU) of GFP fluorescence present in soluble fractions as measured in a Fluorscan Ascent FL fluorometer (LabSystems).

Figure 5 is a Coomassie stained SDS-polyacrylamide gel demonstrating the expression and purification of a human tyrosine kinase as a SUMO fusion protein in E.coli. Tyrosine kinase and the fusion protein SUMO-tyrosine kinase were expressed in the Rossetta pLysS strain (Novagen) of E.coli in LB or minimal media (MM). The right panel shows the Ni-NTA resin purified proteins from the transformed E.coli cells. The left panel has the same lane arrangement as the right panel, but 1/3 of the amount protein was loaded on the SDS-polyacrylamide gel. Numbers indicate molecular weight standards in the first lane.

Figure 6 shows a Coomassie stained SDS-polyacrylamide gel representing purified SUMO hydrolase from E.coli and the partial purification and elution of SUMO-tyrosine kinase fusion protein. E.coli cells were transformed with a vector expressing either SUMO hydrolase Ulp1 or SUMO-tyrosine kinase and cultured in minimal media. Proteins were subsequently purified by Ni-NTA resin. SUMO-

tyrosine kinase was further purified by elution with either 100 mM EDTA or 250 mM imidazole. The gel shows that the current methods yield approximately 90% pure Ulp1 protein.

Figure 7 is a stained SDS-polyacrylamide gel of the expression of the liver X receptor (LXR) ligand binding domain as a fusion protein with SUMO. E.coli cells were transformed with a SUMO-LXR expression vector. The cells were subsequently induced with 1 mM IPTG at 20° C overnight or 37° C for 3 hours. 10 μg of total protein (WC), soluble protein (CS), and insoluble protein (Insol) from each induction were loaded per well of a 12% SDS-polyacrylamide gel.

Figures 8A and 8B display stained SDS-polyacrylamide gels demonstrating the solubility of the SUMO-MAPKAPK2 fusion protein expressed at 37°C (Fig. 8A) and 20°C (Fig. 8B). E.coli cells expressing a SUMO-fusion of MAPKAP2 kinase were induced with 0.1 (lanes 2-4), 0.25 (lanes 5-7), and 0.5 (lanes 8-10) mM IPTG. The original induction sample (I) in addition to the supernatant (S) and resuspended pellet (P) following lysis and centrifugation were analyzed by SDS-PAGE. The first lanes are BioRad low molecular weight markers.

Figure 9 is a Western blot (top panel) of UBL-GFP fusion proteins expressed in yeast cells demonstrating that UBL-GFP fusion proteins are co-translationally cleaved in yeast. Yeast strain BJ1991 was transformed with a vector expressing Ub-GFP, SUMO-GFP, Urm1-GFP, Hub1-GFP, Rub1-GFP, Apg8-GFP, Apg12-GFP or ISG15-GFP under the control of a copper sulfate regulated promoter. Total cell extracts were prepared by boiling the cells in SDS-PAGE buffer and briefly sonicating the sample to reduce viscosity. 20 μg of the total yeast proteins were resolved on 12% SDS-PAGE minigels and analyzed by Western blot with a rabbit polyclonal antibody against GFP and a secondary HRP-conjugated antibody. The arrow indicates the size of unfused GFP. An identical gel (bottom panel) was run in parallel and stained with Coomassie to ensure equal loading of the proteins from all samples.

Figure 10 is a series of Western blots that indicate SUMO-GFP Fusions are co-translationally cleaved in yeast generating novel amino termini. In addition to methionine as the first amino acid of GFP following the C-terminal Gly-Gly sequence of SUMO, we have engineered the remaining 19 amino acids as the amino-terminal residue of GFP in yeast SUMO-(X)20-GFP expression vectors. All expression vectors containing the 20 amino-terminal variants of GFP fusion proteins were expressed in yeast under the control of copper inducible promoter. Yeast lysates were separated by SDS-PAGE and analyzed by Western blot with antibodies against GFP. The "unfused-GFP" lanes represent the expression of GFP alone with no SUMO fusion. The "SUMO-GFP" lanes are bacterially expressed SUMO-GFP.

Figures 11A and 11B are schematic representations of the SUMO (Fig. 11A) and ubiquitin (Fig. 11B) GFP fusion proteins that also contain the gp67 secretory signal. In construct E, only unfused GFP protein is expressed. In construct G, a 7 kDa secretory sequence from gp67 was attached to the N-terminus of GFP. In constructs S and U, SUMO and ubiquitin sequences, respectively, are inserted in frame to the N-terminus of GFP. In constructs GS and GU, gp67 sequences are followed by SUMO and ubiquitin, respectively, and then GFP. In constructs SG and UG, gp67 sequences are inserted in between the C-terminus of SUMO and ubiquitin, repectively, and the N-terminus of GFP.

Figures 12A and 12B are Western blots demonstrating expression of SUMO and ubiquitin fusion proteins in insect cells. Hi-five insect cells were infected with recombinant baculovirus encoding for SUMO or ubiquitin fusion proteins. At 24 hours post-infection, equal amounts of cell lysates (Fig. 12A) and media (Fig. 12B) were separated by SDS-PAGE and analyzed by Western blot with antibodies against GFP. Lane markers: Hi5 is Hi Five cells, E is eGFP, G is gp67-eGFP, U is ubiquitin-eGFP, S is SUMO-eGFP, GU is gp67-ubiquitin-eGFP, UG is ubiquitin-gp67-eGFP, GS is gp67-SUMO-eGFP, SG is SUMO-gp67-eGFP, and eGFP is a positive control.

Figures 13A, 13B, and 13C are Western blots demonstrating expression of SUMO and ubiquitin fusion proteins in insect cells. Hi-five insect cells were infected with recombinant baculovirus encoding for SUMO or ubiquitin fusion proteins. At 48 hours post-infection, equal amounts of cell lysates (Fig. 13A and 13C) and media (Fig. 13B) were separated by SDS-PAGE and analyzed by Western blot with antibodies against GFP. The lanes are: Hi5 is Hi Five cells, E is eGFP, G is gp67-eGFP, U is ubiquitin-eGFP, S is SUMO-eGFP, GU is gp67-ubiquitin-eGFP, UG is ubiquitin-gp67-eGFP, GS is gp67-SUMO-eGFP, SG is SUMO-gp67-eGFP, and S-P is SUMO-proline-GFP.

Figure 14 is a series of micrographs of eGFP expression in Hi-Five cells infected with different eGFP fusion baculoviruses. Pictures were taken with a Leitz Fluovert Inverted Microscope with excitation at 488nm with Hammamatsu Orca Cooled CCD camera.

Figure 15 contains stained SDS-polyacrylamide gels representing the in vitro Ulp1 cleavage of Ni-NTA resin purified His6SUMO-eGFP fusion proteins expressed in E.coli. The purified His6SUMO-eGFP fusions, containing a different amino acid at the +1 position of the Ulp1 cleavage site, were incubated at 30°C for 3 hours with purified Ulp1 hydrolase. The lanes are marked with the single letter code of the +1 amino acid. The negative control (-Ve) is the incubation of His6SUMO-eGFP at 30°C for 3 hours in the absence of enzyme. Low molecular weight markers (LMW) are also provided.

Figure 16 contains a pair of stained SDS-polyacrylamide gels representing the effects of various conditions on Ulp1. Ni-NTA purified His6SUMO-GFP was incubated with Ulp1 under the indicated conditions for one hour at room temperature unless indicated otherwise. Low molecular weight markers (LMW) are also provided.

Figure 17 is a stained SDS-polyacrylamide gel representing the effects of various protease inhibitors on Ulp1. Ni-NTA purified His6SUMO-GFP was incubated with Ulp1 and 10 mM of various protease inhibitors for 1 hour at room temperature. Lane markers: Norm is addition of Ulp1 and N-ethymaleimide (NEM) to the substrate at the same time, Pre is the incubation of Ulp1 with NEM prior to the addition of substrate, +Ve is the absence of any inhibitor, -Ve is in the absence of Ulp1, lane 1 is with E-64, lane 2 is with EDTA, lane 3 is with leupeptin, lane 4 is with NEM, lane 5 is with pepstatin, lane 6 is with TLCK. Low molecular weight markers (LMW) are also provided.

Figure 18 is a stained SDS-polyacrylamide gel showing purification and cleavage of MAPKAP2. E.coli transformed with the expression vector for SUMO-MAPKAP2 where either grown at 37°C and induced with 0.1 mM IPTG (lanes 2-7) or at 20°C and induced with 0.5 mM IPTG (lanes 8-13). Cell lysates were Ni-NTA purified and separated by SDS-PAGE. Lane 1: BioRad low molecular weight marker; lanes 2 and 8: soluble fraction of cell lysates; lanes 3 and 9: flow through from Ni-NTA column; lanes 4 and 10: 15 mM imidazole wash of Ni-NTA column; lanes 5 and 11: 300 mm imidazole elution of Ni-NTA column; lanes 6 and 12: supernatant of 2 hour incubation of elution with SUMO hydrolase at 30°C; and lanes 7 and 13: pellet of hydrolase incubation.

Figure 19 is a stained SDS-polyacrylamide gel showing SUMO hydrolase function at pH 7.5 and 8.0. Purified SUMO-GFP was cleaved using 1/50 diluted purified stock of SUMO hydrolase in sodium phosphate buffer pH 7.5 (lanes 1-6) and 8.0 (lanes 8-13) at room temperature for the following length of times: lanes 1 and 8: 0 minutes, lanes 2 and 9: 1 min, lanes 3 and 10: 2.5 min, lanes 4 and 11: 5 min, lanes 5 and 12: 10 min, and lanes 6 and 13: 20 min. Lane 7 is blank and M is molecular weight markers.

Figure 20 is a stained SDS-polyacrylamide gel indicating SUMO hydrolase cleaves SUMO-β-Galactosidase. Purified SUMO hydrolase was incubated

with E. coli produced SUMO-β-Galactosidase at room temperature for 0 minutes (lane 1), 2 .5 min (lane 2), 5min (lane 3), 10 min (lane 4), and 20 min (lane 5). Molecular weight markers are provided in lane M.

Figure 21 is a stained SDS-polyacrylamide gel showing the cleavage of SUMO-GUS by SUMO Hydrolase in the presence of urea. Ni-NTA purified SUMO-β-GUS was incubated with 1/50 dilution of purified stock of SUMO hydrolase for 1 hour in increasing concentrations of urea at pH 8.0. Lane markers: M is broad range molecular weight marker; lane 1 is SUMO-GUS from soluble E.coli fraction; lane 2: flow through from nickel column; lane 3: wash; lane 4: elution; lanes 5-9: SUMO-GUS and hydrolase with various denaturants, specifically, lane 5: none; lane 6: 1mM DTT; lane 7: 0.5 M Urea; lane 8: 1.0M Urea; lane 9: 2.0M Urea .

Figure 22 is a stained SDS-polyacrylamide gel demonstrating the rapid isolation of a SUMO fusion protein. E. coli cells expressing a single IgG binding domain from Protein G fused to His6Smt3 were lysed with guanidinium chloride lysis buffer. Cell lysate supernatants were purified over Ni-NTA and eluted in a native buffer that allows for cleavage by Ulp1. Lane markers: PMW is molecular weight markers; lane 1 is cellular proteins prior to treatment with guanidinium chloride, lane 2 is guanidinium chloride cell lysates, lane 3 is flow through from Ni-NTA column, lane 4 is elution, and lane 5 is Ulp1 cleavage of elution.

Figures 23 is the amino acid (SEQ ID NO: 1) and nucleotide (SEQ ID NO: 2) sequences of SUMO.

Figures 24A and 25B are the amino acid (SEQ ID NO: 3) and nucleotide (SEQ ID NO: 4) sequences of GFP.

Figures 25A and 25B are the amino acid (SEQ ID NO: 5) and nucleotide (SEQ ID NO: 6) sequences of SUMO-GFP.

Figures 26A and 26B are the amino acid (SEQ ID NO: 7) and nucleotide (SEQ ID NO: 8) sequences of ubiquitin-GFP.

Figures 27A and 27B are the amino acid (SEQ ID NO: 9) and nucleotide (SEQ ID NO: 10) sequences of URM1-GFP.

Figures 28A and 28B are the amino acid (SEQ ID NO: 11) and nucleotide (SEQ ID NO: 12) sequences of HUB1-GFP.

Figures 29A and 29B are the amino acid (SEQ ID NO: 13) and nucleotide (SEQ ID NO: 14) sequences of RUB1-GFP.

Figures 30A and 30B are the amino acid (SEQ ID NO: 15) and nucleotide (SEQ ID NO: 16) sequences of APG8-GFP.

Figures 31A and 31B are the amino acid (SEQ ID NO: 17) and nucleotide (SEQ ID NO: 18) sequences of APG12-GFP.

Figures 32A and 32B are the amino acid (SEQ ID NO: 19) and nucleotide (SEQ ID NO: 20) sequences of ISG15-GFP.

Figure 33 is the amino acid (SEQ ID NO: 21) and nucleotide (SEQ ID NO: 22) sequences of SUMO-Protein G.

Figures 34A, 34B, and 34C are the amino acid (SEQ ID NO: 23) and nucleotide (SEQ ID NO: 24) sequences of SUMO-β GUS.

Figures 35A, 35B, and 35C are the amino acid (SEQ ID NO: 25) and nucleotide (SEQ ID NO: 26) sequences of SUMO-LXRα.

Figures 36A and 36B are the amino acid (SEQ ID NO: 27) and nucleotide (SEQ ID NO: 28) sequences of SUMO-Tyrosine Kinase.

Figure 37A and 37B are the amino acid (SEQ ID NO: 29) and nucleotide (SEQ ID NO: 30) sequences of SUMO-MPAKAP2 Kinase.

Figures 38A, 38B, 38C, 38D, and 38E are the amino acid (SEQ ID NO: 31) and nucleotide (SEQ ID NO: 32) sequences of SUMO-β GAL.

Figure 39 is a circular map of YEpSUMO-eGFP.

Figures 40A, 40B, 40C, 40D, and 40E are the nucleotide sequence (SEQ ID NO: 33) of YEpSUMO-eGFP. Select restriction enzyme sites are indicated.

Figure 41 is a circular map of YEpUbGUS.

Figures 42A, 42B, 42C, 42D, 42E, 42F, and 42G are the nucleotide sequence (SEQ ID NO: 34) of YEpSUMO-eGFP. Select restriction enzyme sites are indicated.

Figure 43 is a circular map of pFastBac SUMO-eGFP.

Figures 44A, 44B, 44C, 44D, and 44E are the nucleotide sequence (SEQ ID NO: 35) of pFastBac SUMO-eGFP. Select restriction enzyme sites are indicated.

Figure 45 is a circular map of pSUMO (pET24d6HisxSUMO).

Figures 46A, 46B, 46C, 46D, and 46E are the nucleotide sequence (SEQ ID NO: 36) of pSUMO (pET24d6HisxSUMO). Select restriction enzyme sites are indicated.

DETAILED DESCRIPTION OF THE INVENTION

There are a number of reasons for the lack of efficient recombinant protein expression in a host, including, for example, short half life, improper folding or compartmentalization and codon bias. While the Human Genome project has successfully created a DNA "map" of the human genome, the development of protein expression technologies that function uniformly in different expression platforms and for all the protein motifs has not yet been achieved.

In accordance with the present invention, it has been discovered that that N-terminal fusion of the ubiquitin homologue SUMO or Smt3 to otherwise unexpressed or poorly expressed proteins remarkably enhances the expression levels of biologically active proteins in both prokaryotes and eukaryotes. The Ubiquitin-Like protein (UBL) family contains many proteins, including for example, SUMO, Rub1, Hub1, ISG15, Apg12, Apg8, Urm1, , Ana 1a and Ana 1b (15, 28). See Table 1. The hallmark of all of these proteins, exept APG12, and URM1, is that they are synthesized as precursors and processed by a hydrolase (or proteases) to generate mature carboxy-terminal sequence. Secondly, all of the UBLs share a common structure.

In E.coli, fusion proteins remained intact while in yeast or insect cells fusion proteins were efficiently cleaved, except when proline was the N-terminal residue of the target protein. While any of the UBLs set forth in Table 1 may be utilized in the compositions and methods of the invention to enhance expression of heterologous fusion proteins of interest, SUMO is exemplified in the gene fusion system provided herein.

Table 1

Properties of Ubiquitin-like Proteins (UBLs)

UBL (yeast)	Function	Knockout phenotype	Substrate	% UB Identity	KDa	Hydro-lase	COOH Residues
UB	Translocation to proteasome for degradation.	not viable	many	100	8.5	UCH/UBPs	LRLRGG (SEQ ID NO: 39)
SUMO (SMT3)	Translocation to nucleus	not viable	Sentrins, RanGap, others	18	11.6	Aut1/Aut2	GG
RUB1 (NEDD8)	Regulation of mitosis.	viable; non-essential.	cullins, cytoskelet .proteins	60	8.7	not known	GG
HUB1	Cell polarization during mating projections.	viable; deficient in mating.	Sph1, Hbt1 cell polarity factors	22	8.2	not known	YY
ISG-15 (UCRP)	Unknown	IFN, LPS hypersensi-tivity; death	many	~30; 28 (two domains)	15.0	UBP43 (USP18)	LRLRGG (SEQ ID NO: 39)
APG12	Autophagy	viable, defective in autophagy	Apg5	18	21.1	not cleaved	FG
URM1	Unknown	ts growth; non-essential.	unknown	20	11.0	not known	GG
APG8 (LC3)	Autophagy	viable; no autophago-cytosis or sporulation	phospatid yl-ethanol-amine	18	13.6	Apg4/Aut2	FG

The SUMO fusion system of the present invention has been successfully applied to express different molecular weight proteins such as 6KDa Protein G domain to 110 KDa β-galactosidase in E.coli and eukaryotic cells. More specifically, the system allows one to: (1) enhance the expression of under-expressed proteins; (2) increase the solubility of proteins that are insoluble; (3) protect candidate proteins from degradation by intracellular proteases by fusing UBLs to their N-termini; (4) cleave the fusion protein to efficiently generate authentic proteins using naturally-present enzymes (5) generate proteins with novel amino termini; and (6) cleave all fusion proteins with remarkable efficiency irrespective of the N-terminal sequence of the fused protein, using UBL hydrolases such as SUMO hydrolase Ulp1. Because UBLs are small molecular weight proteins (~100 amino acids), they can also be used as purification tags as well. These remarkable properties of UBLs make them excellent candidates for enhancing expression and solubility of proteins. The method may also be utilized to generate novel amino termini on proteins of interest for a variety of research, diagnostic and therapeutic applications.

The ultimate fate of ubiquitinated or sumoylated proteins within a cell varies. A protein can be monoubiquitinated or polyubiquitinated. Ubiquitination of protein has multiple functions and gives rise to different fates for the protein within a cell (11). Ubiquitination primarily targets proteins to 26S proteosome for degradation (13). On the other hand, sumoylation of target proteins does not lead to degradation, but, rather, leads directly or indirectly to altered localization of proteins (15) . There are about 17 deubiquitinating enzymes that cleave conjugated ubiquitin from target proteins as well as ubiquitin-ubiquitin and ubiquitin artificial-fusion proteins (1, 35). Thus far it appears that yeast has two cysteinyl proteases, called Ulp1 and Ulp2, that remove SUMO from ε-amino groups of lysine as well from the artificial linear SUMO-fusions(20, 21).

To determine if UBLs and SUMO fusion will enhance expression of recombinant proteins of different sizes and function, we have designed several UBL-GFP fusion proteins in addition to SUMO-fusion proteins and monitored their expression levels in E.coli, yeast and insect cells. In E.coli, the proteins are

expressed as intact fusions, while in eukaryotes, the fusions were efficiently cleaved. A dramatic increase in the yield of proteins after fusion with SUMO and expression in E.coli was observed. In additional studies, SUMO-GFP protein was used as a model fusion for detailed studies in yeast and insect cells. We have designed SUMO-GFP fusion where all the N-terminal methionine residues have been replaced with the rest of the 19 amino acids. We have purified 20 sumo-GFP fusion proteins from E.coli and cleaved them *in vitro* with Ulp1. Ulp1 efficiently cleaved 19 out of the 20 possible amino acid junctions. The proline junction was not cleaved. As compared to deubiquitinating enzyme (3), Ulp1 demonstrated broad specificity and robustness in its digestion properties. Proteins having a wide range of molecular weights were cleaved efficiently by Ulp1. Similarly, in yeast, and insect cells, the fusion proteins were efficiently processed, yielding intact, biologically active proteins. In addition to enhancing protein expression levels, the SUMO-fusion approach can be used to advantage to generate desired N-termini to study novel N-terminal protein functions in the cell. Since SUMO fusion can both enhance recombinant protein yield and generate new N-termini, this technology provides an important tool for post-genomic biotechnology analyses.

The materials and methods set forth below are provided to facilitate the practice of the present invention.

Design and Construction of E. coli Expression Vectors:

The original vector backbone was developed using pET 24d vector from Novagen (see Fig 3 as well as Figures 45-46A-E) . pET24d uses a T7 promoter system that is inducible with IPTG. The vector has a kanamycin selection marker and does not contain any translation terminator.

Construction of variable His6SUMO-GFP fusions:

A N-terminal six his-tagged SUMO (fusion vector was constructed as follows. A PCR product was generated with the primers

5'CCATGGGTCATCACCATCATCATCACGGGTCGGACTCAGAAGTCAATC

AA-3' (SEQ ID NO: 40) and 5'-GGATCCGGTCTCAACCTCCAATC TGTTCGCGGTGAG-3'(SEQ ID NO:41) using yeast Smt3 gene (16) as a template (kind gift of Erica Johnson). The PCR fragment was double digested with Nco I and Bam HI, and then ligated into pET24d, which had been similarly digested. It is important to note that the current invention utilizes a variant of the wild type yeast SUMO sequence. The A nucleotide at position 255 has been replaced with a G nucleotide, thus encoding an alanine instead of a threonine (SEQ ID NOS: 1 and 2). The detailed cloning strategy is provided in Fig 2. The pET24d His6Smt3eGFP fusions, containing each of the twenty different amino acids at the +1 position of the cleavage site were generated as follows. The eGFP sequence was amplified a template, with the primers 5'-GGTCTCAAGGT NNNGTGAGCAAGGGCGAGGAGC-3' (SEQ ID NO:42) and 5'-AAGCTTATTACTTGTACAGCTCGT CCATGCC-3'(SEQ ID NO: 43), where the NNN in the forward primer corresponding to the variable codon encoding one of the twenty amino acids. The PCR products were purified and double digested with Bsa I and Hind III, these were then ligated into the pET24dHisSUMO vector which had been similarly digested. Plasmids from clones containing the variable inserts, were sequenced to confirm the presence of the novel codon in each.

Construction of SUMO-fusion vectors from pSUMO:

The gene encoding the protein of interest is cloned in frame with the SUMO tag, in the pSUMO vector, by utilizing the encoded Bsa I site. Bsa I belongs to the family of Class IIS restriction enzymes, which recognize non-palindromic sequences, and cleave at a site that is separate from their recognition sequences. The latter trait gives Class IIS enzymes two useful properties. First, when a Class IIS enzyme recognition site is engineered at the end of a primer, the site is cleaved when digested. Second, overhangs created by Class IIS enzymes are template-derived and thus unique. This is in clear contrast to regular Class II restriction enzymes such as *Eco*RI, which creates an enzyme-defined overhang that will ligate to any *Eco*RI-digested end. The unique overhangs produced by Class IIS enzymes can be ligated only to their original partner.

It is often preferable to amplify the gene encoding the protein of interest via PCR prior to cloning into the pSUMO vector. The forward primer must contain the additional standard sequence:

5'-GGTCTCAAGGTNNN - 3' (SEQ ID NO:44) where GGTCTC is the Bsa I site and NNN is the first codon of the gene encoding the protein of interest. Additional nucleotides are required for the primer to anneal specifically with the gene of interest during the PCR amplification. The reverse primer may contain another restriction enzyme such as Xho I to allow for directional cloning of a gene into pSUMO. Bsa I can also be employed in the reverse primer to simplify cloning steps, for example, in the following primer:

5' - GGTCTCCTCGAGTTANNN - 3' (SEQ ID NO:45)

The PCR product can be digested with both Xho I and Bsa I. A digestion reaction containing just the latter enzyme generates a product that would directionally ligate into the pSUMO vector between the Bsa I and Xho I sites of the MCS.

Construction of pSUMO-Protein G fusion E.coli expression vector:

The B2 IgG binding domain (9) from streptococcus G148 protein was synthesized by three synthetic oligonucleotides. The sequence of the gene is 5'- GT **CTTAAGA** CTA AGA GGT GGC ACG CCG GCG GTG ACC ACC TAT AAA CTG GTG ATT AAC GGC AAA ACC CTG AAA GGC GAA ACC ACC-3'. (SEQ ID NO:46) The 81 bps oligo sequence is 5'-GCC GTT ATC GTT CGC ATA CTG TTT AAA CGC TTT TTC CGC GGT TTC CGC ATC CAC CGC TTT GGT GGT TTC GCC TTT CAG-3'. (SEQ ID NO:47) The 86 pbs oligo sequence is 5'-CAG TAT GCG AAC GAT AAC GGC GTG GAT GGC GTG TGG ACC TAT GAT GAT GCG ACC AAA ACC TTT ACC GTG ACC GAA TAA **GGT ACC** CC-3'(SEQ ID NO:48). The bolded nucleotides refer to the AflII and Kpn1 sites that flank the protein G domain. ACG is the first amino acid residue of the domain. The above three oligos were annealed using the Life Technologies protocol. The annealed fragments were extended by Pol1 enzyme. The resultant gene was PCR amplified by the following oligo primers G1 forward 5'- CTT GTC TTA AGA

GGT-3' (SEQ ID NO:49) and G2 reverse primer 5'-GCT GGG TAC CTT ATT CGG TCA-3'(SEQ ID NO:50). The above protein G gene was cloned at the AflII and Kpn1 site of the human ubiquitin gene and expressed as ubiquitin-protein G fusion protein in an *E.coli* pET 22 expression vector (Novagen). The protein G sequence was in turn amplified from the ubiquitin-protein G fusion plasmid by using the primers 5'-GGTCTCAAGGTACGCCGGCGGTGACCACCT-3'(SEQ ID NO:51) and 5'- AAGCTTATTATTCGGTCACGGTAAAGGTTT-3'(SEQ ID NO:52) and inserted in pSUMO to generate pSUMO-protein G expression vector.

Construction of E.coli SUMO-β-galactosidase expression vector.

E. coli β-galctosidase was amplified using pfu (Stratagene) a preparation of genomic DNA from BL21(DE3) (Stratagene) as a template and the primers 5'-GGTCTCAAGGTATGACCATGATTACGGATTCACT-3' (SEQ ID NO:53) and 5'-AAGCTTATTATTATTATTTTTGACACCAGACC-3'(SEQ ID NO:54). The PCR products were purified and double digested with Bsa I and Hind III. These were then ligated into the vector pET24d6xHisSUMO, which had been similarly digested.

Construction of E.coli pSUMO-Liver X Receptor (LXR) expression vector:

The PCR products of the LXR from amino acid residue 189 to the end of the protein that spans the ligand binding domain was digested with BsaI and HindIII and ligated into the pSUMO vector, also digested with Bsa1 and HindIII.

Construction of E.coli pSUMO-MAPKAP2 expression vector:

The fragment of MAPKAP2, encoded in the plasmid pMON45641, was amplified by PCR and cloned into pET24d 6HisSUMO vector by designing PCR primers that flank the sequence shown Figures 8A and 8B. The SUMO vector was digested with Bsa I site and Hind III. The cloning procedure yields a fusion protein, which, upon expression, purification and cleavage, generates the desired protein whose first amino acid is a glutamine (CAG).

Construction of E.coli pSUMO-tyrosine kinase expression vector:

For the tyrosine kinase, both, the SUMO fusion and unfused expression vectors were designed. As described above the region of kinase was cloned by PCR flanked with BsaI and Hind III sites that were cloned in to similarly digested pSUMO.

Construction of E.coli pSUMO-β-Glucuronidase expression vector:

E. coli β -glucuronidase was the kind gift of Ben Glick, University of Chicago) and amplified with the primers 5'-GGTCTCAAGGTATGCAGATCTTCGTCAAGACGTT-3'(SEQ ID NO:55) and 5'- AAGC TTATTATTGTTTGCCTCCCTGCTGCG-3'(SEQ ID NO:56).

Construction of E.coli SUMO-hydrolase expression vector:

C-terminal His-tagged SUMO hydrolase/protease Ulp(403-621)p (21) (27) was expressed from pET24d in Rosetta(DE3) pLysS (Novagen). The recombinant protein was purified using Ni-NTA agarose (Qiagen) and buffer exchanged into 20 mM Tris-HCl pH 8.0, 150 mM NaCl and 5 mM β-mercaptoethanol using a PD-10 column (AP Biotech). About 2 ug of the pure protein was analyzed on gels and data shown in Fig 6 lane Ulp1. The protein was almost 90 % pure as judged by SDS-PAGE analysis.

Construction of E.coli UBL-GFP fusion vectors.

DNA sequences encoding ubiquitin (Ub), SUMO, Urm1, Hub1, Rub1, Apg8, and Apg12 were PCR-amplified using Deep-Vent polymerase (NEB) and yeast strain DNA to generate a template. Full-length human ISG15 cDNA was a kind gift of Dr. A. Haas, Medical College of Wisconsin, Milwaukee. A unique NcoI site followed by 6His sequence was introduced by PCR at the 5'-end of each Ubl cDNA. Primer sequence at the 3'-end included unique Esp3I and HindIII sites. PCR products were digested with NcoI/HindIII and inserted into respective sites of pET24d vector (Novagen) as described above. Full length GFP sequence (Clontech Cat # 60610-1) flanked by Esp3I and HindIII sites, respectively, was PCR-amplified

and cloned into pCR4-TOPO-TA vector (Invitrogen). Esp3I/HindIII digested GFP-encoding gene was inserted into respective sites of pET24d-UBLl plasmids, creating final UBL-GFP expression vectors for E.coli. *In toto*, there were nine plasmid constructs coding for the following structures: 6His-Ubl-GFP. All plasmids were sequenced to confirm the expected structure.

Design and Construction of Yeast UBL-Fusion Vectors:

Saccharomyces cerevisiae has been used as a eukaryotic model for all the experiments involving yeast. All of the expression vectors for these studies were designed on multicopy yeast vectors that contain tryptophan or leucine as a selectable marker and 2μ as an origin of replication(22). Proteins were expressed as unfused products or as ubiquitin, SUMO or other UBL fusion proteins.

Construction of the β-glucuronidase Yeast Expression Vectors:

To demonstrate that UBLs increase the level of secretion of the protein to the media, in addition to enhancing the level of expression, expression vectors were constructed with and without ubiquitin. We have also compared ubiquitin fusion and SUMO fusion using GFP as a model protein (see Fig 9 and Fig 10). pRS425-GUS plasmid was produced by cloning the XhoI-SacI fragment (containing *E.coli* β-Glucuronidase (GUS)) from plasmid pGUS1 (25, 22) into the XhoI-SacI sites of plasmid pRS425 (32). The next construction involved addition of a promoter, and resulted in the plasmid pRS425-ADH1p-GUS. The fragment XhoI-HindIII (containing the ADH1 promoter) was inserted into the XhoI-HindIII sites of the plasmid pRS425-GUS. The ADH1 promoter XhoI-HindIII fragment was cloned using polymerase chain reaction (PCR), amplifying the ADH1 promoter from the plasmid pGRIP1(37). The following primers were used to amplify the full length ADH1 promoter: ADH1-XhoI: 5'-gctcgagagcacagatgcttcgttg-3'(SEQ ID NO:57), and ADH1-HindIII: 5'-gcaaagcttggagttgattgtatgc-3'(SEQ ID NO:58). The underlining indicates the nucleotide sequence of the XhoI and HindIII restriction sites. PCR of the DNA fragment involved amplification in 30 cycles (96° C - 30 sec., 54 ° C - 1 min. and 72 ° C - 3 min.) using high replication fidelity Deep Vent

Polymerase (New England Biolabs). The PCR product was then digested with XhoI and HindIII, and subsequently cloned into the XhoI-HindIII sites of pRS425-GUS. Construction of the next set of plasmids involved a change in promoter. The following two plasmids were constructed to give expression vectors containing either a methionine or proline junction between the ubiquitin and the GUS. pRS425-GPDp-Ub(Methionine)-GUS and pRS425-GPDp-Ub(Proline)-GUS were similarly constructed using both pre-constructed plasmids and PCR amplification. The final expression construct was pRS425-CUP1p-SUMO-GUS, which was the only plasmid produced with the CUP1, copper regulated promoter. This plasmid was digested with the enzymes BglII and NsiI, releasing the CUP1 promoter(6). The CUP1 fragment was then ligated to pRS425-GPDp-Ub-GUS, having also been digested with BglII-NsiI.

Construction of SUMO-N-GFP yeast expression vector:

To determine what variety of N-terminal variant amino acids at the junction of SUMO and GFP can be cleaved in yeast we designed SUMO-GFP vectors in which all 20 amino acid residues were encoded at the N-terminus of GFP. Essentially all 20 SUMO-X-GFP vectors designed for E.coli expression were digested with Bsa I - Hind III, and the inserts were purified. The 20 inserts were cloned in Yep12 that was slightly modified. Specifically, YeEpSW was generated by digesting Yep12 with Bam HI and SacI. The CUP1 promoter region was recovered from the fragment by PCR. A polylinker was created at the 3' end of CUP1 with a variety of restriction sites including NcoI and Xho1. All 20 SUMO-GFPs (N end variants) were digested with NcoI-XhoI enzymes and cloned directly YepSW. The resultant vector YepSW-SUMO-eGFP utilizes tryptophan selection and expresses SUMO-GFP proteins under the control of the copper promoter. All vectors were sequenced to ensure correct codons at the junction of SUMO and GFP.

Construction of UBL-GFP fusion yeast expression vectors:

Construction of the UBL-GFP fusion vectors for E.coli has been described above. In order to make UBL yeast expression vector NcoI/XhoI fragments carrying

GFP alone and all the Ubl-GFP fusions were inserted into respective sites of pYEp SW (see above) that was similarly digested with NcoI/XhoI. Insertion of UBL-GFP cassette in Yep SW (See Figures 39 and 40A-40F), allows copper inducible expression of Ubl-GFP fusions in yeast system.

Design and Construction of Recombinant Baculovirus for SUMO and Ubiquitin GFP Fusion Expression:

To demonstrate that attachment of SUMO or ubiquitin to GFP increases its expression and enhances secretion into the media, several GFP fusion vectors were designed with different configurations of gp67 secretory signals. The basic GFP vector for expression is essentially based on E.coli vectors described above. Derivatives of this vector representing each candidate gene have been constructed by designing PCR primers. The construction of GFP plasmid transfer vectors for baculovirus is described. To help appreciate the rationale for the secretory signal in the context of GFP-fusion, see the diagrammatic representation shown in Fig 11. Single letter code refers to unfused GFP (E); gp67-sec signal-GFP (G); ubiquitin-GFP (U); SUMO-GFP (S); gp67-Ub-GFP (GU); Ub-gp67-GFP (UG); gp67-SUMO-GFP (GS); and SUMO-gp67-GFP (SG).

(i) pFastbacE. A synthetic oligonucleotide containing the Esp3I site was inserted between BamHI and EcoRI cloning site of the transfer vector pFastbac1, which had been modified by removing Esp3I site from Gmr region. *(ii) pFastbacG.* The signal sequence of the gp67 gene derived from pACSecG2T was isolated by PCR using 2 primers (f-gp67 and r-gp67), digested with BglII and EcoRI in the next step, and then inserted between BamHI and EcoRI cloning sites of the transfer vector pFastbacE. *(iii) pFastbacS.* A full-length SUMO gene derived from pET SUMO was generated by PCR using 2 primers (f-bacsmt and r-bacsmt), digested with BsaI and EcoRI in the next step, and then inserted between BamHI and EcoRI cloning sites of the transfer vector pFastbacE. *(iv) pFastbacG/S.* The signal sequence of the gp67 gene in the pACSecG2T vector was generated by PCR using 2 primers (f-fusgp67 and r-fusgp67), and inserted between BamHI and EcoRI cloning sites of the transfer vector pFastbacE to create a new pFastbacG, which was used for fusion

with SUMO afterward. A full-length SUMO gene derived from pET SUMO as described above *(iii)* was digested with BsaI and SacI and inserted between Esp3I and SacI cloning sites of the new transfer vector pFastbacG. *(v) pFastbacS/G.* A full-length SUMO gene derived from pET SUMO was generated by PCR using 2 primers (f-fussmt3 and r-fusgp67) and inserted between BamHI and EcoRI cloning sites of the transfer vector pFastbacE to create the new pFastbacS, used for fusion with gp67 afterward. The signal sequence of the gp67 gene derived from pACSecG2T as described above *(ii)* was digested with BsaI and SacI, and then inserted between the Esp3I and SacI cloning sites of the new transfer vector pFastbacS.

Preparation of baculovirus stocks and cell growth.

Transfer vector constructs based on the pFastbac 1 shuttle plasmid (Invitrogen, Inc.) were transposed in DH10Bac *E. coli* competent cells to transfer the respective e-GFP fusion sequences into recombinant virus DNA by site-specific integration. After alkaline lysis of transformed (white colonies) of *E. coli* cells, which contain recombinant virus (bacmid) DNA, and extraction of the recombinant bacmid DNA, the bacmid DNA was used to transfect *Spodoptera frugiperda* (Sf9) insect cells, in which virus replication occurs. The virus was then amplified to produce passage 2 (for long-term storage) and passage 3 virus (for working) stocks by infection of fresh Sf9 cell cultures and used directly to infect cells for fusion protein expression. Virus infectivity (pfu/ml) was determined by titration in Sf9 cells using the BacPAK™ Rapid Titer Kit (BD Sciences Clontech, Inc.). A 50ml culture of Hi-Five cells at concentration of 1x106 cells/ml, was infected with recombinant virus at MOI = 5 in Express Five media (serum free media). The cells were grown in 100ml spinner flask at 27° C. Every 24 hours, cell viability was determined by trypan blue and cell counting. 5ml of the suspension culture was removed at 24 hour intervals, centrifuged at 500 x g at 4°C in 10 minutes. The supernatant was transferred into a fresh tube to monitor any protein that may have been secreted into the media (see below).

Analysis of Proteins from Insect Cell Compartments:

Cell pellets (from above step) were gently washed in 1 ml PBS and recentrifuged at 500 x g at 4 ° C for 10 minutes. All supernatant and pellets are stored at -80° C. The presence of recombinant protein in cells and media was ascertained by SDS-PAGE and Western blotting of supernatant and cell pellets. The total intracellular protein was extracted by M-PER extraction buffer (Pierce), a neutral buffer for protein extraction. The cell pellet was mixed with rapid pipetting and incubated for 1 hour on an orbital shaker. The suspension was centrifuged at 500 x g at 4° C for 10 minutes to remove debris. The supernatant contained extracted cellular proteins that were either analyzed by PAGE or stored at -80° C. To analyze the proteins present in the media, the following procedure was adopted. Trichloroacetic acid was added to 5 ml media to a final concentration of 20%. The suspension was mixed well and left on ice for three hours, and then centrifuged 500 x g at 4 ° C for 10 minutes. The white pellet was washed with 80% ethyl alcohol twice, and then dried. The pellet was suspended in 1 ml of M-PER buffer for PAGE to compare the distribution of control (unfused) and SUMO-fused proteins inside and outside the cell.

Methods for Analysis of Yeast Expressed Fusion Proteins:

Yeast cultures were grown in synthetic or rich media. Standard yeast and *E. coli* media were prepared as described (31) . The yeast strain Y4727: *Matα his3-Δ200 leu2-Δ0 lys2-Δ0 met5-Δ0 trp1-Δ63 ura3-Δ0* was used as a host (gift from Dr. Jeff Boeke) or BJ 1991 . Yeast transformation was performed according to published procedures (8) . Yeast transformants with autonomously replicating plasmids were maintained in yeast selective media. The *E.coli* β-Galactosidase and β-Glucuronidase proteins were expressed under the regulation of either the alcohol dehydrogenase (ADH), or Glyceraldehyde-Phosphate-Dehydrogenase (GPD) promoter or copper metallothioneine (CUP1) promoter in 2μm multicopy plasmids with the *LEU2* selective marker.

Yeast cells were transformed with appropriate expression vectors, and single colonies were grown in synthetic media minus the selectable marker. For each protein, at least two single colonies were independently analyzed for protein expression. Cells were grown in 5 ml culture overnight and, in the morning, the culture was diluted to an O.D. at 600nm of 0.5. If the gene was under the control of copper inducible promoter, copper sulfate was added to 100 uM and the culture was allowed to grow for at least three hours. Cells were pelleted at 2000 x g for 5 minutes, washed with 10 mM Tris-EDTA buffer pH 7.5. If enzymatic assays were performed, cells were disrupted in assay buffer with glass beads, 2 x times the volume of the pellet. Cells were centrifuged and the supernatant was recovered for enzymatic or protein analysis. Alternatively, if the level and the type of protein was analyzed by SDS-PAGE, cell pellet was suspended in SDS-PAGE buffer and boiled for 5 mins. The suspension was centrifuged, and 10-20 ul aliquots were run on 12 % SDS-PAGE.

Measurement of β- GUS activity from yeast:

β-Glucuronidase (GUS) is a 65 kDa protein that is a useful marker for protein trafficking. We have used GUS to determine the role of N-terminal ubiquitin on secretion of GUS in yeast. Yeast cells were transformed with various GUS vectors, grown overnight in selective liquid media at 30 ° C, and diluted in the liquid selective media to 0.1 OD600 (OD culture). Yeast cells were incubated in the presence of inducer in shaker at 30 ° C. After 4 hours of incubation, 100µl of 2 x "Z" Sarcosine-ONPG buffer (120 mM Na2HPO4, 80 mM NaH2PO4, 20mM KCl, 2 mM MgSO4, 100 mM β-mercaptoethanol, pH 7.0, 0.4% lauroyl sarcosine) was added. (The 2x "Z" Sarcosine- buffer is freshly prepared or stored at -20 ° C prior use.) We used a fluorometric assay with 4-methylumbelliferyl β-D-glucuronide as the substrate for *β- GUS* assay. After incubation at 37 ° C for 1 hour (t incubation), the reaction was stopped by adding 100 µl of quenching solution, 0.5 M Na_2CO_3. The GUS activity was determined by reading the plates in a fluorometric plate reader. For calorimetric reactions, relative activity was calculated as following: (1000 x OD reaction)/(t incubation x OD culture).

E. coli growth, compartmentalization and protein expression.

Protein expression studies were carried out in the Rosetta bacterial strain (Novagen). This strain is derived from the lambda DE3 lysogen strain and carries a chromosomal copy of the IPTG inducible T7 RNA polymerase along with tRNAs on a pACYC based plasmid. Cultures were grown in LB as well as minimal media and at growth temperatures of 37° C and 20° C with 100ug/mL ampicillin and 30 ug/mL chloramphenicol. The culture was diluted 50 fold and grown to mid log (OD at 600 nm = 0.5-0.7), at which time the culture was induced with 1mM IPTG. Induction was allowed to proceed for 4-5 hrs. Upon completion of induction, cells were centrifuged and resuspended in a buffer containing 20% sucrose. To analyze protein induction in total cells, SDS-PAGE buffer was added and the protein was analyzed following SDS-PAGE and staining with Coomassie blue.

Separation of soluble and insoluble fractions.

E.coli were harvested by mild centrifugation and washed once with PBS buffer. Cells were resuspended in 4 ml of PBS and ruptured by several pulses of sonication. Unbroken cells were removed by mild centrifugation (5 min at 1500 x g) and supernatants were sonicated again to ensure complete cell lysis. An aliquot (5 µl) was mixed with 2% SDS to ensure that no viscosity is detected owing to lysis of unbroken cells. After ensuring that no unbroken cells remained in the lysate, insoluble material consisting of cell walls, inclusion bodies and membrane fragments was sedimented by centrifugation (18,000 x g for 10 min). The supernatant was considered "Soluble fraction".

The pellets were washed from any remaining soluble proteins, lipids and peptidoglycan as follows. Pellets were resuspended in 600 µl of PBS and to the suspensions 600 µl of solution containing 3 M urea and 1% Triton X100 was added. The suspension was briefly vortexed and insoluble material was collected by centrifugation as above. The PBS/Urea/Triton wash was repeated two more times to ensure complete removal of soluble proteins. The washed pellets, designated as "insoluble fraction," consisted primarily of inclusion bodies formed by over

expressed proteins. Approximately 10μg of protein from each fraction was resolved on 12% SDS-PAGE minigels and stained with Coomassie Brilliant Blue.

Fluorescence (GFP activity) assessment.

GFP fluorescence was measured in soluble fractions (approx. 0.1 mg of soluble protein in a final volume of 40 μl) using Fluoroscan Accent FL fluorometer (LabSystems) with Excitation 485 nm/ Emission 510 nm filter set with the exposure set to 40 sec. The data are presented in Arbitrary Units (AU).

Western blotting.

Twenty μg of total yeast protein per lane were resolved on 12% SDS-PAGE minigel and electro-blotted to nitrocellulose membranes by standard methods. Membranes were blocked with 5% milk in TTBS buffer and incubated with rabbit anti-GFP antibodies (Clontech, cat no. 8367) at 1:100 dilution overnight at 4° C. Secondary HRP-conjugated antibodies were from Amersham. Identical gels were run in parallel and stained with Coomassie to ensure equal loading of the samples.

The various 6HisxSUMO-GFP (16) fusions were expressed in Rosetta(DE3) pLysS (Novagen) using the procedures recommended by the manufacturer. Expression levels in the absence and presence of the fusion proteins was compared by SDS-PAGE analysis. The recombinant proteins were purified using Ni-NTA agarose; (Qiagen) using procedures recommended by the manufacturer.

Cleavage of proteins

For studies in E.coli, an organism that does not possess SUMO or ubiquitin cleaving enzymes, each cleavage reaction contained 100 ul of purified fusion protein, 99 ul of the buffer 20 mM Tris-HCl pH 8.0, 150 mM NaCl , 5 mM β-mercaptoethanol, and 1 ul of enzyme. The reactions were incubated for 3 hours at 30 ° C, and stopped by addition of 6x Laemmli SDS-page loading buffer followed by boiling at 95 ° C for 5 minutes. The products of the cleavage reaction were analyzed by SDS-PAGE.

The following examples are provided to illustrate various embodiments of the present invention. They are not intended to limit the invention in any way.

EXAMPLE I

Attachment of C-Terminus of UBLs to N-Terminus of GFP Enhances the Expression and Solubility of the Protein in E.coli.

The design and construction of all the UBL E.coli expression vectors has been described above. The DNA sequences, accession numbers of the UBL-GFP fusion proteins, and translation frames are shown Figures 25-32. Fig 4A shows the 37 °C expression pattern of GFP, Ub-GFP, SUMO-GFP, Urm1-GFP, Hub1-GFP, Rub1-GFP, Apg8-GFP, Apg12-GFP, ISG15-GFP. Un-fused GFP is generally poorly expressed in E.coli. The data show that all of the UBLs enhance the expression level of GFP to varying degrees. However, the greatest amount of induction was observed with Ub, SUMO, Urm1, Apg8 and Apg12. Induced cells were broken by sonication and soluble proteins were analyzed on SDS-polyacrylamide gels. The stained gel shows (Fig 4A, Soluble Panel) that ubiquitin, SUMO, Urm1, Hub1 and ISG15 were able to solublize the GFP while Rub1, Apg8 and Apg12 fusion proteins were not soluble, however, fusion to these proteins did enhance the level of expression several fold. To determine if the fusion proteins were folded correctly, we determined the fluorescence properties of proteins in the soluble fraction. Fig 4 A also shows GFP fluorescence in approximately 0.1 mg of soluble protein in a final volume of 40 ul using Fluoroscan Accent FL fluorometer (LabSystems) with Excitation 485 nm/ Emission 510 nm filter set with the exposure set to 40 sec. The data are presented in Arbitrary Units (AU) and show that Ub, SUMO, Urm1, Hub1 and ISG15 produced GFP protein that was able to fluoresce and, thus, was folded correctly. Fusions of GFP with Rub1, Apg8 and Apg12 were induced in large amounts but were not soluble and did not show any fluorescence.

In addition, it is shown that ISG15 plays a role in immune response (24). Thus presentation of ISG15 as a fusion protein is a viable tool for novel vaccine

candidates. Similarly, Apg8 and Apg12 translocate protein to compartments in the cell for autophagy (30).

Similar experiments were performed with all the UBL-GFP fusion proteins, but the induction was performed at 26°C overnight. The data shown in Fig 4B confirms the finding in Fig 4A. Almost all of the UBLs except Hub 1 showed dramatically enhanced expression of GFP after fusion. In the case of SUMO, the level of expression was increased about 20 fold. Analysis of soluble fraction showed that Ub, SUMO, Urm and ISG15 were able to solubilize fused GFP (see Fig 4B, Soluble panel). Functional analysis of fusion GFP was performed by fluorescence from the soluble fraction. This data confirms the observation made in Fig 4A. Combining all the data from the induction studies demonstrates that fusion of all the UBLs to GFP enhances expression level from 2-40 fold. In addition, Ub, SUMO, Urm1, Hub1 and ISG15 also increase the solubility of the GFP. These UBLs are therefore capable of producing correctly folded proteins in E.coli.

To gain more insight into the role of UBLs in enhancement of expression and solubility, we have tested the SUMO-fusion systems with other proteins as well. Serine threonine kinases, tyrosine kinase and human nuclear receptor have proven difficult to express in E.coli. Researchers have opted to use tissue culture systems to express soluble kinases of receptors. Fig 5 shows expression 6His-SUMO-Tyr-Kinase and unfused Tyr-Kinase in E.coli using LB or minimal medium (MM), and purified on Ni-NTA resin as described previously. The small fraction of resin was boiled with 1xSDS-PAGE sample buffer and aliquots were resolved on the 12% SDS-PAGE. Equal amounts of E.coli culture were taken for SUMO-Tyr-kinase and unfused Tyr-kinase and purification was performed under identical conditions. The stained gel in Fig 5 shows that SUMO fusion increases the yield of the kinase at least 20 fold, in cells grown in LB media. Fig 6 also shows the pattern of the SUMO-Try kinase that was eluted from Ni-NTA by 100 mM EDTA or 250 mM imidazole. These data further demonstrate that SUMO fusion enhances the expression of difficult to express protein such as Tyr-kinase, and that the expressed fusion protein is soluble.

Human nuclear receptor proteins, such as steroid receptors, contain ligand-binding domains. These proteins have proven hard to express in soluble form in E.coli. We have used human liver X receptor (LXR) ligand binding domain to demonstrate that SUMO fusion promotes solubility of the protein in E.coli. The ligand-binding domain of LXR was expressed as SUMO fusion in Rosetta plysS cell at 20°C or 37°C and the pattern of soluble and insoluble protein was analyzed. Fig 7 shows the stained SDS-polyacrylamide gel demonstrating that about 40% of the LXR protein was solublized by SUMO fusion, see lane CS in 20°C box in Fig 7 (predominant band in 40kDa range). If the cells were induced at 37°C, hardly any SUMO-LXR was soluble although the level of protein induction had increased dramatically. Further proof that SUMO promotes solubility of previously insoluble proteins was gained by expressing MAPKAP2 kinase as a SUMO-fusion in E.coli. Figures 8A and 8B shows induction kinetics in E.coli cells expressing kinase at 20°C and 37°C. Numbers at the top of the gel, 0.1, 0.25 and 0.5 refer to the mM concentration of inducer IPTG, in the culture. The original induced culture (I), supernatant from lysed cells (S) and resuspended pellet (P) were analyzed on 12% SDS-PAGE. The data clearly demonstrate that 90% of the SUMO kinase is soluble when the cells are induced at 20°C with 0.25 mM IPTG. Although induction at 37°C allows greater degree of expression, more than 50% of the kinase is still insoluble under these conditions. Cleavage of SUMO-MAPKKAP2 kinase by SUMO hydrolase is described in Example III. Also see Figure 18.

Overall, these results show that in bacteria, fusion of UBLs to GFP increases the level of expression from 2-40 fold. Some of the UBLs such as Ub, SUMO, Urm1, Hub1, and ISG15 solublize otherwise insoluble proteins. In particular, SUMO has been demonstrated to increase solubility of kinases and LXR α under controlled temperature induction from 50-95% of the total expressed protein.

EXAMPLE II

SUMO-FUSION EXPRESSION IN YEAST AND INSECT CELLS

Fusions of C-terminal UBLs to the N-terminus of GFPs are cleaved in Yeast

To further assess the utility of UBL fusion in eukaryotic cells we expressed all of the UBL-GFP fusions previously described in Fig 4 in yeast. S.cerevisiae BJ1991 strain was transformed with either YEp-GFP or YEp-UBL-GFP fusion constructs using standard procedures. Positive clones were grown in YPD medium and induced with 100 μM $CuSO_4$ at cell density OD600 = 0.2 for 3.5 hours. Total cell extracts were prepared by boiling the yeast cells in SDS-PAGE buffer. Twenty ug of proteins were analyzed on 12% SDS gels. A replica gel was stained in Coomassie blue and another gel was blotted and probed with antibodies against GFP. Data in Fig 9 shows that Ub-GFP, SUMO-GFP and ISG15-GFP fusions were efficiently cleaved in yeast, while Rub1-GFP fusion was partially cleaved. Apg8-GFP fusion was cleaved into two fragments. It is noteworthy that all the UBL-GFP fusions were designed with methionine as the first amino terminus. GFP fusion with Urm1, Hub1 and Apg12 expressed well, but were not cleaved in yeast. There was a modest increase in expression of GFP following fusion with Ub, SUMO, ISG15 and cleavage in yeast. Generally we have observed 10-20 fold increase in the level of protein expression following fusion to UBL in prokaryotes and eukaryotes (see Fig 4B, 10 and 11). The reason for the modest increase in GFP fusion following cleavage is that the cells were grown in induction media containing 100 uM copper sulfate in rich YPD media. Rich media contains many copper binding sites, and less free copper is available to induce the gene. A nearly 100-fold increase in GFP production has been observed with a variety of N-terminal fusions when cells were induced with 100 uM copper sulfate in synthetic media. See Fig 10.

Generation of New Amino Termini:

The identity of the N-terminus of a protein has been proposed to control its half-life (the N-end Rule) (35). Many important biopharmaceuticals such as growth factors, chemokines, and other cellular proteins, require desired N-termini for

therapeutic activity. It has not been possible to generate desired N-termini, as nature initiates translation from methionine, but the SUMO system offers a novel way to accomplish this.

To demonstrate that all N-termini of GFP in SUMO-GFP fusions were efficiently cleaved when expressed in yeast, a comprehensive study of SUMO-GFP with 20 N-termini was carried out. Multi-copy yeast expression plasmids were designed as described above. Plasmids were transformed in yeast strain BJ 1991, four single colonies were selected, and the levels and cleavage patterns of two of the strains were analyzed by SDS-PAGE and western blotting. Data from Western blots of a single colony is presented in Fig 10. These results are in agreement with our in vitro studies of purified SUMO-X-GFPs (from E.coli) and its cleavage pattern of SUMO hydrolase. All of the SUMO-GFP fusions were cleaved efficiently except those containing proline at the junction (see Fig 10, middle panel lane "Pro"). It is also interesting to note that SUMO-Ileu-GFP was partially cleaved during the phase of copper induction. All of the genes are under the control of copper inducible promoter. It is possible that SUMO-Ileu-GFP is resistant to cleavage due to the non-polar nature of the residue at the + 1 active site of SUMO hydrolase. In this respect SUMO-Val-GFP was also partially resistant to cleavage in vivo (see lower most panel lane labeled "Val"). It is clear from these results that SUMO-Pro-GFP fusion was completely resistant to cleavage by yeast SUMO hydrolases as no GFP was observed (see lane "pro" in middle panel of Fig 10). This data is consistent with our previous observations . See Figure 15. Another important aspect of these findings is that fusion of SUMO with various N-termini of GFP appears to increase the expression of almost all the proteins, although to various degrees. For example Cys-GFP, Gly-GFP and His-GFP accumulated in greater amounts as compared to other N-terminal GFPs. A direct comparison of the increase in the level of GFP following fusion to SUMO can be made by comparing the level of un-fused GFP (see last lanes of lower most panel in Fig 10). Although 20 ug of yeast proteins were loaded on SDS-PAGE the GFP signal was not detected. To ensure that we were not dealing with mutation or any artifact, we loaded a protein sample from another single colony that was induced in under similar conditions and the sample was loaded next to the

previous GFP. No signal was detected, suggesting that unfused GFP is made in very small amounts that cannot be detected under the present experimental conditions, (i.e., a four hour induction with copper sulfate). These studies show that fusion with SUMO leads to a dramatic increase in the amount of protein expressed in yeast. All of the N-terminal fusions are cleaved by endogenous SUMO hydrolases except when the N-terminal residue is proline. Thus for enhanced expression of a protein in eukaryotes permanent attachment of SUMO is not required as significant (~100 fold) increased accumulation of the protein was observed even after the cleavage of SUMO. At the same time, SUMO-pro-fusions are also useful as 6xHis-SUMO can be used to purify the protein from yeast, and the SUMO moiety can be removed with 10 times greater amounts of the SUMO hydrolase (see example III).

Previous studies have shown that attachment of ubiquitin to the N-termini of proteins in yeast enhances expression, and protein fusions containing all amino acid at the N-terminal residue, except proline, are efficiently cleaved in yeast (2, 10, 34). However, these technologies have several drawbacks. Firstly, none of the deubiquitinating enzymes (DUBs) have been shown to efficiently cleave ubiquitin fusion proteins of varying sizes and structures (3,1), despite the fact that they were discovered more than 15 years ago (35, 19, 3). Secondly, and perhaps more importantly, ubiquitin predominantly functions as a signal for proteolysis(14). Therefore, for physiological reasons and for the lack of robust cleavage of artificial ubiquitin-fusions by DUBs, the ubiquitin gene fusion system has not been successfully developed for commercial applications. We have observed that the SUMO system appears to perform in a manner that is remarkably superior to that of ubiquitin, as SUMO and other UBL fusions enhance protein expression and solubility in prokaryotes. In addition, many of the UBLs increase expression of GFP, following the cleavage of UBL in yeast. Unlike the ubiquitin-fusion system, which may direct the protein to the ubiquitin proteosome pathway, the current cleavage of fusion-protein in yeast is the result of C-terminal fusion with SUMO, and proteins generated with novel N-termini are not subject to degradation by the ubiquitin-proteosome pathway. This is one of the reasons that large amount of GFP has accumulated in yeast after cleavage of the SUMO fusion (see Fig 10).

N-terminal Attachment of ubiquitin Promotes protein Secretion:

To date, a role for ubiquitin in the secretion of proteins has not been determined. We have assessed whether N-terminal fusion of ubiquitin to a protein promotes its secretion in yeast. Several yeast expression vectors that express E.coli β-glucoronidase (GUS) were designed. All of the yeast GUS expression vectors described in Table 2 are engineered under the control of the strong glycolytic GPD promoter that expresses constitutively. Some of the constructs were also expressed under the control of a copper regulated metallothionein promoter (CUP1) as well. CUP1 promoter driven synthesis of the SUMO-GUS constructs was induced by addition of 100 uM copper sulfate and incubation of 3 hours. To determine the level of GUS from media, cells were harvested by centrifugation at 2000 x g for 10 mins. Supernatant was collected and equal amounts of aliquots were assayed for enzymatic activity or western blot analysis as described above. For the comparative study, all strains were treated identically and grown at the same time to equal O.D, and the assays were performed at the same time. To examine intracellular enzymatic activity, the cells were harvested by centrifugation and washed with Tris EDTA buffer, pH 7.5. The cell pellets were suspended in sarcosine buffer and ruptured with glass beads at 4°C, three times by vigorously vortexing. Supernatant was collected for assay of the enzymatic activity. The amount of protein secretion was determined by estimating relative activity of the enzyme in the media. The data is shown in Table 2.

Table 2

Ubiquitin-GUS Expression and Secretion in Yeast

Vector (pRS425)	Promoter	Signal Sequence	GUS Activity Inside Cell	GUS Activity In Supernatant
ADH1-GUS1	ADH1	-	+++	–
GPD-α-factor-GUS1	GPD	α-factor	++	
GPD-Ub-GUS1	GPD	Ubiquitin	++++	++++
GPD-Ub-α-factor-GUS1	GPD	Ubiquitin-α-factor	++++	–
GPD-α-factor-Ub(pro)-GUS1	GPD	α-factor-Ubiquitin(pro)	++	–
GPD-α-factor-Ub(met)-GUS1	GPD	α-factor-Ubiquitin(met)	++	–
CUP1-Ub-GUS1	CUP1	Ubiquitin	++++	++

GUS activity was measured as described. It was not possible to measure specific units of GUS in the media as yeast grown in synthetic media. Yeast secretes little protein and current methods of protein estimation, BioRad kit cannot estimate the protein, the data was presented as + where one + is equal to 2 units of GUS as described in invention. – Sign means no GUS activity was detected.

The following conclusions are drawn from this study.

1) Fusion of ubiquitin to GUS leads to a several fold increase when yeast extracts were analyzed by enzymatic assays.
2) Insertion of proline at the junction of ubiquitin and GUS did not allow cleavage of the ubiquitin-GUS fusion protein.

3) The attachment of alpha factor secretory sequences to the N-terminus of ubiquitin-fusion did not have show any appreciable increase in secretion of the protein into the media.

4) Presence of alpha factor sequences between ubiquitin and GUS did not lead to any increase in extracellular level of GUS activity.

5) Greatest amount of secretion was observed with ubiquitin-Met-GUS. These observations suggest that endogenous secretory sequences of GUS in the context of ubiquitin promote the best secretion for GUS. To this end the current data from yeast correlates very well with the ubiquitin-GFP protein secretion in insect cells (see Fig 13).

Fusion of SUMO and Ubiquitin to the N-terminus of GFP Promotes Enhanced Expression and Secretion in Insect Cells.

The role of SUMO in enhanced expression and secretion of proteins in cultured cells has also been studied in insect cells. Baculovirus vectors expressing SUMO-GFP constructs and appropriate controls have been described above. See Fig 11A for the orientation gp67 secretory signals in the SUMO-GFP constructs. Data from a 24 hour infection is shown in Fig 12. Panel A shows intracellular protein analysis by Western blots. It is clear that fusion with ubiquitin and SUMO promotes a large increase in the amount of protein (compare lane E with lane U and S). Insertion of gp67 signal sequences to the N-terminus of SUMO leads to further increase in the amount of protein in insect cells (compare unfused GFP lane E with gp67-SUMO-GFP lane GS). On the other hand attachment of gp67 signal sequence to the N-terminus of GFP (lane G, UG or SG) did not increase the level of protein expression, to the contrary there was diminution of signal when gp67 was attached to N-terminus of GFP(lane G) or between SUMO and GFP (lane SG). We estimate that in the level of expression in the context of gp67-SUMO-GFP is 20 x fold higher as compared to unfused GFP (lane E) or 40 x fold higher as compared to gp67-GFP (lane G). No unfused GFP was secreted by any of the constructs at 24 hour post infection, as shown in blot in Fig 12 panel B. These results show that fusion with

SUMO leads to a dramatic increase in expression of GFP in insect cells. Additionally, both SUMO-GFP and gp67-SUMO-GFP were efficiently cleaved by endogenous SUMO hydrolases.

Similar experiments were performed with cells 48 hours post infection. The data in Fig 13 A and B show that the pattern of intracellular expression was similar to the one seen in 24 hours of infection; however, large amounts of ubiquitin and SUMO-GFP protein were secreted at 48 hour post infection. Examination of the blots from media and intracellular protein show that reasonable expression of unfused GFP was observed inside the cell, but hardly any protein was secreted in the media (compare lane E of panel A and panel B in Fig 13). Attachment of gp67 to the N-terminus of SUMO-GFP leads to the greatest amount of protein secreted into the media (see lane GS in panel B). Another important finding is that attachment of ubiquitin without any signal sequences shows very high secretion of GFP in the media. This result is completely consistent with our finding that attachment of ubiquitin to the N-terminus of GUS promotes the greatest amount of secretion of GUS into the yeast media.

We have also discovered that SUMO-Pro-GFP fusion was not cleaved by endogenous SUMO hydrolases in insect cells (Fig 13 C). Although some non-specific degradation of SUMO-Pro-GFP was observed in these experiments (see lane S-P in Fig 13 C), we conclude that unlike SUMO-GFP, SUMO-Pro-GFP is not cleaved in insect cells. This observation is also consistent with the finding in yeast that SUMO-Pro-GFP is not cleaved in cells while other N-terminal GFP fusions are processed in yeast.

Further confirmation of these observations was obtained by fluorescence imaging of the cells expressing GFP fusion proteins. Fig 14 shows that cells expressing GFP and fusion GFP fluoresce intensely. The fluorescence imaging was the strongest and most widely diffused in cell expressing gp67-SUMO-GFP and Ub-GFP. These cells show the largest amount of GFP secreted into the media (Fig 13 panel B). It appears that secretory signal attachment directly the to N-terminus of GFP produces less GFP in the media and inside the cells. This observation is borne out by low fluorescence intensity and granulated pigmented fluorescence (see panel

G-eGFP, S/G-eGFP and U/G-eGFP). These data have led to the following conclusions:

1) The increase in the amount of SUMO-fusion protein expression in insect cells was several-fold higher (20-40 fold) than that of unfused protein, as determined by and Western blot analysis.
2) All of the SUMO-GFP constructs that contain methionine at the +1 position were cleaved except SUMO-Proline-GUS. This aspect of the SUMO-fusion technology allows us to express proteins that are stably sumoylated.
3) Attachment of ubiquitin to the N-terminus of GFP led to dramatic enhancement in secretion of the protein in the media. Ubiquitin promotes secretion of proteins that may or may not have endogenous secretory signal. Thus, N-terminal ubiquitination may be utilized as a tool to enhance secretetion of proteins in eukaryotic cells.
4) N-terminal SUMO also promotes secretion of protein in insect cells.

EXAMPLE III

SUMO Protease ULP1 Cleaves A Variety of SUMO-Fusion Proteins: Properties and Applications in Protein and Peptide Expression and Purification

Yeast cells contain two SUMO proteases, Ulp1 and Ulp2, which cleave sumoylated proteins in the cell. At least eight SUMO hydrolases have been identified in mammalian systems. The yeast SUMO hydrolase Ulp1 catalyzes two reactions. It processes full length SUMO into its mature form and it also de-conjugates SUMO from side chain lysines of target proteins. Examples I and II establish our findings that attachment of SUMO to the N-terminus of under-expressed proteins dramatically enhances their expression in E.coli, yeast and insect cells. To broaden the application of SUMO fusion technology as a tool for expression of proteins and peptides of different sizes and structures, the ability of Ulp1 to cleave a variety of proteins and peptides has been examined. Purified recombinant SUMO-GFPs were efficiently cleaved when any amino acid except Proline is present in the +1 position of the cleavage site. Similar properties of

SUMO hydrolase Ulp1 were observed when Sumo-tyrosine kinase, Sumo-protein G, Sumo-β-GUS, and SUMO MAPKAP2 kinase were used as substrates. The *in vitro* activity of the enzyme showed that it was active under broad ranges of pH, temperature, and salt and imidazole concentration. These findings suggest that the Ulp1 is much more robust in cleavage of the SUMO-fusion proteins as compared to its counterpart, ubiquitin-fusion hydrolase. Broad specificity and highly efficient cleavage properties of the Ulp1 indicate that SUMO-fusion technology can be used as a universal tag to purify a variety of proteins and peptides, which are readily cleaved to render highly pure proteins.

The following materials and methods are provided to facilitate the practice of Example III.

Affinity purification and cleavage of SUMO fusion proteins with SUMO hydrolase.

The following table lists the solutions required for the affinity purification and cleavage procedures:

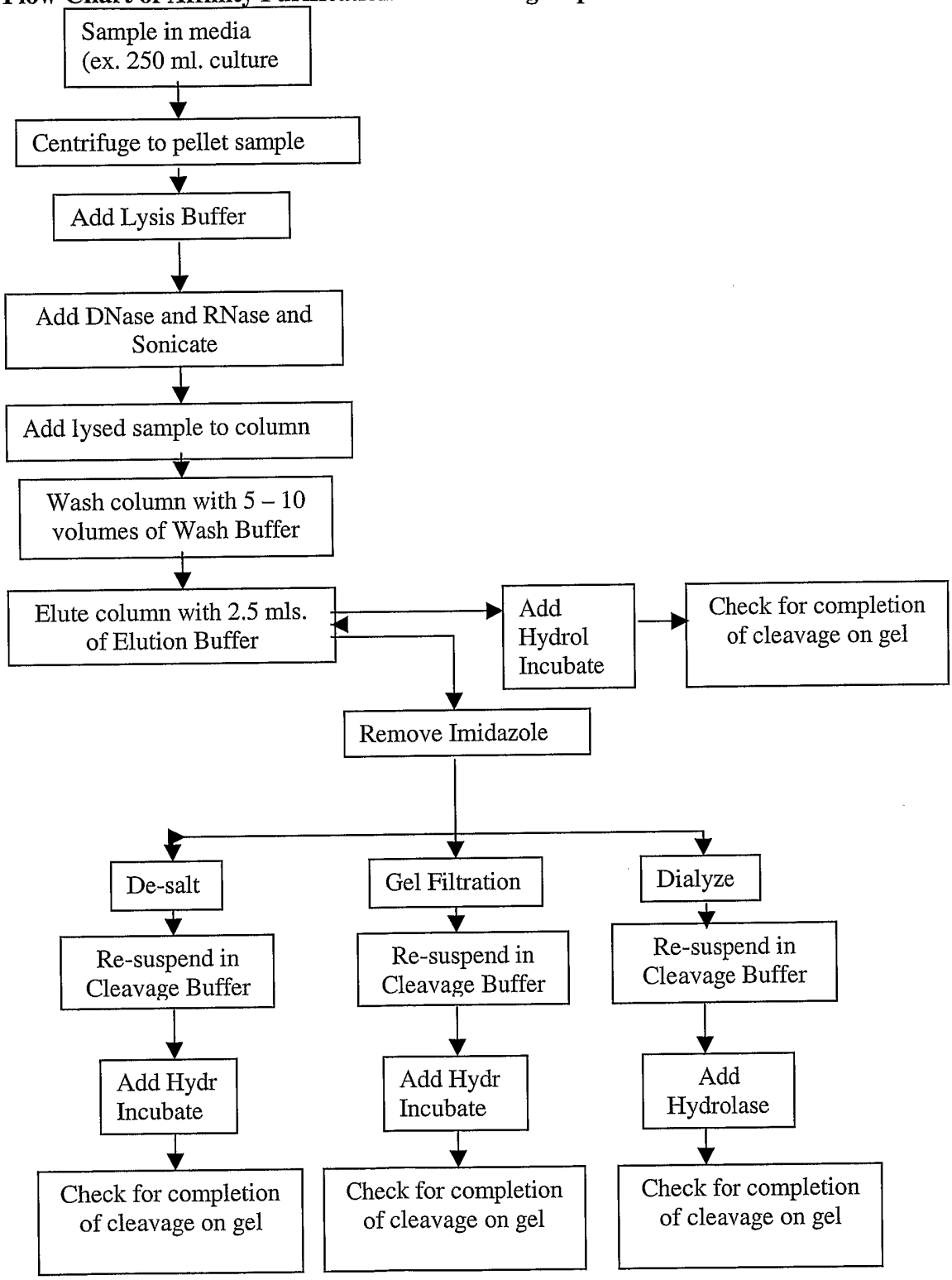
Solution	Components
Lysis buffer	25 mM Tris pH 8.0; 50 mM NaCl
Wash Buffer	25 mM imidazole; 50 mM Tris pH 8.0; 250 mM NaCl; (optional) 5 -10 mM β-mercaptoethanol (protein dependent)
Elution Buffer	300 mM imidazole; 50 mM Tris pH 8.0; 250 mM NaCl; (optional) 5 -10 mM β-mercaptoethanol (protein dependent)
SUMO hydrolase (Ulp1) Cleavage Buffer	50 mM Tris pH 8.0; 250 mM NaCl; 5 mM β-mercaptoethanol (protein dependent)

From typical 250 ml cultures, the samples are pelleted by centrifugation, and supernatants are removed by decanting. Generally, from 250 ml of culture, 1.0 -1.5 grams of wet cells are produced. Pelleted cells are then resuspended in 5-10 ml of lysis buffer. RNase and DNase are added to final concentration of 10 ug/ml lysis solution. Samples are kept on ice throughout the sonication procedure. Using an appropriate tip, the samples are sonicated 3 – 5 times for 10 second pulses at 50% duty cycle. Sonicates are incubated on ice for 30 minutes; if the samples are viscous

after this time, the sonication procedure is repeated. Lysed samples (in lysis solution) are loaded onto 1-ml columns. The columns are washed with 5 to 10 volumes of wash buffer (wash fractions are saved until the procedure is complete). Columns are developed with 2.5 ml of elution buffer, and SUMO hydrolase cleavage is performed by one of two methods: 1) cleavage is performed in elution buffer, with SUMO hydrolase added at 50 ul/ 250 ml buffer, samples incubated at room temperature for 2 hr or overnight at 4°C, and cleavage monitored by gel electrophoresis; 2) imidazole is first removed by dialysis, gel filtration, or desalting, samples are then resuspended in SUMO hydrolase cleavage buffer, SUMO hydrolase is added at 50 ul / 2.5 ml buffer, and samples are incubated at room temperature for 2 hr or at 4°C overnight, with cleavage monitored by gel electrophoresis. Units of SUMO hydrolase are defined as the amount of enzyme that cleaves 1 ug of pure SUMO-Met-GFP (up to 95%) in 50 mM Tris-HCl pH 8.0, 0.5 mM DTT, 150 mM NaCl at room temperature in 60 minutes.

After cleavage, protein can be stored at 4°C, or subjected to purification.

Flow Chart of Affinity Purification and Cleavage Options



The expression and purification of carboxy terminus of Ulp1p is described above.

In vitro cleavage experiments

The various His6smt3XeGFP fusions were expressed in Rosetta (DE3) pLysS (Novagen). The recombinant proteins were purified using Ni-NTA agarose (Qiagen). The comparative *in vitro* cleavage reactions were carried out by first normalizing the amount of the various fusions in each reaction. This was done by measuring the fluorescence properties of the purified fusion proteins using the fluorimeter Fluoriskan II (Lab Systems) and then diluting the more concentrated samples with the Ni-NTA agarose elution buffer (20 mM Tris-HCl pH 8.0, 150 mM NaCl 300 mM Imidazole and 5 mM beta-mercaptoethanol), such that their fluorescence values equaled that of the lowest yielder. Each cleavage reaction contained 100 ul of protein, 99 ul of the buffer 20 mM Tris-HCl pH 8.0, 150 mM NaCl and 5 mM beta-mercaptoethanol and 1 ul of enzyme. The reactions were incubated for 3 hours at 30°C after which they were stopped by addition of 6x Laemmli SDS-page loading buffer followed by boiling at 95°C for 5 minutes. The products of the cleavage reaction were analyzed by SDS-PAGE.

Proline cleavage experiments were carried out in a fashion similar to those described above. The purified His6smt3PeGFP was buffer exchanged into 20 mM Tris-HCl pH 8.0, 150 mM NaCl and 5 mM beta-mercaptoethanol using a PD-10 column. A 10 fold increase in the amount of Ulp1 were added to each reaction. Digestions were incubated for 3 hours at 30°C. All reactions were stopped by addition of Laemmli loading buffer and analyzed by SDS-page.

Fig 15 shows the stained SDS-PAGE analysis of all the SUMO-X-GFPs and their digestion by SUMO hydrolase. The findings clearly show that Ulp1 hydrolase was able to cleave all the SUMO-GFP fusions except proline. These finding are similar to the observations made in yeast (Fig 10) and in insect cells (Fig 13).

Conjugation of ubiquitin and SUMO to its target proteins is a highly regulated and dynamic process. Several deubiquitinating enzymes (DUBs) have

been identified in yeast and other eukaryotic cells(1). Yeast genetics studies show that many of these enzymes are not essential suggesting that an overlapping function is performed by most of these enzymes. DUBs have been most extensively studied and shown to cleave linear ubiquitin fusions as well isopepetide bonds (3, 35). Much less is known about the enzymes that remove SUMO from isopeptide bonds or artificial SUMO-fusion proteins. Hochstrasser and Li have shown that Ulp1 and Ulp2 remove Smt3 and SUMO 1 from proteins and play a role in progression through the G2/M phase and recovery of cells from checkpoint arrest, respectively(20, 21). Ulp1 and Ulp2 cleave C-terminus of SUMO (-GGATY; SEQ ID NO: 59) to mature form (-GG) and de-conjugate Smt3 from the side chains of lysines(20, 21). The sequence similarity of two enzymes is restricted to a 200-amino acid sequence called ULP that contains the catalytically active region. The three-dimensional structure of the ULP domain from Ulp1 has been determined in a complex form with SUMO (Smt3) precursor(27). These studies show that conserved surfaces of SUMO determine the processing and de-conjugation of SUMO. Database searches of the human genome and recent findings suggest that there are at least 7 human ULPs with the size ranging from 238 to 1112 amino acid residues (18, 33, 39) . It is intriguing to note that SUMO Ulps are not related to DUBs, suggesting that SUMO Ulps evolved separately from DUBs. The findings that ULP structure is distantly related to adenovirus processing protease, intracellular pathogen *Chlammydia trachomatis* and other proposed bacterial cystiene protease core domains suggest that this sequence evolved in prokaryotes(20, 21). Detailed properties of the SUMO proteases are provided in described in Table 3.

Table 3

SUMO Hydrolases/Proteases

Enzyme	Properties (MW)	Reference
UB1-specific Protease ULP1	72 KDa. 6 21 residues Cleaves linear fusion and SUMO isopeptides bonds.	Li and Hochstrasser, 1999 (REF 20)
ULP2 (Yeast)	117 KDa, 1034 residues Cleaves linear fusions and SUMO isopeptide structures.	Li and Hochstrasser, 2000 (REF 21)
SUMO-I C-Terminal	30Kda Cleaves linear fusions and SUMO isopeptide structures	Suzuki, et al, 1999 (REF 33)
SUMO-I specific Protease SUSP I (Human)	126 KDa 1112 residues Specific for SUMO-1 fusion but not Smt3 fusion. Does not cleave isopeptide bond.	Kim, et al, 2000 (REF 18)
Sentrin specific Proteases (SENP) SENP1 SENP2 SENP3 SENP4 SENP5 SENP6 SENP7	All of the SENP enzymes have conserved C-terminal region with core catalytic cysteine. The smallest SENP7 is 238 residues and the largest SENP6 is 1112 residues.	Yeh, et al, 2000 (REF 39)

Ulp1 has proven extremely robust in cleaving a variety of SUMO-fusion proteins expressed in E.coli as described in the present example. We have designed SUMO-GFP fusions in which the N-terminal methionine has been replaced with rest of the 19 amino acids. Attachment of 6x His to N-terminus of SUMO afforded easy purification of the 20 SUMO-GFP fusions from E.coli. The enzyme was active under broad ranges of pH, temperature, salts and imidazole concentration and was very effective in cleaving variety of proteins from SUMO fusion that includes BPTI a 6.49 KDa, Protein G a 7 KDa, β-Glucuronidase (GUS) and 110 KDa β-Galactosidase (GAL) genes. These findings suggest that the Ulp1 is much more

robust in cleavage of the SUMO-fusion proteins as compared to its counterpart ubiquitin-fusion hydrolase.

SUMO Protease/Hydrolase is a Robust Enzyme:

Effects of Temperature and Additives

The effects of various additives/conditions and temperature upon the *in vitro* cleavage reaction were determined as follows: His6smt3MeGFP was expressed from pET24d in Rosetta(DE3) pLysS (Novagen). The recombinant protein was purified as before using Ni-NTA agarose (Qiagen) and then buffer exchanged into 20 mM Tris-HCl pH 8.0, 150 mM NaCl and 5 mM β-mercaptoethanol using a PD-10 column (AP Biotech). Cleavage reactions were performed with 100 ug of the purified protein, 0.5 ul of enzyme, the appropriate amount of a stock solution of additive to generate the final concentrations listed in Table 4, plus the exchange buffer up to a final volume of 200 ul. Reactions were incubated for 1 hour at 37°C except for those at 4°C were incubated for 3 hours. The data in Fig 16 shows that Ulp1 was extremely active at 37°C as well as at 4°C. Generally, His tagged proteins are purified on nickel columns and eluted with imidazole. We have discovered that the enzyme was remarkably active at 0-300 mM imidazole concentration. The enzyme was highly active at 0.01% SDS and up to 1% triton X 100. See Table 4. Similarly, chaotropic agents such as urea and did not effect the activity of the enzyme up to 2 M . Ulp1 showed 50% activity at 0.5M concentration of guanadinium hydrochloride (Fig 16 and Table 4). A variety of reagents, including cysteine protease inhibitors, EDTA, PMSF. Pepstatin, Leupeptin, TLCK had no effect on the enzymatic activity (Fig 17 and Table 4). N-ethymaleimide was active only if incubated with the enzyme prior to addition of the substrate. All the data shown in Table 2 demonstrate that this enzyme is extremely robust and thus constitutes a superior reagent for cleaving fusion proteins under variety of conditions.

Table 4

The Effect of Different Conditions on the Ulp1 Hydrolase Activity

Conditions/Additions	Effect
Environmental: Temperature	Ulp1 is active over a broad range of temperatures, cleaving from 4 to 37°C
Salts: Imidazole	Ulp1 shows similar activity in the range of 0 to 300mM
Detergents: SDS Triton-X	0.01% SDS blocks activity Ulp1 shows similar activity on the range of 0 to 0.1%
Chaotrophs Urea Gdm HCl	Ulp1 shows complete activity up to and including a 2M concentration Ulp1 shows 50% activity in 0.5M but is completely inactive in 1M concentrations
Protease inhibitors: E-64 EDTA PMSF Pepstatin Leupeptin TLCK-HCl N-ethylmaleimide	 Cysteine protease inhibitor; no affect Metalloprotease inhibitor; no affect Serine protease inhibitor; no affect Aspartate protease inhibitor; no affect Inhibits serine and cysteine proteases with trypsin-like specificity; no affect Inhibits serine and cysteine proteases with chymotrypsin-like specificity; no affect Cysteine protease inhibitor; on effective if enzyme is preincubated with inhibitor before addition of substrate

Robust Properties of SUMO Hydrolase: Cleavage of Different Size Fusion Proteins Under Broad pH Range:

Fig 18 shows purification of a 40 kDa MAPKAP2 kinase that was difficult to express unless fused to SUMO. We have shown in Example I (Fig 8) that this kinase was expressed in a highly soluble form (95%) as fusion to SUMO. Fig 18 shows that whether purfied from cells expressing at 37°C or 20°C, the SUMO fusion was efficiently cleaved under the conditions described.

The SUMO hydrolase also functions under broad pH range. Fig 19 shows kinetics of cleavage at pH 7.5 and 8.0. The data shows that purified SUMO-GFP was completely digested at room temperature. We have also performed experiments from pH 5.5 to 10. The data (not shown) support the notion that this enzyme is active over broad range of pH.

As discussed above, for broad utility of the system it is important that the enzyme be able to cleave fusion proteins of different sizes and structures *in vitro*. Fig 20 shows the digestion pattern of SUMO-β-galactosidase (β-Gal) a 110 KDa protein. β-Gal enzyme is composed of tetrameric subunits. The digestion pattern demonstrates that in 20 minutes, SUMO hydrolase was able to cleave 100% of the protein.

Among dozens of proteins expressed as SUMO fusions in our lab, only one, β-GUS, proved partially resistant to cleavage by the hydrolase. Configurations of artificial SUMO fusion are bound to occur wherein the structure of the protein will hinder the ability of the enzyme to recognize and bind the cleavage site of the fusion protein. This problem has been solved by adding small concentrations of urea, which does not inhibit the hydrolase, but results in cleavage the fusion that was previously resistant. Fig 21 shows the digestion pattern of purified β-GUS and SUMO hydrolase before and after addition of urea. Lane 6 and 9 contain the same amount of SUMO hydrolase to which 2M urea was added during the incubation. Addition of urea allowed complete cleavage of 65KDa β-GUS in 20 min at room temperature. This data further proves that the SUMO hydrolase cleaves broad spectrum of fusion protein efficiently. Additives such as urea can be added to aid complete cleavage of these structures that are resistant to hydrolase action.

High Throughput Protein Purification of Fusion Proteins: Rapid Peptide Miniprep

We have discovered that, due to the rapid folding properties of SUMO, the fused protein can also be rapidly re-natured after treatment of the crude protein mix with chaotropic agents such as guanidinium hydrochloride or urea. We have developed a simple and rapid procedure to purify SUMO-fused proteins that are expressed in prokaryotes and eukaryotes. This method was tested with SUMO-protein G fusion expressed in E.coli. Cells expressing 6xHis-SUMO-G protein fusion were harvested and frozen until required for protein purification. Three times the weight per volume lysis buffer (6 M Guanidinium Chloride, 20 mM Tris-HCl, 150 mM NaCl, pH 8.0) was added to the cell pellet rapidly lyse the cells. The supernatant was loaded onto a pre-equilibrated column containing Ni-NTA agarose (Qiagen), the flow through was collected for analysis. The column was then washed, first with 2 column volumes (CV) of Lysis buffer, followed by 3 CV of wash buffer (20 mM Tris-HCl, 150 mM NaCl 15 mM Imidazole pH 8.0). The fusion protein was then eluted using 2 CV of elution buffer (20 mM Tris-HCl, 150 mM NaCl 300 mM Imidazole pH 8.0). The purified product is present in a native buffer that allows for cleavage and release of the peptide from the Sumo fusion using Ulp1. See Figure 22. This data demonstrates that it is possible to rapidly purify the fusion protein and cleave it from the resin with Ulp1. It is possible that proteins of higher molecular weights may not rapidly re-nature and be amenable to cleavage by Ulp1. However, since the Ulp1 requires three-dimensional SUMO be intact the purification and cleavage properties are more dependent on the refolding of SUMO. Similar to DNA mini-preps, rapid mini preps for the expression and purification analysis of the fused proteins may be readily employed. Table 5 summarizes the data showing the dramatic enhancement of protein production observed when utilizing the compositions and methods of the present invention. The sequences and vectors utilized in the practice of the invention are shown in Figures 23-46.

Table 5

Fusion with SUMO Enhances Protein Expression

E.coli Expression of UBLs	**All of the fusion have Met N-Termini**
SUMO-GFP Ub-GFP Urm1-GFP Hub1-GFP Rub1-GFP Apg8-GFP Apg12-GFP ISG15-GFP	40 fold 40 fold 50 fold 2 fold 50 fold 40 fold 20 fold 3-5 fold
Yeast	**Met and Various N-Termini**
Various UBLs expressed in rich media.	Copper induction not observed in rich media, however, Ub, SUMO, ISG15 fusions were processed and GFP induced 3-5 fold.
All of the twenty N-terminal variants were expressed in yeast as SUMO-X-GFP fusions. GFP was processed in all cases, except when N-terminal residue was proline.	Dramatic induction of GFP following fusion with SUMO. At least 50-100 fold induction as compared to unfused GFP expression. Under current loading conditions (20 ug) GFP was not detectable.
Insect Cells	**Met as N-termini**
SUMO-GFP gp67-SUMO-GFP gp67-SUMO-GFP Secretion SUMO-GFP Secretion Ub-GFP	10 fold compared to GFP 30 fold compared to gp-GFP 50 fold compared to SUMO-gp67-GFP At least 50 fold compared to GFP At least 50 fold compared to GFP

References:

1. **Amerik, A. Y., S. J. Li, and M. Hochstrasser.** 2000. Analysis of the deubiquitinating enzymes of the yeast Saccharomyces cerevisiae. Biol Chem **381:**981-92.
2. **Bachmair, A., D. Finley, and A. Varshavsky.** 1986. In vivo half-life of a protein is a function of its amino-terminal residue. Science **234:**179-86.
3. **Baker, R. T.** 1996. Protein expression using ubiquitin fusion and cleavage. Curr Opin Biotechnol **7:**541-6.
4. **Bayer, P., A. Arndt, S. Metzger, R. Mahajan, F. Melchior, R. Jaenicke, and J. Becker.** 1998. Structure determination of the small ubiquitin-related modifier SUMO-1. J Mol Biol **280:**275-86.
5. **Butt, T. R., S. Jonnalagadda, B. P. Monia, E. J. Sternberg, J. A. Marsh, J. M. Stadel, D. J. Ecker, and S. T. Crooke.** 1989. Ubiquitin fusion augments the yield of cloned gene products in Escherichia coli. Proc Natl Acad Sci U S A **86:**2540-4.
6. **Butt, T. R., E. J. Sternberg, J. A. Gorman, P. Clark, D. Hamer, M. Rosenberg, and S. T. Crooke.** 1984. Copper metallothionein of yeast, structure of the gene, and regulation of expression. Proc Natl Acad Sci U S A **81:**3332-6.
7. **Ecker, D. J., J. M. Stadel, T. R. Butt, J. A. Marsh, B. P. Monia, D. A. Powers, J. A. Gorman, P. E. Clark, F. Warren, A. Shatzman, and et al.** 1989. Increasing gene expression in yeast by fusion to ubiquitin. J Biol Chem **264:**7715-9.
8. **Gietz, D., A. St. Jean, R. A. Woods, and R. H. Schiestl.** 1992. Improved method for high efficiency transformation of intact yeast cells. Nucleic Acids Res **20:**1425.
9. **Goward, C. R., J. P. Murphy, T. Atkinson, and D. A. Barstow.** 1990. Expression and purification of a truncated recombinant streptococcal protein G. Biochem J **267:**171-7.
10. **Graumann, K., J. L. Wittliff, W. Raffelsberger, L. Miles, A. Jungbauer, and T. R. Butt.** 1996. Structural and functional analysis of N-terminal point mutants of the human estrogen receptor. J Steroid Biochem Mol Biol **57:**293-300.
11. **Hicke, L.** 1997. Ubiquitin-dependent internalization and down-regulation of plasma membrane proteins. Faseb J **11:**1215-26.
12. **Hochstrasser, M.** 2000. Evolution and function of ubiquitin-like protein-conjugation systems. Nat Cell Biol **2:**E153-7.
13. **Hochstrasser, M.** 1995. Ubiquitin, proteasomes, and the regulation of intracellular protein degradation. Curr Opin Cell Biol **7:**215-23.
14. **Hochstrasser, M.** 1996. Ubiquitin-dependent protein degradation. Annu Rev Genet **30:**405-39.
15. **Jentsch, S., and G. Pyrowolakis.** 2000. Ubiquitin and its kin: how close are the family ties? Trends Cell Biol **10:**335-42. _00001785 _00001785.
16. **Johnson, E. S., I. Schwienhorst, R. J. Dohmen, and G. Blobel.** 1997. The ubiquitin-like protein Smt3p is activated for conjugation to other proteins by an Aos1p/Uba2p heterodimer. Embo J **16:**5509-19.
17. **Kapust, R. B., and D. S. Waugh.** 1999. Escherichia coli maltose-binding protein is uncommonly effective at promoting the solubility of polypeptides to which it is fused. Protein Sci **8:**1668-74.
18. **Kim, K. I., S. H. Baek, Y. J. Jeon, S. Nishimori, T. Suzuki, S. Uchida, N. Shimbara, H. Saitoh, K. Tanaka, and C. H. Chung.** 2000. A new SUMO-1-specific protease, SUSP1, that is highly expressed in reproductive organs. J Biol Chem **275:**14102-6.
19. **LaBean, T. H., S. A. Kauffman, and T. R. Butt.** 1995. Libraries of random-sequence polypeptides produced with high yield as carboxy-terminal fusions with ubiquitin. Mol Divers **1:**29-38.
20. **Li, S. J., and M. Hochstrasser.** 1999. A new protease required for cell-cycle progression in yeast. Nature **398:**246-51.
21. **Li, S. J., and M. Hochstrasser.** 2000. The yeast ULP2 (SMT4) gene encodes a novel protease specific for the ubiquitin-like Smt3 protein. Mol Cell Biol **20:**2367-77.

22. **Lyttle, C. R., P. Damian-Matsumura, H. Juul, and T. R. Butt.** 1992. Human estrogen receptor regulation in a yeast model system and studies on receptor agonists and antagonists. J Steroid Biochem Mol Biol **42:**677-85.

23. **Mahajan, R., L. Gerace, and F. Melchior.** 1998. Molecular characterization of the SUMO-1 modification of RanGAP1 and its role in nuclear envelope association. J Cell Biol **140:**259-70.

24. **Malakhova, O., M. Malakhov, C. Hetherington, and D. E. Zhang.** 2002. Lipopolysaccharide activates the expression of ISG15-specific protease UBP43 via interferon regulatory factor 3. J Biol Chem **277:**14703-11.

25. **Marathe, S. V., and J. E. McEwen.** 1995. Vectors with the gus reporter gene for identifying and quantitating promoter regions in Saccharomyces cerevisiae. Gene **154:**105-7.

26. **Matunis, M. J., J. Wu, and G. Blobel.** 1998. SUMO-1 modification and its role in targeting the Ran GTPase-activating protein, RanGAP1, to the nuclear pore complex. J Cell Biol **140:**499-509.

27. **Mossessova, E., and C. D. Lima.** 2000. Ulp1-SUMO crystal structure and genetic analysis reveal conserved interactions and a regulatory element essential for cell growth in yeast. Mol Cell **5:**865-76.

28. **Muller, S., C. Hoege, G. Pyrowolakis, and S. Jentsch.** 2001. SUMO, ubiquitin's mysterious cousin. Nat Rev Mol Cell Biol **2:**202-10.

29. **Muller, S., M. J. Matunis, and A. Dejean.** 1998. Conjugation with the ubiquitin-related modifier SUMO-1 regulates the partitioning of PML within the nucleus. Embo J **17:**61-70.

30. **Ohsumi, Y.** 2001. Molecular dissection of autophagy: two ubiquitin-like systems. Nat Rev Mol Cell Biol **2:**211-6.

31. **Sherman, F., G. Fink, and J. Hicks.** 1986. Methods in yeast genetics. Cold Spring Harbor Laboratory, Cold Spring Harbor, N.Y.

32. **Sikorski, R. S., and P. Hieter.** 1989. A system of shuttle vectors and yeast host strains designed for efficient manipulation of DNA in Saccharomyces cerevisiae. Genetics **122:**19-27.

33. **Suzuki, T., A. Ichiyama, H. Saitoh, T. Kawakami, M. Omata, C. H. Chung, M. Kimura, N. Shimbara, and K. Tanaka.** 1999. A new 30-kDa ubiquitin-related SUMO-1 hydrolase from bovine brain. J Biol Chem **274:**31131-4.

34. **Varshavsky, A.** 1996. The N-end rule: functions, mysteries, uses. Proc Natl Acad Sci U S A **93:**12142-9.

35. **Varshavsky, A.** 2000. Ubiquitin fusion technique and its descendants. Methods Enzymol **327:**578-93.

36. **Waldo, G. S., B. M. Standish, J. Berendzen, and T. C. Terwilliger.** 1999. Rapid protein-folding assay using green fluorescent protein. Nat Biotechnol **17:**691-5.

37. **Walfish, P. G., T. Yoganathan, Y. F. Yang, H. Hong, T. R. Butt, and M. R. Stallcup.** 1997. Yeast hormone response element assays detect and characterize GRIP1 coactivator-dependent activation of transcription by thyroid and retinoid nuclear receptors. Proc Natl Acad Sci U S A **94:**3697-702.

38. **Wright, L. C., J. Seybold, A. Robichaud, I. M. Adcock, and P. J. Barnes.** 1998. Phosphodiesterase expression in human epithelial cells. Am J Physiol **275:**L694-700.

39. **Yeh, E. T., L. Gong, and T. Kamitani.** 2000. Ubiquitin-like proteins: new wines in new bottles. Gene **248:**1-14.

While certain of the preferred embodiments of the present invention have been described and specifically exemplified above, it is not intended that the invention be limited to such embodiments. Various modifications may be made thereto without departing from the scope and spirit of the present invention, as set forth in the following claims.

What is claimed is:

1. A method for enhancing expression levels of a protein of interest in a host cell comprising i) operably linking a nucleic acid sequence encoding molecule selected from the group consisting of SUMO, RUB, HUB, APG8, APG12, URM1, and ISG15 to a nucleic acid sequence encoding said protein of interest thereby generating a construct encoding a fusion protein, ii) introducing said nucleic acid into said host cell, whereby the presence of said molecule in said fusion protein increases the expression level of said protein of interest in said host cell.

2. The method of claim 1, wherein said operably linked molecule is SUMO.

3. The method of claim 1, wherein said host cell is selected from the group consisting of a yeast cell, E. coli, and an insect cell.

4. The method of claim 1, further comprising isolation of said fusion protein.

5. The method of claim 4, further comprising cleavage of said fusion protein to release said protein of interest.

6. The method of claim 3, wherein said host cell is an E. coli cell and said molecule is SUMO, further comprising removal of said SUMO molecule in vitro with a protease.

7. The method of claim 3, wherein said host cell is a yeast cell and said molecule is SUMO, further comprising removal of said SUMO molecule in vitro with a protease.

8. The method of claim 3, wherein said host cell is a yeast cell and said molecule is SUMO, further comprising removal of said SUMO molecule in vivo with a Ulp1.

9. A method for generating an altered amino terminus in a protein of interest in a host cell comprising;

a) providing a nucleic acid sequence encoding said protein;

b) altering the N-terminal amino acid coding sequence in said nucleic acid;

c) operably linking a SUMO molecule to said nucleic acid sequence; and

d) expressing said nucleic acid in a eukaryotic cell, thereby producing said protein of interest in said cell, said eukaryotic cell expressing endogenous SUMO cleaving enzymes, said enzyme effecting cleavage of SUMO the target protein coding sequence, thereby producing a protein of interest having an altered amino terminus.

10. A method for producing a sumolated protein for tracking protein localization within a host cell, comprising;

a) providing a nucleic acid sequence encoding said protein;

b) substituting the N-terminal amino acid coding sequence in said nucleic acid for a codon which encodes proline;

c) operably linking a SUMO molecule to said nucleic acid sequence; and expressing said SUMO linked protein in said host cell.

11. The method of claim 10, further comprising detecting localization of said sumolated protein in said host cell.

12. A method for enhancing secretion levels of a protein of interest from a host cell comprising i) operably linking a nucleic acid sequence encoding molecule selected from the group consisting of SUMO, RUB, HUB, URM1, and ISG15 to a nucleic acid sequence encoding said protein of interest thereby generating a construct encoding a fusion protein, ii) introducing said nucleic acid into said host cell, whereby the presence of said molecule in said fusion protein increases the secretion of said protein of interest from said host cell.

13. The method of claim 12, wherein said operably linked molecule is SUMO.

14. The method of claim 12, wherein said host cell is selected from the group consisting of a yeast cell, E. coli, and an insect cell.

15. The method of claim 12, further comprising isolation of said fusion protein.

16. The method of claim 12, further comprising cleavage of said fusion protein to release said protein of interest.

17. The method of claim 1, wherein said SUMO molecule consists of SEQ ID NO: 2.

18. The method of claim 12, wherein said SUMO molecule consists of SEQ ID NO: 2.

Figure 1

Scheme for Ubiquitin and Ubiquitin-Like Proteins Processing and Conjugation

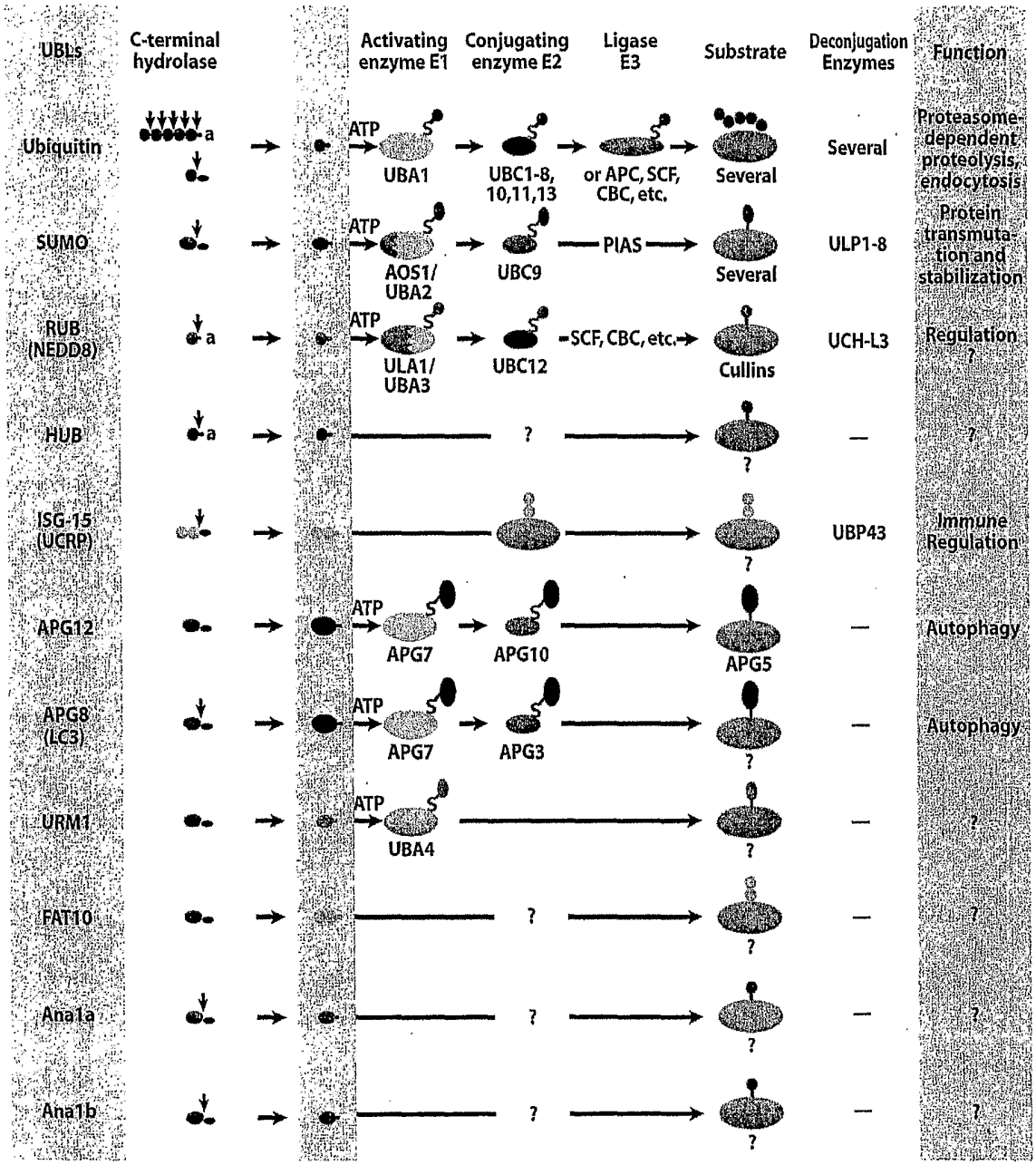


Figure 2

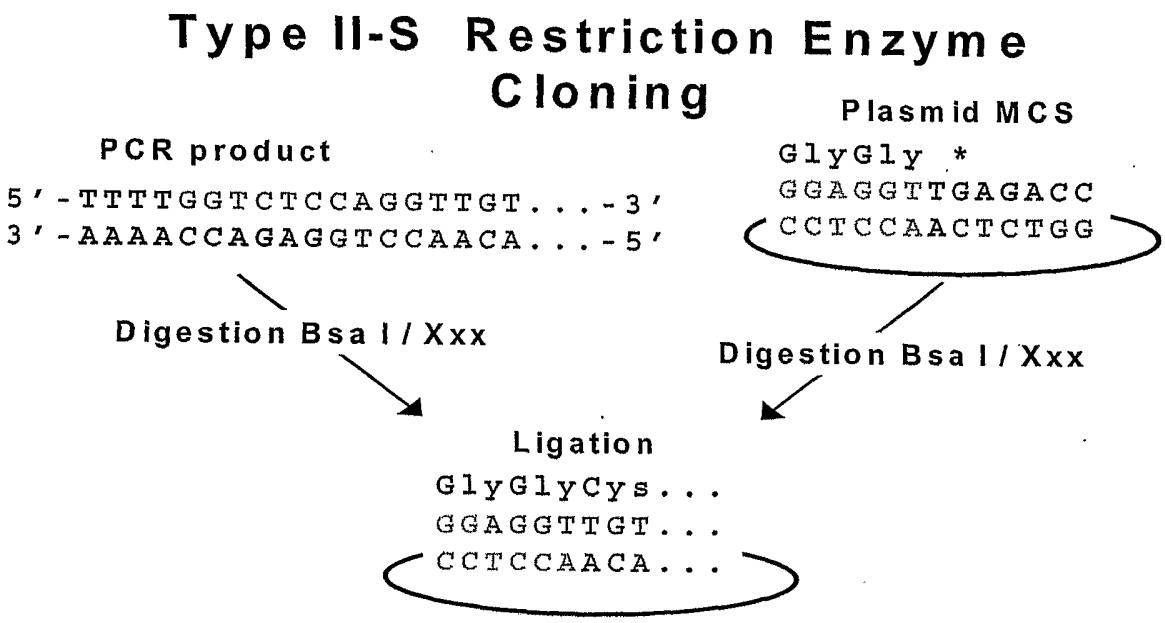
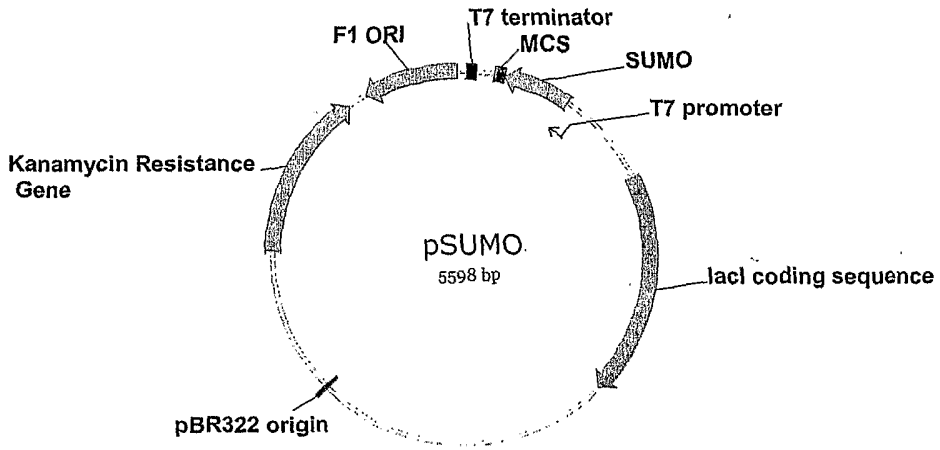

Type II-S Restriction Enzyme Cloning
PCR product
5'-TTTTGGTCTCCAGGTTGT...-3'
3'-AAAACCAGAGGTCCAACA...-5'
Plasmid MCS
GlyGly *
GGAGGTTGAGACC
CCTCCAACTCTGG
Digestion BsaI/Xxx
Digestion BsaI/Xxx
Ligation
GlyGlyCys...
GGAGGTTGT...
CCTCCAACA...

Figure 3



Multiple Cloning Site:

```
        BglII                                                                  XbaI
        ------                                                                 ------
  1     AGATCTCGATCCCGCGAAATTAATACGACTCACTATAGGGGAATTGTGAGCGGATAACAATTCCCCTCTAGAAATAATTTTGTTTAACTTTAAGAAGGAG

              NcoI
             -------
               MetGlyHisHisHisHisHisHisGlySerAspSerGluValAsnGlnGluAlaLysProGluValLysProGluValLysProGluThrHis
 101         ATATACCATGGGTCATCACCATCATCATCACGGGTCGGACTCAGAAGTCAATCAAGAAGCTAAGCCAGAGGTCAAGCCAGAAGTCAAGCCTGAGACTCAC

                                      BglII
                                     ------
        IleAsnLeuLysValSerAspGlySerSerGluIlePhePheLysIleLysLysThrThrProLeuArgArgLeuMetGluAlaPheAlaLysArgGlnGly·
 201    ATCAATTTAAAGGTGTCCGATGGATCTTCAGAGATCTTCTTCAAGATCAAAAAGACCACTCCTTTAAGAAGGCTGATGGAAGCGTTCGCTAAAAGACAGG

                                      EcoRI
                                     ------
        ·GLysGluMetAspSerLeuArgPheLeuTyrAspGlyIleArgIleGlnAlaAspGlnThrProGluAspLeuAspMetGluAspAsnAspIleIleGlu·
 301    GTAAGGAAATGGACTCCTTAAGATTCTTGTACGACGGTATTAGAATTCAAGCTGATCAGACCCCTGAAGATTTGGACATGGAGGATAACGATATTATTGA

                                             SacI   SalI          NotI
                                             ------ ------        --------
                                                                  EagI
                                                                  ------
                                  BsaI BamHI EcoRI                HindIII        XhoI
                                  ------------------              ------         ------
        ·AlaHisArgGluGlnIleGlyGly***
 401    GGCTCACCGCGAACAGATTGGAGGTTGAGACCGGATCCGAATTCGAGCTCCGTCGACAAGCTTGCGGCCGCACTCGAG
                                  ↑
                  Hydrolase Cleavage Site
```

Fig. 4A

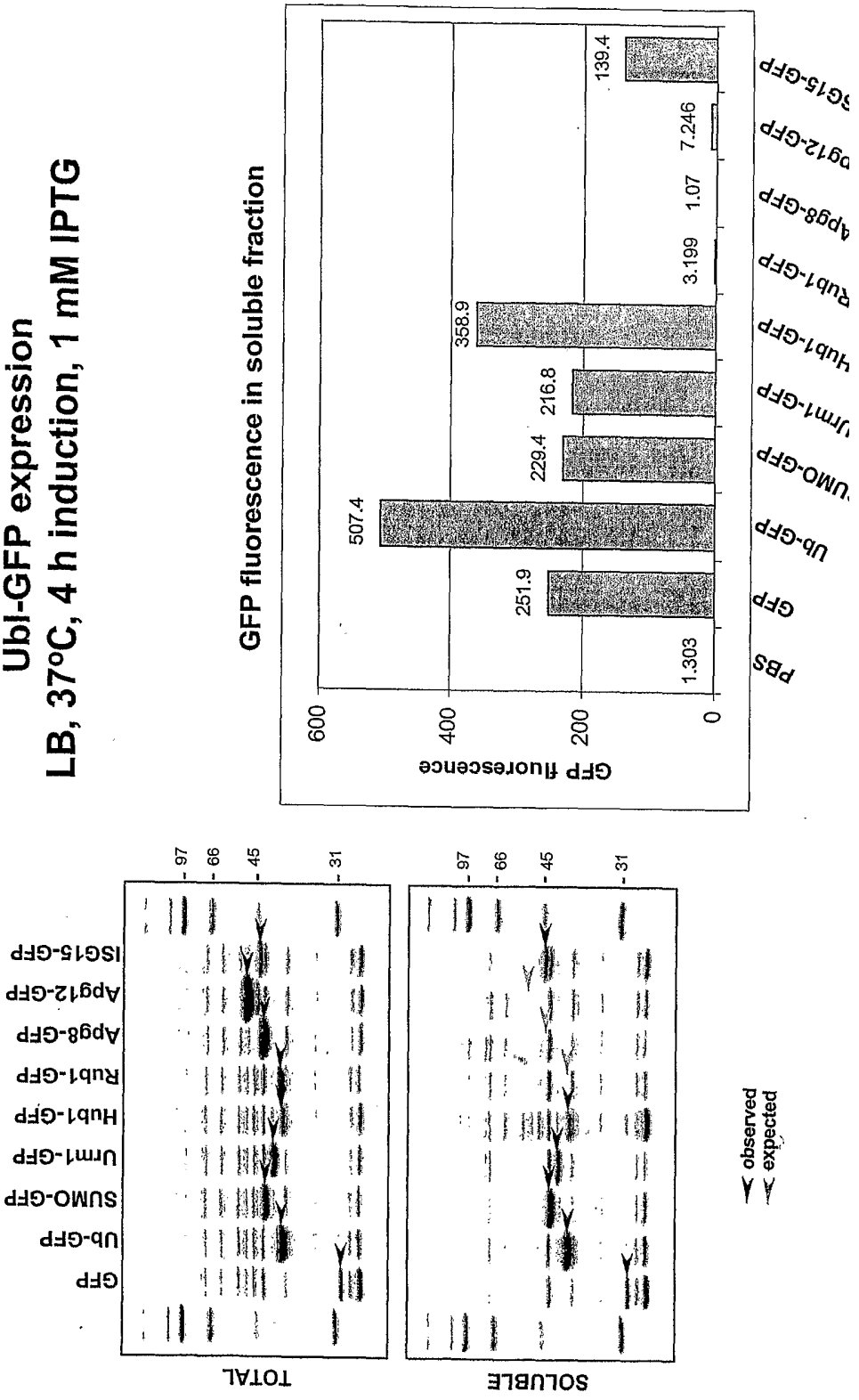
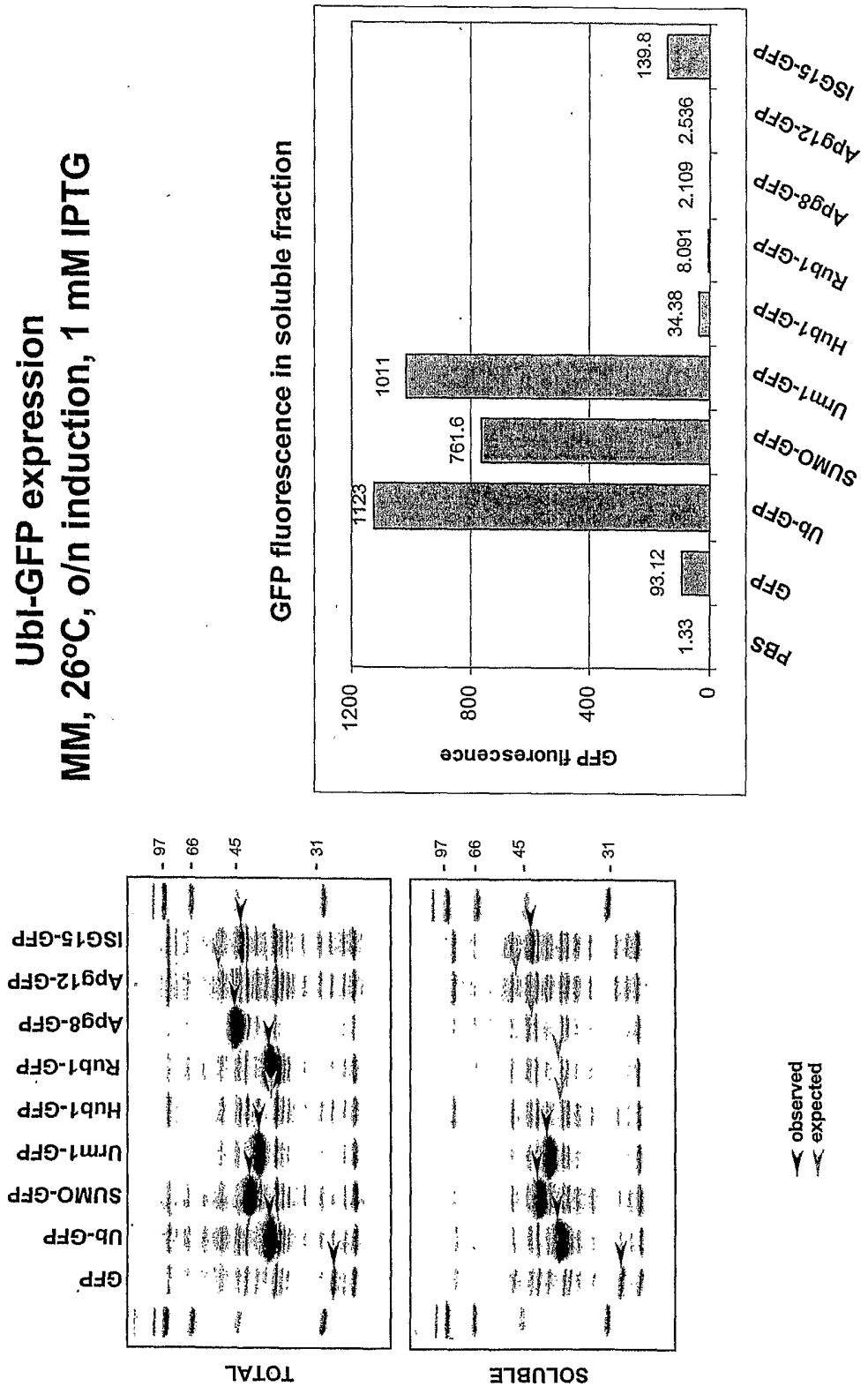

Ubl-GFP expression
LB, 37°C, 4 h induction, 1 mM IPTG
GFP fluorescence in soluble fraction
GFP fluorescence
600
400
200
0
PBS
GFP
Ub-GFP
SUMO-GFP
Urm1-GFP
Hub1-GFP
Rub1-GFP
Apg8-GFP
Apg12-GFP
ISG15-GFP
1.303
251.9
507.4
229.4
216.8
358.9
3.199
1.07
7.246
139.4
TOTAL
SOLUBLE
97
66
45
31
observed
expected

Fig. 4B


Ubl-GFP expression
MM, 26°C, o/n induction, 1 mM IPTG
GFP fluorescence in soluble fraction
GFP fluorescence
1200
800
400
0
PBS
GFP
Ub-GFP
SUMO-GFP
Urm1-GFP
Hub1-GFP
Rub1-GFP
Apg8-GFP
Apg12-GFP
ISG15-GFP
1.33
93.12
1123
761.6
1011
34.38
8.091
2.109
2.536
139.8
TOTAL
SOLUBLE
- 97
- 66
- 45
- 31
observed
expected

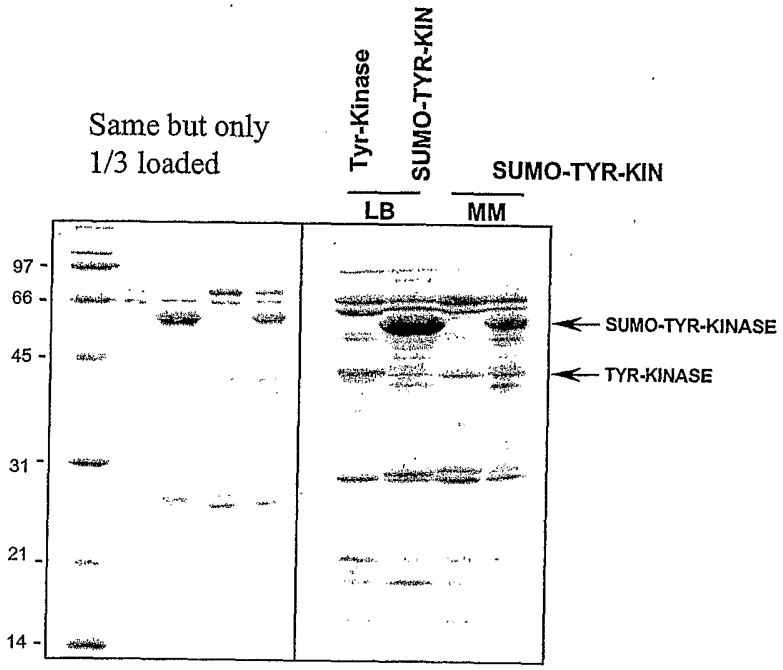

Same but only
1/3 loaded
Tyr-Kinase
SUMO-TYR-KIN
SUMO-TYR-KIN
LB
MM
97
66
45
31
21
14
SUMO-TYR-KINASE
TYR-KINASE

Figure 5

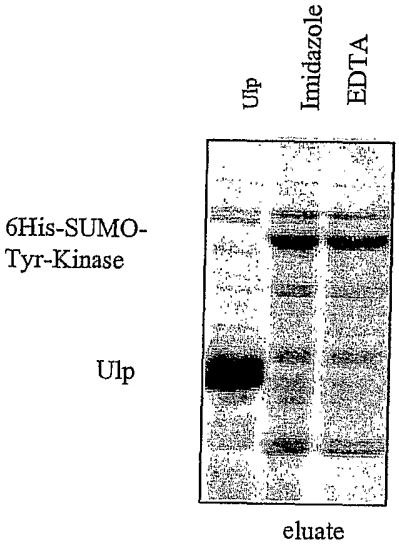


Figure 6

Figure 7

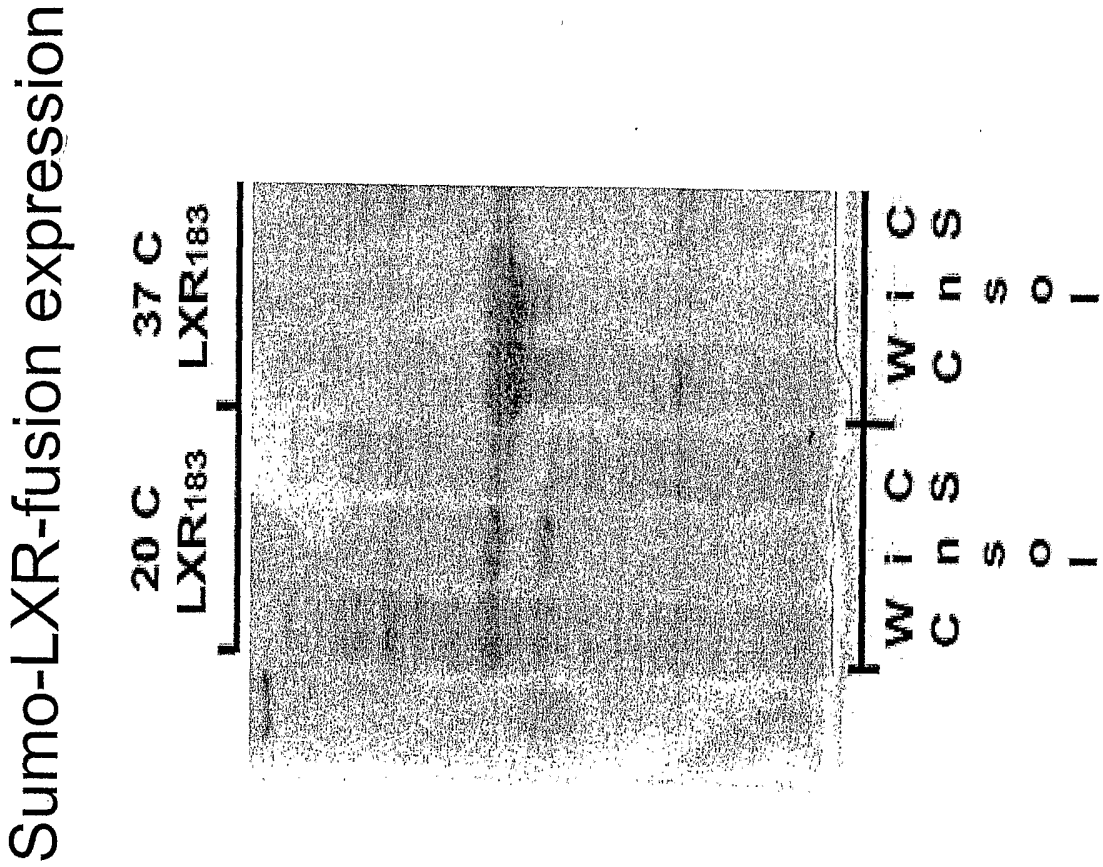

Sumo-LXR-fusion expression
20 C
LXR183
37 C
LXR183
Wi C
Cns Cns
so so
l l

Fig. 8A

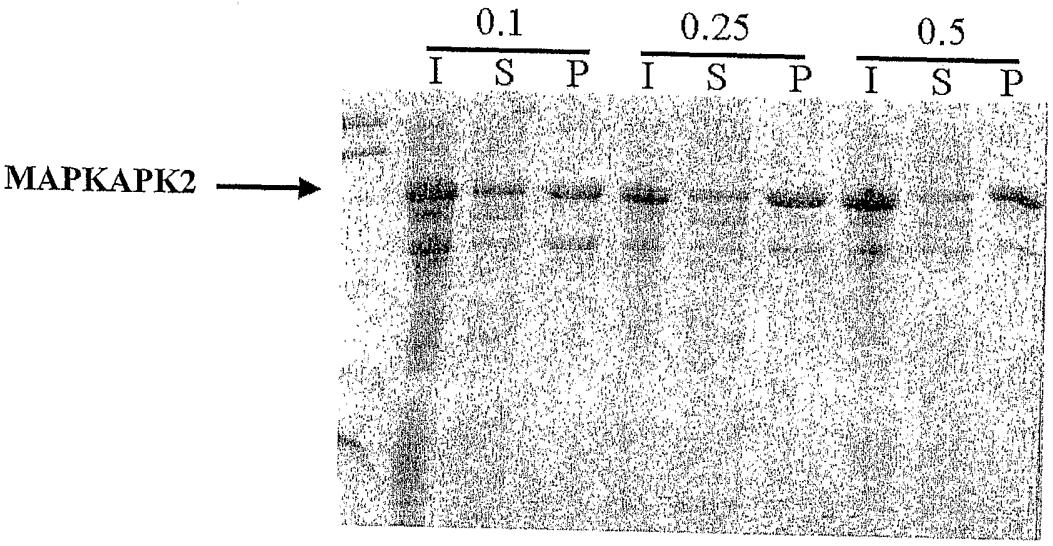


Fig. 8B

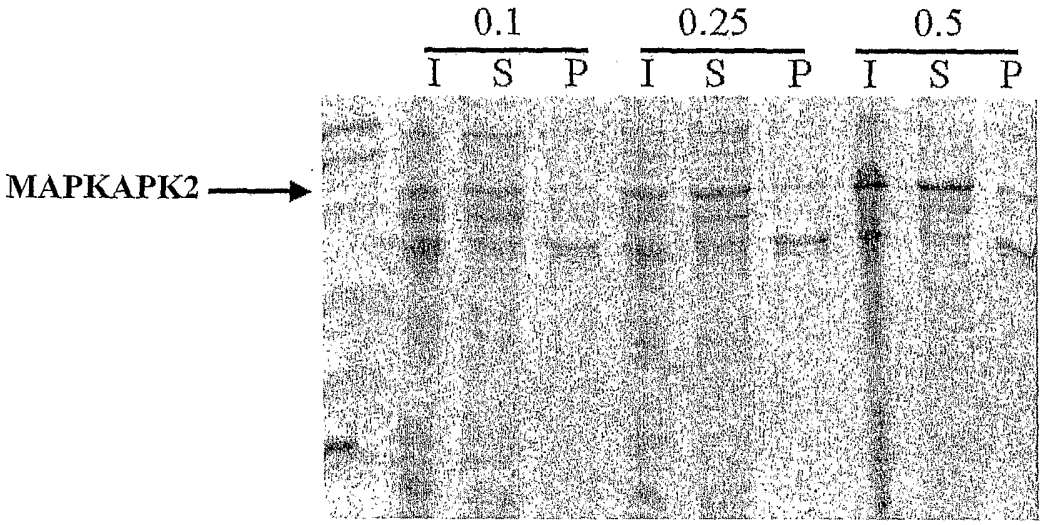


Figure 9

Ubl-GFP co-translational cleavage
YPD, 30°C, 3.5 h induction, 100 μM $CuSO_4$

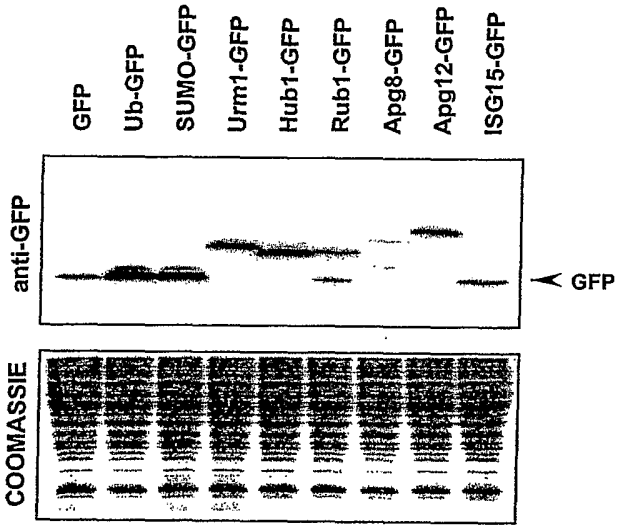


Figure 10

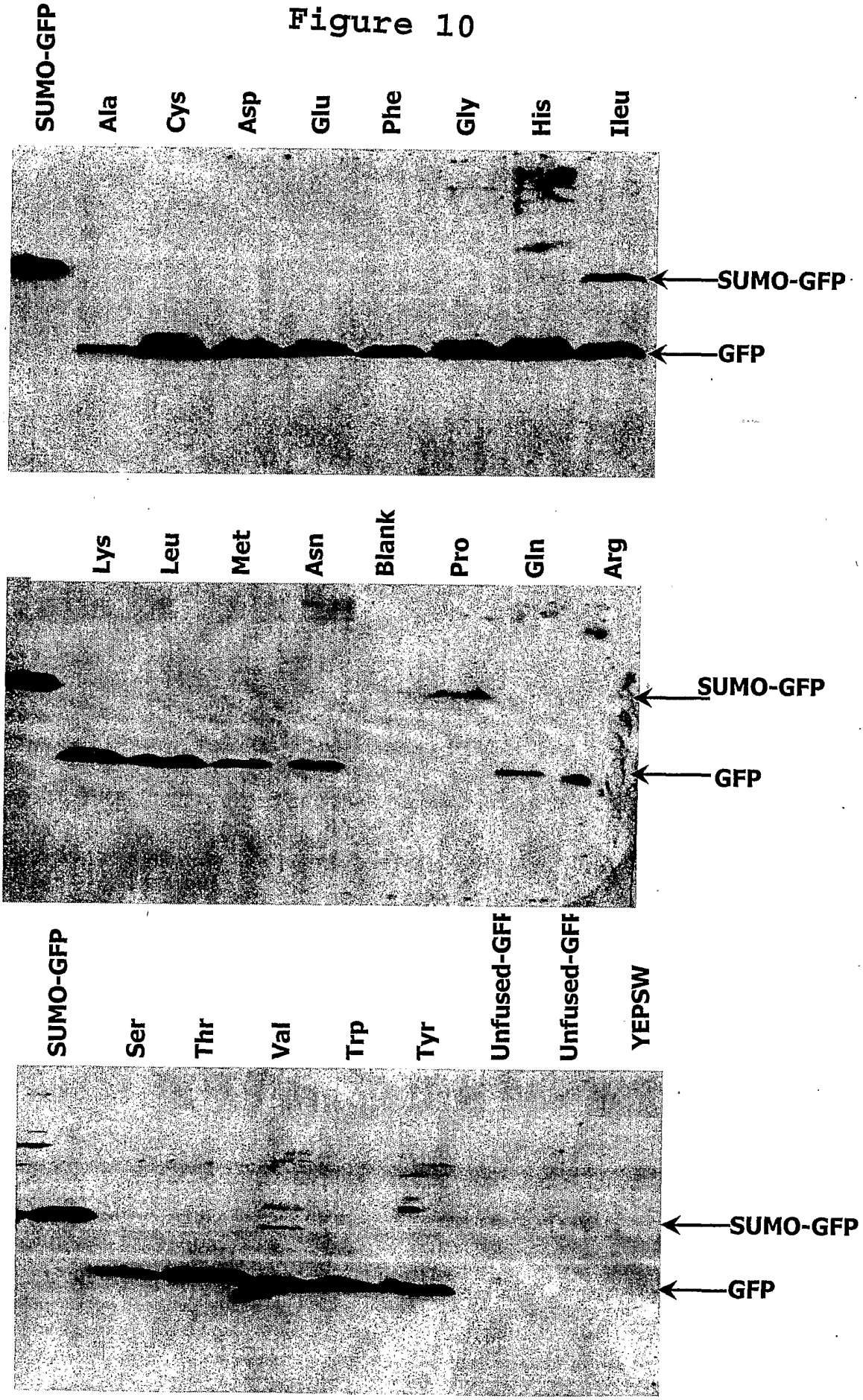
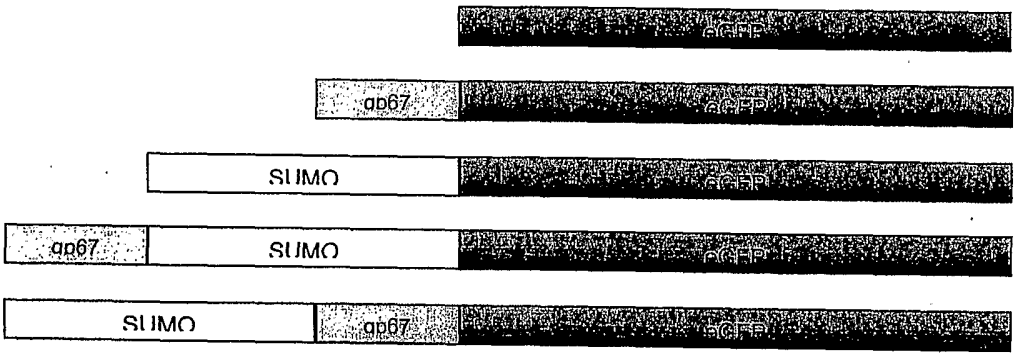
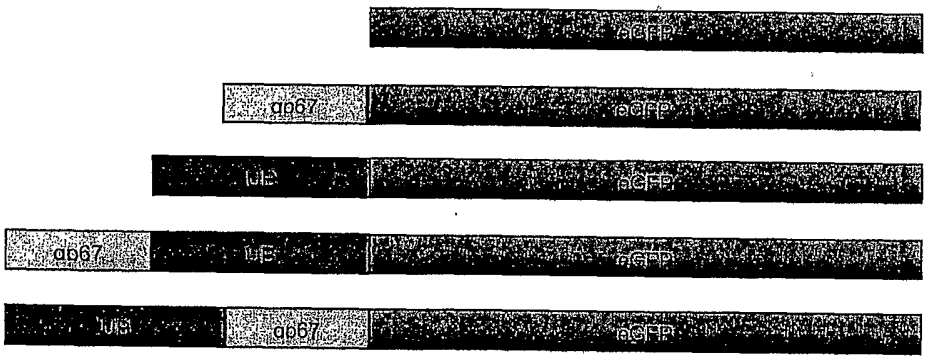

SUMO-GFP
Ala
Cys
Asp
Glu
Phe
Gly
His
Ileu
SUMO-GFP
GFP
Lys
Leu
Met
Asn
Blank
Pro
Gln
Arg
SUMO-GFP
GFP
SUMO-GFP
Ser
Thr
Val
Trp
Tyr
Unfused-GFP
Unfused-GFP
YEPSW
SUMO-GFP
GFP

Fig. 11A



E-eGFP
G-eGFP
S-eGFP
GS-eGFP
SG-eGFP

Fig. 11B



E-eGFP
G-eGFP
U-eGFP
GU-eGFP
UG-eGFP

Fig. 12A

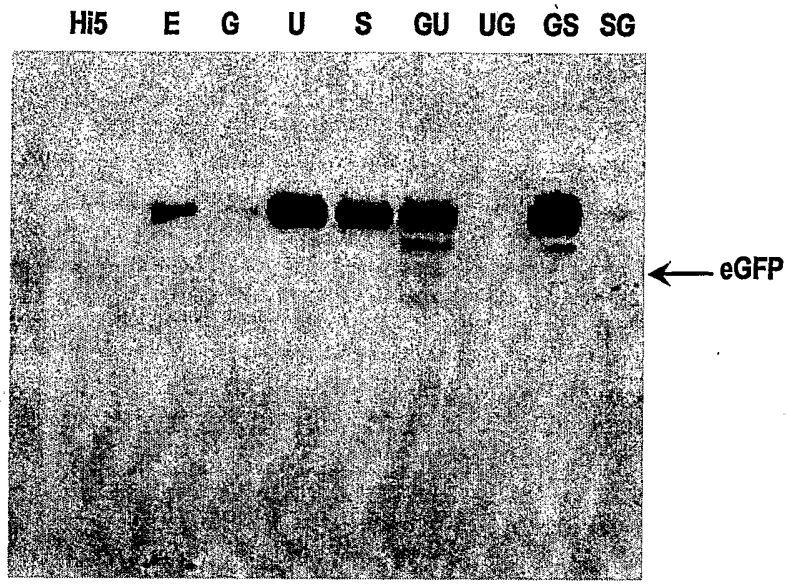


Fig. 12B

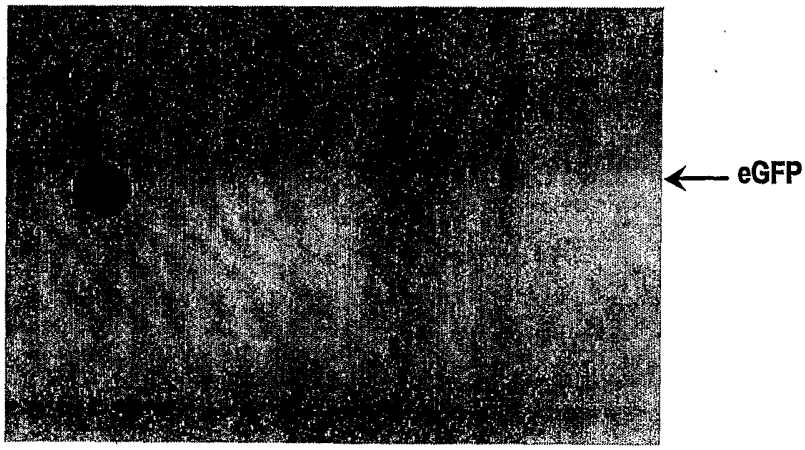


Fig. 13A

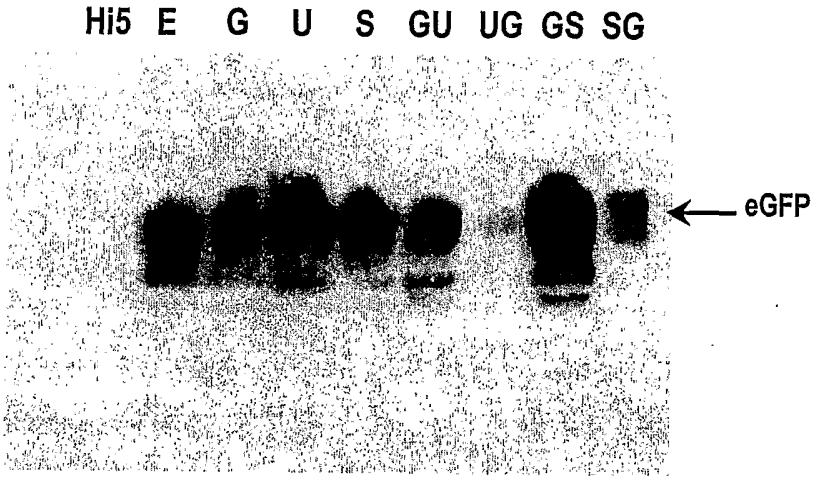


Fig. 13B

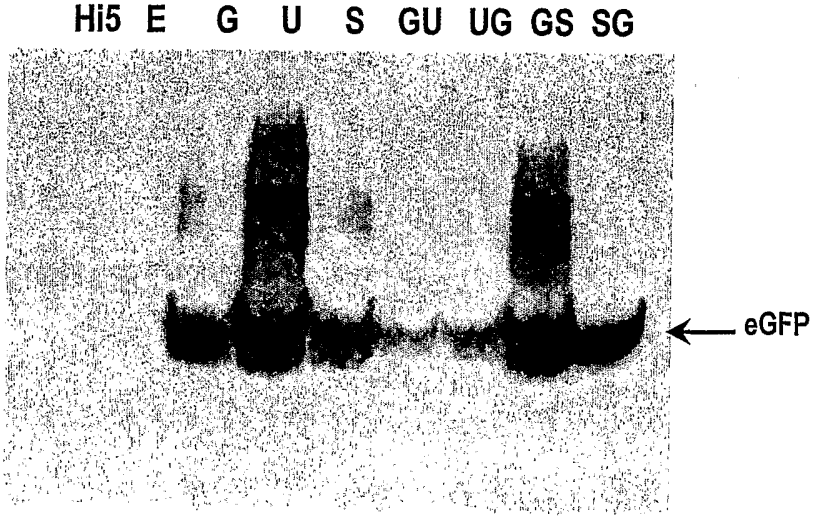


Fig. 13C

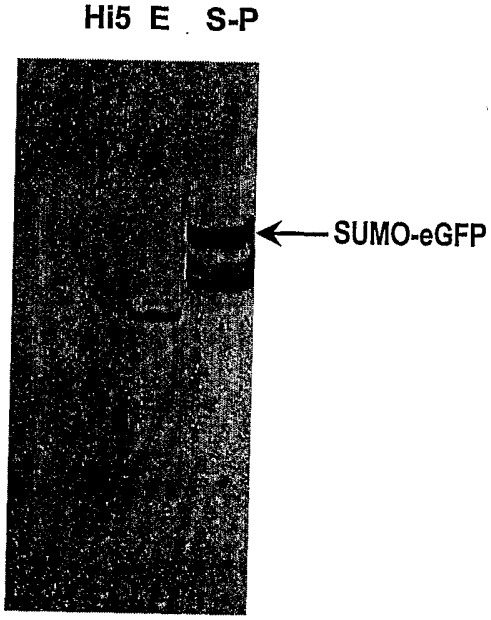
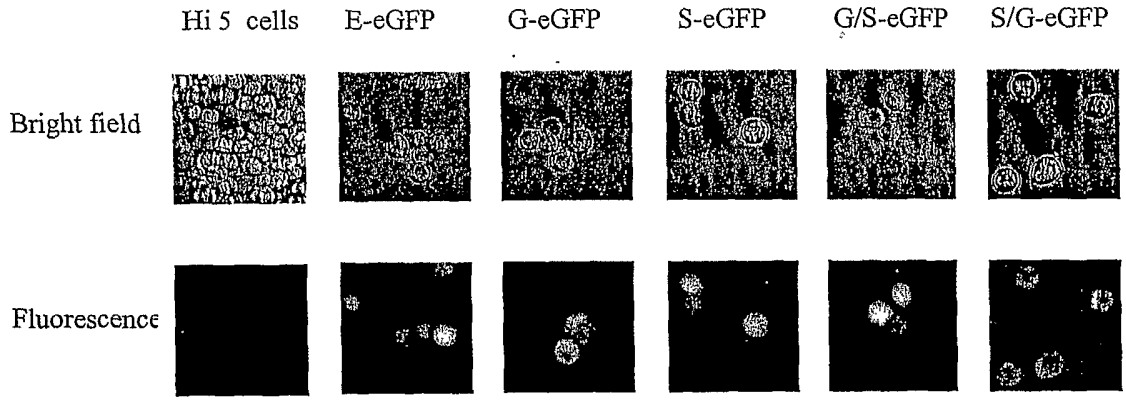


Figure 14

SUMO-GFP fusion proteins expression in Hi-Five cells fluorescence micrographs



Ubiquitin-GFP fusion proteins expression in Hi-Five cells fluorescence micrographs

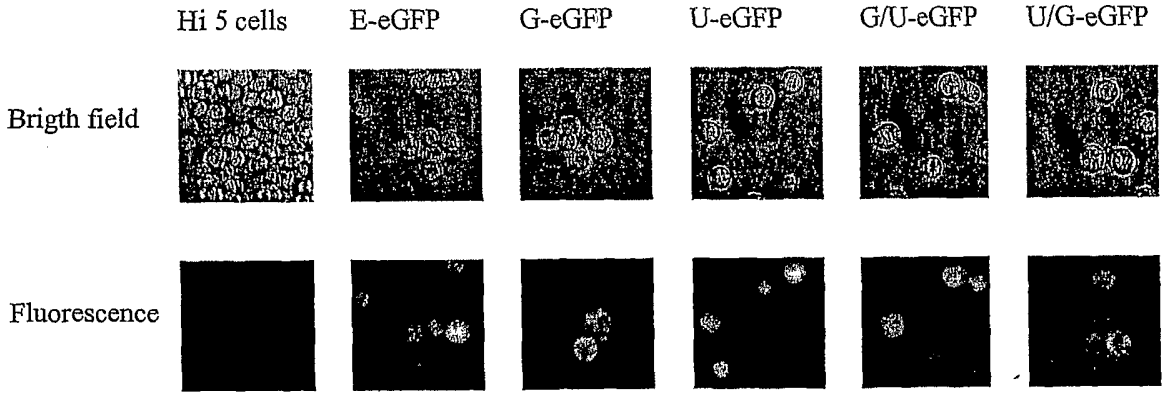


Figure 15

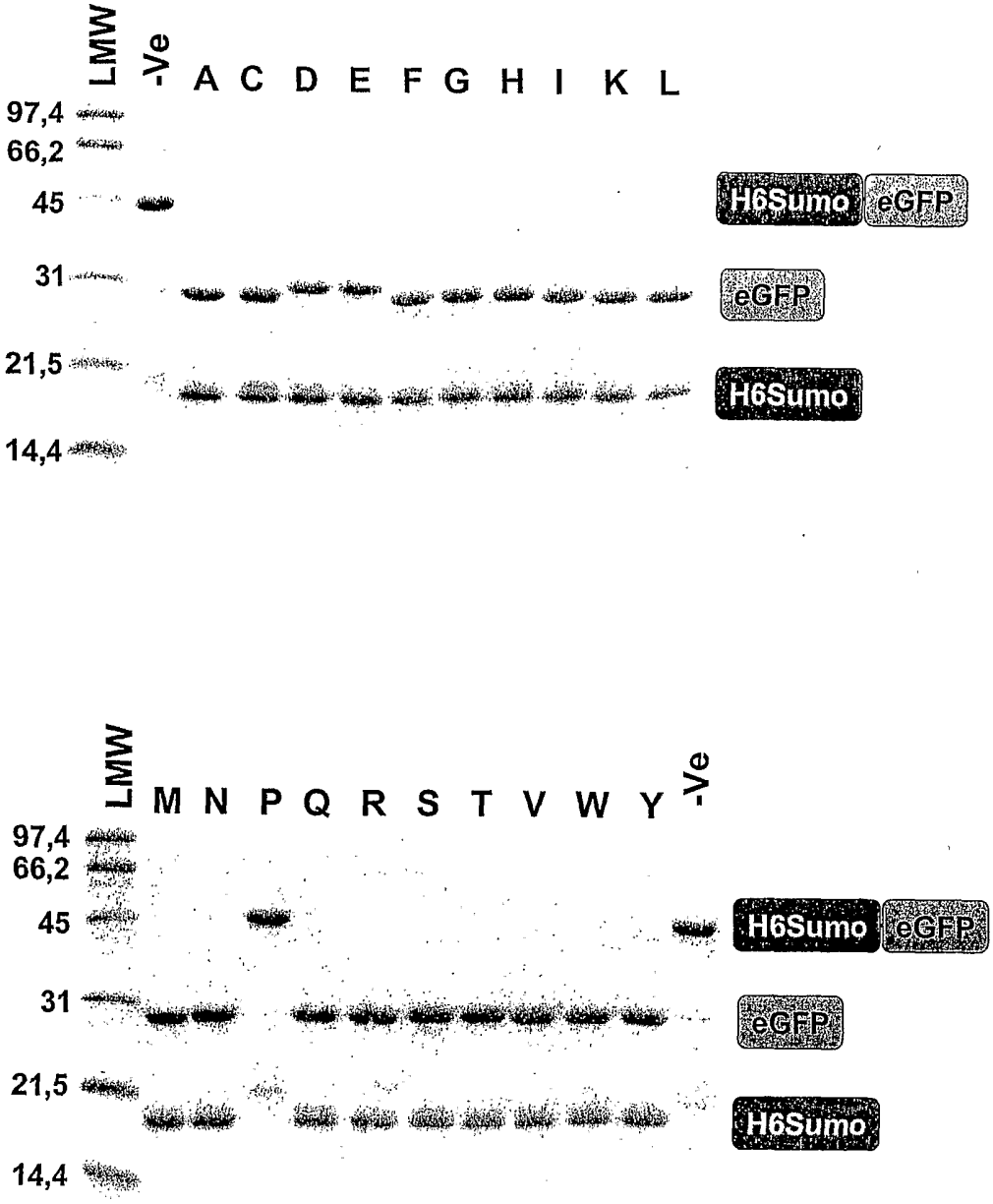

LMW
-Ve
A C D E F G H I K L
97,4
66,2
45
31
21,5
14,4
H6Sumo eGFP
eGFP
H6Sumo
LMW
M N P Q R S T V W Y -Ve
97,4
66,2
45
31
21,5
14,4
H6Sumo eGFP
eGFP
H6Sumo

Figure 16

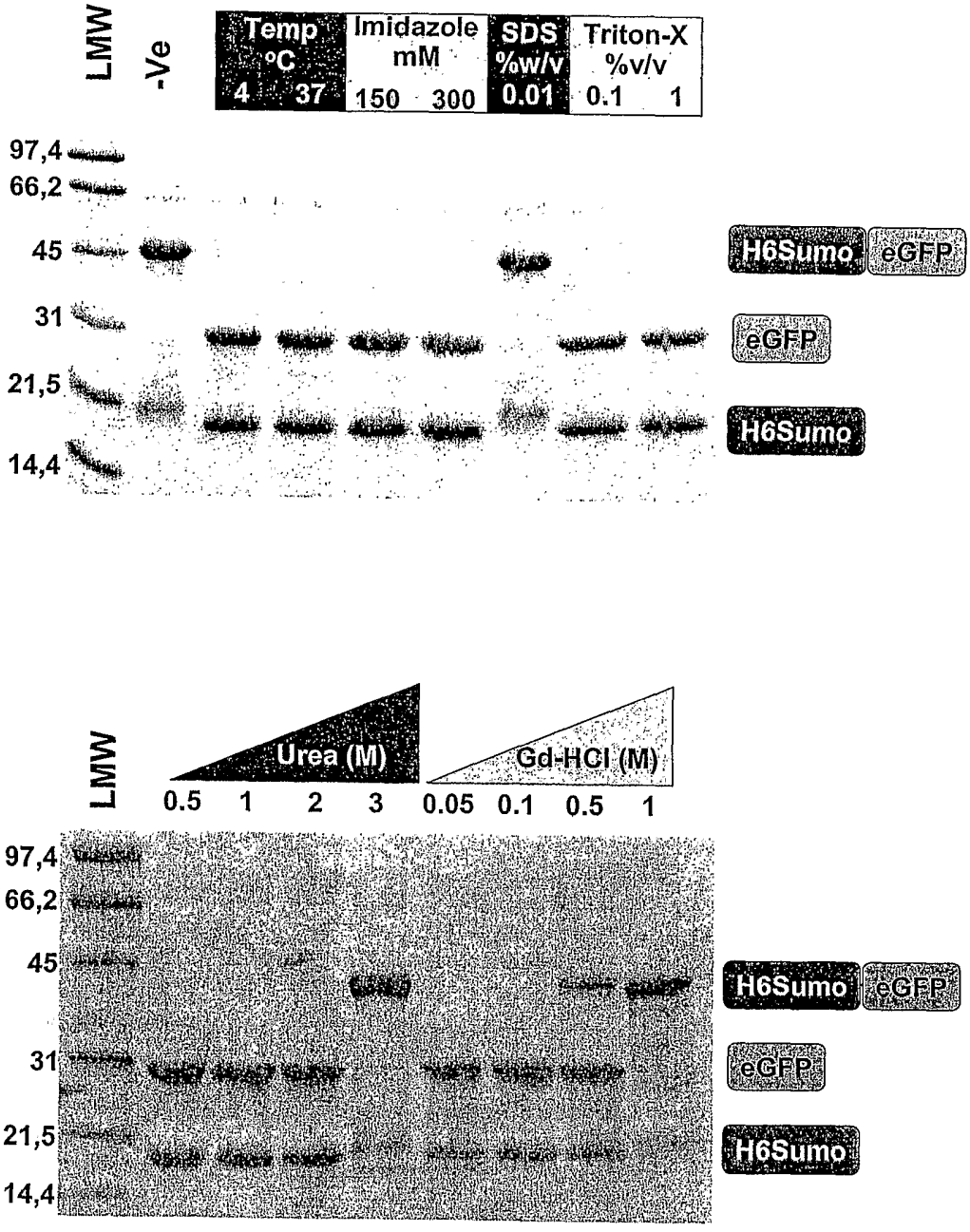

LMW
-Ve
Temp °C
4
37
Imidazole mM
150
300
SDS %w/v
0.01
Triton-X %v/v
0.1
1
97,4
66,2
45
31
21,5
14,4
H6Sumo
eGFP
eGFP
H6Sumo
LMW
Urea (M)
Gd-HCl (M)
0.5
1
2
3
0.05
0.1
0.5
1
97,4
66,2
45
31
21,5
14,4
H6Sumo
eGFP
eGFP
H6Sumo

Figure 17

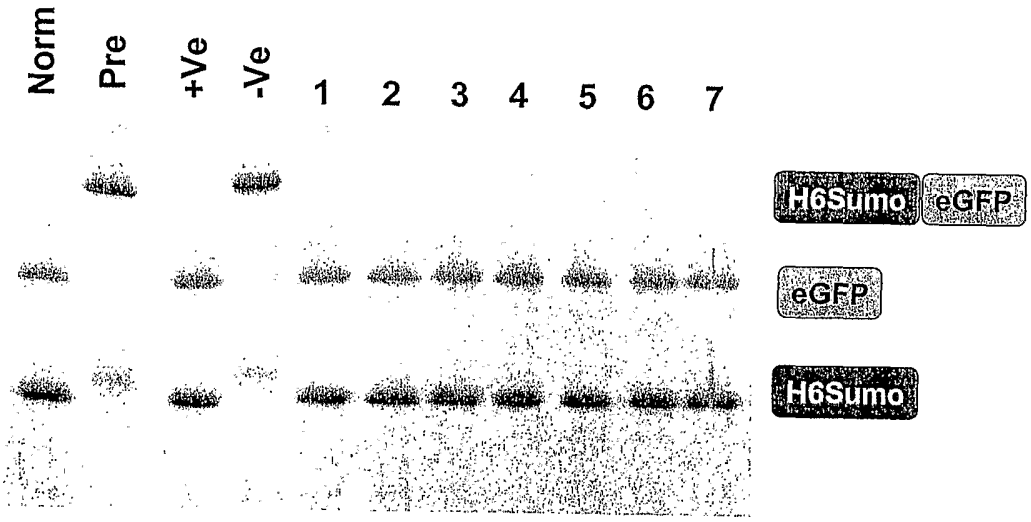


Figure 18

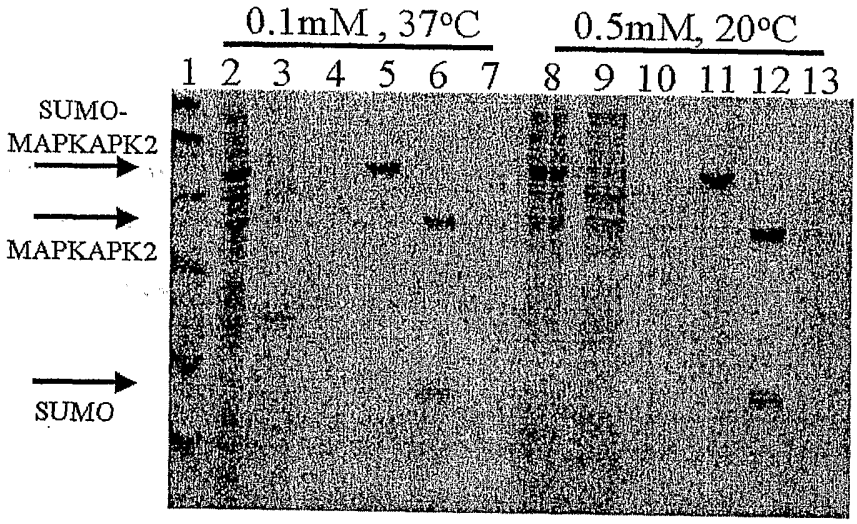

0.1mM, 37°C
0.5mM, 20°C
1 2 3 4 5 6 7 8 9 10 11 12 13
SUMO-MAPKAPK2
MAPKAPK2
SUMO

Figure 19

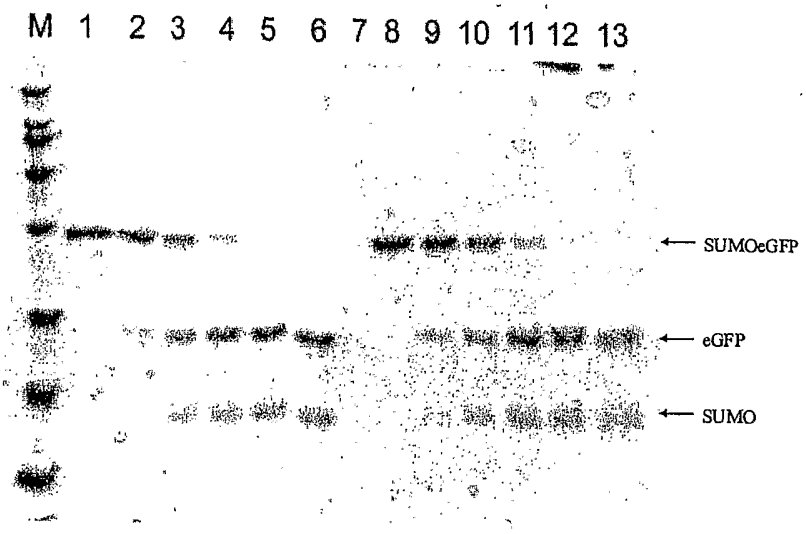


Figure 20

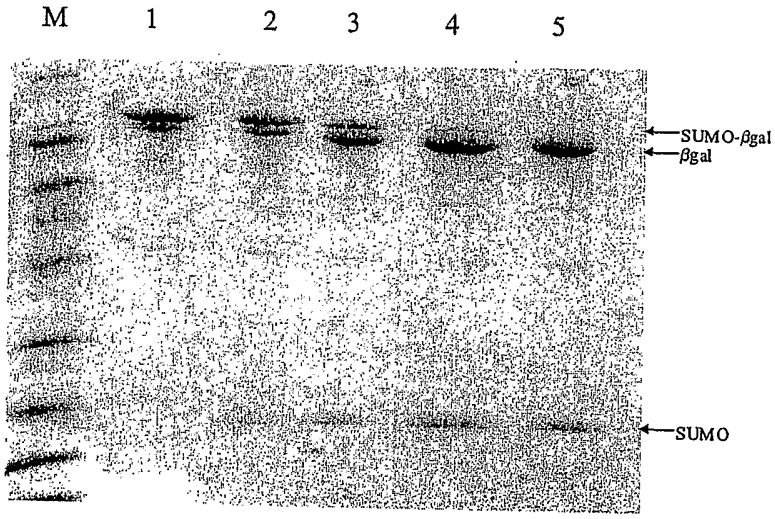


Figure 21

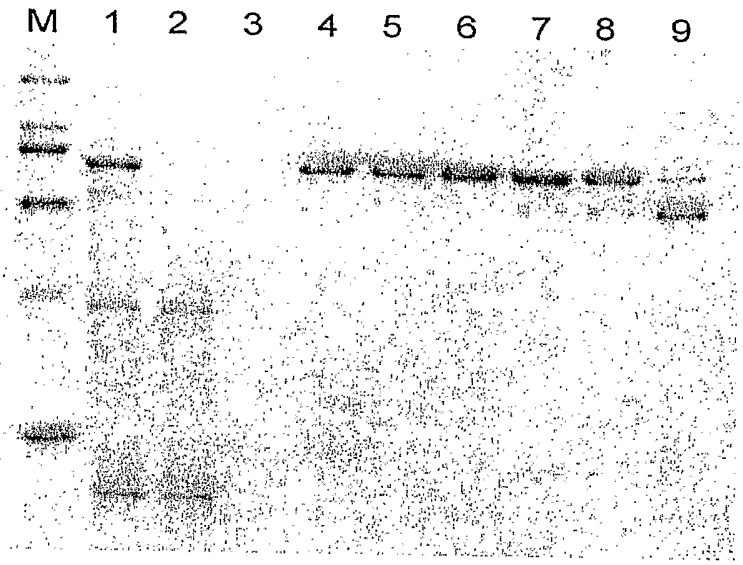
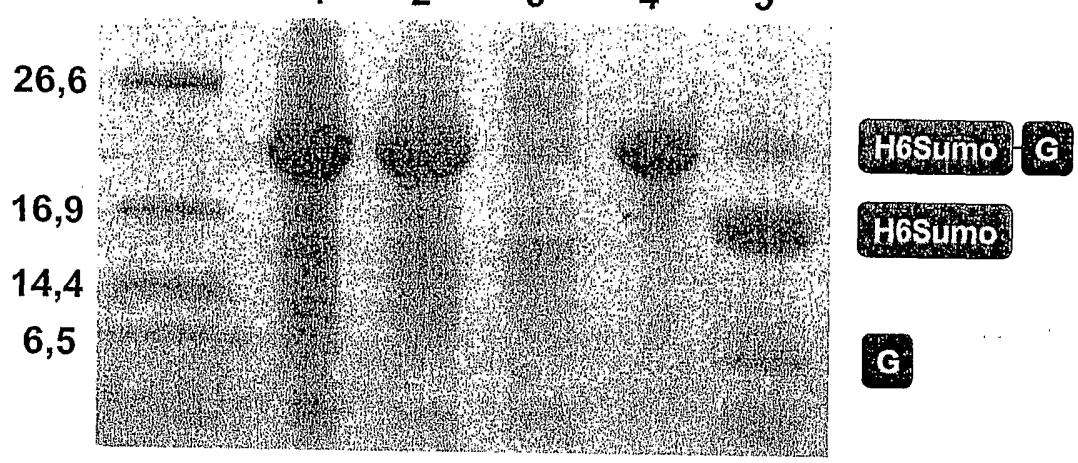


Figure 22



SUMO

SUMO NCBI ACCESSION# Q12306

```
      NcoI
      ~~~~~~
         M  G  H   H  H  H   H  H  G  S   D  S  E   V  N  Q
   1  CCATGGGTCA TCACCATCAT CATCACGGGT CGGACTCAGA AGTCAATCAA
      GGTACCCAGT AGTGGTAGTA GTAGTGCCCA GCCTGAGTCT TCAGTTAGTT
       E  A  K  P   E  V  K   P  E  V   K  P  E  T   H  I  N ·
  51  GAAGCTAAGC CAGAGGTCAA GCCAGAAGTC AAGCCTGAGA CTCACATCAA
      CTTCGATTCG GTCTCCAGTT CGGTCTTCAG TTCGGACTCT GAGTGTAGTT
      · L  K  V   S  D  G  S   S  E  I   F  F  K   I  K  K  T ·
 101  TTTAAAGGTG TCCGATGGAT CTTCAGAGAT CTTCTTCAAG ATCAAAAAGA
      AAATTTCCAC AGGCTACCTA GAAGTCTCTA GAAGAAGTTC TAGTTTTTCT
      · T  P  L   R  R  L   M  E  A  F   A  K  R   Q  G  K
 151  CCACTCCTTT AAGAAGGCTG ATGGAAGCGT TCGCTAAAAG ACAGGGTAAG
      GGTGAGGAAA TTCTTCCGAC TACCTTCGCA AGCGATTTTC TGTCCCATTC
       E  M  D  S   L  R  F   L  Y  D   G  I  R  I   Q  A  D ·
 201  GAAATGGACT CCTTAAGATT CTTGTACGAC GGTATTAGAA TTCAAGCTGA
      CTTTACCTGA GGAATTCTAA GAACATGCTG CCATAATCTT AAGTTCGACT
      · Q  A  P   E  D  L  D   M  E  D   N  D  I   I  E  A  H ·
 251  TCAGGCCCCT GAAGATTTGG ACATGGAGGA TAACGATATT ATTGAGGCTC
      AGTCCGGGGA CTTCTAAACC TGTACCTCCT ATTGCTATAA TAACTCCGAG
      · R  E  Q  · I  G  G
 301  ACCGCGAACA · GATTGGAGGT
      TGGCGCTTGT CTAACCTCCA
```

Figure 23

Figure 24A

GFP

GFP NCBI ACCESSION# P42212

```
      M  V  S  K  G  E  E  L  F  T
1     ATGGTGAGCA AGGGCGAGGA GCTGTTCACC
      TACCACTCGT TCCCGCTCCT CGACAAGTGG
       G  V  V  P  I  L  V  E  L  D  G  D  V  N  G  H  K ·
31    GGGGTGGTGC CCATCCTGGT CGAGCTGGAC GGCGACGTAA ACGGCCACAA
      CCCCACCACG GGTAGGACCA GCTCGACCTG CCGCTGCATT TGCCGGTGTT
      · F  S  V  S  G  E  G  E  G  D  A  T  Y  G  K  L  T ·
81    GTTCAGCGTG TCCGGCGAGG GCGAGGGCGA TGCCACCTAC GGCAAGCTGA
      CAAGTCGCAC AGGCCGCTCC CGCTCCCGCT ACGGTGGATG CCGTTCGACT
      · L  K  F  I  C  T  T  G  K  L  P  V  P  W  P  T
131   CCCTGAAGTT CATCTGCACC ACCGGCAAGC TGCCCGTGCC CTGGCCCACC
      GGGACTTCAA GTAGACGTGG TGGCCGTTCG ACGGGCACGG GACCGGGTGG
       L  V  T  T  L  T  Y  G  V  Q  C  F  S  R  Y  P  D ·
181   CTCGTGACCA CCCTGACCTA CGGCGTGCAG TGCTTCAGCC GCTACCCCGA
      GAGCACTGGT GGGACTGGAT GCCGCACGTC ACGAAGTCGG CGATGGGGCT
      · H  M  K  Q  H  D  F  F  K  S  A  M  P  E  G  Y  V ·
201   CCACATGAAG CAGCACGACT TCTTCAAGTC CGCCATGCCC GAAGGCTACG
      GGTGTACTTC GTCGTGCTGA AGAAGTTCAG GCGGTACGGG CTTCCGATGC
      · Q  E  R  T  I  F  F  K  D  D  G  N  Y  K  T  R
231   TCCAGGAGCG CACCATCTTC TTCAAGGACG ACGGCAACTA CAAGACCCGC
      AGGTCCTCGC GTGGTAGAAG AAGTTCCTGC TGCCGTTGAT GTTCTGGGCG


      A  E  V  K  F  E  G  D  T  L  V  N  R  I  E  L  K ·
281   GCCGAGGTGA AGTTCGAGGG CGACACCCTG GTGAACCGCA TCGAGCTGAA
      CGGCTCCACT TCAAGCTCCC GCTGTGGGAC CACTTGGCGT AGCTCGACTT
      · G  I  D  F  K  E  D  G  N  I  L  G  H  K  L  E  Y ·
331   GGGCATCGAC TTCAAGGAGG ACGGCAACAT CCTGGGGCAC AAGCTGGAGT
      CCCGTAGCTG AAGTTCCTCC TGCCGTTGTA GGACCCCGTG TTCGACCTCA
      · N  Y  N  S  H  N  V  Y  I  M  A  D  K  Q  K  N
381   ACAACTACAA CAGCCACAAC GTCTATATCA TGGCCGACAA GCAGAAGAAC
      TGTTGATGTT GTCGGTGTTG CAGATATAGT ACCGGCTGTT CGTCTTCTTG
       G  I  K  V  N  F  K  I  R  H  N  I  E  D  G  S  V ·
431   GGCATCAAGG TGAACTTCAA GATCCGCCAC AACATCGAGG ACGGCAGCGT
      CCGTAGTTCC ACTTGAAGTT CTAGGCGGTG TTGTAGCTCC TGCCGTCGCA
      · Q  L  A  D  H  Y  Q  Q  N  T  P  I  G  D  G  P  V ·
481   GCAGCTCGCC GACCACTACC AGCAGAACAC CCCCATCGGC GACGGCCCCG
      CGTCGAGCGG CTGGTGATGG TCGTCTTGTG GGGGTAGCCG CTGCCGGGGC
      · L  L  P  D  N  H  Y  L  S  T  Q  S  A  L  S  K
531   TGCTGCTGCC CGACAACCAC TACCTGAGCA CCCAGTCCGC CCTGAGCAAA
      ACGACGACGG GCTGTTGGTG ATGGACTCGT GGGTCAGGCG GGACTCGTTT
```

Figure 24B

```
      D  P  N  E   K  R  D   H  M  V   L  L  E  F   V  T  A
581   GACCCCAACG AGAAGCGCGA TCACATGGTC CTGCTGGAGT TCGTGACCGC
      CTGGGGTTGC TCTTCGCGCT AGTGTACCAG GACGACCTCA AGCACTGGCG
                                                  HindIII
                                                  ~~~~~~
      A  G  I   T  L  G  M   D  E  L   Y  K  *   *
631   CGCCGGGATC ACTCTCGGCA TGGACGAGCT GTACAAGTAA TAAGCTT
      GCGGCCCTAG TGAGAGCCGT ACCTGCTCGA CATGTTCATT ATTCGAA
```

Figure 25A

SUMO-GFP
SUMO NCBI ACCESSION# Q12306

NcoI
~~~~~~

```
      M  G  H   H  H  H   H  H  G  S   D  S  E   V  N  Q
  1 CCATGGGTCA TCACCATCAT CATCACGGGT CGGACTCAGA AGTCAATCAA
    GGTACCCAGT AGTGGTAGTA GTAGTGCCCA GCCTGAGTCT TCAGTTAGTT
     E  A  K  P   E  V  K   P  E  V   K  P  E  T   H  I  N ·
 51 GAAGCTAAGC CAGAGGTCAA GCCAGAAGTC AAGCCTGAGA CTCACATCAA
    CTTCGATTCG GTCTCCAGTT CGGTCTTCAG TTCGGACTCT GAGTGTAGTT
    · L  K  V   S  D  G  S   S  E  I   F  F  K   I  K  K  T ·
101 TTTAAAGGTG TCCGATGGAT CTTCAGAGAT CTTCTTCAAG ATCAAAAAGA
    AAATTTCCAC AGGCTACCTA GAAGTCTCTA GAAGAAGTTC TAGTTTTTCT
    · T  P  L   R  R  L   M  E  A  F   A  K  R   Q  G  K
151 CCACTCCTTT AAGAAGGCTG ATGGAAGCGT TCGCTAAAAG ACAGGGTAAG
    GGTGAGGAAA TTCTTCCGAC TACCTTCGCA AGCGATTTTC TGTCCCATTC
     E  M  D  S   L  R  F   L  Y  D   G  I  R  I   Q  A  D ·
201 GAAATGGACT CCTTAAGATT CTTGTACGAC GGTATTAGAA TTCAAGCTGA
    CTTTACCTGA GGAATTCTAA GAACATGCTG CCATAATCTT AAGTTCGACT
    · Q  A  P   E  D  L  D   M  E  D   N  D  I   I  E  A  H ·
251 TCAGGCCCCT GAAGATTTGG ACATGGAGGA TAACGATATT ATTGAGGCTC
    AGTCCGGGGA CTTCTAAACC TGTACCTCCT ATTGCTATAA TAACTCCGAG
    · R  E  Q   I  G  G   M  V  S  K   G  E  E   L  F  T
301 ACCGCGAACA GATTGGAGGT ATGGTGAGCA AGGGCGAGGA GCTGTTCACC
    TGGCGCTTGT CTAACCTCCA TACCACTCGT TCCCGCTCCT CGACAAGTGG
     G  V  V  P   I  L  V   E  L  D   G  D  V  N   G  H  K ·
351 GGGGTGGTGC CCATCCTGGT CGAGCTGGAC GGCGACGTAA ACGGCCACAA
    CCCCACCACG GGTAGGACCA GCTCGACCTG CCGCTGCATT TGCCGGTGTT
    · F  S  V   S  G  E  G   E  G  D   A  T  Y   G  K  L  T ·
401 GTTCAGCGTG TCCGGCGAGG GCGAGGGCGA TGCCACCTAC GGCAAGCTGA
    CAAGTCGCAC AGGCCGCTCC CGCTCCCGCT ACGGTGGATG CCGTTCGACT
    · L  K  F   I  C  T   T  G  K  L   P  V  P   W  P  T
451 CCCTGAAGTT CATCTGCACC ACCGGCAAGC TGCCCGTGCC CTGGCCCACC
    GGGACTTCAA GTAGACGTGG TGGCCGTTCG ACGGGCACGG GACCGGGTGG
     L  V  T  T   L  T  Y   G  V  Q   C  F  S  R   Y  P  D ·
501 CTCGTGACCA CCCTGACCTA CGGCGTGCAG TGCTTCAGCC GCTACCCCGA
    GAGCACTGGT GGGACTGGAT GCCGCACGTC ACGAAGTCGG CGATGGGGCT
    · H  M  K   Q  H  D  F   F  K  S   A  M  P   E  G  Y  V ·
551 CCACATGAAG CAGCACGACT TCTTCAAGTC CGCCATGCCC GAAGGCTACG
    GGTGTACTTC GTCGTGCTGA AGAAGTTCAG GCGGTACGGG CTTCCGATGC
    · Q  E  R   T  I  F   F  K  D  D   G  N  Y   K  T  R
601 TCCAGGAGCG CACCATCTTC TTCAAGGACG ACGGCAACTA CAAGACCCGC
    AGGTCCTCGC GTGGTAGAAG AAGTTCCTGC TGCCGTTGAT GTTCTGGGCG
     A  E  V  K   F  E  G   D  T  L   V  N  R  I   E  L  K ·
651 GCCGAGGTGA AGTTCGAGGG CGACACCCTG GTGAACCGCA TCGAGCTGAA
    CGGCTCCACT TCAAGCTCCC GCTGTGGGAC CACTTGGCGT AGCTCGACTT
    · G  I  D   F  K  E  D   G  N  I   L  G  H  K   L  E  Y ·
701 GGGCATCGAC TTCAAGGAGG ACGGCAACAT CCTGGGGCAC AAGCTGGAGT
```
~~~~~~

```
      CCCGTAGCTG AAGTTCCTCC TGCCGTTGTA GGACCCCGTG TTCGACCTCA
      · N Y N  S H N  V Y I M  A D K  Q K N .
 751  ACAACTACAA CAGCCACAAC GTCTATATCA TGGCCGACAA GCAGAAGAAC
      TGTTGATGTT GTCGGTGTTG CAGATATAGT ACCGGCTGTT CGTCTTCTTG
       G I K V  N F K  I R H  N I E D  G S V ·
 801  GGCATCAAGG TGAACTTCAA GATCCGCCAC AACATCGAGG ACGGCAGCGT
      CCGTAGTTCC ACTTGAAGTT CTAGGCGGTG TTGTAGCTCC TGCCGTCGCA
      · Q L A  D H Y Q  Q N T  P I G  D G P V ·
 851  GCAGCTCGCC GACCACTACC AGCAGAACAC CCCCATCGGC GACGGCCCCG
      CGTCGAGCGG CTGGTGATGG TCGTCTTGTG GGGGTAGCCG CTGCCGGGGC
      · L L P  D N H  Y L S T  Q S A  L S K
 901  TGCTGCTGCC CGACAACCAC TACCTGAGCA CCCAGTCCGC CCTGAGCAAA
      ACGACGACGG GCTGTTGGTG ATGGACTCGT GGGTCAGGCG GGACTCGTTT
       D P N E  K R D  H M V  L L E F  V T A ·
 951  GACCCCAACG AGAAGCGCGA TCACATGGTC CTGCTGGAGT TCGTGACCGC
      CTGGGGTTGC TCTTCGCGCT AGTGTACCAG GACGACCTCA AGCACTGGCG
                                                 HindIII
                                                 ~~~~~~
      · A G I  T L G M  D E L  Y K *  *
1001  CGCCGGGATC ACTCTCGGCA TGGACGAGCT GTACAAGTAA TAAGCTT
      GCGGCCCTAG TGAGAGCCGT ACCTGCTCGA CATGTTCATT ATTCGAA
```

Figure 25B

Figure 26A

Ub-GFP

Ub NCBI ACCESSION# 751846A

```
      NcoI
      ~~~~~~
         M  G  H  H  H  H  H  H  G  Q  I  F  V  K  T  L
    1 CCATGGGTCA TCACCATCAT CATCACGGGC AGATCTTCGT CAAGACGTTA
      GGTACCCAGT AGTGGTAGTA GTAGTGCCCG TCTAGAAGCA GTTCTGCAAT
         T  G  K  T  I  T  L  E  V  E  P  S  D  T  I  E  N
   51 ACCGGTAAAA CCATAACTCT AGAAGTTGAA CCATCCGATA CCATCGAAAA
      TGGCCATTTT GGTATTGAGA TCTTCAACTT GGTAGGCTAT GGTAGCTTTT
         V  K  A  K  I  Q  D  K  E  G  I  P  P  D  Q  Q  R
  101 CGTTAAGGCT AAAATTCAAG ACAAGGAAGG CATTCCACCT GATCAACAAA
      GCAATTCCGA TTTTAAGTTC TGTTCCTTCC GTAAGGTGGA CTAGTTGTTT
         L  I  F  A  G  K  Q  L  E  D  G  R  T  L  S  D
  151 GATTGATCTT TGCCGGTAAG CAGCTCGAGG ACGGTAGAAC GCTGTCTGAT
      CTAACTAGAA ACGGCCATTC GTCGAGCTCC TGCCATCTTG CGACAGACTA
         Y  N  I  Q  K  E  S  T  L  H  L  V  L  R  L  R  G
  201 TACAACATTC AGAAGGAGTC GACCTTACAT CTTGTCTTAC GCCTACGTGG
      ATGTTGTAAG TCTTCCTCAG CTGGAATGTA GAACAGAATG CGGATGCACC
         G  M  V  S  K  G  E  E  L  F  T  G  V  V  P  I  L
  251 AGGTATGGTG AGCAAGGGCG AGGAGCTGTT CACCGGGGTG GTGCCCATCC
      TCCATACCAC TCGTTCCCGC TCCTCGACAA GTGGCCCCAC CACGGGTAGG
         V  E  L  D  G  D  V  N  G  H  K  F  S  V  S  G
  301 TGGTCGAGCT GGACGGCGAC GTAAACGGCC ACAAGTTCAG CGTGTCCGGC
      ACCAGCTCGA CCTGCCGCTG CATTTGCCGG TGTTCAAGTC GCACAGGCCG
         E  G  E  G  D  A  T  Y  G  K  L  T  L  K  F  I  C
  351 GAGGGCGAGG GCGATGCCAC CTACGGCAAG CTGACCCTGA AGTTCATCTG
      CTCCCGCTCC CGCTACGGTG GATGCCGTTC GACTGGGACT TCAAGTAGAC
         T  T  G  K  L  P  V  P  W  P  T  L  V  T  T  L  T
  401 CACCACCGGC AAGCTGCCCG TGCCCTGGCC CACCCTCGTG ACCACCCTGA
      GTGGTGGCCG TTCGACGGGC ACGGGACCGG GTGGGAGCAC TGGTGGGACT
         Y  G  V  Q  C  F  S  R  Y  P  D  H  M  K  Q  H
  451 CCTACGGCGT GCAGTGCTTC AGCCGCTACC CCGACCACAT GAAGCAGCAC
      GGATGCCGCA CGTCACGAAG TCGGCGATGG GGCTGGTGTA CTTCGTCGTG
         D  F  F  K  S  A  M  P  E  G  Y  V  Q  E  R  T  I
  501 GACTTCTTCA AGTCCGCCAT GCCCGAAGGC TACGTCCAGG AGCGCACCAT
      CTGAAGAAGT TCAGGCGGTA CGGGCTTCCG ATGCAGGTCC TCGCGTGGTA
         F  F  K  D  D  G  N  Y  K  T  R  A  E  V  K  F  E
  551 CTTCTTCAAG GACGACGGCA ACTACAAGAC CCGCGCCGAG GTGAAGTTCG
      GAAGAAGTTC CTGCTGCCGT TGATGTTCTG GGCGCGGCTC CACTTCAAGC
         G  D  T  L  V  N  R  I  E  L  K  G  I  D  F  K
  601 AGGGCGACAC CCTGGTGAAC CGCATCGAGC TGAAGGGCAT CGACTTCAAG
      TCCCGCTGTG GGACCACTTG GCGTAGCTCG ACTTCCCGTA GCTGAAGTTC
         E  D  G  N  I  L  G  H  K  L  E  Y  N  Y  N  S  H
  651 GAGGACGGCA ACATCCTGGG GCACAAGCTG GAGTACAACT ACAACAGCCA
      CTCCTGCCGT TGTAGGACCC CGTGTTCGAC CTCATGTTGA TGTTGTCGGT
         N  V  Y  I  M  A  D  K  Q  K  N  G  I  K  V  N  F
  701 CAACGTCTAT ATCATGGCCG ACAAGCAGAA GAACGGCATC AAGGTGAACT
      GTTGCAGATA TAGTACCGGC TGTTCGTCTT CTTGCCGTAG TTCCACTTGA
         K  I  R  H  N  I  E  D  G  S  V  Q  L  A  D  H
  751 TCAAGATCCG CCACAACATC GAGGACGGCA GCGTGCAGCT CGCCGACCAC
```

```
     AGTTCTAGGC GGTGTTGTAG CTCCTGCCGT CGCACGTCGA GCGGCTGGTG
      Y  Q  Q  N  T  P  I  G  D  G  P  V  L  L  P  D  N ·
801  TACCAGCAGA ACACCCCCAT CGGCGACGGC CCCGTGCTGC TGCCCGACAA
     ATGGTCGTCT TGTGGGGGTA GCCGCTGCCG GGGCACGACG ACGGGCTGTT
     · H  Y  L  S  T  Q  S  A  L  S  K  D  P  N  E  K  R ·
851  CCACTACCTG AGCACCCAGT CCGCCCTGAG CAAAGACCCC AACGAGAAGC
     GGTGATGGAC TCGTGGGTCA GGCGGGACTC GTTTCTGGGG TTGCTCTTCG

     · D  H  M  V  L  L  E  F  V  T  A  A  G  I  T  L
901  GCGATCACAT GGTCCTGCTG GAGTTCGTGA CCGCCGCCGG GATCACTCTC
     CGCTAGTGTA CCAGGACGAC CTCAAGCACT GGCGGCGGCC CTAGTGAGAG
                                HindIII
                                ~~~~~~~
      G  M  D  E  L  Y  K  *  *
951  GGCATGGACG AGCTGTACAA GTAATAAGCT T
     CCGTACCTGC TCGACATGTT CATTATTCGA A
```

Figure 26B

Figure 27A

Urm1-GFP
Urm1 NCBI ACCESSION# NP_587744

```
      NcoI
      ~~~~~~
         M  G  H  H  H  H  H  H  G  V  N  V  K  V  E  F
    1 CCATGGGTCA TCACCATCAT CATCACGGGG TAAACGTGAA AGTGGAGTTT
      GGTACCCAGT AGTGGTAGTA GTAGTGCCCC ATTTGCACTT TCACCTCAAA
       L  G  G  L  D  A  I  F  G  K  Q  R  V  H  K  I  K ·
   51 CTAGGTGGAC TTGATGCTAT TTTTGGAAAA CAAAGAGTAC ATAAAATTAA
      GATCCACCTG AACTACGATA AAAACCTTTT GTTTCTCATG TATTTTAATT
      · M  D  K  E  D  P  V  T  V  G  D  L  I  D  H  I  V ·
  101 GATGGACAAA GAAGATCCTG TCACAGTGGG CGATTTGATT GACCACATTG
      CTACCTGTTT CTTCTAGGAC AGTGTCACCC GCTAAACTAA CTGGTGTAAC
      · S  T  M  I  N  N  P  N  D  V  S  I  F  I  E  D
  151 TATCTACTAT GATCAATAAC CCTAATGACG TTAGTATCTT CATCGAAGAT
      ATAGATGATA CTAGTTATTG GGATTACTGC AATCATAGAA GTAGCTTCTA
       D  S  I  R  P  G  I  I  T  L  I  N  D  T  D  W  E ·
  201 GATTCTATAA GACCCGGTAT CATCACATTA ATCAACGACA CCGACTGGGA
      CTAAGATATT CTGGGCCATA GTAGTGTAAT TAGTTGCTGT GGCTGACCCT
      · L  E  G  E  K  D  Y  I  L  E  D  G  D  I  I  S  F ·
  251 GCTCGAAGGC GAAAAAGACT ACATATTGGA AGACGGTGAC ATCATCTCTT
      CGAGCTTCCG CTTTTTCTGA TGTATAACCT TCTGCCACTG TAGTAGAGAA
      · T  S  T  L  H  G  G  M  V  S  K  G  E  E  L  F
  301 TTACTTCAAC ATTACATGGA GGTATGGTGA GCAAGGGCGA GGAGCTGTTC
      AATGAAGTTG TAATGTACCT CCATACCACT CGTTCCCGCT CCTCGACAAG
       T  G  V  V  P  I  L  V  E  L  D  G  D  V  N  G  H ·
  351 ACCGGGGTGG TGCCCATCCT GGTCGAGCTG GACGGCGACG TAAACGGCCA
      TGGCCCCACC ACGGGTAGGA CCAGCTCGAC CTGCCGCTGC ATTTGCCGGT
      · K  F  S  V  S  G  E  G  E  G  D  A  T  Y  G  K  L ·
  401 CAAGTTCAGC GTGTCCGGCG AGGGCGAGGG CGATGCCACC TACGGCAAGC
      GTTCAAGTCG CACAGGCCGC TCCCGCTCCC GCTACGGTGG ATGCCGTTCG
      · T  L  K  F  I  C  T  T  G  K  L  P  V  P  W  P
  451 TGACCCTGAA GTTCATCTGC ACCACCGGCA AGCTGCCCGT GCCCTGGCCC
      ACTGGGACTT CAAGTAGACG TGGTGGCCGT TCGACGGGCA CGGGACCGGG
       T  L  V  T  T  L  T  Y  G  V  Q  C  F  S  R  Y  P ·
  501 ACCCTCGTGA CCACCCTGAC CTACGGCGTG CAGTGCTTCA GCCGCTACCC
      TGGGAGCACT GGTGGGACTG GATGCCGCAC GTCACGAAGT CGGCGATGGG
      · D  H  M  K  Q  H  D  F  F  K  S  A  M  P  E  G  Y ·
  551 CGACCACATG AAGCAGCACG ACTTCTTCAA GTCCGCCATG CCCGAAGGCT
      GCTGGTGTAC TTCGTCGTGC TGAAGAAGTT CAGGCGGTAC GGGCTTCCGA
      · V  Q  E  R  T  I  F  F  K  D  D  G  N  Y  K  T
  601 ACGTCCAGGA GCGCACCATC TTCTTCAAGG ACGACGGCAA CTACAAGACC
      TGCAGGTCCT CGCGTGGTAG AAGAAGTTCC TGCTGCCGTT GATGTTCTGG
       R  A  E  V  K  F  E  G  D  T  L  V  N  R  I  E  L ·
  651 CGCGCCGAGG TGAAGTTCGA GGGCGACACC CTGGTGAACC GCATCGAGCT
      GCGCGGCTCC ACTTCAAGCT CCCGCTGTGG GACCACTTGG CGTAGCTCGA
      · K  G  I  D  F  K  E  D  G  N  I  L  G  H  K  L  E
  701 GAAGGGCATC GACTTCAAGG AGGACGGCAA CATCCTGGGG CACAAGCTGG
      CTTCCCGTAG CTGAAGTTCC TCCTGCCGTT GTAGGACCCC GTGTTCGACC
```

```
      · Y N Y  N S H  N V Y I  M A D  K Q K
 751  AGTACAACTA CAACAGCCAC AACGTCTATA TCATGGCCGA CAAGCAGAAG
      TCATGTTGAT GTTGTCGGTG TTGCAGATAT AGTACCGGCT GTTCGTCTTC
       N G I K  V N F  K I R  H N I E  D G S ·
 801  AACGGCATCA AGGTGAACTT CAAGATCCGC CACAACATCG AGGACGGCAG
      TTGCCGTAGT TCCACTTGAA GTTCTAGGCG GTGTTGTAGC TCCTGCCGTC
      · V Q L  A D H Y  Q Q N  T P I  G D G P ·
 851  CGTGCAGCTC GCCGACCACT ACCAGCAGAA CACCCCCATC GGCGACGGCC
      GCACGTCGAG CGGCTGGTGA TGGTCGTCTT GTGGGGGTAG CCGCTGCCGG
      · V L L  P D N  H Y L S  T Q S  A L S
 901  CCGTGCTGCT GCCCGACAAC CACTACCTGA GCACCCAGTC CGCCCTGAGC
      GGCACGACGA CGGGCTGTTG GTGATGGACT CGTGGGTCAG GCGGGACTCG
       K D P N  E K R  D H M  V L L E  F V T ·
 951  AAAGACCCCA ACGAGAAGCG CGATCACATG GTCCTGCTGG AGTTCGTGAC
      TTTCTGGGGT TGCTCTTCGC GCTAGTGTAC CAGGACGACC TCAAGCACTG
                                                 HindIII
                                                 ~~~~~~
      · A A G  I T L G  M D E  L Y K  * *
1001  CGCCGCCGGG ATCACTCTCG GCATGGACGA GCTGTACAAG TAATAAGCTT
      GCGGCGGCCC TAGTGAGAGC CGTACCTGCT CGACATGTTC ATTATTCGAA
```

Figure 27B

Figure 28A

Hub1-GFP

Hub1 NCBI ACCESSION# XM_114578

```
      NcoI
      ~~~~~~
         M  G  H     H  Y  H     H  H  G  M     I  E  V     V  V  N
    1 CCATGGGTCA TCACTATCAT CATCACGGGA TGATTGAGGT AGTTGTGAAT
      GGTACCCAGT AGTGATAGTA GTAGTGCCCT ACTAACTCCA TCAACACTTA
       D  R  L  G     K  K  V     R  V  K     C  L  A  E     D  S  V
   51 GACCGATTAG GCAAAAAAGT CAGAGTGAAG TGCCTTGCTG AAGATAGTGT
      CTGGCTAATC CGTTTTTTCA GTCTCACTTC ACGGAACGAC TTCTATCACA
       G  D  F     K  K  V  L     S  L  Q     I  G  T     Q  P  N  K
  101 AGGTGATTTC AAAAAAGTAT TGTCCTTGCA AATTGGCACC CAACCAAACA
      TCCACTAAAG TTTTTTCATA ACAGGAACGT TTAACCGTGG GTTGGTTTGT
       I  V  L     Q  K  G     G  S  V  L     K  D  H     I  S  L
  151 AAATTGTGTT GCAGAAGGGT GGAAGTGTTT TAAAAGACCA TATCTCTCTG
      TTTAACACAA CGTCTTCCCA CCTTCACAAA ATTTTCTGGT ATAGAGAGAC
       E  D  Y  E     V  H  D     Q  T  N     L  E  L  Y     Y  M  V
  201 GAAGATTATG AGGTACATGA TCAGACAAAT TTGGAGCTGT ATTACATGGT
      CTTCTAATAC TCCATGTACT AGTCTGTTTA AACCTCGACA TAATGTACCA
       S  K  G     E  E  L  F     T  G  V     V  P  I     L  V  E  L
  251 GAGCAAGGGC GAGGAGCTGT TCACCGGGGT GGTGCCCATC CTGGTCGAGC
      CTCGTTCCCG CTCCTCGACA AGTGGCCCCA CCACGGGTAG GACCAGCTCG
       D  G  D     V  N  G     H  K  F  S     V  S  G     E  G  E
  301 TGGACGGCGA CGTAAACGGC CACAAGTTCA GCGTGTCCGG CGAGGGCGAG
      ACCTGCCGCT GCATTTGCCG GTGTTCAAGT CGCACAGGCC GCTCCCGCTC
       G  D  A  T     Y  G  K     L  T  L     K  F  I  C     T  T  G
  351 GGCGATGCCA CCTACGGCAA GCTGACCCTG AAGTTCATCT GCACCACCGG
      CCGCTACGGT GGATGCCGTT CGACTGGGAC TTCAAGTAGA CGTGGTGGCC
       K  L  P     V  P  W  P     T  L  V     T  T  L     T  Y  G  V
  401 CAAGCTGCCC GTGCCCTGGC CCACCCTCGT GACCACCCTG ACCTACGGCG
      GTTCGACGGG CACGGGACCG GGTGGGAGCA CTGGTGGGAC TGGATGCCGC
       Q  C  F     S  R  Y     P  D  H  M     K  Q  H     D  F  F
  451 TGCAGTGCTT CAGCCGCTAC CCCGACCACA TGAAGCAGCA CGACTTCTTC
      ACGTCACGAA GTCGGCGATG GGGCTGGTGT ACTTCGTCGT GCTGAAGAAG
       K  S  A  M     P  E  G     Y  V  Q     E  R  T  I     F  F  K
  501 AAGTCCGCCA TGCCCGAAGG CTACGTCCAG GAGCGCACCA TCTTCTTCAA
      TTCAGGCGGT ACGGGCTTCC GATGCAGGTC CTCGCGTGGT AGAAGAAGTT
       D  D  G     N  Y  K  T     R  A  E     V  K  F     E  G  D  T
  551 GGACGACGGC AACTACAAGA CCCGCGCCGA GGTGAAGTTC GAGGGCGACA
      CCTGCTGCCG TTGATGTTCT GGGCGCGGCT CCACTTCAAG CTCCCGCTGT
       L  V  N     R  I  E     L  K  G  I     D  F  K     E  D  G
  601 CCCTGGTGAA CCGCATCGAG CTGAAGGGCA TCGACTTCAA GGAGGACGGC
      GGGACCACTT GGCGTAGCTC GACTTCCCGT AGCTGAAGTT CCTCCTGCCG
       N  I  L  G     H  K  L     E  Y  N     Y  N  S  H     N  V  Y
  651 AACATCCTGG GGCACAAGCT GGAGTACAAC TACAACAGCC ACAACGTCTA
      TTGTAGGACC CCGTGTTCGA CCTCATGTTG ATGTTGTCGG TGTTGCAGAT
       I  M  A     D  K  Q  K     N  G  I     K  V  N     F  K  I  R
  701 TATCATGGCC GACAAGCAGA AGAACGGCAT CAAGGTGAAC TTCAAGATCC
```

```
     ATAGTACCGG CTGTTCGTCT TCTTGCCGTA GTTCCACTTG AAGTTCTAGG
      · H  N  I   E  D  G   S  V  Q  L   A  D  H   Y  Q  Q
751  GCCACAACAT CGAGGACGGC AGCGTGCAGC TCGCCGACCA CTACCAGCAG
     CGGTGTTGTA GCTCCTGCCG TCGCACGTCG AGCGGCTGGT GATGGTCGTC
      N  T  P  I   G  D  G   P  V  L   L  P  D  N   H  Y  L ·
801  AACACCCCCA TCGGCGACGG CCCCGTGCTG CTGCCCGACA ACCACTACCT
     TTGTGGGGGT AGCCGCTGCC GGGGCACGAC GACGGGCTGT TGGTGATGGA
     · S  T  Q   S  A  L  S   K  D  P   N  E  K   R  D  H  M ·
851  GAGCACCCAG TCCGCCCTGA GCAAAGACCC CAACGAGAAG CGCGATCACA
     CTCGTGGGTC AGGCGGGACT CGTTTCTGGG GTTGCTCTTC GCGCTAGTGT
     · V  L  L   E  F  V   T  A  A  G   I  T  L   G  M  D
901  TGGTCCTGCT GGAGTTCGTG ACCGCCGCCG GGATCACTCT CGGCATGGAC
     ACCAGGACGA CCTCAAGCAC TGGCGGCGGC CCTAGTGAGA GCCGTACCTG
                       HindIII
                       ~~~~~~~
      E  L  Y  K   *  *
951  GAGCTGTACA AGTAATAAGC TT
     CTCGACATGT TCATTATTCG AA
```

Figure 28B

Figure 29A

Rub1-GFP
Rub1 NCBI Accession# Y16890

```
      NcoI
      ~~~~~~
         M  G  H     H  H  H     H  H  G  I     V  K  X     K  T  L
   1  CCATGGGTCA TCACCATCAT CATCACGGGA TTGTTAAAGN GAAGACACTG
      GGTACCCAGT AGTGGTAGTA GTAGTGCCCT AACAATTTCN CTTCTGTGAC
       T  G  K  E     I  S  V     E  L  K     E  S  D  L     V  Y  H  ·
  51  ACTGGGAAGG AGATCTCTGT TGAGCTGAAG GAATCAGATC TCGTATATCA
      TGACCCTTCC TCTAGAGACA ACTCGACTTC CTTAGTCTAG AGCATATAGT
      ·  I  K  E     L  L  E  E     K  E  G     I  P  P     S  Q  Q  R  ·
 101  CATCAAGGAA CTTTTGGAGG AAAAAGAAGG GATTCCACCA TCTCAACAAA
      GTAGTTCCTT GAAAACCTCC TTTTTCTTCC CTAAGGTGGT AGAGTTGTTT
      ·  L  I  F     Q  G  K     Q  I  D  D     K  L  T     V  T  D
 151  GACTTATATT CCAGGGAAAA CAAATTGATG ATAAATTAAC AGTAACGGAT
      CTGAATATAA GGTCCCTTTT GTTTAACTAC TATTTAATTG TCATTGCCTA
       A  H  X  V     E  G  M     Q  L  H     L  V  L  T     L  R  G  ·
 201  GCACATNTAG TAGAGGGAAT GCAACTCCAC TTGGTATTAA CACTACGCGG
      CGTGTANATC ATCTCCCTTA CGTTGAGGTG AACCATAATT GTGATGCGCC
      ·  G  M  V     S  K  G  E     E  L  F     T  G  V     V  P  I  L  ·
 251  AGGTATGGTG AGCAAGGGCG AGGAGCTGTT CACCGGGGTG GTGCCCATCC
      TCCATACCAC TCGTTCCCGC TCCTCGACAA GTGGCCCCAC CACGGGTAGG
      ·  V  E  L     D  G  D     V  N  G  H     K  F  S     V  S  G
 301  TGGTCGAGCT GGACGGCGAC GTAAACGGCC ACAAGTTCAG CGTGTCCGGC
      ACCAGCTCGA CCTGCCGCTG CATTTGCCGG TGTTCAAGTC GCACAGGCCG
       E  G  E  G     D  A  T     Y  G  K     L  T  L  K     F  I  C  ·
 351  GAGGGCGAGG GCGATGCCAC CTACGGCAAG CTGACCCTGA AGTTCATCTG
      CTCCCGCTCC CGCTACGGTG GATGCCGTTC GACTGGGACT TCAAGTAGAC
      ·  T  T  G     K  L  P  V     P  W  P     T  L  V     T  T  L  T  ·
 401  CACCACCGGC AAGCTGCCCG TGCCCTGGCC CACCCTCGTG ACCACCCTGA
      GTGGTGGCCG TTCGACGGGC ACGGGACCGG GTGGGAGCAC TGGTGGGACT
      ·  Y  G  V     Q  C  F     S  R  Y  P     D  H  M     K  Q  H
 451  CCTACGGCGT GCAGTGCTTC AGCCGCTACC CCGACCACAT GAAGCAGCAC
      GGATGCCGCA CGTCACGAAG TCGGCGATGG GGCTGGTGTA CTTCGTCGTG
       D  F  F  K     S  A  M     P  E  G     Y  V  Q  E     R  T  I  ·
 501  GACTTCTTCA AGTCCGCCAT GCCCGAAGGC TACGTCCAGG AGCGCACCAT
      CTGAAGAAGT TCAGGCGGTA CGGGCTTCCG ATGCAGGTCC TCGCGTGGTA
      ·  F  F  K     D  D  G  N     Y  K  T     R  A  E     V  K  F  E  ·
 551  CTTCTTCAAG GACGACGGCA ACTACAAGAC CCGCGCCGAG GTGAAGTTCG
      GAAGAAGTTC CTGCTGCCGT TGATGTTCTG GGCGCGGCTC CACTTCAAGC
      ·  G  D  T     L  V  N     R  I  E  L     K  G  I     D  F  K
 601  AGGGCGACAC CCTGGTGAAC CGCATCGAGC TGAAGGGCAT CGACTTCAAG
      TCCCGCTGTG GGACCACTTG GCGTAGCTCG ACTTCCCGTA GCTGAAGTTC
       E  D  G  N     I  L  G     H  K  L     E  Y  N  Y     N  S  H  ·
 651  GAGGACGGCA ACATCCTGGG GCACAAGCTG GAGTACAACT ACAACAGCCA
      CTCCTGCCGT TGTAGGACCC CGTGTTCGAC CTCATGTTGA TGTTGTCGGT
      ·  N  V  Y     I  M  A  D     K  Q  K     N  G  I     K  V  N  F  ·
 701  CAACGTCTAT ATCATGGCCG ACAAGCAGAA GAACGGCATC AAGGTGAACT
```

```
     GTTGCAGATA TAGTACCGGC TGTTCGTCTT CTTGCCGTAG TTCCACTTGA
     · K I R H N I E D G S V Q L A D H
751  TCAAGATCCG CCACAACATC GAGGACGGCA GCGTGCAGCT CGCCGACCAC
     AGTTCTAGGC GGTGTTGTAG CTCCTGCCGT CGCACGTCGA GCGGCTGGTG
      Y Q Q N T P I G D G P V L L P D N ·
801  TACCAGCAGA ACACCCCCAT CGGCGACGGC CCCGTGCTGC TGCCCGACAA
     ATGGTCGTCT TGTGGGGGTA GCCGCTGCCG GGGCACGACG ACGGGCTGTT
     · H Y L S T Q S A L S K D P N E K R ·
851  CCACTACCTG AGCACCCAGT CCGCCCTGAG CAAAGACCCC AACGAGAAGC
     GGTGATGGAC TCGTGGGTCA GGCGGGACTC GTTTCTGGGG TTGCTCTTCG
     · D H M V L L E F V T A A G I T L
901  GCGATCACAT GGTCCTGCTG GAGTTCGTGA CCGCCGCCGG GATCACTCTC
     CGCTAGTGTA CCAGGACGAC CTCAAGCACT GGCGGCGGCC CTAGTGAGAG

                                HindIII
                                ~~~~~~~
      G M D E L Y K * *
951  GGCATGGACG AGCTGTACAA GTAATAAGCT T
     CCGTACCTGC TCGACATGTT CATTATTCGA A
```

Figure 29B

Figure 30A

Apg8-GFP

Apg8 NCBI ACCESSION# P38182

```
          NcoI
          ~~~~
           M  G  H     H  H  H     H  H  G  K     S  T  F     K  S  E
     1    ATGGGTCA TCACCATCAT CATCACGGGA AGTCTACATT TAAGTCTGAA
          TACCCAGT AGTGGTAGTA GTAGTGCCCT TCAGATGTAA ATTCAGACTT
         Y  P  F  E     K  R  K     A  E  S     E  R  I  A     D  R  F  .
    51   TATCCATTTG AAAAAAGGAA GGCGGAGTCG GAGAGGATTG CTGACAGGTT
         ATAGGTAAAC TTTTTTCCTT CCGCCTCAGC CTCTCCTAAC GACTGTCCAA
         .  K  N  R     I  P  V  I     C  E  K     A  E  K     S  D  I  P  .
   101   CAAGAATAGG ATACCTGTGA TTTGCGAAAA AGCTGAAAAG TCAGATATTC
         GTTCTTATCC TATGGACACT AAACGCTTTT TCGACTTTTC AGTCTATAAG
         .  E  I  D     K  R  K     Y  L  V  P     A  D  L     T  V  G
   151   CAGAGATTGA TAAGCGTAAA TATCTAGTTC CTGCTGACCT TACCGTAGGG
         GTCTCTAACT ATTCGCATTT ATAGATCAAG GACGACTGGA ATGGCATCCC
          Q  F  V  Y     V  I  R     K  R  I     M  L  P  P     E  K  A  .
   201   CAATTTGTTT ATGTTATAAG AAAGAGGATT ATGCTACCCC CTGAGAAGGC
         GTTAAACAAA TACAATATTC TTTCTCCTAA TACGATGGGG GACTCTTCCG
         .  I  F  I     F  V  N  D     T  L  P     P  T  A     A  L  M  S  .
   251   CATCTTCATT TTTGTCAATG ATACTTTGCC ACCTACTGCG GCGTTGATGT
         GTAGAAGTAA AAACAGTTAC TATGAAACGG TGGATGACGC CGCAACTACA
         .  A  I  Y     Q  E  H     K  D  K  D     G  F  L     Y  V  T
   301   CTGCCATATA TCAAGAACAC AAGGATAAGG ACGGGTTTTT GTATGTCACT
         GACGGTATAT AGTTCTTGTG TTCCTATTCC TGCCCAAAAA CATACAGTGA
         .  Y  S  G  E     N  T  F     G  M  V     S  K  G  E     E  L  F  .
   351   TACTCAGGAG AAAATACATT TGGTATGGTG AGCAAGGGCG AGGAGCTGTT
         ATGAGTCCTC TTTTATGTAA ACCATACCAC TCGTTCCCGC TCCTCGACAA
         .  T  G  V     V  P  I  L     V  E  L     D  G  D     V  N  G  H  .
   401   CACCGGGGTG GTGCCCATCC TGGTCGAGCT GGACGGCGAC GTAAACGGCC
         GTGGCCCCAC CACGGGTAGG ACCAGCTCGA CCTGCCGCTG CATTTGCCGG
         .  K  F  S     V  S  G     E  G  E  G     D  A  T     Y  G  K
   451   ACAAGTTCAG CGTGTCCGGC GAGGGCGAGG GCGATGCCAC CTACGGCAAG
         TGTTCAAGTC GCACAGGCCG CTCCCGCTCC CGCTACGGTG GATGCCGTTC
          L  T  L  K     F  I  C     T  T  G     K  L  P  V     P  W  P  .
   501   CTGACCCTGA AGTTCATCTG CACCACCGGC AAGCTGCCCG TGCCCTGGCC
         GACTGGGACT TCAAGTAGAC GTGGTGGCCG TTCGACGGGC ACGGGACCGG
         .  T  L  V     T  T  L  T     Y  G  V     Q  C  F     S  R  Y  P  .
   551   CACCCTCGTG ACCACCCTGA CCTACGGCGT GCAGTGCTTC AGCCGCTACC
         GTGGGAGCAC TGGTGGGACT GGATGCCGCA CGTCACGAAG TCGGCGATGG
         .  D  H  M     K  Q  H     D  F  F  K     S  A  M     P  E  G
   601   CCGACCACAT GAAGCAGCAC GACTTCTTCA AGTCCGCCAT GCCCGAAGGC
         GGCTGGTGTA CTTCGTCGTG CTGAAGAAGT TCAGGCGGTA CGGGCTTCCG
          Y  V  Q  E     R  T  I     F  F  K     D  D  G  N     Y  K  T  .
```

```
651  TACGTCCAGG AGCGCACCAT CTTCTTCAAG GACGACGGCA ACTACAAGAC
     ATGCAGGTCC TCGCGTGGTA GAAGAAGTTC CTGCTGCCGT TGATGTTCTG
     · R  A  E   V  K  F  E   G  D  T   L  V  N   R  I  E  L ·
701  CCGCGCCGAG GTGAAGTTCG AGGGCGACAC CCTGGTGAAC CGCATCGAGC
     GGCGCGGCTC CACTTCAAGC TCCCGCTGTG GGACCACTTG GCGTAGCTCG
     · K  G  I   D  F  K   E  D  G  N   I  L  G   H  K  L
751  TGAAGGGCAT CGACTTCAAG GAGGACGGCA ACATCCTGGG GCACAAGCTG
     ACTTCCCGTA GCTGAAGTTC CTCCTGCCGT TGTAGGACCC CGTGTTCGAC
      E  Y  N  Y   N  S  H   N  V  Y   I  M  A  D   K  Q  K ·
801  GAGTACAACT ACAACAGCCA CAACGTCTAT ATCATGGCCG ACAAGCAGAA
     CTCATGTTGA TGTTGTCGGT GTTGCAGATA TAGTACCGGC TGTTCGTCTT
     · N  G  I   K  V  N  F   K  I  R   H  N  I   E  D  G  S ·
851  GAACGGCATC AAGGTGAACT TCAAGATCCG CCACAACATC GAGGACGGCA
     CTTGCCGTAG TTCCACTTGA AGTTCTAGGC GGTGTTGTAG CTCCTGCCGT
     · V  Q  L   A  D  H   Y  Q  Q  N   T  P  I   G  D  G
901  GCGTGCAGCT CGCCGACCAC TACCAGCAGA ACACCCCCAT CGGCGACGGC
     CGCACGTCGA GCGGCTGGTG ATGGTCGTCT TGTGGGGGTA GCCGCTGCCG
      P  V  L  L   P  D  N   H  Y  L   S  T  Q  S   A  L  S ·
951  CCCGTGCTGC TGCCCGACAA CCACTACCTG AGCACCCAGT CCGCCCTGAG
     GGGCACGACG ACGGGCTGTT GGTGATGGAC TCGTGGGTCA GGCGGGACTC
     · K  D  P   N  E  K  R   D  H  M   V  L  L   E  F  V  T ·
1001 CAAAGACCCC AACGAGAAGC GCGATCACAT GGTCCTGCTG GAGTTCGTGA
     GTTTCTGGGG TTGCTCTTCG CGCTAGTGTA CCAGGACGAC CTCAAGCACT
                                                    HindIII
                                                    ~~~~~
     · A  A  G   I  T  L   G  M  D  E   L  Y  K   *  *  A
1051 CCGCCGCCGG GATCACTCTC GGCATGGACG AGCTGTACAA GTAATAAGCTT
     GGCGGCGGCC CTAGTGAGAG CCGTACCTGC TCGACATGTT CATTATTCGAA
```

Figure 30B

Figure 31A

Apg12-GFP
Apg12 NCBI ACCESSION# P38316

NcoI
~~~~~~

```
      M  G  H   H  H  H   H  H  G  S   R  I  L   E  S  E
  1 CCATGGGTCA TCACCATCAT CATCACGGGA GTAGGATCCT AGAGAGCGAA
    GGTACCCAGT AGTGGTAGTA GTAGTGCCCT CATCCTAGGA TCTCTCGCTT
     N  E  T  E   S  D  E   S  S  I   I  S  T  N   N  G  T ·
 51 AATGAAACAG AAAGTGACGA AAGCTCCATC ATATCCACAA ATAATGGAAC
    TTACTTTGTC TTTCACTGCT TTCGAGGTAG TATAGGTGTT TATTACCTTG
    · A  M  E   R  S  R  N   N  Q  E   L  R  S   S  P  H  T ·
101 GGCAATGGAA AGATCCAGAA ATAATCAAGA ATTAAGATCA TCTCCTCATA
    CCGTTACCTT TCTAGGTCTT TATTAGTTCT TAATTCTAGT AGAGGAGTAT
    · V  Q  N   R  L  E   L  F  S  R   R  L  S   Q  L  G
151 CCGTTCAAAA TAGATTGGAA CTTTTTAGCA GGAGATTGTC TCAGCTTGGT
    GGCAAGTTTT ATCTAACCTT GAAAAATCGT CCTCTAACAG AGTCGAACCA
     L  A  S  D   I  S  V   D  Q  Q   V  E  D  S   S  S  G ·
201 TTGGCGAGTG ACATTTCTGT CGACCAGCAA GTTGAAGATT CCTCTAGTGG
    AACCGCTCAC TGTAAAGACA GCTGGTCGTT CAACTTCTAA GGAGATCACC
    · T  Y  E   Q  E  E  T   I  K  T   N  A  Q   T  S  K  Q ·
251 CACTTATGAA CAGGAAGAGA CAATCAAAAC GAATGCACAA ACAAGCAAAC
    GTGAATACTT GTCCTTCTCT GTTAGTTTTG CTTACGTGTT TGTTCGTTTG
    · K  S  H   K  D  E   K  N  I  Q   K  I  Q   I  K  F
301 AAAAAAGCCA TAAAGACGAA AAAAACATAC AAAAGATACA GATAAAATTT
    TTTTTTCGGT ATTTCTGCTT TTTTTGTATG TTTTCTATGT CTATTTTAAA
     Q  P  I  G   S  I  G   Q  L  K   P  S  V  C   K  I  S ·
351 CAGCCCATTG GTTCTATTGG GCAGTTAAAA CCATCTGTTT GTAAAATATC
    GTCGGGTAAC CAAGATAACC CGTCAATTTT GGTAGACAAA CATTTTATAG
    · M  S  Q   S  F  A  M   V  I  L   F  L  K   R  R  L  K ·
401 NATGTCACAG TCTTTTGCAA TGGTTATTTT ATTTCTTAAG AGACGGCTGA
    NTACAGTGTC AGAAAACGTT ACCAATAAAA TAAAGAATTC TCTGCCGACT
    · M  D  H   V  Y  C   Y  I  N  N   S  F  A   P  S  P
451 AAATGGACCA TGTTTATTGT TATATAAATA ATTCGTTTGC GCCAAGTCCG
    TTTACCTGGT ACAAATAACA ATATATTTAT TAAGCAAACG CGGTTCAGGC
     Q  Q  N  I   G  E  L   W  M  X   F  K  T  N   D  E  L ·
501 CAGCAAAATA TTGGTGAACT TTGGATGCNA TTCAAGACTA ATGATGAGCT
    GTCGTTTTAT AACCACTTGA AACCTACGNT AAGTTCTGAT TACTACTCGA
    · I  V  S   Y  C  A  S   V  A  F   G  M  V   S  K  G  E ·
551 TATTGTAAGT TATTGTGCAT CCGTAGCGTT TGGTATGGTG AGCAAGGGCG
    ATAACATTCA ATAACACGTA GGCATCGCAA ACCATACCAC TCGTTCCCGC
    · E  L  F   T  G  V   V  P  I  L   V  E  L   D  G  D
601 AGGAGCTGTT CACCGGGGTG GTGCCCATCC TGGTCGAGCT GGACGGCGAC
    TCCTCGACAA GTGGCCCCAC CACGGGTAGG ACCAGCTCGA CCTGCCGCTG
     V  N  G  H   K  F  S   V  S  G   E  G  E  G   D  A  T ·
651 GTAAACGGCC ACAAGTTCAG CGTGTCCGGC GAGGGCGAGG GCGATGCCAC
    CATTTGCCGG TGTTCAAGTC GCACAGGCCG CTCCCGCTCC CGCTACGGTG
    · Y  G  K   L  T  L  K   F  I  C   T  T  G   K  L  P  V ·
701 CTACGGCAAG CTGACCCTGA AGTTCATCTG CACCACCGGC AAGCTGCCCG
    GATGCCGTTC GACTGGGACT TCAAGTAGAC GTGGTGGCCG TTCGACGGGC
```
~~~~~~

```
      · P W P  T L V  T T L T  Y G V  Q C F
751   TGCCCTGGCC CACCCTCGTG ACCACCCTGA CCTACGGCGT GCAGTGCTTC
      ACGGGACCGG GTGGGAGCAC TGGTGGGACT GGATGCCGCA CGTCACGAAG
       S R Y P  D H M  K Q H  D F F K  S A M ·
801   AGCCGCTACC CCGACCACAT GAAGCAGCAC GACTTCTTCA AGTCCGCCAT
      TCGGCGATGG GGCTGGTGTA CTTCGTCGTG CTGAAGAAGT TCAGGCGGTA
      · P E G  Y V Q  E R T  I F F K  D D G N ·
851   GCCCGAAGGC TACGTCCAGG AGCGCACCAT CTTCTTCAAG GACGACGGCA
      CGGGCTTCCG ATGCAGGTCC TCGCGTGGTA GAAGAAGTTC CTGCTGCCGT
      · Y K T  R A E  V K F E  G D T  L V N
901   ACTACAAGAC CCGCGCCGAG GTGAAGTTCG AGGGCGACAC CCTGGTGAAC
      TGATGTTCTG GGCGCGGCTC CACTTCAAGC TCCCGCTGTG GGACCACTTG
       R I E L  K G I  D F K  E D G N  I L G ·
951   CGCATCGAGC TGAAGGGCAT CGACTTCAAG GAGGACGGCA ACATCCTGGG
      GCGTAGCTCG ACTTCCCGTA GCTGAAGTTC CTCCTGCCGT TGTAGGACCC
      · H K L  E Y N Y  N S H  N V Y  I M A D ·
1001  GCACAAGCTG GAGTACAACT ACAACAGCCA CAACGTCTAT ATCATGGCCG
      CGTGTTCGAC CTCATGTTGA TGTTGTCGGT GTTGCAGATA TAGTACCGGC
      · K Q K  N G I  K V N F  K I R  H N I
1051  ACAAGCAGAA GAACGGCATC AAGGTGAACT TCAAGATCCG CCACAACATC
      TGTTCGTCTT CTTGCCGTAG TTCCACTTGA AGTTCTAGGC GGTGTTGTAG
       E D G S  V Q L  A D H  Y Q Q N  T P I ·
1101  GAGGACGGCA GCGTGCAGCT CGCCGACCAC TACCAGCAGA ACACCCCCAT
      CTCCTGCCGT CGCACGTCGA GCGGCTGGTG ATGGTCGTCT TGTGGGGGTA
      · G D G  P V L L  P D N  H Y L  S T Q S ·
1151  CGGCGACGGC CCCGTGCTGC TGCCCGACAA CCACTACCTG AGCACCCAGT
      GCCGCTGCCG GGGCACGACG ACGGGCTGTT GGTGATGGAC TCGTGGGTCA

      · A L S  K D P  N E K R  D H M  V L L
1201  CCGCCCTGAG CAAAGACCCC AACGAGAAGC GCGATCACAT GGTCCTGCTG
      GGCGGGACTC GTTTCTGGGG TTGCTCTTCG CGCTAGTGTA CCAGGACGAC
       E F V T  A A G  I T L  G M D E  L Y K ·
1251  GAGTTCGTGA CCGCCGCCGG GATCACTCTC GGCATGGACG AGCTGTACAA
      CTCAAGCACT GGCGGCGGCC CTAGTGAGAG CCGTACCTGC TCGACATGTT
          HindIII
          ~~~~~~~
      · * **
1301  GTAATAAGCT T
      CATTATTCGA A
```

Figure 31B

Figure 32A

ISG15-GFP
ISG15 NCBI ACCESSION# P05161

```
      NcoI
      ~~~~~~
         M  G  H     H  H  H     H  H  G  G     W  D  L     T  V  K
    1 CCATGGGTCA TCACCATCAT CATCACGGGG GCTGGGACCT GACGGTGAAG
      GGTACCCAGT AGTGGTAGTA GTAGTGCCCC CGACCCTGGA CTGCCACTTC
       M  L  A  G     N  E  F     Q  V  S     L  S  S  S     M  S  V  ·
   51 ATGCTGGCGG GCAACGAATT CCAGGTGTCC CTGAGCAGCT CCATGTCGGT
      TACGACCGCC CGTTGCTTAA GGTCCACAGG GACTCGTCGA GGTACAGCCA
      ·  S  E  L     K  A  Q  I     T  Q  K     I  G  V     H  A  F  Q  ·
  101 GTCAGAGCTG AAGGCGCAGA TCACCCAGAA GATTGGCGTG CACGCCTTCC
      CAGTCTCGAC TTCCGCGTCT AGTGGGTCTT CTAACCGCAC GTGCGGAAGG
      ·  Q  R  L     A  V  H     P  S  G  V     A  L  Q     D  R  V
  151 AGCAGCGTCT GGCTGTCCAC CCGAGCGGTG TGGCGCTGCA GGACAGGGTC
      TCGTCGCAGA CCGACAGGTG GGCTCGCCAC ACCGCGACGT CCTGTCCCAG
       P  L  A  S     Q  G  L     G  P  G     S  T  V  L     L  V  V  ·
  201 CCCCTTGCCA GCCAGGGCCT GGGCCCTGGC AGCACGGTCC TGCTGGTGGT
      GGGGAACGGT CGGTCCCGGA CCCGGGACCG TCGTGCCAGG ACGACCACCA
      ·  D  K  C     D  E  P  L     S  I  L     V  R  N     N  K  G  R  ·
  251 GGACAAATGC GACGAACCTC TGAGCATCCT GGTGAGGAAT AACAAGGGCC
      CCTGTTTACG CTGCTTGGAG ACTCGTAGGA CCACTCCTTA TTGTTCCCGG
      ·  S  S  T     Y  E  V     R  L  T  Q     T  V  A     H  L  K
  301 GCAGCAGCAC CTACGAGGTC CGGCTGACGC AGACCGTGGC CCACCTGAAG
      CGTCGTCGTG GATGCTCCAG GCCGACTGCG TCTGGCACCG GGTGGACTTC
       Q  Q  V  S     G  L  E     G  V  Q     D  D  L  F     W  L  T  ·
  351 CAGCAAGTGA GCGGGCTGGA GGGTGTGCAG GACGACCTGT TCTGGCTGAC
      GTCGTTCACT CGCCCGACCT CCCACACGTC CTGCTGGACA AGACCGACTG
      ·  F  E  G     K  P  L  E     D  Q  L     P  L  G     E  Y  G  L  ·
  401 CTTCGAGGGG AAGCCCCTGG AGGACCAGCT CCCGCTGGGG GAGTACGGCC
      GAAGCTCCCC TTCGGGGACC TCCTGGTCGA GGGCGACCCC CTCATGCCGG
      ·  K  P  L     S  T  V     F  M  N  L     R  L  R     G  G  G
  451 TCAAGCCCCT GAGCACCGTG TTCATGAATC TGCGCCTGCG GGGAGGCGGC
      AGTTCGGGGA CTCGTGGCAC AAGTACTTAG ACGCGGACGC CCCTCCGCCG
       T  E  P  G     G  M  V     S  K  G     E  E  L  F     T  G  V  ·
  501 ACAGAGCCTG GAGGTATGGT GAGCAAGGGC GAGGAGCTGT TCACCGGGGT
      TGTCTCGGAC CTCCATACCA CTCGTTCCCG CTCCTCGACA AGTGGCCCCA
      ·  V  P  I     L  V  E  L     D  G  D     V  N  G     H  K  F  S  ·
  551 GGTGCCCATC CTGGTCGAGC TGGACGGCGA CGTAAACGGC CACAAGTTCA
      CCACGGGTAG GACCAGCTCG ACCTGCCGCT GCATTTGCCG GTGTTCAAGT
      ·  V  S  G     E  G  E     G  D  A  T     Y  G  K     L  T  L
  601 GCGTGTCCGG CGAGGGCGAG GGCGATGCCA CCTACGGCAA GCTGACCCTG
      CGCACAGGCC GCTCCCGCTC CCGCTACGGT GGATGCCGTT CGACTGGGAC
       K  F  I  C     T  T  G     K  L  P     V  P  W  P     T  L  V  ·
  651 AAGTTCATCT GCACCACCGG CAAGCTGCCC GTGCCCTGGC CCACCCTCGT
      TTCAAGTAGA CGTGGTGGCC GTTCGACGGG CACGGGACCG GGTGGGAGCA
      ·  T  T  L     T  Y  G  V     Q  C  F     S  R  Y     P  D  H  M  ·
  701 GACCACCCTG ACCTACGGCG TGCAGTGCTT CAGCCGCTAC CCCGACCACA
```

```
      CTGGTGGGAC TGGATGCCGC ACGTCACGAA GTCGGCGATG GGGCTGGTGT
      · K Q H  D F F  K S A M  P E G  Y V Q
751   TGAAGCAGCA CGACTTCTTC AAGTCCGCCA TGCCCGAAGG CTACGTCCAG
      ACTTCGTCGT GCTGAAGAAG TTCAGGCGGT ACGGGCTTCC GATGCAGGTC
       E R T I  F F K  D D G  N Y K T  R A E ·
801   GAGCGCACCA TCTTCTTCAA GGACGACGGC AACTACAAGA CCCGCGCCGA
      CTCGCGTGGT AGAAGAAGTT CCTGCTGCCG TTGATGTTCT GGGCGCGGCT
      · V K F  E G D T  L V N  R I E  L K G I ·
851   GGTGAAGTTC GAGGGCGACA CCCTGGTGAA CCGCATCGAG CTGAAGGGCA
      CCACTTCAAG CTCCCGCTGT GGGACCACTT GGCGTAGCTC GACTTCCCGT
      · D F K  E D G  N I L G  H K L  E Y N
901   TCGACTTCAA GGAGGACGGC AACATCCTGG GGCACAAGCT GGAGTACAAC
      AGCTGAAGTT CCTCCTGCCG TTGTAGGACC CCGTGTTCGA CCTCATGTTG
       Y N S H  N V Y  I M A  D K Q K  N G I ·
951   TACAACAGCC ACAACGTCTA TATCATGGCC GACAAGCAGA AGAACGGCAT
      ATGTTGTCGG TGTTGCAGAT ATAGTACCGG CTGTTCGTCT TCTTGCCGTA
      · K V N  F K I R  H N I  E D G  S V Q L ·
1001  CAAGGTGAAC TTCAAGATCC GCCACAACAT CGAGGACGGC AGCGTGCAGC
      GTTCCACTTG AAGTTCTAGG CGGTGTTGTA GCTCCTGCCG TCGCACGTCG
      · A D H  Y Q Q  N T P I  G D G  P V L
1051  TCGCCGACCA CTACCAGCAG AACACCCCCA TCGGCGACGG CCCCGTGCTG
      AGCGGCTGGT GATGGTCGTC TTGTGGGGGT AGCCGCTGCC GGGGCACGAC
       L P D N  H Y L  S T Q  S A L S  K D P ·
1101  CTGCCCGACA ACCACTACCT GAGCACCCAG TCCGCCCTGA GCAAAGACCC
      GACGGGCTGT TGGTGATGGA CTCGTGGGTC AGGCGGGACT CGTTTCTGGG
      · N E K  R D H M  V L L  E F V  T A A G ·
1151  CAACGAGAAG CGCGATCACA TGGTCCTGCT GGAGTTCGTG ACCGCCGCCG
      GTTGCTCTTC GCGCTAGTGT ACCAGGACGA CCTCAAGCAC TGGCGGCGGC
                                          HindIII
                                          ~~~~~~~
      · I T L  G M D  E L Y K  * *
1201  GGATCACTCT CGGCATGGAC GAGCTGTACA AGTAATAAGC TT
      CCTAGTGAGA GCCGTACCTG CTCGACATGT TCATTATTCG AA
```

Figure 32B

Figure 33

SUMO-Protein G

Protein G NCBI Accession# X53324

```
      NcoI
      ~~~~~~
         M  G  H   H  H  H   H  H  G  S   D  S  E   V  N  Q
    1 CCATGGGTCA TCACCATCAT CATCACGGGT CGGACTCAGA AGTCAATCAA
      GGTACCCAGT AGTGGTAGTA GTAGTGCCCA GCCTGAGTCT TCAGTTAGTT
       E  A  K  P   E  V  K   P  E  V   K  P  E  T   H  I  N ·
   51 GAAGCTAAGC CAGAGGTCAA GCCAGAAGTC AAGCCTGAGA CTCACATCAA
      CTTCGATTCG GTCTCCAGTT CGGTCTTCAG TTCGGACTCT GAGTGTAGTT
      · L  K  V   S  D  G  S   S  E  I   F  F  K   I  K  K  T ·
  101 TTTAAAGGTG TCCGATGGAT CTTCAGAGAT CTTCTTCAAG ATCAAAAAGA
      AAATTTCCAC AGGCTACCTA GAAGTCTCTA GAAGAAGTTC TAGTTTTTCT
      · T  P  L   R  R  L   M  E  A  F   A  K  R   Q  G  K
  151 CCACTCCTTT AAGAAGGCTG ATGGAAGCGT TCGCTAAAAG ACAGGGTAAG
      GGTGAGGAAA TTCTTCCGAC TACCTTCGCA AGCGATTTTC TGTCCCATTC
       E  M  D  S   L  R  F   L  Y  D   G  I  R  I   Q  A  D ·
  201 GAAATGGACT CCTTAAGATT CTTGTACGAC GGTATTAGAA TTCAAGCTGA
      CTTTACCTGA GGAATTCTAA GAACATGCTG CCATAATCTT AAGTTCGACT
      · Q  T  P   E  D  L  D   M  E  D   N  D  I   I  E  A  H ·
  251 TCAGACCCCT GAAGATTTGG ACATGGAGGA TAACGATATT ATTGAGGCTC
      AGTCTGGGGA CTTCTAAACC TGTACCTCCT ATTGCTATAA TAACTCCGAG
      · R  E  Q   I  G  G   T  P  A  V   T  T  Y   K  L  V
  301 ACCGCGAACA GATTGGAGGT ACGCCGGCGG TGACCACCTA TAAACTGGTG
      TGGCGCTTGT CTAACCTCCA TGCGGCCGCC ACTGGTGGAT ATTTGACCAC
       I  N  G  K   T  L  K   G  E  T   T  T  K  A   V  D  A ·
  351 ATTAACGGCA AAACCCTGAA AGGCGAAACC ACCACCAAAG CGGTGGATGC
      TAATTGCCGT TTTGGGACTT TCCGCTTTGG TGGTGGTTTC GCCACCTACG
      · E  T  A   E  K  A  F   K  Q  Y   A  N  D   N  G  V  D ·
  401 GGAAACCGCG GAAAAAGCGT TTAAACAGTA TGCGAACGAT AACGGCGTGG
      CCTTTGGCGC CTTTTTCGCA AATTTGTCAT ACGCTTGCTA TTGCCGCACC
      · G  V  W   T  Y  D   D  A  T  K   T  F  T   V  T  E
  451 ATGGCGTGTG GACCTATGAT GATGCGACCA AAACCTTTAC CGTGACCGAA
      TACCGCACAC CTGGATACTA CTACGCTGGT TTTGGAAATG GCACTGGCTT
          HindIII
          ~~~~~~~~
       *  *
  501 TAATAAGCTT
      ATTATTCGAA
```

Figure 34A

SUMO β-GUS
β-GUS NCBI Accession# U12640

```
      M  G  H  H  H  H  H  H  G  S  D  S  E  V  N  Q  E ·
  1 ATGGGTCATC ACCATCATCA TCACGGGTCG GACTCAGAAG TCAATCAAGA
    TACCCAGTAG TGGTAGTAGT AGTGCCCAGC CTGAGTCTTC AGTTAGTTCT
    · A  K  P  E  V  K  P  E  V  K  P  E  T  H  I  N  L ·
 51 AGCTAAGCCA GAGGTCAAGC CAGAAGTCAA GCCTGAGACT CACATCAATT
    TCGATTCGGT CTCCAGTTCG GTCTTCAGTT CGGACTCTGA GTGTAGTTAA
    · K  V  S  D  G  S  S  E  I  F  F  K  I  K  K  T
101 TAAAGGTGTC CGATGGATCT TCAGAGATCT TCTTCAAGAT CAAAAAGACC
    ATTTCCACAG GCTACCTAGA AGTCTCTAGA AGAAGTTCTA GTTTTTCTGG
      T  P  L  R  R  L  M  E  A  F  A  K  R  Q  G  K  E ·
151 ACTCCTTTAA GAAGGCTGAT GGAAGCGTTC GCTAAAAGAC AGGGTAAGGA
    TGAGGAAATT CTTCCGACTA CCTTCGCAAG CGATTTTCTG TCCCATTCCT
    · M  D  S  L  R  F  L  Y  D  G  I  R  I  Q  A  D  Q ·
201 AATGGACTCC TTAAGATTCT TGTACGACGG TATTAGAATT CAAGCTGATC
    TTACCTGAGG AATTCTAAGA ACATGCTGCC ATAATCTTAA GTTCGACTAG
    · T  P  E  D  L  D  M  E  D  N  D  I  I  E  A  H
251 AGACCCCTGA AGATTTGGAC ATGGAGGATA ACGATATTAT TGAGGCTCAC
    TCTGGGGACT TCTAAACCTG TACCTCCTAT TGCTATAATA ACTCCGAGTG
      R  E  Q  I  G  G  M  E  F  M  L  R  P  V  E  T  P ·
301 CGCGAACAGA TTGGAGGTAT GGAATTCATG TTACGTCCTG TAGAAACCCC
    GCGCTTGTCT AACCTCCATA CCTTAAGTAC AATGCAGGAC ATCTTTGGGG
    · T  R  E  I  K  K  L  D  G  L  W  A  F  S  L  D  R ·
351 AACCCGTGAA ATCAAAAAAC TCGACGGCCT GTGGGCATTC AGTCTGGATC
    TTGGGCACTT TAGTTTTTTG AGCTGCCGGA CACCCGTAAG TCAGACCTAG
    · E  N  C  G  I  D  Q  R  W  W  E  S  A  L  Q  E
401 GCGAAAACTG TGGAATTGAT CAGCGTTGGT GGGAAAGCGC GTTACAAGAA
    CGCTTTTGAC ACCTTAACTA GTCGCAACCA CCCTTTCGCG CAATGTTCTT
      S  R  A  I  A  V  P  G  S  F  N  D  Q  F  A  D  A ·
451 AGCCGGGCAA TTGCTGTGCC AGGCAGTTTT AACGATCAGT TCGCCGATGC
    TCGGCCCGTT AACGACACGG TCCGTCAAAA TTGCTAGTCA AGCGGCTACG
    · D  I  R  N  Y  A  G  N  V  W  Y  Q  R  E  V  F  I ·
501 AGATATTCGT AATTATGCGG GCAACGTCTG GTATCAGCGC GAAGTCTTTA
    TCTATAAGCA TTAATACGCC CGTTGCAGAC CATAGTCGCG CTTCAGAAAT
    · P  K  G  W  A  G  Q  R  I  V  L  R  F  D  A  V
551 TACCGAAAGG TTGGGCAGGC CAGCGTATCG TGCTGCGTTT CGATGCGGTC
    ATGGCTTTCC AACCCGTCCG GTCGCATAGC ACGACGCAAA GCTACGCCAG
      T  H  Y  G  K  V  W  V  N  N  Q  E  V  M  E  H  Q ·
601 ACTCATTACG GCAAAGTGTG GGTCAATAAT CAGGAAGTGA TGGAGCATCA
    TGAGTAATGC CGTTTCACAC CCAGTTATTA GTCCTTCACT ACCTCGTAGT
    · G  G  Y  T  P  F  E  A  D  V  T  P  Y  V  I  A  G ·
651 GGGCGGCTAT ACGCCATTTG AAGCCGATGT CACGCCGTAT GTTATTGCCG
    CCCGCCGATA TGCGGTAAAC TTCGGCTACA GTGCGGCATA CAATAACGGC
    · K  S  V  R  I  T  V  C  V  N  N  E  L  N  W  Q
```

Figure 34B

```
701  GGAAAAGTGT ACGTATCACC GTTTGTGTGA ACAACGAACT GAACTGGCAG
     CCTTTTCACA TGCATAGTGG CAAACACACT TGTTGCTTGA CTTGACCGTC
      T  I  P  P  G  M  V  I  T  D  E  N  G  K  K  K  Q ·
751  ACTATCCCGC CGGGAATGGT GATTACCGAC GAAAACGGCA AGAAAAAGCA
     TGATAGGGCG GCCCTTACCA CTAATGGCTG CTTTTGCCGT TCTTTTTCGT
     · S  Y  F  H  D  F  F  N  Y  A  G  I  H  R  S  V  M ·
801  GTCTTACTTC CATGATTTCT TTAACTATGC CGGAATCCAT CGCAGCGTAA
     CAGAATGAAG GTACTAAAGA AATTGATACG GCCTTAGGTA GCGTCGCATT
     · L  Y  T  T  P  N  T  W  V  D  D  I  T  V  V  T
851  TGCTCTACAC CACGCCGAAC ACCTGGGTGG ACGATATCAC CGTGGTGACG
     ACGAGATGTG GTGCGGCTTG TGGACCCACC TGCTATAGTG GCACCACTGC
      H  V  A  Q  D  C  N  H  A  S  V  D  W  Q  V  V  A ·
901  CATGTCGCGC AAGACTGTAA CCACGCGTCT GTTGACTGGC AGGTGGTGGC
     GTACAGCGCG TTCTGACATT GGTGCGCAGA CAACTGACCG TCCACCACCG
     · N  G  D  V  S  V  E  L  R  D  A  D  Q  Q  V  V  A ·
951  CAATGGTGAT GTCAGCGTTG AACTGCGTGA TGCGGATCAA CAGGTGGTTG
     GTTACCACTA CAGTCGCAAC TTGACGCACT ACGCCTAGTT GTCCACCAAC
     · T  G  Q  G  T  S  G  T  L  Q  V  V  N  P  H  L
1001 CAACTGGACA AGGCACTAGC GGGACTTTGC AAGTGGTGAA TCCGCACCTC
     GTTGACCTGT TCCGTGATCG CCCTGAAACG TTCACCACTT AGGCGTGGAG
      W  Q  P  G  E  G  Y  L  Y  E  L  C  V  T  A  K  S ·
1051 TGGCAACCGG GTGAAGGTTA TCTCTATGAA CTGTGCGTCA CAGCCAAAAG
     ACCGTTGGCC CACTTCCAAT AGAGATACTT GACACGCAGT GTCGGTTTTC
     · Q  T  E  C  D  I  Y  P  L  R  V  G  I  R  S  V  A ·
1101 CCAGACAGAG TGTGATATCT ACCCGCTTCG CGTCGGCATC CGGTCAGTGG
     GGTCTGTCTC ACACTATAGA TGGGCGAAGC GCAGCCGTAG GCCAGTCACC
     · V  K  G  Q  Q  F  L  I  N  H  K  P  F  Y  F  T
1151 CAGTGAAGGG CCAACAGTTC CTGATTAACC ACAAACCGTT CTACTTTACT
     GTCACTTCCC GGTTGTCAAG GACTAATTGG TGTTTGGCAA GATGAAATGA
      G  F  G  R  H  E  D  A  D  L  R  G  K  G  F  D  N ·
1201 GGCTTTGGTC GTCATGAAGA TGCGGACTTA CGTGGCAAAG GATTCGATAA
     CCGAAACCAG CAGTACTTCT ACGCCTGAAT GCACCGTTTC CTAAGCTATT
     · V  L  M  V  H  D  H  A  L  M  D  W  I  G  A  N  S ·
1251 CGTGCTGATG GTGCACGACC ACGCATTAAT GGACTGGATT GGGGCCAACT
     GCACGACTAC CACGTGCTGG TGCGTAATTA CCTGACCTAA CCCCGGTTGA
     · Y  R  T  S  H  Y  P  Y  A  E  E  M  L  D  W  A
1301 CCTACCGTAC CTCGCATTAC CCTTACGCTG AAGAGATGCT CGACTGGGCA
     GGATGGCATG GAGCGTAATG GGAATGCGAC TTCTCTACGA GCTGACCCGT
      D  E  H  G  I  V  V  I  D  E  T  A  A  V  G  F  N ·
1351 GATGAACATG GCATCGTGGT GATTGATGAA ACTGCTGCTG TCGGCTTTAA
     CTACTTGTAC CGTAGCACCA CTAACTACTT TGACGACGAC AGCCGAAATT
     · L  S  L  G  I  G  F  E  A  G  N  K  P  K  E  L  Y ·
1401 CCTCTCTTTA GGCATTGGTT TCGAAGCGGG CAACAAGCCG AAAGAACTGT
     GGAGAGAAAT CCGTAACCAA AGCTTCGCCC GTTGTTCGGC TTTCTTGACA
     · S  E  E  A  V  N  G  E  T  Q  Q  A  H  L  Q  A
1451 ACAGCGAAGA GGCAGTCAAC GGGGAAACTC AGCAAGCGCA CTTACAGGCG
     TGTCGCTTCT CCGTCAGTTG CCCCTTTGAG TCGTTCGCGT GAATGTCCGC
      I  K  E  L  I  A  R  D  K  N  H  P  S  V  V  M  W ·
1501 ATTAAAGAGC TGATAGCGCG TGACAAAAAC CACCCAAGCG TGGTGATGTG
     TAATTTCTCG ACTATCGCGC ACTGTTTTTG GTGGGTTCGC ACCACTACAC
     · S  I  A  N  E  P  D  T  R  P  Q  V  H  G  N  I  S ·
1551 GAGTATTGCC AACGAACCGG ATACCCGTCC GCAAGTGCAC GGGAATATTT
```

```
        CTCATAACGG TTGCTTGGCC TATGGGCAGG CGTTCACGTG CCCTTATAAA
        · P L A  E A T  R K L D  P T R  P I T
1601    CGCCACTGGC GGAAGCAACG CGTAAACTCG ACCCGACGCG TCCGATCACC
        GCGGTGACCG CCTTCGTTGC GCATTTGAGC TGGGCTGCGC AGGCTAGTGG

         C V N V  M F C  D A H  T D T I  S D L ·
1651    TGCGTCAATG TAATGTTCTG CGACGCTCAC ACCGATACCA TCAGCGATCT
        ACGCAGTTAC ATTACAAGAC GCTGCGAGTG TGGCTATGGT AGTCGCTAGA
        · F D V  L C L N  R Y Y  G W Y  V Q S G ·
1701    CTTTGATGTG CTGTGCCTGA ACCGTTATTA CGGATGGTAT GTCCAAAGCG
        GAAACTACAC GACACGGACT TGGCAATAAT GCCTACCATA CAGGTTTCGC
        · D L E  T A E  K V L E  K E L  L A W
1751    GCGATTTGGA AACGGCAGAG AAGGTACTGG AAAAAGAACT TCTGGCCTGG
        CGCTAAACCT TTGCCGTCTC TTCCATGACC TTTTTCTTGA AGACCGGACC
         Q E K L  H Q P  I I I  T E Y G  V D T ·
1801    CAGGAGAAAC TGCATCAGCC GATTATCATC ACCGAATACG GCGTGGATAC
        GTCCTCTTTG ACGTAGTCGG CTAATAGTAG TGGCTTATGC CGCACCTATG
        · L A G  L H S M  Y T D  M W S  E E Y Q ·
1851    GTTAGCCGGG CTGCACTCAA TGTACACCGA CATGTGGAGT GAAGAGTATC
        CAATCGGCCC GACGTGAGTT ACATGTGGCT GTACACCTCA CTTCTCATAG
        · C A W  L D M  Y H R V  F D R  V S A
1901    AGTGTGCATG GCTGGATATG TATCACCGCG TCTTTGATCG CGTCAGCGCC
        TCACACGTAC CGACCTATAC ATAGTGGCGC AGAAACTAGC GCAGTCGCGG
         V V G E  Q V W  N F A  D F A T  S Q G ·
1951    GTCGTCGGTG AACAGGTATG GAATTTCGCC GATTTTGCGA CCTCGCAAGG
        CAGCAGCCAC TTGTCCATAC CTTAAAGCGG CTAAAACGCT GGAGCGTTCC
        · I L R  V G G N  K K G  I F T  R D R K ·
2001    CATATTGCGC GTTGGCGGTA ACAAGAAAGG GATCTTCACT CGCGACCGCA
        GTATAACGCG CAACCGCCAT TGTTCTTTCC CTAGAAGTGA GCGCTGGCGT
        · P K S  A A F  L L Q K  R W T  G M N
2051    AACCGAAGTC GGCGGCTTTT CTGCTGCAAA AACGCTGGAC TGGCATGAAC
        TTGGCTTCAG CCGCCGAAAA GACGACGTTT TTGCGACCTG ACCGTACTTG
         F G E K  P Q Q  G G K  Q
2101    TTCGGTGAAA AACCGCAGCA GGGAGGCAAA CAA
        AAGCCACTTT TTGGCGTCGT CCCTCCGTTT GTT
```

Figure 34C

Figure 35A

SUMO-Liver X Receptor α
Liver X Receptor A NCBI Accession# NM_005693

```
      M  G  H  H  H  H  H  H  G  S  D  S  E  V  N  Q  E ·
  1 ATGGGTCATC ACCATCATCA TCACGGGTCG GACTCAGAAG TCAATCAAGA
    TACCCAGTAG TGGTAGTAGT AGTGCCCAGC CTGAGTCTTC AGTTAGTTCT
    ·  A  K  P  E  V  K  P  E  V  K  P  E  T  H  I  N  L ·
 51 AGCTAAGCCA GAGGTCAAGC CAGAAGTCAA GCCTGAGACT CACATCAATT
    TCGATTCGGT CTCCAGTTCG GTCTTCAGTT CGGACTCTGA GTGTAGTTAA
    ·  K  V  S  D  G  S  S  E  I  F  F  K  I  K  K  T
101 TAAAGGTGTC CGATGGATCT TCAGAGATCT TCTTCAAGAT CAAAAAGACC
    ATTTCCACAG GCTACCTAGA AGTCTCTAGA AGAAGTTCTA GTTTTTCTGG
      T  P  L  R  R  L  M  E  A  F  A  K  R  Q  G  K  E ·
151 ACTCCTTTAA GAAGGCTGAT GGAAGCGTTC GCTAAAAGAC AGGGTAAGGA
    TGAGGAAATT CTTCCGACTA CCTTCGCAAG CGATTTTCTG TCCCATTCCT
    ·  M  D  S  L  R  F  L  Y  D  G  I  R  I  Q  A  D  Q ·
201 AATGGACTCC TTAAGATTCT TGTACGACGG TATTAGAATT CAAGCTGATC
    TTACCTGAGG AATTCTAAGA ACATGCTGCC ATAATCTTAA GTTCGACTAG
    ·  T  P  E  D  L  D  M  E  D  N  D  I  I  E  A  H
251 AGACCCCTGA AGATTTGGAC ATGGAGGATA ACGATATTAT TGAGGCTCAC
    TCTGGGGACT TCTAAACCTG TACCTCCTAT TGCTATAATA ACTCCGAGTG
      R  E  Q  I  G  G  M  S  L  W  L  G  A  P  V  P  D ·
301 CGCGAACAGA TTGGAGGTAT GTCCTTGTGG CTGGGGGCCC CTGTGCCTGA
    GCGCTTGTCT AACCTCCATA CAGGAACACC GACCCCCGGG GACACGGACT
    ·  I  P  P  D  S  A  V  E  L  W  K  P  G  A  Q  D  A ·
351 CATTCCTCCT GACTCTGCGG TGGAGCTGTG GAAGCCAGGC GCACAGGATG
    GTAAGGAGGA CTGAGACGCC ACCTCGACAC CTTCGGTCCG CGTGTCCTAC
    ·  S  S  Q  A  Q  G  G  S  S  C  I  L  R  E  E  A
401 CAAGCAGCCA GGCCCAGGGA GGCAGCAGCT GCATCCTCAG AGAGGAAGCC
    GTTCGTCGGT CCGGGTCCCT CCGTCGTCGA CGTAGGAGTC TCTCCTTCGG
      R  M  P  H  S  A  G  G  T  A  G  V  G  L  E  A  A ·
451 AGGATGCCCC ACTCTGCTGG GGGTACTGCA GGGGTGGGGC TGGAGGCTGC
    TCCTACGGGG TGAGACGACC CCCATGACGT CCCCACCCCG ACCTCCGACG
    ·  E  P  T  A  L  L  T  R  A  E  P  P  S  E  P  T  E ·
501 AGAGCCCACA GCCCTGCTCA CCAGGGCAGA GCCCCCTTCA GAACCCACAG
    TCTCGGGTGT CGGGACGAGT GGTCCCGTCT CGGGGGAAGT CTTGGGTGTC
    ·  I  R  P  Q  K  R  K  K  G  P  A  P  K  M  L  G
551 AGATCCGTCC ACAAAAGCGG AAAAAGGGGC CAGCCCCCAA AATGCTGGGG
    TCTAGGCAGG TGTTTTCGCC TTTTTCCCCG GTCGGGGGTT TTACGACCCC
      N  E  L  C  S  V  C  G  D  K  A  S  G  F  H  Y  N ·
601 AACGAGCTAT GCAGCGTGTG TGGGGACAAG GCCTCGGGCT TCCACTACAA
    TTGCTCGATA CGTCGCACAC ACCCCTGTTC CGGAGCCCGA AGGTGATGTT
    ·  V  L  S  C  E  G  C  K  G  F  F  R  R  S  V  I  K ·
651 TGTTCTGAGC TGCGAGGGCT GCAAGGGATT CTTCCGCCGC AGCGTCATCA
    ACAAGACTCG ACGCTCCCGA CGTTCCCTAA GAAGGCGGCG TCGCAGTAGT
    ·  G  A  H  Y  I  C  H  S  G  G  H  C  P  M  D  T
701 AGGGAGCGCA CTACATCTGC CACAGTGGCG GCCACTGCCC CATGGACACC
```

Figure 35B

```
      TCCCTCGCGT GATGTAGACG GTGTCACCGC CGGTGACGGG GTACCTGTGG
       Y  M  R  R  K  C  Q  E  C  R  L  R  K  C  R  Q  A ·
751   TACATGCGTC GCAAGTGCCA GGAGTGTCGG CTTCGCAAAT GCCGTCAGGC
      ATGTACGCAG CGTTCACGGT CCTCACAGCC GAAGCGTTTA CGGCAGTCCG
      · G  M  R  E  E  C  V  L  S  E  E  Q  I  R  L  K  K ·
801   TGGCATGCGG GAGGAGTGTG TCCTGTCAGA AGAACAGATC CGCCTGAAGA
      ACCGTACGCC CTCCTCACAC AGGACAGTCT TCTTGTCTAG GCGGACTTCT
      · L  K  R  Q  E  E  E  Q  A  H  A  T  S  L  P  P
851   AACTGAAGCG GCAAGAGGAG GAACAGGCTC ATGCCACATC CTTGCCCCCC
      TTGACTTCGC CGTTCTCCTC CTTGTCCGAG TACGGTGTAG GAACGGGGGG
       R  R  S  S  P  P  Q  I  L  P  Q  L  S  P  E  Q  L ·
901   AGGCGTTCCT CACCCCCCCA AATCCTGCCC CAGCTCAGCC CGGAACAACT
      TCCGCAAGGA GTGGGGGGGT TTAGGACGGG GTCGAGTCGG GCCTTGTTGA
      · G  M  I  E  K  L  V  A  A  Q  Q  Q  C  N  R  R  S ·
951   GGGCATGATC GAGAAGCTCG TCGCTGCCCA GCAACAGTGT AACCGGCGCT
      CCCGTACTAG CTCTTCGAGC AGCGACGGGT CGTTGTCACA TTGGCCGCGA
      · F  S  D  R  L  R  V  T  P  W  P  M  A  P  D  P
1001  CCTTTTCTGA CCGGCTTCGA GTCACGCCTT GGCCCATGGC ACCAGATCCC
      GGAAAAGACT GGCCGAAGCT CAGTGCGGAA CCGGGTACCG TGGTCTAGGG
       H  S  R  E  A  R  Q  Q  R  F  A  H  F  T  E  L  A ·
1051  CATAGCCGGG AGGCCCGTCA GCAGCGCTTT GCCCACTTCA CTGAGCTGGC
      GTATCGGCCC TCCGGGCAGT CGTCGCGAAA CGGGTGAAGT GACTCGACCG
      · I  V  S  V  Q  E  I  V  D  F  A  K  Q  L  P  G  F ·
1101  CATCGTCTCT GTGCAGGAGA TAGTTGACTT TGCTAAACAG CTACCCGGCT
      GTAGCAGAGA CACGTCCTCT ATCAACTGAA ACGATTTGTC GATGGGCCGA

      · L  Q  L  S  R  E  D  Q  I  A  L  L  K  T  S  A
1151  TCCTGCAGCT CAGCCGGGAG GACCAGATTG CCCTGCTGAA GACCTCTGCG
      AGGACGTCGA GTCGGCCCTC CTGGTCTAAC GGGACGACTT CTGGAGACGC
       I  E  V  M  L  L  E  T  S  R  R  Y  N  P  G  S  E ·
1201  ATCGAGGTGA TGCTTCTGGA GACATCTCGG AGGTACAACC CTGGGAGTGA
      TAGCTCCACT ACGAAGACCT CTGTAGAGCC TCCATGTTGG GACCCTCACT
      · S  I  T  F  L  K  D  F  S  Y  N  R  E  D  F  A  K ·
1251  GAGTATCACC TTCCTCAAGG ATTTCAGTTA TAACCGGGAA GACTTTGCCA
      CTCATAGTGG AAGGAGTTCC TAAAGTCAAT ATTGGCCCTT CTGAAACGGT
      · A  G  L  Q  V  E  F  I  N  P  I  F  E  F  S  R
1301  AAGCAGGGCT GCAAGTGGAA TTCATCAACC CCATCTTCGA GTTCTCCAGG
      TTCGTCCCGA CGTTCACCTT AAGTAGTTGG GGTAGAAGCT CAAGAGGTCC
       A  M  N  E  L  Q  L  N  D  A  E  F  A  L  L  I  A ·
1351  GCCATGAATG AGCTGCAACT CAATGATGCC GAGTTTGCCT TGCTCATTGC
      CGGTACTTAC TCGACGTTGA GTTACTACGG CTCAAACGGA ACGAGTAACG
      · I  S  I  F  S  A  D  R  P  N  V  Q  D  Q  L  Q  V ·
1401  TATCAGCATC TTCTCTGCAG ACCGGCCCAA CGTGCAGGAC CAGCTCCAGG
      ATAGTCGTAG AAGAGACGTC TGGCCGGGTT GCACGTCCTG GTCGAGGTCC
      · E  R  L  Q  H  T  Y  V  E  A  L  H  A  Y  V  S
1451  TGGAGAGGCT GCAGCACACA TATGTGGAAG CCCTGCATGC CTACGTCTCC
      ACCTCTCCGA CGTCGTGTGT ATACACCTTC GGGACGTACG GATGCAGAGG
       I  H  H  P  H  D  R  L  M  F  P  R  M  L  M  K  L ·
1501  ATCCACCATC CCCATGACCG ACTGATGTTC CCACGGATGC TAATGAAACT
      TAGGTGGTAG GGGTACTGGC TGACTACAAG GGTGCCTACG ATTACTTTGA
      · V  S  L  R  T  L  S  S  V  H  S  E  Q  V  F  A  L ·
1551  GGTGAGCCTC CGGACCCTGA GCAGCGTCCA CTCAGAGCAA GTGTTTGCAC
```

```
      CCACTCGGAG GCCTGGGACT CGTCGCAGGT GAGTCTCGTT CACAAACGTG
      · R L Q  D K K  L P P L  L S E  I W D
1601  TGCGTCTGCA GGACAAAAAG CTCCCACCGC TGCTCTCTGA GATCTGGGAT
      ACGCAGACGT CCTGTTTTTC GAGGGTGGCG ACGAGAGACT CTAGACCCTA
       V H E *
1651  GTGCACGAAT GA
      CACGTGCTTA CT
```

Figure 35C

Figure 36A

SUMO Tyrosine Kinase
Tyrosin Kinase NCBI Accession# BC039039

```
      M  G  H  H  H  H  H  H  G  S  D  S  E  V  N  Q  E ·
  1 ATGGGTCATC ACCATCATCA TCACGGGTCG GACTCAGAAG TCAATCAAGA
    TACCCAGTAG TGGTAGTAGT AGTGCCCAGC CTGAGTCTTC AGTTAGTTCT
    · A  K  P  E  V  K  P  E  V  K  P  E  T  H  I  N  L ·
 51 AGCTAAGCCA GAGGTCAAGC CAGAAGTCAA GCCTGAGACT CACATCAATT
    TCGATTCGGT CTCCAGTTCG GTCTTCAGTT CGGACTCTGA GTGTAGTTAA
    · K  V  S  D  G  S  S  E  I  F  F  K  I  K  K  T
101 TAAAGGTGTC CGATGGATCT TCAGAGATCT TCTTCAAGAT CAAAAAGACC
    ATTTCCACAG GCTACCTAGA AGTCTCTAGA AGAAGTTCTA GTTTTTCTGG
      T  P  L  R  R  L  M  E  A  F  A  K  R  Q  G  K  E ·
151 ACTCCTTTAA GAAGGCTGAT GGAAGCGTTC GCTAAAAGAC AGGGTAAGGA
    TGAGGAAATT CTTCCGACTA CCTTCGCAAG CGATTTTCTG TCCCATTCCT
    · M  D  S  L  R  F  L  Y  D  G  I  R  I  Q  A  D  Q ·
201 AATGGACTCC TTAAGATTCT TGTACGACGG TATTAGAATT CAAGCTGATC
    TTACCTGAGG AATTCTAAGA ACATGCTGCC ATAATCTTAA GTTCGACTAG
    · T  P  E  D  L  D  M  E  D  N  D  I  I  E  A  H
251 AGACCCCTGA AGATTTGGAC ATGGAGGATA ACGATATTAT TGAGGCTCAC
    TCTGGGGACT TCTAAACCTG TACCTCCTAT TGCTATAATA ACTCCGAGTG
      R  E  Q  I  G  G  M  C  P  N  S  S  A  S  N  A  S ·
301 CGCGAACAGA TTGGAGGTAT GTGCCCCAAC AGCAGTGCCA GCAACGCCTC
    GCGCTTGTCT AACCTCCATA CACGGGGTTG TCGTCACGGT CGTTGCGGAG
    · G  A  A  A  P  T  L  P  A  H  P  S  T  L  T  H  P ·
351 AGGGGCTGCT GCTCCCACAC TCCCAGCCCA CCCATCCACG TTGACTCATC
    TCCCCGACGA CGAGGGTGTG AGGGTCGGGT GGGTAGGTGC AACTGAGTAG
    · Q  R  R  I  D  T  L  N  S  D  G  Y  T  P  E  P
401 CTCAGAGACG AATCGACACC CTCAACTCAG ATGGATACAC CCCTGAGCCA
    GAGTCTCTGC TTAGCTGTGG GAGTTGAGTC TACCTATGTG GGGACTCGGT
      A  R  I  T  S  P  D  K  P  R  P  M  P  M  D  T  S ·
451 GCACGCATAA CGTCCCCAGA CAAACCGCGG CCGATGCCCA TGGACACGAG
    CGTGCGTATT GCAGGGGTCT GTTTGGCGCC GGCTACGGGT ACCTGTGCTC
    · V  Y  E  S  P  Y  S  D  P  E  E  L  K  D  K  K  L ·
501 CGTGTATGAG AGCCCCTACA GCGACCCAGA GGAGCTCAAG GACAAGAAGC
    GCACATACTC TCGGGGATGT CGCTGGGTCT CCTCGAGTTC CTGTTCTTCG
    · F  L  K  R  D  N  L  L  I  A  D  I  E  L  G  C
551 TCTTCCTGAA GCGCGATAAC CTCCTCATAG CTGACATTGA ACTTGGCTGC
    AGAAGGACTT CGCGCTATTG GAGGAGTATC GACTGTAACT TGAACCGACG
      G  N  F  G  S  V  R  Q  G  V  Y  R  M  R  K  K  Q ·
601 GGCAACTTTG GCTCAGTGCG CCAGGGCGTG TACCGCATGC GCAAGAAGCA
    CCGTTGAAAC CGAGTCACGC GGTCCCGCAC ATGGCGTACG CGTTCTTCGT
    · I  D  V  A  I  K  V  L  K  Q  G  T  E  K  A  D  T ·
651 GATCGACGTG GCCATCAAGG TGCTGAAGCA GGGCACGGAG AAGGCAGACA
    CTAGCTGCAC CGGTAGTTCC ACGACTTCGT CCCGTGCCTC TTCCGTCTGT
    · E  E  M  M  R  E  A  Q  I  M  H  Q  L  D  N  P
701 CGGAAGAGAT GATGCGCGAG GCGCAGATCA TGCACCAGCT GGACAACCCC
    GCCTTCTCTA CTACGCGCTC CGCGTCTAGT ACGTGGTCGA CCTGTTGGGG
      Y  I  V  R  L  I  G  V  C  Q  A  E  A  L  M  L  V ·
751 TACATCGTGC GGCTCATTGG CGTCTGCCAG GCCGAGGCCC TCATGCTGGT
```

```
      ATGTAGCACG CCGAGTAACC GCAGACGGTC CGGCTCCGGG AGTACGACCA
      · M  E  M   A  G  G  G   P  L  H   K  F  L   V  G  K  R ·
 801  CATGGAGATG GCTGGGGGCG GGCCGCTGCA CAAGTTCCTG GTCGGCAAGA
      GTACCTCTAC CGACCCCCGC CCGGCGACGT GTTCAAGGAC CAGCCGTTCT
      · E  E  I   P  V  S   N  V  A  E   L  L  H   Q  V  S
 851  GGGAGGAGAT CCCTGTGAGC AATGTGGCCG AGCTGCTGCA CCAGGTGTCC
      CCCTCCTCTA GGGACACTCG TTACACCGGC TCGACGACGT GGTCCACAGG
       M  G  M  K   Y  L  E   E  K  N   F  V  H  R   D  L  A ·
 901  ATGGGGATGA AGTACCTGGA GGAGAAGAAC TTTGTGCACC GTGACCTGGC
      TACCCCTACT TCATGGACCT CCTCTTCTTG AAACACGTGG CACTGGACCG
      · A  R  N   V  L  L  V   N  R  H   Y  A  K   I  S  D  F ·
 951  GGCCCGCAAC GTCCTGCTGG TTAACCGGCA CTACGCCAAG ATCAGCGACT
      CCGGGCGTTG CAGGACGACC AATTGGCCGT GATGCGGTTC TAGTCGCTGA
      · G  L  S   K  A  L   G  A  D  D   S  Y  Y   T  A  R
1001  TTGGCCTCTC CAAAGCACTG GGTGCCGACG ACAGCTACTA CACTGCCCGC
      AACCGGAGAG GTTTCGTGAC CCACGGCTGC TGTCGATGAT GTGACGGGCG
       S  A  G  K   W  P  L   K  W  Y   A  P  E  C   I  N  F ·
1051  TCAGCAGGGA AGTGGCCGCT CAAGTGGTAC GCACCCGAAT GCATCAACTT
      AGTCGTCCCT TCACCGGCGA GTTCACCATG CGTGGGCTTA CGTAGTTGAA

      · R  K  F   S  S  R  S   D  V  W   S  Y  G   V  T  M  W ·
1101  CCGCAAGTTC TCCAGCCGCA GCGATGTCTG GAGCTATGGG GTCACCATGT
      GGCGTTCAAG AGGTCGGCGT CGCTACAGAC CTCGATACCC CAGTGGTACA
      · E  A  L   S  Y  G   Q  K  P  Y   K  K  M   K  G  P
1151  GGGAGGCCTT GTCCTACGGC CAGAAGCCCT ACAAGAAGAT GAAAGGGCCG
      CCCTCCGGAA CAGGATGCCG GTCTTCGGGA TGTTCTTCTA CTTTCCCGGC
       E  V  M  A   F  I  E   Q  G  K   R  M  E  C   P  P  E ·
1201  GAGGTCATGG CCTTCATCGA GCAGGGCAAG CGGATGGAGT GCCCACCAGA
      CTCCAGTACC GGAAGTAGCT CGTCCCGTTC GCCTACCTCA CGGGTGGTCT
      · C  P  P   E  L  Y  A   L  M  S   D  C  W   I  Y  K  W ·
1251  GTGTCCACCC GAACTGTACG CACTCATGAG TGACTGCTGG ATCTACAAGT
      CACAGGTGGG CTTGACATGC GTGAGTACTC ACTGACGACC TAGATGTTCA
      · E  D  R   P  D  F   L  T  V  E   Q  R  M   R  A  C
1301  GGGAGGATCG CCCCGACTTC CTGACCGTGG AGCAGCGCAT GCGAGCCTGT
      CCCTCCTAGC GGGGCTGAAG GACTGGCACC TCGTCGCGTA CGCTCGGACA
       Y  Y  S  L   A  S  K   V  E  G   P  P  G  S   T  Q  K ·
1351  TACTACAGCC TGGCCAGCAA GGTGGAAGGG CCCCCAGGCA GCACACAGAA
      ATGATGTCGG ACCGGTCGTT CCACCTTCCC GGGGGTCCGT CGTGTGTCTT
      · A  E  A   A  C  A  *
1401  GGCTGAGGCT GCCTGTGCCT GA
      CCGACTCCGA CGGACACGGA CT
```

Figure 36B

Figure 37A

SUMO MAPKAPK2 Kinase

MAPKAPK2 Kinase NCBI Accession# BC036060

```
      M  G  H  H  H  H  H  H  G  S  D  S  E  V  N  Q  E
   1  ATGGGTCATC ACCATCATCA TCACGGGTCG GACTCAGAAG TCAATCAAGA
      TACCCAGTAG TGGTAGTAGT AGTGCCCAGC CTGAGTCTTC AGTTAGTTCT
      A  K  P  E  V  K  P  E  V  K  P  E  T  H  I  N  L
  51  AGCTAAGCCA GAGGTCAAGC CAGAAGTCAA GCCTGAGACT CACATCAATT
      TCGATTCGGT CTCCAGTTCG GTCTTCAGTT CGGACTCTGA GTGTAGTTAA
      K  V  S  D  G  S  S  E  I  F  F  K  I  K  K  T
 101  TAAAGGTGTC CGATGGATCT TCAGAGATCT TCTTCAAGAT CAAAAAGACC
      ATTTCCACAG GCTACCTAGA AGTCTCTAGA AGAAGTTCTA GTTTTTCTGG
      T  P  L  R  R  L  M  E  A  F  A  K  R  Q  G  K  E
 151  ACTCCTTTAA GAAGGCTGAT GGAAGCGTTC GCTAAAAGAC AGGGTAAGGA
      TGAGGAAATT CTTCCGACTA CCTTCGCAAG CGATTTTCTG TCCCATTCCT
      M  D  S  L  R  F  L  Y  D  G  I  R  I  Q  A  D  Q
 201  AATGGACTCC TTAAGATTCT TGTACGACGG TATTAGAATT CAAGCTGATC
      TTACCTGAGG AATTCTAAGA ACATGCTGCC ATAATCTTAA GTTCGACTAG
      T  P  E  D  L  D  M  E  D  N  D  I  I  E  A  H
 251  AGACCCCTGA AGATTTGGAC ATGGAGGATA ACGATATTAT TGAGGCTCAC
      TCTGGGGACT TCTAAACCTG TACCTCCTAT TGCTATAATA ACTCCGAGTG
      R  E  Q  I  G  G  M  Q  F  H  V  K  S  G  L  Q  I
 301  CGCGAACAGA TTGGAGGTAT GCAGTTCCAC GTCAAGTCCG GCCTGCAGAT
      GCGCTTGTCT AACCTCCATA CGTCAAGGTG CAGTTCAGGC CGGACGTCTA
      K  K  N  A  I  I  D  D  Y  K  V  T  S  Q  V  L  G
 351  CAAGAAGAAC GCCATCATCG ATGACTACAA GGTCACCAGC CAGGTCCTGG
      GTTCTTCTTG CGGTAGTAGC TACTGATGTT CCAGTGGTCG GTCCAGGACC
      L  G  I  N  G  K  V  L  Q  I  F  N  K  R  T  Q
 401  GGCTGGGCAT CAACGGCAAA GTTTTGCAGA TCTTCAACAA GAGGACCCAG
      CCGACCCGTA GTTGCCGTTT CAAAACGTCT AGAAGTTGTT CTCCTGGGTC
      E  K  F  A  L  K  M  L  Q  D  C  P  K  A  R  R  E
 451  GAGAAATTCG CCCTCAAAAT GCTTCAGGAC TGCCCCAAGG CCCGCAGGGA
      CTCTTTAAGC GGGAGTTTTA CGAAGTCCTG ACGGGGTTCC GGGCGTCCCT
      V  E  L  H  W  R  A  S  Q  C  P  H  I  V  R  I  V
 501  GGTGGAGCTG CACTGGCGGG CCTCCCAGTG CCCGCACATC GTACGGATCG
      CCACCTCGAC GTGACCGCCC GGAGGGTCAC GGGCGTGTAG CATGCCTAGC
      D  V  Y  E  N  L  Y  A  G  R  K  C  L  L  I  V
 551  TGGATGTGTA CGAGAATCTG TACGCAGGGA GGAAGTGCCT GCTGATTGTC
      ACCTACACAT GCTCTTAGAC ATGCGTCCCT CCTTCACGGA CGACTAACAG
      M  E  C  L  D  G  G  E  L  F  S  R  I  Q  D  R  G
 601  ATGGAATGTT TGGACGGTGG AGAACTCTTT AGCCGAATCC AGGATCGAGG
      TACCTTACAA ACCTGCCACC TCTTGAGAAA TCGGCTTAGG TCCTAGCTCC
      D  Q  A  F  T  E  R  E  A  S  E  I  M  K  S  I  G
 651  AGACCAGGCA TTCACAGAAA GAGAAGCATC CGAAATCATG AAGAGCATCG
      TCTGGTCCGT AAGTGTCTTT CTCTTCGTAG GCTTTAGTAC TTCTCGTAGC
      E  A  I  Q  Y  L  H  S  I  N  I  A  H  R  D  V
 701  GTGAGGCCAT CCAGTATCTG CATTCAATCA ACATTGCCCA TCGGGATGTC
      CACTCCGGTA GGTCATAGAC GTAAGTTAGT TGTAACGGGT AGCCCTACAG
      K  P  E  N  L  L  Y  T  S  K  R  P  N  A  I  L  K
```

```
751  AAGCCTGAGA ATCTCTTATA CACCTCCAAA AGGCCCAACG CCATCCTGAA
     TTCGGACTCT TAGAGAATAT GTGGAGGTTT TCCGGGTTGC GGTAGGACTT
     · L  T  D   F  G  F  A   K  E  T   T  S  H   N  S  L  T ·
801  ACTCACTGAC TTTGGCTTTG CCAAGGAAAC CACCAGCCAC AACTCTTTGA
     TGAGTGACTG AAACCGAAAC GGTTCCTTTG GTGGTCGGTG TTGAGAAACT
     · T  P  C   Y  T  P   Y  Y  V  A   P  E  V   L  G  P
851  CCACTCCTTG TTATACACCG TACTATGTGG CTCCAGAAGT GCTGGGTCCA
     GGTGAGGAAC AATATGTGGC ATGATACACC GAGGTCTTCA CGACCCAGGT
      E  K  Y  D   K  S  C   D  M  W   S  L  G  V   I  M  Y ·
901  GAGAAGTATG ACAAGTCCTG TGACATGTGG TCCCTGGGTG TCATCATGTA
     CTCTTCATAC TGTTCAGGAC ACTGTACACC AGGGACCCAC AGTAGTACAT
     · I  L  L   C  G  Y  P   P  F  Y   S  N  H   G  L  A  I ·
951  CATCCTGCTG TGTGGGTATC CCCCCTTCTA CTCCAACCAC GGCCTTGCCA
     GTAGGACGAC ACACCCATAG GGGGGAAGAT GAGGTTGGTG CCGGAACGGT
     · S  P  G   M  K  T   R  I  R  M   G  Q  Y   E  F  P
1001 TCTCTCCGGG CATGAAGACT CGCATCCGAA TGGGCCAGTA TGAATTTCCC
     AGAGAGGCCC GTACTTCTGA GCGTAGGCTT ACCCGGTCAT ACTTAAAGGG
      N  P  E  W   S  E  V   S  E  E   V  K  M  L   I  R  N ·
1051 AACCCAGAAT GGTCAGAAGT ATCAGAGGAA GTGAAGATGC TCATTCGGAA
     TTGGGTCTTA CCAGTCTTCA TAGTCTCCTT CACTTCTACG AGTAAGCCTT
     · L  L  K   T  E  P  T   Q  R  M   T  I  T   E  F  M  N ·
1101 TCTGCTGAAA ACAGAGCCCA CCCAGAGAAT GACCATCACC GAGTTTATGA
     AGACGACTTT TGTCTCGGGT GGGTCTCTTA CTGGTAGTGG CTCAAATACT
     · H  P  W   I  M  Q   S  T  K  V   P  Q  T   P  L  H
1151 ACCACCCTTG GATCATGCAA TCAACAAAGG TCCCTCAAAC CCCACTGCAC
     TGGTGGGAAC CTAGTACGTT AGTTGTTTCC AGGGAGTTTG GGGTGACGTG
      T  S  R  V   L  K  E   D  K  E   R  W  E  D   V  K  E ·
1201 ACCAGCCGGG TCCTGAAGGA GGACAAGGAG CGGTGGGAGG ATGTCAAGGA
     TGGTCGGCCC AGGACTTCCT CCTGTTCCTC GCCACCCTCC TACAGTTCCT
     · E  M  T   S  A  L  A   T  M  R   V  D  Y   E  Q  I  K ·
1251 GGAGATGACC AGTGCCTTGG CCACAATGCG CGTTGACTAC GAGCAGATCA
     CCTCTACTGG TCACGGAACC GGTGTTACGC GCAACTGATG CTCGTCTAGT

     · *
1301 AGTAA
     TCATT
```

Figure 37B

Figure 38A

SUMO β-Gal

β-Gal NCBI Accession# V00296

```
      M  G  H  H  H  H  H  H  G  S  D  S  E  V  N  Q  E
   1  ATGGGTCATC ACCATCATCA TCACGGGTCG GACTCAGAAG TCAATCAAGA
      TACCCAGTAG TGGTAGTAGT AGTGCCCAGC CTGAGTCTTC AGTTAGTTCT
      A  K  P  E  V  K  P  E  V  K  P  E  T  H  I  N  L
  51  AGCTAAGCCA GAGGTCAAGC CAGAAGTCAA GCCTGAGACT CACATCAATT
      TCGATTCGGT CTCCAGTTCG GTCTTCAGTT CGGACTCTGA GTGTAGTTAA
      K  V  S  D  G  S  S  E  I  F  F  K  I  K  K  T
 101  TAAAGGTGTC CGATGGATCT TCAGAGATCT TCTTCAAGAT CAAAAAGACC
      ATTTCCACAG GCTACCTAGA AGTCTCTAGA AGAAGTTCTA GTTTTTCTGG
      T  P  L  R  R  L  M  E  A  F  A  K  R  Q  G  K  E
 151  ACTCCTTTAA GAAGGCTGAT GGAAGCGTTC GCTAAAAGAC AGGGTAAGGA
      TGAGGAAATT CTTCCGACTA CCTTCGCAAG CGATTTTCTG TCCCATTCCT
      M  D  S  L  R  F  L  Y  D  G  I  R  I  Q  A  D  Q
 201  AATGGACTCC TTAAGATTCT TGTACGACGG TATTAGAATT CAAGCTGATC
      TTACCTGAGG AATTCTAAGA ACATGCTGCC ATAATCTTAA GTTCGACTAG
      T  P  E  D  L  D  M  E  D  N  D  I  I  E  A  H
 251  AGACCCCTGA AGATTTGGAC ATGGAGGATA ACGATATTAT TGAGGCTCAC
      TCTGGGGACT TCTAAACCTG TACCTCCTAT TGCTATAATA ACTCCGAGTG
      R  E  Q  I  G  G  M  T  M  I  T  D  S  L  A  V  V
 301  CGCGAACAGA TTGGAGGTAT GACCATGATT ACGGATTCAC TGGCCGTCGT
      GCGCTTGTCT AACCTCCATA CTGGTACTAA TGCCTAAGTG ACCGGCAGCA
      L  Q  R  R  D  W  E  N  P  G  V  T  Q  L  N  R  L
 351  TTTACAACGT CGTGACTGGG AAAACCCTGG CGTTACCCAA CTTAATCGCC
      AAATGTTGCA GCACTGACCC TTTTGGGACC GCAATGGGTT GAATTAGCGG
      A  A  H  P  P  F  A  S  W  R  N  S  E  E  A  R
 401  TTGCAGCACA TCCCCCTTTC GCCAGCTGGC GTAATAGCGA AGAGGCCCGC
      AACGTCGTGT AGGGGGAAAG CGGTCGACCG CATTATCGCT TCTCCGGGCG
      T  D  R  P  S  Q  Q  L  R  S  L  N  G  E  W  R  F
 451  ACCGATCGCC CTTCCCAACA GTTGCGCAGC CTGAATGGCG AATGGCGCTT
      TGGCTAGCGG GAAGGGTTGT CAACGCGTCG GACTTACCGC TTACCGCGAA
      A  W  F  P  A  P  E  A  V  P  E  S  W  L  E  C  D
 501  TGCCTGGTTT CCGGCACCAG AAGCGGTGCC GGAAAGCTGG CTGGAGTGCG
      ACGGACCAAA GGCCGTGGTC TTCGCCACGG CCTTTCGACC GACCTCACGC
      L  P  E  A  D  T  V  V  V  P  S  N  W  Q  M  H
 551  ATCTTCCTGA GGCCGATACT GTCGTCGTCC CCTCAAACTG GCAGATGCAC
      TAGAAGGACT CCGGCTATGA CAGCAGCAGG GGAGTTTGAC CGTCTACGTG
      G  Y  D  A  P  I  Y  T  N  V  T  Y  P  I  T  V  N
 601  GGTTACGATG CGCCCATCTA CACCAACGTA ACCTATCCCA TTACGGTCAA
      CCAATGCTAC GCGGGTAGAT GTGGTTGCAT TGGATAGGGT AATGCCAGTT
      P  P  F  V  P  T  E  N  P  T  G  C  Y  S  L  T  F
 651  TCCGCCGTTT GTTCCCACGG AGAATCCGAC GGGTTGTTAC TCGCTCACAT
      AGGCGGCAAA CAAGGGTGCC TCTTAGGCTG CCCAACAATG AGCGAGTGTA
      N  V  D  E  S  W  L  Q  E  G  Q  T  R  I  I  F
 701  TTAATGTTGA TGAAAGCTGG CTACAGGAAG GCCAGACGCG AATTATTTTT
      AATTACAACT ACTTTCGACC GATGTCCTTC CGGTCTGCGC TTAATAAAAA
      D  G  V  N  S  A  F  H  L  W  C  N  G  R  W  V  G
 751  GATGGCGTTA ACTCGGCGTT TCATCTGTGG TGCAACGGGC GCTGGGTCGG
      CTACCGCAAT TGAGCCGCAA AGTAGACACC ACGTTGCCCG CGACCCAGCC
```

Figure 38B

```
      · Y  G  Q  D  S  R  L  P  S  E  F  D  L  S  A  F  L ·
 801  TTACGGCCAG GACAGTCGTT TGCCGTCTGA ATTTGACCTG AGCGCATTTT
      AATGCCGGTC CTGTCAGCAA ACGGCAGACT TAAACTGGAC TCGCGTAAAA
      · R  A  G  E  N  R  L  A  V  M  V  L  R  W  S  D
 851  TACGCGCCGG AGAAAACCGC CTCGCGGTGA TGGTGCTGCG TTGGAGTGAC
      ATGCGCGGCC TCTTTTGGCG GAGCGCCACT ACCACGACGC AACCTCACTG
       G  S  Y  L  E  D  Q  D  M  W  R  M  S  G  I  F  R ·
 901  GGCAGTTATC TGGAAGATCA GGATATGTGG CGGATGAGCG GCATTTTCCG
      CCGTCAATAG ACCTTCTAGT CCTATACACC GCCTACTCGC CGTAAAAGGC
      · D  V  S  L  L  H  K  P  T  T  Q  I  S  D  F  H  V ·
 951  TGACGTCTCG TTGCTGCATA AACCGACTAC ACAAATCAGC GATTTCCATG
      ACTGCAGAGC AACGACGTAT TTGGCTGATG TGTTTAGTCG CTAAAGGTAC
      · A  T  R  F  N  D  D  F  S  R  A  V  L  E  A  E
1001  TTGCCACTCG CTTTAATGAT GATTTCAGCC GCGCTGTACT GGAGGCTGAA
      AACGGTGAGC GAAATTACTA CTAAAGTCGG CGCGACATGA CCTCCGACTT
       V  Q  M  C  G  E  L  R  D  Y  L  R  V  T  V  S  L ·
1051  GTTCAGATGT GCGGCGAGTT GCGTGACTAC CTACGGGTAA CAGTTTCTTT
      CAAGTCTACA CGCCGCTCAA CGCACTGATG GATGCCCATT GTCAAAGAAA
      · W  Q  G  E  T  Q  V  A  S  G  T  A  P  F  G  G  E ·
1101  ATGGCAGGGT GAAACGCAGG TCGCCAGCGG CACCGCGCCT TTCGGCGGTG
      TACCGTCCCA CTTTGCGTCC AGCGGTCGCC GTGGCGCGGA AAGCCGCCAC
      · I  I  D  E  R  G  G  Y  A  D  R  V  T  L  R  L
1151  AAATTATCGA TGAGCGTGGT GGTTATGCCG ATCGCGTCAC ACTACGTCTG
      TTTAATAGCT ACTCGCACCA CCAATACGGC TAGCGCAGTG TGATGCAGAC
       N  V  E  N  P  K  L  W  S  A  E  I  P  N  L  Y  R ·
1201  AACGTCGAAA ACCCGAAACT GTGGAGCGCC GAAATCCCGA ATCTCTATCG
      TTGCAGCTTT TGGGCTTTGA CACCTCGCGG CTTTAGGGCT TAGAGATAGC
      · A  V  V  E  L  H  T  A  D  G  T  L  I  E  A  E  A ·
1251  TGCGGTGGTT GAACTGCACA CCGCCGACGG CACGCTGATT GAAGCAGAAG
      ACGCCACCAA CTTGACGTGT GGCGGCTGCC GTGCGACTAA CTTCGTCTTC
      · C  D  V  G  F  R  E  V  R  I  E  N  G  L  L  L
1301  CCTGCGATGT CGGTTTCCGC GAGGTGCGGA TTGAAAATGG TCTGCTGCTG
      GGACGCTACA GCCAAAGGCG CTCCACGCCT AACTTTTACC AGACGACGAC
       L  N  G  K  P  L  L  I  R  G  V  N  R  H  E  H  H ·
1351  CTGAACGGCA AGCCGTTGCT GATTCGAGGC GTTAACCGTC ACGAGCATCA
      GACTTGCCGT TCGGCAACGA CTAAGCTCCG CAATTGGCAG TGCTCGTAGT
      · P  L  H  G  Q  V  M  D  E  Q  T  M  V  Q  D  I  L ·
1401  TCCTCTGCAT GGTCAGGTCA TGGATGAGCA GACGATGGTG CAGGATATCC
      AGGAGACGTA CCAGTCCAGT ACCTACTCGT CTGCTACCAC GTCCTATAGG
      · L  M  K  Q  N  N  F  N  A  V  R  C  S  H  Y  P
1451  TGCTGATGAA GCAGAACAAC TTTAACGCCG TGCGCTGTTC GCATTATCCG
      ACGACTACTT CGTCTTGTTG AAATTGCGGC ACGCGACAAG CGTAATAGGC
       N  H  P  L  W  Y  T  L  C  D  R  Y  G  L  Y  V  V ·
1501  AACCATCCGC TGTGGTACAC GCTGTGCGAC CGCTACGGCC TGTATGTGGT
      TTGGTAGGCG ACACCATGTG CGACACGCTG GCGATGCCGG ACATACACCA
      · D  E  A  N  I  E  T  H  G  M  V  P  M  N  R  L  T ·
1551  GGATGAAGCC AATATTGAAA CCCACGGCAT GGTGCCAATG AATCGTCTGA
      CCTACTTCGG TTATAACTTT GGGTGCCGTA CCACGGTTAC TTAGCAGACT

      · D  D  P  R  W  L  P  A  M  S  E  R  V  T  R  M
1601  CCGATGATCC GCGCTGGCTA CCGGCGATGA GCGAACGCGT AACGCGAATG
      GGCTACTAGG CGCGACCGAT GGCCGCTACT CGCTTGCGCA TTGCGCTTAC
```

Figure 38C

```
          V  Q  R  D  R  N  H  P  S  V  I  I  W  S  L  G  N ·
1651  GTGCAGCGCG ATCGTAATCA CCCGAGTGTG ATCATCTGGT CGCTGGGGAA
      CACGTCGCGC TAGCATTAGT GGGCTCACAC TAGTAGACCA GCGACCCCTT
      · E  S  G  H  G  A  N  H  D  A  L  Y  R  W  I  K  S ·
1701  TGAATCAGGC CACGGCGCTA ATCACGACGC GCTGTATCGC TGGATCAAAT
      ACTTAGTCCG GTGCCGCGAT TAGTGCTGCG CGACATAGCG ACCTAGTTTA
      · V  D  P  S  R  P  V  Q  Y  E  G  G  G  A  D  T
1751  CTGTCGATCC TTCCCGCCCG GTGCAGTATG AAGGCGGCGG AGCCGACACC
      GACAGCTAGG AAGGGCGGGC CACGTCATAC TTCCGCCGCC TCGGCTGTGG
       T  A  T  D  I  I  C  P  M  Y  A  R  V  D  E  D  Q ·
1801  ACGGCCACCG ATATTATTTG CCCGATGTAC GCGCGCGTGG ATGAAGACCA
      TGCCGGTGGC TATAATAAAC GGGCTACATG CGCGCGCACC TACTTCTGGT
      · P  F  P  A  V  P  K  W  S  I  K  K  W  L  S  L  P ·
1851  GCCCTTCCCG GCTGTGCCGA AATGGTCCAT CAAAAAATGG CTTTCGCTAC
      CGGGAAGGGC CGACACGGCT TTACCAGGTA GTTTTTTACC GAAAGCGATG
      · G  E  T  R  P  L  I  L  C  E  Y  A  H  A  M  G
1901  CTGGAGAGAC GCGCCCGCTG ATCCTTTGCG AATACGCCCA CGCGATGGGT
      GACCTCTCTG CGCGGGCGAC TAGGAAACGC TTATGCGGGT GCGCTACCCA
       N  S  L  G  G  F  A  K  Y  W  Q  A  F  R  Q  Y  P ·
1951  AACAGTCTTG GCGGTTTCGC TAAATACTGG CAGGCGTTTC GTCAGTATCC
      TTGTCAGAAC CGCCAAAGCG ATTTATGACC GTCCGCAAAG CAGTCATAGG
      · R  L  Q  G  G  F  V  W  D  W  V  D  Q  S  L  I  K ·
2001  CCGTTTACAG GGCGGCTTCG TCTGGGACTG GGTGGATCAG TCGCTGATTA
      GGCAAATGTC CCGCCGAAGC AGACCCTGAC CCACCTAGTC AGCGACTAAT
      · Y  D  E  N  G  N  P  W  S  A  Y  G  G  D  F  G
2051  AATATGATGA AAACGGCAAC CCGTGGTCGG CTTACGGCGG TGATTTTGGC
      TTATACTACT TTTGCCGTTG GGCACCAGCC GAATGCCGCC ACTAAAACCG
       D  T  P  N  D  R  Q  F  C  M  N  G  L  V  F  A  D ·
2101  GATACGCCGA ACGATCGCCA GTTCTGTATG AACGGTCTGG TCTTTGCCGA
      CTATGCGGCT TGCTAGCGGT CAAGACATAC TTGCCAGACC AGAAACGGCT
      · R  T  P  H  P  A  L  T  E  A  K  H  Q  Q  Q  F  F ·
2151  CCGCACGCCG CATCCAGCGC TGACGGAAGC AAAACACCAG CAGCAGTTTT
      GGCGTGCGGC GTAGGTCGCG ACTGCCTTCG TTTTGTGGTC GTCGTCAAAA
      · Q  F  R  L  S  G  Q  T  I  E  V  T  S  E  Y  L
2201  TCCAGTTCCG TTTATCCGGG CAAACCATCG AAGTGACCAG CGAATACCTG
      AGGTCAAGGC AAATAGGCCC GTTTGGTAGC TTCACTGGTC GCTTATGGAC
       F  R  H  S  D  N  E  L  L  H  W  M  V  A  L  D  G ·
2251  TTCCGTCATA GCGATAACGA GCTCCTGCAC TGGATGGTGG CGCTGGATGG
      AAGGCAGTAT CGCTATTGCT CGAGGACGTG ACCTACCACC GCGACCTACC
      · K  P  L  A  S  G  E  V  P  L  D  V  A  P  Q  G  K ·
2301  TAAGCCGCTG GCAAGCGGTG AAGTGCCTCT GGATGTCGCT CCACAAGGTA
      ATTCGGCGAC CGTTCGCCAC TTCACGGAGA CCTACAGCGA GGTGTTCCAT
      · Q  L  I  E  L  P  E  L  P  Q  P  E  S  A  G  Q
2351  AACAGTTGAT TGAACTGCCT GAACTACCGC AGCCGGAGAG CGCCGGGCAA
      TTGTCAACTA ACTTGACGGA CTTGATGGCG TCGGCCTCTC GCGGCCCGTT
       L  W  L  T  V  R  V  V  Q  P  N  A  T  A  W  S  E ·
2401  CTCTGGCTCA CAGTACGCGT AGTGCAACCG AACGCGACCG CATGGTCAGA
      GAGACCGAGT GTCATGCGCA TCACGTTGGC TTGCGCTGGC GTACCAGTCT

      · A  G  H  I  S  A  W  Q  Q  W  R  L  A  E  N  L  S ·
2451  AGCCGGGCAC ATCAGCGCCT GGCAGCAGTG GCGTCTGGCG GAAAACCTCA
      TCGGCCCGTG TAGTCGCGGA CCGTCGTCAC CGCAGACCGC CTTTTGGAGT
```

Figure 38D

```
      · V T L P A A S H A I P H L T T S
2501  GTGTGACGCT CCCCGCCGCG TCCCACGCCA TCCCGCATCT GACCACCAGC
      CACACTGCGA GGGGCGGCGC AGGGTGCGGT AGGGCGTAGA CTGGTGGTCG
       E M D F C I E L G N K R W Q F N R ·
2551  GAAATGGATT TTTGCATCGA GCTGGGTAAT AAGCGTTGGC AATTTAACCG
      CTTTACCTAA AAACGTAGCT CGACCCATTA TTCGCAACCG TTAAATTGGC
      · Q S G F L S Q M W I G D K K Q L L ·
2601  CCAGTCAGGC TTTCTTTCAC AGATGTGGAT TGGCGATAAA AAACAACTGC
      GGTCAGTCCG AAAGAAAGTG TCTACACCTA ACCGCTATTT TTTGTTGACG
      · T P L R D Q F T R A P L D N D I
2651  TGACGCCGCT GCGCGATCAG TTCACCCGTG CACCGCTGGA TAACGACATT
      ACTGCGGCGA CGCGCTAGTC AAGTGGGCAC GTGGCGACCT ATTGCTGTAA
       G V S E A T R I D P N A W V E R W ·
2701  GGCGTAAGTG AAGCGACCCG CATTGACCCT AACGCCTGGG TCGAACGCTG
      CCGCATTCAC TTCGCTGGGC GTAACTGGGA TTGCGGACCC AGCTTGCGAC
      · K A A G H Y Q A E A A L L Q C T A ·
2751  GAAGGCGGCG GGCCATTACC AGGCCGAAGC AGCGTTGTTG CAGTGCACGG
      CTTCCGCCGC CCGGTAATGG TCCGGCTTCG TCGCAACAAC GTCACGTGCC
      · D T L A D A V L I T T A H A W Q
2801  CAGATACACT TGCTGATGCG GTGCTGATTA CGACCGCTCA CGCGTGGCAG
      GTCTATGTGA ACGACTACGC CACGACTAAT GCTGGCGAGT GCGCACCGTC
       H Q G K T L F I S R K T Y R I D G ·
2851  CATCAGGGGA AAACCTTATT TATCAGCCGG AAAACCTACC GGATTGATGG
      GTAGTCCCCT TTTGGAATAA ATAGTCGGCC TTTTGGATGG CCTAACTACC
      · S G Q M A I T V D V E V A S D T P ·
2901  TAGTGGTCAA ATGGCGATTA CCGTTGATGT TGAAGTGGCG AGCGATACAC
      ATCACCAGTT TACCGCTAAT GGCAACTACA ACTTCACCGC TCGCTATGTG
      · H P A R I G L N C Q L A Q V A E
2951  CGCATCCGGC GCGGATTGGC CTGAACTGCC AGCTGGCGCA GGTAGCAGAG
      GCGTAGGCCG CGCCTAACCG GACTTGACGG TCGACCGCGT CCATCGTCTC
       R V N W L G L G P Q E N Y P D R L ·
3001  CGGGTAAACT GGCTCGGATT AGGGCCGCAA GAAAACTATC CCGACCGCCT
      GCCCATTTGA CCGAGCCTAA TCCCGGCGTT CTTTTGATAG GGCTGGCGGA
      · T A A C F D R W D L P L S D M Y T ·
3051  TACTGCCGCC TGTTTTGACC GCTGGGATCT GCCATTGTCA GACATGTATA
      ATGACGGCGG ACAAAACTGG CGACCCTAGA CGGTAACAGT CTGTACATAT
      · P Y V F P S E N G L R C G T R E
3101  CCCCGTACGT CTTCCCGAGC GAAAACGGTC TGCGCTGCGG GACGCGCGAA
      GGGGCATGCA GAAGGGCTCG CTTTTGCCAG ACGCGACGCC CTGCGCGCTT
       L N Y G P H Q W R G D F Q F N I S ·
3151  TTGAATTATG GCCCACACCA GTGGCGCGGC GACTTCCAGT TCAACATCAG
      AACTTAATAC CGGGTGTGGT CACCGCGCCG CTGAAGGTCA AGTTGTAGTC
      · R Y S Q Q Q L M E T S H R H L L H ·
3201  CCGCTACAGT CAACAGCAAC TGATGGAAAC CAGCCATCGC CATCTGCTGC
      GGCGATGTCA GTTGTCGTTG ACTACCTTTG GTCGGTAGCG GTAGACGACG
      · A E E G T W L N I D G F H M G I
3251  ACGCGGAAGA AGGCACATGG CTGAATATCG ACGGTTTCCA TATGGGGATT
      TGCGCCTTCT TCCGTGTACC GACTTATAGC TGCCAAAGGT ATACCCCTAA

       G G D D S W S P S V S A E F Q L S ·
3301  GGTGGCGACG ACTCCTGGAG CCCGTCAGTA TCGGCGGAAT TCCAGCTGAG
      CCACCGCTGC TGAGGACCTC GGGCAGTCAT AGCCGCCTTA AGGTCGACTC
```

```
      · A  G  R   Y  H  Y  Q   L  V  W   C  Q  K   *  *
3351  CGCCGGTCGC TACCATTACC AGTTGGTCTG GTGTCAAAAA TAATAA
      GCGGCCAGCG ATGGTAATGG TCAACCAGAC CACAGTTTTT ATTATT
```

Figure 38E

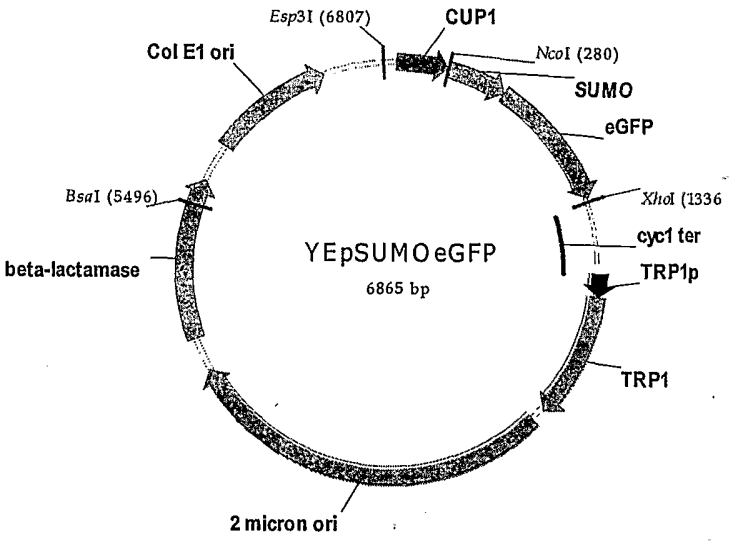

Esp3I (6807)
CUP1
Col E1 ori
NcoI (280)
SUMO
eGFP
BsaI (5496)
XhoI (1336
YEpSUMOeGFP
6865 bp
cyc1 ter
beta-lactamase
TRP1p
TRP1
2 micron ori

Figure 39

Figure 40A

```
   1  CGCCTTGTTA CTAGTTAGAA AAAGACATTT TTGCTGTCAG TCACTGTCAA
      GCGGAACAAT GATCAATCTT TTTCTGTAAA AACGACAGTC AGTGACAGTT
  51  GAGATTCTTT TGCTGGCATT TCTTCTAGAA GCAAAAAGAG CGATGCGTCT
      CTCTAAGAAA ACGACCGTAA AGAAGATCTT CGTTTTTCTC GCTACGCAGA
 101  TTTCCGCTGA ACCGTTCCAG CAAAAAAGAC TACCAACGCA ATATGGATTG
      AAAGGCGACT TGGCAAGGTC GTTTTTTCTG ATGGTTGCGT TATACCTAAC
 151  TCAGAATCAT ATAAAAGAGA AGCAAATAAC TCCTTGTCTT GTATCAATTG
      AGTCTTAGTA TATTTTCTCT TCGTTTATTG AGGAACAGAA CATAGTTAAC
 201  CATTATAATA TCTTCTTGTT AGTGCAATAT CATATAGAAG TCATCGAAAT
      GTAATATTAT AGAAGAACAA TCACGTTATA GTATATCTTC AGTAGCTTTA
                                      NcoI
                                      ~~~~~~~
 251  AGATATTAAG AAAAACAAAC TGTACAATCC ATGGGTCATC ACCATCATCA
      TCTATAATTC TTTTTGTTTG ACATGTTAGG TACCCAGTAG TGGTAGTAGT
 301  TCACGGGTCG GACTCAGAAG TCAATCAAGA AGCTAAGCCA GAGGTCAAGC
      AGTGCCCAGC CTGAGTCTTC AGTTAGTTCT TCGATTCGGT CTCCAGTTCG
 351  CAGAAGTCAA GCCTGAGACT CACATCAATT TAAAGGTGTC CGATGGATCT
      GTCTTCAGTT CGGACTCTGA GTGTAGTTAA ATTTCCACAG GCTACCTAGA
 401  TCAGAGATCT TCTTCAAGAT CAAAAAGACC ACTCCTTTAA GAAGGCTGAT
      AGTCTCTAGA AGAAGTTCTA GTTTTTCTGG TGAGGAAATT CTTCCGACTA
 451  GGAAGCGTTC GCTAAAAGAC AGGGTAAGGA AATGGACTCC TTAAGATTCT
      CCTTCGCAAG CGATTTTCTG TCCCATTCCT TTACCTGAGG AATTCTAAGA
 501  TGTACGACGG TATTAGAATT CAAGCTGATC AGACCCCTGA AGATTTGGAC
      ACATGCTGCC ATAATCTTAA GTTCGACTAG TCTGGGGACT TCTAAACCTG
 551  ATGGAGGATA ACGATATTAT TGAGGCTCAC CGCGAACAGA TTGGAGGTAT
      TACCTCCTAT TGCTATAATA ACTCCGAGTG GCGCTTGTCT AACCTCCATA
 601  GGTGAGCAAG GGCGAGGAGC TGTTCACCGG GGTGGTGCCC ATCCTGGTCG
      CCACTCGTTC CCGCTCCTCG ACAAGTGGCC CCACCACGGG TAGGACCAGC
 651  AGCTGGACGG CGACGTAAAC GGCCACAAGT TCAGCGTGTC CGGCGAGGGC
      TCGACCTGCC GCTGCATTTG CCGGTGTTCA AGTCGCACAG GCCGCTCCCG
 701  GAGGGCGATG CCACCTACGG CAAGCTGACC CTGAAGTTCA TCTGCACCAC
      CTCCCGCTAC GGTGGATGCC GTTCGACTGG GACTTCAAGT AGACGTGGTG
 751  CGGCAAGCTG CCCGTGCCCT GGCCCACCCT CGTGACCACC CTGACCTACG
      GCCGTTCGAC GGGCACGGGA CCGGGTGGGA GCACTGGTGG GACTGGATGC
 801  GCGTGCAGTG CTTCAGCCGC TACCCCGACC ACATGAAGCA GCACGACTTC
      CGCACGTCAC GAAGTCGGCG ATGGGGCTGG TGTACTTCGT CGTGCTGAAG
 851  TTCAAGTCCG CCATGCCCGA AGGCTACGTC CAGGAGCGCA CCATCTTCTT
      AAGTTCAGGC GGTACGGGCT TCCGATGCAG GTCCTCGCGT GGTAGAAGAA
 901  CAAGGACGAC GGCAACTACA AGACCCGCGC CGAGGTGAAG TTCGAGGGCG
      GTTCCTGCTG CCGTTGATGT TCTGGGCGCG GCTCCACTTC AAGCTCCCGC
 951  ACACCCTGGT GAACCGCATC GAGCTGAAGG GCATCGACTT CAAGGAGGAC
      TGTGGGACCA CTTGGCGTAG CTCGACTTCC CGTAGCTGAA GTTCCTCCTG
1001  GGCAACATCC TGGGGCACAA GCTGGAGTAC AACTACAACA GCCACAACGT
      CCGTTGTAGG ACCCCGTGTT CGACCTCATG TTGATGTTGT CGGTGTTGCA
1051  CTATATCATG GCCGACAAGC AGAAGAACGG CATCAAGGTG AACTTCAAGA
      GATATAGTAC CGGCTGTTCG TCTTCTTGCC GTAGTTCCAC TTGAAGTTCT
1101  TCCGCCACAA CATCGAGGAC GGCAGCGTGC AGCTCGCCGA CCACTACCAG
      AGGCGGTGTT GTAGCTCCTG CCGTCGCACG TCGAGCGGCT GGTGATGGTC
1151  CAGAACACCC CCATCGGCGA CGGCCCCGTG CTGCTGCCCG ACAACCACTA
      GTCTTGTGGG GGTAGCCGCT GCCGGGGCAC GACGACGGGC TGTTGGTGAT
```

Figure 40B

```
1201  CCTGAGCACC CAGTCCGCCC TGAGCAAAGA CCCCAACGAG AAGCGCGATC
      GGACTCGTGG GTCAGGCGGG ACTCGTTTCT GGGGTTGCTC TTCGCGCTAG
1251  ACATGGTCCT GCTGGAGTTC GTGACCGCCG CCGGGATCAC TCTCGGCATG
      TGTACCAGGA CGACCTCAAG CACTGGCGGC GGCCCTAGTG AGAGCCGTAC
                                          XhoI
                                          ~~~~~~~
1301  GACGAGCTGT ACAAGTAATA AGCTTGCGGC CGCACTCGAG GAGCTCCCTG
      CTGCTCGACA TGTTCATTAT TCGAACGCCG GCGTGAGCTC CTCGAGGGAC
1351  GCGAATTGTA CCAAGATGGC CTTTGGTGGG TTGAAGAAGG AAAAAGACAG
      CGCTTAACAT GGTTCTACCG GAAACCACCC AACTTCTTCC TTTTTCTGTC
1401  AAACGACTTA ATTACCTACT TGAAAAAAGC CTGTGAGTAA ACAGGCCCCT
      TTTGCTGAAT TAATGGATGA ACTTTTTTCG GACACTCATT TGTCCGGGGA
1451  TTTCCTTTGT CGATATCATG TAATTAGTTA TGTCACGCTT ACATTCACGC
      AAAGGAAACA GCTATAGTAC ATTAATCAAT ACAGTGCGAA TGTAAGTGCG
1501  CCTCCCCCCA CATCCGCTCT AACCGAAAAG GAAGGAGTTA GACAACCTGA
      GGAGGGGGGT GTAGGCGAGA TTGGCTTTTC CTTCCTCAAT CTGTTGGACT
1551  AGTCTAGGTC CCTATTTATT TTTTTATAGT TATGTTAGTA TTAAGAACGT
      TCAGATCCAG GGATAAATAA AAAAATATCA ATACAATCAT AATTCTTGCA
1601  TATTTATATT TCAAATTTTT CTTTTTTTTC TGTACAGACG CGTGTACGCA
      ATAAATATAA AGTTTAAAAA GAAAAAAAAG ACATGTCTGC GCACATGCGT
1651  TGTAACATTA TACTGAAAAC CTTGCTTGAG AAGGTTTTGG GACGCTCGAA
      ACATTGTAAT ATGACTTTTG GAACGAACTC TTCCAAAACC CTGCGAGCTT
1701  GGCTTTAATT TGCAAGCTTA TCGATGATAA GCTGTCAAAC ATGAGAATTC
      CCGAAATTAA ACGTTCGAAT AGCTACTATT CGACAGTTTG TACTCTTAAG
1751  GGTCGAAAAA AGAAAAGGAG AGGGCCAAGA GGGAGGGCAT TGGTGACTAT
      CCAGCTTTTT TCTTTTCCTC TCCCGGTTCT CCCTCCCGTA ACCACTGATA
1801  TGAGCACGTG AGTATACGTG ATTAAGCACA CAAAGGCAGC TTGGAGTATG
      ACTCGTGCAC TCATATGCAC TAATTCGTGT GTTTCCGTCG AACCTCATAC
1851  TCTGTTATTA ATTTCACAGG TAGTTCTGGT CCATTGGTGA AAGTTTGCGG
      AGACAATAAT TAAAGTGTCC ATCAAGACCA GGTAACCACT TTCAAACGCC
1901  CTTGCAGAGC ACAGAGGCCG CAGAATGTGC TCTAGATTCC GATGCTGACT
      GAACGTCTCG TGTCTCCGGC GTCTTACACG AGATCTAAGG CTACGACTGA
1951  TGCTGGGTAT TATATGTGTG CCCAATAGAA AGAGAACAAT TGACCCGGTT
      ACGACCCATA ATATACACAC GGGTTATCTT TCTCTTGTTA ACTGGGCCAA
2001  ATTGCAAGGA AAATTTCAAG TCTTGTAAAA GCATATAAAA ATAGTTCAGG
      TAACGTTCCT TTTAAAGTTC AGAACATTTT CGTATATTTT TATCAAGTCC
2051  CACTCCGAAA TACTTGGTTG GCGTGTTTCG TAATCAACCT AAGGAGGATG
      GTGAGGCTTT ATGAACCAAC CGCACAAAGC ATTAGTTGGA TTCCTCCTAC
2101  TTTTGGCTCT GGTCAATGAT TACGGCATTG ATATCGTCCA ACTGCATGGA
      AAAACCGAGA CCAGTTACTA ATGCCGTAAC TATAGCAGGT TGACGTACCT
2151  GATGAGTCGT GGCAAGAATA CCAAGAGTTC CTCGGTTTGC CAGTTATTAA
      CTACTCAGCA CCGTTCTTAT GGTTCTCAAG GAGCCAAACG GTCAATAATT
2201  AAGACTCGTA TTTCCAAAAG ACTGCAACAT ACTACTCAGT GCAGCTTCAC
      TTCTGAGCAT AAAGGTTTTC TGACGTTGTA TGATGAGTCA CGTCGAAGTG
2251  AGAAACCTCA TTCGTTTATT CCCTTGTTTG ATTCAGAAGC AGGTGGGACA
      TCTTTGGAGT AAGCAAATAA GGGAACAAAC TAAGTCTTCG TCCACCCTGT
2301  GGTGAACTTT TGGATTGGAA CTCGATTTCT GACTGGGTTG GAAGGCAAGA
      CCACTTGAAA ACCTAACCTT GAGCTAAAGA CTGACCCAAC CTTCCGTTCT
2351  GAGCCCCGAA AGCTTACATT TTATGTTAGC TGGTGGACTG ACGCCAGAAA
      CTCGGGGCTT TCGAATGTAA AATACAATCG ACCACCTGAC TGCGGTCTTT
2401  ATGTTGGTGA TGCGCTTAGA TTAAATGGCG TTATTGGTGT TGATGTAAGC
```

Figure 40C

```
      TACAACCACT ACGCGAATCT AATTTACCGC AATAACCACA ACTACATTCG
2451  GGAGGTGTGG AGACAAATGG TGTAAAAGAC TCTAACAAAA TAGCAAATTT
      CCTCCACACC TCTGTTTACC ACATTTTCTG AGATTGTTTT ATCGTTTAAA
2501  CGTCAAAAAT GCTAAGAAAT AGGTTATTAC TGAGTAGTAT TTATTTAAGT
      GCAGTTTTTA CGATTCTTTA TCCAATAATG ACTCATCATA AATAAATTCA
2551  ATTGTTTGTG CACTTGCCTG CAGCTTCTCA ATGATATTCG AATACGCTTT
      TAACAAACAC GTGAACGGAC GTCGAAGAGT TACTATAAGC TTATGCGAAA
2601  GAGGAGATAC AGCCTAATAT CCGACAAACT GTTTTACAGA TTTACGATCG
      CTCCTCTATG TCGGATTATA GGCTGTTTGA CAAAATGTCT AAATGCTAGC
2651  TACTTGTTAC CCATCATTGA ATTTTGAACA TCCGAACCTG GGAGTTTTCC
      ATGAACAATG GGTAGTAACT TAAAACTTGT AGGCTTGGAC CCTCAAAAGG
2701  CTGAAACAGA TAGTATATTT GAACCTGTAT AATAATATAT AGTCTAGCGC
      GACTTTGTCT ATCATATAAA CTTGGACATA TTATTATATA TCAGATCGCG
2751  TTTACGGAAG ACAATGTATG TATTTCGGTT CCTGGAGAAA CTATTGCATC
      AAATGCCTTC TGTTACATAC ATAAAGCCAA GGACCTCTTT GATAACGTAG
2801  TATTGCATAG GTAATCTTGC ACGTCGCATC CCCGGTTCAT TTTCTGCGTT
      ATAACGTATC CATTAGAACG TGCAGCGTAG GGGCCAAGTA AAAGACGCAA
2851  TCCATCTTGC ACTTCAATAG CATATCTTTG TTAACGAAGC ATCTGTGCTT
      AGGTAGAACG TGAAGTTATC GTATAGAAAC AATTGCTTCG TAGACACGAA
2901  CATTTTGTAG AACAAAAATG CAACGCGAGA GCGCTAATTT TTCAAACAAA
      GTAAAACATC TTGTTTTTAC GTTGCGCTCT CGCGATTAAA AAGTTTGTTT
2951  GAATCTGAGC TGCATTTTTA CAGAACAGAA ATGCAACGCG AAAGCGCTAT
      CTTAGACTCG ACGTAAAAAT GTCTTGTCTT TACGTTGCGC TTTCGCGATA
3001  TTTACCAACG AAGAATCTGT GCTTCATTTT TGTAAAACAA AAATGCAACG
      AAATGGTTGC TTCTTAGACA CGAAGTAAAA ACATTTTGTT TTTACGTTGC
3051  CGAGAGCGCT AATTTTTCAA ACAAAGAATC TGAGCTGCAT TTTTACAGAA
      GCTCTCGCGA TTAAAAAGTT TGTTTCTTAG ACTCGACGTA AAAATGTCTT
3101  CAGAAATGCA ACGCGAGAGC GCTATTTTAC CAACAAAGAA TCTATACTTC
      GTCTTTACGT TGCGCTCTCG CGATAAAATG GTTGTTTCTT AGATATGAAG
3151  TTTTTTGTTC TACAAAAATG CATCCCGAGA GCGCTATTTT TCTAACAAAG
      AAAAAACAAG ATGTTTTTAC GTAGGGCTCT CGCGATAAAA AGATTGTTTC
3201  CATCTTAGAT TACTTTTTTT CTCCTTTGTG CGCTCTATAA TGCAGTCTCT
      GTAGAATCTA ATGAAAAAAA GAGGAAACAC GCGAGATATT ACGTCAGAGA
3251  TGATAACTTT TTGCACTGTA GGTCCGTTAA GGTTAGAAGA AGGCTACTTT
      ACTATTGAAA AACGTGACAT CCAGGCAATT CCAATCTTCT TCCGATGAAA
3301  GGTGTCTATT TTCTCTTCCA TAAAAAAAGC CTGACTCCAC TTCCCGCGTT
      CCACAGATAA AAGAGAAGGT ATTTTTTTCG GACTGAGGTG AAGGGCGCAA
3351  TACTGATTAC TAGCGAAGCT GCGGGTGCAT TTTTTCAAGA TAAAGGCATC
      ATGACTAATG ATCGCTTCGA CGCCCACGTA AAAAAGTTCT ATTTCCGTAG
3401  CCCGATTATA TTCTATACCG ATGTGGATTG CGCATACTTT GTGAACAGAA
      GGGCTAATAT AAGATATGGC TACACCTAAC GCGTATGAAA CACTTGTCTT
3451  AGTGATAGCG TTGATGATTC TTCATTGGTC AGAAAATTAT GAACGGTTTC
      TCACTATCGC AACTACTAAG AAGTAACCAG TCTTTTAATA CTTGCCAAAG
3501  TTCTATTTTG TCTCTATATA CTACGTATAG GAAATGTTTA CATTTTCGTA
      AAGATAAAAC AGAGATATAT GATGCATATC CTTTACAAAT GTAAAAGCAT
3551  TTGTTTTCGA TTCACTCTAT GAATAGTTCT TACTACAATT TTTTTGTCTA
      AACAAAAGCT AAGTGAGATA CTTATCAAGA ATGATGTTAA AAAAACAGAT
3601  AAGAGTAATA CTAGAGATAA ACATAAAAAA TGTAGAGGTC GAGTTTAGAT
      TTCTCATTAT GATCTCTATT TGTATTTTTT ACATCTCCAG CTCAAATCTA
3651  GCAAGTTCAA GGAGCGAAAG GTGGATGGGT AGGTTATATA GGGATATAGC
      CGTTCAAGTT CCTCGCTTTC CACCTACCCA TCCAATATAT CCCTATATCG
```

Figure 40D

```
3701  ACAGAGATAT ATAGCAAAGA GATACTTTTG AGCAATGTTT GTGGAAGCGG
      TGTCTCTATA TATCGTTTCT CTATGAAAAC TCGTTACAAA CACCTTCGCC
3751  TATTCGCAAT ATTTTAGTAG CTCGTTACAG TCCGGTGCGT TTTTGGTTTT
      ATAAGCGTTA TAAAATCATC GAGCAATGTC AGGCCACGCA AAAACCAAAA
3801  TTGAAAGTGC GTCTTCAGAG CGCTTTTGGT TTTCAAAAGC GCTCTGAAGT
      AACTTTCACG CAGAAGTCTC GCGAAAACCA AAAGTTTTCG CGAGACTTCA
3851  TCCTATACTT TCTAGAGAAT AGGAACTTCG GAATAGGAAC TTCAAAGCGT
      AGGATATGAA AGATCTCTTA TCCTTGAAGC CTTATCCTTG AAGTTTCGCA
3901  TTCCGAAAAC GAGCGCTTCC GAAAATGCAA CGCGAGCTGC GCACATACAG
      AAGGCTTTTG CTCGCGAAGG CTTTTACGTT GCGCTCGACG CGTGTATGTC
3951  CTCACTGTTC ACGTCGCACC TATATCTGCG TGTTGCCTGT ATATATATAT
      GAGTGACAAG TGCAGCGTGG ATATAGACGC ACAACGGACA TATATATATA
4001  ACATGAGAAG AACGGCATAG TGCGTGTTTA TGCTTAAATG CGTACTTATA
      TGTACTCTTC TTGCCGTATC ACGCACAAAT ACGAATTTAC GCATGAATAT
4051  TGCGTCTATT TATGTAGGAT GAAAGGTAGT CTAGTACCTC CTGTGATATT
      ACGCAGATAA ATACATCCTA CTTTCCATCA GATCATGGAG GACACTATAA
4101  ATCCCATTCC ATGCGGGGTA TCGTATGCTT CCTTCAGCAC TACCCTTTAG
      TAGGGTAAGG TACGCCCCAT AGCATACGAA GGAAGTCGTG ATGGGAAATC
4151  CTGTTCTATA TGCTGCCACT CCTCAATTGG ATTAGTCTCA TCCTTCAATG
      GACAAGATAT ACGACGGTGA GGAGTTAACC TAATCAGAGT AGGAAGTTAC
4201  CTATCATTTC CTTTGATATT GGATCATATG CATAGTACCG AGAAACTAGT
      GATAGTAAAG GAAACTATAA CCTAGTATAC GTATCATGGC TCTTTGATCA
4251  GCGAAGTAGT GATCAGGTAT TGCTGTTATC TGATGAGTAT ACGTTGTCCT
      CGCTTCATCA CTAGTCCATA ACGACAATAG ACTACTCATA TGCAACAGGA
4301  GGCCACGGCA GAAGCACGCT TATCGCTCCA ATTTCCCACA ACATTAGTCA
      CCGGTGCCGT CTTCGTGCGA ATAGCGAGGT TAAAGGGTGT TGTAATCAGT
4351  ACTCCGTTAG GCCCTTCATT GAAAGAAATG AGGTCATCAA ATGTCTTCCA
      TGAGGCAATC CGGGAAGTAA CTTTCTTTAC TCCAGTAGTT TACAGAAGGT
4401  ATGTGAGATT TTGGGCCATT TTTTATAGCA AAGATTGAAT AAGGCGCATT
      TACACTCTAA AACCCGGTAA AAAATATCGT TTCTAACTTA TTCCGCGTAA
4451  TTTCTTCAAA GCTTTATTGT ACGATCTGAC TAAGTTATCT TTTAATAATT
      AAAGAAGTTT CGAAATAACA TGCTAGACTG ATTCAATAGA AAATTATTAA
4501  GGTATTCCTG TTTATTGCTT GAAGAATTGC CGGTCCTATT TACTCGTTTT
      CCATAAGGAC AAATAACGAA CTTCTTAACG GCCAGGATAA ATGAGCAAAA
4551  AGGACTGGTT CAGAATTCTT GAAGACGAAA GGGCCTCGTG ATACGCCTAT
      TCCTGACCAA GTCTTAAGAA CTTCTGCTTT CCCGGAGCAC TATGCGGATA
4601  TTTTATAGGT TAATGTCATG ATAATAATGG TTTCTTAGAC GTCAGGTGGC
      AAAATATCCA ATTACAGTAC TATTATTACC AAAGAATCTG CAGTCCACCG
4651  ACTTTTCGGG GAAATGTGCG CGGAACCCCT ATTTGTTTAT TTTTCTAAAT
      TGAAAAGCCC CTTTACACGC GCCTTGGGGA TAAACAAATA AAAAGATTTA
4701  ACATTCAAAT ATGTATCCGC TCATGAGACA ATAACCCTGA TAAATGCTTC
      TGTAAGTTTA TACATAGGCG AGTACTCTGT TATTGGGACT ATTTACGAAG
4751  AATAATATTG AAAAAGGAAG AGTATGAGTA TTCAACATTT CCGTGTCGCC
      TTATTATAAC TTTTTCCTTC TCATACTCAT AAGTTGTAAA GGCACAGCGG
4801  CTTATTCCCT TTTTTGCGGC ATTTTGCCTT CCTGTTTTTG CTCACCCAGA
      GAATAAGGGA AAAAACGCCG TAAAACGGAA GGACAAAAAC GAGTGGGTCT
4851  AACGCTGGTG AAAGTAAAAG ATGCTGAAGA TCAGTTGGGT GCACGAGTGG
      TTGCGACCAC TTTCATTTTC TACGACTTCT AGTCAACCCA CGTGCTCACC
4901  GTTACATCGA ACTGGATCTC AACAGCGGTA AGATCCTTGA GAGTTTTCGC
      CAATGTAGCT TGACCTAGAG TTGTCGCCAT TCTAGGAACT CTCAAAAGCG
4951  CCCGAAGAAC GTTTTCCAAT GATGAGCACT TTTAAAGTTC TGCTATGTGG
```

Figure 40E

```
      GGGCTTCTTG CAAAAGGTTA CTACTCGTGA AAATTTCAAG ACGATACACC
5001  CGCGGTATTA TCCCGTGTTG ACGCCGGGCA AGAGCAACTC GGTCGCCGCA
      GCGCCATAAT AGGGCACAAC TGCGGCCCGT TCTCGTTGAG CCAGCGGCGT
5051  TACACTATTC TCAGAATGAC TTGGTTGAGT ACTCACCAGT CACAGAAAAG
      ATGTGATAAG AGTCTTACTG AACCAACTCA TGAGTGGTCA GTGTCTTTTC
5101  CATCTTACGG ATGGCATGAC AGTAAGAGAA TTATGCAGTG CTGCCATAAC
      GTAGAATGCC TACCGTACTG TCATTCTCTT AATACGTCAC GACGGTATTG
5151  CATGAGTGAT AACACTGCGG CCAACTTACT TCTGACAACG ATCGGAGGAC
      GTACTCACTA TTGTGACGCC GGTTGAATGA AGACTGTTGC TAGCCTCCTG
5201  CGAAGGAGCT AACCGCTTTT TTGCACAACA TGGGGGATCA TGTAACTCGC
      GCTTCCTCGA TTGGCGAAAA AACGTGTTGT ACCCCCTAGT ACATTGAGCG
5251  CTTGATCGTT GGGAACCGGA GCTGAATGAA GCCATACCAA ACGACGAGCG
      GAACTAGCAA CCCTTGGCCT CGACTTACTT CGGTATGGTT TGCTGCTCGC
5301  TGACACCACG ATGCCTGCAG CAATGGCAAC AACGTTGCGC AAACTATTAA
      ACTGTGGTGC TACGGACGTC GTTACCGTTG TTGCAACGCG TTTGATAATT
5351  CTGGCGAACT ACTTACTCTA GCTTCCCGGC AACAATTAAT AGACTGGATG
      GACCGCTTGA TGAATGAGAT CGAAGGGCCG TTGTTAATTA TCTGACCTAC
5401  GAGGCGGATA AGTTGCAGG ACCACTTCTG CGCTCGGCCC TTCCGGCTGG
      CTCCGCCTAT TTCAACGTCC TGGTGAAGAC GCGAGCCGGG AAGGCCGACC
                                                  BsaI
                                                  ~~~~~~~
5451  CTGGTTTATT GCTGATAAAT CTGGAGCCGG TGAGCGTGGG TCTCGCGGTA
      GACCAAATAA CGACTATTTA GACCTCGGCC ACTCGCACCC AGAGCGCCAT
5501  TCATTGCAGC ACTGGGGCCA GATGGTAAGC CCTCCCGTAT CGTAGTTATC
      AGTAACGTCG TGACCCCGGT CTACCATTCG GGAGGGCATA GCATCAATAG
5551  TACACGACGG GGAGTCAGGC AACTATGGAT GAACGAAATA GACAGATCGC
      ATGTGCTGCC CCTCAGTCCG TTGATACCTA CTTGCTTTAT CTGTCTAGCG
5601  TGAGATAGGT GCCTCACTGA TTAAGCATTG GTAACTGTCA GACCAAGTTT
      ACTCTATCCA CGGAGTGACT AATTCGTAAC CATTGACAGT CTGGTTCAAA
5651  ACTCATATAT ACTTTAGATT GATTTAAAAC TTCATTTTTA ATTTAAAAGG
      TGAGTATATA TGAAATCTAA CTAAATTTTG AAGTAAAAAT TAAATTTTCC
5701  ATCTAGGTGA AGATCCTTTT TGATAATCTC ATGACCAAAA TCCCTTAACG
      TAGATCCACT TCTAGGAAAA ACTATTAGAG TACTGGTTTT AGGGAATTGC
5751  TGAGTTTTCG TTCCACTGAG CGTCAGACCC CGTAGAAAAG ATCAAAGGAT
      ACTCAAAAGC AAGGTGACTC GCAGTCTGGG GCATCTTTTC TAGTTTCCTA
5801  CTTCTTGAGA TCCTTTTTTT CTGCGCGTAA TCTGCTGCTT GCAAACAAAA
      GAAGAACTCT AGGAAAAAAA GACGCGCATT AGACGACGAA CGTTTGTTTT
5851  AAACCACCGC TACCAGCGGT GGTTTGTTTG CCGGATCAAG AGCTACCAAC
      TTTGGTGGCG ATGGTCGCCA CCAAACAAAC GGCCTAGTTC TCGATGGTTG
5901  TCTTTTTCCG AAGGTAACTG GCTTCAGCAG AGCGCAGATA CCAAATACTG
      AGAAAAAGGC TTCCATTGAC CGAAGTCGTC TCGCGTCTAT GGTTTATGAC
5951  TCCTTCTAGT GTAGCCGTAG TTAGGCCACC ACTTCAAGAA CTCTGTAGCA
      AGGAAGATCA CATCGGCATC AATCCGGTGG TGAAGTTCTT GAGACATCGT
6001  CCGCCTACAT ACCTCGCTCT GCTAATCCTG TTACCAGTGG CTGCTGCCAG
      GGCGGATGTA TGGAGCGAGA CGATTAGGAC AATGGTCACC GACGACGGTC
6051  TGGCGATAAG TCGTGTCTTA CCGGGTTGGA CTCAAGACGA TAGTTACCGG
      ACCGCTATTC AGCACAGAAT GGCCCAACCT GAGTTCTGCT ATCAATGGCC
6101  ATAAGGCGCA GCGGTCGGGC TGAACGGGGG GTTCGTGCAC ACAGCCCAGC
      TATTCCGCGT CGCCAGCCCG ACTTGCCCCC CAAGCACGTG TGTCGGGTCG
6151  TTGGAGCGAA CGACCTACAC CGAACTGAGA TACCTACAGC GTGAGCTATG
      AACCTCGCTT GCTGGATGTG GCTTGACTCT ATGGATGTCG CACTCGATAC
```

Figure 40F

```
6201  AGAAAGCGCC ACGCTTCCCG AAGGGAGAAA GGCGGACAGG TATCCGGTAA
      TCTTTCGCGG TGCGAAGGGC TTCCCTCTTT CCGCCTGTCC ATAGGCCATT
6251  GCGGCAGGGT CGGAACAGGA GAGCGCACGA GGGAGCTTCC AGGGGGAAAC
      CGCCGTCCCA GCCTTGTCCT CTCGCGTGCT CCCTCGAAGG TCCCCCTTTG
6301  GCCTGGTATC TTTATAGTCC TGTCGGGTTT CGCCACCTCT GACTTGAGCG
      CGGACCATAG AAATATCAGG ACAGCCCAAA GCGGTGGAGA CTGAACTCGC
6351  TCGATTTTTG TGATGCTCGT CAGGGGGGCG GAGCCTATGG AAAAACGCCA
      AGCTAAAAAC ACTACGAGCA GTCCCCCCGC CTCGGATACC TTTTTGCGGT
6401  GCAACGCGGC CTTTTTACGG TTCCTGGCCT TTTGCTGGCC TTTTGCTCAC
      CGTTGCGCCG GAAAAATGCC AAGGACCGGA AAACGACCGG AAAACGAGTG
6451  ATGTTCTTTC CTGCGTTATC CCCTGATTCT GTGGATAACC GTATTACCGC
      TACAAGAAAG GACGCAATAG GGGACTAAGA CACCTATTGG CATAATGGCG
6501  CTTTGAGTGA GCTGATACCG CTCGCCGCAG CCGAACGACC GAGCGCAGCG
      GAAACTCACT CGACTATGGC GAGCGGCGTC GGCTTGCTGG CTCGCGTCGC
6551  AGTCAGTGAG CGAGGAAGCG GAAGAGCGCC TGATGCGGTA TTTTCTCCTT
      TCAGTCACTC GCTCCTTCGC CTTCTCGCGG ACTACGCCAT AAAAGAGGAA
6601  ACGCATCTGT GCGGTATTTC ACACCGCATA TGGTGCACTC TCAGTACAAT
      TGCGTAGACA CGCCATAAAG TGTGGCGTAT ACCACGTGAG AGTCATGTTA
6651  CTGCTCTGAT GCCGCATAGT TAAGCCAGTA TACACTCCGC TATCGCTACG
      GACGAGACTA CGGCGTATCA ATTCGGTCAT ATGTGAGGCG ATAGCGATGC
6701  TGACTGGGTC ATGGCTGCGC CCCGACACCC GCCAACACCC GCTGACGCGC
      ACTGACCCAG TACCGACGCG GGGCTGTGGG CGGTTGTGGG CGACTGCGCG
                                                        Esp3I
                                                        ~
6751  CCTGACGGGC TTGTCTGCTC CCGGCATCCG CTTACAGACA AGCTGTGACC
      GGACTGCCCG AACAGACGAG GGCCGTAGGC GAATGTCTGT TCGACACTGG
      Esp3I
      ~~~~~
6801  GTCTCCGGGA GCTGCATGTG TCAGAGGTTT TCACCGTCAT CACCGAAACG
      CAGAGGCCCT CGACGTACAC AGTCTCCAAA AGTGGCAGTA GTGGCTTTGC
6851  CGCGAGGCAG GGATC
      GCGCTCCGTC CCTAG
```

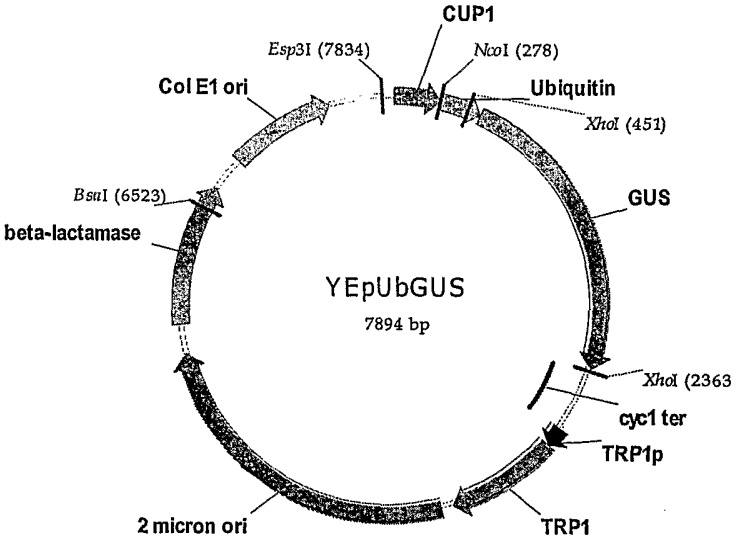

CUP1
Esp3I (7834)
NcoI (278)
Col E1 ori
Ubiquitin
XhoI (451)
GUS
BsaI (6523)
beta-lactamase
YEpUbGUS
7894 bp
XhoI (2363
cyc1 ter
TRP1p
2 micron ori
TRP1

Figure 41

Figure 42A

```
1     CCTTGTTACT AGTTAGAAAA AGACATTTTT GCTGTCAGTC ACTGTCAAGA
      GGAACAATGA TCAATCTTTT TCTGTAAAAA CGACAGTCAG TGACAGTTCT
 51   GATTCTTTTG CTGGCATTTC TTCTAGAAGC AAAAAGAGCG ATGCGTCTTT
      CTAAGAAAAC GACCGTAAAG AAGATCTTCG TTTTTCTCGC TACGCAGAAA
101   TCCGCTGAAC CGTTCCAGCA AAAAAGACTA CCAACGCAAT ATGGATTGTC
      AGGCGACTTG GCAAGGTCGT TTTTTCTGAT GGTTGCGTTA TACCTAACAG
151   AGAATCATAT AAAAGAGAAG CAAATAACTC CTTGTCTTGT ATCAATTGCA
      TCTTAGTATA TTTTCTCTTC GTTTATTGAG GAACAGAACA TAGTTAACGT
201   TTATAATATC TTCTTGTTAG TGCAATATCA TATAGAAGTC ATCGAAATAG
      AATATTATAG AAGAACAATC ACGTTATAGT ATATCTTCAG TAGCTTTATC
                                  NcoI
                                  ~~~~~~~
251   ATATTAAGAA AAACAAACTG TACAATCCAT GGGTCATCAC CATCATCATC
      TATAATTCTT TTTGTTTGAC ATGTTAGGTA CCCAGTAGTG GTAGTAGTAG
301   ACGGGCAGAT CTTCGTCAAG ACGTTAACCG GTAAAACCAT AACTCTAGAA
      TGCCCGTCTA GAAGCAGTTC TGCAATTGGC CATTTTGGTA TTGAGATCTT
351   GTTGAACCAT CCGATACCAT CGAAAACGTT AAGGCTAAAA TTCAAGACAA
      CAACTTGGTA GGCTATGGTA GCTTTTGCAA TTCCGATTTT AAGTTCTGTT
                                                        XhoI
                                                        ~
401   GGAAGGCATT CCACCTGATC AACAAAGATT GATCTTTGCC GGTAAGCAGC
      CCTTCCGTAA GGTGGACTAG TTGTTTCTAA CTAGAAACGG CCATTCGTCG
      XhoI
      ~~~~~
451   TCGAGGACGG TAGAACGCTG TCTGATTACA ACATTCAGAA GGAGTCGACC
      AGCTCCTGCC ATCTTGCGAC AGACTAATGT TGTAAGTCTT CCTCAGCTGG
501   TTACATCTTG TCTTACGCCT ACGTGGAGGT ATGGAATTCA TGTTACGTCC
      AATGTAGAAC AGAATGCGGA TGCACCTCCA TACCTTAAGT ACAATGCAGG
551   TGTAGAAACC CCAACCCGTG AAATCAAAAA ACTCGACGGC CTGTGGGCAT
      ACATCTTTGG GGTTGGGCAC TTTAGTTTTT TGAGCTGCCG GACACCCGTA
601   TCAGTCTGGA TCGCGAAAAC TGTGGAATTG ATCAGCGTTG GTGGGAAAGC
      AGTCAGACCT AGCGCTTTTG ACACCTTAAC TAGTCGCAAC CACCCTTTCG
651   GCGTTACAAG AAAGCCGGGC AATTGCTGTG CCAGGCAGTT TTAACGATCA
      CGCAATGTTC TTTCGGCCCG TTAACGACAC GGTCCGTCAA AATTGCTAGT
701   GTTCGCCGAT GCAGATATTC GTAATTATGC GGGCAACGTC TGGTATCAGC
      CAAGCGGCTA CGTCTATAAG CATTAATACG CCCGTTGCAG ACCATAGTCG
751   GCGAAGTCTT TATACCGAAA GGTTGGGCAG GCCAGCGTAT CGTGCTGCGT
      CGCTTCAGAA ATATGGCTTT CCAACCCGTC CGGTCGCATA GCACGACGCA
801   TTCGATGCGG TCACTCATTA CGGCAAAGTG TGGGTCAATA ATCAGGAAGT
      AAGCTACGCC AGTGAGTAAT GCCGTTTCAC ACCCAGTTAT TAGTCCTTCA
851   GATGGAGCAT CAGGGCGGCT ATACGCCATT TGAAGCCGAT GTCACGCCGT
      CTACCTCGTA GTCCCGCCGA TATGCGGTAA ACTTCGGCTA CAGTGCGGCA
901   ATGTTATTGC CGGGAAAAGT GTACGTATCA CCGTTTGTGT GAACAACGAA
      TACAATAACG GCCCTTTTCA CATGCATAGT GGCAAACACA CTTGTTGCTT
951   CTGAACTGGC AGACTATCCC GCCGGGAATG GTGATTACCG ACGAAAACGG
      GACTTGACCG TCTGATAGGG CGGCCCTTAC CACTAATGGC TGCTTTTGCC
1001  CAAGAAAAAG CAGTCTTACT TCCATGATTT CTTTAACTAT GCCGGAATCC
      GTTCTTTTTC GTCAGAATGA AGGTACTAAA GAAATTGATA CGGCCTTAGG
1051  ATCGCAGCGT AATGCTCTAC ACCACGCCGA ACACCTGGGT GGACGATATC
      TAGCGTCGCA TTACGAGATG TGGTGCGGCT TGTGGACCCA CCTGCTATAG
```

Figure 42B

```
1101  ACCGTGGTGA CGCATGTCGC GCAAGACTGT AACCACGCGT CTGTTGACTG
      TGGCACCACT GCGTACAGCG CGTTCTGACA TTGGTGCGCA GACAACTGAC
1151  GCAGGTGGTG GCCAATGGTG ATGTCAGCGT TGAACTGCGT GATGCGGATC
      CGTCCACCAC CGGTTACCAC TACAGTCGCA ACTTGACGCA CTACGCCTAG
1201  AACAGGTGGT TGCAACTGGA CAAGGCACTA GCGGGACTTT GCAAGTGGTG
      TTGTCCACCA ACGTTGACCT GTTCCGTGAT CGCCCTGAAA CGTTCACCAC
1251  AATCCGCACC TCTGGCAACC GGGTGAAGGT TATCTCTATG AACTGTGCGT
      TTAGGCGTGG AGACCGTTGG CCCACTTCCA ATAGAGATAC TTGACACGCA
1301  CACAGCCAAA AGCCAGACAG AGTGTGATAT CTACCCGCTT CGCGTCGGCA
      GTGTCGGTTT TCGGTCTGTC TCACACTATA GATGGGCGAA GCGCAGCCGT
1351  TCCGGTCAGT GGCAGTGAAG GGCCAACAGT TCCTGATTAA CCACAAACCG
      AGGCCAGTCA CCGTCACTTC CCGGTTGTCA AGGACTAATT GGTGTTTGGC
1401  TTCTACTTTA CTGGCTTTGG TCGTCATGAA GATGCGGACT TACGTGGCAA
      AAGATGAAAT GACCGAAACC AGCAGTACTT CTACGCCTGA ATGCACCGTT
1451  AGGATTCGAT AACGTGCTGA TGGTGCACGA CCACGCATTA ATGGACTGGA
      TCCTAAGCTA TTGCACGACT ACCACGTGCT GGTGCGTAAT TACCTGACCT
1501  TTGGGGCCAA CTCCTACCGT ACCTCGCATT ACCCTTACGC TGAAGAGATG
      AACCCCGGTT GAGGATGGCA TGGAGCGTAA TGGGAATGCG ACTTCTCTAC
1551  CTCGACTGGG CAGATGAACA TGGCATCGTG GTGATTGATG AAACTGCTGC
      GAGCTGACCC GTCTACTTGT ACCGTAGCAC CACTAACTAC TTTGACGACG
1601  TGTCGGCTTT AACCTCTCTT TAGGCATTGG TTTCGAAGCG GGCAACAAGC
      ACAGCCGAAA TTGGAGAGAA ATCCGTAACC AAAGCTTCGC CCGTTGTTCG
1651  CGAAAGAACT GTACAGCGAA GAGGCAGTCA ACGGGGAAAC TCAGCAAGCG
      GCTTTCTTGA CATGTCGCTT CTCCGTCAGT TGCCCCTTTG AGTCGTTCGC
1701  CACTTACAGG CGATTAAAGA GCTGATAGCG CGTGACAAAA ACCACCCAAG
      GTGAATGTCC GCTAATTTCT CGACTATCGC GCACTGTTTT TGGTGGGTTC
1751  CGTGGTGATG TGGAGTATTG CCAACGAACC GGATACCCGT CCGCAAGTGC
      GCACCACTAC ACCTCATAAC GGTTGCTTGG CCTATGGGCA GGCGTTCACG
1801  ACGGGAATAT TTCGCCACTG GCGGAAGCAA CGCGTAAACT CGACCCGACG
      TGCCCTTATA AAGCGGTGAC CGCCTTCGTT GCGCATTTGA GCTGGGCTGC
1851  CGTCCGATCA CCTGCGTCAA TGTAATGTTC TGCGACGCTC ACACCGATAC
      GCAGGCTAGT GGACGCAGTT ACATTACAAG ACGCTGCGAG TGTGGCTATG
1901  CATCAGCGAT CTCTTTGATG TGCTGTGCCT GAACCGTTAT TACGGATGGT
      GTAGTCGCTA GAGAAACTAC ACGACACGGA CTTGGCAATA ATGCCTACCA
1951  ATGTCCAAAG CGGCGATTTG GAAACGGCAG AGAAGGTACT GGAAAAAGAA
      TACAGGTTTC GCCGCTAAAC CTTTGCCGTC TCTTCCATGA CCTTTTTCTT
2001  CTTCTGGCCT GGCAGGAGAA ACTGCATCAG CCGATTATCA TCACCGAATA
      GAAGACCGGA CCGTCCTCTT TGACGTAGTC GGCTAATAGT AGTGGCTTAT
2051  CGGCGTGGAT ACGTTAGCCG GGCTGCACTC AATGTACACC GACATGTGGA
      GCCGCACCTA TGCAATCGGC CCGACGTGAG TTACATGTGG CTGTACACCT
2101  GTGAAGAGTA TCAGTGTGCA TGGCTGGATA TGTATCACCG CGTCTTTGAT
      CACTTCTCAT AGTCACACGT ACCGACCTAT ACATAGTGGC GCAGAAACTA
2151  CGCGTCAGCG CCGTCGTCGG TGAACAGGTA TGGAATTTCG CCGATTTTGC
      GCGCAGTCGC GGCAGCAGCC ACTTGTCCAT ACCTTAAAGC GGCTAAAACG
2201  GACCTCGCAA GGCATATTGC GCGTTGGCGG TAACAAGAAA GGGATCTTCA
      CTGGAGCGTT CCGTATAACG CGCAACCGCC ATTGTTCTTT CCCTAGAAGT
2251  CTCGCGACCG CAAACCGAAG TCGGCGGCTT TTCTGCTGCA AAAACGCTGG
      GAGCGCTGGC GTTTGGCTTC AGCCGCCGAA AAGACGACGT TTTTGCGACC
2301  ACTGGCATGA ACTTCGGTGA AAAACCGCAG CAGGGAGGCA AACAATAAGC
      TGACCGTACT TGAAGCCACT TTTTGGCGTC GTCCCTCCGT TTGTTATTCG
                  XhoI
```

Figure 42C

```
                    ~~~~~~
2351  TTGCGGCCGC ACTCGAGGAG CTCCCTGGCG AATTGTACCA AGATGGCCTT
      AACGCCGGCG TGAGCTCCTC GAGGGACCGC TTAACATGGT TCTACCGGAA
2401  TGGTGGGTTG AAGAAGGAAA AAGACAGAAA CGACTTAATT ACCTACTTGA
      ACCACCCAAC TTCTTCCTTT TTCTGTCTTT GCTGAATTAA TGGATGAACT
2451  AAAAAGCCTG TGAGTAAACA GGCCCCTTTT CCTTTGTCGA TATCATGTAA
      TTTTTCGGAC ACTCATTTGT CCGGGGAAAA GGAAACAGCT ATAGTACATT
2501  TTAGTTATGT CACGCTTACA TTCACGCCCT CCCCCCACAT CCGCTCTAAC
      AATCAATACA GTGCGAATGT AAGTGCGGGA GGGGGGTGTA GGCGAGATTG
2551  CGAAAAGGAA GGAGTTAGAC AACCTGAAGT CTAGGTCCCT ATTTATTTTT
      GCTTTTCCTT CCTCAATCTG TTGGACTTCA GATCCAGGGA TAAATAAAAA
2601  TTATAGTTAT GTTAGTATTA AGAACGTTAT TTATATTTCA AATTTTTCTT
      AATATCAATA CAATCATAAT TCTTGCAATA AATATAAAGT TTAAAAAGAA
2651  TTTTTTCTGT ACAGACGCGT GTACGCATGT AACATTATAC TGAAAACCTT
      AAAAAAGACA TGTCTGCGCA CATGCGTACA TTGTAATATG ACTTTTGGAA
2701  GCTTGAGAAG GTTTTGGGAC GCTCGAAGGC TTTAATTTGC AAGCTTATCG
      CGAACTCTTC CAAAACCCTG CGAGCTTCCG AAATTAAACG TTCGAATAGC
2751  ATGATAAGCT GTCAAACATG AGAATTCGGT CGAAAAAAGA AAAGGAGAGG
      TACTATTCGA CAGTTTGTAC TCTTAAGCCA GCTTTTTTCT TTTCCTCTCC
2801  GCCAAGAGGG AGGGCATTGG TGACTATTGA GCACGTGAGT ATACGTGATT
      CGGTTCTCCC TCCCGTAACC ACTGATAACT CGTGCACTCA TATGCACTAA
2851  AAGCACACAA AGGCAGCTTG GAGTATGTCT GTTATTAATT TCACAGGTAG
      TTCGTGTGTT TCCGTCGAAC CTCATACAGA CAATAATTAA AGTGTCCATC
2901  TTCTGGTCCA TTGGTGAAAG TTTGCGGCTT GCAGAGCACA GAGGCCGCAG
      AAGACCAGGT AACCACTTTC AAACGCCGAA CGTCTCGTGT CTCCGGCGTC
2951  AATGTGCTCT AGATTCCGAT GCTGACTTGC TGGGTATTAT ATGTGTGCCC
      TTACACGAGA TCTAAGGCTA CGACTGAACG ACCCATAATA TACACACGGG
3001  AATAGAAAGA GAACAATTGA CCCGGTTATT GCAAGGAAAA TTTCAAGTCT
      TTATCTTTCT CTTGTTAACT GGGCCAATAA CGTTCCTTTT AAAGTTCAGA
3051  TGTAAAAGCA TATAAAAATA GTTCAGGCAC TCCGAAATAC TTGGTTGGCG
      ACATTTTCGT ATATTTTTAT CAAGTCCGTG AGGCTTTATG AACCAACCGC
3101  TGTTTCGTAA TCAACCTAAG GAGGATGTTT TGGCTCTGGT CAATGATTAC
      ACAAAGCATT AGTTGGATTC CTCCTACAAA ACCGAGACCA GTTACTAATG
3151  GGCATTGATA TCGTCCAACT GCATGGAGAT GAGTCGTGGC AAGAATACCA
      CCGTAACTAT AGCAGGTTGA CGTACCTCTA CTCAGCACCG TTCTTATGGT
3201  AGAGTTCCTC GGTTTGCCAG TTATTAAAAG ACTCGTATTT CCAAAAGACT
      TCTCAAGGAG CCAAACGGTC AATAATTTTC TGAGCATAAA GGTTTTCTGA
3251  GCAACATACT ACTCAGTGCA GCTTCACAGA AACCTCATTC GTTTATTCCC
      CGTTGTATGA TGAGTCACGT CGAAGTGTCT TTGGAGTAAG CAAATAAGGG
3301  TTGTTTGATT CAGAAGCAGG TGGGACAGGT GAACTTTTGG ATTGGAACTC
      AACAAACTAA GTCTTCGTCC ACCCTGTCCA CTTGAAAACC TAACCTTGAG
3351  GATTTCTGAC TGGGTTGGAA GGCAAGAGAG CCCCGAAAGC TTACATTTTA
      CTAAAGACTG ACCCAACCTT CCGTTCTCTC GGGGCTTTCG AATGTAAAAT
3401  TGTTAGCTGG TGGACTGACG CCAGAAAATG TTGGTGATGC GCTTAGATTA
      ACAATCGACC ACCTGACTGC GGTCTTTTAC AACCACTACG CGAATCTAAT
3451  AATGGCGTTA TTGGTGTTGA TGTAAGCGGA GGTGTGGAGA CAAATGGTGT
      TTACCGCAAT AACCACAACT ACATTCGCCT CCACACCTCT GTTTACCACA
3501  AAAAGACTCT AACAAAATAG CAAATTTCGT CAAAAATGCT AAGAAATAGG
      TTTTCTGAGA TTGTTTTATC GTTTAAAGCA GTTTTTACGA TTCTTTATCC
3551  TTATTACTGA GTAGTATTTA TTTAAGTATT GTTTGTGCAC TTGCCTGCAG
      AATAATGACT CATCATAAAT AAATTCATAA CAAACACGTG AACGGACGTC
```

Figure 42D

```
3601  CTTCTCAATG ATATTCGAAT ACGCTTTGAG GAGATACAGC CTAATATCCG
      GAAGAGTTAC TATAAGCTTA TGCGAAACTC CTCTATGTCG GATTATAGGC
3651  ACAAACTGTT TTACAGATTT ACGATCGTAC TTGTTACCCA TCATTGAATT
      TGTTTGACAA AATGTCTAAA TGCTAGCATG AACAATGGGT AGTAACTTAA
3701  TTGAACATCC GAACCTGGGA GTTTTCCCTG AAACAGATAG TATATTTGAA
      AACTTGTAGG CTTGGACCCT CAAAAGGGAC TTTGTCTATC ATATAAACTT
3751  CCTGTATAAT AATATATAGT CTAGCGCTTT ACGGAAGACA ATGTATGTAT
      GGACATATTA TTATATATCA GATCGCGAAA TGCCTTCTGT TACATACATA
3801  TTCGGTTCCT GGAGAAACTA TTGCATCTAT TGCATAGGTA ATCTTGCACG
      AAGCCAAGGA CCTCTTTGAT AACGTAGATA ACGTATCCAT TAGAACGTGC
3851  TCGCATCCCC GGTTCATTTT CTGCGTTTCC ATCTTGCACT TCAATAGCAT
      AGCGTAGGGG CCAAGTAAAA GACGCAAAGG TAGAACGTGA AGTTATCGTA
3901  ATCTTTGTTA ACGAAGCATC TGTGCTTCAT TTTGTAGAAC AAAAATGCAA
      TAGAAACAAT TGCTTCGTAG ACACGAAGTA AAACATCTTG TTTTTACGTT
3951  CGCGAGAGCG CTAATTTTTC AAACAAAGAA TCTGAGCTGC ATTTTTACAG
      GCGCTCTCGC GATTAAAAAG TTTGTTTCTT AGACTCGACG TAAAAATGTC
4001  AACAGAAATG CAACGCGAAA GCGCTATTTT ACCAACGAAG AATCTGTGCT
      TTGTCTTTAC GTTGCGCTTT CGCGATAAAA TGGTTGCTTC TTAGACACGA
4051  TCATTTTTGT AAAACAAAAA TGCAACGCGA GAGCGCTAAT TTTTCAAACA
      AGTAAAAACA TTTTGTTTTT ACGTTGCGCT CTCGCGATTA AAAAGTTTGT
4101  AAGAATCTGA GCTGCATTTT TACAGAACAG AAATGCAACG CGAGAGCGCT
      TTCTTAGACT CGACGTAAAA ATGTCTTGTC TTTACGTTGC GCTCTCGCGA
4151  ATTTTACCAA CAAAGAATCT ATACTTCTTT TTTGTTCTAC AAAAATGCAT
      TAAAATGGTT GTTTCTTAGA TATGAAGAAA AAACAAGATG TTTTTACGTA
4201  CCCGAGAGCG CTATTTTTCT AACAAAGCAT CTTAGATTAC TTTTTTTCTC
      GGGCTCTCGC GATAAAAAGA TTGTTTCGTA GAATCTAATG AAAAAAAGAG
4251  CTTTGTGCGC TCTATAATGC AGTCTCTTGA TAACTTTTTG CACTGTAGGT
      GAAACACGCG AGATATTACG TCAGAGAACT ATTGAAAAAC GTGACATCCA
4301  CCGTTAAGGT TAGAAGAAGG CTACTTTGGT GTCTATTTTC TCTTCCATAA
      GGCAATTCCA ATCTTCTTCC GATGAAACCA CAGATAAAAG AGAAGGTATT
4351  AAAAAGCCTG ACTCCACTTC CCGCGTTTAC TGATTACTAG CGAAGCTGCG
      TTTTTCGGAC TGAGGTGAAG GGCGCAAATG ACTAATGATC GCTTCGACGC
4401  GGTGCATTTT TTCAAGATAA AGGCATCCCC GATTATATTC TATACCGATG
      CCACGTAAAA AAGTTCTATT TCCGTAGGGG CTAATATAAG ATATGGCTAC
4451  TGGATTGCGC ATACTTTGTG AACAGAAAGT GATAGCGTTG ATGATTCTTC
      ACCTAACGCG TATGAAACAC TTGTCTTTCA CTATCGCAAC TACTAAGAAG
4501  ATTGGTCAGA AAATTATGAA CGGTTTCTTC TATTTTGTCT CTATATACTA
      TAACCAGTCT TTTAATACTT GCCAAAGAAG ATAAAACAGA GATATATGAT
4551  CGTATAGGAA ATGTTTACAT TTTCGTATTG TTTTCGATTC ACTCTATGAA
      GCATATCCTT TACAAATGTA AAAGCATAAC AAAAGCTAAG TGAGATACTT
4601  TAGTTCTTAC TACAATTTTT TTGTCTAAAG AGTAATACTA GAGATAAACA
      ATCAAGAATG ATGTTAAAAA AACAGATTTC TCATTATGAT CTCTATTTGT
4651  TAAAAAATGT AGAGGTCGAG TTTAGATGCA AGTTCAAGGA GCGAAAGGTG
      ATTTTTTACA TCTCCAGCTC AAATCTACGT TCAAGTTCCT CGCTTTCCAC
4701  GATGGGTAGG TTATATAGGG ATATAGCACA GAGATATATA GCAAAGAGAT
      CTACCCATCC AATATATCCC TATATCGTGT CTCTATATAT CGTTTCTCTA
4751  ACTTTTGAGC AATGTTTGTG GAAGCGGTAT TCGCAATATT TTAGTAGCTC
      TGAAAACTCG TTACAAACAC CTTCGCCATA AGCGTTATAA AATCATCGAG
4801  GTTACAGTCC GGTGCGTTTT TGGTTTTTTG AAAGTGCGTC TTCAGAGCGC
      CAATGTCAGG CCACGCAAAA ACCAAAAAAC TTTCACGCAG AAGTCTCGCG
4851  TTTTGGTTTT CAAAAGCGCT CTGAAGTTCC TATACTTTCT AGAGAATAGG
```

Figure 42E

```
      AAAACCAAAA GTTTTCGCGA GACTTCAAGG ATATGAAAGA TCTCTTATCC
4901  AACTTCGGAA TAGGAACTTC AAAGCGTTTC CGAAAACGAG CGCTTCCGAA
      TTGAAGCCTT ATCCTTGAAG TTTCGCAAAG GCTTTTGCTC GCGAAGGCTT
4951  AATGCAACGC GAGCTGCGCA CATACAGCTC ACTGTTCACG TCGCACCTAT
      TTACGTTGCG CTCGACGCGT GTATGTCGAG TGACAAGTGC AGCGTGGATA
5001  ATCTGCGTGT TGCCTGTATA TATATATACA TGAGAAGAAC GGCATAGTGC
      TAGACGCACA ACGGACATAT ATATATATGT ACTCTTCTTG CCGTATCACG
5051  GTGTTTATGC TTAAATGCGT ACTTATATGC GTCTATTTAT GTAGGATGAA
      CACAAATACG AATTTACGCA TGAATATACG CAGATAAATA CATCCTACTT
5101  AGGTAGTCTA GTACCTCCTG TGATATTATC CCATTCCATG CGGGGTATCG
      TCCATCAGAT CATGGAGGAC ACTATAATAG GGTAAGGTAC GCCCCATAGC
5151  TATGCTTCCT TCAGCACTAC CCTTTAGCTG TTCTATATGC TGCCACTCCT
      ATACGAAGGA AGTCGTGATG GGAAATCGAC AAGATATACG ACGGTGAGGA
5201  CAATTGGATT AGTCTCATCC TTCAATGCTA TCATTTCCTT TGATATTGGA
      GTTAACCTAA TCAGAGTAGG AAGTTACGAT AGTAAAGGAA ACTATAACCT
5251  TCATATGCAT AGTACCGAGA AACTAGTGCG AAGTAGTGAT CAGGTATTGC
      AGTATACGTA TCATGGCTCT TTGATCACGC TTCATCACTA GTCCATAACG
5301  TGTTATCTGA TGAGTATACG TTGTCCTGGC CACGGCAGAA GCACGCTTAT
      ACAATAGACT ACTCATATGC AACAGGACCG GTGCCGTCTT CGTGCGAATA
5351  CGCTCCAATT TCCCACAACA TTAGTCAACT CCGTTAGGCC CTTCATTGAA
      GCGAGGTTAA AGGGTGTTGT AATCAGTTGA GGCAATCCGG GAAGTAACTT
5401  AGAAATGAGG TCATCAAATG TCTTCCAATG TGAGATTTTG GGCCATTTTT
      TCTTTACTCC AGTAGTTTAC AGAAGGTTAC ACTCTAAAAC CCGGTAAAAA
5451  TATAGCAAAG ATTGAATAAG GCGCATTTTT CTTCAAAGCT TTATTGTACG
      ATATCGTTTC TAACTTATTC CGCGTAAAAA GAAGTTTCGA AATAACATGC
5501  ATCTGACTAA GTTATCTTTT AATAATTGGT ATTCCTGTTT ATTGCTTGAA
      TAGACTGATT CAATAGAAAA TTATTAACCA TAAGGACAAA TAACGAACTT
5551  GAATTGCCGG TCCTATTTAC TCGTTTTAGG ACTGGTTCAG AATTCTTGAA
      CTTAACGGCC AGGATAAATG AGCAAAATCC TGACCAAGTC TTAAGAACTT
5601  GACGAAAGGG CCTCGTGATA CGCCTATTTT TATAGGTTAA TGTCATGATA
      CTGCTTTCCC GGAGCACTAT GCGGATAAAA ATATCCAATT ACAGTACTAT
5651  ATAATGGTTT CTTAGACGTC AGGTGGCACT TTTCGGGGAA ATGTGCGCGG
      TATTACCAAA GAATCTGCAG TCCACCGTGA AAAGCCCCTT TACACGCGCC
5701  AACCCCTATT TGTTTATTTT TCTAAATACA TTCAAATATG TATCCGCTCA
      TTGGGGATAA ACAAATAAAA AGATTTATGT AAGTTTATAC ATAGGCGAGT
5751  TGAGACAATA ACCCTGATAA ATGCTTCAAT AATATTGAAA AAGGAAGAGT
      ACTCTGTTAT TGGGACTATT TACGAAGTTA TTATAACTTT TTCCTTCTCA
5801  ATGAGTATTC AACATTTCCG TGTCGCCCTT ATTCCCTTTT TTGCGGCATT
      TACTCATAAG TTGTAAAGGC ACAGCGGGAA TAAGGGAAAA AACGCCGTAA
5851  TTGCCTTCCT GTTTTTGCTC ACCCAGAAAC GCTGGTGAAA GTAAAAGATG
      AACGGAAGGA CAAAAACGAG TGGGTCTTTG CGACCACTTT CATTTTCTAC
5901  CTGAAGATCA GTTGGGTGCA CGAGTGGGTT ACATCGAACT GGATCTCAAC
      GACTTCTAGT CAACCCACGT GCTCACCCAA TGTAGCTTGA CCTAGAGTTG
5951  AGCGGTAAGA TCCTTGAGAG TTTTCGCCCC GAAGAACGTT TTCCAATGAT
      TCGCCATTCT AGGAACTCTC AAAAGCGGGG CTTCTTGCAA AAGGTTACTA
6001  GAGCACTTTT AAAGTTCTGC TATGTGGCGC GGTATTATCC CGTGTTGACG
      CTCGTGAAAA TTTCAAGACG ATACACCGCG CCATAATAGG GCACAACTGC
6051  CCGGGCAAGA GCAACTCGGT CGCCGCATAC ACTATTCTCA GAATGACTTG
      GGCCCGTTCT CGTTGAGCCA GCGGCGTATG TGATAAGAGT CTTACTGAAC
6101  GTTGAGTACT CACCAGTCAC AGAAAAGCAT CTTACGGATG GCATGACAGT
      CAACTCATGA GTGGTCAGTG TCTTTTCGTA GAATGCCTAC CGTACTGTCA
```

Figure 42F

```
6151  AAGAGAATTA TGCAGTGCTG CCATAACCAT GAGTGATAAC ACTGCGGCCA
      TTCTCTTAAT ACGTCACGAC GGTATTGGTA CTCACTATTG TGACGCCGGT
6201  ACTTACTTCT GACAACGATC GGAGGACCGA AGGAGCTAAC CGCTTTTTTG
      TGAATGAAGA CTGTTGCTAG CCTCCTGGCT TCCTCGATTG GCGAAAAAAC
6251  CACAACATGG GGGATCATGT AACTCGCCTT GATCGTTGGG AACCGGAGCT
      GTGTTGTACC CCCTAGTACA TTGAGCGGAA CTAGCAACCC TTGGCCTCGA
6301  GAATGAAGCC ATACCAAACG ACGAGCGTGA CACCACGATG CCTGCAGCAA
      CTTACTTCGG TATGGTTTGC TGCTCGCACT GTGGTGCTAC GGACGTCGTT
6351  TGGCAACAAC GTTGCGCAAA CTATTAACTG GCGAACTACT TACTCTAGCT
      ACCGTTGTTG CAACGCGTTT GATAATTGAC CGCTTGATGA ATGAGATCGA
6401  TCCCGGCAAC AATTAATAGA CTGGATGGAG GCGGATAAAG TTGCAGGACC
      AGGGCCGTTG TTAATTATCT GACCTACCTC CGCCTATTTC AACGTCCTGG
6451  ACTTCTGCGC TCGGCCCTTC CGGCTGGCTG GTTTATTGCT GATAAATCTG
      TGAAGACGCG AGCCGGGAAG GCCGACCGAC CAAATAACGA CTATTTAGAC
                          BsaI
                         ~~~~~~~
6501  GAGCCGGTGA GCGTGGGTCT CGCGGTATCA TTGCAGCACT GGGGCCAGAT
      CTCGGCCACT CGCACCCAGA GCGCCATAGT AACGTCGTGA CCCCGGTCTA
6551  GGTAAGCCCT CCCGTATCGT AGTTATCTAC ACGACGGGGA GTCAGGCAAC
      CCATTCGGGA GGGCATAGCA TCAATAGATG TGCTGCCCCT CAGTCCGTTG
6601  TATGGATGAA CGAAATAGAC AGATCGCTGA GATAGGTGCC TCACTGATTA
      ATACCTACTT GCTTTATCTG TCTAGCGACT CTATCCACGG AGTGACTAAT
6651  AGCATTGGTA ACTGTCAGAC CAAGTTTACT CATATATACT TTAGATTGAT
      TCGTAACCAT TGACAGTCTG GTTCAAATGA GTATATATGA AATCTAACTA
6701  TTAAAACTTC ATTTTTAATT TAAAAGGATC TAGGTGAAGA TCCTTTTTGA
      AATTTTGAAG TAAAAATTAA ATTTTCCTAG ATCCACTTCT AGGAAAAACT
6751  TAATCTCATG ACCAAAATCC CTTAACGTGA GTTTTCGTTC CACTGAGCGT
      ATTAGAGTAC TGGTTTTAGG GAATTGCACT CAAAAGCAAG GTGACTCGCA
6801  CAGACCCCGT AGAAAAGATC AAAGGATCTT CTTGAGATCC TTTTTTTCTG
      GTCTGGGGCA TCTTTTCTAG TTTCCTAGAA GAACTCTAGG AAAAAAAGAC
6851  CGCGTAATCT GCTGCTTGCA AACAAAAAAA CCACCGCTAC CAGCGGTGGT
      GCGCATTAGA CGACGAACGT TTGTTTTTTT GGTGGCGATG GTCGCCACCA
6901  TTGTTTGCCG GATCAAGAGC TACCAACTCT TTTTCCGAAG GTAACTGGCT
      AACAAACGGC CTAGTTCTCG ATGGTTGAGA AAAAGGCTTC CATTGACCGA
6951  TCAGCAGAGC GCAGATACCA AATACTGTCC TTCTAGTGTA GCCGTAGTTA
      AGTCGTCTCG CGTCTATGGT TTATGACAGG AAGATCACAT CGGCATCAAT
7001  GGCCACCACT TCAAGAACTC TGTAGCACCG CCTACATACC TCGCTCTGCT
      CCGGTGGTGA AGTTCTTGAG ACATCGTGGC GGATGTATGG AGCGAGACGA
7051  AATCCTGTTA CCAGTGGCTG CTGCCAGTGG CGATAAGTCG TGTCTTACCG
      TTAGGACAAT GGTCACCGAC GACGGTCACC GCTATTCAGC ACAGAATGGC
7101  GGTTGGACTC AAGACGATAG TTACCGGATA AGGCGCAGCG GTCGGGCTGA
      CCAACCTGAG TTCTGCTATC AATGGCCTAT TCCGCGTCGC CAGCCCGACT
7151  ACGGGGGGTT CGTGCACACA GCCCAGCTTG GAGCGAACGA CCTACACCGA
      TGCCCCCCAA GCACGTGTGT CGGGTCGAAC CTCGCTTGCT GGATGTGGCT
7201  ACTGAGATAC CTACAGCGTG AGCTATGAGA AAGCGCCACG CTTCCCGAAG
      TGACTCTATG GATGTCGCAC TCGATACTCT TTCGCGGTGC GAAGGGCTTC
7251  GGAGAAAGGC GGACAGGTAT CCGGTAAGCG GCAGGGTCGG AACAGGAGAG
      CCTCTTTCCG CCTGTCCATA GGCCATTCGC CGTCCCAGCC TTGTCCTCTC
7301  CGCACGAGGG AGCTTCCAGG GGGAAACGCC TGGTATCTTT ATAGTCCTGT
      GCGTGCTCCC TCGAAGGTCC CCCTTTGCGG ACCATAGAAA TATCAGGACA
7351  CGGGTTTCGC CACCTCTGAC TTGAGCGTCG ATTTTTGTGA TGCTCGTCAG
```

Figure 42G

```
      GCCCAAAGCG GTGGAGACTG AACTCGCAGC TAAAAACACT ACGAGCAGTC
7401  GGGGGCGGAG CCTATGGAAA AACGCCAGCA ACGCGGCCTT TTTACGGTTC
      CCCCCGCCTC GGATACCTTT TTGCGGTCGT TGCGCCGGAA AAATGCCAAG
7451  CTGGCCTTTT GCTGGCCTTT TGCTCACATG TTCTTTCCTG CGTTATCCCC
      GACCGGAAAA CGACCGGAAA ACGAGTGTAC AAGAAAGGAC GCAATAGGGG
7501  TGATTCTGTG GATAACCGTA TTACCGCCTT TGAGTGAGCT GATACCGCTC
      ACTAAGACAC CTATTGGCAT AATGGCGGAA ACTCACTCGA CTATGGCGAG
7551  GCCGCAGCCG AACGACCGAG CGCAGCGAGT CAGTGAGCGA GGAAGCGGAA
      CGGCGTCGGC TTGCTGGCTC GCGTCGCTCA GTCACTCGCT CCTTCGCCTT
7601  GAGCGCCTGA TGCGGTATTT TCTCCTTACG CATCTGTGCG GTATTTCACA
      CTCGCGGACT ACGCCATAAA AGAGGAATGC GTAGACACGC CATAAAGTGT
7651  CCGCATATGG TGCACTCTCA GTACAATCTG CTCTGATGCC GCATAGTTAA
      GGCGTATACC ACGTGAGAGT CATGTTAGAC GAGACTACGG CGTATCAATT
7701  GCCAGTATAC ACTCCGCTAT CGCTACGTGA CTGGGTCATG GCTGCGCCCC
      CGGTCATATG TGAGGCGATA GCGATGCACT GACCCAGTAC CGACGCGGGG
7751  GACACCCGCC AACACCCGCT GACGCGCCCT GACGGGCTTG TCTGCTCCCG
      CTGTGGGCGG TTGTGGGCGA CTGCGCGGGA CTGCCCGAAC AGACGAGGGC
                                    Esp3I
                                    ~~~~~~~
7801  GCATCCGCTT ACAGACAAGC TGTGACCGTC TCCGGGAGCT GCATGTGTCA
      CGTAGGCGAA TGTCTGTTCG ACACTGGCAG AGGCCCTCGA CGTACACAGT
7851  GAGGTTTTCA CCGTCATCAC CGAAACGCGC GAGGCAGGGA TCCG
      CTCCAAAAGT GGCAGTAGTG GCTTTGCGCG CTCCGTCCCT AGGC
```

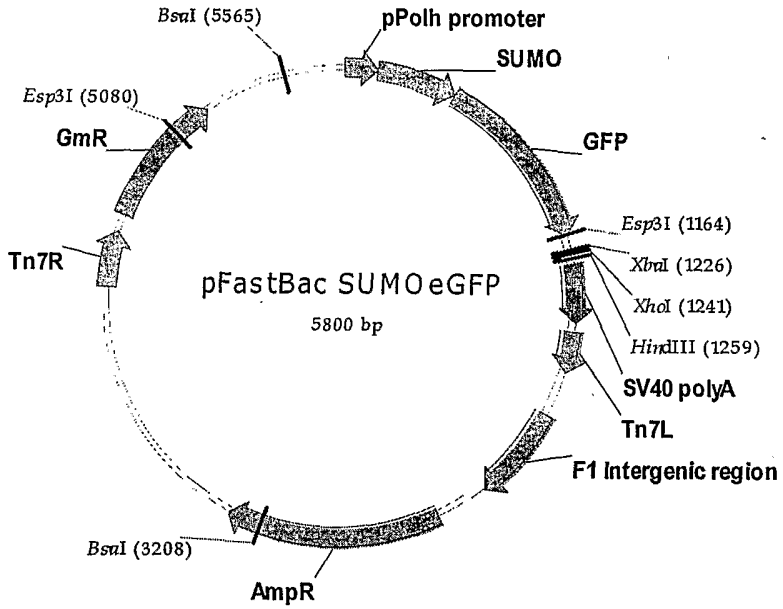

BsmI (5565)
pPolh promoter
SUMO
Esp3I (5080)
GmR
GFP
Esp3I (1164)
XbaI (1226)
XhoI (1241)
HindIII (1259)
Tn7R
pFastBac SUMOeGFP
5800 bp
SV40 polyA
Tn7L
F1 intergenic region
BsmI (3208)
AmpR

Figure 43

Figure 44A

```
   1  ATCATGGAGA TAATTAAAAT GATAACCATC TCGCAAATAA ATAAGTATTT
      TAGTACCTCT ATTAATTTTA CTATTGGTAG AGCGTTTATT TATTCATAAA
  51  TACTGTTTTC GTAACAGTTT TGTAATAAAA AAACCTATAA ATATTCCGGA
      ATGACAAAAG CATTGTCAAA ACATTATTTT TTTGGATATT TATAAGGCCT
 101  TTATTCATAC CGTCCCACCA TCGGGCGCGA TGGGTCATCA CCATCATCAT
      AATAAGTATG GCAGGGTGGT AGCCCGCGCT ACCCAGTAGT GGTAGTAGTA
 151  CACGGGTCGG ACTCAGAAGT CAATCAAGAA GCTAAGCCAG AGGTCAAGCC
      GTGCCCAGCC TGAGTCTTCA GTTAGTTCTT CGATTCGGTC TCCAGTTCGG
 201  AGAAGTCAAG CCTGAGACTC ACATCAATTT AAAGGTGTCC GATGGATCTT
      TCTTCAGTTC GGACTCTGAG TGTAGTTAAA TTTCCACAGG CTACCTAGAA
 251  CAGAGATCTT CTTCAAGATC AAAAAGACCA CTCCTTTAAG AAGGCTGATG
      GTCTCTAGAA GAAGTTCTAG TTTTTCTGGT GAGGAAATTC TTCCGACTAC
 301  GAAGCGTTCG CTAAAAGACA GGGTAAGGAA ATGGACTCCT TAAGATTCTT
      CTTCGCAAGC GATTTTCTGT CCCATTCCTT TACCTGAGGA ATTCTAAGAA
 351  GTACGACGGT ATTAGAATTC AAGCTGATCA GACCCCTGAA GATTTGGACA
      CATGCTGCCA TAATCTTAAG TTCGACTAGT CTGGGGACTT CTAAACCTGT
 401  TGGAGGATAA CGATATTATT GAGGCTCACC GCGAACAGAT TGGAGGTATG
      ACCTCCTATT GCTATAATAA CTCCGAGTGG CGCTTGTCTA ACCTCCATAC
 451  GTGAGCAAGG GCGAGGAGCT GTTCACCGGG GTGGTGCCCA TCCTGGTCGA
      CACTCGTTCC CGCTCCTCGA CAAGTGGCCC CACCACGGGT AGGACCAGCT
 501  GCTGGACGGC GACGTAAACG GCCACAAGTT CAGCGTGTCC GGCGAGGGCG
      CGACCTGCCG CTGCATTTGC CGGTGTTCAA GTCGCACAGG CCGCTCCCGC
 551  AGGGCGATGC CACCTACGGC AAGCTGACCC TGAAGTTCAT CTGCACCACC
      TCCCGCTACG GTGGATGCCG TTCGACTGGG ACTTCAAGTA GACGTGGTGG
 601  GGCAAGCTGC CCGTGCCCTG GCCCACCCTC GTGACCACCC TGACCTACGG
      CCGTTCGACG GGCACGGGAC CGGGTGGGAG CACTGGTGGG ACTGGATGCC
 651  CGTGCAGTGC TTCAGCCGCT ACCCCGACCA CATGAAGCAG CACGACTTCT
      GCACGTCACG AAGTCGGCGA TGGGGCTGGT GTACTTCGTC GTGCTGAAGA
 701  TCAAGTCCGC CATGCCCGAA GGCTACGTCC AGGAGCGCAC CATCTTCTTC
      AGTTCAGGCG GTACGGGCTT CCGATGCAGG TCCTCGCGTG GTAGAAGAAG
 751  AAGGACGACG GCAACTACAA GACCCGCGCC GAGGTGAAGT TCGAGGGCGA
      TTCCTGCTGC CGTTGATGTT CTGGGCGCGG CTCCACTTCA AGCTCCCGCT
 801  CACCCTGGTG AACCGCATCG AGCTGAAGGG CATCGACTTC AAGGAGGACG
      GTGGGACCAC TTGGCGTAGC TCGACTTCCC GTAGCTGAAG TTCCTCCTGC
 851  GCAACATCCT GGGGCACAAG CTGGAGTACA ACTACAACAG CCACAACGTC
      CGTTGTAGGA CCCCGTGTTC GACCTCATGT TGATGTTGTC GGTGTTGCAG
 901  TATATCATGG CCGACAAGCA GAAGAACGGC ATCAAGGTGA ACTTCAAGAT
      ATATAGTACC GGCTGTTCGT CTTCTTGCCG TAGTTCCACT TGAAGTTCTA
 951  CCGCCACAAC ATCGAGGACG GCAGCGTGCA GCTCGCCGAC CACTACCAGC
      GGCGGTGTTG TAGCTCCTGC CGTCGCACGT CGAGCGGCTG GTGATGGTCG
1001  AGAACACCCC CATCGGCGAC GGCCCCGTGC TGCTGCCCGA CAACCACTAC
      TCTTGTGGGG GTAGCCGCTG CCGGGGCACG ACGACGGGCT GTTGGTGATG
1051  CTGAGCACCC AGTCCGCCCT GAGCAAAGAC CCCAACGAGA AGCGCGATCA
      GACTCGTGGG TCAGGCGGGA CTCGTTTCTG GGGTTGCTCT TCGCGCTAGT
1101  CATGGTCCTG CTGGAGTTCG TGACCGCCGC CGGGATCACT CTCGGCATGG
      GTACCAGGAC GACCTCAAGC ACTGGCGGCG GCCCTAGTGA GAGCCGTACC
                          Esp3I
                          ~~~~~~~
1151  ACGAGCTGTA CAAGTAATGA GACGGAATTC AAAGGCCTAC GTCGACGAGC
      TGCTCGACAT GTTCATTACT CTGCCTTAAG TTTCCGGATG CAGCTGCTCG
```

Figure 44B

```
                                          XbaI                XhoI
                                          ~~~~~~~             ~~~~~~~
1201  TCACTAGTCG CGGCCGCTTT CGAATCTAGA GCCTGCAGTC TCGAGGCATG
      AGTGATCAGC GCCGGCGAAA GCTTAGATCT CGGACGTCAG AGCTCCGTAC
             HindIII
             ~~~~~~~
1251  CGGTACCAAG CTTGTCGAGA AGTACTAGAG GATCATAATC AGCCATACCA
      GCCATGGTTC GAACAGCTCT TCATGATCTC CTAGTATTAG TCGGTATGGT
1301  CATTTGTAGA GGTTTTACTT GCTTTAAAAA ACCTCCCACA CCTCCCCCTG
      GTAAACATCT CCAAAATGAA CGAAATTTTT TGGAGGGTGT GGAGGGGGAC
1351  AACCTGAAAC ATAAAATGAA TGCAATTGTT GTTGTTAACT TGTTTATTGC
      TTGGACTTTG TATTTTACTT ACGTTAACAA CAACAATTGA ACAAATAACG
1401  AGCTTATAAT GGTTACAAAT AAAGCAATAG CATCACAAAT TTCACAAATA
      TCGAATATTA CCAATGTTTA TTTCGTTATC GTAGTGTTTA AAGTGTTTAT
1451  AAGCATTTTT TTCACTGCAT TCTAGTTGTG GTTTGTCCAA ACTCATCAAT
      TTCGTAAAAA AAGTGACGTA AGATCAACAC CAAACAGGTT TGAGTAGTTA
1501  GTATCTTATC ATGTCTGGAT CTGATCACTG CTTGAGCCTA GGAGATCCGA
      CATAGAATAG TACAGACCTA GACTAGTGAC GAACTCGGAT CCTCTAGGCT
1551  ACCAGATAAG TGAAATCTAG TTCCAAACTA TTTTGTCATT TTTAATTTTC
      TGGTCTATTC ACTTTAGATC AAGGTTTGAT AAAACAGTAA AAATTAAAAG
1601  GTATTAGCTT ACGACGCTAC ACCCAGTTCC CATCTATTTT GTCACTCTTC
      CATAATCGAA TGCTGCGATG TGGGTCAAGG GTAGATAAAA CAGTGAGAAG
1651  CCTAAATAAT CCTTAAAAAC TCCATTTCCA CCCCTCCCAG TTCCCAACTA
      GGATTTATTA GGAATTTTTG AGGTAAAGGT GGGGAGGGTC AAGGGTTGAT
1701  TTTTGTCCGC CCACAGCGGG GCATTTTTCT TCCTGTTATG TTTTTAATCA
      AAAACAGGCG GGTGTCGCCC CGTAAAAAGA AGGACAATAC AAAAATTAGT
1751  AACATCCTGC CAACTCCATG TGACAAACCG TCATCTTCGG CTACTTTTTC
      TTGTAGGACG GTTGAGGTAC ACTGTTTGGC AGTAGAAGCC GATGAAAAAG
1801  TCTGTCACAG AATGAAAATT TTTCTGTCAT CTCTTCGTTA TTAATGTTTG
      AGACAGTGTC TTACTTTTAA AAAGACAGTA GAGAAGCAAT AATTACAAAC
1851  TAATTGACTG AATATCAACG CTTATTTGCA GCCTGAATGG CGAATGGGAC
      ATTAACTGAC TTATAGTTGC GAATAAACGT CGGACTTACC GCTTACCCTG
1901  GCGCCCTGTA GCGGCGCATT AAGCGCGGCG GGTGTGGTGG TTACGCGCAG
      CGCGGGACAT CGCCGCGTAA TTCGCGCCGC CCACACCACC AATGCGCGTC
1951  CGTGACCGCT ACACTTGCCA GCGCCCTAGC GCCCGCTCCT TTCGCTTTCT
      GCACTGGCGA TGTGAACGGT CGCGGGATCG CGGGCGAGGA AAGCGAAAGA
2001  TCCCTTCCTT TCTCGCCACG TTCGCCGGCT TTCCCCGTCA AGCTCTAAAT
      AGGGAAGGAA AGAGCGGTGC AAGCGGCCGA AAGGGGCAGT TCGAGATTTA
2051  CGGGGGCTCC CTTTAGGGTT CCGATTTAGT GCTTTACGGC ACCTCGACCC
      GCCCCCGAGG GAAATCCCAA GGCTAAATCA CGAAATGCCG TGGAGCTGGG
2101  CAAAAAACTT GATTAGGGTG ATGGTTCACG TAGTGGGCCA TCGCCCTGAT
      GTTTTTTGAA CTAATCCCAC TACCAAGTGC ATCACCCGGT AGCGGGACTA
2151  AGACGGTTTT TCGCCCTTTG ACGTTGGAGT CCACGTTCTT TAATAGTGGA
      TCTGCCAAAA AGCGGGAAAC TGCAACCTCA GGTGCAAGAA ATTATCACCT
2201  CTCTTGTTCC AAACTGGAAC AACACTCAAC CCTATCTCGG TCTATTCTTT
      GAGAACAAGG TTTGACCTTG TTGTGAGTTG GGATAGAGCC AGATAAGAAA
2251  TGATTTATAA GGGATTTTGC CGATTTCGGC CTATTGGTTA AAAAATGAGC
      ACTAAATATT CCCTAAAACG GCTAAAGCCG GATAACCAAT TTTTTACTCG
2301  TGATTTAACA AAAATTTAAC GCGAATTTTA ACAAAATATT AACGTTTACA
      ACTAAATTGT TTTTAAATTG CGCTTAAAAT TGTTTTATAA TTGCAAATGT
2351  ATTTCAGGTG GCACTTTTCG GGGAAATGTG CGCGGAACCC CTATTTGTTT
```

Figure 44C

```
      TAAAGTCCAC CGTGAAAAGC CCCTTTACAC GCGCCTTGGG GATAAACAAA
2401  ATTTTTCTAA ATACATTCAA ATATGTATCC GCTCATGAGA CAATAACCCT
      TAAAAAGATT TATGTAAGTT TATACATAGG CGAGTACTCT GTTATTGGGA
2451  GATAAATGCT TCAATAATAT TGAAAAAGGA AGAGTATGAG TATTCAACAT
      CTATTTACGA AGTTATTATA ACTTTTTCCT TCTCATACTC ATAAGTTGTA
2501  TTCCGTGTCG CCCTTATTCC CTTTTTTGCG GCATTTTGCC TTCCTGTTTT
      AAGGCACAGC GGGAATAAGG GAAAAAACGC CGTAAAACGG AAGGACAAAA
2551  TGCTCACCCA GAAACGCTGG TGAAAGTAAA AGATGCTGAA GATCAGTTGG
      ACGAGTGGGT CTTTGCGACC ACTTTCATTT TCTACGACTT CTAGTCAACC
2601  GTGCACGAGT GGGTTACATC GAACTGGATC TCAACAGCGG TAAGATCCTT
      CACGTGCTCA CCCAATGTAG CTTGACCTAG AGTTGTCGCC ATTCTAGGAA
2651  GAGAGTTTTC GCCCCGAAGA ACGTTTTCCA ATGATGAGCA CTTTTAAAGT
      CTCTCAAAAG CGGGGCTTCT TGCAAAAGGT TACTACTCGT GAAAATTTCA
2701  TCTGCTATGT GGCGCGGTAT TATCCCGTAT TGACGCCGGG CAAGAGCAAC
      AGACGATACA CCGCGCCATA ATAGGGCATA ACTGCGGCCC GTTCTCGTTG
2751  TCGGTCGCCG CATACACTAT TCTCAGAATG ACTTGGTTGA GTACTCACCA
      AGCCAGCGGC GTATGTGATA AGAGTCTTAC TGAACCAACT CATGAGTGGT
2801  GTCACAGAAA AGCATCTTAC GGATGGCATG ACAGTAAGAG AATTATGCAG
      CAGTGTCTTT TCGTAGAATG CCTACCGTAC TGTCATTCTC TTAATACGTC
2851  TGCTGCCATA ACCATGAGTG ATAACACTGC GGCCAACTTA CTTCTGACAA
      ACGACGGTAT TGGTACTCAC TATTGTGACG CCGGTTGAAT GAAGACTGTT
2901  CGATCGGAGG ACCGAAGGAG CTAACCGCTT TTTTGCACAA CATGGGGGAT
      GCTAGCCTCC TGGCTTCCTC GATTGGCGAA AAAACGTGTT GTACCCCCTA
2951  CATGTAACTC GCCTTGATCG TTGGGAACCG GAGCTGAATG AAGCCATACC
      GTACATTGAG CGGAACTAGC AACCCTTGGC CTCGACTTAC TTCGGTATGG
3001  AAACGACGAG CGTGACACCA CGATGCCTGT AGCAATGGCA ACAACGTTGC
      TTTGCTGCTC GCACTGTGGT GCTACGGACA TCGTTACCGT TGTTGCAACG
3051  GCAAACTATT AACTGGCGAA CTACTTACTC TAGCTTCCCG GCAACAATTA
      CGTTTGATAA TTGACCGCTT GATGAATGAG ATCGAAGGGC CGTTGTTAAT
3101  ATAGACTGGA TGGAGGCGGA TAAAGTTGCA GGACCACTTC TGCGCTCGGC
      TATCTGACCT ACCTCCGCCT ATTTCAACGT CCTGGTGAAG ACGCGAGCCG
3151  CCTTCCGGCT GGCTGGTTTA TTGCTGATAA ATCTGGAGCC GGTGAGCGTG
      GGAAGGCCGA CCGACCAAAT AACGACTATT TAGACCTCGG CCACTCGCAC
       BsaI
      ~~~~~~
3201  GGTCTCGCGG TATCATTGCA GCACTGGGGC CAGATGGTAA GCCCTCCCGT
      CCAGAGCGCC ATAGTAACGT CGTGACCCCG GTCTACCATT CGGGAGGGCA
3251  ATCGTAGTTA TCTACACGAC GGGGAGTCAG GCAACTATGG ATGAACGAAA
      TAGCATCAAT AGATGTGCTG CCCCTCAGTC CGTTGATACC TACTTGCTTT
3301  TAGACAGATC GCTGAGATAG GTGCCTCACT GATTAAGCAT TGGTAACTGT
      ATCTGTCTAG CGACTCTATC CACGGAGTGA CTAATTCGTA ACCATTGACA
3351  CAGACCAAGT TTACTCATAT ATACTTTAGA TTGATTTAAA ACTTCATTTT
      GTCTGGTTCA AATGAGTATA TATGAAATCT AACTAAATTT TGAAGTAAAA
3401  TAATTTAAAA GGATCTAGGT GAAGATCCTT TTTGATAATC TCATGACCAA
      ATTAAATTTT CCTAGATCCA CTTCTAGGAA AAACTATTAG AGTACTGGTT
3451  AATCCCTTAA CGTGAGTTTT CGTTCCACTG AGCGTCAGAC CCCGTAGAAA
      TTAGGGAATT GCACTCAAAA GCAAGGTGAC TCGCAGTCTG GGGCATCTTT
3501  AGATCAAAGG ATCTTCTTGA GATCCTTTTT TTCTGCGCGT AATCTGCTGC
      TCTAGTTTCC TAGAAGAACT CTAGGAAAAA AAGACGCGCA TTAGACGACG
3551  TTGCAAACAA AAAAACCACC GCTACCAGCG GTGGTTTGTT TGCCGGATCA
      AACGTTTGTT TTTTTGGTGG CGATGGTCGC CACCAAACAA ACGGCCTAGT
```

Figure 44D

```
3601  AGAGCTACCA ACTCTTTTTC CGAAGGTAAC TGGCTTCAGC AGAGCGCAGA
      TCTCGATGGT TGAGAAAAAG GCTTCCATTG ACCGAAGTCG TCTCGCGTCT
3651  TACCAAATAC TGTCCTTCTA GTGTAGCCGT AGTTAGGCCA CCACTTCAAG
      ATGGTTTATG ACAGGAAGAT CACATCGGCA TCAATCCGGT GGTGAAGTTC
3701  AACTCTGTAG CACCGCCTAC ATACCTCGCT CTGCTAATCC TGTTACCAGT
      TTGAGACATC GTGGCGGATG TATGGAGCGA GACGATTAGG ACAATGGTCA
3751  GGCTGCTGCC AGTGGCGATA AGTCGTGTCT TACCGGGTTG GACTCAAGAC
      CCGACGACGG TCACCGCTAT TCAGCACAGA ATGGCCCAAC CTGAGTTCTG
3801  GATAGTTACC GGATAAGGCG CAGCGGTCGG GCTGAACGGG GGGTTCGTGC
      CTATCAATGG CCTATTCCGC GTCGCCAGCC CGACTTGCCC CCCAAGCACG
3851  ACACAGCCCA GCTTGGAGCG AACGACCTAC ACCGAACTGA GATACCTACA
      TGTGTCGGGT CGAACCTCGC TTGCTGGATG TGGCTTGACT CTATGGATGT
3901  GCGTGAGCAT TGAGAAAGCG CCACGCTTCC CGAAGGGAGA AAGGCGGACA
      CGCACTCGTA ACTCTTTCGC GGTGCGAAGG GCTTCCCTCT TTCCGCCTGT
3951  GGTATCCGGT AAGCGGCAGG GTCGGAACAG GAGAGCGCAC GAGGGAGCTT
      CCATAGGCCA TTCGCCGTCC CAGCCTTGTC CTCTCGCGTG CTCCCTCGAA
4001  CCAGGGGGAA ACGCCTGGTA TCTTTATAGT CCTGTCGGGT TTCGCCACCT
      GGTCCCCCTT TGCGGACCAT AGAAATATCA GGACAGCCCA AAGCGGTGGA
4051  CTGACTTGAG CGTCGATTTT TGTGATGCTC GTCAGGGGGG CGGAGCCTAT
      GACTGAACTC GCAGCTAAAA ACACTACGAG CAGTCCCCCC GCCTCGGATA
4101  GGAAAAACGC CAGCAACGCG GCCTTTTTAC GGTTCCTGGC CTTTTGCTGG
      CCTTTTTGCG GTCGTTGCGC CGGAAAAATG CCAAGGACCG GAAAACGACC
4151  CCTTTTGCTC ACATGTTCTT TCCTGCGTTA TCCCCTGATT CTGTGGATAA
      GGAAAACGAG TGTACAAGAA AGGACGCAAT AGGGGACTAA GACACCTATT
4201  CCGTATTACC GCCTTTGAGT GAGCTGATAC CGCTCGCCGC AGCCGAACGA
      GGCATAATGG CGGAAACTCA CTCGACTATG GCGAGCGGCG TCGGCTTGCT
4251  CCGAGCGCAG CGAGTCAGTG AGCGAGGAAG CGGAAGAGCG CCTGATGCGG
      GGCTCGCGTC GCTCAGTCAC TCGCTCCTTC GCCTTCTCGC GGACTACGCC
4301  TATTTTCTCC TTACGCATCT GTGCGGTATT TCACACCGCA GACCAGCCGC
      ATAAAAGAGG AATGCGTAGA CACGCCATAA AGTGTGGCGT CTGGTCGGCG
4351  GTAACCTGGC AAAATCGGTT ACGGTTGAGT AATAAATGGA TGCCCTGCGT
      CATTGGACCG TTTTAGCCAA TGCCAACTCA TTATTTACCT ACGGGACGCA
4401  AAGCGGGTGT GGGCGGACAA TAAAGTCTTA AACTGAACAA AATAGATCTA
      TTCGCCCACA CCCGCCTGTT ATTTCAGAAT TTGACTTGTT TTATCTAGAT
4451  AACTATGACA ATAAAGTCTT AAACTAGACA GAATAGTTGT AAACTGAAAT
      TTGATACTGT TATTTCAGAA TTTGATCTGT CTTATCAACA TTTGACTTTA
4501  CAGTCCAGTT ATGCTGTGAA AAAGCATACT GGACTTTTGT TATGGCTAAA
      GTCAGGTCAA TACGACACTT TTTCGTATGA CCTGAAAACA ATACCGATTT
4551  GCAAACTCTT CATTTTCTGA AGTGCAAATT GCCCGTCGTA TTAAAGAGGG
      CGTTTGAGAA GTAAAAGACT TCACGTTTAA CGGGCAGCAT AATTTCTCCC
4601  GCGTGGCCAA GGGCATGGTA AAGACTATAT TCGCGGCGTT GTGACAATTT
      CGCACCGGTT CCCGTACCAT TTCTGATATA AGCGCCGCAA CACTGTTAAA
4651  ACCGAACAAC TCCGCGGCCG GGAAGCCGAT CTCGGCTTGA ACGAATTGTT
      TGGCTTGTTG AGGCGCCGGC CCTTCGGCTA GAGCCGAACT TGCTTAACAA
4701  AGGTGGCGGT ACTTGGGTCG ATATCAAAGT GCATCACTTC TTCCCGTATG
      TCCACCGCCA TGAACCCAGC TATAGTTTCA CGTAGTGAAG AAGGGCATAC
4751  CCCAACTTTG TATAGAGAGC CACTGCGGGA TCGTCACCGT AATCTGCTTG
      GGGTTGAAAC ATATCTCTCG GTGACGCCCT AGCAGTGGCA TTAGACGAAC
4801  CACGTAGATC ACATAAGCAC CAAGCGCGTT GGCCTCATGC TTGAGGAGAT
      GTGCATCTAG TGTATTCGTG GTTCGCGCAA CCGGAGTACG AACTCCTCTA
4851  TGATGAGCGC GGTGGCAATG CCCTGCCTCC GGTGCTCGCC GGAGACTGCG
```

Figure 44E

```
      ACTACTCGCG CCACCGTTAC GGGACGGAGG CCACGAGCGG CCTCTGACGC
4901  AGATCATAGA TATAGATCTC ACTACGCGGC TGCTCAAACC TGGGCAGAAC
      TCTAGTATCT ATATCTAGAG TGATGCGCCG ACGAGTTTGG ACCCGTCTTG
4951  GTAAGCCGCG AGAGCGCCAA CAACCGCTTC TTGGTCGAAG GCAGCAAGCG
      CATTCGGCGC TCTCGCGGTT GTTGGCGAAG AACCAGCTTC CGTCGTTCGC
5001  CGATGAATGT CTTACTACGG AGCAAGTTCC CGAGGTAATC GGAGTCCGGC
      GCTACTTACA GAATGATGCC TCGTTCAAGG GCTCCATTAG CCTCAGGCCG
                              Esp3I
                              ~~~~~~
5051  TGATGTTGGG AGTAGGTGGC TACGTCTCCG AACTCACGAC CGAAAAGATC
      ACTACAACCC TCATCCACCG ATGCAGAGGC TTGAGTGCTG GCTTTTCTAG
5101  AAGAGCAGCC CGCATGGATT TGACTTGGTC AGGGCCGAGC CTACATGTGC
      TTCTCGTCGG GCGTACCTAA ACTGAACCAG TCCCGGCTCG GATGTACACG
5151  GAATGATGCC CATACTTGAG CCACCTAACT TTGTTTTAGG GCGACTGCCC
      CTTACTACGG GTATGAACTC GGTGGATTGA AACAAAATCC CGCTGACGGG
5201  TGCTGCGTAA CATCGTTGCT GCTGCGTAAC ATCGTTGCTG CTCCATAACA
      ACGACGCATT GTAGCAACGA CGACGCATTG TAGCAACGAC GAGGTATTGT
5251  TCAAACATCG ACCCACGGCG TAACGCGCTT GCTGCTTGGA TGCCCGAGGC
      AGTTTGTAGC TGGGTGCCGC ATTGCGCGAA CGACGAACCT ACGGGCTCCG
5301  ATAGACTGTA CAAAAAAACA GTCATAACAA GCCATGAAAA CCGCCACTGC
      TATCTGACAT GTTTTTTTGT CAGTATTGTT CGGTACTTTT GGCGGTGACG
5351  GCCGTTACCA CCGCTGCGTT CGGTCAAGGT TCTGGACCAG TTGCGTGAGC
      CGGCAATGGT GGCGACGCAA GCCAGTTCCA AGACCTGGTC AACGCACTCG
5401  GCATACGCTA CTTGCATTAC AGTTTACGAA CCGAACAGGC TTATGTCAAC
      CGTATGCGAT GAACGTAATG TCAAATGCTT GGCTTGTCCG AATACAGTTG
5451  TGGGTTCGTG CCTTCATCCG TTTCCACGGT GTGCGTCACC CGGCAACCTT
      ACCCAAGCAC GGAAGTAGGC AAAGGTGCCA CACGCAGTGG GCCGTTGGAA
5501  GGGCAGCAGC GAAGTCGAGG CATTTCTGTC CTGGCTGGCG AACGAGCGCA
      CCCGTCGTCG CTTCAGCTCC GTAAAGACAG GACCGACCGC TTGCTCGCGT
              BsaI
              ~~~~~~~
5551  AGGTTTCGGT CTCCACGCAT CGTCAGGCAT TGGCGGCCTT GCTGTTCTTC
      TCCAAAGCCA GAGGTGCGTA GCAGTCCGTA ACCGCCGGAA CGACAAGAAG
5601  TACGGCAAGG TGCTGTGCAC GGATCTGCCC TGGCTTCAGG AGATCGGAAG
      ATGCCGTTCC ACGACACGTG CCTAGACGGG ACCGAAGTCC TCTAGCCTTC
5651  ACCTCGGCCG TCGCGGCGCT TGCCGGTGGT GCTGACCCCG GATGAAGTGG
      TGGAGCCGGC AGCGCCGCGA ACGGCCACCA CGACTGGGGC CTACTTCACC
5701  TTCGCATCCT CGGTTTTCTG GAAGGCGAGC ATCGTTTGTT CGCCCAGGAC
      AAGCGTAGGA GCCAAAAGAC CTTCCGCTCG TAGCAAACAA GCGGGTCCTG
5751  TCTAGCTATA GTTCTAGTGG TTGGCTACGT ATACTCCGGA ATATTAATAG
      AGATCGATAT CAAGATCACC AACCGATGCA TATGAGGCCT TATAATTATC
```

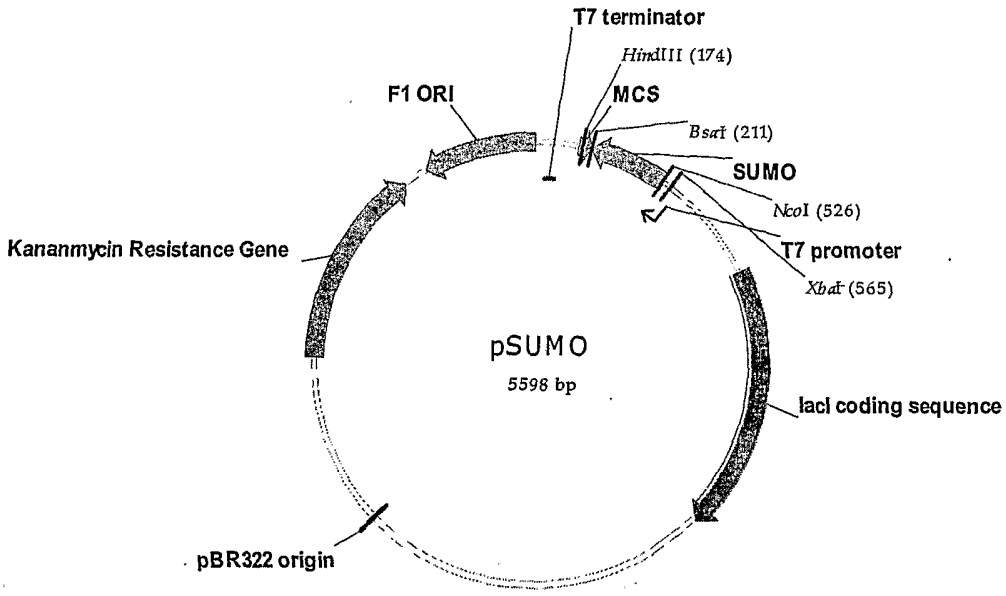

T7 terminator
HindIII (174)
MCS
F1 ORI
BsaI (211)
SUMO
NcoI (526)
T7 promoter
XbaI (565)
Kananmycin Resistance Gene
pSUMO
5598 bp
lacI coding sequence
pBR322 origin

Figure 45

Figure 46A

```
   1  ATCCGGATAT AGTTCCTCCT TTCAGCAAAA AACCCCTCAA GACCCGTTTA
      TAGGCCTATA TCAAGGAGGA AAGTCGTTTT TTGGGGAGTT CTGGGCAAAT
  51  GAGGCCCCAA GGGGTTATGC TAGTTATTGC TCAGCGGTGG CAGCAGCCAA
      CTCCGGGGTT CCCCAATACG ATCAATAACG AGTCGCCACC GTCGTCGGTT
 101  CTCAGCTTCC TTTCGGGCTT TGTTAGCAGC CGGATCTCAG TGGTGGTGGT
      GAGTCGAAGG AAAGCCCGAA ACAATCGTCG GCCTAGAGTC ACCACCACCA
                                HindIII
                                ~~~~~~
 151  GGTGGTGCTC GAGTGCGGCC GCAAGCTTGT CGACGGAGCT CGAATTCGGA
      CCACCACGAG CTCACGCCGG CGTTCGAACA GCTGCCTCGA GCTTAAGCCT
          BsaI
          ~~~~~~
 201  TCCGGTCTCA ACCTCCAATC TGTTCGCGGT GAGCCTCAAT AATATCGTTA
      AGGCCAGAGT TGGAGGTTAG ACAAGCGCCA CTCGGAGTTA TTATAGCAAT
 251  TCCTCCATGT CCAAATCTTC AGGGGTCTGA TCAGCTTGAA TTCTAATACC
      AGGAGGTACA GGTTTAGAAG TCCCCAGACT AGTCGAACTT AAGATTATGG
 301  GTCGTACAAG AATCTTAAGG AGTCCATTTC CTTACCCTGT CTTTTAGCGA
      CAGCATGTTC TTAGAATTCC TCAGGTAAAG GAATGGGACA GAAAATCGCT
 351  ACGCTTCCAT CAGCCTTCTT AAAGGAGTGG TCTTTTTGAT CTTGAAGAAG
      TGCGAAGGTA GTCGGAAGAA TTTCCTCACC AGAAAAACTA GAACTTCTTC
 401  ATCTCTGAAG ATCCATCGGA CACCTTTAAA TTGATGTGAG TCTCAGGCTT
      TAGAGACTTC TAGGTAGCCT GTGGAAATTT AACTACACTC AGAGTCCGAA
 451  GACTTCTGGC TTGACCTCTG GCTTAGCTTC TTGATTGACT TCTGAGTCCG
      CTGAAGACCG AACTGGAGAC CGAATCGAAG AACTAACTGA AGACTCAGGC
                                  NcoI
                                  ~~~~~~~
 501  ACCCGTGATG ATGATGGTGA TGACCCATGG TATATCTCCT TCTTAAAGTT
      TGGGCACTAC TACTACCACT ACTGGGTACC ATATAGAGGA AGAATTTCAA
                        XbaI
                        ~~~~~~
 551  AAACAAAATT ATTTCTAGAG GGGAATTGTT ATCCGCTCAC AATTCCCCTA
      TTTGTTTTAA TAAAGATCTC CCCTTAACAA TAGGCGAGTG TTAAGGGGAT
 601  TAGTGAGTCG TATTAATTTC GCGGGATCGA GATCTCGATC CTCTACGCCG
      ATCACTCAGC ATAATTAAAG CGCCCTAGCT CTAGAGCTAG GAGATGCGGC
 651  GACGCATCGT GGCCGGCATC ACCGGCGCCA CAGGTGCGGT TGCTGGCGCC
      CTGCGTAGCA CCGGCCGTAG TGGCCGCGGT GTCCACGCCA ACGACCGCGG
 701  TATATCGCCG ACATCACCGA TGGGGAAGAT CGGGCTCGCC ACTTCGGGCT
      ATATAGCGGC TGTAGTGGCT ACCCCTTCTA GCCCGAGCGG TGAAGCCCGA
 751  CATGAGCGCT TGTTTCGGCG TGGGTATGGT GGCAGGCCCC GTGGCCGGGG
      GTACTCGCGA ACAAAGCCGC ACCCATACCA CCGTCCGGGG CACCGGCCCC
 801  GACTGTTGGG CGCCATCTCC TTGCATGCAC CATTCCTTGC GGCGGCGGTG
      CTGACAACCC GCGGTAGAGG AACGTACGTG GTAAGGAACG CCGCCGCCAC
 851  CTCAACGGCC TCAACCTACT ACTGGGCTGC TTCCTAATGC AGGAGTCGCA
      GAGTTGCCGG AGTTGGATGA TGACCCGACG AAGGATTACG TCCTCAGCGT
 901  TAAGGGAGAG CGTCGAGATC CCGGACACCA TCGAATGGCG CAAAACCTTT
      ATTCCCTCTC GCAGCTCTAG GGCCTGTGGT AGCTTACCGC GTTTTGGAAA
 951  CGCGGTATGG CATGATAGCG CCCGGAAGAG AGTCAATTCA GGGTGGTGAA
      GCGCCATACC GTACTATCGC GGGCCTTCTC TCAGTTAAGT CCCACCACTT
1001  TGTGAAACCA GTAACGTTAT ACGATGTCGC AGAGTATGCC GGTGTCTCTT
```

Figure 46B

```
       ACACTTTGGT CATTGCAATA TGCTACAGCG TCTCATACGG CCACAGAGAA
1051   ATCAGACCGT TTCCCGCGTG GTGAACCAGG CCAGCCACGT TTCTGCGAAA
       TAGTCTGGCA AAGGGCGCAC CACTTGGTCC GGTCGGTGCA AAGACGCTTT
1101   ACGCGGGAAA AAGTGGAAGC GGCGATGGCG GAGCTGAATT ACATTCCCAA
       TGCGCCCTTT TTCACCTTCG CCGCTACCGC CTCGACTTAA TGTAAGGGTT
1151   CCGCGTGGCA CAACAACTGG CGGGCAAACA GTCGTTGCTG ATTGGCGTTG
       GGCGCACCGT GTTGTTGACC GCCCGTTTGT CAGCAACGAC TAACCGCAAC
1201   CCACCTCCAG TCTGGCCCTG CACGCGCCGT CGCAAATTGT CGCGGCGATT
       GGTGGAGGTC AGACCGGGAC GTGCGCGGCA GCGTTTAACA GCGCCGCTAA
1251   AAATCTCGCG CCGATCAACT GGGTGCCAGC GTGGTGGTGT CGATGGTAGA
       TTTAGAGCGC GGCTAGTTGA CCCACGGTCG CACCACCACA GCTACCATCT
1301   ACGAAGCGGC GTCGAAGCCT GTAAAGCGGC GGTGCACAAT CTTCTCGCGC
       TGCTTCGCCG CAGCTTCGGA CATTTCGCCG CCACGTGTTA GAAGAGCGCG
1351   AACGCGTCAG TGGGCTGATC ATTAACTATC CGCTGGATGA CCAGGATGCC
       TTGCGCAGTC ACCCGACTAG TAATTGATAG GCGACCTACT GGTCCTACGG
1401   ATTGCTGTGG AAGCTGCCTG CACTAATGTT CCGGCGTTAT TTCTTGATGT
       TAACGACACC TTCGACGGAC GTGATTACAA GGCCGCAATA AAGAACTACA
1451   CTCTGACCAG ACACCCATCA ACAGTATTAT TTTCTCCCAT GAAGACGGTA
       GAGACTGGTC TGTGGGTAGT TGTCATAATA AAAGAGGGTA CTTCTGCCAT
1501   CGCGACTGGG CGTGGAGCAT CTGGTCGCAT TGGGTCACCA GCAAATCGCG
       GCGCTGACCC GCACCTCGTA GACCAGCGTA ACCCAGTGGT CGTTTAGCGC
1551   CTGTTAGCGG GCCCATTAAG TTCTGTCTCG GCGCGTCTGC GTCTGGCTGG
       GACAATCGCC CGGGTAATTC AAGACAGAGC CGCGCAGACG CAGACCGACC
1601   CTGGCATAAA TATCTCACTC GCAATCAAAT TCAGCCGATA GCGGAACGGG
       GACCGTATTT ATAGAGTGAG CGTTAGTTTA AGTCGGCTAT CGCCTTGCCC
1651   AAGGCGACTG GAGTGCCATG TCCGGTTTTC AACAAACCAT GCAAATGCTG
       TTCCGCTGAC CTCACGGTAC AGGCCAAAAG TTGTTTGGTA CGTTTACGAC
1701   AATGAGGGCA TCGTTCCCAC TGCGATGCTG GTTGCCAACG ATCAGATGGC
       TTACTCCCGT AGCAAGGGTG ACGCTACGAC CAACGGTTGC TAGTCTACCG
1751   GCTGGGCGCA ATGCGCGCCA TTACCGAGTC CGGGCTGCGC GTTGGTGCGG
       CGACCCGCGT TACGCGCGGT AATGGCTCAG GCCCGACGCG CAACCACGCC
1801   ATATCTCGGT AGTGGGATAC GACGATACCG AAGACAGCTC ATGTTATATC
       TATAGAGCCA TCACCCTATG CTGCTATGGC TTCTGTCGAG TACAATATAG
1851   CCGCCGTTAA CCACCATCAA ACAGGATTTT CGCCTGCTGG GGCAAACCAG
       GGCGGCAATT GGTGGTAGTT TGTCCTAAAA GCGGACGACC CCGTTTGGTC
1901   CGTGGACCGC TTGCTGCAAC TCTCTCAGGG CCAGGCGGTG AAGGGCAATC
       GCACCTGGCG AACGACGTTG AGAGAGTCCC GGTCCGCCAC TTCCCGTTAG
1951   AGCTGTTGCC CGTCTCACTG GTGAAAAGAA AAACCACCCT GGCGCCCAAT
       TCGACAACGG GCAGAGTGAC CACTTTTCTT TTTGGTGGGA CCGCGGGTTA
2001   ACGCAAACCG CCTCTCCCCG CGCGTTGGCC GATTCATTAA TGCAGCTGGC
       TGCGTTTGGC GGAGAGGGGC GCGCAACCGG CTAAGTAATT ACGTCGACCG
2051   ACGACAGGTT TCCCGACTGG AAAGCGGGCA GTGAGCGCAA CGCAATTAAT
       TGCTGTCCAA AGGGCTGACC TTTCGCCCGT CACTCGCGTT GCGTTAATTA
2101   GTAAGTTAGC TCACTCATTA GGCACCGGGA TCTCGACCGA TGCCCTTGAG
       CATTCAATCG AGTGAGTAAT CCGTGGCCCT AGAGCTGGCT ACGGGAACTC
2151   AGCCTTCAAC CCAGTCAGCT CCTTCCGGTG GGCGCGGGGC ATGACTATCG
       TCGGAAGTTG GGTCAGTCGA GGAAGGCCAC CCGCGCCCCG TACTGATAGC
2201   TCGCCGCACT TATGACTGTC TTCTTTATCA TGCAACTCGT AGGACAGGTG
       AGCGGCGTGA ATACTGACAG AAGAAATAGT ACGTTGAGCA TCCTGTCCAC
2251   CCGGCAGCGC TCTGGGTCAT TTTCGGCGAG GACCGCTTTC GCTGGAGCGC
       GGCCGTCGCG AGACCCAGTA AAAGCCGCTC CTGGCGAAAG CGACCTCGCG
```

Figure 46C

```
2301  GACGATGATC GGCCTGTCGC TTGCGGTATT CGGAATCTTG CACGCCCTCG
      CTGCTACTAG CCGGACAGCG AACGCCATAA GCCTTAGAAC GTGCGGGAGC
2351  CTCAAGCCTT CGTCACTGGT CCCGCCACCA AACGTTTCGG CGAGAAGCAG
      GAGTTCGGAA GCAGTGACCA GGGCGGTGGT TTGCAAAGCC GCTCTTCGTC
2401  GCCATTATCG CCGGCATGGC GGCCCCACGG GTGCGCATGA TCGTGCTCCT
      CGGTAATAGC GGCCGTACCG CCGGGGTGCC CACGCGTACT AGCACGAGGA
2451  GTCGTTGAGG ACCCGGCTAG GCTGGCGGGG TTGCCTTACT GGTTAGCAGA
      CAGCAACTCC TGGGCCGATC CGACCGCCCC AACGGAATGA CCAATCGTCT
2501  ATGAATCACC GATACGCGAG CGAACGTGAA GCGACTGCTG CTGCAAAACG
      TACTTAGTGG CTATGCGCTC GCTTGCACTT CGCTGACGAC GACGTTTTGC
2551  TCTGCGACCT GAGCAACAAC ATGAATGGTC TTCGGTTTCC GTGTTTCGTA
      AGACGCTGGA CTCGTTGTTG TACTTACCAG AAGCCAAAGG CACAAAGCAT
2601  AAGTCTGGAA ACGCGGAAGT CAGCGCCCTG CACCATTATG TTCCGGATCT
      TTCAGACCTT TGCGCCTTCA GTCGCGGGAC GTGGTAATAC AAGGCCTAGA
2651  GCATCGCAGG ATGCTGCTGG CTACCCTGTG GAACACCTAC ATCTGTATTA
      CGTAGCGTCC TACGACGACC GATGGGACAC CTTGTGGATG TAGACATAAT
2701  ACGAAGCGCT GGCATTGACC CTGAGTGATT TTTCTCTGGT CCCGCCGCAT
      TGCTTCGCGA CCGTAACTGG GACTCACTAA AAAGAGACCA GGGCGGCGTA
2751  CCATACCGCC AGTTGTTTAC CCTCACAACG TTCCAGTAAC CGGGCATGTT
      GGTATGGCGG TCAACAAATG GGAGTGTTGC AAGGTCATTG GCCCGTACAA
2801  CATCATCAGT AACCCGTATC GTGAGCATCC TCTCTCGTTT CATCGGTATC
      GTAGTAGTCA TTGGGCATAG CACTCGTAGG AGAGAGCAAA GTAGCCATAG
2851  ATTACCCCCA TGAACAGAAA TCCCCCTTAC ACGGAGGCAT CAGTGACCAA
      TAATGGGGGT ACTTGTCTTT AGGGGGAATG TGCCTCCGTA GTCACTGGTT
2901  ACAGGAAAAA ACCGCCCTTA ACATGGCCCG CTTTATCAGA AGCCAGACAT
      TGTCCTTTTT TGGCGGGAAT TGTACCGGGC GAAATAGTCT TCGGTCTGTA
2951  TAACGCTTCT GGAGAAACTC AACGAGCTGG ACGCGGATGA ACAGGCAGAC
      ATTGCGAAGA CCTCTTTGAG TTGCTCGACC TGCGCCTACT TGTCCGTCTG
3001  ATCTGTGAAT CGCTTCACGA CCACGCTGAT GAGCTTTACC GCAGCTGCCT
      TAGACACTTA GCGAAGTGCT GGTGCGACTA CTCGAAATGG CGTCGACGGA
3051  CGCGCGTTTC GGTGATGACG GTGAAAACCT CTGACACATG CAGCTCCCGG
      GCGCGCAAAG CCACTACTGC CACTTTTGGA GACTGTGTAC GTCGAGGGCC
3101  AGACGGTCAC AGCTTGTCTG TAAGCGGATG CCGGGAGCAG ACAAGCCCGT
      TCTGCCAGTG TCGAACAGAC ATTCGCCTAC GGCCCTCGTC TGTTCGGGCA
3151  CAGGGCGCGT CAGCGGGTGT TGGCGGGTGT CGGGGCGCAG CCATGACCCA
      GTCCCGCGCA GTCGCCCACA ACCGCCCACA GCCCCGCGTC GGTACTGGGT
3201  GTCACGTAGC GATAGCGGAG TGTATACTGG CTTAACTATG CGGCATCAGA
      CAGTGCATCG CTATCGCCTC ACATATGACC GAATTGATAC GCCGTAGTCT
3251  GCAGATTGTA CTGAGAGTGC ACCATATATG CGGTGTGAAA TACCGCACAG
      CGTCTAACAT GACTCTCACG TGGTATATAC GCCACACTTT ATGGCGTGTC
3301  ATGCGTAAGG AGAAAATACC GCATCAGGCG CTCTTCCGCT TCCTCGCTCA
      TACGCATTCC TCTTTTATGG CGTAGTCCGC GAGAAGGCGA AGGAGCGAGT
3351  CTGACTCGCT GCGCTCGGTC GTTCGGCTGC GGCGAGCGGT ATCAGCTCAC
      GACTGAGCGA CGCGAGCCAG CAAGCCGACG CCGCTCGCCA TAGTCGAGTG
3401  TCAAAGGCGG TAATACGGTT ATCCACAGAA TCAGGGGATA ACGCAGGAAA
      AGTTTCCGCC ATTATGCCAA TAGGTGTCTT AGTCCCCTAT TGCGTCCTTT
3451  GAACATGTGA GCAAAAGGCC AGCAAAAGGC CAGGAACCGT AAAAAGGCCG
      CTTGTACACT CGTTTTCCGG TCGTTTTCCG GTCCTTGGCA TTTTTCCGGC
3501  CGTTGCTGGC GTTTTTCCAT AGGCTCCGCC CCCCTGACGA GCATCACAAA
      GCAACGACCG CAAAAAGGTA TCCGAGGCGG GGGGACTGCT CGTAGTGTTT
3551  AATCGACGCT CAAGTCAGAG GTGGCGAAAC CCGACAGGAC TATAAAGATA
```

Figure. 46D

```
      TTAGCTGCGA GTTCAGTCTC CACCGCTTTG GGCTGTCCTG ATATTTCTAT
3601  CCAGGCGTTT CCCCCTGGAA GCTCCCTCGT GCGCTCTCCT GTTCCGACCC
      GGTCCGCAAA GGGGGACCTT CGAGGGAGCA CGCGAGAGGA CAAGGCTGGG
3651  TGCCGCTTAC CGGATACCTG TCCGCCTTTC TCCCTTCGGG AAGCGTGGCG
      ACGGCGAATG GCCTATGGAC AGGCGGAAAG AGGGAAGCCC TTCGCACCGC
3701  CTTTCTCATA GCTCACGCTG TAGGTATCTC AGTTCGGTGT AGGTCGTTCG
      GAAAGAGTAT CGAGTGCGAC ATCCATAGAG TCAAGCCACA TCCAGCAAGC
3751  CTCCAAGCTG GGCTGTGTGC ACGAACCCCC CGTTCAGCCC GACCGCTGCG
      GAGGTTCGAC CCGACACACG TGCTTGGGGG GCAAGTCGGG CTGGCGACGC
3801  CCTTATCCGG TAACTATCGT CTTGAGTCCA ACCCGGTAAG ACACGACTTA
      GGAATAGGCC ATTGATAGCA GAACTCAGGT TGGGCCATTC TGTGCTGAAT
3851  TCGCCACTGG CAGCAGCCAC TGGTAACAGG ATTAGCAGAG CGAGGTATGT
      AGCGGTGACC GTCGTCGGTG ACCATTGTCC TAATCGTCTC GCTCCATACA
3901  AGGCGGTGCT ACAGAGTTCT TGAAGTGGTG GCCTAACTAC GGCTACACTA
      TCCGCCACGA TGTCTCAAGA ACTTCACCAC CGGATTGATG CCGATGTGAT
3951  GAAGGACAGT ATTTGGTATC TGCGCTCTGC TGAAGCCAGT TACCTTCGGA
      CTTCCTGTCA TAAACCATAG ACGCGAGACG ACTTCGGTCA ATGGAAGCCT
4001  AAAAGAGTTG GTAGCTCTTG ATCCGGCAAA CAAACCACCG CTGGTAGCGG
      TTTTCTCAAC CATCGAGAAC TAGGCCGTTT GTTTGGTGGC GACCATCGCC
4051  TGGTTTTTTT GTTTGCAAGC AGCAGATTAC GCGCAGAAAA AAAGGATCTC
      ACCAAAAAAA CAAACGTTCG TCGTCTAATG CGCGTCTTTT TTTCCTAGAG
4101  AAGAAGATCC TTTGATCTTT TCTACGGGGT CTGACGCTCA GTGGAACGAA
      TTCTTCTAGG AAACTAGAAA AGATGCCCCA GACTGCGAGT CACCTTGCTT
4151  AACTCACGTT AAGGGATTTT GGTCATGAAC AATAAAACTG TCTGCTTACA
      TTGAGTGCAA TTCCCTAAAA CCAGTACTTG TTATTTTGAC AGACGAATGT
4201  TAAACAGTAA TACAAGGGGT GTTATGAGCC ATATTCAACG GGAAACGTCT
      ATTTGTCATT ATGTTCCCCA CAATACTCGG TATAAGTTGC CCTTTGCAGA
4251  TGCTCTAGGC CGCGATTAAA TTCCAACATG GATGCTGATT TATATGGGTA
      ACGAGATCCG GCGCTAATTT AAGGTTGTAC CTACGACTAA ATATACCCAT
4301  TAAATGGGCT CGCGATAATG TCGGGCAATC AGGTGCGACA ATCTATCGAT
      ATTTACCCGA GCGCTATTAC AGCCCGTTAG TCCACGCTGT TAGATAGCTA
4351  TGTATGGGAA GCCCGATGCG CCAGAGTTGT TTCTGAAACA TGGCAAAGGT
      ACATACCCTT CGGGCTACGC GGTCTCAACA AAGACTTTGT ACCGTTTCCA
4401  AGCGTTGCCA ATGATGTTAC AGATGAGATG GTCAGACTAA ACTGGCTGAC
      TCGCAACGGT TACTACAATG TCTACTCTAC CAGTCTGATT TGACCGACTG
4451  GGAATTTATG CCTCTTCCGA CCATCAAGCA TTTTATCCGT ACTCCTGATG
      CCTTAAATAC GGAGAAGGCT GGTAGTTCGT AAAATAGGCA TGAGGACTAC
4501  ATGCATGGTT ACTCACCACT GCGATCCCCG GGAAAACAGC ATTCCAGGTA
      TACGTACCAA TGAGTGGTGA CGCTAGGGGC CCTTTTGTCG TAAGGTCCAT
4551  TTAGAAGAAT ATCCTGATTC AGGTGAAAAT ATTGTTGATG CGCTGGCAGT
      AATCTTCTTA TAGGACTAAG TCCACTTTTA TAACAACTAC GCGACCGTCA
4601  GTTCCTGCGC CGGTTGCATT CGATTCCTGT TTGTAATTGT CCTTTTAACA
      CAAGGACGCG GCCAACGTAA GCTAAGGACA AACATTAACA GGAAAATTGT
4651  GCGATCGCGT ATTTCGTCTC GCTCAGGCGC AATCACGAAT GAATAACGGT
      CGCTAGCGCA TAAAGCAGAG CGAGTCCGCG TTAGTGCTTA CTTATTGCCA
4701  TTGGTTGATG CGAGTGATTT TGATGACGAG CGTAATGGCT GGCCTGTTGA
      AACCAACTAC GCTCACTAAA ACTACTGCTC GCATTACCGA CCGGACAACT
4751  ACAAGTCTGG AAAGAAATGC ATAAACTTTT GCCATTCTCA CCGGATTCAG
      TGTTCAGACC TTTCTTTACG TATTTGAAAA CGGTAAGAGT GGCCTAAGTC
4801  TCGTCACTCA TGGTGATTTC TCACTTGATA ACCTTATTTT TGACGAGGGG
      AGCAGTGAGT ACCACTAAAG AGTGAACTAT TGGAATAAAA ACTGCTCCCC
```

Figure 46E

```
4851  AAATTAATAG GTTGTATTGA TGTTGGACGA GTCGGAATCG CAGACCGATA
      TTTAATTATC CAACATAACT ACAACCTGCT CAGCCTTAGC GTCTGGCTAT
4901  CCAGGATCTT GCCATCCTAT GGAACTGCCT CGGTGAGTTT TCTCCTTCAT
      GGTCCTAGAA CGGTAGGATA CCTTGACGGA GCCACTCAAA AGAGGAAGTA
4951  TACAGAAACG GCTTTTTCAA AAATATGGTA TTGATAATCC TGATATGAAT
      ATGTCTTTGC CGAAAAAGTT TTTATACCAT AACTATTAGG ACTATACTTA
5001  AAATTGCAGT TTCATTTGAT GCTCGATGAG TTTTTCTAAG AATTAATTCA
      TTTAACGTCA AAGTAAACTA CGAGCTACTC AAAAAGATTC TTAATTAAGT
5051  TGAGCGGATA CATATTTGAA TGTATTTAGA AAAATAAACA AATAGGGGTT
      ACTCGCCTAT GTATAAACTT ACATAAATCT TTTTATTTGT TTATCCCCAA
5101  CCGCGCACAT TTCCCCGAAA AGTGCCACCT GAAATTGTAA ACGTTAATAT
      GGCGCGTGTA AAGGGGCTTT TCACGGTGGA CTTTAACATT TGCAATTATA
5151  TTTGTTAAAA TTCGCGTTAA ATTTTTGTTA AATCAGCTCA TTTTTTAACC
      AAACAATTTT AAGCGCAATT TAAAAACAAT TTAGTCGAGT AAAAAATTGG
5201  AATAGGCCGA AATCGGCAAA ATCCCTTATA AATCAAAAGA ATAGACCGAG
      TTATCCGGCT TTAGCCGTTT TAGGGAATAT TTAGTTTTCT TATCTGGCTC
5251  ATAGGGTTGA GTGTTGTTCC AGTTTGGAAC AAGAGTCCAC TATTAAAGAA
      TATCCCAACT CACAACAAGG TCAAACCTTG TTCTCAGGTG ATAATTTCTT
5301  CGTGGACTCC AACGTCAAAG GGCGAAAAAC CGTCTATCAG GGCGATGGCC
      GCACCTGAGG TTGCAGTTTC CCGCTTTTTG GCAGATAGTC CCGCTACCGG
5351  CACTACGTGA ACCATCACCC TAATCAAGTT TTTTGGGGTC GAGGTGCCGT
      GTGATGCACT TGGTAGTGGG ATTAGTTCAA AAAACCCCAG CTCCACGGCA
5401  AAAGCACTAA ATCGGAACCC TAAAGGGAGC CCCCGATTTA GAGCTTGACG
      TTTCGTGATT TAGCCTTGGG ATTTCCCTCG GGGGCTAAAT CTCGAACTGC
5451  GGGAAAGCCG GCGAACGTGG CGAGAAAGGA AGGGAAGAAA GCGAAAGGAG
      CCCTTTCGGC CGCTTGCACC GCTCTTTCCT TCCCTTCTTT CGCTTTCCTC
5501  CGGGCGCTAG GGCGCTGGCA AGTGTAGCGG TCACGCTGCG CGTAACCACC
      GCCCGCGATC CCGCGACCGT TCACATCGCC AGTGCGACGC GCATTGGTGG
5551  ACACCCGCCG CGCTTAATGC GCCGCTACAG GGCGCGTCCC ATTCGCCA
      TGTGGGCGGC GCGAATTACG CGGCGATGTC CCGCGCAGGG TAAGCGGT
```