Published in final edited form as:

Protein Expr Purif. 2011 March ; 76(1): 65–71. doi:10.1016/j.pep.2010.08.008.

A SUMO-GROUCHO Q DOMAIN FUSION PROTEIN: CHARACTERIZATION AND IN VIVO ULP1-MEDIATED CLEAVAGE

Dennis Kuo¹, Minghua Nie², Peter De Hoff², Michael Chambers¹, Martin Phillips¹, Ann M. Hirsch^{2,3}, and Albert J. Courey^{1,2,*}

¹Department of Chemistry and Biochemistry, UCLA, Los Angeles, CA 90095

²Molecular Biology Institute, UCLA, Los Angeles, CA 90095

³Department of Molecular, Cell, and Developmental Biology, UCLA, Los Angeles, CA 90095

Abstract

We describe here a system for the expression and purification of small ubiquitin-related modifier (SUMO) fusion proteins, which often exhibit dramatically increased solubility and stability during expression in bacteria relative to unfused proteins. The vector described here allows expression of a His-tagged protein of interest fused at its N-terminus to SUMO. Using this vector, we have produced a polypeptide consisting of SUMO fused to the Q-domain of *Drosophila* Groucho in a concentrated soluble form. Hydrodynamic analysis shows that, consistent with previous studies on full-length Groucho, the fusion protein forms an elongated tetramer, as well as higher order oligomers. After expressing a protein as a fusion to SUMO, it is often desirable to cleave the SUMO off of the fusion protein using a SUMO-specific protease such as Ulp1. To facilitate such processing, we have constructed a dual expression vector encoding two fusion proteins: one consisting of SUMO fused to Ulp1 and a second consisting of SUMO fused to a His-tagged protein of interest. The SUMO-Ulp1 cleaves both itself and the other SUMO fusion protein in the bacterial cells prior to lysis, and the proteins retain solubility after cleavage.

INTRODUCTION

Small ubiquitin-related modifier (SUMO) is a ubiquitin family protein that becomes covalently conjugated to a large variety of target proteins via isopeptide bonds to internal lysine side chains in the target proteins. SUMO modification appears to alter protein activity in a number of ways, for example, by influencing subcellular localization and non-covalent protein:protein interactions. SUMO appears to have a chaperone-like activity, preventing protein aggregation, although it is not clear if this is related to its biological functions. Because of the ability of SUMO to prevent aggregation, the translational fusion of SUMO to other proteins often results in significantly improved solubility of those proteins during expression in *E. coli* [1–5].

© 2010 Elsevier Inc. All rights reserved.

Address correspondence to: Albert J. Courey, PhD, 607 Charles E. Young Drive, East, Los Angeles, CA, 90095-1569. Fax: 310-206-4038; courey@chem.ucla.edu.

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

Groucho (Gro)/transducin-like enhancer of split (TLE) family corepressors interact with many types of transcriptional repressors to regulate a myriad of developmental processes and signaling pathways. For example, *Drosophila* Gro helps mediate the transcriptional response to the Notch, Wnt, receptor tyrosine kinase, and Dpp signal transduction cascades and is critically required for anteroposterior and dorsoventral patterning of the embryo, sex determination, and multiple aspects of imaginal development such as eye and wing patterning [6–10]. The Q domain, which constitutes the N-terminal 133 amino acids of Gro is a highly conserved protein domain that mediates Gro oligomerization, a function that appears to be critical for efficient repression, perhaps by allowing Gro to spread along a chromatin template and establish a large transcriptionally silent domain.

We have developed an expression vector that allows production of a protein of interest with SUMO and an intervening 6 His tag fused to its N-terminus. Using this vector, we produced large quantities of a SUMO-GroQ domain fusion protein (which we term S-GroQ). In contrast to the unfused Q domain, S-GroQ is expressed at high levels and is highly soluble. Analysis of the hydrodynamic properties of S-GroQ by gel filtration chromatography, velocity sedimentation, and equilibrium sedimentation reveals that the Q domain forms an elongated tetramer, in addition to higher order oligomers.

For many applications involving proteins initially expressed as SUMO fusions, it is desirable to cleave off the SUMO moiety using a SUMO-specific protease such as Ulp1. Therefore, to increase the utility of the SUMO fusion protein expression system, we have developed a system to express simultaneously a His-tagged protein of interest fused to SUMO and Ulp1 also fused to SUMO. This leads to processing of the SUMO fusion proteins in the bacterial cells liberating the His-tagged protein of interest, which can be purified by nickel affinity chromatography. For the two SUMO fusion proteins that we have subjected to *in vivo* Ulp1-mediated cleavage (SUMO-Groucho and SUMO-Ulp1), the protein of interest remains soluble even after SUMO is removed suggesting that this is likely to be a generally applicable approach.

EXPERIMENTAL PROCEDURES

Expression constructs

To construct pMP-SUMO-H₆, sequences encoding the mature form of *Drosophila* SUMO lacking the final two amino acids that are cleaved off as a part of the normal maturation process were inserted into the pRSFDuet-1 (Novagen) at the Nco I site. The S-GroQ expression vector was constructed by inserting sequences encoding the Q domain (Groucho isoform A) into pMP-SUMO-H₆ between the Bam HI and Hind III sites. The resulting plasmid encodes a protein consisting of SUMO followed by a 6xHis tag followed by the Q domain. Expression is under control of the T7 promoter.

pMP-SUMO-H₆/SUMO-Ulp1 was made by inserting sequences encoding full-length SUMO into pMP-SUMO-H₆ between the Nde I and Eco RV sites, and then inserting sequences encoding the Ulp1 C-terminal 227 amino acids between the Kpn I and Xho I sites. The vector for coexpression of S-GroQ with SUMO-Ulp1 was made by inserting sequences encoding the Q domain (Groucho isoform A) into pMP-SUMO-H₆/SUMO-Ulp1 between the Bam HI and Hind III sites.

The vector for expression of SUMO fused to His-tagged Ulp1 was made as follows: sequences encoding the Ulp1 C-terminal 227 amino acids were inserted into pET16b between the Nco I and Xho I sites to create pET-Ulp1. Sequences encoding the mature form of *Drosophila* SUMO were inserted into pET-Ulp1 at the NcoI site to generate pET-SUMO-H₆-Ulp1.

Protein Expression and Purification

Media and solutions for auto-induction were as previously described [11]. For expression of the SUMO-GroQ fusion protein, *E. coli* BL21 (DE3) harboring the expression vectors described above encoding S-GroQ with or without sequences encoding SUMO-Ulp1 was grown at 37°C in non-inducing medium ZYM-505 containing kanamycin. Once the cultures reached an OD₆₀₀ of 0.6, the cells were diluted 250-fold into auto-inducing medium ZYM-5052, and allowed to grow overnight at 30° C at which point the OD₆₀₀ was ~10. Cells from 250 ml of culture were pelleted at 4000xg, 4°C and then resuspended in 25 ml of ice-cold lysis buffer (50 mM NaH₂PO₄, 300 mM NaCl, 10% glycerol, 0.1% NP-40, pH 8.0). The cells were lysed by two passages through a French press. The sample was then centrifuged at 15,000xg, 4°C. The pellet was discarded, and 1 ml of nickel nitrilotriacetic acid (Ni-NTA) resin was added to the supernatant. Binding was for 2 hours at 4° C with continuous mixing. The slurry was then poured into a column apparatus and the flow through was collected. The resin was washed twice with 20 ml of wash buffer (50 mM NaH₂PO₄, 300 mM NaCl, 10% glycerol, pH 8.0) containing 30 mM imidazole and then with 2 ml of the same buffer containing 50 mM imidazole. The His-tagged protein was eluted from the beads with elution buffer (50 mM Tris HCl, 100 mM NaCl, pH 8.0) containing increasing concentrations of imidazole (150 mM, 200 mM, 250 mM). Two 1 ml elutions were performed at each imidazole concentration. The purity of S-GroQ was assessed by SDS-PAGE, and the concentration was evaluated by Bradford assay using BSA as a standard.

The expression and purification of Ulp1 was essentially the same as that of S-GroQ, with the exception that cells harboring pET-SUMO-H₆-Ulp1 were grown in ZYM-5052 medium overnight to OD₆₀₀ > 10 at 25° C.

SUMO deconjugation assays

EPRS-rep (rep) and EPRS-rep covalently modified with SUMO (rep-SUMO) were expressed as GST-fusion proteins and purified as described previously [12]. Approximately 200 ng of the purified rep or rep-SUMO was incubated with 50 ng of recombinant Ulp1 in Ulp1 reaction buffer (50 mM Tris, pH 7.5, 150 mM NaCl, 1 mM DTT) at room temperature for 30min. The reaction was terminated by the addition of SDS-loading buffer.

Far-UV circular dichroism spectroscopy

Circular dichroism spectroscopy was performed on a Jasco J-715 spectrophotometer. S-GroQ protein was dialyzed to remove imidazole, and then diluted to 0.2 mg/ml with 100 mM NaCl, 50 mM Tris, pH 8.0. Spectra were analyzed using Softsec (Version 1.2). Secondary structure composition was calculated by the self-consistent method [13].

In-gel trypsinolysis and mass spectroscopy

In-gel trypsinolysis with sequencing-grade Trypsin Gold (Promega) was carried out as described in the PromegaTM manual. The LC-MS/MS spectra of the tryptic digests were obtained on an Applied Biosystems QSTAR (quadrupole-time-of-flight) mass spectrometer coupled with an LC Packings nanoflow HPLC system, through a nanoelectrospray ionization source (Protana). Product ion (MS/MS) spectra of the fragments were recorded and then submitted to the Mascot database search engine (Matrix Science) for protein identification. Positive protein identification was based on standard Mascot scoring criteria (p<0.05) for statistical analysis of the LC-MS/MS data.

Gel filtration analysis

Gel filtration was performed on an FPLC apparatus using a 24-ml Superose 6 column calibrated with the following standards: aldolase (Stokes radius = 48 Å), catalase (52 Å), ferritin (61 Å), thyroglobulin (85 Å), and blue dextran 2000. S-GroQ (240 µl) was chromatographed at 4° C in 200 mM NaCl, 50 mM Tris, pH 7.4. The Stokes radius of the unknown was determined by comparison to a standard curve of K_{av} (equal to $(V_e - V_o)/(V_t - V_o)$, where V_e is elution volume, V_o is the void volume, V_t is the total volume of the column), vs. the logarithm of the Stokes radius.

Velocity sedimentation

Velocity sedimentation analysis was performed in a Beckman Optima™ XL-A analytical ultracentrifuge using absorption optics at 280 nm. Twelve mm pathlength double sector cells were used. Concentration of S-GroQ was adjusted to ~0.4–0.5 mg/ml in 100 mM NaCl, 50 mM Tris, pH 8.0. Runs were at 50,000 rpm, and data were recorded every 2 minutes. The sedimentation coefficient was determined from the maximum of a sedimentation coefficient distribution plot, i.e., a $g(s)$ versus s plot, calculated from the following equations: 1. $s = (1/\omega^2 t) \ln(r/r_m)$, where s is the sedimentation coefficient, ω is the angular velocity, r is the radius, r_m is the radius of the meniscus, and t is time. 2. $g(s) = (dc/dt)(1/c_o)(\omega^2 t^2 / \ln(r_m/r))(r^2/r_m^2)$, where c is the concentration and c_o is the initial concentration of S-GroQ.

Equilibrium sedimentation

Equilibrium sedimentation analysis was performed on a Beckman Optima XL-A analytical ultracentrifuge using absorption optics at 280 nm and a 12 mm pathlength double sector cell. S-GroQ was adjusted to a concentration of 0.5 mg/ml in 100 mM NaCl, 50 mM Tris, pH 8.0. Sedimentation equilibrium profiles were determined at 4° C at speeds of 8000, 10,000, and 14,000 rpm, allowing 28 hours for equilibrium to be established at each speed. The data were initially fitted with a nonlinear least-squares exponential fit for a single ideal species using the Beckman Origin-based software (Version 3.01). The Beckman global analysis software (the "multifit" option of the above mentioned software) was then used to analyze multiple scans, corresponding to the different speeds, simultaneously for various models. A partial specific volume of 0.722 ml/g calculated from the amino acid composition and corrected to 4°C was used [14,15]. The apparent equilibrium constants determined for the monomer to tetramer and the monomer to 16-mer transitions were $7.30 \times 10^{14} \text{ M}^{-3}$ and $1.44 \times 10^{74} \text{ M}^{-15}$, respectively. Using these equilibrium constants and the total initial concentration of S-GroQ, we calculated the concentrations of the monomer, tetramer, and 16-mer.

RESULTS

A SUMO fusion protein expression vector and its use to express S-GroQ

We created a T7 promoter expression vector for the production of SUMO fusion proteins containing sequences encoding the mature form of SUMO followed by a 6xHis tag and a multiple cloning site [Figure 1A]. Sequences encoding the Gro Q domain (133 amino acids) were then cloned into the MCS to create an in-frame fusion of SUMO to the Q domain termed S-GroQ [Figure 1B], which was expressed in BL-21 cells. Unlike the unfused Q domain, which is largely insoluble (data not shown), S-GroQ was found in the soluble fraction after cell lysis in a French Press [Figure 1C]. These findings are consistent with other studies showing that translational fusion of proteins to SUMO greatly improves their solubility and yield [1–5], and opens up the possibility of future structural studies of the Q domain by NMR or X-ray crystallography.

S-GroQ was purified by Ni-NTA chromatography and its identity confirmed by tryptic digestion followed by mass spectrophotometric (MS) analysis of the peptides [Figure 1C]. SDS-PAGE followed by Coomassie staining indicated that the fusion protein was at least 95% pure. A Bradford assay indicates a yield of approximately 1 mg of purified fusion protein per gram of wet *E. coli*. The only significant contaminants represent breakdown products of the fusion protein as indicated by both MS analysis and Western blotting [Figure 1C and data not shown]. Although we did not attempt further preparative fractionation steps, additional purification as might be required for atomic resolution structure determination could probably be achieved by an additional chromatographic step (e.g., a finishing ion exchange or size exclusion step). The CD spectrum of the S-GroQ preparation indicates that the fusion protein is folded and contains 35% α -helix, 13% β -sheet, 18% β -turn, and 33% random coil [data not shown].

Gel filtration and velocity sedimentation suggest that S-GroQ is a tetramer with an eccentric shape

To determine the native size of S-GroQ oligomers, the purified samples were analyzed by gel filtration chromatography. Two independent S-GroQ samples were analyzed under identical conditions to verify reproducibility [Figure 2, red and green traces]. Both traces reveal a major peak corresponding to a Stokes radius of 72 Å. Assuming spherical symmetry, this corresponds to an apparent molecular mass of ~510 kDa.

Velocity sedimentation was also used to characterize the oligomerization state of S-GroQ revealing a major species with an apparent sedimentation coefficient of 4.4 S [Figure 3A and B]. Assuming spherical symmetry, this corresponds to a molecular mass of ~60–70 kDa.

The discrepancy between the apparent molecular masses as determined by gel filtration and velocity sedimentation indicates a significant departure from spherical symmetry. Combining the Stokes radius with the sedimentation coefficient yields a calculated molecular mass of 129 kDa and a frictional coefficient ratio (f/f_0) of 2.16 [Table 1]. Since the subunit molecular mass of S-GroQ is 27 kDa, the calculated molecular mass is in reasonable agreement with previous measurements of the native size of full-length Groucho, which showed that it was mostly likely a tetramer. The frictional coefficient ratio (which would be unity for a sphere) suggests a large departure from spherical symmetry, and, for example, could be accounted for by a prolate ellipsoid with an axial ratio of 24:1, or a hydrated prolate ellipsoid with an axial ratio of 10:1 [Table 1].

Equilibrium sedimentation analysis suggests that S-GroQ is predominantly tetrameric, but also forms higher order oligomers

S-GroQ was further analyzed by equilibrium sedimentation [Figure 4]. In attempting to fit the data to a single exponential, the best fit was obtained by setting the molecular mass at of S-GroQ at 103 kDa [Figure 4A]. This is in reasonable agreement with the mass calculated by combining the Stokes radius and sedimentation coefficient measurements, and is once again consistent with a tetramer. However, the non-random residuals to the fit clearly indicate molecular weight heterogeneity. The fit improved only slightly when we assumed a monomer/tetramer mixture (Figure 4B), but improved dramatically when we assumed a mixture of monomers, tetramers, and a higher order oligomer, of which 16-mers gave the best fit [Figure 4C]. The apparent equilibrium constants for the monomer-tetramer and monomer-16-mer transitions were $7.30 \times 10^{14} \text{ M}^{-3}$ and $1.44 \times 10^{74} \text{ M}^{-15}$, respectively. Using these values, we calculate a ratio by mass of monomer to tetramer to 16-mer in the starting material of 3:6:1.

Expression of a SUMO-Ulp1 fusion protein and its use to process S-GroQ in bacterial cells

SUMO specific proteases such as Ulp1 recognize the C-terminal end of SUMO and hydrolyze peptide or isopeptide bonds to the carboxyl group of the C-terminal glycine residue [16]. Therefore, such proteases provide a convenient way of processing SUMO fusion proteins. In our hands, *Drosophila* Ulp1 is difficult to express in *E. coli*, forming insoluble aggregates (data not shown). We therefore explored the possibility of expressing Ulp1 as a SUMO fusion protein to improve the yield and activity. For these studies, we expressed Ulp1 fused at its N-terminus to SUMO with an intervening 6xHis tag. We found that the fusion protein catalyzed its own cleavage such that very little unprocessed fusion protein was observed in cell lysates. The liberated His-tagged Ulp1 could be readily purified from the soluble fraction after cell lysis [Figure 5A]. It was active as shown in an activity assay using the SUMO conjugated rep domain from glutamyl-prolyl tRNA synthetase (EPRS) as a substrate (Figure 5B).

Since the SUMO-Ulp1 fusion protein can cleave itself in bacteria while remaining soluble, it seemed likely that co-expression of SUMO-Ulp1 with another SUMO fusion protein of interest might be a useful strategy for expression of proteins in bacteria. We therefore generated a dual protein expression vector for simultaneous expression of SUMO-Ulp1 and a His-tagged SUMO fusion protein of interest [Figure 6A]. Using this vector, we coexpressed S-GroQ and SUMO-Ulp1 in BL21 cells and then purified the His-tagged GroQ by nickel affinity chromatography. To allow direct comparison of the yields, we carried out a parallel experiment in which we expressed and purified S-GroQ from BL21 cells not coexpressing SUMO-Ulp1. Yields of the processed and unprocessed forms of S-GroQ after purification from the soluble fraction of the cell lysates were comparable indicating that Q domain remains soluble after cleavage and purification to remove the SUMO moiety [Figure 6B].

DISCUSSION

Fusion of SUMO to other proteins significantly increases the solubility of proteins to which it is attached as shown here and by others [1–5]. It is possible that this reflects one of the normal roles of SUMO in modulating protein function, although there is no evidence that directly supports this notion. For example, sumoylated proteins such as Polycomb group proteins and promyelocytic leukemia protein have been found associated with dense protein bodies in the nucleus such as the PML oncogenic domains (PODs, also termed nuclear bodies) and Polycomb bodies [17,18]. Association of proteins with these bodies may serve to activate them or to sequester them until some signaling event triggers their release. One role of SUMO in these bodies may be to prevent the highly concentrated proteins from forming non-native irreversible aggregates.

The mechanism by which SUMO prevents protein aggregation is unknown. One possibility is that the high solubility of SUMO itself counteracts the propensity of less soluble proteins to which it is attached to aggregate. However, if this were true, then we might expect cleavage of SUMO from these proteins during expression in *E. coli* to lead to aggregation. As we have shown here, cleavage of SUMO from either of two proteins (Ulp1 or Gro) during expression does not lead to aggregation. Therefore, we favor the notion that SUMO serves to prevent aggregation of proteins during their initial folding.

The findings presented here that the Q domain mediates the formation of elongated tetramers as well as higher order oligomers is consistent with the notion that Gro is recruited to DNA as a tetramer (the most stable form of Gro in solution) and then spreads along chromatin, which serves to template the formation of higher order oligomers [19]. The mechanism by which the Q domain forms multiple size oligomers is unknown, but one

likely possibility is “domain swapping” [20]. In this process, one part of a protein:protein interaction motif (for example, one of the amphipathic α -helices in a Q domain coiled-coil motif) exchanges places with an identical domain from another subunit. This process can mediate the growth of oligomers of indefinite length. Further assessment of this model may be greatly facilitated by efforts to determine the atomic resolution structure of Q domain oligomers.

Acknowledgments

We thank Rachel Loo for technical assistance in mass spectrometry. This work was supported by NIH grants GM63596 and GM44522 to AJC; and by UC-BIOSTAR grant S98-86 and a grant from Shanbrom Foundation to AMH. PDH was a trainee on an NIH Genetics training grant to UCLA.

REFERENCES

1. Butt TR, Edavettal SC, Hall JP, Mattern MR. SUMO fusion technology for difficult-to-express proteins. *Protein Expr Purif* 2005;43:1–9. [PubMed: 16084395]
2. Malakhov MP, Mattern MR, Malakhova OA, Drinker M, Weeks SD, Butt TR. SUMO fusions and SUMO-specific protease for efficient expression and purification of proteins. *J Struct Funct Genomics* 2004;5:75–86. [PubMed: 15263846]
3. Marblestone JG, Edavettal SC, Lim Y, Lim P, Zuo X, Butt TR. Comparison of SUMO fusion technology with traditional gene fusion systems: enhanced expression and solubility with SUMO. *Protein Sci* 2006;15:182–189. [PubMed: 16322573]
4. Panavas T, Sanders C, Butt TR. SUMO fusion technology for enhanced protein production in prokaryotic and eukaryotic expression systems. *Methods Mol Biol* 2009;497:303–317. [PubMed: 19107426]
5. Zuo X, Mattern MR, Tan R, Li S, Hall J, Sterner DE, Shoo J, Tran H, Lim P, Sarafianos SG, Kazi L, Navas-Martin S, Weiss SR, Butt TR. Expression and purification of SARS coronavirus proteins using SUMO-fusions. *Protein Expr Purif* 2005;42:100–110. [PubMed: 15939295]
6. Parkhurst SM. Groucho: making its Marx as a transcriptional co-repressor. *Trends Genet* 1998;14:130–132. [PubMed: 9594657]
7. Paroush Z, Finley RL Jr, Kidd T, Wainwright SM, Ingham PW, Brent R, Ish-Horowicz D. Groucho is required for Drosophila neurogenesis, segmentation, and sex determination and interacts directly with hairy-related bHLH proteins. *Cell* 1994;79:805–815. [PubMed: 8001118]
8. Chen G, Courey AJ. Groucho/TLE family proteins and transcriptional repression. *Gene* 2000;249:1–16. [PubMed: 10831834]
9. Fischer-Vize JA, Vize PD, Rubin GM. A unique mutation in the Enhancer of split gene complex affects the fates of the mystery cells in the developing Drosophila eye. *Development* 1992;115:89–101. [PubMed: 1638995]
10. Buscarlet M, Stifani S. The 'Marx' of Groucho on development and disease. *Trends Cell Biol* 2007;17:353–361. [PubMed: 17643306]
11. Studier FW. Protein production by auto-induction in high density shaking cultures. *Protein Expr Purif* 2005;41:207–234. [PubMed: 15915565]
12. Nie M, Xie Y, Loo JA, Courey AJ. Genetic and proteomic evidence for roles of Drosophila SUMO in cell cycle control, Ras signaling, and early pattern formation. *PLoS One* 2009;4:e5905. [PubMed: 19529778]
13. Sreerama N, Woody RW. A self-consistent method for the analysis of protein secondary structure from circular dichroism. *Anal Biochem* 1993;209:32–44. [PubMed: 8465960]
14. Cohn, EJ.; Edsall, JT. *Proteins, Amino Acids and Peptides as Ions and Dipolar Ions*. New York: Reinhold Publishing Corporation; 1943.
15. Laue, TM.; Shah, BD.; Ridgeway, TM.; Pelletier, SL. *Analytical Ultracentrifugation in Biochemistry and Polymer Science*. Cambridge, Great Britain: The Royal Society of Chemistry; 1992.

16. Kim JH, Baek SH. Emerging roles of desumoylating enzymes. *Biochim Biophys Acta* 2009;1792:155–162. [PubMed: 19162180]
17. Kagey MH, Melhuish TA, Wotton D. The polycomb protein Pc2 is a SUMO E3. *Cell* 2003;113:127–137. [PubMed: 12679040]
18. Bernardi R, Pandolfi PP. Structure, dynamics and functions of promyelocytic leukaemia nuclear bodies. *Nat Rev Mol Cell Biol* 2007;8:1006–1016. [PubMed: 17928811]
19. Courey AJ, Jia S. Transcriptional repression: the long and the short of it. *Genes Dev* 2001;15:2786–2796. [PubMed: 11691830]
20. Liu Y, Eisenberg D. 3D domain swapping: as domains continue to swap. *Protein Sci* 2002;11:1285–1299. [PubMed: 12021428]
21. Siegel LM, Monty KJ. Determination of molecular weights and frictional ratios of proteins in impure systems by use of gel filtration and density gradient centrifugation. Application to crude preparations of sulfite and hydroxylamine reductases. *Biochim Biophys Acta* 1966;112:346–362. [PubMed: 5329026]
22. Freifelder, DM. *Physical Biochemistry: Applications to Biochemistry and Molecular Biology* W.H. Freeman New York: 1982.

Fig. 1.

Expression and purification of S-GroQ. **A.** Diagram showing the essential features of pMP-SUMO-H₆, which is derived from the pRSFDuet-1 expression vector (Novagen). The indicated multiple cloning site (MCS) is MCS1 from that vector. **B.** Domain structure of Gro and the S-Gro-Q fusion protein. **C.** S-GroQ purification by Ni-NTA chromatography was monitored by 12% SDS PAGE. The lanes are: 1) markers (sizes labeled in kilodaltons); 2) uninduced culture; 3) induced culture; 4) flow through; 5–13) imidazole washes and elutions carried out as described in Materials and Methods using the following concentrations of imidazole: 30 mM (lanes 5 and 6), 50 mM (lane 7), 150 mM (lanes 8 and 9), 200 mM (lanes 10 and 11), and 250 mM (lanes 12 and 13). Bands were excised from lanes 8–13 and

analyzed by mass spectrometry after in-gel trypsinolysis. Tryptic fragments identified in each band are indicated next to the gel. The major species has an apparent molecular mass (32 kDa) that agrees well with the predicted molecular mass of the S-GroQ monomer (27 kDa), and the MS analysis confirms that this is the intact fusion protein. The slower migrating bands are likely to be SDS resistant oligomers. The fast migrating band appears to be a C-terminally truncated form of S-GroQ. The S-GroQ peptide sequence is shown below the gel: green lettering indicates SUMO sequence, blue lettering indicates Gro sequence, tryptic peptides (abcedf) identified by MS are boxed.

Fig. 2.

Gel filtration analysis. Two independent preparations of S-GroQ were analyzed by Superose 6 gel filtration chromatography. Positions of native protein standards as determined in a separate run are indicated by the arrows, which are labeled with the Stokes radii (Å): thyroglobulin 669 kDa (Stokes radius: 85 Å); ferritin, 440 kDa (61 Å); catalase, 232 kDa (52 Å); aldolase, 158 kDa (48 Å). As shown at the bottom, fractions were subjected to SDS-PAGE and Coomassie staining showing that the major peak corresponds to S-GroQ. The left-hand lane shows molecular mass markers labeled in kilodaltons.

Fig. 3. Velocity sedimentation. *A.* Velocity sedimentation was performed at 50,000 rpm, 20°C. Data are shown as absorbance at 280 nm vs. the distance from the center of rotation. Measurements were taken in intervals of 2 minutes, but only data taken at 10 minute intervals (10 – 120 minutes) during sedimentation at 50,000 rpm are shown. *B.* The velocity sedimentation coefficient distribution ($g(s)$ vs. sedimentation coefficient) for S-GroQ. The mathematical quantity $g(s)$ is proportional to concentration (see Materials and Methods). The most prominent species in the sample possesses a sedimentation coefficient of 4.4S.

Fig. 4.

Equilibrium sedimentation. Data were fit assuming *A.* A single species with a mass of 103,000 daltons, which approximately corresponds to a tetramer; *B.* a mixture of monomer and tetramer; or *C.* a mixture of monomer, tetramer, and 16-mer. Sedimentation equilibrium profiles depicted were measured at 8000 rpm. The lower plot in each panel depicts absorbance as a function of distance from the center of rotation. The upper plot in each panel depicts the deviation of the experimental data from the exponential fits.

Fig. 5.

Expression and purification of His-tagged Ulp1. **A.** His-tagged Ulp1 was expressed and purified by Ni-NTA affinity chromatography as described in Materials and Methods. Purification was monitored by SDS PAGE and Sypro Ruby staining. The lanes are: 1) Input; 2) Flow through (FT); 3–4) 30 mM imidazole washes; 5) marker (sizes labeled in kilodaltons); 6–10) elutions with the following concentrations of imidazole: 50 mM (lane 6), 100 mM (lane 7), 150 mM (lane 8), 200 mM (lane 9), 250 mM (lane 10); 11) BSA (1 µg). **B.** SUMO deconjugation with His-tagged Ulp1. The SUMO-conjugated EPRS rep domain was incubated with His-tagged Ulp1 and the products visualized by SDS-PAGE and Sypro Ruby staining. The lanes are: 1) rep domain; 2) SUMO-conjugated rep domain; 3) marker (sizes labeled in kilodaltons); 4) rep domain after incubation with His-tagged Ulp1; 5) SUMO-conjugated rep domain after incubation with His-tagged Ulp1.

Fig. 6.

Coexpression of S-GroQ and SUMO-Ulp1. **A.** Diagram showing the essential features of pMP-SUMO-H₆/SUMO-Ulp1, which is derived from the pRSFDuet-1 expression vector (Novagen). The indicated multiple cloning site (MCS) is MCS1 from that vector. **B.** S-GroQ was expressed without SUMO-Ulp1 coexpression using pMP-SUMO-H₆ (lane 2), or with SUMO-Ulp1 coexpression pMP-SUMO-H₆/SUMO-Ulp1 (lane 3). The products were purified by Ni-NTA affinity chromatography and displayed by SDS-PAGE followed by Coomassie staining. The arrow indicates a band of the expected size for unprocessed S-GroQ, while the arrowhead indicates a band of the expected size for GroQ generated by Ulp1 cleavage of S-GroQ. Lane 1 show molecular mass markers labeled with sizes in

kilodaltons. The purity of this material is lower than that of the material shown in Figure 1 (and used in the analysis in Figures 2–4), because these are small scale preps and the conditions were not fully optimized as they were for the prep shown in Figure 1.

Table 1

Molecular mass and shape of S-GroQ from gel filtration and velocity sedimentation analysis

Stokes radius (Å)	Sedimentation coefficient (S)	Molecular Mass (kDa) ^a	f/f_0 ^b	Axial ratio (prolate ellipsoid) ^c
72	4.4	129	2.16	24

^aThe molecular mass was calculated from $M=6\pi\eta N a s/(1-\bar{v}\rho)$ where M is the molar mass, η is the viscosity, N is Avogadro's number, a is the Stokes radius, s is the sedimentation coefficient, $\bar{v}$ is the protein partial specific volume (0.722 ml/g), and ρ is the density of the solvent (1.006 g/ml) [21].

^bThe frictional coefficient ratio was calculated from $f/f_0 = a/(3\bar{v}M/4\pi N)^{1/3}$ [21].

^cThe axial ratio was determined as described elsewhere, assuming no hydration. If a value of 0.6 g water/g protein is assumed, the axial ratio of the corresponding prolate ellipsoid is 10 [22].