Discovery and engineering of enhanced SUMO protease enzymes

Received for publication, May 26, 2018, and in revised form, June 28, 2018 Published, Papers in Press, July 5, 2018, DOI 10.1074/jbc.RA118.004146

Yue-Ting K. Lau†,‡, Vladimir Baytstok†, Tessa A. Howard§, Brooke M. Fiala‡, JayLee M. Johnson†, Lauren P. Carter‡, David Baker¶,  Christopher D. Lima***, and Christopher D. Bahl††‡‡

From the †Institute for Protein Design, ‡Department of Biochemistry, and §Howard Hughes Medical Institute, University of Washington, Seattle, Washington 98195, the ¶Structural Biology Program and ***Howard Hughes Medical Institute, Sloan Kettering Institute, Memorial Sloan Kettering Cancer Center, New York, New York 10065, and the ‡‡Institute for Protein Innovation, Boston, Massachusetts 02115

Edited by George N. DeMartino

Small ubiquitin-like modifier (SUMO) is commonly used as a protein fusion domain to facilitate expression and purification of recombinant proteins, and a SUMO-specific protease is then used to remove SUMO from these proteins. Although this protease is highly specific, its limited solubility and stability hamper its utility as an in vitro reagent. Here, we report improved SUMO protease enzymes obtained via two approaches. First, we developed a computational method and used it to re-engineer WT Ulp1 from Saccharomyces cerevisiae to improve protein solubility. Second, we discovered an improved SUMO protease via genomic mining of the thermophilic fungus Chaetomium thermophilum, as proteins from thermophilic organisms are commonly employed as reagent enzymes. Following expression in Escherichia coli, we found that these re-engineered enzymes can be more thermostable and up to 12 times more soluble, all while retaining WT-or-better levels of SUMO protease activity. The computational method we developed to design solubility-enhancing substitutions is based on the RosettaScripts application for the macromolecular modeling suite Rosetta, and it is broadly applicable for the improvement of solution properties of other proteins. Moreover, we determined the X-ray crystal structure of a SUMO protease from C. thermophilum to 1.44 Å resolution. This structure revealed that this enzyme exhibits structural and functional conservation with the S. cerevisiae SUMO protease, despite exhibiting only 28% sequence identity. In summary, by re-engineering the Ulp1 protease and discovering a SUMO protease from C. thermophilum, we have obtained proteases that are more soluble, more thermostable, and more efficient than the current commercially available Ulp1 enzyme.

The ability to express and purify recombinant proteins is critical to studying their structure and function. Genetic fusion to the C terminus of small ubiquitin-like modifier (SUMO) can chaperone folding and increase the soluble yield obtained from heterologous protein expression (1–4). An affinity tag, such as polyhistidine, is generally fused to the N terminus of SUMO to facilitate purification. The SUMO fusion is readily removed via digestion with the highly specific SUMO protease, which can leave the target protein intact without residual amino acids at its N terminus (i.e., no “scar” residues remain).

The most commonly employed SUMO fusion system utilizes the Saccharomyces cerevisiae SUMO protein Smt3 and SUMO protease Ulp1 (1, 2). In vivo, Ulp1 functions as an essential isopeptidase that is responsible for processing pre-Smt3 and generating mature Smt3 that ends with the canonical di-glycine motif. Ulp1 also functions by cleaving Smt3 off of proteins that have been conjugated post-translationally (3). Natively, Ulp1 is a 621-amino acid protein, and residues 403–621 constitute a conserved protease domain (2). This fragment can be purified without the N-terminal residues, is constitutively active, and can be used for removing genetically fused SUMO domains from recombinant protein in vitro in reactions that are analogous to the processing of pre-Smt3 (2); we will refer to this construct as Ulp1_WT. The active site is characteristic of papain-like cysteine proteases, and the catalytic triad consists of a cysteine nucleophile (Cys-580) coordinated by a histidine (His-514) and an acid (Asp-531). When functioning as a peptidase, Ulp1_WT is highly tolerant to sequence diversity at P residue positions; the only restriction is that the P1 residue cannot be a proline (6).

Commercially available Ulp1_WT is prone to precipitation at room temperature and includes detergent to maintain solubility (Thermo Fisher Scientific, catalog no. 12588018). Precipitation of the enzyme can lead to incomplete digestion of SUMO fusion constructs as well as potentially nucleate
aggregation of target proteins. To uncover determinants of Ulp1_WT that lead to poor solubility, we analyzed the structure of Ulp1_WT (PDB code 1EUV) using Rosetta and identified 10 hydrophobic residues that project outward toward solvent from the protein surface, not including residues involved in SUMO binding or the active site (Fig. 1A and Table 1). These residues are far enough from functional sites that they are unlikely to contribute to protease activity in vitro.

In this study, we sought to improve the current state-of-the-art protease by two orthogonal approaches. First, we used computational protein design to engineer Ulp1_WT and remove the solvent-exposed hydrophobic surfaces by mutating nonpolar amino acids to polar amino acids. Second, we mined the genome of a thermophilic organism for a homologous enzyme. We assessed these new proteases for solubility, thermostability, and enzyme activity. The enzymes reported here exhibit improved behavior as in vitro reagents when compared with Ulp1_WT.

**Results and discussion**

**Computational design of solubility-enhancing mutations to Ulp1**

Previous studies that leveraged computational methods to enhance protein solubility and minimize aggregation focused on restricting large contiguous hydrophobic patches during design of the entire protein surface (8) or on modulation of surface charge (9). For this application, our goal was to minimize the amount of nonessential hydrophobic surface exposed to solvent while introducing the fewest possible number of mutations, as preserving enzymatic activity was paramount.

Using Rosetta, we developed a generally applicable computational method that identifies hydrophobic residue positions on the surface of a protein and determines amino acid substitutions to polar residues that yield low-energy solutions (see supporting information for the RosettaScripts XML protocol). To do this, the algorithm performs iterative rounds of flexible backbone design (10), and the positions of all Ca atoms are constrained to favor retention of the starting coordinates. We utilized the previously reported crystal structure of Ulp1_WT in complex with Smt3 (PDB code 1EUV) as the starting model (2). In total, 10 hydrophobic residues that project toward solvent from the protein surface were detected and designed (Table 1). Residues on Ulp1_WT that form the interface with the substrate Smt3 were detected by the algorithm, and the catalytic triad residues were manually specified; these residues were not permitted to mutate (Fig. 1B).

We generated 10 designs and selected four for testing in the wet laboratory (which we have named Ulp1_R1–Ulp1_R4). These four variants were selected because they were maximally different from one another with regard to amino acid sequence and exhibited high sequence similarity at designed positions with human and mouse SUMO proteases (see supporting information for the designed structural models). The algorithm’s results converged at four of the designed residue positions, and in each case, this amino acid is predicted to form a new salt bridge. Mutation of Ile-567 to Arg can form a salt bridge with Glu; mutation of Ile-605 to Glu enables a salt bridge with Arg; and mutation of Ala-539 and Phe-542 to Asp and Lys, respectively, enables these two residues to form a salt bridge. These charge–charge interactions are favored by the energy function over mutations that do not result in the formation of an ionic bond.

**Table 1**

| Residue | Ulp1_WT | Ulp1_R1 | Ulp1_R2 | Ulp1_R3 | Ulp1_R4 |
|---------|---------|---------|---------|---------|---------|
| 420     | Ala     | Asp     | Asn     | Asp     | Glu     |
| 444     | Ala     | Glu     | Glu     | Glu     | Arg     |
| 539     | Ala     | Asp     | Asp     | Asp     | Glu     |
| 540     | Met     | Ser     | Asp     | Ser     | Ser     |
| 542     | Phe     | Lys     | Lys     | Lys     | Lys     |
| 543     | Ala     | Gln     | Arg     | Gln     | Gln     |
| 553     | Met     | Lys     | Arg     | Gln     | Lys     |
| 567     | Ile     | Arg     | Arg     | Arg     | Arg     |
| 601     | Tyr     | Gln     | Gln     | Ser     | Gln     |
| 605     | Ile     | Glu     | Glu     | Glu     | Glu     |

*Residues are numbered according to PDB entry 1EUV.

**Figure 1.** The SUMO protease Ulp1_WT (colored) is shown in complex with a substrate SUMO protein Smt3 (gray). Nonpolar amino acids of Ulp1_WT selected for computational design are colored orange, and residues that contact SUMO (yellow) or are part of the enzyme catalytic triad (green) were not permitted to change during design; all other residues are colored cyan. Models for Ulp1_WT and Smt3 are from PDB entry 1EUV. A, A molecular surface rendition. B, Smt3 is shown as a semitransparent molecular surface. The Ulp1_WT main chain is shown as a cartoon model, and amino acid side chains are shown as sticks. The side chains of the polar amino acid substitutions from Ulp1_R1–Ulp1_R4 are colored magenta; these and the corresponding side chains of Ulp1_WT are shown in boldface type for emphasis.
SUMO protease engineering

Identification of a SUMO protease from thermophilic fungi

The SUMO system is unique to eukaryotes, and whereas there are no known hyperthermophilic eukaryotes, the single-celled fungus Chaetomium thermophilum, originally isolated from compost, is a thermophile that can grow at temperatures up to 60 °C (11). This organism has generated considerable interest within the structural biology community, as proteins from organisms that grow at high temperature are generally more stable and more amenable to crystallization than proteins from mesophiles (12).

The genome of C. thermophilum was recently sequenced (13), and currently there is no annotated SUMO protease gene. To identify candidates, we performed a BLASTp search (14) with Ulp1_WT as the query sequence. The top scoring alignment was to UniProtKB accession number G0RZV7, which is annotated as “specific protease-like protein” and exhibits 28% sequence identity over the aligned region with Ulp1_WT (Fig. 2A). To further investigate, we predicted the three-dimensional structure of the domain that aligns to Ulp1_WT using Robetta (15). The predicted model indicates this sequence adopts a structure that is highly similar to Ulp1_WT (Fig. 2C). The catalytic triad and active site residues were structurally conserved, and surprisingly, many of the residues that mediate binding to the SUMO substrate were also conserved (see supporting information for the predicted model). This suggested that the sequence encoded a SUMO protease, which we will refer to henceforth as Cth SUMO protease (or Cth), and it seemed likely that this enzyme could recognize the S. cerevisiae SUMO protease substrate Smt3.

SUMO protease variants exhibit enhanced solubility and thermostability

We began characterization of the protease variants by expressing each protein in Escherichia coli with a polyhistidine tag, followed by purification via immobilized metal-affinity chromatography and size-exclusion chromatography. The overall expression yields were similar; 10–35 mg of purified protein was obtained via log-linear regression analysis of protein solubility over Ulp1_WT; Ulp1_R3 is the most soluble, with over a 12-fold increase in maximum concentration (Fig. 3). The least soluble of the designed proteases is Ulp1_R4, which exhibits nearly a 2-fold increase in solubility (Table 2). Cth SUMO protease is also more soluble than Ulp1_WT, with a one-third increase in solubility. These data are consistent with the model that solvent-exposed hydrophobic residues drive lower solubility for Ulp1_WT in aqueous solution.

Next, we sought to determine the effect of design on the structure and fold of the Rosetta-engineered Ulp1 variants and to further characterize the putative Cth SUMO protease. Steady-state CD spectra revealed that all of the proteases contain a mixed α/β character at 20 °C (Fig. 4A). As expected, the Cth SUMO protease melts at higher temperature than its mesophilic homologue Ulp1_WT, and upon heating, Ulp1_WT and Cth SUMO protease both undergo a cooperative unfolding transition with melting temperatures at 40 and 50 °C, respectively (Fig. 4B). A visible protein precipitate was observable in the cuvette following thermal denaturation for Ulp1_WT and Cth SUMO protease. Similar to other Ulp1-like SUMO proteases, the full-length Cth SUMO protease includes a large N-terminal domain that may be required for protein stability at the higher temperatures at which C. thermophilum can grow. Curiously, only Ulp1_R4 of the Rosetta-designed variants undergoes a clear, cooperative unfolding transition upon heating. For all designs, some of the structural character reported by the far-UV wavelength measurements appears to melt with a Tm that ranges from 31 to 38 °C when monitoring the CD signal at 208 nm (Fig. 4B), and this may represent a soluble, partially unfolded state, as we did not observe precipitate in the cuvette for any of the designed variants after heating.

Enzyme kinetics of SUMO protease variants

Sequence changes can impact enzymatic function, even when distal to the catalytic machinery (19). Thus, we sought to determine the effect of mutation on the rate of proteolysis for the Ulp_R1–4 variants and to determine whether Cth SUMO protease exhibited activity on the S. cerevisiae substrate protein Smt3. To accomplish this, we used an EGFP-Smt3-mCherry linear fusion construct whereby the cleavage after Smt3 is monitored by a loss of FRET (20, 21).

Kinetic analysis revealed that the Rosetta design Ulp1 variants all exhibit Vmax and Km values comparable with those of the WT enzyme, indicating that solubility enhancement had little effect on enzyme activity (Fig. 5A). By extension, the Rosetta variants likely retain WT specificity for Smt3, because a change in substrate specificity should result in a change to the measured specificity constant (kcat/Km), whereas the Rosetta variants exhibit similar specificity constants for the cleavage of EGFP-Smt3-mCherry as compared with Ulp1_WT (Table 2). The Cth SUMO protease has roughly an order of magnitude higher Km but also a Vmax that is over 4-fold faster than Ulp1_WT (Fig. 5B, Table 2). Although the resulting lower specificity constant of the Cth SUMO protease may be due to intrinsic properties of the enzyme, it is more likely the result of S. cerevisiae Smt3 being a suboptimal substrate. When used in vitro with high substrate concentration, the weak Km may not be problematic, and the higher maximum rate could make Cth SUMO protease the superior reagent enzyme. At low substrate concentrations, or when the reaction needs to go to completion quickly, the Rosetta-designed Ulp1 variants would be the better reagents. Regardless, these data clearly demonstrate that Cth SUMO protease (UniProtKB accession number G0RZV7) encodes a bona fide SUMO protease.
X-ray crystal structure of Cth SUMO protease

To determine the structural basis for Cth SUMO protease activity and to assess the accuracy of the Robetta structure prediction, we crystallized the protease domain of the Cth SUMO protease. We solved the X-ray crystal structure by molecular replacement using the structure of Ulp1_WT as the search model after removal of loops and side chains that differed based on sequence alignments (PDB code 1EUV). Molecular replacement yielded a solution that was improved by iterative rounds of building and refinement. The structure revealed one molecule of Cth SUMO protease per asymmetric unit. The model was refined to 1.44 Å with R\text{work} and R\text{free} values of 11.9 and 14.8%, respectively, and excellent geometry (Fig. 6; also see Table 3).

Overall, the structure of Cth SUMO protease closely resembles previously determined Ulp1/SENP structures (Fig. 6A) (2, 21–25) and aligns to the S. cerevisiae Ulp1_WT-Smt3 structure (PDB code 1EUV, chain A) with an RMSD of 1.7 Å (Cα over 187...
residues; Fig. 6A). The catalytic Asp-His-Cys triad and the glutamine that contributes to the formation of the oxyanion hole exhibit similar conformations when Cth and Ulp1_WT structures are compared (Fig. 6B). The conserved tryptophan and glycine residues (Trp-992, Trp-1071, and Gly-1132) that line the narrow catalytic tunnel occupy similar positions in the Cth structure as they do in several other structures of Ulp1/SENPs complexed with Smt3/SUMO. However, a notable exception occurs with Cth Trp-992, which would clash with the modeled C-terminal tail of Smt3 (Fig. 6B). This clash is most likely because SUMO is absent in our Cth structure, as the analogous Trp residue in SENP-1 (Trp-410) also rotates to accommodate SUMO-2 upon binding (24).

Although the SUMO-interacting surface of Cth SUMO protease resembles those of other Ulp1/SENPs, several substitutions in Cth indicate potential differences in interaction with its cognate SUMO(s), possibly explaining the weaker interaction with S. cerevisiae Smt3 that we observe in our FRET assay (as judged by the increase in $K_m$). In Ulp1_WT, Glu-434 is within hydrogen-bonding distance of Smt3 Gln-73, whereas it is substituted to Pro-976 in the Cth SUMO protease (Fig. 6C). Whereas this residue is not strictly conserved in the Ulp1/SENP
family, Ulp1_WT and SENP1, -2, and -6 have polar residues in this position (Fig. 2A). The probable *C. thermophilum* SUMO homolog (gene ID: CHTT_0059470) as well as human SUMO-2 and SUMO-3 also have a glutamine/asparagine in the corresponding position, although in human SUMO-1, this residue is substituted to alanine (Fig. 2B). The salt bridge between Ulp1_WT Glu-455 and Arg-71 of Smt3 appears absent in *Cth* as Glu-455 is substituted to Ile-999 (Fig. 6C). This residue is conserved as charged or polar side chains in all members of the Ulp1/SENP family (Fig. 2A). Arg-71 of Smt3 is conserved in human SUMO-1 and the *Cth* SUMO homolog but is substituted to a proline in SUMO-2 and SUMO-3 (Fig. 2B). Both Thr-477 and Asn-509 of Ulp1_WT are within hydrogen-bonding distance of Gln-95 of Smt3, with Asn-509 within hydrogen-bonding distance of the backbone of Ile-96 of Smt3 (Fig. 6C). Thr-477 is substituted to Pro-1031, and Asn-509 corresponds to Cys-1066 in *Cth* (Fig. 6C), thus suggesting that hydrogen-bonding interactions observed in Ulp1_WT-Smt3 are not present between *Cth* and its cognate SUMO substrate(s). Pro-1031 is conserved as charged or polar in all Ulp1/SENP with the exception of SENP-2, where it is also a proline (Fig. 2A). *Cth* Cys-1066 is conserved as an asparagine or histidine in all Ulp1/SENP (Fig. 2A). All SUMO variants, including that of *Cth*, have a glutamine in this position (Fig. 2B).

Several other differences between the SUMO-binding interfaces of Ulp1_WT and *Cth* SUMO protease are evident from the structure-based alignment. A universally conserved tryptophan in other Ulp1/SENP is substituted to leucine in *Cth* (Leu-1044). In structures of Ulp1/SENP proteases, this tryptophan packs against a highly conserved glycine in Smt3 and contributes hydrogen bonds to the backbone of Smt3, SUMO-1, and SUMO-2/3 (Figs. 2A and B) and 6C (21–24). Mutation of this tryptophan to alanine in Ulp1_WT results in temperature-sensitive growth defects in *S. cerevisiae* (2). Thus, the presence of leucine in this position in *Cth* suggests the possibility of different packing interactions with cognate *C. thermophilum* SUMO variant(s), although at least one of these variants, CHTT_0059470, retains the conserved glycine (Fig. 2B).

Another Ulp1_WT residue that presumably contributes to Smt3 recognition is Gln-512, which is within hydrogen-bonding distance to Arg-93 and the backbone of Smt3. In *Cth*, this residue is substituted to Ser-1068 and appears to face away from the modeled Smt3 C terminus (Fig. 6C). There is poor conservation of this residue among other SENPs, indicating that it may not be important for Smt3/SUMO recognition (Fig. 2A). Finally, it is worth noting that the *Cth* SUMO protease is more closely related to *S. cerevisiae* Ulp1 (i.e. Ulp1_WT) and human SENP1, -2, -3, and -5 is quite different from SENP6 and SENP7, which have unique structural features and are more adept at deconjugating SUMO chains (Fig. 2A) (25).

**Conclusion**

We have obtained proteases that are more soluble, more thermostable, and more efficient than the current commercially available state-of-the-art Ulp1_WT enzyme both by successful re-engineering of Ulp1_WT and discovery of a new SUMO protease from *C. thermophilum* (Table 2). To accomplish this, we created an automated computational design method that is generally applicable to improving the solubility of proteins. Combining bioinformatic analysis with structure prediction allowed us to successfully identify a functional homologue to Ulp1_WT despite low sequence identity, and this could be a generalizable approach for assigning function to many unannotated genes.

The solubility of each protease we assessed roughly tracks with the amount of hydrophobic surface area that is exposed to solvent (Table 4). However, it is unclear what drives the 5-fold difference in *S*o between Ulp1_R2 and Ulp1_R4, as these proteins exhibit the same amount of hydrophobic surface. The surface area calculations for Ulp1_WT and *Cth* SUMO protease are based on models determined via X-ray crystallography, whereas Ulp1_R1–4 are based on the Rosetta models, which may not represent the solution structure of these proteins as accurately as the experimentally determined models. However, we hypothesize that there are additional factors contributing to the observed solubility differences between these constructs, and the computational design protocol and assays described here will provide an ideal model system for future investigation.

For researchers looking to utilize an improved SUMO protease for preparation of recombinant protein, we recommend Ulp1_R3 or *Cth* SUMO protease. Expression plasmids for the engineered protease constructs reported in this study have been made available via Addgene.
**SUMO protease engineering**

**Experimental procedures**

**Computational protein design**

Generally applicable computational methods to design solubility-enhancing mutations and to calculate the solvent-accessible surface areas (total, hydrophobic, and nonpolar) were developed using the RosettaScripts application programming interface ([26](#)) for Rosetta ([27](#)). The xml protocols needed to run these methods are available in the supporting material.

**Molecular cloning**

*E. coli* codon optimization was performed using DNAworks ([28](#)). The gene for Ulp1_WT was cloned into vector pET28b (Novagen) with an N-terminal hexahistidine tag followed by a thrombin cleavage site. Ulp1_R1–4 were cloned into vector pET29b with a hexahistidine tag at both the N and C termini; these are plasmids pCDB325–328. *Cth* SUMO protease was cloned into vector pCDB98 with a C-terminal decahistidine tag; this plasmid is pCDB302.

**Protein expression and purification**

All Ulp1 variants and *Cth* SUMO protease were expressed and purified as follows. BL21(DE3)RIL cells were grown to optical density 0.6–0.8 at 37 °C and then induced by the addition of 0.5 mM isopropyl β-D-1-thiogalactopyranoside and grown at 18 °C for an additional 16–20 h (1 liter/construct). Cells were harvested by centrifugation and resuspended in 10 ml of buffer A (50 mM Tris, pH 7.5, 500 mM NaCl, 20 mM imidazole, 0.5 mM EDTA, 1 mM DTT) supplemented with 1.5 μl of benzonase (Sigma, E1014-25kU), lysed by sonication at 4 °C, and centrifuged at 39,000 × g at 4 °C for 25 min. The supernatant was applied to 2.5 ml of Ni²⁺-NTA beads prewashed with buffer B (50 mM Tris, pH 7.5, 500 mM NaCl, 20 mM imidazole, 0.25 mM EDTA, 1 mM DTT) incubated at 4 °C for 20 min. The Ni²⁺-NTA beads were washed with 100 ml of buffer B, and sample was eluted with 10 ml of buffer C (50 mM Tris, pH 7.5, 500 mM NaCl, 250 mM imidazole, 0.25 mM EDTA, 0.5 mM DTT). The eluted sample was applied to a Superdex 75 26/60 column equilibrated in buffer D (50 mM Tris, pH 7.5, 200 mM

![Figure 6. Structure of Cth SUMO protease.](#)

A, two orthogonal views of the *Cth* SUMO protease (maroon; PDB code 6DG4) superimposed onto the Ulp1_WT structure (cyan; PDB code 1EUV, chain A) with the SUMO-binding interfaces colored in yellow and cyan, respectively, and depicted in stick representation. The arrow points to the catalytic cysteine in the view on the left. B, superimposed catalytic pockets of Ulp1_WT (cyan) and the apo-*Cth* SUMO protease (maroon/yellow) with several key residues shown in stick representation. Numbering is for the *Cth* protease. C, C-terminal region of Smt3 (purple) and Smt3-binding surface of Ulp1_WT (cyan) superimposed onto apo-*Cth* structure (maroon/yellow). Several side chains that differ between *Cth* and Ulp1_WT in the Ulp1_WT-Smt3 interface are shown in stick representation.
NaCl, 5% (v/v) glycerol, 2 mM DTT, 1 mM EDTA). Fractions containing the highest purity of Ulp1 as judged by SDS-PAGE were concentrated down to 1–1.5 ml using spin columns, and concentration was measured by A\textsubscript{280} using a NanoDrop (Thermo Scientific). Concentrated protein was aliquoted and flash-frozen in liquid N\textsubscript{2} and stored at −80 °C for later use.

EGFP-Smt3\textsuperscript{GGGGG}-mCherry was expressed and purified as follows. BL21(DE3)RIL cells (4 liters) were grown to optical density 0.6–0.8 at 37 °C and then induced by the addition of isopropyl \(\beta\)-D-1-thiogalactopyranoside to 1 mM and grown for 20–24 h. The resulting data were log10-transformed, plotted against the concentration of PEG, and fit to a least-squares regression line. The indicated errors in \(S\alpha\) are the S.E. of the \(y\) intercept as calculated by the S.E. of the regression (\(S\)), where \(S\) is found by dividing the sum of the squares of the deviation from the best-fit line by the number of data points beyond the minimum two required to fit the specified linear curve (29).

### CD and thermal denaturation

CD measurements were performed using a JASCO J-1500 CD spectrophotometer and a quartz cuvette with a 1-mm light path. Protein samples were prepared in 10 mM sodium phosphate, pH 7.4. Steady-state wavelength spectra were recorded in 0.1-nm increments from 200 to 260 nm at 20 and 95 °C. Thermal denaturation was assessed by measuring the CD signal at 208-, 218-, and 222-nm wavelengths in 1 °C increments from 20 to 95 °C.

**FRET assay for EGFP-Smt3\textsuperscript{GGGGG}-mCherry cleavage by Ulp1**

All FRET assays were performed in buffer J (25 mM Tris, pH 7.5, 150 mM NaCl, 0.1% (v/v) Tween 20, 2 mM DTT, 0.5 mM
SUMO protease engineering

EDTA) at 30 °C. Varying concentrations of EGFP-Smt3GGGGG-mCherry were preincubated in 96-well half-area plates (Corning, catalog no. 3686) for 10 min at 30 °C, and SUMO protease was added to initiate the reaction (0.1–2 nm enzyme). The time course of the cleavage reaction was monitored by loss of FRET (excitation, 450 nm; emission, 655 nm) with 15 reads/well and 4–5-s read intervals on a SpectraMax M5 plate reader (Molecular Devices). Rates at each concentration of substrate were calculated by measuring the initial slope of reaction, dividing this value by the amplitude of the signal change, and multiplying by the concentration of substrate. Obtained values were then divided by the concentration of protease enzyme. Rates from three or more replicates for each concentration of substrate were averaged, and the S.D. was calculated to generate the error bars. The resulting data are plotted against the concentration of EGFP-Smt3GGGGG-mCherry and fit to the Michaelis–Menten equation, rate = (V max × [S])/(K m + [S]), using the nonlinear least-squares fit in Prism (GraphPad Software). The indicated errors in V max and K m values are S.E. of the fit as reported by Prism.

Structure determination by X-ray crystallography

His 6-tagged Cth SUMO protease was crystallized via the hanging-drop method by mixing 2 μl of 10 mg/ml protein in buffer D with 2 μl of reservoir solution (0.1 m HEPES, pH 7.0, 2.1 m ammonium sulfate) and incubating over 500 μl of reservoir solution. Crystals were grown at 18 °C, harvested after 2 days, and cryo-protected by dipping into cryo-solution (0.1 m HEPES, pH 7.0, 2.5 m ammonium sulfate, 15% (v/v) glycerol) for 10–15 s before freezing in liquid N2. X-ray diffraction data were collected at the Advanced Photon Source 24-ID-E beam line at a wavelength of 0.97918 Å. Data were processed with HKL2000 (30), and a molecular replacement solution was obtained using Phaser (31) as part of the Phenix suite (32) with PDB entry 1EUV, chain A, as the search model. The molecular replacement solution was manually rebuilt using COOT (7) based on electron density and refined in Phenix.

References

1. Lima, C. D., and Moossessova, E. (March 22, 2011) Rapidly cleavable sumo fusion protein expression system for difficult to express proteins. U. S. Patent US7910364B2
2. Moossessova, E., and Lima, C. D. (2000) Ulp1-SUMO crystal structure and genetic analysis reveal conserved interactions and a regulatory element essential for cell growth in yeast. Mol. Cell. 5, 865–876 CrossRef Medline
3. Butt, T. R., Edavettal, S. C., Hall, J. P., and Mattern, M. R. (2005) SUMO fusion technology for difficult-to-express proteins. Protein Expr. Purif. 43, 1–9 CrossRef Medline
4. Marblestone, J. G., Edavettal, S. C., Lim, Y., Lim, P., Zuo, X., and Butt, T. R. (2006) Comparison of SUMO fusion technology with traditional gene fusion systems: enhanced expression and solubility with SUMO. Protein Sci. 15, 182–189 CrossRef Medline
5. Li, S. J., and Hochstrasser, M. (1999) A new protease required for cell-cycle progression in yeast. Nature 398, 246–251 CrossRef Medline
6. Owerbach, D., McKay, E. M., Yeh, E. T. H., Gabbay, K. H., and Bohren, K. M. (2005) A proline-90 residue unique to SUMO-4 prevents maturation and sumoylation. Biochem. Biophys. Res. Commun. 337, 517–520 CrossRef Medline
7. Emsley, P., Lohkamp, B., Scott, W. G., and Cowtan, K. (2010) Features and development of Coot. Acta Crystallogr. D Biol. Crystallogr. 66, 486–501 CrossRef Medline
8. Jacak, R., Leaver-Fay, A., and Kuhlman, B. (2012) Computational protein design with explicit consideration of surface hydrophobic patches. Proteins 80, 825–838 CrossRef Medline
9. Raghunathan, G., Sokalingam, S., Soundrarajan, N., Madan, B., Munussami, G., and Lee, S.-G. (2013) Modulation of protein stability and aggregation properties by surface charge engineering. Mol. Biosyst. 9, 2379–2389 CrossRef Medline
10. Bhardwaj, G., Mulligan, V. K., Bahl, C. D., Gilmore, J. M., Harvey, P. J., Chenewal, O., Bachko, G. W., Pulavarti, S. V., S. R. K., Kaas, Q., Eleksty, A., Huang, P.-S., Johnsen, W. A., Greisen, P. I., Rocklin, G. J., Song, Y., et al. (2016) Accurate de novo design of hyperstable constrained peptides. Nature 538, 329–335 CrossRef Medline
11. Amlacher, S., Sarges, P., Fleming, D., van Noort, V., Kunze, R., Devos, D. P., Arumugam, M., Bork, P., and Hurt, E. (2011) Insight into structure and assembly of the nuclear pore complex by utilizing the genome of an eukaryotic thermophile. Cell 146, 277–289 CrossRef Medline
12. Kellner, N., Schwarz, J., Sturm, M., Fernandez-Martinez, J., Griesel, S., Zhang, W., Chait, B. T., Rout, M. P., Kück, U., and Hurt, E. (2016) Developing genetic tools to exploit Chaetomium thermophilum for biochemical analyses of eukaryotic macromolecular assemblies. Sci. Rep. 6, 20937 CrossRef Medline
13. Bock, T., Chen, W.-H., Ari, O., Malik, N., Silva-Martín, N., Huerta-Cepas, J., Powell, S. T., Kastritis, P. L., Smyshlyaev, G., Vonkova, I., Kirkpatrick, J., Doerks, T., Nesme, L., Bassler, J., Kos, M., et al. (2014) An integrated approach for genome annotation of the eukaryotic thermophile Chaetomium thermophilum. Nucleic Acids Res. 42, 13525–13533 CrossRef Medline
14. Boratyń, G. M., Schäffer, A. A., Agarwala, R., Altschul, S. F., Lipman, D. J., and Madden, T. L. (2012) Domain enhanced lookup time accelerated BLAST. Biol. Direct. 7, 12 CrossRef Medline
15. Kim, D. E., Chivian, D., and Baker, D. (2004) Protein structure prediction and analysis using the Robetta server. Nucleic Acids Res. 32, W526–W531 CrossRef Medline
16. Middag, C. R., Tisel, W. A., Haire, R. N., and Rosenberg, A. (1979) Determination of the apparent thermodynamic activities of saturated protein solutions. J. Biol. Chem. 254, 367–370 Medline
17. Kramer, R. M., Shende, V. R., Motl, N., Pace, C. N., and Scholtz, J. M. (2012) Toward a molecular understanding of protein solubility: increased negative surface charge correlates with increased solubility. Biophys. J. 102, 1907–1915 CrossRef Medline
18. Li, L., Kantor, A., and Warne, N. (2013) Application of a PEG precipitation method for solubility screening: a tool for developing high protein concentration formulations. Protein Sci. 22, 1118–1123 CrossRef Medline
19. Fraser, J. S., Clarkson, M. W., Degnan, S. C., Erion, R., Kern, D., and Alber, T. (2009) Hidden alternative structures of proline isomerase essential for catalysis. Nature 462, 669–673 CrossRef Medline
20. Clegg, R. M. (1995) Fluorescence resonance energy transfer. Curr. Opin. Biotechnol. 6, 103–110 CrossRef Medline
21. Shen, L., Tatham, M. H., Dong, C., Zagórska, A., Naismith, J. H., and Hay, R. T. (2006) SUMO protease SENP1 induces isomerization of the scissile peptide bond. Nat. Struct. Mol. Biol. 13, 1069–1077 CrossRef Medline
22. Reverter, D., and Lima, C. D. (2004) A basis for SUMO protease specificity provided by analysis of human Senp2 and a Senp2-SUMO complex. Structure 12, 1519–1531 CrossRef Medline
23. Reverter, D., and Lima, C. D. (2006) Structural basis for SENP2 protease interactions with SUMO precursors and conjugated substrates. Nat. Struct. Mol. Biol. 13, 1060–1068 CrossRef Medline
24. Shen, L. N., Dong, C., Liu, H., Naismith, J. H., and Hay, R. T. (2006) The structure of SENP1-SUMO-2 complex suggests a structural basis for discrimination between SUMO paralogues during processing. Biochem. J. 397, 279–288 CrossRef Medline
25. Lima, C. D., and Reverter, D. (2008) Structure of the human SENP7 catalytic domain and poly-SUMO deconjugation activities for SENP6 and SENP7. J. Biol. Chem. 283, 32045–32055 CrossRef Medline
26. Fleishman, S. J., Leaver-Fay, A., Corn, J. E., Strauch, E.-M., Khare, S. D., Koga, N., Ashworth, J., Murphy, P., Richter, F., Lemmon, G., Meier, J., and Baker, D. (2011) RosettaScripts: a scripting language interface to the Rosetta macromolecular modeling suite. PLoS One 6, e20161 CrossRef Medline
27. Kaufmann, K. W., Lemmon, G. H., Deluca, S. L., Sheehan, J. H., and Meier, J. (2010) Practically useful: what the Rosetta protein modeling suite can do for you. Biochemistry 49, 2987–2998 CrossRef Medline
28. Hoover, D. M., and Lubkowski, J. (2002) DNAWorks: an automated method for designing oligonucleotides for PCR-based gene synthesis. Nucleic Acids Res. 30, e43 CrossRef Medline
29. Devore, J. L. (1987) Probability and Statistics for Engineering and the Sciences, 2nd Ed., Brooks/Cole Publishing Co., Boston
30. Otwinowski, Z., and Minor, W. (1997) Processing of X-ray diffraction data collected in oscillation mode. Methods Enzymol. 276, 307–326 CrossRef Medline
31. McCoy, A. J., Grosse-Kunstleve, R. W., Adams, P. D., Winn, M. D., Storoni, L. C., and Read, R. J. (2007) Phaser crystallographic software. J. Appl. Crystallogr. 40, 658–674 CrossRef Medline
32. Adams, P. D., Afonine, P. V., Bunkóczi, G., Chen, V. B., Davis, I. W., Echols, N., Headd, J. J., Hung, L.-W., Kapral, G. J., Grosse-Kunstleve, R. W., McCoy, A. J., Moriarty, N. W., Oeffner, R., Read, R. J., Richardson, D. C., et al. (2010) PHENIX: a comprehensive Python-based system for macromolecular structure solution. Acta Crystallogr. D Biol. Crystallogr. 66, 213–221 CrossRef Medline