A novel approach for production of an active N-terminally truncated Ulp1 (SUMO protease 1) catalytic domain from Escherichia coli inclusion bodies

Linova, Marina Y.; Risør, Michael W.; Jørgensen, Sanne E.; Mansour, Zohra; Kaya, Jacob; Sigurdarson, Jens J.; Enghild, Jan J.; Karring, Henrik

Published in:
Protein Expression and Purification

DOI:
10.1016/j.pep.2019.105507

Publication date:
2020

Document version:
Accepted manuscript

Document license:
CC BY-NC-ND

Citation for published version (APA):

Go to publication entry in University of Southern Denmark's Research Portal

Terms of use
This work is brought to you by the University of Southern Denmark. Unless otherwise specified it has been shared according to the terms for self-archiving. If no other license is stated, these terms apply:

• You may download this work for personal use only.
• You may not further distribute the material or use it for any profit-making activity or commercial gain.
• You may freely distribute the URL identifying this open access version.

If you believe that this document breaches copyright please contact us providing details and we will investigate your claim. Please direct all enquiries to puresupport@bib.sdu.dk
A novel approach for production of an active N-terminally truncated Ulp1 (SUMO protease 1) catalytic domain from *Escherichia coli* inclusion bodies

Marina Y. Linova a, Michael W. Risør b, Sanne E. Jørgensen a, Zohra Mansour a, Jacob Kaya a, Jens J. Sigurdarson a, Jan J. Enghild b, Henrik Karring a*

a Department of Chemical Engineering, Biotechnology and Environmental Technology, University of Southern Denmark, Campusvej 55, DK-5230 Odense M, Denmark

b Department of Molecular Biology and Genetics, Aarhus University, Gustav Wieds Vej 10C, DK-8000 Aarhus C, Denmark

*Corresponding author. Address: Department of Chemical Engineering, Biotechnology and Environmental Technology, University of Southern Denmark, Campusvej 55, DK-5230 Odense M, Denmark

E-mail address: hka@kbm.sdu.dk

Tel: +45 21356350

Current address:

Zohra Mansour; Novo Nordisk Pharmatech A/S, Københavnsvej 216, 4600 Køge, Denmark

Jacob Kaya; Chr. Hansen A/S, Bege Alle 10, 2970 Hørsholm, Denmark

Michael W. Risør; BASF enzymes LLC, 3550 John Hopkins Court 92121 San Diego

Declaration of interest: none
Abstract

The SUMO fusion system is widely used to facilitate recombinant expression and production of difficult-to-express proteins. After purification of the recombinant fusion protein, removal of the SUMO-tag is accomplished by the yeast cysteine protease, SUMO protease 1 (Ulp1), which specifically recognizes the tertiary fold of the SUMO domain. At present, the expression of the catalytic domain, residues 403-621, is used for obtaining soluble and biologically active Ulp1. However, we have observed that the soluble and catalytically active Ulp1\textsubscript{403-621} inhibits the growth of \textit{E. coli} host cells. In the current study, we demonstrate an alternative route for producing active Ulp1 catalytic domain from a His-tagged N-terminally truncated variant, residues 416-621, which is expressed in \textit{E. coli} inclusion bodies and subsequently refolded. Expressing the insoluble Ulp1\textsubscript{416-621} variant is advantageous for achieving higher production yields. Approximately 285 mg of recombinant Ulp1\textsubscript{416-621} was recovered from inclusion bodies isolated from one liter of high cell-density \textit{E. coli} batch fermentation culture. After Ni\textsuperscript{2+}-affinity purification of inactive and denatured Ulp1\textsubscript{416-621} in 7.5 M urea, different refolding conditions with varying L-arginine concentration, pH, and temperature were tested. We have successfully refolded the enzyme in 0.25 M L-arginine and 0.5 M Tris-HCl (pH 7) at room temperature. Approximately 80 mg of active Ulp1\textsubscript{416-621} catalytic domain can be produced from one liter of high cell-density \textit{E. coli} culture. We discuss the applicability of inclusion body-directed expression and considerations for obtaining high expression yields and efficient refolding conditions to reconstitute the active protein fold.

Keywords: ubiquitin-like protease (Ulp); small ubiquitin-like modifier (SUMO); SUMO-specific protease; inclusion bodies; refolding; L-arginine

1. Introduction

Recombinant protein production in bacteria presents several challenges, such as low expression levels, protein misfolding, and partial degradation [1-3]. Among other advancements, fusion protein systems have shown to be efficient in solving these issues [4, 5]. Several fusion systems have been developed, of which small ubiquitin-like modifier (SUMO) fusion system has proven to be highly effective [1]. SUMO-like proteins are present in all eukaryotic cells and are highly conserved in the evolution from yeast to humans. It is the second best characterized ubiquitin-like (UBL) family after ubiquitin. \textit{Saccharomyces cerevisiae} comprises a single SUMO homologue gene \textit{SMT3}, whereas there are three homologous members in the mammalian SUMO family (\textit{SUMO-1, SUMO-2, SUMO-3}).
Proteins from the SUMO family attach covalently to other proteins via an isopeptide bond formed between the C-terminal carboxyl group in SUMO and the ε-amino group of an lysine side chain in the conjugated proteins [3]. The de-conjugation (desumoylation) of the SUMO moiety from the modified protein is catalyzed by cysteine proteases, which recognize the SUMO-fold and cleave the isopeptide bond. These enzymes, known as SUMO-specific proteases, have been identified in different eukaryotic species including human and yeast [6]. In *S. cerevisiae*, two SUMO-specific proteases have been characterized and referred to as ubiquitin-like specific proteases (Ulp1 and Ulp2) [7, 8].

In several studies, the yeast SUMO (Smt3) and SUMO-specific protease 1 (also called SUMO protease 1 and Ulp1) have successfully been used in a SUMO fusion system for difficult-to-express proteins [2, 3, 9]. When fused to the SUMO solubility tag otherwise insoluble proteins are expressed as soluble and properly folded proteins. Ulp1 overcomes a major disadvantage of protein fusion technologies, by cleaving the SUMO-tag with high specificity and over a wide range of conditions [3]. In comparison to site-specific proteases such as thrombin and TEV protease [1] which are used in other protein fusion systems to remove affinity and solubility tags, Ulp1 recognizes the unique tertiary structure of the SUMO-tag. It cleaves specifically at the C-terminus of the glycine-glycine motif in the C-terminal end of the SUMO tag, allowing for any desired N-terminus on the C-terminal target protein [3, 10]. This property of the SUMO-tag/Ulp1 system is advantageous because the cleavage leaves no additional residues on the target protein's N-terminus.

Full-length Ulp1 is composed of a weakly conserved N-terminal domain (residues 1-432), and a conserved C-terminal protease fold (residues 432-621; numbering *S. cerevisiae* Ulp1) [7]. The active site of Ulp1 is similar to other cysteine proteases, comprising the catalytic triad (Cys-580, His-514, Asp-531). However, unlike other cysteine proteases, Ulp1 catalytic domain has a narrow hydrophobic tunnel, recognizing the di-glycine C-terminal motif, thus responsible for the high selectivity of the enzyme [10].

Li and Hochstrasser discovered through limited proteolysis of yeast Ulp1, that the fragment comprised of residues 403-621 is as active as the full-length protease [3, 7, 10]. Thus, recombinant Ulp1403-621 has already been used for digesting SUMO-tagged proteins [2, 3, 9, 11]. It is known that the enzyme kinetics and the optimal enzyme to substrate ratio vary and are very dependent on the SUMO-fused protein [3]. These findings suggest that in some cases, in order to achieve good digestion within a reasonable period of time, higher amount of the Ulp1 catalytic domain is needed.
In protein research such as drug discovery, protein crystallization, and NMR structure determination requiring significant amounts of pure and native protein, larger quantities of SUMO protease 1 may be needed.

Due to the commonly employed use of SUMO protease 1 together with the yeast SUMO (Smt3), ways of expression and purification of the soluble protease, as well as strategies for improving its stability and solubility have been investigated [2, 9, 12]. Here, an alternative approach towards the production of active Ulp1 have been applied. We have identified an N-terminally truncated version of the catalytic domain of *S. cerevisiae* Ulp1 (Ulp1_{416-621}) which expresses in an insoluble state in *E. coli*. In the present study, we show that the expression of the truncated SUMO protease 1 domain, lacking residues 403-415 (LVPELNEKDDDQV), is considerably higher than that of the intact and soluble catalytic Ulp1 domain (Ulp1_{403-621}). We have observed that the expression of the soluble Ulp1_{403-621} variant inhibits the growth of *E. coli* after induction, which limits the recombinant protein yield. In contrast, expression of the insoluble Ulp1_{416-621} does not affect the growth of *E. Coli* and offers a high yield alternative strategy to Ulp1 expression. We present a simple refolding procedure, which eliminates the bottleneck for the production of biologically active Ulp1_{416-621} from *E. coli* inclusion bodies. Based on the performed optimization, we conclude that L-arginine mediated refolding of Ulp1_{416-621} provides a novel and efficient route for producing soluble and proteolytically active Ulp1 catalytic domain.

2. Materials and methods

2.1 Materials

HisTrap HP columns were purchased from GE Life Science. Protein marker (14-116) was from AppliChem (A5238). The plasmid pFGET19_Ulp1 was a gift from Hideo Iwai (Addgene plasmid # 64697) [9]. All other chemicals were purchased from Sigma Aldrich.

2.2 Cloning of Ulp1_{416-621} and transformation of expression strains

Codon-optimized cDNA of Ulp1_{416-621} was synthesized at Genscript® and sub-cloned into the pET28a(+) vector using NdeI and XhoI restriction sites, providing an in-frame fusion to the N-terminal and C-terminal His-tags of the pET28a(+) vector. The construct included an 11-residue N-terminal linker (SSGLVPRGSHM) and a Glu residue linker in its C-terminus. The resulting plasmid pET28a_Ulp1(416-621) was transformed into *E. coli* BL21(DE3) for expression of H6-Ulp1_{416-621}
H6. The plasmid pFGET19_Ulp1 (pET28b(+) encoding for N-terminal His-tagged Ulp1\textsubscript{403-621} with an N-terminal GLVPRGS linker), was also transformed into \textit{E. coli} BL21(DE3). Both plasmids code for kanamycin resistance and expression of genes inserted in the multiple cloning site is under the control of a T7 promotor.

2.3 Comparison of expression levels of H6-Ulp1\textsubscript{416-621}-H6 and H6-Ulp1\textsubscript{403-416} in \textit{E. coli} BL21(DE3)

Vectors pET28a\textsubscript{Ulp1}(416-621) and pFGET19\textsubscript{Ulp1}, encoding H6-Ulp1\textsubscript{416-621}-H6 and H6-Ulp1\textsubscript{403-621} respectively, were transformed into \textit{E. coli} strain BL21(DE3). Protein expression of the Ulp1 catalytic domain from the two constructs was carried out in triplicate at 37 °C and 200 rpm shaking. Each flask containing 200 ml Lysogeny Broth (LB) medium \cite{11}, supplemented with 34 µg/ml kanamycin, was inoculated with 10 ml of fresh overnight culture. Protein expression was induced at OD\textsubscript{600}~0.6 by adding isopropyl β-D-1-thiogalactopyranoside (IPTG) to a final concentration of 1 mM. Expression was carried out for 3 hours during which the OD\textsubscript{600} was monitored every 30 minutes. Samples for sodium dodecyl sulfate-polyacrylamide gel electrophoresis (SDS-PAGE) analysis were withdrawn from the expression culture before induction and every 60 minutes during the expression.

2.4 Optimisation of H6-Ulp1\textsubscript{416-621}-H6 expression

Expression of H6-Ulp1\textsubscript{416-621}-H6 in \textit{E. coli} strain BL21(DE3) was tested in Terrific Broth (TB) medium \cite{13} and M9 minimal medium \cite{14} supplemented with 34 µg/ml kanamycin, and at three different temperatures (23 °C, 30 °C and 37 °C). In both media, 0.4% (V/V) glycerol was used as a carbon source. Each expression culture was initiated by addition of 0.5 ml of fresh pre-culture to 50 mL of the corresponding medium and growth conditions. Due to a very low cell-density, cultures incubated at 23 °C were inoculated with a 37 °C grown pre-culture. Expression was induced at OD\textsubscript{600}~0.5-0.6 and carried out as described in the previous section.

2.5 Determination of total protein expression and solubility levels

To determine the total amount of expressed Ulp1\textsubscript{403-621} and Ulp1\textsubscript{416-621}, and the fraction of soluble and insoluble proteins, samples were withdrawn from the expression cultures and centrifuged at 16,126 x g for 2 minutes at room temperature before the medium was carefully discarded. The resulting cell pellet was prepared for analyses of total protein (TP), soluble protein (SP), or insoluble protein (IP). For TP analyses the cell pellets were stored at -20 °C for later use, while for SP and IP analyses they were suspended in a lysis buffer (50 mM Tris-HCl, pH 8, 50 mM NaCl, 5 mM EDTA, 1 mg/ml lysozyme, 4 µg/ml DNase I) and sonicated five times for 10 seconds with 5 minutes intervals using...
a Branson Digital Sonifier. Samples were kept on ice between the sonications. Subsequently, the
lysed cells were centrifuged at 16,126 x g for 10 minutes at RT (23 ºC). For SP analyses some of the
soluble lysate was transferred to another tube and stored at -20 ºC. For IP analyses the cell debris
after centrifugation was washed again with lysis buffer and centrifuged for 10 minutes. After the
second centrifugation, the supernatant was removed and the resulting pellet constituted the IP sample.
All resulting samples (TP, SP, IP) were analyzed by SDS-PAGE.

2.6 High-level expression of His-Ulp1416-621-His in E. coli BL21(DE3) by high cell-density batch
fermentation

E. coli BL21(DE3) transformed with pET28a_Ulp1(416-621) was grown overnight in TB medium
with kanamycin (34 µg/ml) at 37 ºC in a shaking incubator at 150 rpm. Subsequently, 100 ml of
overnight culture was added to a 5 L bioreactor (Applikon Biotechnology) containing 1.9 L aerated
TB medium with kanamycin (34 µg/ml), supplemented with 300 µl antifoam. The airflow was set to
4 L/min and the level of dissolved oxygen was kept above 10% by regulation of agitation (500-1500
rpm). The temperature was kept at 37 ºC and the pH was maintained at 7 by addition of 1 M NaOH
or 1 M HCl. Protein expression was induced by the addition of 1 mM IPTG, when OD$_{600}$ reached
approximately 1. OD$_{600}$ was measured every 30 minutes and samples for SDS-PAGE analysis were
taken immediately before induction and every hour after induction. Approximately 4.5 hours after
induction the cells were harvested by centrifugation at 7,000 x g for 15 minutes at 4 ºC. Finally, the
cells were washed in 20 mM Tris-HCl, pH 7.5 before centrifugation at 7,000 x g for 15 minutes at 4
ºC, and stored at -20 ºC for later use.

2.7 Cell lysis and isolation of inclusion bodies

The harvested cells were suspended in lysis buffer (50 mM Tris-HCl, pH 8, 50 mM NaCl, 5 mM
EDTA, 1 mg/ml lysozyme, 4 µg/ml DNase I) and sonicated twice (Branson Digital Sonifier) for 20
seconds at 15 % amplitude with 15 minutes in between while being kept on ice. After centrifugation,
the cell debris was washed twice with a washing buffer (50 mM Tris-HCl, 5 mM EDTA, 5% Triton
X-100, pH 8). At each washing step the inclusion bodies were kept for 10 minutes at room
temperature, followed by 15 minutes of centrifugation at 16,162 x g. Finally, the supernatant was
discarded and the pellets were stored at -20 ºC for later use.

2.8 Purification of His-tagged Ulp1416-621 from inclusion bodies
His-tagged Ulp1\textsubscript{416-621} from inclusion bodies, isolated from 50 ml of cell culture, was purified under denaturing conditions using an ÄKTA start purification system. Buffers were composed of 7.5 M urea, 50 mM NaH\textsubscript{2}PO\textsubscript{4}/Na\textsubscript{2}HPO\textsubscript{4}, 500 mM NaCl and 5 mM 2-mercaptoethanol and varying in pH (buffer A, pH 7.4; buffer B, pH 6.3; buffer C, pH 3.5). The inclusion bodies were solubilized in buffer A, centrifuged for 15 minutes at 20,000 rpm, and filtered through a 0.2 µm filter. The resulting 4.5 ml filtrate was loaded on a 5 ml HisTrap HP column equilibrated with buffer A. The column was washed with 50 ml buffer B and the bound protein was eluted using a 100 ml pH gradient from pH 6.3 (buffer B) to pH 3.5 (buffer C). Selected 1 ml fractions were analyzed by SDS-PAGE. Fractions containing the recombinant Ulp1 catalytic domain were pooled. Protein concentrations of the pooled fractions was determined by Bradford protein assay and the purity was assessed by SDS-PAGE.

2.9 Expression and purification of SUMO(Smt3)-FAS1-4 fusion protein

N-terminally His-tagged SUMO-FAS1-4 (WT) fusion protein was expressed in \textit{E. coli} BL21(DE3) using the Champion\textsuperscript{TM} pET SUMO Expression System (Invitrogen). Cells were grown at 37 °C in LB medium containing 34 µg/ml kanamycin while shaking at 160 rpm. When OD\textsubscript{600} reached approximately 0.7, the expression of the fusion protein was induced by the addition of IPTG to a final 1 mM concentration. After 2 hours of induction the cells were harvested by centrifugation at 6,000 x g for 15 minutes and stored at -20 °C prior to cell lysis. The harvested cells were lysed and the SUMO-FAS1-4 (WT) fusion protein was purified as previously described [15] and dialyzed against a buffer comprising 50 mM Tris-HCl, pH 7.6 and 95 mM NaCl. When stored for extended periods, the SUMO-FAS1-4 protein was flash-frozen with the addition of 10 % glycerol and stored at -80 °C.

2.10 Refolding of purified Ulp1\textsubscript{416-621}

Refolding of purified Ulp1\textsubscript{416-621} into its active form was tested at 18 selected refolding conditions all containing 0.1 mM EDTA and 0.5 mM dithiothreitol (DTT) but differing in pH (6, 7, or 8), temperature (RT or 4 °C), and concentration of L-arginine (0 M, 0.25 M, or 0.50 M).
Table 1. Refolding conditions tested to form active Ulp1416-621 catalytic domain. Condition sets A (0 M L-arginine), B (0.25 M L-arginine) and C (0.5 M L-arginine) show the different combinations of pH and temperature used in the experiment.

<table>
<thead>
<tr>
<th>Condition Set</th>
<th>pH 6</th>
<th>pH 7</th>
<th>pH 8</th>
</tr>
</thead>
<tbody>
<tr>
<td>A</td>
<td>0 M L-arginine</td>
<td>0 M L-arginine</td>
<td>0 M L-arginine</td>
</tr>
<tr>
<td></td>
<td>RT</td>
<td>4°C</td>
<td>RT</td>
</tr>
<tr>
<td>B</td>
<td>0.25 M L-arginine</td>
<td>0.25 M L-arginine</td>
<td>0.25 M L-arginine</td>
</tr>
<tr>
<td></td>
<td>RT</td>
<td>4°C</td>
<td>RT</td>
</tr>
<tr>
<td>C</td>
<td>0.5 M L-arginine</td>
<td>0.5 M L-arginine</td>
<td>0.5 M L-arginine</td>
</tr>
<tr>
<td></td>
<td>RT</td>
<td>4°C</td>
<td>RT</td>
</tr>
</tbody>
</table>

Refolding buffers with no L-arginine (Table 1, set A) were prepared using 50 mM Na₂HPO₄/NaH₂PO₄ while refolding buffers with 0.25 M and 0.5 M L-arginine (Table 1, sets B and C) contained 0.5 M Tris-HCl. All combinations of L-arginine concentration, pH, and temperature were tested in triplicates. The refolding experiments were performed by ten-fold rapid dilutions of purified Ulp1416-621. For this purpose, 10 µl of purified Ulp1416-621 was mixed with 90 µl of refolding buffer and immediately vortexed for 2-3 seconds. For refolding performed at 4 °C, the buffers were cooled to this temperature prior to the fast dilutions. As purified Ulp1416-621 was kept in a buffer having a concentration of 7.5 M urea and 500 mM NaCl, the final concentration of these components in the 100 µl refolding mixture was 0.75 M and 50 mM, respectively. The refolding mixtures were incubated for 24 hours before they were centrifuged for 15 min at 16,162 x g and samples for SDS-PAGE were taken from the supernatant. The amount of refolded (soluble) Ulp1416-621 obtained for each condition was quantified by SDS-PAGE densitometry and compared to that of a control sample containing affinity-purified Ulp1416-621 diluted ten-fold in denaturing buffer (7.5 M urea, 50 mM NaH₂PO₄/Na₂HPO₄, pH 7.4) (Equation 1).

\[
\text{Refolded Ulp1416-621} = \frac{\text{Int(Refolded Ulp1416-621)}}{\text{Int(Denatured Ulp1416-621)}} \times 100, \% \quad (1)
\]
The proteolytic activity of the refolded Ulp1{416-621} was evaluated by testing its ability to cleave off the SUMO-tag from the purified SUMO-FAS1-4 (WT) fusion protein. Unless otherwise specified, digestions were performed in molar ratio 1:110 (Ulp1{416-621}:SUMO-FAS1-4) by adding a small fraction of the refolded Ulp1 into the SUMO-FAS1-4 preparation before it was incubated for 24 hours at room temperature. Samples for SDS-PAGE were taken at the end of the digestion and the intensity of the protein bands was quantified by densitometry. For each activity assay, a sample containing intact SUMO-FAS1-4 fusion protein was used as a control. The activity of refolded Ulp1{416-621} for each condition was determined by calculating the amount of “Cleaved substrate” using Equation 2.

\[
\text{Cleaved substrate} = \frac{\text{Int}(\text{FAS1-4}) + \text{Int}(\text{SUMO-tag})}{\text{Int}(\text{Uncleaved substrate}) + \text{Int}(\text{FAS1-4}) + \text{Int}(\text{SUMO-tag})} \times 100, \%
\]  

(2)

In order to better estimate the ability of the different conditions to refold active Ulp1{416-621}, the obtained values for “Cleaved substrate” were normalized with the corresponding relative amounts of refolded (soluble) Ulp1{416-621}, as noted in Equation 3.

\[
\text{Ulp1 activity} = \frac{\text{Cleaved substrate}, \%}{\text{Refolded Ulp1}_{416-621}, \%}
\]  

(3)

Comparison of the activity of Ulp1{416-621} refolded under the two most promising conditions was carried out by estimating the percentage of cleaved SUMO-fused protein after 30 minutes of incubation. The proteolytic activity of Ulp1{416-621}, refolded under the best performing condition, was also analyzed at 1h, 8h and 24h. The activity of Ulp1{416-621} was also examined at 2mM DTT and enzyme:substrate ratios ranging from 1:110 to 1:2000. The protease activity was also tested in an upscaled 5 ml cleavage reaction, using 2 mM DTT and 1:2000 enzyme:substrate ratio.

### 2.1.1 SDS-PAGE analyses

Protein samples were added SDS-PAGE sample buffer containing a final concentration of 5 mM DTT and SDS-PAGE was performed using hand-cast 8-20% gradient gels and the SDS-PAGE Buffer System [16]. The separated proteins were visualized by Coomassie Brilliant Blue staining using a Blue Silver staining protocol [17]. Stained gels were documented using a ChemiDoc XRS+ system (Bio-Rad) and quantification of protein bands were achieved by densitometry using the Image Lab™ software from Bio-Rad. As an alternative Adobe Photoshop CS5.1 software (version 12.1x 32) was
used for quantification of protein bands for the experimental work described in section 2.4 and 2.6 [18].

3. Results

3.1 Comparison of expression levels of H$_6$-Ulp1$_{416-621}$-H$_6$ and H$_6$-Ulp1$_{403-621}$ in E. coli BL21(DE3)

To evaluate the production of N-terminally truncated SUMO protease 1 catalytic domain in E. coli, the expression level of His-tagged Ulp1$_{416-621}$ from plasmid pET28a_Ulp1(416-621) was compared with that of His-tagged Ulp1$_{403-621}$ from pFGET19_Ulp1 (Figure 1A). The growth of the two different transformed E. coli BL21(DE3) cultures were very similar before the recombinant protein expression was induced by addition of IPTG at approximately 1.5 hours after inoculation. E. coli harboring pFGET19_Ulp1 reached a stationary phase of OD$_{600}$~4 after 1 hour of induction while the growth of E. coli transformed with pET28a_Ulp1(416-621) continued the next 3.5 hours reaching an OD$_{600}$~2.5 (Figure 1A). Analyses of the expression levels of total protein (TP) reveal the induction of an approximately 27 kDa protein in both cultures. The theoretical molecular weights of H$_6$-Ulp1$_{416-621}$-H$_6$ and H$_6$-Ulp1$_{403-621}$ are 27.2 kDa and 27.4 kDa, respectively (Figure 1B and C). The target protein yield per total protein was higher for the Ulp1$_{416-621}$ variant than for the intact SUMO protease catalytic domain Ulp1$_{403-621}$ (Figure 1A). Thus, the yield per total protein reached 31.9 ± 1.39 % and 10.8 ± 2.44 % at 3 hours post-induction for Ulp1$_{416-621}$ and Ulp1$_{403-621}$, respectively (Figure 1A). SDS-PAGE analyses of the soluble protein (SP) and insoluble protein (IP) at 3 hours after induction confirmed that Ulp1$_{403-621}$ is in the soluble fraction with only a small fraction observed in the insoluble fraction. In contrast, the truncated SUMO protease domain Ulp1$_{416-621}$ is only expressed as insoluble protein (Figure 1B and C). The superior yield observed from the truncated SUMO protease domain (Ulp1$_{416-621}$) expression along with the observed growth inhibition for Ulp1$_{403-621}$ prompted us to further optimize the conditions for producing Ulp1$_{416-621}$. 
Figure 1. Expression of the soluble Ulp1\textsubscript{403-621} variant inhibits \textit{E. coli} growth. A) Growth curves of \textit{E. coli} BL21(DE3) transformed with pFGET19\_Ulp1 (encoding His-tagged Ulp1\textsubscript{403-621}) and \textit{E. coli} BL21(DE3) transformed with pET28\_Ulp1(416-621) (encoding His-tagged Ulp1\textsubscript{416-621}). Columns show the level of Ulp1 catalytic domain per total protein. The obtained results for OD\textsubscript{600} and Ulp1 catalytic domain expression levels are mean values from triplicate measurements. Error bars represent standard deviations. B) SDS-PAGE analysis of proteins from \textit{E. coli} BL21(DE3) cells expressing the Ulp1\textsubscript{403-621} (left gel) and Ulp1\textsubscript{416-621} (right gel). The expected molecular weight of the two variants is 27.4 kDa and 27.2 kDa, respectively. Samples for total protein (TP) quantification were withdrawn at 1, 2, and 3 hours after induction, whereas those for soluble (SP) and insoluble protein (IP) analyses were collected at 3 hours after induction. Bands of the Ulp1 catalytic domain variants are indicated with arrows. Ulp1\textsubscript{403-621} is found both in the soluble and insoluble fractions, while Ulp1\textsubscript{416-621} is only found in the insoluble fraction.
3.2 High-level heterologous expression and purification of H₆-Ulp1₄₁₆-₆₂₁-H₆ from inclusion bodies in E. coli.

We determined the optimal expression conditions of Ulp1₄₁₆-₆₂₁ to be in TB medium at 37 °C after small-scale scouting experiments in both TB and M9 medium at 23, 30, and 37 °C (data not shown). The process was transferred to high cell-density batch fermentation to produce considerable amounts of His-tagged Ulp1₄₁₆-₆₂₁ and to determine process yield and purity. The induction of Ulp1₄₁₆-₆₂₁ expression was successful and the cell culture reached an OD₆₀₀ of 12.2 at 3.5 hours post IPTG induction (Figure 2). The weight of the harvested wet cell pellet at the end of the fermentation was 44.83 g from a 2 L batch fermentation and total protein and Ulp1₄₁₆-₆₂₁ amounts were tracked through inclusion body preparation, Ni²⁺-affinity purification, and refolding (Table 2). Each step improved the purity of Ulp1₄₁₆-₆₂₁ from an initial 21.3% to a final 73.3%, thereby obtaining 6.9 mg Ulp1₄₁₆-₆₂₁ from the total 14.3 mg Ulp1₄₁₆-₆₂₁ present in the denatured inclusion bodies isolated from 50 mL culture. Based on this, the yield from 1 L of high cell-density E. coli culture was estimated (Table 2). The purity of Ulp1₄₁₆-₆₂₁ after each step in the recovery procedure was calculated based on SDS-PAGE densitometry. Despite the relative low purity estimate (Table 2) the SDS-PAGE showed no major contaminants in the preparation of refolded Ulp1₄₁₆-₆₂₁ (Figure 3A). Ulp1₄₁₆-₆₂₁ can be stored in refolding buffer at -20 °C for several months or as inclusion bodies pellet for extended periods.

![Figure 2. High-level expression of recombinant Ulp1₄₁₆-₆₂₁ in E. coli.](image.png)

Growth conditions and cell density were monitored during the 2 L batch fermentation. The cells were grown in TB media at 37 °C and pH 7. The level of dissolved oxygen (DO, %) was kept above 10% during most of the growth. OD₆₀₀ was measured every 30 minutes. The addition of IPTG and the time of cell harvest are marked with vertical lines and arrows.
Table 2. Recovery of Ulp1416-621 from 1 L of high cell-density E. coli batch fermentation at the different steps of the developed procedure.

<table>
<thead>
<tr>
<th>Step</th>
<th>Total protein (mg) a</th>
<th>Purity (%) b</th>
<th>Step yield (mg Ulp1416-621)</th>
<th>Overall yield (%)</th>
<th>Relative purification</th>
</tr>
</thead>
<tbody>
<tr>
<td>Bacterial lysate</td>
<td>3300</td>
<td>21.3</td>
<td>702.9</td>
<td>100</td>
<td>1</td>
</tr>
<tr>
<td>Inclusion bodies</td>
<td>577</td>
<td>49.7</td>
<td>286.8</td>
<td>40.8</td>
<td>2.3</td>
</tr>
<tr>
<td>Ni²⁺-affinity chromatography</td>
<td>188</td>
<td>73.3</td>
<td>137.8</td>
<td>19.6</td>
<td>3.4</td>
</tr>
<tr>
<td>Refolding c</td>
<td>114</td>
<td>72</td>
<td>82.1</td>
<td>11.7</td>
<td>3.4</td>
</tr>
</tbody>
</table>

aProtein concentration is determined by Bradford assay and BSA as a standard protein
bProtein purity is determined by SDS-PAGE densitometry
cRefolding at optimized conditions (room temperature, 0.25 M L-arginine, 0.5 M Tris-HCl, pH 7).

3.3 Identifying the best conditions to refold active Ulp1416-621

The denatured and affinity purified Ulp1416-621 was diluted rapidly into nine different buffers at 4 °C and RT to identify the optimal refolding conditions (Table 1). The degree of refolding at the 18 different conditions was determined by the amount of solubilized Ulp1416-621 (Equation 1) and its ability to digest the SUMO-FAS1-4 (WT) fusion protein substrate (Equation 2). Furthermore, Ulp1 activity was evaluated based on the amount of cleaved substrate, normalized by the available refolded (soluble) Ulp1416-621 (Equation 3). When refolded without L-arginine Ulp1416-621 exhibited very low values of solubility, ranging from 0% to 26.2 ± 9.66% (data not shown). In contrast, the solubility results from refolding Ulp1416-621 in buffers containing 0.25 M and 0.5 M L-arginine showed levels of solubilized Ulp1416-621 in the range of 60–80% (Figure 3A and C). The activity of the refolded Ulp1416-621 was assessed by its ability to cleave the SUMO-tag off the SUMO-FAS1-4 fusion protein. The close molecular weight values of the SUMO-tag (13.4 kDa) and the FAS1-4 domain (16.9 kDa) caused the two proteins to migrate similarly when analyzed by SDS-PAGE (Figure 3B). Noteworthy, the two bands show a bit slower migration than those expected considering their actual molecular masses. When Ulp1416-621 is refolded at pH 6 less cleavage is observed in the performed digestions compared to the rest of the refolding conditions, where the recovered Ulp1416-621 cleaved between 50 and 75% of the available substrate (Figure 3D). The estimated Ulp1 activities (Figure 3E) indicate that the two
best conditions to refold active Ulp1_{416-621}, based on the calculated mean values, are 0.25 M L-arginine, pH 7, RT (1.24 ± 0.07) and 0.25 M L-arginine, pH 8, 4 °C (1.21 ± 0.13). As those two conditions have very similar activities when the SUMO-cleaved substrate was measured after 24 hours of digestion, the proteolysis was also measured only 30 minutes after the beginning of the reaction to indicate the initial rates of the cleavage (Figure 3F).

**Figure 3.** L-arginine mediated refolding and activity of Ulp1_{416-621}. The indicated pH values, L-arginine concentrations, and temperatures correspond to the different refolding conditions of Ulp1_{416-621}. Standard
deviations are shown with error bars. A) SDS-PAGE analyses of Ulp1416-621 refolded at 0.25 M L-arginine, pH 7 and RT. From left to right: denatured Ulp1416-621 (lane 1) and refolded Ulp1416-621 performed in triplicates (lanes 2, 3 and 4). B) SDS-PAGE analyses of representative activity assays of Ulp1416-621 refolded at 0.25 M L-arginine and pH 7. The digestion reactions were incubated for 24 hours at RT. From left to right: undigested SUMO-FAS1-4 fusion protein substrate (lane 1), digest of SUMO-FAS1-4 fusion protein with Ulp1416-621 refolded at 4 °C (lanes 2, 3, and 4), and at RT (lanes 5, 6, and 7). C) Soluble Ulp1416-621 after refolding under the tested conditions. D) Levels of SUMO-fused substrate cleaved with Ulp1416-621 refolded under the tested conditions. E) Ulp1 activity presented by the ratio between the amount of cleaved SUMO-fused substrate (shown in panel D) and the solubility of refolded Ulp1416-621 (shown in panel C). Columns represent the ratio of the mean values (from panels C and D) and error bars indicate the propagated standard deviations. F) Levels of SUMO-fused substrate cleaved after 30 minutes of reaction with Ulp1416-621 refolded at 0.25 M L-arginine, pH 7, RT and 0.25 M L-arginine, pH 8, 4 °C. Columns in panels C, D) and E) represent the mean values of measurements performed in triplicates and analyzed by SDS-PAGE densitometry.

The digestion with Ulp1416-621 refolded at 4 °C in 0.25 M L-arginine buffer (pH 8) shows no proteolytic activity after 30 minutes, while the reaction using Ulp1416-621 refolded at pH 7 and RT has cleaved a significant fraction of the substrate fusion protein after 30 minutes (Figure 3F). Therefore, refolding Ulp1416-621 at pH 7 and RT in the presence of 0.25 M L-arginine is optimal from the tested refolding conditions. The catalytic activity of the refolded enzyme is comparable to the one of commercially available Ulp1, when the level of SUMO-cleaved substrate is measured (Figure S1). We were also able to refold 1 ml volumes of the His-tagged Ulp1416-621 using ten-fold rapid dilution and the identified optimal conditions (data not shown).

The digestion of SUMO-FAS1-4 fusion protein with Ulp1416-621, refolded using the optimal conditions, was measured at three more time points (1, 8 and 24 h) over the course of 24 hours. The reaction reached a plateau of approximately 90% cleaved substrate before 8 h, indicating that the proteolysis has reached completion (Figure 4A). The level of cleaved substrate was moderately enhanced when the cleavage reactions were performed at 2 mM DTT (Figure 4B). Furthermore, using these conditions the activity of Ulp1416-621 in the presence of relatively more SUMO-fused substrate and in larger volumes was tested. In an upscaled 5 ml cleavage reaction, and a molar enzyme:substrate ratio of 1:2000, 95 ± 0.6 % of SUMO-FAS 1-4 was proteolyzed (Figure 4C).
Figure 4. Digestion of SUMO-FAS1-4 fusion protein with Ulp1<sub>416-621</sub> refolded at optimal conditions (0.25 M L-arginine, pH 7, RT). A) Levels of cleaved SUMO-fused substrate at different time points. The digestions with refolded Ulp1<sub>416-621</sub> in molar ratio 1:110 (Ulp1<sub>416-621</sub>:SUMO-FAS1-4) were performed at room temperature in triplicates and analyzed at 1, 8, and 24 hours by SDS-PAGE densitometry. The data points represent mean values while error bars indicate the corresponding standard deviations. B) SDS-PAGE gel showing the effect of DTT on the activity of refolded Ulp1<sub>416-621</sub>. From left to right: undigested SUMO-FAS1-4 fusion protein substrate (lane 1), digests of SUMO-FAS1-4 fusion protein without added DTT (lanes 2, 3 and 4) and at 2 mM DTT (lanes 5, 6 and 7). C) Upscaled (5 ml) proteolysis of SUMO-FAS1-4 fusion protein substrate. From left to right: undigested SUMO-FAS1-4 fusion protein substrate (lane 1), cleavage of SUMO-FAS1-4 fusion protein at 1:2000 enzyme-to-substrate molar ratio incubated for 24 hours and RT (lane 2, 3 and 4). In panels B) and C) the bands of the SUMO-tag and the FAS1-4 domain appear as one merged band due to their close migration on the SDS-PAGE gels.

4. Discussion

The catalytic domain of SUMO protease is essential in SUMO fusion systems for cleaving off the SUMO-tag from the target protein. *S. cerevisiae* SUMO protease 1 (Ulp1) and SUMO (Smt3) are widely used for recombinant protein production. We have developed a novel and alternative method to produce active Ulp1 catalytic domain, avoiding the observed growth inhibition when expressing soluble Ulp1 catalytic domain in *E. coli.*

4.1 Expressing soluble Ulp1<sub>403-621</sub> is growth inhibitory to *E. coli*

In this study, we observed that the expression of the soluble Ulp1<sub>403-621</sub> catalytic domain in *E. coli* inhibits growth, while production of the insoluble and N-terminally truncated Ulp1<sub>416-621</sub> variant does
not inhibit growth. Thus, these results indicate that residues L403-V415 in Ulp1, which comprises approximately half of helix A and partly the preceding sequence [10], are important for the folding of a soluble and active SUMO protease 1 catalytic domain. When a full-length or catalytically active deletion-mutants of Ulp1 have been overexpressed in yeast, cell growth is impaired and the levels of bulk-sumoylated proteins are decreased [19]. This suggests catalytically active Ulp1 at elevated concentrations causes a shift in the equilibrium towards desumoylation thereby impairing the biological role of the sumoylation processes in the cell. Even though SUMO proteins are found only in eukaryotic cells, the presence of other ubiquitin-like proteins has been reported in prokaryotic cells [20, 21]. Thus, the reduction in E. coli growth, after inducing expression of soluble Ulp1403-621 catalytic domain, suggests that the active Ulp1 proteolytic domain is somehow toxic to the cell, which could be due to interaction with and possible cleavage of an endogenous E. coli protein. To our knowledge an E. coli endogenous Ulp1 substrate has never been reported. The lower cell-density obtained when expressing the soluble Ulp1403-621 catalytic domain presents a bottleneck for obtaining a high Ulp1 catalytic domain yield (Figure 1).

4.2 An alternative route for the production of active recombinant Ulp1 catalytic domain (residues 416-621)

Formation of inclusion bodies is often the result of an imbalance between in vivo protein aggregation and solubilization. Inclusion bodies formed during heterologous expression primarily contain the target protein in an inactive form [22, 23], which sometimes can be an advantage if the heterologous protein is toxic to the host cell or susceptible to processing by endogenous proteases [22, 23]. Thus, targeting protein expression in inclusion bodies can result in high levels of enriched heterologous protein which is protected from proteolytic degradation in the host cell. In addition, inclusion bodies can be easily separated from the endogenous cellular proteins through centrifugation, facilitating subsequent protein purification [24]. However, the main challenge when using this protein production strategy is often the recovery of a correctly folded, soluble and active target protein from the inclusion bodies.

The high-yield expression of Ulp1416-621 and the developed simple-to-use refolding procedure offer an alternative route for producing a considerable amount of active Ulp1 catalytic domain, which can be used for cleaving off the SUMO-tag from recombinant SUMO-fusion proteins (Figure 5). We anticipate that optimization of the intermediate steps in the procedure including isolation of inclusion
bodies and Ni\textsuperscript{2+}-affinity purification will increase the final yield of active Ulp1\textsubscript{416-621} per liter of cell culture beyond what we have reported in the present study.

**Figure 5. Procedure for production of active recombinant Ulp1\textsubscript{416-621} catalytic domain from inclusion bodies in E. coli.** The flowchart gives an overview of the different steps in producing active recombinant His-tagged Ulp1\textsubscript{416-621} catalytic domain from *E. coli.*

The effects of L-arginine concentration, pH, and temperature on the refolding were tested in small-scale fast dilution experiments. The experiments show that L-arginine greatly increased the solubility and activity of Ulp1\textsubscript{416-621} catalytic domain while pH had different effects on the solubility and activity. Coutard *et al.* have reported an inverse correlation between the pI of a protein and the optimal
refolding pH [25]. According to that observation, the Ulp1416-621 variant (theoretical pI is approximately 8.7) should refold best under acidic conditions. However, we found that Ulp1416-621 actually refolded better at pH 7 and 8, than at pH 6. The temperature is also known to affect refolding [26] but in our experiments, the temperature effect was secondary to both L-arginine concentration and pH effects.

4.3 L-arginine plays a key role in the refolding of Ulp1416-621

Our results show that the refolding of Ulp1416-621 is a relatively robust process. Successful refolding was achieved at 73.3% protein purity, and although there were notable differences between the various refolding conditions, active SUMO protease 1 catalytic domain was obtained from all refolding experiments. In addition, the results show that the addition of L-arginine to the refolding buffer improves the yield dramatically. Ulp1416-621 recovered from all 0.25 M L-arginine-containing buffers had almost equal yields in terms of soluble protease. However, when the ability of refolded Ulp1 catalytic domain to cleave SUMO-fused protein was compared, Ulp1416-621 refolded at pH 6 had insufficient activity compared to the domain refolded at pH 7 and pH 8. Previous reports imply that L-arginine influences the refolding process by suppressing protein aggregation. A proposed mechanism involves the formation of L-arginine clusters with a hydrophobic methylene layer that interacts with the hydrophobic surface of the unfolded inactive protein [27]. This shields the protein's aggregation-prone zones from making intermolecular hydrophobic interactions. However, the solubility-enhancing effect of L-arginine can also lead to aberrantly folded and inactive soluble protein [28]. This ability of L-arginine to trap a protein in soluble intermediate states which are not able to form proper interactions could be a factor in our observed activity reduction for the soluble Ulp1 refolded at pH 6 [29]. We also note that the 0.75 M urea present in all refolding experiments might have contributed to suppress unwanted protein aggregation during refolding. In fact, L-arginine and urea can have complementary effects on protein refolding by suppressing different types of hydrophobic interactions [28].

5. Conclusion

We report an inclusion body-directed method for expressing truncated SUMO protease catalytic domain. Our Ulp1416-621 variant demonstrates superior production yields in E. coli flask and circumvents the growth inhibition observed with expression of the canonical and soluble Ulp1 catalytic domain, Ulp1403-621. Active Ulp1416-621 protease can be recovered through efficient L-
arginine-mediated refolding. Based on solubility and activity measurements we conclude that the best parameters for refolding recombinant Ulp1416-621 are 0.25 M L-arginine, pH 7 and room temperature.

After optimizing the conditions for the cleavage reactions, we show that the reaction can be upscaled and the ratio between the SUMO-fused substrate and Ulp1416-621 can be modulated in order to e.g. alter the cleavage time. Using the procedure outlined in the present work (Figure 5), approximately 80 mg of active recombinant Ulp1416-621 catalytic domain can be produced from one liter of high cell-density E. coli culture. Our study demonstrates the feasibility of an inclusion body-directed expression strategy for a well-known research-relevant protease. The final yield of active protease can likely be increased by further optimization of the inclusion body recovery and protein purification steps.

Acknowledgements

This work was partly supported by a grant from the Fight for Sight Denmark foundation. The foundation had no role in experimental design, data collection and analysis, preparation of the manuscript, or decision to publish. We would like to thank Dr. Hideo Iwai, University of Helsinki, Institute of Biotechnology for providing the plasmid pFGET19_Ulp1 (Addgene plasmid #64697) [9].

References


Supplementary Information

for

A novel approach for production of an active N-terminally truncated Ulp1 (SUMO protease 1) catalytic domain from Escherichia coli inclusion bodies

Marina Y. Linova a, Michael W. Risør b, Sanne E. Jørgensen a, Zohra Mansour a, Jacob Kaya a, Jens J. Sigurdarson a, Jan J. Enghild b, Henrik Karring a*

a Department of Chemical Engineering, Biotechnology and Environmental Technology, University of Southern Denmark, Campusvej 55, DK-5230 Odense M, Denmark

b Department of Molecular Biology and Genetics, Aarhus University, Gustav Wieds Vej 10C, DK-8000 Aarhus C, Denmark

*Corresponding author. Address: Department of Chemical Engineering, Biotechnology and Environmental Technology, University of Southern Denmark, Campusvej 55, DK-5230 Odense M, Denmark

E-mail address: hka@kbm.sdu.dk

Tel: +45 21356350
Comparative proteolytic analysis

1. Materials and methods

SUMO-FAS1-4 proteolysis with Ulp1<sub>416-621</sub> refolded at optimal conditions (0.25 M L-arginine, pH 7, RT) and commercial Ulp1 (Invitrogen - Cat. No. 12588-018) was followed at 0.5, 1, and 8 hours. As previously, the cleavage reactions were performed in molar ratio 1:110 (enzyme:substrate) and at room temperature.

2. Results

The difference between the observed level of cleaved SUMO-fused substrate after 8 hours of digestion with Ulp1<sub>416-621</sub> (88.97 ± 7.63 %) and commercial Ulp1 (92.8 ± 0.54 %) was close to negligible. The results obtained at 0.5 and 1 hour for the two enzymes showed different degree of substrate proteolysis. However, the standard deviations for measurements at those time points, and in particular for the refolded Ulp1<sub>416-621</sub>, exhibited high standard deviations.

```
Figure S1. Digestion of SUMO-FAS1-4 with refolded Ulp1<sub>416-621</sub> (□) and commercial Ulp1 (○). The levels of cleaved SUMO-fused substrate after digestion with the two proteases were analyzed. The digestions were performed in triplicates and analyzed at 0.5, 1, and 8 hours by SDS-PAGE densitometry. The data points represent mean values while error bars indicate the corresponding standard deviations.
```

---

The levels of cleaved SUMO-fused substrate after digestion with the two proteases were analyzed. The digestions were performed in triplicates and analyzed at 0.5, 1, and 8 hours by SDS-PAGE densitometry. The data points represent mean values while error bars indicate the corresponding standard deviations.