Al-Qadisiyah Journal of Pure Science
ISSN(Printed): 1997-2490 ISSN(Online): 2411-3514
DOI: 10.29350/jops
Al-Qadisiyah Journal of Pure Science

Construction, expression and characterization of TEV protease mutants engineered for improved solubility

George Kostallas, Lisa Sandersjöö, Mohammed Al-Askeri and Patrik Samuelson

Division of Molecular Biotechnology, School of Biotechnology, AlbaNova University Center, Royal Institute of Technology (KTH), SE-106 91 Stockholm, Sweden

(*) Corresponding author. E-mail address: patrik@biotech.kth.se

abstract

In recent years, the highly sequence specific tobacco etch virus protease (TEVp) has emerged as one of the most popular and widely used reagents for removal of fusion tags from target proteins. Its use, however, has been hampered due to relatively poor solubility and inefficient expression in E. coli. Although a lot of progress has been made, there is still need for new and improved TEVp variants. Recently, two different gain-of-function TEVp mutants were described; one containing the substitutions L56V/S135G, which conferred improved solubility and activity in vitro, while the other mutant, containing the substitutions T17S/N68D/I77V, was claimed to yield more soluble protease than the wild-type (wt) protease upon overexpression in E. coli. Here, we analyzed if the L56V/S135G substitutions could promote increased solubility also in vivo, as that would be beneficial to TEVp production and had never been investigated before. We also intended to create a novel, and hopefully superior, TEVp variant with all five mutations combined (T17S/L56V/N68D/S135G/I77V) in a single protease molecule. This variant and the two parental TEVp variants as well as the wt protease, were all expressed in E. coli and characterized with respect to the expression levels, solubility and activity using several different techniques; among them, a newly developed fluorescence-assisted whole-cell assay that directly reports on the apparent protease activity in vivo. Our results show that the L56V/S135G substitutions improve the solubility not only in vitro but also in vivo, which did hold true for the activity as well. Disappointingly, the protease variant containing all five substitutions (T17S/L56V/N68D/S135G/I77V) did not show the best performance, which instead the L56V/S135G variant did. In contrast to an earlier report, we show that the substitutions T17S/N68D/I77V, did not improve the TEVp solubility. In fact, they reduced the activity, and even appeared to have a slightly negative effect on solubility, of all protease constructs in which they were present. Thus, the best current and most promising TEVp variant for future protease engineering efforts, towards improved expression properties and enhanced catalytic efficiency, are those containing the L56V/S135G substitutions.

keywords: TEV protease; tobacco etch virus protease; solubility; E. coli; protein engineering; GFP; mutagenesis

1. Introduction

Heterologous protein expression through the use of recombinant DNA technology is one of today's most popular and cost-efficient strategies for production of high-value target proteins. Frequently, these products are produced as fusion proteins, where the partner serves to (i) facilitate the detection and purification process, (ii) improve the stability, solubility and yield of the target protein, or (iii) function as a localization tag [1-3]. However, as fusion tags may alter the protein properties they can also be deleterious to protein function and structure [4,5], and are therefore often removed prior to final use of the target molecule. Typically, this is accomplished through the use of site-specific proteases that recognize and cleave a substrate

peptide positioned between the target and the fusion domain. Traditionally, proteases such as enterokinase, thrombin and factor Xa have been the primary choice. Lately, however, a class of proteases that exhibit an even more refined sequence specificity has emerged as an attractive alternative, and includes the viral proteases from tobacco mottling virus [6,7], human rhinovirus 3C [8] and tobacco etch virus [9]. Today, tobacco etch virus protease (TEVp) is one of the most widely used, and there are several reasons to its popularity. First, TEVp is highly stringent with respect to its recognition sequence, ENLYFQG/S. Second, upon cleavage it leaves only one amino acid (position P1') at the new N-terminal end; where glycine (G) or serine (S) that normally occupies P1' in the substrate sequence can be substituted by several other amino acids and still generate functional cleavage sites (Kapust et al, 2001). Third, the protease can operate under a wide range of conditions with respect to temperature and buffer composition [10].

TEVp, which usually refers to the catalytically active 27 kDa C-terminal domain of the nuclear inclusion a (NIa) proteinase from tobacco etch virus [11], can be produced in Escherichia coli. However, functional expression in this host has been difficult. Mainly due to the presence of several rare arginine codons in the TEVp gene, low solubility in vivo, and that the protease is subject to auto-inactivation [12,13], which result in low yields of active TEVp. Over the years, considerable efforts have been invested to overcome these issues. For instance, the solubility has been improved through low temperature expression [14], co-expression with chaperones [14], incorporation of fusion partners [15] and alternative E. coli strains [16]. Auto-inactivation has been decreased through certain amino acid substitutions at position 219 [12,17] while the codon-bias problem has been solved through codon optimization [13]. Recently, van den Berg et al. applied directed evolution techniques to improve the protease solubility [18]. More specifically, flow cytometry screening of E. coli cells expressing TEVp random mutant libraries fused to GFP (as a folding reporter) enabled the identification of TEVp variants that conferred increased whole-cell fluorescence intensity, which was interpreted as gain-of-function mutants with respect to the intrinsic protease solubility [18]. Their best variant TEVSH, containing the amino acid substitutions T17S/N68D/I77V, resulted in > five-fold increase in the yield of purified protease [18]. Despite all these valuable advances that collectively have resulted in > 65-fold increase in the volumetric productivity of TEVp in shake-flask cultures [16,19], the recombinant protease still suffers from relatively low solubility in solution (generally below 1 mg/mL) as well as in vivo. Cabrita et al. addressed this issue in a recent article in which they describe an in silico design approach that enabled the identification of several single-site mutations predicted to improve the stability of TEVp [20]. Two of these mutations, L56V and S135G, were also highly interesting candidates to enhance the protease solubility as they increased the overall surface polarity of the protease [20]. This was also confirmed in subsequent (expression and characterization) experiments, where a fully active double mutant (L56V/S135G) remained soluble in vitro at concentrations > 40 mg/mL. However, whether it also exhibits improved solubility in vivo has remained an open question.

In spite of all these advancements, both the use of TEVp in different applications and its production would benefit from further improvements in the solubility, activity and stability of the protease. Therefore, based on these two different TEVp mutants, we decided to create a novel TEVp variant in which all five mutations had been combined. We hoped that this version would constitute a superior protease to be used in various biotechnical applications as well as a promising starting point for further protease engineering efforts.

The novel and the two parental TEVp variants, as well as the wild-type (wt) protease, were expressed in E. coli and characterized with respect to their expression levels, solubility and activity. This was done through a variety of different techniques, including a newly developed fluorescence-assisted whole-cell assay that directly reports on the apparent protease activity in vivo.

2. Results

2.1. TEVp solubility analyzed in vivo by using a folding interference assay

The highly sequence specific protease from tobacco etch virus has emerged as one of the most popular and efficient reagents for removal of fusion partners from recombinantly produced target proteins. In recent articles, two different TEVp mutants, engineered for improved solubility and stability, have been described one contained two amino acid substitutions (L56V/S135G; Cabrita et al. [20]) while the other had three substitutions (T17S/N68D/I77V; van den Berg et al. [18];). Mapping these mutations onto the TEVp crystal structure [21] located them either at or close to the protease surface. Therefore, we wanted to investigate whether they could be combined into a single protease molecule for further improvements in the solubility without compromising the proteolytic activity.

We introduced all relevant mutations into a previously constructed protease variant [22,23], here called TEVpwt, which already (i) contained a S219V substitution to minimize auto-inactivation [12] and (ii) had all six rare arginine (R) codons in the gene (R49, R50, R80, R101, R105, R159) substituted for more frequently used synonymous codons [13]. Thus, in the end we had three new protease variants besides TEVpwt, containing two (L56V/S135G), three (T17S/N68D/I77V) or all five mutations (T17S/L56V/N68D/S135G/I77V), which we have called TEVp2m, TEVp3m and TEVp5m, respectively.

The use of GFP as a reporter for the successful folding of a target protein fused at the N-terminal end of GFP is becoming increasingly popular, and has also proven effective in several protein solubility engineering projects [24]. Therefore, as an initial test, we thought it would justify checking the solubility of the different protease variants in vivo as fusions to GFP; especially since (i) the original TEVp3m-variant first had been identified as a gain-of-function mutant from a GFP-fused random mutant TEVp library, and (ii) the in silico designed variant by Cabrita et al., corresponding to TEVp2m, only had been tested in solution previously, where the solubility had increased from 1.5 mg/mL to > 40 mg/mL compared to the wild-type TEVp [20].

Thus, all four protease variants were expressed in individual E. coli cultures, and the whole-cell fluorescence intensity was then analyzed on a flow cytometer. As can be seen in Fig. 1, both TEVp2m (mean fluorescence intensity, MFI=1478) and TEVp5m (MFI=1277) yielded significantly higher whole-cell fluorescence intensities than TEVpwt (MFI=725) and TEVp3m (MFI=684). TEVp2m actually proved to be the best protease variant of all four tested (Fig. 1). Moreover, in comparison with TEVpwt, TEVp3m did not seem to be improved at all with respect to the in vivo solubility as they generated essentially the same fluorescence intensities (Fig. 1). To our surprise, the engineered variant, TEVp3m, in fact seemed to be the least soluble protease variant of all four, as judged from the whole-cell fluorescence intensity data (Fig. 1). This result was highly unanticipated, especially as this variant once originated as a gain-of-function mutant from a GFP-TEVp library [18].

This initial experiment then demonstrated that the L56V/S135G substitutions present in TEVp2m and TEVp5m not only improve the protease's solubility in solution but also in vivo, which has not been shown

before. Moreover, in contrast to the previous study by van den Berg et al., their mutations (T17S/N68D/I77V) actually appeared to have a slightly negative effect on the protease solubility, at least in our system and experimental conditions.

2.2. Substrate processing efficiency analyzed by using a novel fluorescence-assisted whole-cell assay

Recently, we created a novel function-based protease assay that reports on the apparent catalytic activity in vivo [22,23]. In brief, the assay utilizes short-lived plasmid-encoded fluorescent substrates as a quantitative reporter system in E. coli. More specifically, the reporter consists of ssrA-tagged green fluorescent protein (GFP) that contains a protease cleavage site (PS) between the GFP and the ssrA-moiety. As long as the ssrA-tag is present, the cytoplasmic ClpXP degradation machinery will capture and destroy the reporter protein [22,23,25,26], and the cells appear dark. However, upon removal of the degradation tag, which can be achieved through cleavage by a coexpressed site-specific protease, GFP will be saved and the fluorescence intensity of the entire cell is increased [22,23], which can be monitored on a flow cytometer.

We have recently used this technique to (i) analyze and discriminate among protease variants that exhibit differences with respect to their in vivo solubility [22], and (ii) through high-throughput screening of large combinatorial substrate libraries identified the substrate profile of TEVp [23]. Here, we used the assay to examine the apparent catalytic activity in vivo of the four TEVp variants. In this particular case, the reporter protein contained the natural substrate peptide for TEVp, namely ENLYFQG (denoted subG), which is one of the best TEVp substrates known [27]. Individual clones, each coexpressing a unique TEVp variant and the reporter protein GFP-subG-ssrANY [22], were subjected to flow cytometry analysis and ranked on the basis of their whole-cell fluorescence intensities (Fig. 2).

Moreover, since maltose binding protein (MBP) fused to TEVp has proven to have a marked positive effect on the protease solubility [15,22], we thought it would be interesting to see how such a MBP-TEVp fusion compares to the solubility-engineered variants constructed and used in this study. Therefore, we also included cells expressing TEVpwt as a transient fusion to MBP (encoded from pMal-TEV2) [22]; the MBP-moiety is cleaved off by the protease itself in the bacterial cytoplasm due to the presence of the natural TEVp substrate peptide (subG) between the MBP and TEVpwt domains [15,22]. For comparison and as a negative control, we also used cells that instead coexpressed that same protease variant but together with a non-cleavable reporter GFP-ssrANY [22] lacking subG.

All TEVp variants seemed functional in vivo as they generated whole-cell fluorescence intensities greater than the negative control (MFI=9) (Fig. 2). We feel confident that this gain-of-fluorescence represents true substrate processing as we in a previous study have shown that only catalytically active protease induced a significant shift in the fluorescence intensity [22].

We only observed very small differences in fluorescence intensity from the different protease variants. Nonetheless, TEVp2m yielded the highest fluorescence intensity (MFI=28) closely followed by TEVp5m (MFI=25), TEVpwt (MFI=23), and TEVp3m (MFI=23) (Fig. 2). Thus, the order was consistent with the data from the initial in vivo solubility experiment in which the corresponding protease variants were fused to GFP (Fig. 1), suggesting that the observed differences in fluorescence intensity (i.e., apparent substrate processing efficiency) to some degree might be related to dissimilarities in the proteases' in vivo solubility. However, we could not rule out the possibility that the actual reason instead was due to differences in the specific

activity of each variant or a combination thereof. Actually, when Cabrita et al. investigated the in vitro activity of their protease variant, containing the L56V/S135G substitutions (i.e., equivalent to TEVp2m in our study), it proved to be slightly improved over the wt protease, with a kcat/Km value of 7.08 vs 6.11, respectively [20]. Therefore, it is not farfetched to believe that the elevated fluorescence intensities we observed for TEVp2m and TEVp5m were the net result from alterations in the proteases' solubility and catalytic efficiency.

Although the solubility-engineered TEVp variants containing the L56V/S135G substitutions (TEVp2m and TEVp5m) exhibited an improved apparent processing efficiency, they could not by far compare to TEVpwt expressed as a transient fusion to the solubility enhancing MBP-moiety (MFI=322) (Fig. 2). Thus, MBP appears to be much more efficient at increasing the protease's solubility than the solubility enhancing mutations, L56V/S135G.

2.3. Characterization of substrate hydrolysis in vitro

To find out what effect the residue substitutions have on the proteases' catalytic efficiency, we performed an in vitro cleavage assay. This would also allow us to elucidate how the hydrolysis data from the previous in vivo activity assay relates to those measured in vitro. The substrate used was ABP-subG-ZZ, which is a fusion between an albumin binding protein (ABP) derived from streptococcal protein G and the IgG binding protein A-derivative, Z, connected by the TEVp substrate peptide subG [23].

We made sure that we only used soluble and presumably therefore also active protease in the assay. To this end, the different TEVp variants had first been produced as transient MBP-fusions from a T7 driven promoter, and the soluble protease fraction (devoid of the MBP-moiety) was purified on an immobilized metal ion affinity (IMAC) column and quantified prior use.

For the cleavage assay, the protease and substrate were mixed at a molar ratio of 1:30 in TEVp reaction buffer and cleavage was followed over time at 37°C. The cleavage reactions were separated on an SDS-polyacrylamide gel (Fig. 3) and stained with Gelcode Blue Stain Reagent (Pierce). Cleavage proved to be rapid and close to 100% efficient for all protease variants over the 6 h incubation time (Fig. 3), and they showed similar but not identical cleavage efficiencies. More specifically, TEVp2m performed better than TEVpwt, which was expected since an earlier study established that a corresponding variant containing the L56V/S135G substitutions exhibited higher catalytic efficiency than the wt protease in vitro, with a kcat/Km of 7.08 vs 6.11, respectively [20]. Not only was TEVp2m better than TEVpwt, but it also turned out to be the best variant of all that we tested (Fig. 3). TEVp5m showed the second best performance, just slightly better than TEVpwt, while TEVp3m again turned out to be the least efficient variant. The fact that TEVp3m demonstrated the poorest catalytic efficiency was a bit of a surprise, especially since van den Berg et al. claimed that it showed similar activity as that of the wt protease in their cleavage assay [18]. However, the substrate conversion is fast and almost complete after just 1h, and since their earliest sampling point was 1h, it is likely that they may have missed any differences that would have been apparent at an earlier time-point.

Nevertheless, overall, our in vitro cleavage data were consistent with the results from the in vivo substrate processing whole-cell assay (Fig. 2), and, interestingly, also the in vivo solubility assay (Fig. 1). Collectively, the data thus implies that the substitutions L56V/S135G and T17S/N68D/I77V have a positive and negative effect, respectively, on the catalytic efficiency and solubility, as all TEVp variants in which they appeared

performed either better or worse, respectively, than their counterparts in which they were absent (Fig. 1, Fig. 2 and Fig. 3).

2.4. Assessment of the protease expression levels and distribution between the soluble and insoluble fraction in cells overexpressing the different TEVp variants

Two series of experiments were performed in which the proteases were over-expressed (at 25°C from a T7 driven promoter), either with or without a transiently attached MBP-moiety. The insoluble and soluble protein fractions from the cell lysates were then examined using SDS-PAGE and subsequent protein staining.

Expression with the transiently attached MBP proved to be crucial for getting (any significant amounts of) soluble protease (Fig. 4) When the different TEVp variants were expressed without the solubility enhancing MBP, they all ended up, almost exclusively, in the insoluble fraction (Fig. 4A). On the contrary, the presence of the MBP clearly shifted the equilibrium towards more soluble protease being produced (Fig. 4B), which was consistent with an earlier report [15]. Surprisingly, we could not detect any significant differences between the various TEVp variants, neither in terms of the solubility nor the productivity, regardless of the specific expression strategy used (i.e., with or without MBP).

3. Discussion

TEVp has attracted a lot of interest as an efficient tool for site-specific removal of fusions tags from recombinant target proteins and as an indispensable tool in biomedical research. But there is a need for new and improved TEVp variants not suffering from the low intrinsic solubility and poor expression characteristics still usually associated with the current best TEVp variants available. Here, we investigated if certain amino acid substitutions, previously described as beneficial to the protease's solubility and stability [18,20], could be combined in a single molecule to generate a superior protease variant. Hopefully, this would then facilitate not only the production of the protease but also its use in different settings. Potentially, such a variant could also constitute a promising starting point for further protease engineering efforts.

Four different TEVp variants were constructed and expressed in E. coli and characterized with respect to the solubility and activity both in vivo and in vitro. For the first time, we show that the substitutions, L56V/S135G, not only improve the in vitro solubility of the protease [20] but also its solubility in vivo (Fig 1). For example, the variants containing these substitutions (TEVp2m and TEVp5m) yielded higher whole-cell fluorescence intensities as GFP-fusions than their counterparts lacking the specified mutations (TEVp and TEVp3m) (Fig. 1). Moreover, we confirmed that the L56V/S135G substitutions also were beneficial for the catalytic activity (Fig. 2 and Fig. 3), consistent with the results by Cabrita et al. [20].

Interestingly, in contrast to what was reported by van den Berg et al., the mutations T17S/N68D/I77V did not at all improve the protease solubility, at least not in our experiments. Actually, the substitutions seemed to have a slightly negative effect on solubility and were clearly detrimental to the activity of all protease constructs in which they were present (TEVp3m and TEVp5m) (Fig. 1, Fig. 2 and Fig. 3). To us, this came as a surprise, especially since these substitutions originally were identified in a gain-of-function mutant that appeared when van den Berg et al. screened a TEVp random mutant library fused to GFP [18]. However, upon closer examination of their report, it turns out that they never reexamined this variant as a GFP-fusion on a flow cytometer, and it is therefore difficult to say whether or not it represents a "true" gain-of-function mutant (with respect to the in vivo solubility as a GFP-fusion). Nevertheless, when they overexpressed it as

a fusion to a C-terminally attached V5 epitope and hexahistidine tag, it yielded approximately 5.5 times more purified and soluble TEVp than the corresponding wt protease fusion did [18]. However, it is not uncommon that a particular protein's solubility is context dependent, which actually seems to be the case with their engineered TEVp variant that was less soluble as a GFP-fusion than when fused to the V5-hexahistidine tag [18]. Unfortunately, as they never disclosed that data in any greater detail, it is difficult to estimate how the reported improvement in solubility as a V5-fusion translates to the GFP-fused protease, other than it at least should be less than 5.5 times.

At present, there is no obvious explanation to why this set of substitutions (T17S/N68D/I77V) gave different results in the studies by van den Berg et al. and us, respectively. However, there are a number of dissimilarities between the two studies, which may account for some of the discrepancies observed. First, their solubility-improved variant was isolated from a TEVp-GFP library expressed and screened at a temperature of 20°C as opposed to our experiments that were conducted at 25°C (for TEVp overexpression), 30°C (for the in vivo processing assay) or 37°C (for the in vitro cleavage assay as well as the in vivo solubility assay as GFP-fusions). We therefore speculate that the T17S/N68D/I77V substitutions may confer improved folding and solubility at 20°C but not necessarily at the elevated temperatures used in our experiments.

Second, another significant difference is that although being highly similar, the protease variants used in the two studies were not entirely identical. For example; (i) the serine (S) in position 219, which affects the auto-inactivation propensity [12,17], were substituted with asparagine (N) and valine (V) in the proteases used by van den Berg et al. and us, respectively, (ii) our protease variants had all rare arginine codons substituted for more frequently used synonymous codons, (iii) the peptide linker joining the protease and GFP were also different in the two studies, (iv) in both studies, the overexpressed proteases were fused to polyhistidine tags, which were attached to the C-terminal and N-terminal ends of the protease used by van den Berg et al. and us, respectively, and (v) besides the specified mutations, T17S/N68D/I77V, their TEVp gene also contained two silent mutations (unfortunately, they do not specify what type and where) that were not present in our system. Naturally, all these differences result in slightly different mRNA sequences that may affect the (i) the mRNA stability and conformation [28] and (ii) the codon bias (i.e., a non-frequently used codon can be exchanged for a synonymous that is used more often, or vice versa), which ultimately can have bearing on the total expression level, the solubility as well as the folding-state of a protein [29-31]; in our particular case then the different TEVp variants. Interestingly, Kimchi-Sarfaty et al. recently observed that already very small differences between alternative gene variants, involving synonymous single-nucleotide polymorphisms (SNPs) but preserved polypeptide chain, can result in a protein with altered structure and function [32]. In particular, SNPs that involve frequent-to-rare (or vice versa) codon substitutions are important, as that may affect the ribosome stalling propensity due to changes in the concentration of available cognate tRNAs or alterations of the RNA structure, which could be the case here. It was recently proposed that long-enough ribosomal pause time scales actually might lead to alternative folding pathways and distinct minima in the folding free energy landscape where alternative protein structures can be kinetically trapped [31]. Thus, the silent mutations in the TEVp variant by van den Berg et al. perhaps promoted soluble expression, but on this we can only speculate.

Nevertheless, what we do know for sure is that our TEVp genes have been subject to rare-to-frequent arginine codon optimizations, and thereby a no-pause case may occur for the ribosome during the co-translational folding of the protease. Perhaps, this result in higher expression level, but not necessarily more soluble protease, especially not as domain swapping that can affect the aggregation propensity is more likely in a no-pause case [31]. Thus, should this be the case for us, then the potentially positive effect of the mutations on solubility could be obscured by an increased tendency to form insoluble aggregates or other kinetically trapped inactive protease conformations. However, irrespective of the reasons behind the discrepancy in results obtained by us and by van den Berg et al., respectively, a recent study by Blommel and Fox is in favour of our results regarding the solubility effect of the T17S/N68D/I77V-mutations [33]. They could not detect any solubility improving effect from the mutations, but were unable to explicitly unravel the impact of these mutations on their experiments.

Upon examination of the SDS-PAGE data presented by van den Berg et al., there is no doubt that their engineered TEVp variant generated more protease in the soluble fraction compared to the parental protease variant, but the actual reasons behind this remain unclear. To us it appears as if their mutant not only yielded more soluble protease but also an increased total expression level, which then at least in part could explain their results. However, based on the different amounts of endogenous proteins that can be seen in the gel, we are uncertain of whether they actually have loaded directly comparable amounts representing the soluble and insoluble protein fractions for the different protease variants. Hence, it is difficult to dissect the specific contributions from enhanced intrinsic solubility and increased total expression level, respectively.

When van den Berg et al. characterized the TEVp mutant containing the substitutions T17S/N68D/I77V, they found that it exhibited cleavage efficiency comparable to that of the wt protease [18]. However, in our experiments the two protease variants did not show the same performance; in fact, the mutant protease proved to be less efficient. We believe that they never detected this difference in catalytic efficiency due to infrequent sampling on their hand; the cleavage was almost complete already at their first sampled time-point, one hour.

Although all TEVp variants could be expressed in E. coli, the protease solubility was enhanced significantly when produced as MBP-fusions. Interestingly, the solubility promoting effect of MBP even proved to be greater than the current best solubility-improving mutations, L56V/S135G; TEVpwt expressed as a transient fusion to MBP yielded higher whole-cell fluorescence intensity than cells expressing the best protease mutant, TEVp2m (Fig. 2).

Unfortunately, our efforts to create a greatly improved TEVp variant based on previously described solubility-improving mutations [18,20] failed. Most likely, since the mutations T17S/N68D/I77V in our experiments not were found to improve the solubility, but actually seemed to decrease the catalytic activity and the solubility. Thus, our data suggest that the current best TEVp variants are those containing the L56V/S135G substitutions as they improve both the solubility and the activity, and consequently hold the greatest promise for future protease engineering efforts towards improved and desired functionality.

4. Materials and methods

4.1. Bacterial strains and reagents

E. coli strain RR1ΔM15 [34] was used as host during construction of plasmids. E. coli strain DH5α (Gibco) was used for fluorescence-based analysis of TEVp activity in vivo. E. coli strain Rosetta (DE3)pLysS

(Novagen) was used for production of ABP-subG-ZZ and different TEVp variants, and also for flow cytometry analysis of various TEVp-GFP fusions. Culture media and chemicals were from Merck and Sigma-Aldrich, respectively. DNA modifying enzymes were all from New England Biolabs and used according to the manufacturer's recommendations. Primers were purchased from Eurofins MWG Operon (for a list of all oligonucleotides used in this study, see Supplementary Table S1).

4.2. Plasmid constructions

For an overview of the relevant important plasmids constructed and/or used in this study, see Supplementary Figure S1. Three vectors encoding different TEVp variants that contained different combinations of amino acid substitutions reported to improve the stability and solubility of the protease [18,20], were created. One with two amino acid substitutions (L56V and S135G), another with three substitutions (T17S, N68D, and I77V) and a third with these mutations combined (T17S, L56V, N68D, I77V and S135G) were created. The mutations were introduced step by step into the vector pTEV [23] using a QuikChange Site-Directed Mutagenesis Kit (Stratagene) together with the relevant primer pairs (L56Vfor/L56Vrev, S135Gfor/S135Grev, T17Sfor/T17Srev, N68Dfor/N68Drev, or I77Vfor/I77Vrev). All these steps finally resulted in pTEV2m (encoding the L56V/S135G-variant), pTEV3m (encoding the T17S/N68D/I77V-variant) and pTEV5m (encoding the T17S/L56V/N68D/I77V/S135G-variant).

Originally, pTEV encodes an "optimized" version of the protease where the autoproteolysis-site has been inactivated (S219V) [12,19] and all rare arginine codons exchanged for more common ones; a variant that we here refer to as the wild-type (wt) TEV protease, TEVpwt.

For production of the different TEVp variants, the corresponding genes were cloned into pTEVprod that originally encodes the TEVpwt as a fusion protein under the control of a T7-promoter. More specifically, the fusion consists of an N-terminally attached solubility-enhancing maltose binding protein (MBP), a TEVp-specific substrate peptide (ENLYFQG), a polyhistidine motif and the protease as the C-terminal domain (i.e., MBP-ENLYFQG-His6-TEVp). Thus, soluble protease will be released from the fusion partner upon expression in the bacterial cytoplasm as the protease itself cleaves within the substrate peptide present in the fusion protein.

The TEVp genes were transferred from pTEV2m, pTEV3m, and pTEV5m as KpnI/BamHI-fragments and ligated into the backbone KpnI/BamHI-digested pTEVprod, thus yielding pTEV2mprod, pTEV3mprod, and pTEV5mprod, respectively. In addition, based on these vectors including pTEV, we also created a set of TEVp production plasmids that did not contain the MBP encoding gene. The relevant gene fragments from pTEV, pTEV2m, pTEV3m, and pTEV5m, were PCR-amplified using primers GEKO32 and GEKO51 and the purified amplicons were digested with NcoI/XhoI and subsequently ligated into NcoI/XhoI-digested pAff8eGFP [35], thus yielding pTEVprodΔMBP, pTEV2mprodΔMBP, pTEV3mprodΔMBP, and pTEV5mprodΔMBP.

All three TEVp variants as well as the wt protease were also cloned as N-terminal fusions to green fluorescent protein (GFP) to allow for subsequent assessment of the protease solubility in vivo [36,37]. This was done by PCR-amplification of the relevant gene fragments from pTEV, pTEV2m, pTEV3m, and pTEV5m, using GEKO32 and GEKO33 as primers, and then ligating the purified NcoI/AscI-digested amplicons into NcoI/AscI-digested pAff8eGFP, thereby yielding pTEV-GFP, pTEV2m-GFP, pTEV3m-GFP, and

pTEV5m-GFP, respectively. All plasmid constructs were verified by standard DNA sequencing using Big Dye Terminator 3.1 Cycle Sequencing Kit and ABI Prisma 3700 sequencer (Applied Biosystems).

4.3. Flow cytometry analysis to evaluate the TEVp activity and solubility in vivo

For individual clone analysis to investigate the apparent catalytic activity of the different TEVp variants in vivo, overnight cultures of DH5α cells harboring a TEVp substrate reporter vector (pGFP-subG-ssrANY or pGFP-ssrANY [22,23] and a plasmid encoding the relevant TEVp variants (pTEV, pTEV2m, pTEV3m, or pTEV5m) were subcultured by dilution (1:75) into fresh LB broth containing 100 µg/ml ampicillin and 20 µg/ml chloramphenicol. The cultures were incubated at 37°C in a rotary shaker set at 150 rpm and when they reached a cell density of OD600 ≈ 0.5, IPTG was added to a final concentration of 0.1 mM to initiate TEVp expression (from a tac-promoter). The cultures were now placed in a shaker set at 30°C, 150 rpm, and after 30 minutes, expression of the reporter construct was induced by adding L(+)-arabinose to a final concentration of 0.2%. Two hours later, 1 ml of each culture was placed on ice and 5-10 µl from each sample was diluted with 1 ml ice-cold 1xPBS (11,68 g NaCl; 9,44 g Na2HPO4; 5,28 g NaH2PO4·2H2O; 1,000 ml MilliQ; pH 7,2) and kept on ice until analyzed on a FACSVantage SETM flow cytometer (Becton Dickinson). The throughput rate for the analysis was 300 events/sec with 488 nm excitation wavelength (argon ion laser), emission detection between 510 and 530 nm, and 10,000 events were recorded for each sample. For comparison and as a control, we also used cells that harbored pGFP-subG-ssrANY and a vector, pMal-TEV2 [22,23] that promotes the expression of soluble protease; pMal-TEV2 is practically identical to pTEV, except that the protease is expressed as a transient fusion to the solubility enhancing MBP-moiety [15]. Rosetta (DE3)pLysS cells expressing the different TEVp variants as N-terminal fusions to GFP (Rosetta (DE3)pLysS/pTEVwt-GFP, Rosetta (DE3)pLysS/pTEV2m-GFP, Rosetta (DE3)pLysS/pTEV3m-GFP, and Rosetta (DE3)pLysS/pTEV5m-GFP) were cultured and analyzed in a similar fashion. However, the culture media was now supplemented with 50 µg/ml kanamycin and expression of the protease-GFP fusions were induced by addition of IPTG at a final concentration of 1 mM. 3 hours later, the cultures were analyzed on a flow cytometer.

4.4. Protein expression, purification and fractionation

E. coli Rosetta (DE3)pLysS cells harboring the TEVp production vectors (either of pTEVprod, pTEV2mprod, pTEV3mprod, pTEV5mprod, pTEVprodΔMBP pTEV2mprodΔMBP, pTEV3mprodΔMBP, or pTEV5mprodΔMBP) were cultured overnight in a rotary shaker at 30°C, 150 rpm in 500 ml tryptic soy broth (TSB) supplemented with 50 µg/ml kanamycin and 20 µg/ml chloramphenicol. Five milliliters of the overnight culture was re-inoculated in 500 ml fresh TSB medium containing the same antibiotics and left to grow at 30°C, 150 rpm until OD600 reached 0.5, at which time the protease expression was induced through addition of IPTG (0.5 mM final concentration). The cultures were left to grow for another four hours at 30°C, 150 rpm before the cells were harvested by centrifugation (2,500xg, 10 min) and then resuspended in 60 ml TALON buffer (50 mM NaH2PO4, 300 NaCl, pH 7.5). The cells were subjected to sonication (4 min, 40% effect) to release the intracellular protein content and the cell debris was removed by centrifugation (30 min, 40,000xg). The supernatant was loaded on a 2 ml TALON metal affinity column (Clontech) and treated according to the manufacturer's recommendations. For elution of the immobilized TEVp, 5x1ml 1.5 M imidazole (pH 7) was added to the column. The eluted fractions, containing the target protein, were pooled,

and the buffer exchanged to TEVp reaction buffer (100 mM Tris-HCl (pH 8.0), 25 mM NaCl, 0.5 mM EDTA, 1 mM DTT) using a PD-10 desalting column (Amersham). The protease concentration was determined spectrophotometrically and with a BCA Protein Assay Kit (Thermo Scientific Pierce). The quality of the purified TEVp variants was analyzed by sodium dodecyl sulphate-polyacrylamide gel electrophoresis (SDS-PAGE) and subsequent staining with GelCode Blue Stain Reagent (Pierce). The purified protease variants were eventually used in an in vitro cleavage assay to analyze the conversion of a recombinant fusion substrate, ABP-subG-ZZ (where subG is ENLYFQG).

The distribution between the soluble and insoluble fraction of TEV protease upon overexpression in E. coli was also analyzed. Identical amount of cells (corresponding to 5 OD-equivalents) from each culture were treated with BugBuster Protein Extraction Reagent (Merck) according to the manufacturer's protocol. The insoluble fraction was finally resuspended in TEVp reaction buffer (100 mM Tris-HCl (pH 8.0), 25 mM NaCl, 0.5 mM EDTA, 1 mM DTT) using a volume corresponding to that of the soluble fraction.

Expression of the fusion protein, ABP-subG-ZZ, containing a TEVp specific substrate peptide between the albumin binding protein domain, ABP, and the protein A-derivative, ZZ, was performed essentially as described for the protease production. However, the harvested cells (E. coli Rosetta (DE3)pLysS/pABP-PS-ZZ [23] where PS is subG, i.e., ENLYFQG) were instead resuspended in 20 ml denaturing lysis buffer (7 M guanidiumchloride, 47 mM Na2HPO4, 2.67 mM NaH2PO4, 10 mM Tris-HCl, 100 mM NaCl, 20 mM β-mercaptoethanol, pH 8.0) and incubated for 2h at 37°C. The cell lysate was centrifuged at 35,300xg for 30 minutes and the supernatant loaded on an ASPEC XL4 (Gilson) automated protein purification system equipped with columns filled with 1 mL Talon metal affinity chromatography resin (Clontech), and purified according to the protocol described by Steen et al. [38]. The buffer of the purified protein fractions was exchanged into TEVp reaction buffer by using PD-10 desalting columns.

4.5. In vitro cleavage of a fusion protein substrate

Enzymatic cleavage reactions, containing 10 µg recombinant ABP-subG-ZZ fusion protein and 0.3 µg protease (TEVpwt, TEVp2m, TEVp3m, or TEVp5m) in 30 µl TEVp reaction buffer (100 mM Tris-HCl (pH 8.0), 25 mM NaCl, 0.5 mM EDTA, 1 mM DTT), were incubated at 37°C for different incubation times (0, 5 min, 15 min, 30 min, 45 min, 60 min, 90 min, 2h, 3h, 4h, 5h and 6h), and then terminated through addition of 3.3 µl 1% SDS. 10 µl 3xSDS denaturing buffer (150 mM TRIS, 300 mM DTT, 6% SDS, 0.3 % bromophenol blue and 30% glycerol) was then added to 20 µl of the stopped reactions and the samples were heat denatured at 96°C for 7 min before being loaded on an SDS-PAGE gel (Novex 4-12 % Tris-glycine gradient gel, Invitrogen). The separated protein fragments were stained in the gel using GelCode Blue Stain Reagent (Pierce). All experiments were executed in three independent replicates.

Acknowledgements

This work was supported by grant 621-2004-4647 from the Swedish Research Council.

References

[1] Georgiou G, Valax P (1996) Expression of correctly folded proteins in Escherichia coli. Current opinion in biotechnology 7: 190-197.

[2] Waugh DS (2005) Making the most of affinity tags. Trends in biotechnology 23: 316-320.

[3] Nilsson J, Stahl S, Lundeberg J, Uhlen M, Nygren PA (1997) Affinity fusion strategies for detection, purification, and immobilization of recombinant proteins. Protein expression and purification 11: 1-16.

[4] Chant A, Kraemer-Pecore CM, Watkin R, Kneale GG (2005) Attachment of a histidine tag to the minimal zinc finger protein of the Aspergillus nidulans gene regulatory protein AreA causes a conformational change at the DNA-binding site. Protein expression and purification 39: 152-159.
[5] Smyth DR, Mrozkiewicz MK, McGrath WJ, Listwan P, Kobe B (2003) Crystal structures of fusion proteins with large-affinity tags. Protein science: a publication of the Protein Society 12: 1313-1322.
[6] Nallamsetty S, Kapust RB, Tozser J, Cherry S, Tropea JE, et al. (2004) Efficient site-specific processing of fusion proteins by tobacco vein mottling virus protease in vivo and in vitro. Protein expression and purification 38: 108-115.
[7] Tözsér J, Tropea JE, Cherry S, Bagossi P, Copeland TD, et al. (2005) Comparison of the substrate specificity of two potyvirus proteases. FEBS J 272: 514-523.
[8] Cordingley MG, Callahan PL, Sardana VV, Garsky VM, Colonno RJ (1990) Substrate requirements of human rhinovirus 3C protease for peptide cleavage in vitro. The Journal of Biological Chemistry 265: 9062-9065.
[9] Dougherty WG, Cary SM, Parks TD (1989) Molecular genetic analysis of a plant virus polyprotein cleavage site: a model. Virology 171: 356-364.
[10] Lundbäck A-K, van den Berg S, Hebert H, Berglund H, Eshaghi S (2008) Exploring the activity of tobacco etch virus protease in detergent solutions. Analytical Biochemistry 382: 69-71.
[11] Dougherty WG, Parks TD, Cary SM, Bazan JF, Fletterick RJ (1989) Characterization of the catalytic residues of the tobacco etch virus 49-kDa proteinase. Virology 172: 302-310.
[12] Kapust RB, Tözsér J, Fox JD, Anderson DE, Cherry S, et al. (2001) Tobacco etch virus protease: mechanism of autolysis and rational design of stable mutants with wild-type catalytic proficiency. Protein Eng 14: 993-1000.
[13] Kapust RB, Routzahn KM, Waugh DS (2002) Processive degradation of nascent polypeptides, triggered by tandem AGA codons, limits the accumulation of recombinant tobacco etch virus protease in Escherichia coli BL21(DE3). Protein Expr Purif 24: 61-70.
[14] Fang L, Jia K-Z, Tang Y-L, Ma D-Y, Yu M, et al. (2007) An improved strategy for high-level production of TEV protease in Escherichia coli and its purification and characterization. Protein Expr Purif 51: 102-109.
[15] Kapust RB, Waugh DS (1999) Escherichia coli maltose-binding protein is uncommonly effective at promoting the solubility of polypeptides to which it is fused. Protein Sci 8: 1668-1674.
[16] Miladi B, Bouallagui H, Dridi C, El Marjou A, Boeuf G, et al. (2011) A new tagged-TEV protease: construction, optimisation of production, purification and test activity. Protein expression and purification 75: 75-82.
[17] Lucast LJ, Batey RT, Doudna JA (2001) Large-scale purification of a stable form of recombinant tobacco etch virus protease. BioTechniques 30: 544-546, 548, 550 passim.
[18] van den Berg S, Löfdahl P-A, Härd T, Berglund H (2006) Improved solubility of TEV protease by directed evolution. J Biotechnol 121: 291-298.
[19] Parks TD, Howard ED, Wolpert TJ, Arp DJ, Dougherty WG (1995) Expression and purification of a recombinant tobacco etch virus NIa proteinase: biochemical analyses of the full-length and a naturally occurring truncated proteinase form. Virology 210: 194-201.
[20] Cabrita LD, Gilis D, Robertson AL, Dehouck Y, Rooman M, et al. (2007) Enhancing the stability and solubility of TEV protease using in silico design. Protein Science 16: 2360-2367.
[21] Phan J, Zdanov A, Evdokimov AG, Tropea JE, Peters HK, et al. (2002) Structural basis for the substrate specificity of tobacco etch virus protease. J Biol Chem 277: 50564-50572.
[22] Kostallas G, Samuelson P (2010) Novel fluorescence-assisted whole-cell assay for engineering and characterization of proteases and their substrates. Appl Environ Microbiol 76: 7500-7508.
[23] Kostallas G, Löfdahl P-Å, Samuelson P (2011) Substrate profiling of tobacco etch virus protease using a novel fluorescence-assisted whole-cell assay. PLoS ONE 6: e16136.
[24] Waldo GS (2003) Genetic screens and directed evolution for protein solubility. Current Opinion in Chemical Biology 7: 33-38.
[25] Karzai AW, Roche ED, Sauer RT (2000) The SsrA-SmpB system for protein tagging, directed degradation and ribosome rescue. Nat Struct Biol 7: 449-455.
[26] DeLisa MP, Samuelson P, Palmer T, Georgiou G (2002) Genetic analysis of the twin arginine translocator secretion pathway in bacteria. J Biol Chem 277: 29825-29831.
[27] Kapust RB, Tözsér J, Copeland TD, Waugh DS (2002) The P1' specificity of tobacco etch virus protease. Biochem Biophys Res Commun 294: 949-955.
[28] Griswold KE, Mahmood NA, Iverson BL, Georgiou G (2003) Effects of codon usage versus putative 5'-mRNA structure on the expression of Fusarium solani cutinase in the Escherichia coli cytoplasm. Protein expression and purification 27: 134-142.
[29] Welch M, Villalobos A, Gustafsson C, Minshull J (2009) You're one in a googol: optimizing genes for protein expression. Journal of the Royal Society, Interface / the Royal Society 6 Suppl 4: S467-476.

[30] Gustafsson C, Govindarajan S, Minshull J (2004) Codon bias and heterologous protein expression. Trends in biotechnology 22: 346-353.
[31] Tsai CJ, Sauna ZE, Kimchi-Sarfaty C, Ambudkar SV, Gottesman MM, et al. (2008) Synonymous mutations and ribosome stalling can lead to altered folding pathways and distinct minima. Journal of molecular biology 383: 281-291.
[32] Kimchi-Sarfaty C, Oh JM, Kim IW, Sauna ZE, Calcagno AM, et al. (2007) A "silent" polymorphism in the MDR1 gene changes substrate specificity. Science 315: 525-528.
[33] Blommel PG, Fox BG (2007) A combined approach to improving large-scale production of tobacco etch virus protease. Protein expression and purification 55: 53-68.
[34] Rüther U (1982) pUR 250 allows rapid chemical sequencing of both DNA strands of its inserts. Nucleic Acids Res 10: 5765-5772.
[35] Hedhammar M, Stenvall M, Lonneborg R, Nord O, Sjolin O, et al. (2005) A novel flow cytometry-based method for analysis of expression levels in Escherichia coli, giving information about precipitated and soluble protein. Journal of biotechnology 119: 133-146.
[36] Pédelacq J-D, Piltch E, Liong EC, Berendzen J, Kim C-Y, et al. (2002) Engineering soluble proteins for structural genomics. Nat Biotechnol 20: 927-932.
[37] Waldo GS, Standish BM, Berendzen J, Terwilliger TC (1999) Rapid protein-folding assay using green fluorescent protein. Nat Biotechnol 17: 691-695.
[38] Steen J, Uhlén M, Hober S, Ottosson J (2006) High-throughput protein purification using an automated set-up for high-yield affinity chromatography. Protein Expr Purif 46: 173-178.

5. Figures and tables

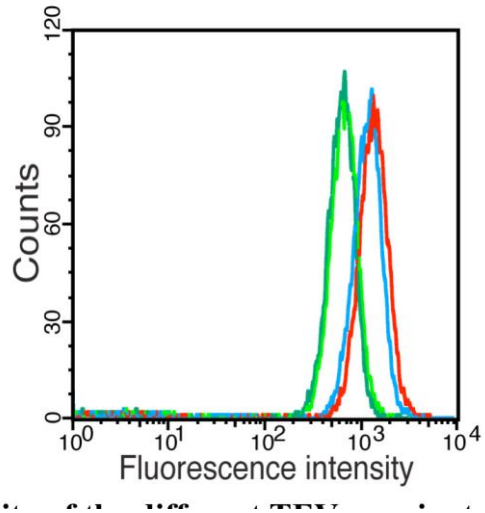


Figure 1. In vivo solubility of the different TEVp-variants as GFP-fusions. Each histogram corresponds to a pure culture of Rosetta DE3(pLys) cells expressing TEVpwt-GFP (light green) or TEVp2m-GFP (red) or TEVp3m-GFP (dark green) or TEVp5m-GFP) (blue), 3h after induction (1 mM IPTG) of the T7-promoter.

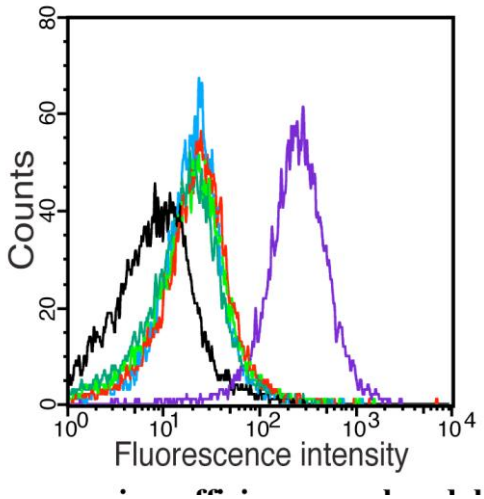


Figure 2. Substrate processing efficiency analyzed by a whole-cell assay for proteolysis. Flow cytometry analysis of E. coli DH5α cells that coexpress the proteolysis reporter GFP-subG-ssrANY and TEVpwt (light green) or TEVp2m (red) or TEVp3m (dark green) or TEVp5m (blue). As a positive and negative control, cells expressed MBP-subG-TEVpwt together with GFP-subG-ssrANY (purple) and GFP-ssrANY (black), respectively. The cultures were analyzed 2.5h after induction (0.1 mM IPTG, 0.2% arabinose).

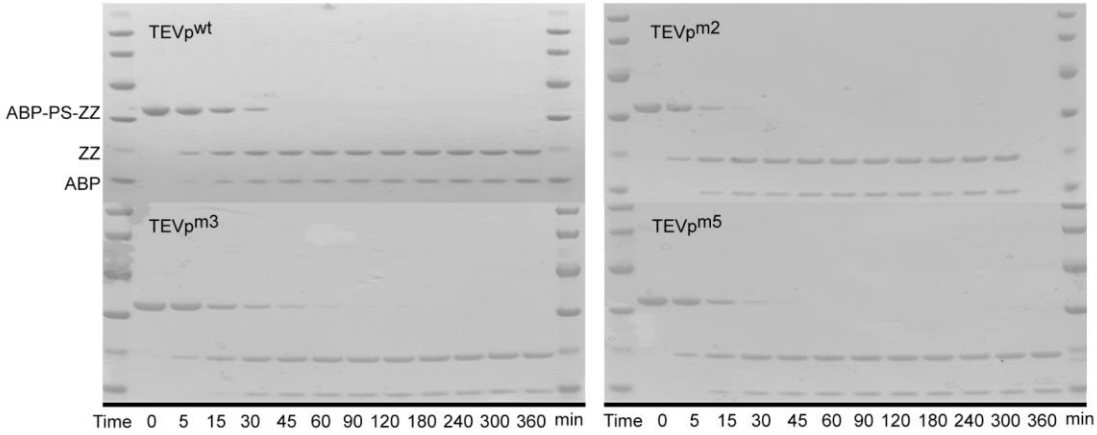


Figure 3. Substrate conversion in vitro of ABP-subG-ZZ fusion protein by the different recombinantly produced TEVp-variants. Each TEVp-variant (0.3 µg) was incubated with ABP-subG-ZZ (10 µg) in TEVp reaction buffer at 37°C for different incubation times (0, 5 min, 15 min, 30 min, 45 min, 60 min, 90 min, 2h, 3h, 4h, 5h and 6h). At each time-point, a small sample was withdrawn from each reaction and analyzed by SDS-PAGE.

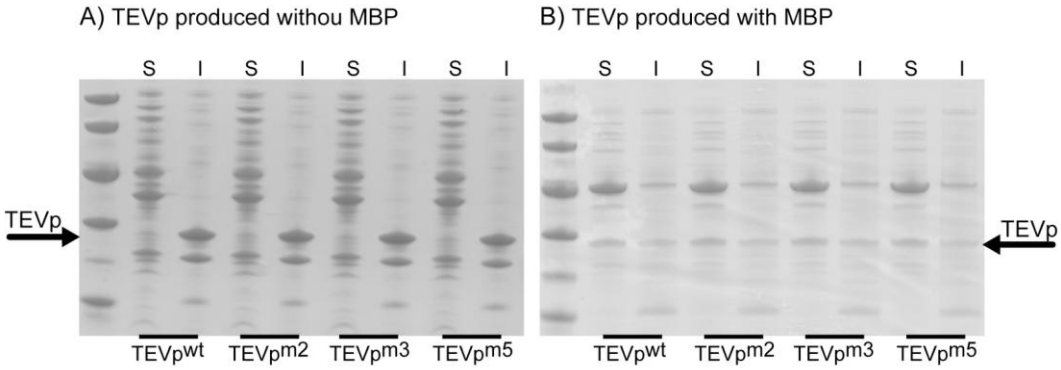
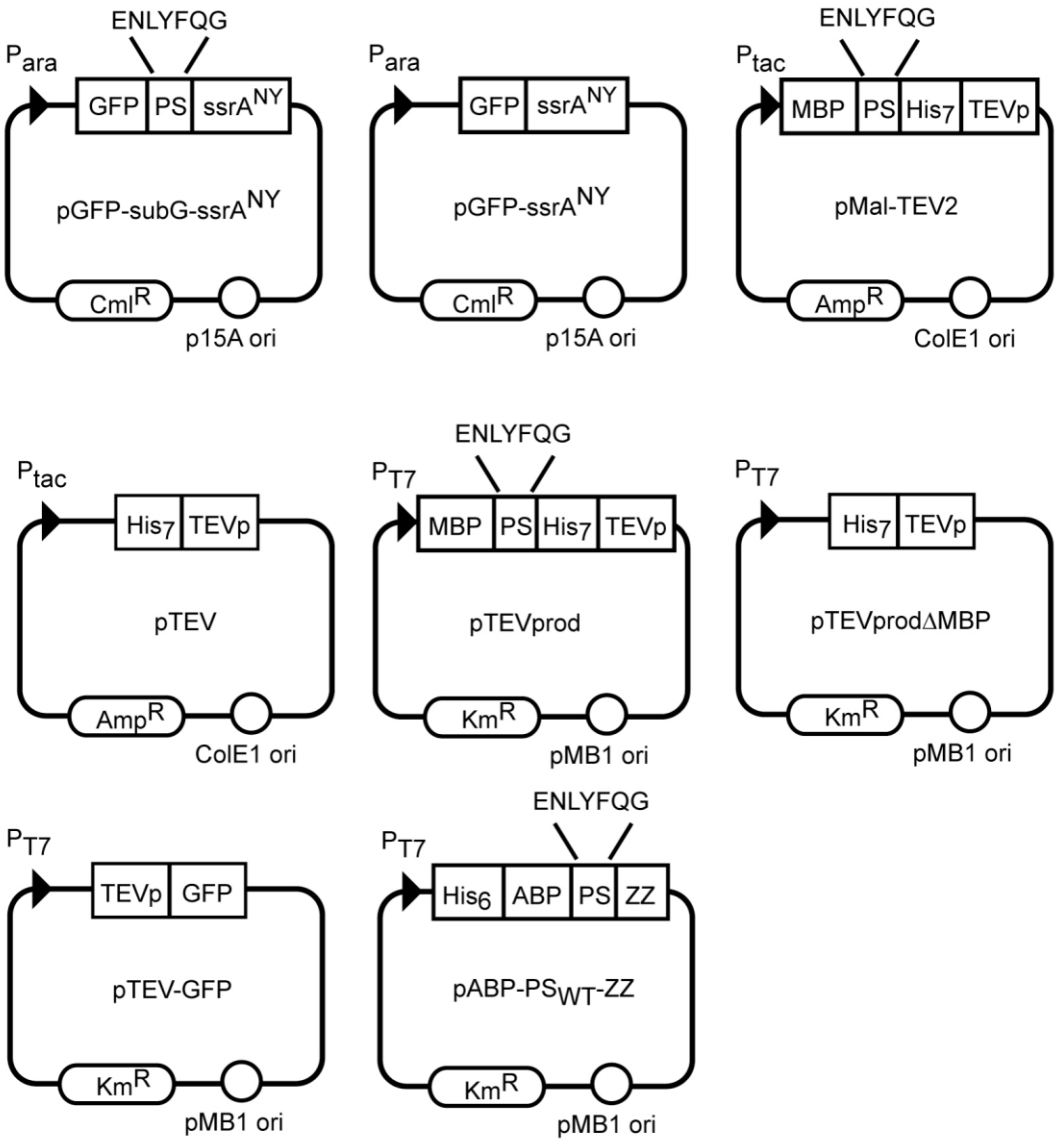


Figure 4. Expression analysis of the different TEVp variants produced without or with a solubility increasing MBP-moiety. SDS-PAGE analysis showing the soluble (S) and insoluble protein fractions (I) from whole-cell lysates of Rosetta (DE3)pLys cells overexpressing the different TEVp variants (as indicated in the figure) from a T7-promoter at 25°C. (A) The proteases were expressed without MBP-fusion. (B) The proteases were expressed with a transiently attached MBP-fusion.



Supplementary Figure 1. Plasmids used in this study

Supplementary Table 1. Oligonucleotides used in this study

Name	Sequence [5' to 3']
T17Sfor	GTGATTACAACCCGATATCGAGCTCCATTTGTCATTTGACGAATG
T17Srev	CATTCGTCAAATGACAAATGGAGCTCGATATCGGGTTGTAATCAC
L56Vfor	CGCCGCAATAATGGAACACTGGTGGTCCAATCACTACATGGTGTA
L56Vrev	TACACCATGTAGTGATTGGACCACCAGTGTTCCATTATTGCGGCG
N68Dfor	CATGGTGTATTCAAGGTCAAGGACACCACGACTTTGCAACAACAC
N68Drev	GTGTTGTTGCAAAGTCGTGGTGTCCTTGACCTTGAATACACCATG
I77Vfor	CTTTGCAACAACACCTC GTT GATGGGCGCGACATG
I77Vrev	CATGTCGCGCCCATCAA CGA GGTGTTGTTGCAAAG
S135Gfor	ACTAGTTGCACATTCCCTTCAGGCGATGGCATATTCTGGAAGCAT
S135Grev	ATGCTTCCAGAATATGCCATCGCCTGAAGGGAATGTGCAACTAGT
GEKO32	ACTGTACCATGGTGGGTACCCATCATCATCATC
GEKO33	ACTGTAGGCGCGCCGCGACGGCGACGACGATTCA
GEKO51	ACTGTACTCGAGTTATTAGCGACGGCGACGACGAT