Surface supercharged human enteropeptidase light chain shows improved solubility and refolding yield

Peter Simeonov, Renate Berger-Hoffmann, Ralf Hoffmann, Norbert Sträter and Thole Zuchner

Ultrasensitive Protein Detection Unit, Institute of Bioanalytical Chemistry, Center for Biotechnology and Biomedicine, University of Leipzig, Deutscher Platz 5, 04103 Leipzig, Germany

1To whom correspondence should be addressed. E-mail: Thole.Zuechner@bbz.uni-leipzig.de

Received May 12, 2010; revised October 14, 2010; accepted October 19, 2010

Edited by Arne Skerra

Enteropeptidase is a serine protease used in different biotechnological applications. For many applications the smaller light chain can be used to avoid the expression of the rather large holoenzyme. Recombinant human enteropeptidase light chain (hEPL) shows high activity but low solubility and refolding yields, currently limiting its use in biotechnological applications. Here we describe several protein modifications that lead to improved solubility and refolding yield of human hEPL whilst retaining the enzyme activity. Specifically, protein surface supercharging (N6D, G21D, G22D, N141D, K209E) of the protein increased the solubility more than 100-fold. Replacement of a free cysteine residue with serine (C112S) improved the refolding yield by 50%. The heat stability of this C112S variant was also significantly improved by supercharging. This study shows that even mild protein surface supercharging can have pronounced effects on protein solubility and stability.

Keywords: enteropeptidase/human enteropeptidase/protein solubility/refolding yield/surface supercharging

Introduction

Enteropeptidase (EC 3.4.21.9) is an enzyme of the digestive tract. It is a highly specific membrane-bound serine protease of the small intestine, which cleaves its native substrate trypsinogen at the specific recognition site (Asp)114-Lys. Trypsinogen is activated by this cleavage of the N-terminus of the protein. The resulting trypsin itself activates other pancreatin-like motifs and up to 10 intramolecular disulfide bridges, is responsible for specific macromolecular substrate recognition (Lu et al., 1997; Mikhailova et al., 2007). The light chain is connected to the heavy chain via one disulfide bridge and contains the classical catalytic triad (His57, Asp102 and Ser195 in chymotrypsin numbering) with up to four intramolecular disulfide bridges. In humans, different mutations of the proenteropeptidase gene are directly related to congenital enteropeptidase deficiency (Hadorn et al., 1969; Haworth et al., 1971)

Because of its ability to cleave fusion proteins over a wide pH and temperature range as well as its tolerance towards various detergents and denaturants, the highly site-specific enteropeptidase represents a very attractive tool in biotechnology. The small and catalytically active light chain offers especially high potential for biotechnological applications, and several attempts to understand and improve the functionality of this protease have been made (Lu et al., 1999; Liew et al., 2007; Shahravan et al., 2008). Most investigations have focused on the bovine light chain enteropeptidase due to its high availability. Up to now a number of expression and purification methods have been described for the recombinant soluble production of the catalytic subunit in Escherichia coli (Yuan and Hua, 2002; Huang et al., 2004; Tan et al., 2007), Pichia pastoris (Fang et al., 2004; Peng et al., 2004), Aspergillus niger (Svetcina et al., 2000), Saccharomyces cerevisiae (Choi et al., 2001) and CHO cells (LaVallie et al., 1993). Furthermore, Lu et al. (1997) have determined the crystal structure of a bovine light chain complex with a trypsinogen activation peptide analogue at a resolution of 2.3 Å. Human enteropeptidase has been recombinantly expressed in E.coli by Gasparian et al. (2003), although this resulted in insoluble aggregates and no crystal structure has yet been reported. The subsequent refolding via 6 M guanidinium chloride resulted in a total refolding yield of 2% after two cycles of renaturation.

One strategy to increase the refolding yield is to remove unpaired cysteines, which can disturb the refolding process or interact unspecifically in secondary reactions. Ivavenkov et al. (2003) demonstrated that the replacement of a free cysteine by serine or the protein ectonucleoside triphosphate diphosphohydrolase 6 (NTPDase 6) reduces the amount of mispaired disulfide bonds resulting in increased refolding efficiencies as well as increased protein stability. Furthermore, it was shown that the replacement of cysteine did not influence the catalytic activity of the purified protein.

Until today, most recombinant enteropeptidases studied were the bovine or porcine enzymes. Soluble expression was described in several studies (Huang et al., 2004; Peng et al.,...
expressions as well as from other species (Gasparian et al., 2006). In conclusion, expression of hEPL seems desirable in terms of achieving higher activities.

For various biotechnological applications, it is necessary to couple, bind or attach the enzyme of choice to another protein or immobilise it on a surface. Usually these chemical reactions require low salt conditions and preferably high protein concentrations to be effective. However, several enzymes, including enteropeptidase, show low to very low solubility under low salt concentrations or non-physiological pH. Lawrence et al. (2007) introduced the concept of ‘supercharging’ in order to overcome such problems. This method increases the net charge of a protein by replacing polar and non-polar solvent-exposed residues with charged residues to significantly decrease aggregation properties while the thermodynamic stability and activity is largely unaffected. In this study, monomeric and multimeric proteins of varying structures and functions were tested, including green fluorescence protein, glutathione-S-transferase and streptavidin variants. ‘Supercharging’ may therefore represent a useful and generally applicable approach to reduce the aggrega-
tion tendency and improve the solubility of some proteins.

For mutagenesis the Qiagen quick change method was used with heating for 5 min at 95 °C, followed by 20 cycles of 95 °C for 30 s, 54 °C for 30 s, 72 °C for 7 min and an additional extension time of 5 min at 72 °C. The hEPL gene was mutated to create the amino acid substitutions N6D(forward primer 5′-atgtaggaagtaatgcaagagggcc-3′ and reverse primer 5′-gcctgcttacagcataccaaat-3′), G21D and G22D (forward primer 5′-gtgggtggtttatagatgacagcagctgtcggc-3′ and reverse primer 5′-gcctgacagacagcc-3′) with two isolated Asp 4-Lys cleavage sites in between. The two cleavage sites result in the cleavage of the Trx-tag on the one hand and ensure the correct formation of the native N-terminus of the hEPL on the other hand. After transforma-
tion in Dh5α and subsequent plasmid preparation the vector pET32-hEPL was confirmed by DNA sequencing with T7 promoter and T7 terminator primer.

Expression, renaturation and purification of hEPL in E.coli

The thioredoxin-fusion constructs containing wild-type and hEPL mutants were transformed into the expression strain E.coli BL21 (DE3) codon plus RIL by heat shock and grown on LB agar plates supplemented with 100 µg/ml ampicillin overnight. One colony from the plate was incubated in 5 ml LB ampicillin (37 °C, 5 g) for a minimum of 12 h. The pre-
culture was then diluted 50-fold into 500 ml TB medium containing 100 µg/ml ampicillin. The culture was allowed to grow with shaking (5 g) at 37 °C for 2.5–3 h. Upon reaching a cell density corresponding to OD<sub>600</sub>=0.6, IPTG was added to a final concentration of 0.1 mM and cultivation was continued (14 h, 25 °C). The cells were collected by centrifuga-
tion (10 min, 5000g) and afterwards the formed insoluble aggregates were purified by resuspending the cell pellet in lysis buffer [100 mM Tris–HCl (pH 7.0), 1 mM EDTA].

Cells were lysed by lysozyme (0.1 mg/ml) and using a French press. Between each of the following two washing steps with washing buffer I [20 mM EDTA (pH 8.0), 500 mM NaCl and 2% Triton X-100] and washing buffer II (20 mM EDTA, 100 mM Tris–HCl (pH 7.0)] the inclusion bodies were sedimented by centrifugation (20 000g, 4 °C, 30 min). The subsequent refolding was performed as follows.
After solubilising the aggregated thioredoxin-fusion protein in 40 ml per g cell weight of 6 M guanidine-HCl, 100 mM Tris–HCl (pH 8.0). Trx-hEPI was reduced with 100 mM DTT (60 °C, 20 min). The solution was dialyzed against 3 M guanidine, 5 mM HCl and then centrifuged (48 000 g, 1 h). An equivalent volume of oxidation buffer (6 M guanidine-HCl, 100 mM oxidised glutathione and 100 mM Tris–HCl pH 8.0) was added to the supernatant to allow mixed disulfide formation overnight. The buffer was changed to 3 M guanidine-HCl (pH 8.0). The refolding was done by fast dilution of the protein in 0.7 M arginine-HCl (pH 8.5), 15% (v/v) glycerol, 3 mM reduced glutathione and 1 mM EDTA. After 72 h at 4 °C the refolding solution was dialyzed for 8 h against 50 mM Tris–HCl (pH 8.0) to facilitate complete autocatalytic activation by cleaving the fusion tag. The active hEPI was purified by affinity chromatography on 4 ml STI agarose. The column was equilibrated with 50 mM Tris–HCl (pH 8.0), 500 mM NaCl and the sample was loaded at a flow rate of 1 ml/min. The column was washed with 50 mM Tris–HCl (pH 8.0), 500 mM NaCl (10 column volumes) and bound active hEPI was eluted with 50 mM glycine-HCl (pH 3.0) in 1 ml fractions which were immediately neutralised [25 μl 2 M Tris–HCl (pH 8.0)].

The purified hEPI variants were analysed by SDS-PAGE and the bands were analysed by MALDI-TOF/TOF-MS after in-gel tryptic digestion (sequence coverage of ≈56%). Protein concentrations were determined spectrophotometrically at 280 nm using an extinction coefficient of 55 390 M⁻¹ cm⁻¹ for the product 3-carboxy-4-nitrophenoxide (n = 3).

Enzyme kinetics

The specific activities of the recombinant hEPL variants were obtained by a slightly modified literature procedure according to Green and Shaw (1979). Assays were performed in 750 μl of 100 mM Tris–HCl (pH 8.0), 260 μM DTNB and 20 μM to 400 μM Z-Lys-SBzI at room temperature. The reaction was started by adding hEPL variants (30–80 ng enzyme). The reaction was monitored continuously for 3 min by the increasing absorbance at 412 nm using an extinction coefficient of 136 000 M⁻¹ cm⁻¹ for the product 3-carboxy-4-nitrophenoxide.

Solubility testing and heat stability

For solubility tests, all hEPI variants were tested under exactly identical buffer conditions to avoid influences of different salt concentrations. In detail, all four hEPL variants were directly eluted from the columns into 50 mM Tris/50 mM glycine, pH 8.0 buffer. Protein concentrations were determined by absorption at 280 nm and resulted in values from 450 to 650 μg/ml. After 2 h, samples were centrifuged for 1 min at 16 000 g. The soluble protein concentration of the resulting supernatant was determined by spectrophotometric measurement at 280 nm. The samples of hEPL sc and hEPL sc C112S were lyophilised in a Speed-Vac centrifugation system (Concentrator 5301, Eppendorf) and subsequently dialysed against 50 mM Tris/50 mM glycine (pH 8.0). Finally, samples were centrifuged (1 min, 16 000 g) and the soluble protein concentrations of the resulting supernatants were determined by spectrophotometric measurement at 280 nm. Even at the highest concentrations tested with the available amount of protein, no precipitation was observed. Therefore, these data are the lower limit of the solubility of these variants.

For heat stability experiments, all four hEPL variants were diluted to a protein concentration of 140 μg/ml in 50 mM Tris/50 mM glycine (pH 8.0). Sample volumes were 10 μl per sample. Samples were heated for 90 s at either 37 °C, 40 °C, 45 °C, 50 °C, 55 °C, 60 °C, 65 °C, 70 °C or 100 °C. The negative control contained 4% Tween 20 to ensure complete inactivation. Afterwards, samples were incubated at room temperature for 2 h. Enzyme activities were measured as described above.

SDS-PAGE

The protein content of samples was measured as described by Lowry et al. (1951). Proteins (10 μg) were separated by sodium dodecyl sulfate–polyacrylamide gel electrophoresis mini-gels (15%) (Miniprotein; Bio-Rad Laboratories, Hercules, CA, USA). Gels were stained with Coomassie brilliant blue.

Isoelectric focusing

Isoelectric focusing (IEF) was performed on a Phast System apparatus from GE Healthcare (Uppsala, Sweden) using a precast IEF gel (pH 3–9) and standard protocols. Gels were silver stained.

Protein digests

To test for a loss of specificity in the different hEPL variants, actin, bovine serum albumin, casein and ovalbumin were digested in 50 mM Tris–HCl (pH 8.0) at 37 °C. hEPL variants in a molar ratio of 1–100 (hEPL to protein substrates) were used for digests. The reaction was stopped by adding SDS sample buffer and immediate heating to 95°C for 10 min. Samples were analysed via standard SDS-PAGE and the optical densities of the bands were measured using the ImageJ software.

Software

The following programs were used: T-Coffee (Notredame et al., 2000) and Bioedit ( www.mbio.ncsu.edu/BioEdit) for the sequence alignments of the different enteropeptidase catalytic light chains, PyMOL ( www.pymol.org) with APBS Tools ( http://pymolwiki.org/index.php/APBS) for the visualisation of the protein structure and the surface potential, OriginPro 8G (Originlab, Northhampton, USA) for kinetic parameters and ProtParam from expasy tools (Gasteiger et al., 2005) for physicochemical parameter determination.

Results

For biotechnological applications of enzymes, high activities, high refolding yields after expression and purification as well as good stability and solubility are crucial. In the case of the enteropeptidase light chain, the human version has been demonstrated to show a 10-fold higher activity compared with bovine versions or other species (Gasparian et al., 2006). However, expression of the hEPL in sufficient amounts is difficult, mainly due to very low refolding yields (Gasparian et al., 2003, 2006). Furthermore, the solubility and stability of hEPL are suboptimal for biotechnological purposes. We investigated the effect of several modifications of hEPL, on its refolding yield, stability and solubility.
The design of these modifications is based on a sequence comparison of enteropeptidases from several species and the crystal structure of bovine EPL (1EKB) (Figs 1 and 2). As a first step, we aimed to increase the refolding yield by replacing the unpaired cysteine residue by a serine (C112S) (Fig. 1), thereby reducing the number of mispaired disulfide bridges and thus the proportion of misfolded hEPL. We also replaced five residues on the protein surface with negatively charged amino acids of the highest possible amino acid similarity to non-human sequences (Fig. 1: N6D, G21D, G22D, N141D, K209E). The increased net charge of the protein from -3 to -9 at neutral pH was performed to improve both the solubility and the stability. We refer to this process with the term ‘supercharging’ (Lawrence et al., 2007). To prevent an undesirable loss of enzyme activity or specificity by this supercharging, it was necessary to engineer the protein in a way that would minimise influences on the active centre of the protein. To this end, both the sequence comparison with enteropeptidase light chain variants from other species and molecular modelling calculations of the

---

**Fig. 1.** Sequence comparison of hEPL with mouse, rat, porcine and bovine enteropeptidase. Mutated positions of hEPL were located at N6D, G21D, G22D, N141D, K209E. Red: negatively charged amino acids, blue: positively charged amino acids, yellow: unpaired cysteine, green: catalytic triad.

**Fig. 2.** Models of the molecular surface of (A) wild-type hEPL and (B) surface supercharged hEPL. The surface is coloured according to the electrostatic potential (red: negative potential, blue: positive potential). The locations of the five mutations are marked with circles.
designed variants were performed beforehand. The sequence comparison showed 11 amino acid positions that are charged in other species but not in hEPL. Among these 11 amino acids, those 5 amino acids which showed the strongest conservation in other species in combination with the most optimal location on the protein surface were chosen for subsequent modification. The optimal location of these amino acids on the protein surface was judged by a structural model of hEPL based on the crystal structure of bovine enteropeptidase (Fig. 2). For modification, amino acids were chosen which are located far from the active centre of the enzyme and which are solvent accessible at the protein surface and not involved in specific interactions with other protein residues.

In the following, four variants of hEPL are described: wild-type (hEPL WT), Cys112Ser (hEPL C112S), supercharged hEPL (hEPL sc) and supercharged hEPL carrying the Cys112Ser mutation (hEPL sc C112S).

These constructs were cloned and expressed in E.coli BL21 (DE3). After refolding and purification, SDS-PAGE showed a single distinct band of the expected molecular weight for each of the four constructs (Fig. 3). Tryptic digests of the bands confirmed their identities in MALDI-TOF/TOF-MS (data not shown).

The refolding yields were significantly higher for the C112S mutation (Fig. 4). In detail, a total amount of 9 mg/l (refolding yield of 3%) could be achieved for the variants carrying the C112S mutation whereas the wild-type enzyme and the supercharged variants (without C112S) resulted in only 6 mg/l enzyme (refolding yield of 2%). Supercharging did not have an effect on the refolding yield.

Table I. Kinetic parameters observed for different hEPL variants

<table>
<thead>
<tr>
<th>Variant</th>
<th>$K_M$ (mM)</th>
<th>$k_{cat}$ (s$^{-1}$)</th>
<th>$k_{cat}/K_M$ (mM$^{-1}$ s$^{-1}$)</th>
</tr>
</thead>
<tbody>
<tr>
<td>hEPL wt</td>
<td>0.118 ± 0.014</td>
<td>111 ± 5</td>
<td>941</td>
</tr>
<tr>
<td>hEPL sc</td>
<td>0.083 ± 0.012</td>
<td>134 ± 5</td>
<td>1614</td>
</tr>
<tr>
<td>hEPL C112S</td>
<td>0.100 ± 0.016</td>
<td>121 ± 7</td>
<td>1210</td>
</tr>
<tr>
<td>hEPL sc C112S</td>
<td>0.128 ± 0.009</td>
<td>159 ± 4</td>
<td>1242</td>
</tr>
<tr>
<td>hEPL wt*</td>
<td>0.140 ± 0.007</td>
<td>133 ± 3</td>
<td>950</td>
</tr>
</tbody>
</table>

The $k_{cat}$ values for hEPL sc C112S was significantly higher than all other versions ($P < 0.05$, unpaired two-sided t-test). The abbreviations correspond to Fig. 3.

*Parameters as given by Gasparian et al., 2003.
100-fold compared with the non-supercharged variants hEPL and hEPL C112S. The highest solubility was achieved with hEPL sc C112S (>17.3 mg/ml), representing at least 119-fold increased solubility compared with hEPL wt. hEPL sc showed a solubility of >15.4 mg/ml. Due to technical limitations, the observed improved solubilities represent minimum estimates, the true values are likely to be higher than this.

Furthermore, we investigated the heat stability of the variants by subjecting them to various temperatures for 90 s and then incubating at room temperature for 2 h to allow sufficient time for refolding to occur. Subsequent measurement of the residual enzyme activity revealed significant differences (Fig. 6). The mutation with the most pronounced effect was C112S, which had a significantly reduced heat stability compared with the wild-type protein. After 90 s incubation at 55°C, hEPL C112S showed only 41% activity compared with the 37°C sample. At the same temperature, the activity of the supercharged version of the C112S variant (hEPL sc C112S) is reduced to only 93% compared with the value at 37°C. This is similar to the activity observed for the hEPL wt variant at this temperature. Thus, the dramatic decrease in heat stability caused by the C112S mutation could be recovered almost entirely by supercharging of the enzyme. Even at higher temperatures (70°C), both supercharged versions of hEPL still showed activities of 15 and 8% compared with only 1–2% for the non-supercharged versions of hEPL. Therefore, supercharging was accompanied by improved heat stabilities. This effect may be the result of an increased melting point of the variants and/or of an improved refolding behaviour of the heat-denatured enzyme. Finally, an IEF of the different hEPL variants was performed to verify the effect of supercharging on the isoelectric points of the enzymes. The pI values were as follows: hEPL wt = 5.7, hEPL C112S = 5.8; hEPL sc = 4.8; hEPL sc C112S = 4.9 (Fig. 7) and were in agreement with the expected changes.

**Discussion**

Enteropeptidase is an enzyme of biotechnological interest. It cleaves substrates with high specificity and can be expressed in different forms. Biotechnological applications include its use in ELISA detection systems (Lee et al., 2009) as well as in site-specific cleavage of recombinant fusion proteins (see e.g. Tenno et al., 2004; Poh et al., 2009). Especially for ELISA detection systems, the enteropeptidase light chain has advantages in its smaller size and easier expression. It has been demonstrated previously that the enteropeptidase light chain has an activity, which is similar to the holoenzyme and that the role of the heavy chain of the enzyme is mainly to increase specificity by a factor of around 5 (Liew et al., 2007; Mikahailova et al., 2007). However, the activities of various enteropeptidase light chain forms from different species are low and the expression of the human form of enteropeptidase light chain with its 10-fold higher activity is desirable. In this paper, we optimised hEPL for biotechnological applications in several ways: as a first step, the refolding yield was increased by 50% by replacing a free cysteine

**Fig. 6.** Heat stability of the hEPL variants. Enzymes were heated to the stated temperatures for 90 s and activity was measured after an additional 2 h. (A) Heat stability of hEPL sc as compared with hEPL wt was significantly improved at 65°C and 70°C (P < 0.05, two-sided unpaired t-test). (B) Heat stability of hEPL sc C112S as compared with hEPL C112S was strongly improved at 50–70°C (P < 0.05, two-sided unpaired t-test). The abbreviations correspond to Fig. 3.

**Fig. 7.** IEF gel of different human enteropeptidase light chain variants. The pI values were as follows: hEPL wt = 5.7; hEPL C112S = 5.8; hEPL sc = 4.8; hEPL sc C112S = 4.9 and are in good agreement with the expected changes due to the mutations. The abbreviations correspond to Fig. 3.
Protein engineering of human enteropeptidase light chain

In conclusion, we were able to improve the solubility and refolding yields of hEPL wild type. In hEPL C112S, supercharging in addition had a dramatic effect on the heat stability. To the best of our knowledge, this is the first study demonstrating that even modest protein surface supercharging can have dramatic effects on protein solubility and stability. The resulting bioengineered enzyme is now available for biotechnological applications such as ELISA detection systems. Protein solubility and aggregation is also a common obstacle in sample preparation for protein structure analysis by NMR as well as crystallography, for which high protein concentrations of 1 mM or 10 mg/ml, respectively, are usually required.

Supplementary data
Supplementary data are available at PEDS online.

Acknowledgements
We would like to thank Dr. Matthias Zebisch for helpful discussions, Dr. Bartholomeus Küttner for help with the modelling studies, Anita Bhandari for help with gel electrophoresis and Dr. Andrew Hagan for proofreading. We gratefully acknowledge the expert technical assistance of Susen Joswiak and Katrin Müller.

Funding
This work was supported by BMBF Innoprofile, Project Number 03IP604.

References