



Published in final edited form as:

Protein Expr Purif. 2006 July ; 48(1): 104–111.

Expression of human cationic trypsinogen with an authentic N terminus using intein-mediated splicing in aminopeptidase P (*pepP*) deficient *Escherichia coli*

Orsolya Király^{1,¶}, Lan Guan^{2,¶}, Edit Szepessy¹, Miklós Tóth³, Zoltán Kukor³, and Miklós Sahin-Tóth^{1,*}

¹Department of Molecular and Cell Biology, Goldman School of Dental Medicine, Boston University, Boston, MA, 02118;

²Department of Physiology, University of California Los Angeles, Los Angeles, CA 90095;

³Department of Medical Chemistry, Molecular Biology and Pathobiochemistry, Semmelweis University, Budapest, Hungary.

Abstract

High-level expression of human trypsinogens as inclusion bodies in *Escherichia coli* requires deletion of the secretory signal sequence and placement of an initiator methionine at the N terminus. Trypsinogen preparations obtained this way contain a mixture of abnormal N termini, as a result of processing by cytoplasmic aminopeptidases. Here we describe an expression system that produces recombinant human cationic trypsinogen with a native, intact N terminus, using intein-mediated protein splicing and an aminopeptidase P (*pepP*) deficient *Escherichia coli* strain. As a first application of this system, the effect of the pancreatitis-associated mutation A16V on the autoactivation of human cationic trypsinogen was characterized. The use of the novel *pepP* knock-out *Escherichia coli* strain should be generally applicable to the expression of recombinant proteins, which undergo unwanted N-terminal trimming by aminopeptidase P.

The human pancreas produces the digestive pro-enzyme trypsinogen in three isoforms. On the basis of their relative isoelectric points and electrophoretic mobility, these are commonly referred to as cationic trypsinogen, anionic trypsinogen, and mesotrypsinogen. The isoenzymes are encoded by separate genes, the *PRSS1* (protease, serine, 1), *PRSS2* and *PRSS3* genes (for a recent review see [1] and references therein). Cationic trypsinogen (*PRSS1*) and anionic trypsinogen (*PRSS2*) make up the bulk of secreted trypsinogens in the pancreatic juice, while mesotrypsinogen (*PRSS3*) accounts for 2–10 % [2–6]. Trypsinogens are synthesized as pre-pro-enzymes (pre-trypsinogens) with a signal peptide of 15 amino acids, followed by the 8 amino acid long pro-peptide, the trypsinogen activation peptide. The signal-peptide is removed upon entry into the endoplasmic reticulum lumen and the pro-enzymes are packaged into zymogen granules and eventually secreted into the pancreatic juice. Activation of trypsinogen to trypsin is achieved by proteolytic removal of the pro-peptide, which normally takes place in the duodenum catalyzed by enteropeptidase (enterokinase), a highly specialized serine protease in the brush-border membrane of enterocytes. Trypsin can also activate trypsinogen, a process termed autoactivation, which in the duodenum may have a physiological role in facilitating zymogen activation, whereas inappropriate autoactivation in the pancreas might

Mailing address: 715 Albany Street, Evans-433; Boston, MA 02118, Tel: (617) 414-1070; Fax: (617) 414-1041; E-mail: miklos@bu.edu.

¶The first two authors contributed equally to the project.

* corresponding author

This work was supported by NIH grant DK058088 to M. S.-T.

cause pancreatitis. Mutations in the *PRSS1* gene have been identified in patients with hereditary pancreatitis, familial or sporadic chronic pancreatitis [7–11]. To date, 22 *PRSS1* gene variants have been described, which affect 18 different amino-acid positions in trypsinogen, and result in 20 different amino-acid substitutions. When classified according to frequency of occurrence, R122H (~70 %), N29I (~25 %) and A16V (~4 %) are the three most prevalent mutations [8–11].

In order to elucidate how mutations in cationic trypsinogen cause pancreatitis, biochemical analysis of the mutant proteins is essential. Pancreatic tissue or juice is not readily available in significant amounts from patients carrying *PRSS1* mutations. In addition to limited availability, isolation of the heterozygously expressed mutant trypsinogen from the complex protein mixture poses another technical challenge. Therefore, heterologous recombinant expression in *E. coli* has been the method of choice to generate wild-type and mutant trypsinogen proteins for biochemical analysis. For reasons that are not apparent, cationic trypsinogen is not secreted to the periplasmic space of *E. coli* when expressed as a fusion with the alkaline-phosphatase signal-sequence, even though fusions with human anionic trypsinogen, mesotrypsinogen or rat anionic trypsinogen are secreted to low levels [12]. Similarly, expression of full length native pre-trypsinogen in *E. coli* results in no detectable zymogen secretion (M. S-T., unpublished observations). To obtain high levels of human cationic trypsinogen expression in *E. coli*, the signal-sequence is replaced with an initiator methionine residue and trypsinogen is expressed in the cytoplasm, where it accumulates in inclusion bodies, which are isolated and re-natured in vitro. Re-folding procedures for denatured trypsinogen have been developed and perfected from the 1970s and applied in the expression-purification protocols of bovine and human isoforms [12–22]. In our laboratory, re-folded trypsinogen is purified to homogeneity by affinity chromatography using immobilized ecotin [23]. Ecotin is a protease inhibitor with broad specificity, isolated from the periplasm of *E. coli*. Cationic trypsin preparations produced recombinantly or purified from pancreatic juice exhibit essentially identical catalytic activity, indicating that recombinant trypsinogen is a suitable experimental alternative to its native counterpart [18]. However, here we report that subtle structural differences do exist between recombinant and native trypsinogen preparations, because cytoplasmic aminopeptidases process the N terminus of the recombinant protein in *E. coli*. This raises the possibility that the biochemical properties of pancreatitis-associated mutations characterized in the context of recombinant trypsinogen preparations may not accurately reflect the in vivo situation. Furthermore, the N-terminal heterogeneity prevents the analysis of the biochemical effects of the A16V mutation, which alters the very N-terminal amino-acid of mature trypsinogen. Therefore, development of an expression system which allows high-level production of secretory trypsinogens with authentic N termini would be valuable. Here we describe that utilization of self-splicing mini-intein fusions combined with deletion of the *pepP* gene circumvents the deleterious effects of aminopeptidases and allows the expression of human cationic trypsinogen with an authentic, native N-terminal sequence. As an application, wild-type and A16V mutant proteins were produced, purified, and properties of autoactivation were characterized. The use of the *pepP* knock-out *E. coli* strain should be generally applicable to the expression of other proteins, which undergo unwanted N-terminal trimming by aminopeptidase P.

METHODS

Nomenclature

Numbering of amino-acids of human cationic trypsinogen follows the genetic convention; and starts with the initiator methionine (Met¹) of pre-trypsinogen. Because the secretory signal-peptide comprising the first 15 amino-acids is removed in the secretory pathway, the first amino-acid of mature trypsinogen is Ala¹⁶.

Plasmid construction

Construction of the pTrapT7-PRSS1 plasmid harboring the human cationic trypsinogen gene was described previously [12,17]. Note that the nucleotide sequence of this gene construct differs at several positions from the *PRSSI* cDNA sequence, but encodes the same protein (see ref. 12 for details). Intein fusion constructs were generated by overlap extension PCR. The ~460 nt DnaB mini-intein gene from the cyanobacterium *Synechocystis sp.* was amplified from the pTwin2 plasmid (New England Biolabs) using the sense “primer A” 5'-CGG GAG TCC ATG GCT ATC TCT GGC GAT AGT CTG ATC AGC-3' (the Nco I site within the primer sequence is underlined) and the antisense “primer B” 5'-GAA AGG AGC GTT GTG TAC AAT GAT GTC ATT CGC-3. A ~230 nt portion of the recombinant human cationic trypsinogen gene (corresponding to amino-acids Ala¹⁶ – Gln⁸⁶) was amplified from the pTrapT7-PRSS1 template with sense “primer C” 5'-ATT GTA CAC AAC GCT CCT TTC GAT GAT GAT GAC AAG-3' and antisense “primer D” 5'-CTG CTC ATT GCC CTC AAG GAC-3'. Primer D anneals downstream of an EcoR I site within the cationic trypsinogen sequence (Fig 1 and online supplementary material). The underlined sequences in primers B and C are overlapping and thus allow the PCR products obtained in the first round of amplifications to hybridize. PCR products from the two reactions were gel purified, mixed and extended using primers A and D. The resulting PCR product was digested with Nco I and EcoR I and ligated into the pTrapT7-PRSS1 plasmid digested with the same enzymes (Fig 1 and online supplementary material). The sequence of the fusion construct has been deposited with GenBank under accession number DQ371396. Mutation N29I was introduced into the intein-*PRSSI* fusion gene by PCR mutagenesis using the antisense mutagenic primer AAC AGA ATT CTC TTC GCA GAT GTA TCC TCC-3' and “primer A”. Mutation A16V was created by 2-step overlap extension PCR mutagenesis. Mutagenic (“internal”) primers were A16Vsense 5'-GTA CAC AAC GTT CCT TTC GAT-3; and A16Vantisense 5'-ATC GAA AGG AAC GTT GTG TAC-3'. External sense and antisense primers were “primer A” and “primer D”. The PCR products were digested with Nco I and EcoR I and ligated into the pTrapT7-intein-PRSS1 plasmid digested with the same enzymes.

Deletion of the *pepP* gene from *E. coli*

The *pepP* gene (GenBank accession number D00398) encoding aminopeptidase P was deleted from the *E. coli* chromosome using the method described by Datsenko and Wanner (2000) [24]. Strains and plasmids used were kind gifts from Barry L. Wanner. Briefly, we PCR-amplified the kanamycin resistance gene using primers with 44–45 nucleotide extensions homologous to sequences flanking the *pepP* gene (Fig 2). The template was plasmid pKD4. PCR products were treated with *Dpn* I to eliminate methylated template DNA. Electrocompetent *E. coli* BW25113 cells harboring the helper plasmid pKD46 encoding the λ phage Red recombinase were prepared and PCR products containing the kanamycin resistance gene flanked by two FLP recognition target (FRT) sites were introduced by electroporation. Transformants were selected by kanamycin resistance. Red recombinase catalyzes homologous recombination between the PCR product and the bacterial chromosome, resulting in the exchange of the *pepP* gene with the kanamycin resistance gene. The helper plasmid was eliminated by growing at 37 °C and the new strain was named LG-1. Plasmid pCP20, harboring the FLP recombinase gene, was introduced into kanamycin resistant LG-1 cells by electroporation. Transformants were selected by growing at 30 °C in the presence of ampicillin. FLP recombinase excises the *kan* gene between the two FRT sites, leaving one FRT site. Plasmid pCP20 was then eliminated by growing at 43 °C, and the resulting aminopeptidase P deficient *E. coli* strain was designated LG-2. Finally, plasmid pGP1-2, harboring the T7 RNA polymerase gene under the control of a temperature-inducible λ promoter [25], was introduced into LG-2 cells by electroporation. The resulting strain was designated LG-3 and was used for the expression of the intein-trypsinogen fusions.

Expression and purification of trypsinogens

Intein fusion constructs harboring wild type and mutant trypsinogens were transformed into the aminopeptidase P deficient *E. coli* strain LG-3. Transformants were grown in 200 mL culture volumes at 30 °C. When culture density reached OD 0.5 at 600 nm (OD₆₀₀), expression of trypsinogen was induced by shifting the incubation temperature to 42 °C for 30 minutes and adding *isopropyl* 1-thio β, D-galactopyranoside (IPTG) to a final concentration of 1 mM. The temperature shift induces the expression of T7 RNA polymerase encoded on plasmid pGP1-2 under the control of a temperature-sensitive λ repressor. IPTG was added because the pTrapT7 plasmid also contained a lac operator between the T7 promoter and the intein-trypsinogen fusion gene [12]. After induction, cells were grown for an additional 5 hours at 30 °C. Cells were harvested, resuspended in 0.1 M Tris-HCl (pH 8.0), 5 mM K-EDTA at 0.8 mL per 10 mL original culture volume, and inclusion bodies were isolated by sonication (3 × 20 sec, Heat Systems Ultrasonics cell disruptor, Model W-200R with a microtip probe, continuous mode, power-setting 4) and centrifugation (13,200 rpm, 5 min, Eppendorf microcentrifuge, 4 °C). The pellet was washed twice with 1 mL 0.1 M Tris-HCl (pH 8.0), 5 mM K-EDTA and dissolved in 500 μL 4 M guanidine-HCl, 0.1 M Tris-HCl (pH 8.0), 2 mM K-EDTA, 30 mM dithiothreitol (final concentrations) and incubated at 37 °C for 30 minutes to reduce trypsinogens. Denatured trypsinogens were rapidly diluted in 50 mL refolding buffer (0.9 M guanidine-HCl, 0.1 M Tris-HCl (pH 8.0), 2 mM K-EDTA, 1 mM L-cysteine, 1 mM L-cystine), stirred under argon for 5 minutes and incubated overnight at 4 °C. The solution was diluted with 50 mL 0.4 M NaCl, centrifuged for 15 minutes at 15,000 rpm and loaded onto a 2 mL ecotin affinity column [23]. The column was washed with 20 mM Tris-HCl (pH 8.0), 0.2 M NaCl and trypsinogens were eluted with 50 mM HCl. The ecotin-eluate contained about 90 % trypsinogen and 10 % uncleaved intein-trypsinogen fusion protein, which were separated in a final chromatographic step using a MonoS HR 5/5 strong cation-exchange column. Concentrations of the purified trypsinogen solutions were calculated from their ultraviolet absorbance at 280 nm, using a theoretical extinction coefficient of 36,160 M⁻¹ cm⁻¹. Trypsinogens eluted from the MonoS column were prone to spontaneous autoactivation, therefore, functional analysis was performed immediately after the chromatography.

Autoactivation of trypsinogens

MonoS-purified trypsinogens diluted to 2 μM final concentration were incubated at 37 °C in a final volume of 100 μL. For assays at pH 5.0, column fractions in 20 mM sodium acetate buffer (pH 5.0) were diluted with the same buffer. For assays at pH 8.0, column fractions were diluted with Tris-HCl buffer (pH 8.0) to a final buffer concentration of 0.1 M and 1 mM CaCl₂ (final concentration) was also included. Reactions also contained ~0.1 M NaCl carried over from the MonoS chromatography and 2 mg/mL bovine serum albumin as inert protein to prevent binding of trypsinogens to plastic surfaces. To initiate the autoactivation reaction, 10 nM trypsin was added (final concentration). At the indicated times, 2 μL aliquots were removed from the reaction mixture and trypsin activity was determined using the synthetic chromogenic substrate N-CBZ-Gly-Pro-Arg-p-nitroanilide. Kinetics of the chromophore release was followed at 405 nm in 0.1 M Tris-HCl (pH 8.0), 1 mM CaCl₂, at 22 °C.

RESULTS AND DISCUSSION

N-terminal processing of recombinant human cationic trypsinogen by aminopeptidases in *E. coli*

The initial aim of our study was to test the effect of the pancreatitis-associated A16V mutation on human cationic trypsinogen. The mutation affects the first amino acid (Ala¹⁶) of the mature zymogen. First, trypsinogens cloned under the control of the T7 promoter in the pTrap-T7 plasmid were expressed in the *E. coli* Rosetta (DE3) strain (Novagen), which produces the T7 RNA polymerase in an IPTG-inducible fashion. The secretory signal sequence was deleted

from the trypsinogen gene and an initiator methionine was inserted upstream of the native Ala¹⁶-Pro¹⁷-Phe¹⁸- N terminus [12,17]. Unexpectedly, N-terminal protein sequencing of purified trypsinogens revealed an inconsistent N terminus with approximately 30% of trypsinogens having a Pro¹⁷-Phe¹⁸ sequence, whereas circa 70 % of trypsinogens contained the Met-Ala¹⁶-Pro¹⁷-Phe¹⁸- sequence (Table 1). Essentially no trypsinogen was recovered with an authentic, native-like N terminus.

Post-translational removal of the initiator methionine is a well-documented phenomenon in *E. coli* [26]. Methionine aminopeptidase cleaves off the N-terminal methionine of cytoplasmic proteins, provided the next amino acid has a small, uncharged side chain. However, when recombinant proteins are expressed to high levels in *E. coli*, methionine aminopeptidase becomes saturated and only a fraction of the recombinant proteins is processed. This problem is alleviated by co-expression of methionine aminopeptidase [27,28]. Clearly, in our experiments a large fraction (~70%, Table 1) of the over-expressed trypsinogen remained unprocessed, indicating that saturation of methionine aminopeptidase has occurred. Consistent with this interpretation, when a high-expressing mutant of cationic trypsinogen (D22A, see ref 29) was produced, more than 90% of recombinant trypsinogens retained the initiator methionine (not shown).

Processing of cationic trypsinogen by methionine aminopeptidase should result in the accumulation of a species with the native-like Ala¹⁶-Pro¹⁷-Phe¹⁸- N-terminal sequence. However, we did not find this protein in the purified trypsinogen fraction. Instead, approximately 30 % of trypsinogens contained the Pro¹⁷-Phe¹⁸- sequence at the N terminus, suggesting that recombinant cationic trypsinogen was further processed by a putative aminopeptidase P enzyme (Table 1). Generally, proline aminopeptidases excise N-terminal amino acids before a proline residue [30]; in this case Ala¹⁶ before Pro¹⁷. Indeed, a gene (*pepP*) coding for an aminopeptidase P enzyme was identified and cloned from *Escherichia coli* HB101 [31,32], suggesting that the *pepP* gene product was responsible for N-terminal processing of recombinant cationic trypsinogen.

Expression of human cationic trypsinogen with an authentic, native N terminus

To study the biochemical effect(s) of the mutation A16V, it was imperative to use a trypsinogen preparation with native, intact N terminus. Therefore, we employed a two-step method to avoid cleavages by the two aminopeptidases mentioned above. First, we eliminated the need for an initiator methionine by fusing the N terminus of the trypsinogen gene with the C terminus of the DnaB mini-intein from *Synechocystis sp* (Fig 1 and online supplementary material). Inteins are self-splicing protein elements with a range of experimental applications in molecular biology [33–35]. Protein splicing is defined as the excision of an intervening sequence (the intein) from a protein precursor and the concomitant ligation of the flanking protein fragments (the exteins) to form a mature host protein (extein) and the free intein [36]. Accordingly, inteins contain an N-terminal and a C-terminal splicing region which catalyze the respective splicing reactions via different mechanisms (for a more detailed description of the splicing mechanisms and references see the New England Biolabs website at <http://www.neb.com/neb/inteins.html>). In our fusion construct, the N-terminal splicing was disabled by replacement of the catalytic N-terminal Cys by an Ala residue. The C-terminal splicing region was directly fused “in-frame” with the Ala¹⁶-Pro¹⁷-Phe¹⁸- sequence of the trypsinogen N terminus, which now serves as the extein in this particular context (Fig 1 and online supplementary material). *In vitro*, the C-terminal cleavage reaction is usually induced by incubation of the fusion proteins at pH 7.0, 25 °C, however, cleavage can also spontaneously occur during expression *in vivo*. As expected, after C-terminal self-cleavage of the intein, a trypsinogen with the native N terminus Ala¹⁶-Pro¹⁷-Phe¹⁸- was released from the intein-trypsinogen fusion protein. However, this N terminus was still subject to cleavage by

aminopeptidase P, as revealed by N terminal sequencing of intein fusion proteins expressed in Rosetta (DE3) cells (Table 1). In the next step, we set out to prevent cleavage by aminopeptidase P. We created an aminopeptidase P deficient *E. coli* strain by removing the gene coding for aminopeptidase P (*pepP*) [31,32] with the method described by Datsenko and Wanner (2000) [24]. Absence of the *pepP* gene was demonstrated by direct sequencing of the region, which revealed the characteristic 'scar' sequence, normally remaining after the procedure (Fig 2). Growth curves in LB medium for the knock-out strain as well as the parental strain BW25113 were essentially identical; indicating that *pepP* is not an essential gene under typical laboratory conditions. Importantly, all trypsinogens expressed as intein fusions in the aminopeptidase P deficient *E. coli* had authentic, native-like N termini (Table 1).

Expression and purification of intein-trypsinogen fusions in the *pepP* negative *E. coli* strain

Intein fusion constructs with wild type and mutant trypsinogens were expressed in *E. coli* LG-3 as inclusion bodies (Fig 3A). SDS-PAGE analysis of solubilized inclusion bodies revealed that a large fraction of the fusion proteins have already been processed to mature trypsinogens *in vivo*. After *in vitro* re-folding, trypsinogens were purified by ecotin affinity chromatography [23]. On SDS-PAGE gels, the majority of the preparation consisted of cleaved, intein-free trypsinogen with slower migrating bands corresponding to unprocessed fusion proteins (Fig. 3B). To remove the small fraction of unprocessed fusion proteins, ion exchange chromatography was employed. Samples were loaded onto a Mono S HR 5/5 column and proteins were eluted with a gradient of 0–0.5 M NaCl. Eluted trypsinogens were free from unprocessed forms (Fig. 3C).

Assessing the effect of the A16V mutation on trypsinogen autoactivation

The A16V mutation of cationic trypsinogen was identified by Witt et al. [37] as the 3rd most common genetic variant associated with chronic pancreatitis. Other groups have also found this mutation in their samples [10,38–40], and to date more than 25 affected carriers have been documented worldwide. Previous biochemical analysis of the more frequent pancreatitis-associated R122H and N29I mutations indicated that the common phenotypic change is an increased propensity for autoactivation [12,17,18]. Therefore, it seemed logical to assume that the A16V mutation might exert its effect via a similar mechanism. On the other hand, the mutation is 7 amino-acids away from the Lys²³-Ile²⁴ cleavage site (see Fig 1) and it affects a region which is largely disordered in crystal structures [41,42], suggesting it may have little, if any, effect on trypsinogen activation. The development of the unique expression system described in this communication allows these assumptions to be tested. In the experiments presented in Fig 4, we compared autoactivation of cationic trypsinogen harboring the A16V mutation to that of the wild-type zymogen. Mutation N29I, previously shown to enhance autoactivation [12,17,18], was included as a positive control. Autoactivation kinetics was determined under two conditions, at pH 5.0 and at pH 8.0 in the presence of 1 mM CaCl₂. As observed previously, the N29I mutant exhibited moderately increased autoactivation under both conditions. Mutation A16V, on the other hand, had essentially no effect on autoactivation at pH 5.0 and caused a slight decrease at pH 8.0. These results indicate that enhanced autoactivation can be ruled out as a mechanism underlying the association of mutation A16V and chronic pancreatitis. Other mechanisms, such as an effect on the intracellular transport of pre-trypsinogen, as speculated by Witt et al., [37] may be operational.

Summary

We have created an expression system to yield recombinant trypsinogen with a native, intact N terminus, using intein-mediated protein splicing and an aminopeptidase P deficient *Escherichia coli*. As a first application of this system, the effect of the pancreatitis-associated mutation A16V on the autoactivation of human cationic trypsinogen was characterized. The

use of the *pepP* knock-out *E. coli* strain should be generally applicable to the expression of other proteins, which undergo unwanted N-terminal trimming by aminopeptidase P.

Acknowledgements

The authors thankfully acknowledge Barry L. Wanner (Department of Biological Sciences; Purdue University, West Lafayette, IN 47907) for sharing the strains and plasmids of the Red-recombinase based gene deletion system. Special thanks to Ron Kaback for his support. Protein sequencing was carried out by David McCourt (Midwest Analytical, Inc).

References

1. Trypsinogen genes: evolution. In: Cooper, DN., editor. Nature encyclopedia of the human genome. Macmillan; London: 2003. p. 645-650.
2. Guy O, Lombardo D, Bartelt DC, Amic J, Figarella C. Two human trypsinogens. Purification, molecular properties, and N-terminal sequences. *Biochemistry* 1978;17:1669-1675. [PubMed: 656395]
3. Rinderknecht H, Renner IG, Carmack C. Trypsinogen variants in pancreatic juice of healthy volunteers, chronic alcoholics and patients with pancreatitis and cancer of the pancreas. *Gut* 1979;20:886-891. [PubMed: 533700]
4. Scheele G, Bartelt D, Bieger W. Characterization of human exocrine pancreatic proteins by two-dimensional isoelectric focusing/sodium dodecyl sulfate gel electrophoresis. *Gastroenterology* 1981;80:461-473. [PubMed: 6969677]
5. Rinderknecht H, Stace NH, Renner IG. Effects of chronic alcohol abuse on exocrine pancreatic secretion in man. *Dig Dis Sci* 1985;30:65-71. [PubMed: 3965275]
6. Rinderknecht H, Renner IG, Abramson SB, Carmack C. Mesotrypsin: a new inhibitor-resistant protease from a zymogen in human pancreatic tissue and fluid. *Gastroenterology* 1984;86:681-692. [PubMed: 6698368]
7. Whitcomb DC, Gorry MC, Preston RA, Furey W, Sossenheimer MJ, Ulrich CD, Martin SP, Gates LK Jr, Amann ST, Toskes PP, Liddle R, McGrath K, Uomo G, Post JC, Ehrlich GD. Hereditary pancreatitis is caused by a mutation in the cationic trypsinogen gene. *Nat Genet* 1996;14:141-145. [PubMed: 8841182]
8. Applebaum-Shapiro SE, Finch R, Pfützer RH, Hepp LA, Gates L, Amann S, Martin S, Ulrich CD II, Whitcomb DC. Hereditary pancreatitis in North America: The Pittsburgh-Midwest Multi-Center Pancreatic Study Group study. *Pancreatology* 1991;1:439-443. [PubMed: 12120221]
9. Keim V, Witt H, Bauer N, Bodeker H, Rosendahl J, Teich N, Mössner J. The course of genetically-determined chronic pancreatitis. *JOP* 2003;4:146-154. [PubMed: 12853682]
10. Howes N, Lerch MM, Greenhalf W, Stocken DD, Ellis I, Simon P, Truninger K, Ammann R, Cavallini G, Charnley RM, Uomo G, Delhaye M, Spicak J, Drumm B, Jansen J, Mountford R, Whitcomb DC, Neoptolemos JP. European Registry of Hereditary Pancreatitis and Pancreatic Cancer (EUROPAC). Clinical and genetic characteristics of hereditary pancreatitis in Europe. *Clin Gastroenterol Hepatol* 2004;2:252-261. [PubMed: 15017610]
11. Otsuki M, Nishimori I, Hayakawa T, Hirota M, Ogawa M, Shimosegawa T. Research Committee on Intractable Disease of the Pancreas. Hereditary pancreatitis: clinical characteristics and diagnostic criteria in Japan. *Pancreas* 2004;28:200-206. [PubMed: 15028953]
12. Sahin-Tóth M. Human cationic trypsinogen. Role of Asn-21 in zymogen activation and implications in hereditary pancreatitis. *J Biol Chem* 2000;275:22750-22755. [PubMed: 10801865]
13. Sinha NK, Light A. Refolding of reduced, denatured trypsinogen and trypsin immobilized on agarose beads. *J Biol Chem* 1975;250:8624-8629. [PubMed: 241750]
14. Orsini G, Goldberg ME. The renaturation of reduced chymotrypsinogen A in guanidine HCl. Refolding versus aggregation. *J Biol Chem* 1978;253:3453-3458. [PubMed: 649582]
15. Odorzynski TW, Light A. Refolding of the mixed disulfide of bovine trypsinogen and glutathione. *J Biol Chem* 1979;254:4291-4295. [PubMed: 438188]
16. Light, Duda CT, Odorzynski TW, Moore WG. Refolding of serine proteinases. *J Cell Biochem* 1986;31:19-26. [PubMed: 3522609]

17. Sahin-Tóth M, Tóth M. Gain-of-function mutations associated with hereditary pancreatitis enhance autoactivation of human cationic trypsinogen. *Biochem Biophys Res Commun* 2000;278:286–289. [PubMed: 11097832]
18. Szilágyi L, Kénesi E, Katona G, Kaslik G, Juhász G, Gráf L. Comparative *in vitro* studies on native and recombinant human cationic trypsins. Cathepsin B is a possible pathological activator of trypsinogen in pancreatitis. *J Biol Chem* 2001;276:24574–24580. [PubMed: 11312265]
19. Peterson FC, Gordon NC, Gettins PG. High-level bacterial expression and ¹⁵N-alanine-labeling of bovine trypsin. Application to the study of trypsin-inhibitor complexes and trypsinogen activation by NMR spectroscopy. *Biochemistry* 2001;40:6275–6283. [PubMed: 11371189]
20. Kukor Z, Tóth M, Sahin-Tóth M. Human anionic trypsinogen. Properties of autocatalytic activation and degradation and implications in pancreatic diseases. *Eur J Biochem* 2003;270:2047–2058. [PubMed: 12709065]
21. Szmola R, Kukor Z, Sahin-Tóth M. Human mesotrypsin is a unique digestive protease specialized for the degradation of trypsin inhibitors. *J Biol Chem* 2003;278:48580–48589. [PubMed: 14507909]
22. Hohenblum H, Vorauer-Uhl K, Katinger H, Mattanovich D. Bacterial expression and refolding of human trypsinogen. *J Biotechnol* 2004;109:3–11. [PubMed: 15063609]
23. Lengyel Z, Pál G, Sahin-Tóth M. Affinity purification of recombinant trypsinogen using immobilized ecotin. *Protein Express Purif* 1998;12:291–294.
24. Datsenko KA, Wanner BL. One-step inactivation of chromosomal genes in *Escherichia coli* K-12 using PCR products. *Proc Natl Acad Sci USA* 2000;97:6640–6645. [PubMed: 10829079]
25. Tabor S, Richardson CC. A bacteriophage T7 RNA polymerase/promoter system for controlled exclusive expression of specific genes. *Proc Natl Acad Sci USA* 1985;82:1074–1078. [PubMed: 3156376]
26. Ben-Bassat A, Bauer K, Chang SY, Myambo K, Boosman A, Chang S. Processing of the initiation methionine from proteins: properties of the *Escherichia coli* methionine aminopeptidase and its gene structure. *J Bacteriol* 1987;169:751–757. [PubMed: 3027045]
27. Shen TJ, Ho NT, Simplaceanu V, Zou M, Green BN, Tam MF, Ho C. Production of unmodified human adult hemoglobin in *Escherichia coli*. *Proc Natl Acad Sci USA* 1993;90:8108–8112. [PubMed: 8367471]
28. Van Valkenburgh HA, Kahn RA. Coexpression of proteins with methionine aminopeptidase and/or N-myristoyltransferase in *Escherichia coli* to increase acylation and homogeneity of protein preparations. *Methods Enzymol* 2002;344:186–193. [PubMed: 11771383]
29. Nemoda Z, Sahin-Tóth M. The tetra-aspartate motif in the activation peptide of human cationic trypsinogen is essential for autoactivation control but not for enteropeptidase recognition. *J Biol Chem* 2005;280:29645–29652. [PubMed: 15970597]
30. Yaron A, Mlynar D. Aminopeptidase-P. *Biochem Biophys Res Commun* 1968;32:658–663. [PubMed: 4878817]
31. Yoshimoto T, Murayama N, Honda T, Tone H, Tsuru D. Cloning and expression of aminopeptidase P gene from *Escherichia coli* HB101 and characterization of expressed enzyme. *J Biochem (Tokyo)* 1988;104:93–97. [PubMed: 2851590]
32. Yoshimoto T, Tone H, Honda T, Osatomi K, Kobayashi R, Tsuru D. Sequencing and high expression of aminopeptidase P gene from *Escherichia coli* HB101. *J Biochem (Tokyo)* 1989;105:412–416. [PubMed: 2659585]
33. Perler FB, Adam E. Protein splicing and its applications. *Curr Opin Biotechnol* 2000;11:377–383. [PubMed: 10975457]
34. Paulus H. Protein splicing and related forms of protein autoprocessing. *Annu Rev Biochem* 2000;69:447–496. [PubMed: 10966466]
35. Xu MQ, Evans TC Jr. Intein-mediated ligation and cyclization of expressed proteins. *Methods* 2001;24:257–277. [PubMed: 11403575]
36. Perler FB, Davis EO, Dean GE, Gimble FS, Jack WE, Neff N, Noren CJ, Thorner J, Belfort M. Protein splicing elements: inteins and exteins--a definition of terms and recommended nomenclature. *Nucleic Acids Res* 1994;22:1125–1127. [PubMed: 8165123]

37. Witt H, Luck W, Becker M. A signal peptide cleavage site mutation in the cationic trypsinogen gene is strongly associated with chronic pancreatitis. *Gastroenterology* 1999;117:7–10. [PubMed: 10381903]
38. Pfützer RH, Whitcomb DC. Trypsinogen mutations in chronic pancreatitis. *Gastroenterology* 1999;117:1507–1508. [PubMed: 10610342]
39. Chen JM, Ragueneas O, Ferec C, Deprez PH, Verellen-Dumoulin C, Andriulli A. The A16V signal peptide cleavage site mutation in the cationic trypsinogen gene and chronic pancreatitis. *Gastroenterology* 1999;117:1508–1509. [PubMed: 10610343]
40. Teich N, Bauer N, Mossner J, Keim V. Mutational screening of patients with nonalcoholic chronic pancreatitis: identification of further trypsinogen variants. *Am J Gastroenterol* 2002;97:341–6. [PubMed: 11866271]
41. Kossiakoff AA, Chambers JL, Kay LM, Stroud RM. Structure of bovine trypsinogen at 1.9 Å resolution. *Biochemistry* 1977;16:654–664. [PubMed: 556951]
42. Pasternak A, Ringe D, Hedstrom L. Comparison of anionic and cationic trypsinogens: the anionic activation domain is more flexible in solution and differs in its mode of BPTI binding in the crystal structure. *Protein Sci* 1999;8:253–258. [PubMed: 10210204]

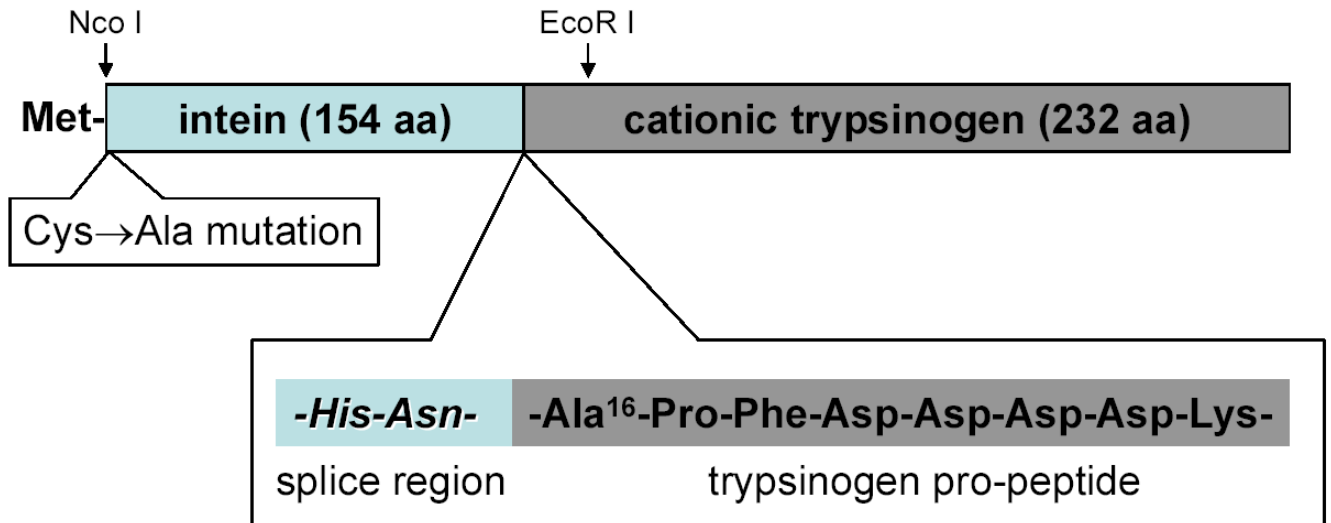


Figure 1.

Primary structure of the intein-trypsinogen fusion protein. An initiator methionine was placed upstream of the 154 amino-acid long mini-intein, which was then fused in-frame with the 232 amino-acid long human cationic trypsinogen. At the fusion junction, the C-terminal Asn of the splice region is joined to Ala¹⁶ of the trypsinogen pro-peptide. The N-terminal cysteine of the mini-intein has been mutated to Ala to disable splicing at the N terminus. The positions of the restriction endonuclease sites used for the construction of the gene-fusion are also indicated. See text for additional details. The sequence of the fusion construct has been deposited with GenBank under accession number DQ371396.

P1-sense primer

5' -ACTCTACACTAAAAACAAAAACGTAAGGAGAGTGTATGAGTGTGTAGGCTGGAGCTGCTTC-3'

P2-antisense primer

5' -AGCGCCAGCGTCGCGCCCGCCATGCCGCCACCGACGATGATTACGCATATGAATATCCTCCTTA-3'

pepP gene

CCGCAACCGACCGCCAGAAAGTACAAAAACCGACTCTACACTAAAAACAAAAACGTAAGGAGAGTGTATGAGTGT
 AGATATCCCAGCAAGAGTTTCAGCGTCGCGGTGAGCCCTGGTGGAGCAAATGCAACCCGCGAGCGCCGCGCTGATT
 TTTGCTGCACCAGAAGTAACACGTAGCGCCGACAGCGAATACCCCTATCGTCAGAACAGTGACTTCTGGTACTTCAC
 CGGCTTTAACGAACCGGAAGCGGTGCTGGTGTGATTTAAAAGCGATGACACTCATAACCACAGCGTTCTGTTTAAAC
 GCGTTCGCGACCTGACGGCGGAGATCTGGTTTGGCCGTCGCTTAGGCCAGGATGCCGCGCCAGAGAACTGGGCGTT
 GACCGCGCACTGGCATTTCAGCGAAATCAATCAGCAACTTTATCAACTACTTAACGGCTGGATGTGGTTTACCATGC
 CCAGGGCGAATATGCATATGCTGATGTAATCGTGAACAGTGCCTGGAAAACTGCGTAAAGGTTCCGCGGCAAAATC
 TCACCGCACCGCAACGATGATCGACTGGCGTCTGTTGTTTCATGAAATGCGCCTGTTCAAATCGCCAGAAGAGATT
 GCCGTACTCCGCGCGCGGGAGAAATCACCGCCATGGCACATACAGGGCGATGGAAAAATGCCGTCCGGGAATGTT
 CGAGTACCATCTGGAAGGCGAAATTCACCACGAATTTAACCGCCACGGTGCAGCGCTATCCGTCTATAACACCATTG
 TCGGCAGCGGTGAAAACGGCTGCATTCTGCACTACACCGAAAACGAGTGTGAAATGCGCGACGGCGACCTGGTGTG
 ATTGACGCGGGTTGTGAATACAAAGGTTACGCTGGCGATATTACCGCACCTTCCCGTCAACGGCAAATTCACCCA
 GGCCAGCGTGAAATCTACGACATTGTGCTGGAGTCTCTCGAAACCAGCCTGCGCCTGTATCGTCCGGGAATTC
 TTCTGGAAGTCACTGGTGAAGTGGTGCATCATGGTTAGCGGCTGGTAAAACTCGGCATCCTGAAAGGTGATGTT
 GATGAACTGATCGCTCAGAACGCCATCGTCTTTCTTTATGCATGGCCTTAGCCACTGGTTAGGACTGGATGTCCA
 TGACGTGGGTGTTTATGGTCAGGATCGCTCGCGCATTCTGGAACCGGCATGGTACTGACCGTAGAGCCAGGGCTGT
 ATATTGCGCCGGATGCAGAAGTCCAGAACAATATCGCGGTATCGGCATTGATGAAAGACGACATTGTGATTACC
 GAAACCGGTAACGAAAACCTCACCGCCAGCGTGGTGA AAAAGCCGGAAGAAATCGAAGCGTTGATGGTTGCTGCGAG
 AAAGCAA**TGA**CGTAATCATCGTCCGTGGCGGCATGGCGGGCGCGACGCTGGCGCTGGCTATTTCCCGGTTAAGTCA
 CGGGCGCTGCCGGTACATTTG

Scar sequence

CGACTCTACACTAAAAACAAAAACGTAAGGAGAGTGTATGAGTGTGTAGGCTGGAGCTGCTTC**AAGTTCCTA**
ACTTTCTAGAG**AAATAGGAAC**TGGGAATAGGAAC**TAAGGAGGATATTCATATG**CGTAATCATCGTCCGTGGCGGCAT
 GGCGGGCGCGACGCTGGCGCTGGCTATTTCCCGGTTAAGTCACGGGGCGCTGCCGGT

Figure 2.

Deletion of the *pepP* gene from the *E. coli* chromosome. The two primers (P1 sense and P2 antisense) used for the amplification of the kanamycin resistance gene from plasmid pKD4 are shown. The primer sequences indicated in blue anneal to the kanamycin resistance gene. The “non-annealing overhangs” of the primers are shown in grey and correspond to the highlighted sequences in the *pepP* gene (GenBank accession number D00398). The initiator methionine codon (ATG) and the stop codon (TGA) within the *pepP* gene are emboldened and underlined. The scar sequence that remained after the deletion of the *pepP* gene contains the two primer sequences. The positions of the primers used to amplify the scar sequence are highlighted in yellow. The lower case bold “c” indicates an unexpected nucleotide deletion in the scar

sequence. The short sequences highlighted in red are the nearly perfect inverted repeats in the FLP recognition target (FRT) site. See “*Methods*” and reference [24] for details.

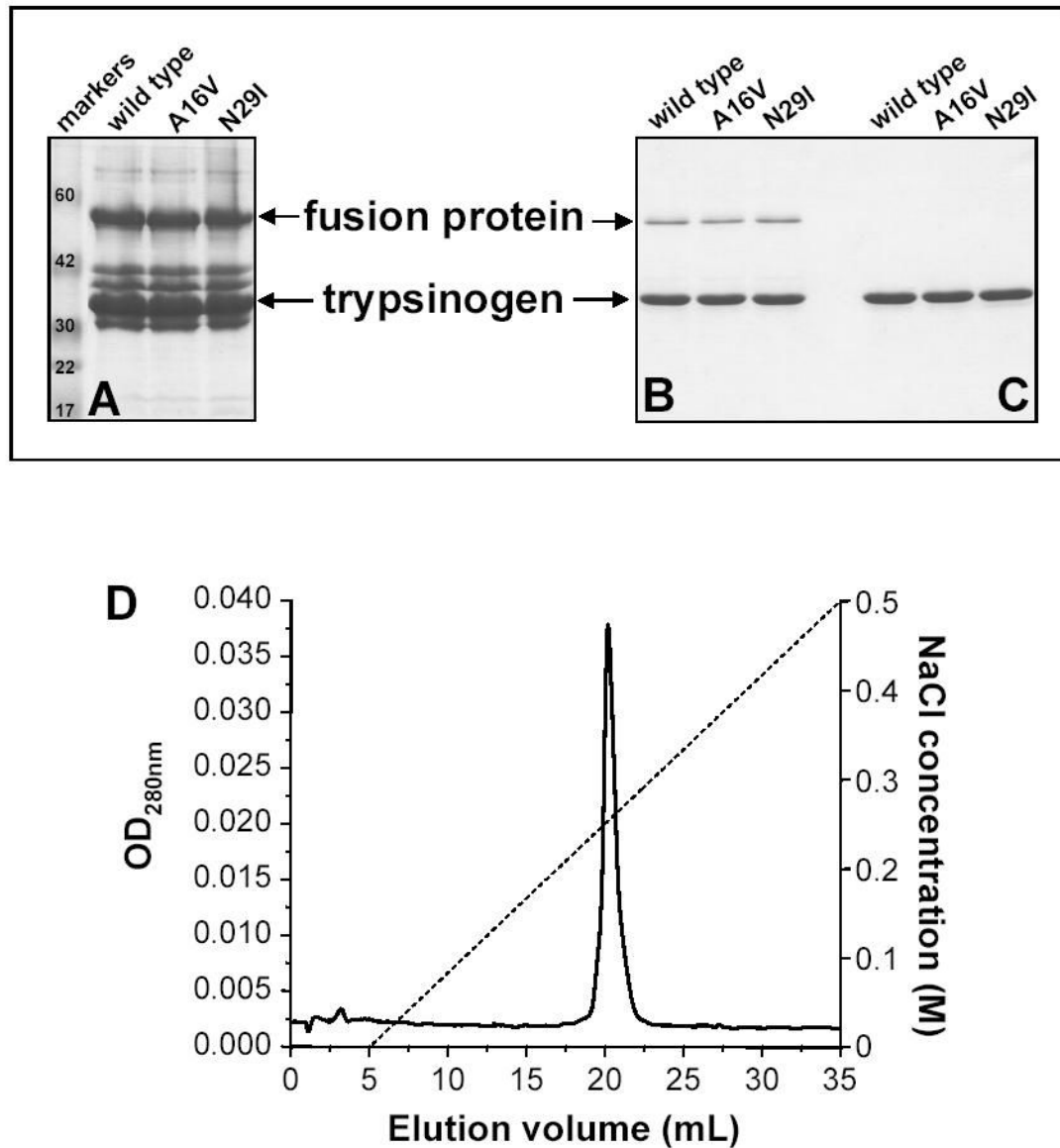


Figure 3. Expression and purification of intein-trypsinogen fusions. SDS-PAGE analysis of **A.** inclusion bodies from LG-3 cells expressing intein-trypsinogen fusions; **B.** trypsinogen eluted from the ecotin-affinity column; and **C.** trypsinogen fractions after MonoS chromatography. The inclusion body fraction prepared from 1.8 mL LG-3 culture (OD_{600} 1.6) or ~5 μ g purified protein was loaded per lane. Samples were solubilized in reducing Laemmli sample buffer, heat-denatured, electrophoresed on 13 % minigels, and stained with Coomassie blue. The molecular weight markers were MultiMark multi-colored standards (Invitrogen). **D.** MonoS chromatography of cationic trypsinogen eluted from the ecotin-affinity column. The column was developed with a 0–0.5 M gradient of NaCl at a flow rate of 1 mL/min. Under these conditions only pure trypsinogen is eluted, while unprocessed fusion proteins are not recovered.

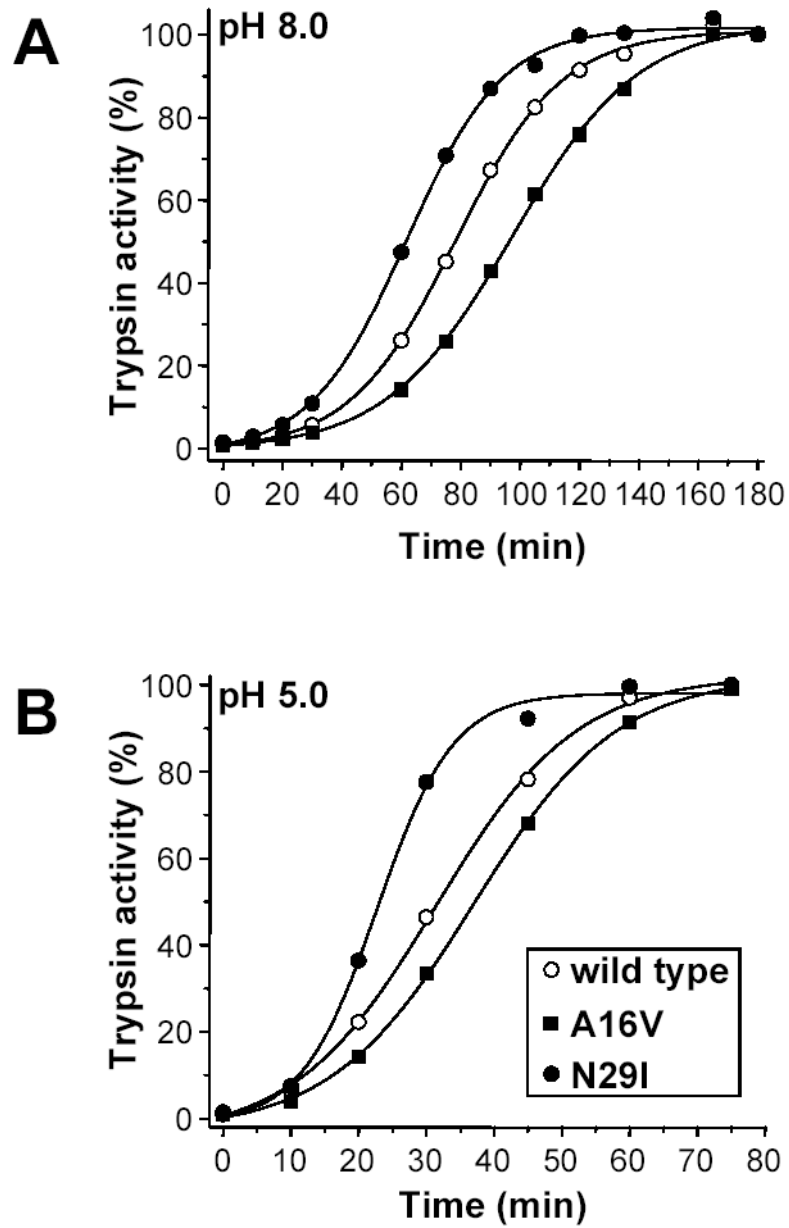


Figure 4. Effect of the pancreatitis-associated A16V mutation on the autoactivation of human cationic trypsinogen. Experiments were carried out at pH 8.0 (A) and pH 5.0 (B), as described in “Methods”. Trypsin activity was expressed as percent of potential maximal activity, which was determined by enteropeptidase activation.

```

      Nco I
cc ATG GCT ATC TCT GGC GAT AGT CTG ATC AGC CTG GCT AGC ACA GGA AAA AGA GTT TCT
MET Ala Ile Ser Gly Asp Ser Leu Ile Ser Leu Ala Ser Thr Gly Lys Arg Val Ser

ATT AAA GAT TTG TTA GAT GAA AAA GAT TTT GAA ATA TGG GCA ATT AAT GAA CAG ACG ATG
Ile Lys Asp Leu Leu Asp Glu Lys Asp Phe Glu Ile Trp Ala Ile Asn Glu Gln Thr MET

AAG CTA GAA TCA GCT AAA GTT AGT CGT GTA TTT TGT ACT GGC AAA AAG CTA GTT TAT ATT
Lys Leu Glu Ser Ala Lys Val Ser Arg Val Phe Cys Thr Gly Lys Lys Leu Val Tyr Ile

CTA AAA ACT CGA CTA GGT AGA ACT ATC AAG GCA ACA GCA AAT CAT AGA TTT TTA ACT ATT
Leu Lys Thr Arg Leu Gly Arg Thr Ile Lys Ala Thr Ala Asn His Arg Phe Leu Thr Ile

GAT GGT TGG AAA AGA TTA GAT GAG CTA TCT TTA AAA GAG CAT ATT GCT CTA CCC CGT AAA
Asp Gly Trp Lys Arg Leu Asp Glu Leu Ser Leu Lys Glu His Ile Ala Leu Pro Arg Lys

CTA GAA AGC TCC TCT TTA CAA TTG TCA CCA GAA ATA GAA AAG TTG TCT CAG AGT GAT ATT
Leu Glu Ser Ser Ser Leu Gln Leu Ser Pro Glu Ile Glu Lys Leu Ser Gln Ser Asp Ile

TAC TGG GAC TCC ATC GTT TCT ATT ACG GAG ACT GGA GTC GAA GAG GTT TTT GAT TTG ACT
Tyr Trp Asp Ser Ile Val Ser Ile Thr Glu Thr Gly Val Glu Glu Val Phe Asp Leu Thr
16

GTG CCA GGA CCA CAT AAC TTT GTC GCG AAT GAC ATC ATT GTA CAC AAC GCT CCT TTC GAT
Val Pro Gly Pro His Asn Phe Val Ala Asn Asp Ile Ile Val His Asn Ala Pro Phe Asp
29

GAT GAT GAC AAG ATC GTT GGA GGA TAC AAC TGC GAA GAG AAT TCT GTT CCC TAC CAA GTG
Asp Asp Asp Lys Ile Val Gly Gly Tyr Asn Cys Glu Glu Asn Ser Val Pro Tyr Gln Val
EcoR I

TCC CTG AAC TCT GGC TAC CAC TTC TGT GGA GGT TCT CTC ATC AAT GAA CAG TGG GTG GTG
Ser Leu Asn Ser Gly Tyr His Phe Cys Gly Gly Ser Leu Ile Asn Glu Gln Trp Val Val

TCT GCA GGT CAC TGC TAT AAG TCC CGC ATC CAA GTG AGA CTG GGA GAG CAC AAC ATC GAA
Ser Ala Gly His Cys Tyr Lys Ser Arg Ile Gln Val Arg Leu Gly Glu His Asn Ile Glu

GTC CTT GAG GGC AAT GAG CAG TTT ATC AAT GCT GCC AAG ATC ATC AGG CAT CCC CAA TAC
Val Leu Glu Gly Asn Glu Gln Phe Ile Asn Ala Ala Lys Ile Ile Arg His Pro Gln Tyr

GAT AGG AAG ACC CTG AAC AAT GAC ATC ATG CTG ATC AAG CTC TCT TCC CGT GCG GTA ATC
Asp Arg Lys Thr Leu Asn Asn Asp Ile MET Leu Ile Lys Leu Ser Ser Arg Ala Val Ile

AAT GCT CGA GTG TCC ACC ATC TCT CTG CCC ACT GCC CCT CCA GCT ACT GGC ACC AAG TGC
Asn Ala Arg Val Ser Thr Ile Ser Leu Pro Thr Ala Pro Pro Ala Thr Gly Thr Lys Cys

CTC ATC TCC GGC TGG GGC AAC ACT GCG AGT TCT GGT GCC GAC TAC CCA GAC GAG CTG CAG
Leu Ile Ser Gly Trp Gly Asn Thr Ala Ser Ser Gly Ala Asp Tyr Pro Asp Glu Leu Gln

TGC CTG GAT GCT CCT GTG CTG AGC CAG GCT AAG TGT GAA GCC TCC TAC CCT GGA AAG ATT
Cys Leu Asp Ala Pro Val Leu Ser Gln Ala Lys Cys Glu Ala Ser Tyr Pro Gly Lys Ile

ACC AGC AAC ATG TTC TGT GTT GGC TTC CTA GAG GGA GGC AAG GAT TCC TGC CAG GGT GAC
Thr Ser Asn MET Phe Cys Val Gly Phe Leu Glu Gly Gly Lys Asp Ser Cys Gln Gly Asp

TCT GGT GGC CCT GTG GTC TGT AAT GGA CAG CTC CAA GGA GTT GTC TCC TGG GGC GAT GGC
Ser Gly Gly Pro Val Val Cys Asn Gly Gln Leu Gln Gly Val Val Ser Trp Gly Asp Gly

TGT GCC CAG AAG AAC AAG CCT GGA GTC TAC ACC AAG GTC TAC AAC TAT GTG AAA TGG ATT
Cys Ala Gln Lys Asn Lys Pro Gly Val Tyr Thr Lys Val Tyr Asn Tyr Val Lys Trp Ile

AAG AAC ACC ATA GCT GCC AAC AGC TAG
Lys Asn Thr Ile Ala Ala Asn Ser TER

```

Online supplementary material

Nucleotide and amino-acid sequences of the intein-trypsinogen fusion construct. The restriction endonuclease sites used for engineering the gene-fusion and positions of Ala¹⁶ and Asn²⁹ in human cationic trypsinogen are indicated. Note that the nucleotide sequence of cationic trypsinogen in this gene construct differs at several positions from the *PRSSI* cDNA sequence, but encodes the same protein (see ref. 12 for details). GenBank accession number DQ371396.

Table 1

N-terminal sequences of recombinant trypsinogen preparations expressed in Rosetta(DE3) cells or the aminopeptidase P-deficient LG-3 cells. Purified trypsinogens were electrophoresed on 13 % SDS-PAGE minigels, transferred to PVDF membranes and subjected to Edman degradation. N.D., not determined. See text for details.

<i>E. coli</i> strain	PRSS1	Intein-PRSS1
Rosetta(DE3)	Met-Ala ¹⁶ -Pro ¹⁷ -Phe ¹⁸ - ~70% Pro ¹⁷ -Phe ¹⁸ - ~30%	Ala ¹⁶ -Pro ¹⁷ -Phe ¹⁸ - ~50% Pro ¹⁷ -Phe ¹⁸ - ~50%
LG-3 (<i>pepP</i> ⁻)	N.D.	Ala ¹⁶ -Pro ¹⁷ -Phe ¹⁸ - 100%