Available online at www.sciencedirect.com
SCIENCE DIRECT®
Protein Expression and Purification 29 (2003) 185–192
Protein Expression & Purification
www.elsevier.com/locate/yprep

The first semi-synthetic serine protease made by native chemical ligation

Gábor Pál,[a] Fanny Santamaria,[b] Anthony A. Kossiakoff,[a] and Wuyuan Lu[b,*]

[a] *Department of Biochemistry and Molecular Biology, the University of Chicago, 920 East 58th Street, Chicago, IL 60637, USA*
[b] *Institute of Human Virology, University of Maryland Biotechnology Institute, 725 West Lombard Street, Baltimore, MD 21201, USA*

Received 31 October 2002, and in revised form 9 January 2003

Abstract

Selective incorporation of non-natural amino acid residues into proteins is a powerful approach to delineate structure–function relationships. Although many methodologies are available for chemistry-based protein engineering, more facile methods are needed to make this approach suitable for routine laboratory practice. Here, we describe a new strategy and provide a proof of concept for engineering semi-synthetic proteins. We chose a serine protease *Streptomyces griseus* trypsin (SGT) for this study to show that it is possible to efficiently couple a synthetic peptide containing a catalytically critical residue to a recombinant fragment containing the other active site residues. The 223-residue hybrid SGT molecule was prepared by fusing a chemically synthesized N-terminal peptide to a large C-terminal fragment of recombinant origin using native chemical ligation. This C-terminal polypeptide was produced from full-length SGT by cyanogen bromide cleavage at a genetically engineered Met57 position. This semi-synthetic hybrid trypsin is fully active, showing kinetics identical to the wild-type enzyme. Thus, we believe that it is an ideal model enzyme for studying the catalytic mechanisms of serine proteases by providing a straightforward approach to incorporate non-natural amino acids in the N-terminal region of the protein. In particular, this strategy will allow for replacement of the catalytic His57 residue and the buried N-terminus, which is thought to help align the active site, with synthetic analogs. Our approach relies on readily available recombinant proteins and small synthetic peptides, thus having general applications in chemical engineering of large proteins where the N-terminal region is the focal interest.

Engineering the functional properties of proteins using non-natural building blocks is becoming a widely used tool of protein chemists. Several quite different methods for incorporating non-coded amino acids into proteins are currently in practice including total chemical synthesis [1–3], in vitro translation [4–6], auxotrophic bacterial expression [7–10], enzymatic semi-synthesis [11–13], native chemical ligation [14–17], and more recently expression in *Escherichia coli* using novel orthogonal tRNA/synthetase pairs [18,19]. However, there are a variety of technical challenges associated with each of these technologies that have limited their widespread uses by most protein engineers.

Native chemical ligation, developed by Kent and colleagues, allows for two fully unprotected synthetic peptides to react chemoselectively in aqueous solution to assemble a longer polypeptide chain linked by a new peptide bond [14]. The method requires the presence of a C-terminal α-thioester moiety on the N-terminal peptide and an N-terminal cysteine residue on the C-terminal peptide. Chemically synthesizing proteins by native chemical ligation is a robust approach to produce molecules up to typically about 120 amino acid residues. Larger protein molecules can be made using multi-step ligation strategies either in solution or on solid phase [15,20]; however, synthesis of multiple individual peptides is required and elaborate chemistries associated with the assembly process are involved. Consequently, synthesis of large proteins still presents significant challenges limiting its use in practice.

Although there are many small protein systems that offer opportunities to dissect important structure–function parameters through the standard chemical synthesis approach, many of the most exciting problems involve proteins of sizes that are not readily accessible by

* Corresponding author. Fax: 1-410-706-7583.
E-mail address: luw@umbi.umd.edu (W. Lu).

1046-5928/03/$ - see front matter © 2003 Elsevier Science (USA). All rights reserved.
doi:10.1016/S1046-5928(03)00022-6

standard methods. Since generally only a few non-natural amino acid residues would be introduced into any one molecule, the most efficient method would be the one that employs both chemical and recombinant arms combined in ways to maximize their individual advantages. In this regard, the Muir group developed an intein-mediated expression system to generate α-thioester-bearing protein molecules for subsequent native chemical ligation reactions [21]. This expressed protein ligation approach has been used successfully to prepare large chimeric proteins that contain a C-terminal synthetic moiety of interest, thus significantly expanding the scope of applications of chemical protein engineering [22–24]. However, while well suited for preparing large multi-domain semi-synthetic proteins whose C-terminal region is the target for chemical modifications, expressed protein ligation has limited applications in studies where the N-terminal region of the protein is the focal interest. This is because the N-terminal segment used in expressed protein ligation is produced through cell-based expression, which would be inherently difficult to allow for straightforward incorporation of non-natural building blocks.

The approach we present here is complementary to the intein-mediated synthesis method. It also couples chemical synthesis and recombinant approaches to produce a semi-synthetic protein, but in this case the non-natural amino acids can be introduced into the N-terminal section of the molecule. The C-terminal portion is derived through recombinant procedures and, thus, it can be very large, its size being limited only by the applied expression system. The only requirement for the C-terminal peptide fragment is an N-terminal Cys residue. The N-terminal piece is chemically synthesized, has an α-thioester carboxyl terminus moiety for subsequent native chemical ligation, and can contain multiple non-natural amino acid substitutions. We have previously shown through developing the biosynthetic phage display methodology [25] that this approach is viable for making functionally active semi-synthetic proteins. However, in the case of the phage display application the conditions were less stringent; small amounts and low concentrations of the recombinant peptides were presented on phage that prevented aggregation of the unfolded C-terminal fragments. The efficiency of producing folded semi-synthetic products on phage is much less than what we wished to achieve here. Our goal here was to translate this protein engineering concept into a robust user-friendly method whereby large quantities of semi-synthetic proteins can be produced for detailed characterization by structural, biochemical, and biophysical approaches.

Here, we report preparation and characterization of a fully active, hybrid *Streptomyces griseus* trypsin (SGT) of which the first segment of 37 residues at the N-terminus, (16–57)SGT (chymotrypsinogen numbering), was made by solid phase peptide synthesis while the C-terminal segment of 186 residues, (58–245)SGT bearing an N-terminal Cys residue, was generated by cyanogen bromide (CNBr) cleavage of a recombinant SGT molecule over-expressed in *E. coli*. The C-terminal fragment, in principle, could be generated by proteolytic processing. In fact, we made an attempt to use a genetically engineered subtilisin variant, Genenase, as the processing agent. However, poor solubility of the recombinant product that contained a Genenase cleavage site preceding Cys^{58} rendered the enzymatic processing unfeasible (for details see Discussion). Therefore, we applied a chemical processing approach where use of high concentrations of denaturants circumvents the solubility problem. Our strategy will allow for virtually unlimited non-natural modifications to the chemically synthesized part of the hybrid trypsin molecule. Since His^{57}–Cys^{58} is completely conserved in the chymotrypsin family of serine proteases, this approach can be applied directly for many other serine proteases therefore promising an ideal tool for better understanding their catalytic mechanisms. The method presented here is generally applicable for the semi-synthesis of large proteins where the N-terminal region is the focal interest.

Materials and methods

Cloning strategy

A plasmid containing the coding sequence for *Streptomyces griseus* trypsin was provided by Dr. Thor Borgford from Simon Fraiser University. For the expression of the full-length (16–245)SGT molecule, the corresponding DNA fragment was produced by PCR using the following primers:

(a) SGTR-5′
5′-AAAAAAAACATATG*TTCGCGCACTAC*GTC
GTCGGCGGAACCCG-3′
HisMet*PheAlaHisTyr*ValValGlyGlyThr
(b) SGTR-3′
5′-AAAAAACTCGAGTTATTAGAGCGTGCGG
GCGGCCGAGGCGAATGGCGG-3′

Cleavage sites for the restriction endonucleases *Nde*I (in oligo a) and *Xho*I (in oligo b) are underlined. The 5′ primer also codes for a peptide segment that serves as recognition and cleavage site for Genenase [26,27]. The Genenase site coding portion of primer a as well as the corresponding translated peptide sequence is indicated as *italic*. Genenase should cleave after the Tyr residue of the PheAlaHisTyr segment. A similar construct (not shown) was also produced to yield truncated (58–245)SGT with an N-terminal Cys residue after the Genenase processing. However, since proteolytic processing failed to produce this segment, an alternative

chemical approach had to be explored (refer to Discussion for details). The PCR product was digested with the indicated restriction endonucleases and ligated into a pET-28a(+) vector (Novagen) precut with the same restriction enzymes. The construct pGenSGTr was transformed into XL1-blue *E. coli* cells (Stratagene). Plasmids from individual clones were isolated and sequenced.

Expression and purification of recombinant SGTr

The pGenSGTr expression vector was transformed into BL21(DE3) *E. coli* cells (Novagen). The combination of the BL21(DE3) cell line and the pET-28a(+) vector ensured robust IPTG-inducible expression of the gene from a T7 promoter and also provided an N-terminal His tag for the gene product. Cells harboring the expression vector were grown at 37 °C in LB medium containing 35 mg/ml kanamycin. At $OD_{600} = 2.0$, the cells were induced by 1 mM final concentration of IPTG and incubated for an additional 4 h. Then, the cells were harvested by centrifuging (10 min at 8000*g*) and the pellet was frozen and stored overnight at −70 °C. The cells were resuspended in 30 ml lysis buffer (50 mM Tris–HCl, 1 mM EDTA, and 100 mM NaCl) and 800 μl of 10 mg/ml lysozyme (Boehringer) was added. After thorough sonication, 1.5 ml of 100 mM $MgCl_2$ and 800 U Bensonaze (Sigma) were added to remove nucleic acid contamination. The suspension was incubated at 37 °C for 1 h. The inclusion body fraction was collected by centrifugation (10 min at 10,000*g*) and washed several times with repeated steps of resuspension in lysis buffer and centrifugation. Although the recombinant construct had N-terminal His tag, no nickel column had to be applied, because the above protocol yielded very large amount (~800 mg protein from 1 L *E. coli* culture) and high purity inclusion body fraction.

Constructing an SGTr variant for CNBr cleavage

To use CNBr as a processing agent to produce a (58–245)SGT fragment with Cys58 as its N-terminal residue, the His57Met mutation was designed. Since the wild-type SGTr sequence has a methionine residue, Met199, we mutated it to leucine to avoid CNBr cleavage at this position. The mutations were introduced by the Kunkel method [28]. CJ236 cells from Bio-Rad were transformed with the pGenSGTr vector and VCS M13 helper phage from Stratagene was used to isolate uracil-containing single-stranded DNA as template for the mutagenesis reactions. The following mutagenesis primers were used:

His57Met-SGTr

5′-CCGATCCGCTCACGCACATGGCCGCGGTGAGGACG-3′

Met199Leu-SGTr

5′-GGCGTTGTCCTTGCGGAAGAGCGGGCCGCCGGAG-3′

The positions of the mutations are underlined. Note that the f1 origin in pET-28a(+) is oriented such that the resultant single-stranded DNA corresponds to the coding strand. Thus, the complementary mutagenesis primers represent the non-coding strand. XL1-Blue cells were transformed with the mutagenesis products and positive clones containing the desired mutations were identified with DNA sequencing. BL21 (DE3) cells were transformed with the His57Met; Met199Leu SGT constructs. The expression and purification were done by the same protocol that has been described above.

Cleavage of recombinant SGT by CNBr

Inclusion bodies of (16–245)$Met^{57}Leu^{199}$-SGT were dissolved at 20 mg/ml in 0.5 M Tris–HCl buffer containing 6 M guanidine hydrochloride (GuHCl), pH 9.2, to which oxidized glutathione (GSSG) was added to a final concentration of 100 mg/ml. After 48 h, the reaction was stopped by acidification with acetic acid and the solution was dialyzed in water (MWCO 8000). The resultant protein precipitates were collected by centrifugation and re-suspended at 20 mg/ml in 0.1 M HCl containing 6 M GuHCl and 80 mg/ml CNBr. The cleavage reaction proceeded overnight, followed by dialysis and lyophilization. The major cleavage product, (58–245)Leu^{199}-SGT containing five glutathione molecules disulfide bridged with Cys residues, was verified by electrospray ionization mass spectrometric (ESI-MS) analysis.

Synthesis, renaturation, and purification of hybrid SGT

The N-terminal thioester peptide (16–57)α$COSCH_2$CO–Leu–OH was manually synthesized on Boc-His(Bom)-SCH_2CO-Leu-OCH_2-Pam resin using the published in situ neutralization/HBTU activation protocol for Boc chemistry [29]. After cleavage and deprotection in HF, the peptide was purified by preparative C18 reversed-phase HPLC and its molecular weight was ascertained by ESI-MS. Native chemical ligation of (16–57)α$COSCH_2$CO-Leu-OH and crude (58–245)Leu^{199}-SGT was carried out at 20 mg/ml in 0.2 M phosphate buffer containing 6 M GuHCl and 2% thiophenol, pH 7.5. After 24 h, β-mercaptoethanol was added (25% v/v) to quench the ligation reaction, followed by extensive dialysis in acidified water before lyophilization. The observed molecular weight of the ligation product (16–245)Leu^{199}-SGT was 23088.6 Da, in agreement with the expected value of 23087.9 Da that was calculated based on average isotope compositions.

Folding of SGT was performed by dissolving the crude ligation product at 1 mg/ml in 50 mM Tris–HCl buffer containing 6 M GuHCl, pH 9.2, followed by a

1000-fold dilution using 1 M diethylamine buffer containing 20 mM $CaCl_2$, 3 mM reduced glutathione (GSH), and 0.3 mM GSSG, pH 8.5. The folding proceeded for 3 h at room temperature and the solution was loaded onto a soybean trypsin inhibitor (STI) immobilized affinity column. Elution of the bound, folded enzyme from STI-Sepharose was achieved using 10 mM HCl containing 0.2 M NaCl and 10 mM $CaCl_2$. The molecular weight of the folded product was determined to be 23082.2 by ESI-MS.

Enzyme kinetics analysis of SGT

Enzyme kinetics were carried out in 100 mM Bis–Tris propane buffer containing 20 mM $CaCl_2$ and 0.005% Triton X-100, pH 8.3. Chromogenic substrates Bz-Arg-*p*NA and Bz-Phe-Val-Arg-*p*NA were dissolved in DMF at 10 and 20 mM, respectively. Hydrolysis was monitored by following absorbance changes at 380 nm as different concentrations of substrates [S] were applied in a series of 3-ml cuvettes containing enzyme solutions at a fixed concentration, $[E]_0$. The enzyme concentration was determined by titration with standardized bovine pancreatic trypsin inhibitor [30]. The slopes (s^{-1}) of hydrolysis were converted to initial reaction rates, r (M/s), based on an experimentally determined value of the molar extinction coefficient for the released compound *p*-nitroanilide (*p*NA). Values of k_{cat} and K_m were obtained by non-linear regression analysis using the Michaelis–Menten equation: $r = k_{cat}[S][E]_0/(K_m + [S])$.

Results

The strategy for the preparation of hybrid SGT is illustrated in Fig. 1. In the native sequence of SGT, the catalytic residue His^{57} is conveniently located N-terminal to a neighboring cysteine residue, making an ideal position for native chemical ligation. To obtain the (58–245) fragment from recombinant SGT, His^{57} was mutated to Met so that CNBr could be used to remove the N-terminal peptide, (16–57)SGT. In addition, Met^{199} was replaced by Leu to eliminate undesired chemical cleavage by CNBr at this position of the molecule. The recombinant protein (16–245)$Met^{57}Leu^{199}$-SGT was expressed as inclusion bodies in *E. coli* and obtained in extremely high yield (more than 800 mg/L culture). After protection of all the six Cys residues of SGT by formation of mixed disulfide bridges with oxidized glutathione, (16–245)$Met^{57}Leu^{199}$-SGT was quantitatively cleaved by CNBr, resulting in (58–245)Leu^{199}-SGT. The remaining five mixed disulfide bonds located in the C-terminal segment of SGT were reduced by thiophenol during the subsequent ligation reaction. It should be pointed out that no special purification steps other than centrifugation and dialysis were involved in the entire workup, i.e., inclusion body isolation, CNBr cleavage, and native chemical ligation.

The hybrid polypeptide (16–245)Leu^{199}-SGT was folded and oxidized to the active conformation using a modified version of the previously published conditions [31], yielding greater than 50% activity recovery based on a parallel folding experiment using native enzyme. After affinity purification on STI-Sepharose, the final product was analyzed by C4 reversed-phase HPLC and characterized by ESI-MS (Figs. 2 and 3). The observed mass 23082.2 Da is in excellent agreement with the expected value of 23081.9 Da calculated based on average isotopic compositions for folded SGT. For reduced hybrid SGT prior to oxidative folding, the observed mass was determined to be 23088.6 Da, 6 mass units higher than that of the folded enzyme, indicative of the formation of three disulfide bridges in the final product.

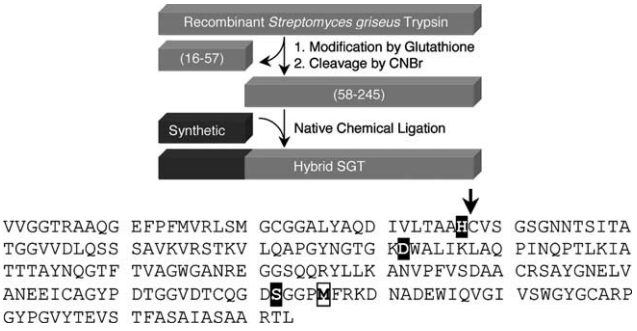

Fig. 1. Strategy for the semi-synthesis of *Streptomyces griseus* trypsin by native chemical ligation. As shown in the amino acid sequence of SGT, the three catalytic residues, His^{57}, Asp^{102}, and Ser^{195}, are shaded in black. The ligation site His^{57}–Cys^{58} is indicated by an arrow. In the recombinant SGT construct, His^{57} was mutated to Met and the only Met residue (boxed) in the C-terminal fragment was replaced by Leu.

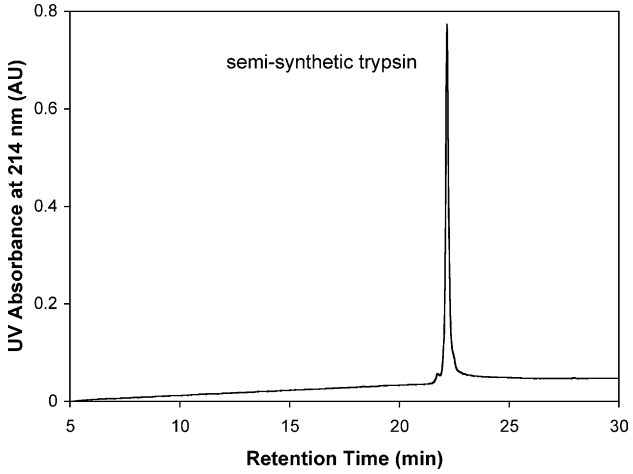

Fig. 2. Folded and purified hybrid *Streptomyces griseus* trypsin on reversed-phase HPLC. The analysis was performed at 40 °C on a Vydac analytical C4 column (4.6 × 150 mm) using a linear gradient of acetonitrile containing 0.1% trifluoroacetic acid, 5–65% over 30 min at a flow rate of 1 ml/min.

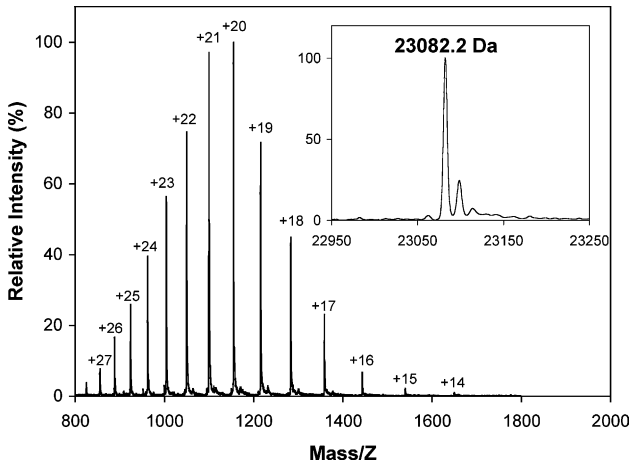

Fig. 3. Electrospray ionization mass spectrum of hybrid *Streptomyces griseus* trypsin. The data were collected on a Micromass ZQ-4000 quadruple mass spectrometer. Samples were diluted in methanol/water (1:1) containing 1% acetic acid and infused by a syringe pump at 10 μl/min.

To characterize the enzymatic activity of hybrid SGT, two chromogenic trypsin substrates were used for kinetic measurements: Bz-Arg-*p*NA and Bz-Phe-Val-Arg-*p*NA. For comparison, naturally occurring SGT was purified to homogeneity from commercial Pronase (Sigma) by STI-Sepharose affinity chromatography and SP-Sepharose cation-exchange purification (data not shown), and an identical activity assay was also performed. The experimental results are tabulated in Table 1.

Discussion

The most straightforward approach to yield a recombinant protein fragment with an N-terminal Cys would be to express the corresponding portion of the gene in the cytoplasm of *E. coli*, where an in vivo process removes the initiator Met residue from the nascent polypeptide. The efficiency of this enzymatic process depends mostly on the size of the residue adjacent to Met; the smaller the side-chain of the residue the more efficient the removal of the N-terminal Met [32]. In the case of a Met–Cys N-terminal processing site, the processing is reported to be relatively efficient [32]. Consequently, our first attempt to produce the 186 amino acid C-terminal polypeptide was to express the Met-(58–245)SGT fragment in the cytoplasm of *E. coli*. Although the pET-27b(+) vector (Novagen) used is optimized for robust transcription and provides optimal Shine–Dalgarno (SD) sequence for efficient translation initiation, no appreciable amount of the recombinant product was obtained. It has recently been recognized that in addition to the SD sequence upstream to the initiator codon, there is another regulatory element called downstream box (DB) that greatly affects translation initiation and expression yield [33,34]. This sequence overlaps with the N-terminal coding region in the mRNA and is thought to be one of the important indirect factors responsible for the long observed dependency of the expression yield on the N-terminal sequence of the protein. Even though codon degeneracy provides some limited choices for silent mutations for a better DB sequence with the original amino acid sequence preserved, we decided to follow a more general approach by elongating the N-terminus with a vector-derived, optimized amino acid sequence. Obviously, this strategy comes with the necessity for a proteolytic cleavage site between the vector-derived sequence and the desired N-terminal Cys residue.

In principle, preparation of (58–245)SGT can be achieved through enzymatic cleavage of either a full-length or a truncated SGT construct and a variety of processing enzymes are available for this purpose. For example, factor Xa, a serine protease with stringent sequence requirements for cleavage site recognition, has been widely used to cleave the linker peptide in recombinant fusion proteins or to remove extra amino acid residues from the N-terminus of a folded protein structure. In fact, it was the preferred enzyme to generate N-terminal Cys-containing protein domains for expressed protein ligation [23,24]. In anticipation of potential solubility problems with SGT, however, we decided to use Genenase, an engineered variant of subtilisin as the processing enzyme, because of its superior stability in moderately high concentrations of denaturants [26,27].

For bacterial expression of the truncated SGT containing the Genenase recognition site, i.e., –Phe^{54}Ala^{55} His^{56}Tyr^{57} ↓ Cys^{58}–, we chose the pET-28a(+) vector that provided a T7 promoter for high level transcription, as well as an optimal ribosome binding site and

Table 1
Kinetic measurements for wild-type and hybrid *Streptomyces griseus* trypsin in 0.1 M Bis–Tris propane buffer containing 20 mM $CaCl_2$ and 0.005% Triton X-100, pH 8.3

Substrates and enzymes		k_{cat} (s^{-1})	K_m (μM)	k_{cat}/K_m ($M^{-1}s^{-1}$)
Bz-R-*p*NA	Hybrid SGT	12.4	110	1.1×10^5
	Wild-type SGT	4.97	41.5	1.2×10^5
Bz-FVR-*p*NA	Hybrid SGT	208	12.7	1.6×10^7
	Wild-type SGT	63.7	3.35	1.9×10^7

N-terminal peptide sequence for high-efficiency translation. Indeed, the construct demonstrated high levels of expression, more than 800 mg of the protein could be purified from 1 L culture. However, the concentration of the denaturants needed to solubilize the protein turned out to be greater than 4 M for GuHCl and even higher for urea, which rendered Genenase totally inactive in the buffer solution.

This problem, in principle, could be overcome by refolding the full-length product to yield a soluble protein. However, in addition to the fact that introduction of the Genenase cleavage site and Tyr^{57} in particular in the highly conserved catalytic cleft of trypsin might disrupt oxidative refolding of the full-length SGT construct, subsequent enzymatic cleavage would be unlikely to occur because the recognition site appears to be structurally inaccessible to Genenase, as judged from the crystal structure of SGT [35]. Therefore, we concluded that enzymatic cleavage to produce the 58–245 fragment is impractical not only from a truncated SGT, but from the full-length construct as well.

This conclusion generally applies to systems where solubility of the protein to be cleaved and/or accessibility of the cleavage site becomes problematic. Specific chemical cleavage using CNBr represents a good alternative because the reaction can be carried out under solvent conditions that ensure complete solubility of the protein. One drawback of the approach, however, is that use of CNBr precludes the presence of any Met residues other than the one N-terminal to the designed cleavage site in the sequence. Fortunately, Met is a relatively rare amino acid and can be readily substituted by residues such as Leu without adverse structural and functional effects, as clearly demonstrated by our findings with SGT. It is worth noting that the side-chains of Met and Leu have similar surface sizes and hydrophobicity scales. In fact, it has been well documented in the literature that replacement of Met by Leu in the hydrophobic packing of a number of extensively studied protein systems causes little or no energetic consequences [36–39].

First introduced by Gross and Witkop, CNBr cleavage of the peptide bond C-terminal to Met has a wide range of applications from processing recombinant fusion proteins to protein sequence analysis [40,41]. Although highly selective and nearly quantitative for most Met–Xxx bonds, the cleavage reaction is nonetheless slow and of low yield when the residues adjacent to Met are Ser, Thr, or Cys. This is because the neighboring hydroxyl or sulfhydryl group participates in a side reaction, resulting primarily in a non-cleavable side product with Met converted to homoserine (Hse) [42]. The side reaction can be particularly problematic for the Met–Cys bond presumably due to the fact that the sulfhydryl group is more nucleophilic as compared with the hydroxyl group of Ser or Thr. In fact, when fully reduced (16–245)$Met^{58}Leu^{199}$-SGT was cleaved by CNBr, the major resultant product detected by mass spectrometry contained the intact Hse^{58}–Cys^{58} peptide bond, which is consistent with the published observations [43]. Interestingly, as Xu et al. reported, treatment of Met–Cys-containing peptides with CNBr resulted in cleavage of the bond but caused oxidation of the Cys residue [24]. Thus, to ensure a productive cleavage of the Met–Cys bond in proteins, the Cys side-chain has to be protected.

We have determined that modification of the free sulfhydryl groups by mixed disulfide bridge formation using oxidized glutathione completely alleviated the side reaction and resulted in a quantitative cleavage of the Met–Cys peptide bond in trypsin. It should also be pointed out that we used a non-conventional medium (0.1 M HCl, 6 M GuHCl) for the cleavage reaction as opposed to the standard 70% formic acid [40]. This alteration of the conditions was applied for two reasons. First, an increase in water content in the reaction medium has been reported to enhance the productive cleavage in the case of Met–Ser and Met–Thr [42]. Second, formic acid often causes formylation of and irreversible damage to proteins [44,45].

Besides being crucial for the high-efficiency cleavage, the glutathione modification of the Cys residues had another unexpected beneficial effect; it significantly improved the chromatographic behavior of SGT on reversed-phase HPLC. The unmodified, denatured SGT and its C-terminal (58–245) fragment are soluble only at high concentrations of urea or GuHCl and their purification by HPLC is not feasible due to permanent retardation of these forms on preparative reversed-phase columns. However, the formation of mixed disulfide bonds with glutathione dramatically increased the solubility of these polypeptides. Even though no chromatographic purification steps were needed to prepare the hybrid SGT molecule prior to folding due to high purity of the inclusion bodies and completeness of the subsequent cleavage and ligation reactions, such modifications should be beneficial in general when purification is needed for protein fragments that are otherwise strongly retained on reversed-phase HPLC.

The hybrid SGT differs from the wild-type enzyme by a single amino acid residue (Met^{199}-to-Leu^{199}). Residue 199 in trypsins is highly variable among different groups of living organisms. In vertebrate trypsins, Val is the predominant residue at this position, whereas Leu^{199} is found in most invertebrate enzymes. Other residues such as Ile, Phe, and Ala, although rare, are also found at position 199 in trypsins. The turn-over number of the enzyme (k_{cat}) and the apparent dissociation constant (K_m) for the enzyme–substrate complexes are roughly 3-fold higher for hybrid Leu^{199}-SGT than those for wild-type SGT, resulting in almost indistinguishable specificity constants (k_{cat}/K_m) for the two enzymes. The

differences in the k_{cat} and K_m values for the native and hybrid SGT as a result of the Met199Leu substitution are in fact much smaller than variations among trypsins from different species.

Serine proteases play key roles in a great variety of biological processes from simple protein degradation in the digestive track to specific cleavages in the highly regulated blood clotting system. As the best-known member, trypsin has been a classic model enzyme in the research aimed at understanding the catalytic mechanisms of serine proteases [46,47]. Arguably most intensively scrutinized and best understood, the mechanisms of serine proteases nonetheless have been the center of controversy among enzymologists for decades. At issues are the precise roles played by each member of the catalytic triad His^{57}–Asp^{102}–Ser^{195} [48]. Recent reports on the nature of the interaction between His^{57} and Asp^{102} have rekindled an intense and yet unsettled debate on the existence of low-barrier high-energy H-bonds in enzyme catalysis [49–53]. This has motivated us to develop a methodology by which carefully designed non-natural analogs of His^{57} can be readily incorporated into trypsin to study how catalysis can be influenced by altered electrostatic interactions in the serine protease. Clearly, the hybrid Met199Leu SGT is an ideal model enzyme for investigating the role of His^{57} in atomic detail by using synthetic His analogs.

In conclusion, we have developed a facile approach for the biosynthesis of a hybrid trypsin, which will allow for convenient incorporation of non-coded amino acid analogs into the N-terminal region of the protein. This will provide the possibility of new approaches to the efforts to better understand the molecular mechanisms of catalysis for serine proteases by using non-natural protein building blocks. Incorporation of novel His57 analogs to systematically alter the basicity of the group is currently underway in our laboratory. Our strategy to process a recombinant protein for native chemical ligation via CNBr cleavage of a genetically engineered and subsequently glutathione protected Met–Cys peptide bond is simple and robust, and can be generalized for many different protein systems where the N-terminus is the focal interest. When multi-segment chemical ligation is used, the N-terminal region targeted by chemical modifications can be easily extended to cover more than 100 amino acid residues. The hybridization approach, in many ways, is complementary to the expressed protein ligation technique [21], which is better suited for substitutions in the C-terminal region of proteins.

Acknowledgments

We thank Dr. Thor Borgford from Simon Fraiser University for providing the cloning plasmid for SGT.

References

[1] S.B. Kent, Chemical synthesis of peptides and proteins, Annu. Rev Biochem. 57 (1988) 957–989.
[2] A. Wlodawer, M. Miller, M. Jaskolski, B.K. Sathyanarayana, E. Baldwin, I.T. Weber, L.M. Selk, L. Clawson, J. Schneider, S.B. Kent, Conserved folding in retroviral proteases: crystal structure of a synthetic HIV-1 protease, Science 245 (1989) 616–621.
[3] R.C. Milton, S.C. Milton, S.B. Kent, Total chemical synthesis of a D-enzyme: the enantiomers of HIV-1 protease show reciprocal chiral substrate specificity [corrected], Science 256 (1992) 1445–1448.
[4] C.J. Noren, S.J. Anthony_Cahill, M.C. Griffith, P.G. Schultz, A general method for site-specific incorporation of unnatural amino acids into proteins, Science 244 (1989) 182–188.
[5] D. Mendel, J.A. Ellman, Z. Chang, D.L. Veenstra, P.A. Kollman, P.G. Schultz, Probing protein stability with unnatural amino acids, Science 256 (1992) 1798–1802.
[6] J.D. Bain, C. Switzer, A.R. Chamberlin, S.A. Benner, Ribosome-mediated incorporation of a non-standard amino acid into a peptide through expansion of the genetic code, Nature 356 (1992) 537–539.
[7] S.H. Beiboer, B. van_den_Berg, N. Dekker, R.C. Cox, H.M. Verheij, Incorporation of an unnatural amino acid in the active site of porcine pancreatic phospholipase A2. Substitution of histidine by 1,2,4-triazole-3-alanine yields an enzyme with high activity at acidic pH, Protein Eng. 9 (1996) 345–352.
[8] N. Budisa, C. Minks, S. Alefelder, W. Wenger, F. Dong, L. Moroder, R. Huber, Toward the experimental codon reassignment in vivo: protein building with an expanded amino acid repertoire, FASEB J. 13 (1999) 41–51.
[9] F. Hamano_Takaku, T. Iwama, S. Saito_Yano, K. Takaku, Y. Monden, M. Kitabatake, D. Soll, S. Nishimura, A mutant *Escherichia coli* tyrosyl-tRNA synthetase utilizes the unnatural amino acid azatyrosine more efficiently than tyrosine, J. Biol. Chem. 275 (2000) 40324–40328.
[10] J.C.M. van Hest, K.L. Kiick, D.A. Tirrell, Efficient incorporation of unsaturated methionine analogues into proteins in vivo, J. Am. Chem. Soc. 122 (2000) 1282–1288.
[11] T.L. Bigler, W. Lu, S.J. Park, M. Tashiro, M. Wieczorek, R. Wynn, M. Laskowski, Binding of amino acid side chains to preformed cavities: interaction of serine proteinases with turkey ovomucoid third domains with coded and noncoded P1 residues, Protein Sci. 2 (1993) 786–799.
[12] D.Y. Jackson, J. Burnier, C. Quan, M. Stanley, J. Tom, J.A. Wells, A designed peptide ligase for total synthesis of ribonuclease A with unnatural catalytic residues, Science 266 (1994) 243–247.
[13] C.J. Wallace, Peptide ligation and semisynthesis, Curr. Opin. Biotechnol. 6 (1995) 403–410.
[14] P.E. Dawson, T.W. Muir, I. Clark_Lewis, S.B. Kent, Synthesis of proteins by native chemical ligation, Science 266 (1994) 776–779.
[15] P.E. Dawson, S.B. Kent, Synthesis of native proteins by chemical ligation, Annu. Rev. Biochem. 69 (2000) 923–960.
[16] D.W. Low, M.G. Hill, M.R. Carrasco, S.B. Kent, P. Botti, Total synthesis of cytochrome b562 by native chemical ligation using a removable auxiliary, Proc. Natl. Acad. Sci. USA 98 (2001) 6554–6559.
[17] J.P. Tam, Y.A. Lu, C.F. Liu, J. Shao, Peptide synthesis using unprotected peptides through orthogonal coupling methods, Proc. Natl. Acad. Sci. USA 92 (1995) 12485–12489.
[18] L. Wang, A. Brock, B. Herberich, P.G. Schultz, Expanding the genetic code of *Escherichia coli*, Science 292 (2001) 498–500.
[19] L. Wang, A. Brock, P.G. Schultz, Adding L-3-(2-naphthyl)alanine to the genetic code of *E. coli*, J. Am. Chem. Soc. 124 (2002) 1836–1837.

[20] L.E. Canne, P. Botti, R.J. Simon, Y. Chen, E.A. Dennis, S.B. Kent, Chemical protein synthesis by solid phase ligation of unprotected peptide segments, J. Am. Chem. Soc. 121 (1999) 8720–8727.
[21] T.W. Muir, D. Sondhi, P.A. Cole, Expressed protein ligation: a general method for protein engineering, Proc. Natl. Acad. Sci. USA 95 (1998) 6705–6710.
[22] G.J. Cotton, T.W. Muir, Peptide ligation and its application to protein engineering, Chem. Biol. 6 (1999) R247–R256.
[23] G.J. Cotton, T.W. Muir, Generation of a dual-labeled fluorescence biosensor for Crk-II phosphorylation using solid-phase expressed protein ligation, Chem. Biol. 7 (2000) 253–261.
[24] R. Xu, B. Ayers, D. Cowburn, T.W. Muir, Chemical ligation of folded recombinant proteins: segmental isotopic labeling of domains for NMR studies, Proc. Natl. Acad. Sci. USA 96 (1999) 388–393.
[25] M.A. Dwyer, W. Lu, J.J. Dwyer, A.A. Kossiakoff, Biosynthetic phage display: a novel protein engineering tool combining chemical and genetic diversity, Chem. Biol. 7 (2000) 263–274.
[26] P. Carter, B. Nilsson, J.P. Burnier, D. Burdick, J.A. Wells, Engineering subtilisin BPN' for site-specific proteolysis, Proteins 6 (1989) 240–248.
[27] P. Carter, J.A. Wells, Engineering enzyme specificity by "substrate-assisted catalysis", Science 237 (1987) 394–399.
[28] T.A. Kunkel, J.D. Roberts, R.A. Zakour, Rapid and efficient site-specific mutagenesis without phenotypic selection, Methods Enzymol. 154 (1987) 367–382.
[29] M. Schnolzer, P. Alewood, A. Jones, D. Alewood, S.B. Kent, In situ neutralization in Boc-chemistry solid phase peptide synthesis. Rapid, high yield assembly of difficult sequences, Int. J. Pept. Protein Res. 40 (1992) 180–193.
[30] W. Lu, I. Apostol, M.A. Qasim, N. Warne, R. Wynn, W.L. Zhang, S. Anderson, Y.W. Chiang, E. Ogin, I. Rothberg, K. Ryan, M. Laskowski, Binding of amino acid side-chains to S1 cavities of serine proteinases, J. Mol. Biol. 266 (1997) 441–461.
[31] D. Nohara, H. Sugiura, H. Sakakibara, M. Matsubara, S. Kojima, K. Miura, T. Sakai, High performance in refolding of *Streptomyces griseus* trypsin by the aid of a mutant of *Streptomyces subtilisin* inhibitor designed as trypsin inhibitor, J. Biochem. 125 (1999) 343–347.
[32] P.H. Hirel, M.J. Schmitter, P. Dessen, G. Fayat, S. Blanquet, Extent of N-terminal methionine excision from *Escherichia coli* proteins is governed by the side-chain length of the penultimate amino acid, Proc. Natl. Acad. Sci. USA 86 (1989) 8247–8251.
[33] F. Baneyx, Recombinant protein expression in *Escherichia coli*, Curr. Opin. Biotechnol. 10 (1999) 411–421.
[34] J.P. Etchegaray, M. Inouye, Translational enhancement by an element downstream of the initiation codon in *Escherichia coli*, J. Biol. Chem. 274 (1999) 10079–10085.
[35] R.J. Read, M.N. James, Refined crystal structure of *Streptomyces griseus* trypsin at 1.7 A resolution, J. Mol. Biol. 200 (1988) 523–551.
[36] D. Shortle, W.E. Stites, A.K. Meeker, Contributions of the large hydrophobic amino acids to the stability of staphylococcal nuclease, Biochemistry 29 (1990) 8033–8041.
[37] L. Serrano, J.T. Kellis, P. Cann, A. Matouschek, A.R. Fersht, The folding of an enzyme. II. Substructure of barnase and the contribution of different interactions to protein stability, J. Mol. Biol. 224 (1992) 783–804.
[38] J. Thomson, G.S. Ratnaparkhi, R. Varadarajan, J.M. Sturtevant, F.M. Richards, Thermodynamic and structural consequences of changing a sulfur atom to a methylene group in the M13Nle mutation in ribonuclease-S, Biochemistry 33 (1994) 8587–8593.
[39] B.W. Matthews, Studies on protein stability with T4 lysozyme, Adv. Protein Chem. 46 (1995) 249–278.
[40] E. Gross, Methods Enzymol. 11 (1967) 238–255.
[41] E. Gross, B. Witkop, J. Am. Chem. Soc. 83 (1961) 1510–1511.
[42] R. Kaiser, L. Metzka, Enhancement of cyanogen bromide cleavage yields for methionyl-serine and methionyl-threonine peptide bonds, Anal. Biochem. 266 (1999) 1–8.
[43] N. Doyen, C. Lapresle, Partial non-cleavage by cyanogen bromide of a methionine–cystine bond from human serum albumin and bovine alpha-lactalbumin, Biochem. J. 177 (1979) 251–254.
[44] D.R. Goodlett, F.B. Armstrong, R.J. Creech, R.B. van_Breemen, Formylated peptides from cyanogen bromide digests identified by fast atom bombardment mass spectrometry, Anal. Biochem. 186 (1990) 116–120.
[45] H.S. Duewel, J.F. Honek, CNBr/formic acid reactions of methionine- and trifluoromethionine-containing λ lysozyme: probing chemical and positional reactivity and formylation side reactions by mass spectrometry, J. Protein Chem. 17 (1998) 337–350.
[46] J.J. Perona, C.S. Craik, Structural basis of substrate specificity in the serine proteases, Protein Sci. 4 (1995) 337–360.
[47] J.J. Perona, C.S. Craik, Evolutionary divergence of substrate specificity within the chymotrypsin-like serine protease fold, J. Biol. Chem. 272 (1997) 29987–29990.
[48] G. Dodson, A. Wlodawer, Catalytic triads and their relatives, Trends Biochem. Sci. 23 (1998) 347–352.
[49] P.A. Frey, S.A. Whitt, J.B. Tobin, A low-barrier hydrogen bond in the catalytic triad of serine proteases, Science 264 (1994) 1927–1930.
[50] E.L. Ash, J.L. Sudmeier, E.C. De_Fabo, W.W. Bachovchin, A low-barrier hydrogen bond in the catalytic triad of serine proteases?. Theory versus experiment, Science 278 (1997) 1128–1132.
[51] W.W. Cleland, P.A. Frey, J.A. Gerlt, The low barrier hydrogen bond in enzymatic catalysis, J. Biol. Chem. 273 (1998) 25529–25532.
[52] A. Warshel, Electrostatic origin of the catalytic power of enzymes and the role of preorganized active sites, J. Biol. Chem. 273 (1998) 27035–27038.
[53] A. Warshel, A. Papazyan, Energy considerations show that low-barrier hydrogen bonds do not offer a catalytic advantage over ordinary hydrogen bonds, Proc. Natl. Acad. Sci. USA 93 (1996) 13665–13670.