Engineering the Primary Substrate Specificity of *Streptomyces griseus* Trypsin†

Michael J. Page,‡ Sui-Lam Wong,§ Jeff Hewitt,‡ Natalie C. J. Strynadka,‡ and Ross T. A. MacGillivray*,‡

Department of Biochemistry and Molecular Biology, University of British Columbia, Vancouver, British Columbia V6T 1Z3, Canada, and Department of Biological Sciences, University of Calgary, Calgary, Alberta T2N 1N4, Canada

Received March 17, 2003; Revised Manuscript Received May 22, 2003

ABSTRACT: *Streptomyces griseus* trypsin (SGT) was chosen as a model scaffold for the development of serine proteases with enhanced substrate specificity. Recombinant SGT has been produced in a *Bacillus subtilis* expression system in a soluble active form and purified to homogeneity. The recombinant and native proteases have nearly identical enzymatic properties and structures. Four SGT mutants with alterations in the S1 substrate binding pocket (T190A, T190P, T190S, and T190V) were also expressed. The T190P mutant demonstrated the largest shift to a preference for Arg versus Lys in the P1 site. This was shown by a minor reduction in catalytic activity toward an Arg-containing substrate (k_{cat} reduction of 25%). The crystal structures of the recombinant SGT and the T190P mutant in a complex with the inhibitor benzamidine were obtained at high resolution (∼1.9 Å). The increase in P1 specificity, achieved with minimal effect on the catalytic efficiency, demonstrates that the T190P mutant is an ideal candidate for the design of additional substrate specificity engineered into the S2 to S4 binding pockets.

Substrate specificity is a key concept in the analysis of serine proteases. Sequence analysis studies show that the S1 family of trypsin-like enzymes probably evolved from a common ancestral gene (*1*). The culmination of incremental evolutionary steps led to the appearance of a number of highly specific proteases such as those found in the vertebrate blood coagulation cascades. These proteases fulfill regulatory roles in cellular processes that are distinct from their more primitive roles as degradative and protective enzymes (*2*). Much data have been collected on the specificity determinants of serine proteases, resulting in a classification system based on their primary specificity [denoted as the S1 pocket (*3*)].[1] The specificity of trypsin-like enzymes at the S1 pocket is largely defined by the presence of a negatively charged side chain at position 189 (*4*). Optimal binding of the positively charged substrate (Arg or Lys side chains) to residue 189 is mediated by residue 190 (*5*, *6*). In trypsin-like serine proteases, position 190 is occupied by a limited number of amino acids. Degradative proteases with low primary specificity (Arg:Lys preference of 4:1) display Gln, Thr, or Ser at position 190, whereas proteases with high primary specificity (Arg:Lys preferences greater than 7:1) contain Ala or Ser at this position (*7*). Previous work shows that mutagenesis of position 190 can be used to manipulate the substrate specificity of trypsin to favor cleavage after either Arg or Lys side chains (*3*, *6*, *8*).

Streptomyces griseus trypsin (SGT)[2] was initially purified from Pronase, a commercial preparation of secreted proteases, and characterized on the basis of its hydrolysis of N^{α}-benzoyl-L-arginine ethyl ester (BAEE) and casein and its inhibition by soybean trypsin inhibitor (*9*). Further characterization of its specificity and inhibition identified SGT as a typical broad specificity trypsin-like serine protease of the S1 family that hydrolyzes polypeptide chains on the C-terminal side of basic residues (Arg and Lys). On the basis of sequence alignments, SGT is more similar to bovine trypsin than to other bacterial serine proteases (*10*−*12*). The structure of SGT was subsequently determined by X-ray crystallography and refined to 1.7 Å and showed a three-dimensional fold that is also more similar to mammalian serine proteases than bacterial proteases (*13*, *14*). Like other bacterial proteases, SGT is synthesized as a precursor that is cleaved into a zymogen form during the secretion process. A short four amino acid propeptide in SGT must then be removed to generate the active protease. Processing of the SGT precursor is likely nonautocatalytic as the carboxy-terminal residue of this region is proline and is not a good substrate for the protease (*12*). Although similarities exist at the sequence and structural level, SGT differs from its mammalian homologues in its reduced number of amino acid insertions in the polypeptide chain when aligned by either sequence or structure. Moreover, SGT contains only three disulfide bonds rather than the five or six typically observed in the S1 family of proteases. These differences suggest that SGT could be used as a model scaffold to study the substrate specificity and other properties demonstrated by mammalian proteases.

† M.J.P. was supported by a graduate fellowship from Canadian Blood Services. This work was supported in part by a grant from the CBS−CIHR Program in Blood Utilization and Conservation Research awarded to R.T.A.M.

* Corresponding author. Tel: (604) 822-3027. Fax: (604) 822-4364. E-mail: macg@interchange.ubc.ca.

‡ University of British Columbia.

§ University of Calgary.

[1] Chymotrypsin numbering system is used for amino acid positions (*3*).

[2] Abbreviations: AMC, 7-amino-4-methylcoumarin; SGT, native *Streptomyces griseus* trypsin; bSGT, *Bacillus subtilis* derived recombinant *S. griseus* trypsin; PMSF, phenylmethanesulfonyl fluoride.

10.1021/bi0344230 CCC: $25.00
Published on Web 07/12/2003

Similar to degradative vertebrate trypsin-like enzymes, SGT demonstrates a primary substrate preference of Arg:Lys of 4:1. In this study, we utilized a *Bacillus subtilis* expression system to produce soluble SGT that is fully active. We also constructed mutants of SGT that change the arginine to lysine preference (Arg:Lys) of the enzyme. The T190P mutant of SGT is considerably more active and less Arg-specific when compared with the previously published S190P mutant created in rat anionic trypsin. The previously reported mutant demonstrated a high degree of Arg selectivity (Arg:Lys 135:1) with a 10-fold reduction in catalytic efficiency (*5*). These results further our understanding of the primary substrate specificity of trypsin-like enzymes. On the basis of the ease of production and purification of the recombinant protein in our *B. subtilis* expression system, SGT is an ideal scaffold for the introduction of additional mutations to enhance the substrate specificity of the S2 to S4 binding pockets.

MATERIALS AND METHODS

Plasmids, Bacterial Strains, and Growth Conditions. *Escherichia coli* was grown using standard methods (*15*). *B. subtilis* strain WB700 (*16*) was grown in super-rich medium (*17*) or on tryptose blood agar base (Difco) at 37 °C. For the *B. subtilis* carrying plasmid pWB980 (*18*), kanamycin was added to a final concentration of 10 μg mL^{-1} in liquid and solid media.

DNA Manipulation. Procedures for genomic [*S. griseus* (ATCC 10137)] and plasmid DNA manipulation were carried out using established protocols (*15*). Plasmid DNA was purified using a QIAprep spin miniprep kit (Qiagen). Enzymes were obtained from New England Biolabs and Roche Molecular Biochemicals. For PCR amplification of the *SprT* gene encoding SGT, the following oligonucleotides were designed to maintain the reading frame of the sacB signal peptide present in plasmid pWB980: 5′-ggaagcttttgcaGTCGTCGGCGGAACCCGCGCGG-3′ and 5′-ggtctagattaGAGCGTGCGGGCGGCCGAGG-3′ (restriction sites are underlined, and the *SprT* gene specific sequences are given in upper case). The PCR fragment was first cloned into pBluescript KS+ (Stratagene) and then subcloned into pWB980 using the *Hin*dIII and *Xba*I restriction enzyme sites contained in the oligonucleotide primers. Transformation of *B. subtilis* strain WB700 was performed as described previously (*19*). Mutagenesis was performed on the SGT cloned into pBluescript using a QuikChange site-directed mutagenesis kit (Stratagene) as described by the manufacturer. DNA sequence analysis of the cloned gene and mutants was performed using the Big Dye terminator kit and analyzed on an ABI 3700 DNA sequencer (Applied Biosystems).

Protein Purification. To purify the native and recombinant proteases to homogeneity, a purification strategy was developed at low pH to minimize autolysis. After centrifugation to remove cellular debris (5000*g*, 1 h), recombinant SGT was purified from the supernatant of 1 L of *B. subtilis* WB700 culture. Sequential ammonium sulfate fractionation was carried out at 30% and 85% saturation. The 85% $(NH_4)_2SO_4$ fraction pellet was resuspended in 20 mM sodium acetate buffer, pH 4.5, dialyzed against the same buffer, and applied to a column (15 cm × 1.5 cm) of SP-Sepharose Fast Flow (Amersham Pharmacia). After extensive washing with 20 mM sodium acetate buffer containing 50 mM NaCl, pH 4.5, the bound proteins were eluted with 20 mM sodium acetate buffer, pH 4.5, containing 150 mM NaCl. The active fractions were pooled and applied to a benzamidine−Sepharose 4 Fast Flow column (8 cm × 0.75 cm) (Amersham Pharmacia). The column was washed with 20 mM sodium acetate buffer containing 500 mM NaCl, pH 4.5, and the enzyme was eluted in the same buffer containing in addition 40 mM benzamidine hydrochloride (Sigma). Fractions containing active protease were pooled, concentrated using a 10000 NMWL Ultrafree-4 centrifugal filter unit (Millipore) and dialyzed against 10 mM Tris-HCl buffer containing 150 mM NaCl and 20 mM $CaCl_2$, pH 7.6. Gel filtration through a column (45 cm × 0.75 cm) of Sephadex G-75 (Amersham Pharmacia) was performed using 10 mM Tris-HCl containing 150 mM NaCl and 20 mM $CaCl_2$, pH 7.6. Native SGT was isolated from 1 g of extracellular filtrate from *S. griseus* (Sigma) using the same protocol. The final protein concentration was determined by UV absorbance at 280 nm, using the extinction coefficient of 37100 M^{-1} cm^{-1} (*20*), or by a BCA protein assay kit (Pierce). Active site titration was performed using 4-nitrophenyl *p*′-guanidinobenzoate and a standard curve of *p*-nitrophenol (Sigma). Sodium dodecyl sulfate−polyacrylamide gel electrophoresis and Coomassie Blue staining were performed according to standard procedures (*15*). N-Terminal protein microsequence analysis was performed by the University of Victoria Genome BC Proteomics Centre (Victoria, Canada). Electrospray−mass spectrometry was carried out on a PE-Sciex API 300 triple quadrupole mass spectrometer (Sciex) equipped with an Ionspray ion source.

Kinetic Analysis. Kinetic analysis was performed in 10 mM Tris-HCl buffer containing 150 mM NaCl, 20 mM $CaCl_2$, and 0.1% PEG 8000, pH 7.6. A standard of 7-amino 4-methylcoumarin (AMC) was used to quantify the rates of hydrolysis of the fluorogenic substrates Tos-Gly-Pro-Arg-AMC and Tos-Gly-Pro-Lys-AMC (Bachem). A minimum of six substrate concentrations ranging from 1 to 500 μM was used. Enzyme concentrations ranged from 0.5 to 10 nM. Benzamidine concentrations ranged from 5 to 200 μM. Nonlinear regression of the initial reaction rates and calculation of the kinetic parameters were performed using the Graphpad Prism 3.0 software (Graphpad).

Crystallization. Previous crystals of the native SGT were obtained through batch crystallization using $(NH_4)_2SO_4$ (*14*). In this study, proteins were crystallized using similar conditions [10−15 mg/mL protein, 1.5 M $(NH_4)_2SO_4$, 10 mM calcium acetate, pH 6.2] except hanging drop vapor diffusion was utilized where the reservoir contained 1.55 M $(NH_4)_2SO_4$. The T190P mutant was crystallized using similar conditions except the sample buffer contained in addition 25 mM benzamidine hydrochloride (Sigma). Crystals appeared in 2−3 weeks to dimensions of approximately 0.3 × 0.3 × 0.3 mm. Data were collected at 100 K (Oxford Cryostream) with an Mar345 detector mounted on a Rigaku RU-200 X-ray generator (50 kV, 100 mA) with Osmic focusing mirrors. Crystals were soaked briefly in 20% glycerol and 2.2 M $(NH_4)_2SO_4$ for cryoprotection prior to data collection. Data were processed using the HKL package and refined using CNS version 1.1 in combination with Xtalview (*21*−*23*). The previously reported native SGT structure (PDB entry 1SGT) was used as a model for rigid body refinement of both structures (*14*).

Table 1: Purification Table for Recombinant SGT (bSGT) from *B. subtilis* Extracellular Supernatant[a]

	vol (mL)	total protein (mg)	total activity (mmol/s)	specific activity (μmol s^{-} mg^{-1})	purification (*x*-fold)	yield (%)
media	1000	16000	59	4	1	100
$(NH_4)_2SO_4$ precipitate	50	400	53	134	34	90
SP-Sepharose	15	18.3	51	2774	694	87
benzamidine−Sepharose	5	12.4	49	3856	964	81
G-75 Superdex	4	12.2	48	3955	989	81

[a] Similar yields were obtained for each mutant studied. Activity was measured by the hydrolysis of the chromogenic substrate Bz-Ile-Glu-Arg-pNA at 40 μM in 10 mM Tris-HCl and 20 mM $CaCl_2$, pH 7.6.

Table 2: Electrospray Mass Spectrometry Analysis of Native (SGT), Recombinant (bSGT), and the Four Mutants of SGT

	theoretical mass (amu)	observed mass (amu)
SGT	23105.9	23106.9
bSGT	23105.9	23107.0
T190S	23091.9	23090.0
T190V	23103.9	23102.0
T190A	23075.9	23079.0
T190P	23101.9	23098.8

RESULTS

Production and Purification of Recombinant SGT. Initial attempts to produce recombinant SGT in various strains of *E. coli* yielded insoluble protein in the form of inclusion bodies. *B. subtilis* strain WB700 was used to produce soluble protein and was found to secrete active protease into the extracellular medium. For both wild-type and mutant enzymes, yields of >15 mg/L of culture medium were obtained within 24 h of growth at 37 °C. The four-step purification typically produced 10−15 mg/L of *B. subtilis* culture with an overall yield of 80% (Table 1).

Recombinant proteases were purified to homogeneity prior to analysis. Purity was assessed by several different criteria. By electrospray ionization mass spectrometry, the expected and observed masses were nearly identical and confirmed the presence of the mutations previously shown by DNA sequence analysis (Table 2). For each protein the integrated data from the m/z^+ fragments yielded a single unambiguous peak with minimal background. In addition, when the protease was treated with phenylmethanesulfonyl fluoride (PMSF), SDS−PAGE showed a single band of approximately 23 kDa, which is similar the expected molecular mass of the protein. In the absence of the inhibitor two major autolytic fragments, 15.1 and 12.8 kDa, were evident (Figure 1). Amino-terminal sequence analysis revealed that the wild-type enzyme and mutants had a single unambiguous sequence, NH_2-Val-Val-Gly-Gly-Thr-Arg, corresponding to the published SGT sequence (*12*). Together with the protein assays and active site titration data, these results suggest that the final protein preparation was greater than 99% pure.

Kinetic Analysis. Two fluorogenic peptide substrates, Tos-Gly-Pro-Arg-AMC and Tos-Gly-Pro-Lys-AMC, were used to monitor the P1 Arg:Lys substrate discrimination of the native, recombinant ,and mutant proteases (Table 3). Mutants T190P and T190A demonstrated a significant increase, compared to the wild-type enzyme, in P1 preference for Arg over Lys. The relative Arg:Lys catalytic efficiencies are 18:1 and 8:1, respectively. All four mutants showed increased K_m values for the substrates tested, suggesting that the S1 pocket of SGT is optimized for substrate binding, a feature that has been observed in other trypsin-like enzymes (*5*, *6*, *8*). A similar trend of fold differences for the K_i value of the small molecule inhibitor benzamidine was observed, with the exception of the T190P mutant (Table 4).

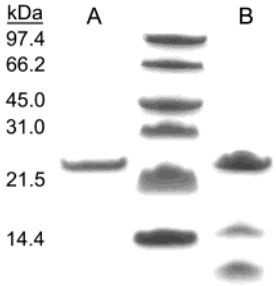

FIGURE 1: SDS−PAGE of purified recombinant SGT. Lane A: bSGT (0.5 μg) inactivated with PMSF prior to addition of SDS−PAGE loading buffer and boiling. Lane B: bSGT without PMSF inhibitor prior to addition of SDS−PAGE loading buffer and boiling. The central lane of the gel contains low-range protein molecular mass standards (Bio-Rad) whose masses are given on the left-hand side of the gel. The gel was stained with Coomassie Brilliant Blue.

Crystallization. X-ray diffraction data were obtained for the recombinant wild-type and T190P mutant of SGT at 1.5 and 1.9 Å resolution, respectively. Data collection and refinement statistics are given in Table 5. Both recombinant proteases crystallized in the $C222_1$ space group and contained one molecule per asymmetric unit and a Matthews coefficient of 2.2−2.3 $Å^3$/Da. The structures are deposited in the PDB database as 1OSS and 1OS8, for the wild-type recombinant and T190P mutant, respectively. During refinement, low R_{cryst} and R_{free} values were obtained (0.19 and 0.22 for the recombinant wild type and 0.17 and 0.21 for the T190P mutant, respectively). These were accompanied by excellent stereochemistry, indicating high quality models. Inspection of the two Ramachandran plots revealed that all non-glycine backbone atoms are in allowed regions, with only Asn178 adopting a conformation in the generously allowed region. The overall *B*-factors for the polypeptide atoms were low (~13 $Å^2$), and regions with high *B*-factors were limited to solvent-exposed regions that are not involved in crystal packing. The structures of SGT and mutant are of the highest quality to date likely due to the decreased radiation damage as the present crystals were analyzed under cryogenic conditions with a shorter collection time.

Comparison of the Two Crystal Structures to Pronase-Derived SGT (1SGT). Crystal structures of the wild-type recombinant SGT and T190P mutant were solved to 1.5 and 1.9 Å, respectively. All amino acid residues in the two models were clearly identified and positioned. In 1SGT (*14*), several residues had weak or absent side chain density. In the present structures, most of these densities were clearly resolved, although several residues lacked density in the terminal atoms of their side chains (Thr20, Gln75, Lys82, Thr98, Ser236, and Arg243), indicating disorder of these solvent-exposed atoms. Residues 77 and 79 were incorrectly modeled as Gly and Ala in 1SGT but are two Ser residues by DNA sequence analysis (*12*). These Ser residues were clearly resolved in the current electron density maps. One

Table 3: Preferences for Arginine over Lysine among Mutant SGT Enzymes[a]

	Tos-Gly-Pro-Arg-AMC			Tos-Gly-Pro-Lys-AMC			
	k_{cat} (min^{-1})	K_m (μM)	k_{cat}/K_m ($min^{-1}\,\mu M^{-1}$)	k_{cat} (min^{-1})	K_m (μM)	k_{cat}/K_m ($min^{-1}\,\mu M^{-1}$)	S_R/S_K[b]
SGT	4880 ± 410	2.3 ± 0.2	2122	1670 ± 120	3.6 ± 0.4	464	4.6
bSGT	4570 ± 1210	2.0 ± 0.2	2285	1520 ± 60	3.2 ± 0.2	475	4.8
T190S	6036 ± 561	6.1 ± 0.6	990	2584 ± 141	10.1 ± 1.2	256	3.9
T190V	2300 ± 177	224 ± 20	11	791 ± 13	231 ± 2	3.4	3.2
T190A	4950 ± 470	12.9 ± 1.0	384	192 ± 32	4.4 ± 0.6	44	8.7
T190P	3610 ± 130	67 ± 5	54	527 ± 2	166 ± 2	3	18

[a] Arg:Lys preference was measured by amidolytic activity of the native (SGT), recombinant (bSGT) and mutants of SGT using two fluorogenic peptides, Tos-Gly-Pro-Arg-AMC and Tos-Gly-Pro-Lys-AMC ($n \geq 3$). Errors provided are standard deviation values. [b] S_R/S_K = (Tos-Gly-Pro-Arg-AMC k_{cat}/K_m)/(Tos-Gly-Pro-Lys-AMC k_{cat}/K_m).

Table 4: K_i Values of Benzamidine Bound to Recombinant SGT and the Four Mutant Forms ($n \geq 3$)[a]

	K_i (μM)
bSGT	2.7 ± 0.8
T190S	4.8 ± 0.8
T190V	197 ± 37
T190A	9.9 ± 0.7
T190P	16.4 ± 2.2

[a] Errors provided are standard deviation values.

Table 5: Data Collection and Refinement Statistics[a]

	bSGT	T190P
data collection		
resolution (Å)	1.5 (1.55−1.65)	1.9 (1.93−2.05)
total observations	25923	14471
completeness (%)	84.1 (74.1)	89.7 (79.1)
average redundancy	2.8	7.1
$I/\sigma I$	20.4 (5.1)	44.4 (25.2)
R_{merge} (%)	4.1 (18.2)	3.1 (5.6)
refinement statistics		
space group	$C222_1$	$C222_1$
cell dimensions (a, b, c) (Å)[b]	50.04, 69.82, 119.65	50.08, 69.60, 119.83
molecules per asym unit	1	1
R_{cryst}	0.196	0.167
R_{free}	0.222	0.212
protein atoms[c]	1623	1632
solvent atoms per asym unit	211	193
av B-factor for protein (Å^2)	12.9	12.8
av B-factor for water (Å^2)	21.0	21.4
occupancy of Ca^{2+} (B Å^2)	0.53 (14.98)	0.43 (17.50)
bond length deviations (Å)	0.007	0.007
bond angle deviations (deg)	1.4	1.4

[a] Statistics for the highest resolution shell are given in parentheses. [b] $\alpha = \beta = \gamma = 90°$. [c] Including alternate side chain conformations.

and two sulfate ions from the crystallization buffer were visible in the electron density maps in the recombinant wild-type and T190P mutant, respectively, and were included in the final model. In both structures, a sulfate ion was observed in the oxyanion hole of the substrate binding site. The position of this sulfate is conserved in anionic salmon trypsin (PDB entry 1BIT), bovine trypsin (1TLD), and porcine pancreatic elastase (3EST) (*24*−*26*). A second sulfate ion was observed in the T190P structure and forms an electrostatic interaction with Arg145. Alternate conformations were observed for Gln192, which either points into the solvent or forms a pair of hydrogen bonds with the backbone NH of Gly148 of an adjacent SGT molecule. In 1SGT the same residue was noted as having high mobility (*14*). The overall differences between the native (1SGT) and wild-type recombinant enzyme are minor (Figure 2), with a root mean square deviation of all C_α backbone atoms of 0.27 Å.

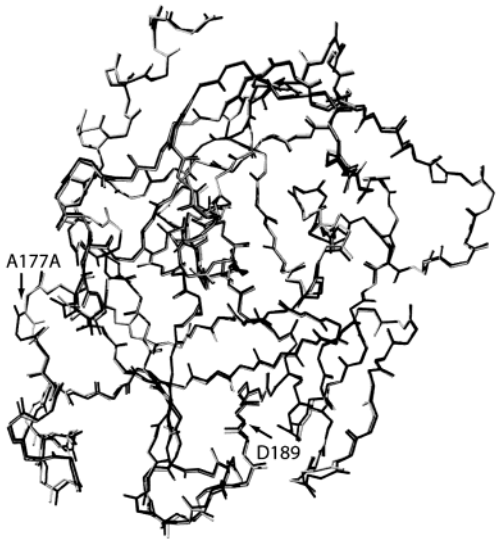

FIGURE 2: Superimposition of the C_α traces of native (1SGT) and recombinant wild-type SGT. The rms deviation between the C_α backbone atoms of the two structures is 0.27 Å. The largest deviation in the backbone occurs at the 174-loop where Ala177a adopts a 180° conformational change comparative to the native structure.

Ca^{2+} Binding Site of B. subtilis Derived SGT. In the previously reported structure (1SGT) the calcium binding site consisted of the Asp165 and Glu230 carboxylate groups, two well-ordered water molecules, and the carbonyl oxygen atoms of residues Ala177a and Glu180 (Figure 3) (*14*). In the present structures, Asp165 was shown to adopt a different conformation than observed previously. In the wild-type recombinant SGT structure, the carboxylate of Asp165 forms a bidentate electrostatic interaction with Arg169 rather than the structural Ca^{2+} ion. In the T190P model, Asp165 adopts a pair of alternative conformations, either facing the Ca^{2+} ion or toward Arg169. Modeling of the two positions at half-occupancy generates a lower B-factor for the conformation involved in the electrostatic interaction with Arg169. In both structures, the carbonyl oxygen of Ala177a was oriented away from the Ca^{2+} ion and does not appear to be involved in the interaction. Three ordered water molecules are associated with the ion in the present wild-type structure, rather than the two previously observed. In contrast, a single disordered water molecule with a high B-factor (41.4 Å^2) was found near the Ca^{2+} ion in the T190P crystal structure. In other trypsin-like proteases, Ca^{2+} ions have been found

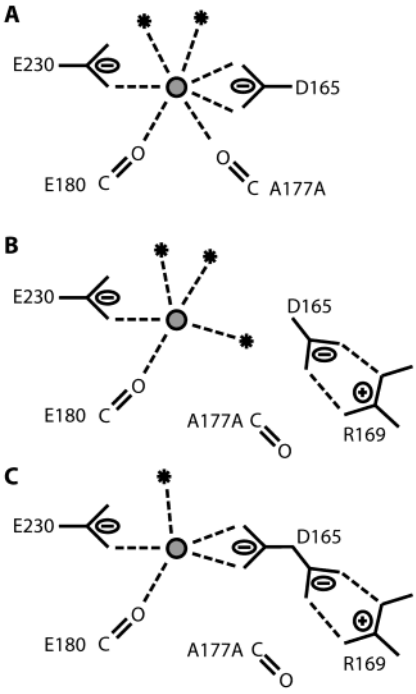

FIGURE 3: Comparison of the Ca^{2+} binding site in SGT enzymes: (A) native, (B) wild-type recombinant, and (C) T190P mutant of SGT. The number of water molecules (*) that coordinate the structural calcium ion (○), as well as the conformation of the amino acid ligands, differs in all three structures.

previously with high *B*-factors relative to the overall structure, suggesting less than full occupancy (*27*). The reduced occupancy of the ion is not surprising as a related enzyme, *Streptomyces erythraeus* trypsin, lacks Glu230, and no calcium is evident in the structure (*28*). Moreover, the suggested Ca^{2+} binding site in SGT is completely different than that observed in mammalian trypsin-like enzymes (*14*). When taken with previous observations that the calcium ion plays no role in catalysis but plays a role only in structural stability in low/high pH solutions, these data suggest that the calcium binding site has weak affinity for the ion (*10*).

Wild-Type Native and Recombinant SGT Substrate Binding. The base of the S1 pocket bears the negatively charged D189 residue that confers the primary specificity of the enzyme toward positively charged Arg or Lys side chains (*4*). On the basis of the structure of 1SGT, D189 is located at the base of a narrow cylindrical cleft that can accommodate these side chains (*14*). The slight preference for Arg over Lys in this pocket is due to the requirement for a bridging water molecule between the shorter lysyl side chain and D189 (*5*). The γ-OH of residue T190 interacts directly with the substrate via hydrogen bonding. Both Lys and Arg side chains adopt favorable conformations for interaction with D189 and the hydroxyl group of T190. Kinetic analysis of the native and recombinant SGT proteases demonstrated an Arg:Lys preference of 4:1. The kinetic parameters of the native and recombinant wild-type SGT are similar to previously reported values using the same pair of substrates (*5*).

DISCUSSION

Expression of Recombinant SGT. Previous studies demonstrated the ability to produce soluble trypsin-like enzymes in the periplasmic space of *E. coli* (*29*). In our experiments, however, *E. coli* was incapable of generating soluble SGT despite using a variety of plasmid constructs in a number of bacterial host strains. The presence of three disulfide bonds, high G+C% content of the *SprT* gene (70%), and toxicity of the recombinant protein are possible reasons for the lack of production of recombinant proteins in *E. coli* (*30*). To overcome these limitations, *B. subtilis* WB700, a strain that is deficient in seven proteases, was used to produce sufficient yields of recombinant SGT for kinetic and structural analysis. Unlike *E. coli*, the Gram-positive *B. subtilis* is capable of secreting proteins into the extracellular environment that facilitates rapid detection, purification, and analysis of recombinant proteins. The results of this study confirm the utility of *B. subtilis* as an alternative prokaryotic expression system to *E. coli* (*31*).

Expression levels of the four mutant proteases were comparable with the wild-type SGT, suggesting that the mutations were not detrimental to the overall stability and folding of the proteins. The four-step purification scheme was carried out at pH 4.5 to minimize autolytic breakdown during the processing and resulted in high yield (>80%) and purity (>99%) for all six enzymes characterized (Table 1). Expression levels of 10−15 mg/L of bacterial culture are significantly higher than the previously reported values for soluble protein expression of similar trypsin-like enzymes in *E. coli* (*29*, *32*). Moreover, a previous study reported expression of the recombinant protease as an inactive zymogen that requires activation by enterokinase (*33*). Due to low or absent yields of soluble trypsin in *E. coli*, recombinant trypsin-like enzymes have been produced previously in the form of inactive and insoluble inclusion bodies at typically >100 mg/L of culture. Although the insoluble protein is more abundant and readily purified, the need for refolding the recombinant protein in vitro limits these systems for high-throughput mutagenesis and screening of mutant proteases. Future studies will utilize a tagged version of the enzyme (such as a hexahistidine tag) to facilitate the rapid purification and analysis of mutant proteases.

T190S and the Loss of γ-CH_3. Degradative proteases involved in digestive and protective functions typically possess Ser or Thr residues at position 190. However, proteases exhibiting higher substrate specificity may also possess these residues (*7*). The T190S mutant of SGT demonstrated no significant increase in overall substrate specificity, yet a minor increase in catalytic activity (k_{cat} increase of 25%) and a 3-fold increase in K_m (Table 3) were observed for both the Arg- and Lys-containing substrates. Loss of the γ-CH_3 is unlikely to significantly increase the solvent accessibility of D189, and it is more likely that the increased mobility of the γ-OH results in the increased K_m values for both substrates. In eukaryotic trypsin-like enzymes that contain Ser at position 190, the space that would be occupied by a γ-CH_3 is filled by methyl groups from either residue 16 or 138 usually in the form of isoleucine side chains. In SGT, both of these residues are valine, which possesses one fewer methyl group.

T190V and the Effect of a Branched Side Chain. In contrast to the kinetics of T190S, the T190V mutant demonstrated a 2-fold reduction in k_{cat} in combination with nearly a 100-fold increase in K_m for both Arg- and Lys-containing substrates (Table 3). Replacement of the γ-OH with a methyl group removes the hydrogen-bonding capacity of the residue

and reduces the solvent accessibility of D189. This results in a destabilized transition state complex relative to the wild-type protease and a weaker electrostatic interaction between the substrate and enzyme. The 100-fold increase in K_m for Arg-containing substrates compared to the 70-fold increase in K_m for Lys-containing substrates indicates that the small increase in excluded volume has a more significant effect on the longer and bulkier arginyl side chain. The D189 and V190 pair has not been observed in naturally occurring trypsin-like enzymes likely due to the poor catalytic efficiency of this combination of side chains.

T190A and the Loss of γ-OH. A majority of vertebrate enzymes involved in physiological regulation, such as the coagulation factor serine proteases, possess Ala at position 190 and exhibit Arg:Lys substrate specificities ranging from 7:1 for coagulation factor Xa to greater than 14:1 for bovine thrombin (*6*). The T190A mutation in SGT demonstrates the molecular basis for the predominance of Ala at this position in highly specific proteases favoring a P1 Arg residue. Optimal rates of catalysis at low concentrations of substrate tend to be requisite characteristics of these vertebrate enzymes. Kinetic analysis of the T190A mutant revealed no change in k_{cat} for Arg-containing substrates but rather a 10-fold reduction in k_{cat} for Lys-containing substrates (Table 3). As noted previously, the increase in solvent accessibility of D189 should stabilize the transition state complex. However, the loss of hydrogen bonding (previously provided by the γ-OH) exhibits a more pronounced effect on the Lys-containing substrate due to its requirement of an ordered bridging water molecule to D189.

T190P. Unlike the T190A mutation, whose effects were predominantly on the k_{cat} of the Lys-containing substrate, the T190P mutation affected the K_m significantly. The K_m for the Arg-containing substrate was 35-fold higher than the wild type, compared to the 46-fold increase for the Lys-containing substrate (Table 3). Similarly, the reduction in k_{cat} for the Arg substrate (25%) was significantly less than the Lys substrate (66%). Together, these changes generate an overall Arg to Lys preference of 18 to 1. A previous report (*5*) analyzed the S190P mutant of rat anionic trypsin, which is analogous to the T190P constructed in SGT. When using the same pair of substrates used in this study, wild-type rat anionic trypsin exhibits a similar primary substrate specificity to SGT. However, the S190P mutation in rat anionic trypsin results in a highly specific protease that favors the Arg substrate 135-fold over the Lys-containing substrate. This increase in substrate specificity was combined with a greater than 10-fold reduction in k_{cat} for both Arg and Lys substrates. These authors suggested that Tyr228 may be involved in steric clashing with the proline ring at position 190, leading to reduced activity. In SGT, residue 228 is also Tyr, suggesting an alternate binding mode of this mutant. To address these discrepancies, the crystal structure of the T190P mutant in complex with the small molecule inhibitor benzamidine was investigated.

Binding of the benzamidine inhibitor to the T190P mutant is nearly identical to that observed in other trypsin-like proteases (*27*). The proline ring of residue 190 does not adopt a conformation that occludes the negatively charged carboxylate group of Asp189, nor does it conflict with Tyr228, suggesting that the previously characterized specificity of the S190P mutant in rat anionic trypsin was the result of

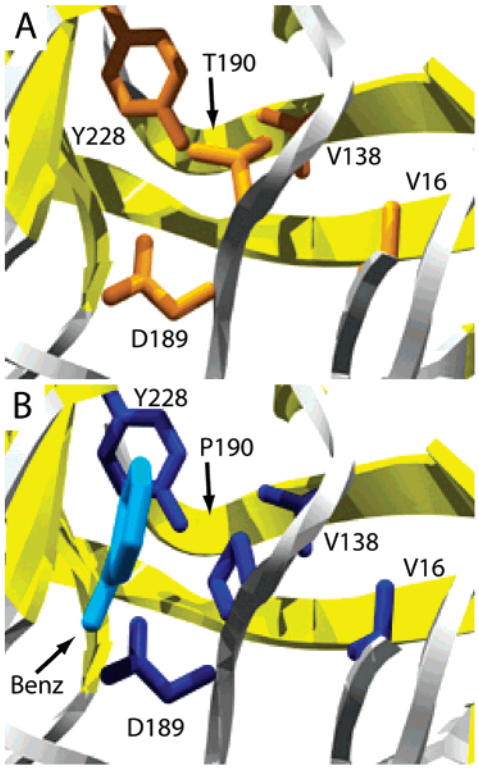

FIGURE 4: Comparison of the S1 binding pocket in SGT enzymes: (A) the recombinant wild-type structure and (B) the T190P mutant of SGT complexed with the benzamidine inhibitor (Benz). In the wild-type structure, the γ-OH points toward the S1 pocket and provides a H-bonding group for the substrate. Mutation of residue 190 to proline removes this H-bonding capacity without disrupting the critical D189.

second shell residues at positions 16 or 138 (Figure 4). The mutation does not significantly affect the conformation of any of the residues surrounding T190P, including the critical Asp189. The local rms deviation is low (0.70 Å) for all 96 atoms within a 5 Å radius of residue 190. However, the backbone carbonyl group of Asp189 is rotated 45° relative to the wild-type structure due the steric constraints of the proline residue. Rotation of this carbonyl group does not disrupt the hydrogen bond with the backbone nitrogen of residue 17. Hence, the moderate increase in Arg to Lys substrate specificity of this mutant is likely the result of the strengthened interaction between the substrate and Asp189 in the more hydrophobic environment, which enhances electrostatic interactions. The effect on lysyl side chains is predominant due to the lack of hydrogen bonding of the proline ring to the substrate or a bridging water molecule and would reduce the rate of association of the side chain with Asp189.

Although Asp189 adopts a conformation similar to that found in the wild-type protease, it is possible that the binding of the benzamidine inhibitor stabilizes the conformation of this side chain through formation of the electrostatic interaction. Analysis of the inhibition constants of benzamidine with the wild-type recombinant protease and four mutants reveals the basis for the increased specificity without loss of catalytic activity. Similar differences for the K_m values for the peptide substrates and K_i values relative to the wild type are observed for all mutants except T190P. The K_i value is 6-fold higher in the T190P mutant than in the wild-type enzyme, whereas the K_m values for the Arg- and Lys-containing substrates

increase 35-fold and 46-fold, respectively (Tables 3 and 4). As the inhibitor forms a direct electrostatic interaction with Asp189, the minor difference of the inhibitory constants suggests that in the T190P mutant the interaction occurs in a more hydrophobic environment and is not accompanied by structural rearrangement of the S1 binding pocket.

Whereas T190A affects interactions at the S1 binding pocket in the majority of proteases with a P1 Arg preference, the kinetic analysis of T190P suggests a potential intermediate in the evolution of the vertebrate coagulation cascade. Only two natural proteases have been identified that possess Asp189 and Pro190, hagfish prothrombin and kallikrein 10, and neither protein has been characterized with respect to substrate specificity. Both genes have been characterized by DNA sequence analysis of a number of overlapping cDNA library clones (*34*−*36*). The presence of Pro at this position suggests a specificity similar to that observed for the T190P mutant of SGT. Moreover, the residues surrounding residue 190 are identical to that in SGT. DNA sequence analysis of the hagfish prothrombin gene reveals that many of the features attributed to substrate specificity, such as the 60- and 99-loops, are identical. Hence, the reduced catalytic activity and higher K_m values caused by the proline at this position may be compensated by an increased concentration of the enzyme or substrate in the blood stream of this primitive vertebrate.

Second Shell Residues. A number of catalytic or structural studies involving the structure-based design of enzyme properties have demonstrated an important role for second shell residues surrounding the mutation of interest (*37*). In serine proteases, residue 190 interacts directly with the side chains of residues 16 (the N-terminus of the protein), 138, and 228. Residue 228 is a highly conserved tyrosine in known trypsin-like enzymes. Residues 16 and 138 are restricted to hydrophobic side chains (Val, Ile, and Leu). The differences in kinetic parameters observed between mutations made in this study, rat anionic trypsin and human trypsin (type I), are likely due to the presence or absence of methyl groups within this pair of residues (*5*, *6*, *8*). Moreover, mutagenesis of residue 16 in rat anionic trypsinogen II has been demonstrated to affect the primary substrate specificity of the enzyme (*38*). A more detailed investigation of these residues is required to understand the basis of substrate specificity within this family of important enzymes.

Future Directions. On the basis of the ease of production of SGT in the *B. subtilis* expression system, the ability to enhance specificity at the S2 to S4 binding pockets by either structure-based design or directed evolution strategies is possible. Introduction of an affinity tag, such as a hexahistidine tag, will speed the purification process of the protein and facilitate characterization of recombinant mutant proteases. The ability to crystallize this protein readily supports SGT as a model scaffold for the development of highly specific proteases and for understanding the mechanisms of substrate specificity in the highly evolved serine protease family.

ACKNOWLEDGMENT

The authors thank M. Boulanger for helpful discussion, S. He and S. Withers for the electrospray−MS analysis, R. Harrigan for making available DNA sequencing facilities, and R. Read for advice on crystallization conditions.

REFERENCES

1. Rypniewski, W. R., Perrakis, A., Vorgias, C. E., and Wilson, K. S. (1994) *Protein Eng. 7*, 57−64.
2. Perona, J. J., and Craik, C. S. (1997) *J. Biol. Chem. 272*, 29987−29990.
3. Schecter, I., and Berger, A. (1967) *Biochem. Biophys. Res. Commun. 27*, 157−162.
4. Graf, L., Jancso, A., Szilagy, L., Hegy, G., Pinter, K., Naray-Szabo, G., Hepp, J., Medzihradszky, K., and Rutter, W. J. (1988) *Proc. Natl. Acad. Sci. U.S.A. 85*, 4961−4965.
5. Evnin, L. B., Vásquez, J. R., and Craik, C. S. (1990) *Proc. Natl. Acad. Sci. U.S.A. 87*, 6659−6663.
6. Sichler, K., Hopfner, K. P., Kopetzki, E., Huber, R., Bode, W., and Brandstetter, H. (2002) *FEBS Lett. 530*, 220−224.
7. Rose, T., and Di Cera, E. (2002) *J. Biol. Chem. 277*, 19243−19246.
8. Perona, J. J., Evnin, L. B., and Craik, C. S. (1993) *Gene 137*, 121−126.
9. Trop, M., and Birk, Y. (1968) *Biochem. J. 109*, 475−476.
10. Olafson, R. W., and Smillie, L. B. (1975) *Biochemistry 14*, 1161−1167.
11. Olafson, R. W., Jurásek, L., Carpenter, M. R., and Smillie, L. B. (1975) *Biochemistry 14*, 1168−1176.
12. Kim, J. C., Cha, S. C., Jeon, S. T., Oh, S. K., and Byun, S. M. (1991) *Biochem. Biophys. Res. Commun. 181*, 707−713.
13. Read, R. J, Brayer, G. D., Jurásek, L., and James, M. N. G. (1984) *Biochemistry 23*, 6570−6575.
14. Read, R. J., and James, M. N. G. (1988) *J. Mol. Biol. 200*, 523−551.
15. Sambrook, J., Fritsch, E. F., and Maniatis, T. (1989) *Molecular Cloning: A Laboratory Manual*, 2nd ed., Cold Spring Harbor Laboratory, Cold Spring Harbor, NY.
16. Ye, R., Yang, L. P., and Wong, S.-L. (1996) *Proceedings of the International Symposium on Recent Advances in Bioindustry*, pp 160−169, Korean Society for Applied Microbiology, Seoul, Korea.
17. Wong, S.-L., Kawamura, F., and Doi, R. H. (1986) *J. Bacteriol. 168*, 1005−1009.
18. Wu, S. C., and Wong, S. L. (1999) *J. Biotechnol. 72*(1−2), 185−195.
19. Spizizen, J. (1958) *Proc. Natl. Acad. Sci. U.S.A. 44*, 1072−1078.
20. Hatanka, Y., Tsunematsu, H., Mizusaki, K., and Makisumi, S. (1985) *Biochim. Biophys. Acta 832*, 274−279.
21. Otwinowski, Z., and Minor, W. (1997) *Methods Enzymol. 276*, 307−326.
22. Brunger A. T., Adams, P. D., Clore, G. M., DeLano, W. L., Gros, P., Grosse-Kunstleve, R. W., Jiang, J. S., Kuszewski, J., Nilges, M., Pannu, N. S., Read, R. J., Rice, L. M., Simonson, T., and Warren, G. L. (1998) *Acta Crystallogr. D54*, 905−921.
23. McRee, D. E. (1999) *J. Struct. Biol. 125*, 156−165.
24. Meyer, E., Cole, G., Radhakrishnan, R., and Epp, O. (1988) *Acta Crystallogr. B44*, 26−38.
25. Bartunik, H. D., Summers, L. J., and Bartsch, H. H. (1989) *J. Mol. Biol. 210*, 813−828.
26. Berglund, G. I., Smalas, A. O., Hordvik, A., and Willassen, N. P. (1995) *Acta Crystallogr. D51*, 725−730.
27. Schroder, H. K., Willassen, N. P., and Smalas, A. O. (1998) *Acta Crystallogr. D54*, 780−798.
28. Yamane, T., Iwasaki, A., Suzuki, A., Ashida, T., and Kawata, Y. (1995) *J. Biochem. (Tokyo) 118*, 882−894.
29. Vásquez, J. R., Evnin, L. B., Higaki, J. N., and Craik, C. S. (1989) *J. Cell. Biochem. 39*, 265−276.
30. Hannig, G., and Makrides, S. C. (1998) *Trends Biotechnol. 16*, 54−60.
31. Wong, S.-L. (1995) *Curr. Opin. Biotechnol. 6*, 517−522.
32. Yuan, L. D., and Hua, Z. C. (2002) *Protein Expression Purif. 25*, 300−304.
33. Hedstrom, L., Szilagyi, L., and Rutter, W. J. (1992) *Science 255*, 1249−1253.
34. Banfield, D. K., and MacGillivray, R. T. A. (1992) *Proc. Natl. Acad. Sci. U.S.A. 89*, 2779−2783.
35. Banfield, D. K., Irwin, D. M., Walz, D. A., and MacGillivray, R. T. A. (1994) *J. Mol. Evol. 38*, 177−187.
36. Liu, X. L., Wazer, D. E., Watanabe, K., and Band, V. (1996) *Cancer Res. 56*, 3371−3379.
37. Tobin, M. B., Gustafsson, C., and Huisman, G. W. (2000) *Curr. Opin. Struct. Biol. 10*, 421−427.
38. Hedstrom, L., Lin, T. Y., and Fast, W. J. (1996) *Biochemistry 35*, 4515−4523.

BI0344230