Trypsin Specificity Increased through Substrate-Assisted Catalysis†

D. R. Corey,‡ W. S. Willett,§ G. S. Coombs,‡ and C. S. Craik*,§

Howard Hughes Medical Institute, Department of Pharmacology, and Molecular Biophysics Graduate Program, University of Texas Southwestern Medical Center at Dallas, 5323 Harry Hines Boulevard, Dallas, Texas 75235, and Department of Pharmaceutical Chemistry, University of California, San Francisco, California 94143-0446

Received March 7, 1995; Revised Manuscript Received May 30, 1995⊗

ABSTRACT: Histidine 57 of the catalytic triad of trypsin was replaced with alanine to determine whether the resulting variant would be capable of substrate-assisted catalysis [Carter, P., & Wells, J. A. (1987) *Science 237*, 394−9]. A 2.5-fold increase in k_{cat}/K_m was observed on tri- or tetrapeptide substrates containing *p*-nitroanilide leaving groups and histidine at P2. In contrast, hydrolysis of peptide substrates extending from P6 to P6′ is improved 70−300-fold by histidine in the P2 or P1′ position. This preference creates new protease specificities for sequences HR↓, R↓H, HK↓, and K↓H. The ability of histidine from either the P2 or the P1′ position of substrate to participate in catalysis emphasizes the considerable variability of proteolytically active orientations which can be assumed by the catalytic triad. Trypsin H57A is able to hydrolyze fully folded ornithine decarboxylase with complete specificity at a site containing the sequence HRH. Trypsin H57A was compared to enteropeptidase in its ability to cleave a propeptide from trypsinogen. Trypsin H57A cleaved the propeptide of a variant trypsinogen containing an introduced FPVDDDHR cleavage site only 100-fold slower than enteropeptidase cleaved trypsinogen. The selective cleavage of folded proteins suggests that trypsin H57A can be used for specific peptide and protein cleavage. The extension of substrate-assisted catalysis to the chymotrypsin family of proteolytic enzymes indicates that it may be possible to apply this strategy to a wide range of serine proteases and thereby develop various unique specificities for peptide and protein hydrolysis.

Proteases are valuable tools for the manipulation of proteins and peptides. However, in comparison to the stringent specificities of restriction endonucleases, most proteases lack selectivity for unique primary sequences. Highly selective proteases do exist which possess the ability to cleave proteins at unique sites, including factor Xa (Nagai & Thogersen, 1987), enteropeptidase (Kitamoto et al., 1994; Light et al., 1980), tobacco etch virus (TEV) protease (Parks et al., 1994), and picornavirus 3C protease (Walker et al., 1994). Yet their collective specificities are few, and the discrimination against closely related sequences is poor. As a result, there are applications for which no adequate protease is available. While the range of protease specificities will continually expand as naturally occurring proteases are identified, the current lack of selective proteases has made the development of new protease specificities a target for protein design.

Various strategies have been developed to confer novel substrate specificities on serine proteases. Plasminogen activators have been coupled to anti-fibrin antibodies to selectively target blood clots (Haber et al., 1989; Yang et al., 1994), and chymotrypsin has been biotinylated to localize proteolysis to avidin-bound substrates (Bayer et al., 1990). Proteases have been altered in the primary specificity pocket by rational design or by random mutagenesis [for a review, see Perona and Craik (1995)], engineered in the extended binding subsites so that a metal ion bridges the enzyme to the substrate (Willett et al., 1995), and modified by the removal of a catalytic residue that is subsequently provided by the substrate (Carter & Wells, 1987). The last of these strategies was applied to subtilisin by removing the active site histidine, making the enzyme specific for sequences containing histidine adjacent to small amino acids. Subtilisin BPN′ was altered by substituting histidine 64 with alanine to make optimal catalysis dependent on the presence of histidine at either the P2 or P1′ position (Schechter & Berger, 1968) of the substrate (Matthews & Wells, 1993). The resulting subtilisin H64A was specific for hydrolysis at HX↓ or X↓H peptide linkages where X is a small hydrophobic amino acid. This variant has been shown to cleave a recombinant parathyroid hormone fusion protein more efficiently than thrombin or enteropeptidase (Forsberg et al., 1991). Additional site-directed mutagenesis of subtilisin H64A was used to increase the hydrolysis rate of the enzyme for particular substrates (Carter et al., 1991). Subsequently, phage display was used to identify especially labile substrate sequences which are cleaved more efficiently by the enzyme (Matthews & Wells, 1993).

We applied the substrate-assisted catalysis strategy to the chymotrypsin family of serine proteases. Although the three-dimensional folds of the chymotrypsin and subtilisin families are unrelated, the catalytic triads of the enzymes superimpose with a root-mean-square difference of less than 0.2 Å (Corey & Craik, 1992, and references therein). The enzymes of the chymotrypsin family range from the digestive enzymes trypsin, chymotrypsin, and elastase, whose specificities are

†This work was supported in part by NSF MCB-9219806 to C.S.C. D.R.C. is an Assistant Investigator with the Howard Hughes Medical Institute. W.S.W. is supported by the University of California Biotechnology Research and Education Program. G.S.C. is supported by NIH Biophysics Predoctoral Training Program 5P32GM08287.

* Author to whom correspondence should be addressed at the Department of Pharmaceutical Chemistry, University of California at San Francisco, Box 0446, San Francisco, CA, 94143−0446. Phone: (415) 476−8146; FAX: (415) 476−0688; e-mail: craik@cgl.ucsf.edu.

‡ University of Texas Southwestern Medical Center at Dallas.

§ University of California, San Francisco.

⊗ Abstract published in *Advance ACS Abstracts,* August 1, 1995.

0006-2960/95/0434-11521$09.00/0

determined primarily by their S1 subsites [for a review, see Perona and Craik (1995)], to the highly selective proteases of the blood coagulation pathway such as thrombin (Vu et al., 1991), tissue plasminogen activator (Madison et al., 1995), and complement factor D (Narayana et al., 1994), which interact more extensively with their substrates. If substrate-assisted catalysis were to prove general for serine proteases, their diverse substrate preferences would allow the development of various stringent specificities.

Histidine 57 of the catalytic triad of rat anionic trypsin was replaced with alanine, and the catalytic properties of the variant enzyme were characterized on both synthetic and natural substrates. Trypsin was chosen as an archetypal member of the structurally related chymotrypsin family of serine proteases. Trypsin is specific for the cleavage of Arg and Lys at P1, so that introduction of histidine specificity at P2 by trypsin H57A would yield H-R/K or R/K-H specificities not possessed by subtilisin H64A or any other known protease. The catalytic properties of trypsin H57A and trypsin were compared on extended peptide substrates which closely mimic target substrate sequences within fusion proteins. Trypsin H57A was also analyzed for its ability to efficiently hydrolyze folded proteins at unique sites. In spite of the dissimilarity in tertiary structure between trypsin and subtilisin, substrate-assisted catalysis by trypsin H57A occurs with both peptide and protein substrates. This result demonstrates the potential to generate new protease specificities by adapting the chymotrypsin family of serine proteases for substrate-assisted catalysis.

MATERIALS AND METHODS

Materials. Peptides were synthesized by solid-phase synthesis on an Applied Biosystems (Foster City, CA) Model 430A or a Symphony Multiplex (Rainin, Woburn MA) peptide synthesizer using Fmoc (*N*-fluorenylmethoxycarbonyl) chemistry. Each peptide was desalted using a Whatman C-4 cartridge (Fisher). The molecular weights of the peptides and their proteolyzed fragments were confirmed by mass spectral analysis using a VG (Altrincham, England) 30−250 quadrapole mass spectrometer. Wild-type rat anionic trypsin and trypsin H57A were obtained as described (Corey et al., 1992) from a bacterial expression system or from a yeast expression system (Hedstrom et al., 1992). Proteases obtained from either expression system were indistinguishable. Purified recombinant ornithine decarboxylase from *Leishmania donovani* (LODC) (Osterman et al., 1995) was obtained from Dr. Margaret Phillips (University of Texas Southwestern Medical Center at Dallas). Automated Edman degradation of trypsin H57A cleavage products of ornithine decarboxylase was performed on an Applied Biosystems Model 477A amino acid sequencer with an online Model 120A pTH amino acid analyzer using standard conditions. The peptidyl *p*-nitroanilide (pNA) substrate succinyl-Ala-His-Arg-pNA (sucAHR-pNA) was obtained from Dr. Paul Carter (Genentech, South San Francisco, CA), and succinyl-Ala-Ala-Ala-Arg-pNA (suc-AAAR-pNA) was obtained from Bachem.

Computer Modeling Methods. The Insight II program (Biosym Technologies) running on a Silicon Graphics Indigo XS24 was used to display the X-ray crystal structure of rat anionic trypsin D102N complexed to the bacterial serine protease inhibitor ecotin (McGrath et al., 1994). The trypsin/ecotin interface at the active site is used as a model of the enzyme/substrate interaction with ecotin representing bound substrate. To model the potential of histidine from a substrate to participate in catalysis by trypsin H57A, the amino acid at either the P2 or the P1′ position of ecotin (Thr83 or Met85) was replaced with histidine, and the P1 amino acid (Met84) was replaced with a lysine. Histidine 57 on trypsin was replaced with alanine to create a model of trypsin H57A. The histidine side chains on the substrate model were rotated manually to find a position that formed suitable hydrogen bonds with Ser195 of trypsin. Criteria for good hydrogen bonds were bond distances between 2.5 and 3.5 Å, and a nearly coplanar arrangement of the imidazole ring nitrogens of the histidine from the substrate and the hydroxyl proton from Ser195 of trypsin.

Kinetic Assays. The cleavage of the peptides by trypsin or trypsin H57A was monitored at 220 nm by reverse-phase HPLC (Rainin) using a Microsorb 5-μm 300-Å reversed-phase column (4.6 mm × 25 cm) (Rainin) and a 10−50% gradient of 0.1% trifluoroacetic acid in doubly distilled water (buffer A) and 0.08% trifluoroacetic acid in 95% acetonitrile/5% doubly distilled water (buffer B) (Corey & Craik, 1992). Cleavage assays were performed at 37 °C in 100 mM NaCl/20 mM $CaCl_2$/50 mM Tris-HCl, pH 8.0, buffer. Assays with trypsin required 2−20 min and were terminated through the addition of 0.5 vol of 1% aqueous trifluoroacetic acid. Assays with trypsin H57A required 2 h for substrates containing histidine at P2 or P1′ and 48 h for substrates lacking histidine at these positions. Assays with trypsin H57A were terminated by freezing on dry ice. The integrated areas of the cleavage products and starting material after HPLC analysis were used to determine the rate of cleavage at varied substrate concentrations. Eadie−Hofstee analysis of the initial rate of peptide cleavage was used to determine k_{cat} and K_m values. The location of cleavage within the peptides was confirmed by mass spectral analysis of the peptide fragments recovered after enzymatic digestion and HPLC purification. Assays for the hydrolysis of peptides containing D-histidine were performed as described above and were analyzed after a 48-h incubation at 37 °C.

Cleavage of LODC by trypsin H57A was performed at 37 °C in a buffer consisting of 20 mM Tris-HCl, pH 7.5, 2 μM β-mercaptoethanol, 0.02% Brij-35, 0.5 mM EDTA, and 20 μM pyridoxal phosphate. The cleavage reactions were stopped periodically over 1−48 h by freezing on dry ice, and the appearance of hydrolyzed products was monitored periodically over 72 h by SDS−PAGE. Kinetic constants for the cleavage of LODC by trypsin H57A were determined using LODC concentrations of 0.5, 2, 3, 4, 5, and 6 μM and 0.2 μM trypsin H57A at 37 °C in a buffer containing 25 mM Tris-HCl, pH 7.5, 0.02 Brij-35, 2.5 mM DTT, and 20 μM pyridoxal phosphate, for times ranging from 3.5 to 6.5 h. These concentrations resulted in rates in the linear range of the reaction and involved less than 20% conversion of substrate. Reactions were stopped by freezing at −70 °C, and cleavage products were separated by SDS−PAGE using 12% polyacrylamide. Gels were stained with Coomassie brilliant blue R-250 for 1 h and then destained overnight in 10% acetic acid/5% methanol/85% water. Intensities of substrate and product bands were quantitated on a Model 300A scanning densitometer from Molecular Dynamics which was operated with Image Quant 3.0 software. The percentage of LODC cleaved at a given LODC concentration

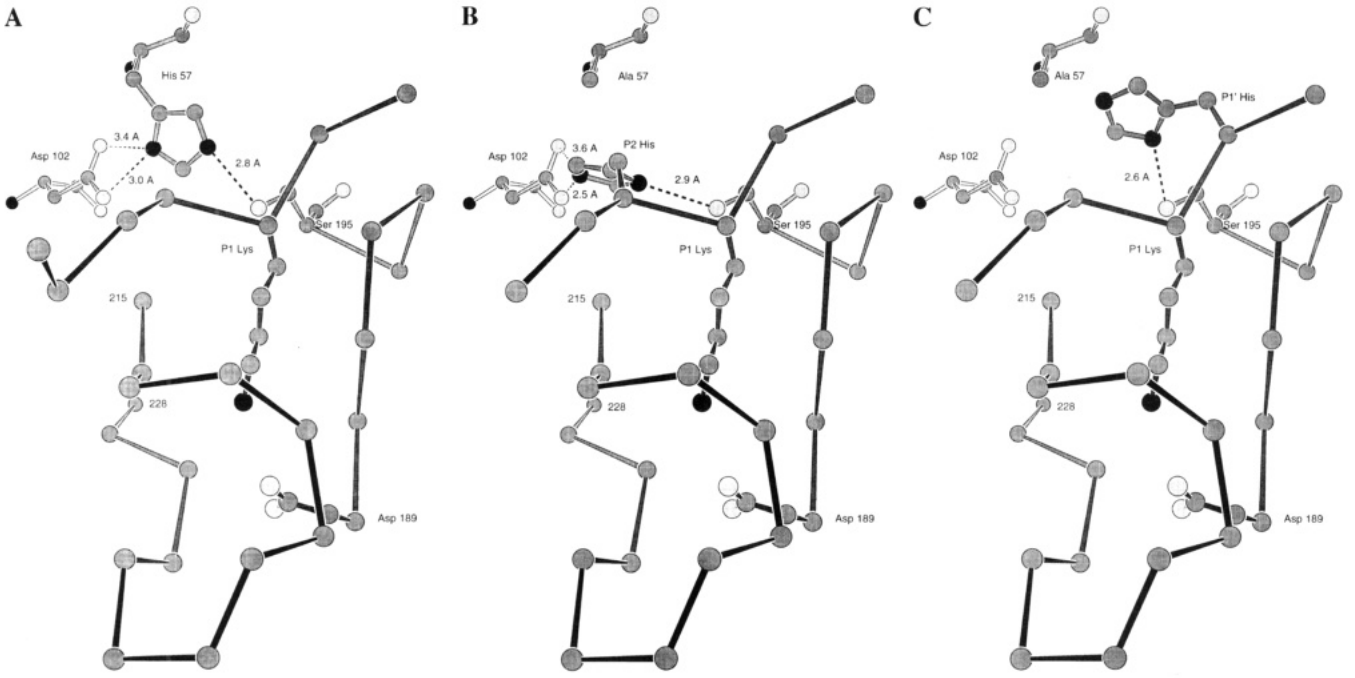

FIGURE 1: Modeling substrate-assisted catalysis in trypsin. (A) Model of the trypsin active site displaying the catalytic triad and the S1 subsite. (B) Modeled interactions of substrate-assisted catalysis in trypsin H57A showing bound substrate contributing the catalytic histidine from the P2 position. (C) Substrate contributing the catalytic histidine from the P1′ position. Carbon is shown in dark gray, oxygen in light gray, and nitrogen in black. Hydrogen bonds are represented by dashed lines, and distances are in angstroms.

was determined by calculating $P/(S + P)$ where P was the intensity of the product bands and S was the intensity of the substrate band. These ratios were used to calculate initial velocities of cleavage. The kinetic parameters k_{cat} and K_m were derived from Eadie−Hofstee analysis and were done in duplicate. Cleavage of substrates sucAHR-pNA and sucAAPR-pNA was monitored by UV spectrophotometry as described (Corey & Craik, 1992).

Production of Recombinant Trypsinogen or Trypsin. Trypsinogen and trypsinogen variants were constructed and expressed in yeast as described previously (Willett et al., 1995) using site-directed mutagenesis and the following synthetic oligonucleotides to mutate the Bluescriptbased mutagenesis plasmid, pST: GGAGGCAAG CACAGCTGCCAGGG, D189H oligo; GGTGTCTGCGG CCGCATG CTATAA, H57A oligo; and GTGGATGATGA TCATAGAATCGTTGGAGG, D14H/K15R oligo. The mismatched bases are underlined. The DNA encoding the recombinant proteins was sequenced in its entirety in pST to confirm the desired mutations and ensure that unwanted mutations did not occur elsewhere in the gene. Mutant forms of the protein were expressed in bacteria for a rapid determination of kinetic parameters as described previously (Corey et al., 1992). To obtain larger amounts (>10 mg) of the recombinant proteins, the yeast expression system was used. Concentrations of trypsin or trypsinogen were calculated from absorbance at 280 nm using an extinction coefficient of 34 300 M^{-1}.

Cleavage of Trypsinogen. Cleavage of trypsinogen by trypsin H57A was performed in 100 mM Tris and 1 mM $CaCl_2$, pH 8.0, at 37 °C. A typical reaction mixture contained 20 μM trypsinogen and 1 μM trypsin H57A in a 50-μL volume in a 0.5-mL siliconized Eppendorf tube. Cleavage of trypsinogen by enteropeptidase was accomplished using the same reaction conditions except that the enteropeptidase concentration was 50 nM in the reaction mixture. Ten-microliter samples were taken from the reaction mixture at various times and analyzed by Coomassie-stained SDS−PAGE on gels containing 12% acrylamide. Cleavage of trypsinogen yields trypsin, which upon SDS−PAGE analysis is indistinguishable from the trypsin H57A used to cleave the propeptide. As a result, the extent of cleavage by trypsin H57A was monitored by the disappearance of trypsinogen. Trypsinogen was well-resolved from mature trypsin, allowing indirect observation of its loss during conversion to trypsin. The intensity of the Coomassie-stained trypsinogen band was quantitated on a Model 300A scanning densitometer from Molecular Dynamics which was operated with Image Quant 3.0 software and was used to evaluate cleavage by trypsin H57A.

RESULTS

Computer Modeling. Replacement of amino acids at P2 and P1′ with histidines from a potential substrate was modeled from the structure of trypsin complexed with ecotin (McGrath et al., 1994). The modeling suggests a molecular basis for substrate-assisted catalysis by trypsin (Figure 1). The histidine introduced at P2 has a $\chi_1 = -139.5°$ and a $\chi_2 = -71.3°$. The histidine introduced at P1′ has a $\chi_1 = -19.3°$ and a $\chi_2 = -51.6°$. With the exception of the $-19.3°$ χ_1 angle observed in the P1′ model, these χ angles fall in the observed range of histidine rotamers (Ponder & Richards, 1987). The modeled histidines make no short contacts less than 3.0 Å other than hydrogen bonds. Implicit in the modeling is the hypothesis that the histidine brought to the active site by the substrate can participate in the reaction by abstracting the proton from Ser195 to allow nucleophilic attack of the scissile bond in the substrate. The modeling suggests that the trypsin scaffold can allow a histidine in either the P2 or the P1′ position of the substrate to bind in a conformation placing one of the histidine ring nitrogens in a proper position to abstract the Ser195 proton.

Table 1: Synthetic Substrate Kinetics[a]

enzyme	substrate	k_{cat} (s^{-1})	K_m (μM)	k_{cat}/K_m (M^{-1} min^{-1})	rel k_{cat}/K_m
trypsin	sucAAPR-pNA	43 ± 1	22 ± 1	1.9×10^6	29 500
	sucAHR-pNA	36 ± 1	41 ± 11	0.9×10^6	13 250
trypsin H57A	sucAAHR-pNA	0.002 ± 0.0002	31 ± 12	67	1
	sucAHR-pNA	0.007 ± 0.001	40 ± 13	167	2.5

[a] Assays were carried out in 1 mL of 100 mMTris/100 mM NaCl/1 mM $CaCl_2$, pH 8.0, at 25 °C. [trypsin] = 2 nM; [trypsin H57A] = 500 nM. Activity was measured by increase in absorbance at 410 nm, and kinetic constants were calculated by fitting the data to the Michaelis–Menten equation.

Table 2: Peptide Kinetics of Trypsin H57A[a]

	trypsin			trypsin H57A		
substrate	k_{cat} (s^{-1})	K_m (μM)	k_{cat}/K_m (M^{-1} s^{-1})	k_{cat} (s^{-1})	K_m (μM)	k_{cat}/K_m (M^{-1} s^{-1})
(I) YLVG**HR**GFFYDA	16	15	1.1×10^6	4.7×10^{-3}	47	100
(II) YLVG**PR**GFFYDA	31	12	2.6×10^6	5.0×10^{-5}	34	1.5
(III) YLVG**HK**GFFYDA	1.2	12	1.0×10^5	5.2×10^{-3}	46	110
(IV) YLVG**PK**GFFYDA	1.8	63	2.9×10^4		nd	
(V) YLVG**HRH**FFYDA	2.2	78	2.8×10^4	2.1×10^{-2}	47	466
(VI) YLVG**PRH**FFYDA	2.3	11	2.2×10^5	2.8×10^{-3}	25	110
(VII) GGSGPF**GRH**ALVPE	7.5	3.9	1.9×10^6	1.3×10^{-2}	38	350
(VIII) YLVG**PRGH**FYDA	5.0	33	1.5×10^5	1.2×10^{-4}	130	0.9
(IX) YLVG(D)**HR**GFFYDA		na			na	
(X) YLVG**PR**(D)**H**FFYDA		na			na	

[a] The error in these determinations was ±20% for k_{cat} and K_m. nd, not determined; activity observed was too low over a 24-h period to calculate accurate kinetic constants. na, not active; no hydrolytic activity was observed over a 24-h period.

Numerous structural and functional studies of trypsin show that His57 has its Nδ1 proton hydrogen bonded to the Asp102 carboxylate at a distance of 3.0–3.5 Å and its Nε2 unprotonated, poised to accept the Ser195 hydroxyl proton from a distance of 2.8 Å (Figure 1A). Substituting a histidine at P2 on the substrate now allows Nε2 of the imidazole ring to hydrogen bond to Asp102 at distances of 3.6 and 2.5 Å to the carboxylate oxygens, and Nδ1 to accept the Ser195 proton from a distance of 2.9 Å (Figure 1B). Yet if the P2 histidine were to donate the proton back to the leaving group amine as His 57 in trypsin does during the acylation reaction, it would have to rotate about the Cα–Cβ bond to move nearly 2 Å. Subsequently, to participate in the deacylation reaction, the histidine would have to move back closer to its original modeled position to accept a proton from water. This would be a large directed movement for a histidine side chain which lacks the structural constraints of His57 in trypsin.

The Nδ1 from histidine substituted at P1′ can form a hydrogen bond with Ser195 at a distance of 2.6 Å, but the Nε2 is 6.7 and 6.4 Å from the oxygens of the Asp 102 carboxylate group, too far away to benefit from a potential hydrogen bond (Figure 1C). A histidine at P1′ on the substrate can only participate in the acylation reaction because it is on the leaving group side of the scissile bond. It is not known whether the histidine donates a proton to its own incipient N-terminus or whether the proton comes from the solvent. The model shows the P1′ His Nδ1 2.7 Å from its backbone nitrogen and 65° out of the plane of the imidazole ring, suggesting that it may be able to donate the necessary proton. Presumably, hydroxide ion from water attacks the acyl enzyme intermediate to catalyze the deacylation reaction. Modeling of D-histidine at either P2 or P1′ resulted in sub-van der Waals contacts with several adjacent residues and distances from the imidazole nitrogens to the Ser195 hydroxyl group that did not allow formation of hydrogen bonds.

Activity of Trypsin H57A toward Labile Amide Substrates. The efficiency of substrate-assisted catalysis by trypsin H57A was initially assayed using tetrapeptide (P4 to P1) or tripeptide (P3 to P1) substrates containing *p*-nitroanilide (pNA) leaving groups (Table 1). The substrates were sucAHR-pNA, which contains a histidine at P2 that could potentially participate in substrate-assisted catalysis, and sucAAPR-pNA, which lacks histidine and would not be expected to participate in catalysis. Trypsin H57A catalyzed the hydrolysis of both substrates much more slowly than did trypsin (Table 1), by factors of 29 500 for sucAAPR-pNA and 5300 for sucAHR-pNA. Hydrolysis by trypsin H57A of sucAHR-pNA was accelerated 2.5-fold relative to the hydrolysis by trypsin H57A of sucAAPR-pNA. This slight acceleration compares to a 2-fold decrease in the activity of trypsin for sucAHR-pNA relative to sucAAPR-pNA.

Activity of Trypsin H57A toward Peptide Substrates Extending from P6 to P6′. We had shown previously that the catalysis of labile amide linkages by variant trypsins with altered catalytic triads did not always accurately predict the level of catalysis toward extended peptides with normal peptide bonds (Corey & Craik, 1992). Therefore, to further our evaluation of trypsin H57A, a series of extended peptides were synthesized to probe (1) the extent of catalysis by trypsin H57A towards peptides, (2) the relative importance of Lys or Arg at P1, and (3) the possibility that histidine at P1′ of the substrate might also be able to participate in catalysis. Direct comparison of cleavage by trypsin H57A of **I** (Table 2), which contained a histidine at P2, and **II** which contained a proline at P2, revealed that k_{cat}/K_m for the cleavage of extended peptide substrates was increased 70-fold by the presence of a histidine at P2. This increase contrasts with a 2-fold decrease in the activity of trypsin upon the alteration of the P2 proline of **II** to a P2 histidine in **I**. The observed rate enhancement was more dramatic when lysine was located at P1. Trypsin H57A cleaved the P2-histidine, P1-lysine-containing substrate **III** as rapidly as

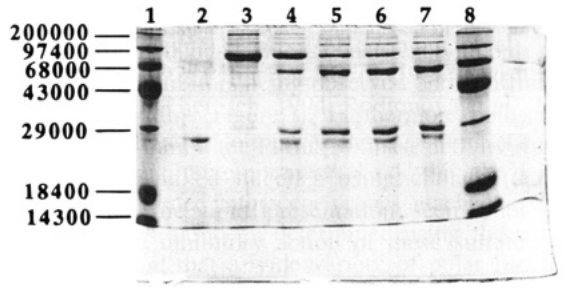

FIGURE 2: Cleavage of LODC by trypsin H57A. Cleavage reaction mixtures contained trypsin H57A (1.7 μM) and LODC (3.2 μM). Reactions were performed as described in Materials and Methods. Lanes 1 and 8, molecular weight markers; lane 2, trypsin H57A; lane 3, LODC; lanes 4−7, incubation of LODC with trypsin H57A for 4, 17, 27, and 42 h.

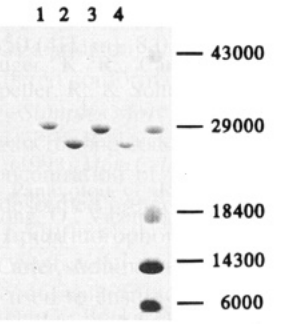

FIGURE 3: Coomassie-stained SDS−PAGE analysis of trypsin H57A cleavage of DDDHR trypsinogen. Lane 1, trypsinogen D189H; lane 2, trypsinogen D189H cleaved by enteropeptidase; lane 3, trypsinogen D14H/K15R/D189H; lane 4, trypsinogen D14H/K15R/D189H cleaved by trypsin H57A; rightmost lane, low molecular weight markers.

its arginine-containing counterpart **I**, while trypsin H57A failed to yield measurable cleavage of peptide **IV**, which contained a lysine at P1 but which lacked histidine.

It had previously been shown that subtilisin H64A exhibited an enhanced cleavage of peptides containing histidine at P1′ (Matthews & Wells, 1993). To assay whether trypsin H57A would also possess this activity, peptides **V** and **VI** were designed to contain histidine either at P1′ and P2 or at P1′ alone. Assays toward peptide **VI**, which contains a histidine at P1′, revealed an 80-fold rate enhancement relative to peptide **II**, which contains a glycine at P1′, an acceleration which was similar to that observed for hydrolysis by trypsin H57A of substrate **I**, which contained a P2 histidine. Substrate **V** contained histidine at both P1′ and P2 and was a 4−5-fold better substrate in terms of k_{cat}/K_m than either of the single-histidine-containing substrates (**I** or **VI**). The positions from which substrate imidazole rings may substitute for the histidine of the catalytic triad is limited, however, as substrates **IX** and **X** containing D-histidine at either the P2 or the P1′ position were not hydrolyzed by trypsin H57A. This is consistent with the modeling results showing that the D-histidine residues are unable to adopt a catalytically active conformation. Similarly, no rate enhancement was observed for the hydrolysis by trypsin H57A of **VIII**, a substrate containing L-histidine at P2′. Efficient hydrolysis could also be achieved by using substrates containing optimized subsite interactions. Peptide **VII**, which contained histidine at P2 and whose P4, P3, and P1′−P4′ residues had been identified within a particularly labile trypsin substrate (Ding et al., 1995), was processed by trypsin H57A with a k_{cat}/K_m which was 3−4 fold greater than for the other single-histidine-containing substrates, **I** and **IV**.

Cleavage of Protein Substrates. To examine whether trypsin H57A could cleave a folded protein with complete specificity, trypsin H57A-catalyzed hydrolysis of ornithine decarboxylase from *Leishmania donovani* (LODC) was analyzed. LODC is a 708 amino acid protein containing 37 arginines and 11 lysines (Hanson et al., 1992). There is an HRH site at amino acids 246−248 and an RH site at amino acids 325−326. Treatment of LODC with trypsin H57A yielded two fragments of 247 and 461 amino acids in length (Figure 2, lanes 4−7). No other products were evident after incubation for 42 h (lane 7). N-Terminal amino acid sequence analysis confirmed that cleavage was occurring at the HRH site between R247 and H248. Trypsin H57A did not cleave at the secondary RH site at residues 325−326, emphasizing that the structure of the protein substrate may obscure particular cleavage sequences. The HRH site occurs at a junction between the C-terminal region common to all parasite decarboxylases and an N-terminal extension which has only been observed in LODC (Hanson et al., 1992) and is presumably more solvent-exposed than the R325−H326 site.

The initial rate of cleavage was examined as a function of substrate concentration to determine the kinetic constants for the specific hydrolysis of LODC by trypsin H57A. The turnover number, k_{cat}, was 2.0×10^{-4} s^{-1}; the binding constant, K_m, was 2.4 μM; and k_{cat}/K_m was 83 M^{-1} s^{-1}. Limited proteolysis of LODC by trypsin also afforded the 461 amino acid product, but there were many other proteolysis products as well, and the 247 amino acid product was not present. Further proteolysis by trypsin resulted in the total degradation of LODC, whereas extended digestions with trypsin H57A yielded complete digestion at the target site with no appearance of secondary hydrolysis.

Trypsin H57A was also used to cleave the propeptide of a variant trypsinogen. Three mutations were introduced into trypsinogen so that it could be used as a substrate for trypsin H57A (Figure 3). The D189H mutation was introduced into the primary specificity pocket so that, upon activation, the resultant trypsin would be inactive, thereby removing the possibility of autoactivation and subsequent autolysis products. We have previously shown that trypsin D189H is 250-fold reduced in k_{cat}/K_m on synthetic lysine-containing ester substrates, with no detectable amidase activity (Willett et al., 1995). Two other mutations, D15H and K16R, were introduced at the propeptide/enzyme junction to create a substrate sequence for trypsin H57A (DDD**HR**-IVGG vs DDDDK-IVGG in trypsinogen). Trypsin H57A cleaved the variant propeptide with complete specificity with a k_{cat}/K_m of 1.9 M^{-1} s^{-1}. This was ~100-fold slower than enteropeptidase cleaved trypsinogen as judged from Coomassie-stained gels of reaction products. Although there are several Lys and Arg residues located on surface loops of trypsin and one other KH sequence at position 90−91, there was no nonselective cleavage observed by trypsin H57A of either the wild-type or the variant trypsinogen (Figure 3). In contrast to the stringent specificity of trypsin H57A, the more relaxed specificity of enteropeptidase permitted cleavage of the variant trypsinogen propeptide (D_3HR) as well as the wild-type propeptide (D_4K) with comparable efficiency.

Trypsin also cleaved the mutant trypsinogen, but yielded a ladder of unwanted proteolytic byproducts (data not shown).

DISCUSSION

A goal of protein engineering is the creation of variant proteins which afford a useful perspective on the structure and function of the wild-type protein and on related enzymes. A test of whether this perspective has been obtained is the ability to alter function in a prescribed fashion. If successful, these alterations might also yield catalysts that possess practical value as experimental tools or therapeutic agents. To be of practical value, modified enzymes must combine a turnover rate that is sufficient for a given experimental purpose with a high degree of specificity. Engineered proteases which possess these properties would have a number of applications. Novel proteases with altered specificities could supplement existing proteases in the mapping of proteins much as restriction endonucleases with eight-base recognition sequences complement those which recognize six-base sites for the mapping of nucleic acids. Modified proteases could also be used to cleave folded proteins at naturally occurring cleavage sites to generate desired truncated products or to isolate protein domains. Alternatively, recombinant fusion proteins can be cleaved at engineered sites to liberate the desired native protein. Eventually these principles will allow us to create therapeutic proteases capable of cleaving target sites *in vivo*.

Site-directed mutagenesis was used to replace the catalytic histidine of trypsin with alanine to produce trypsin H57A. The ability of this variant to participate in substrate-assisted catalysis was examined using synthetic amide, extended peptide, and protein substrates. Catalysis by trypsin H57A toward synthetic substrates containing labile amide bonds showed a 2.5-fold increase in k_{cat}/K_m when histidine was present at P2. This differential reactivity was much less than had been observed earlier for subtilisin H64A, where the introduction of histidine at P2 resulted in a 200-fold increase in k_{cat}/K_m. Activity of trypsin H57A was then examined using a series of extended peptide substrates containing histidine at various subsites relative to lysine or arginine at P1. These assays revealed a greater than 70-fold increase in k_{cat}/K_m for the hydrolysis of peptides containing histidine at P2 or P1′ by trypsin H57A. The increase in catalysis caused by the presence of a P2 or P1′ histidine in the substrate resulted primarily from an increased k_{cat} supporting the hypothesis that histidine enhances hydrolysis through direct participation in the catalytic mechanism.

The efficiency of substrate-assisted catalysis could be increased either by the simultaneous inclusion of P2 and P1′ histidines in the substrate or by using a substrate whose subsite occupancy is known to yield efficient catalysis by trypsin. The k_{cat}/K_m of trypsin H57A toward a peptide which contained histidine at P2 and P1′ was only 55-fold lower than that of trypsin on the same peptide, an impressive result for a variant trypsin which lacks a member of the catalytic triad. This shows that substrate specificity can be varied without a dramatic reduction in catalytic efficiency. The improved rate enhancement observed with this peptide substrate may have been due to the ability of the two histidines in the substrate to increase the effective concentration of the general base required for catalysis. The restoration of activity to both trypsin H57A and subtilisin H64A through the presence of histidine at either the P2 or the P1′ position of substrate is clear evidence for the diversity of catalytically competent orientations that are possible for histidine to assume. The highest level of catalysis observed by trypsin H57A on a peptide with one histidine was achieved with GGSGPFG**RH**ALVPE. This peptide contains a histidine at P1′ and amino acids that form favorable subsite interactions yielding a low K_m and a high k_{cat}/K_m (Ding et al., 1995). This shows that it is possible to optimize the rate of substrate cleavage through the selection of labile substrate sequences.

A possible reason for the low level of activity by trypsin H57A on synthetic amide substrates relative to peptide substrates is the absence of extended subsite interactions C-terminal to the scissile bond. The peptide substrate occupies these subsites, allowing the binding energy to be translated into increased catalysis. For pNA substrates, the S1′ subsite of the protease is not filled by an amino acid but by a chromogenic or fluorogenic leaving group. The other prime sites, S2′, S3′, S4′, and so on, are not filled at all, so that the binding energy contributed by the prime-side residues is not available for catalysis. Hence, exclusive use of synthetic substrates may mask increased activity of a variant trypsin toward peptide bonds (Corey & Craik, 1992). The binding energy contributed by the P′ residues may induce a conformational change in the enzyme necessary for efficient catalysis. For example, trypsin D102S/S214D, which relocates the negative charge of the catalytic triad, exhibits only a 10-fold increase in k_{cat} relative to the D102S variant for the hydrolysis of labile synthetic substrates, but exhibits a 7200-fold increase in k_{cat} for the hydrolysis of peptide linkages relative to trypsin D102S (Corey & Craik, 1992, 1993). This effect is consistent with the increase in serine protease activity observed toward peptide substrates as additional P′ amino acid residues are present (Bauer et al., 1981; Bizzozero & Dutler, 1987) or as they are varied (Schellenberger et al., 1993).

Just as the hydrolysis of labile synthetic amide linkages by proteases does not always reflect their ability to hydrolyze the unactivated amide linkages of peptides, the hydrolysis of peptides may not always reflect the capacity of a protease to cleave folded proteins. Folded secondary and tertiary structure can obstruct potential cleavage sites and can prevent protein hydrolysis. Trypsin H57A hydrolyzed an internal amide linkage in ornithine decarboxylase from *Leishmania donovoni* (LODC) and a variant propeptide from recombinant trypsinogen. Measurement of the kinetic constants for the cleavage of LODC by trypsin H57A revealed that both K_m and k_{cat} were lower than had been observed for the cleavage of peptide substrates. The measured k_{cat}/K_m was 82 M^{-1} s^{-1}, a rate which is quite slow relative to cleavage of peptide substrates by wild-type trypsin. Cleavage by trypsin H57A of the propeptide from the recombinant trypsinogen occurred with a k_{cat}/K_m of 1.9 M^{-1} s^{-1}, at least 2 orders of magnitude down from enteropeptidase cleavage of the propeptide. Whereas enteropeptidase has evolved to cleave the D_4K sequence, trypsin is a poor enzyme against this substrate, probably due to the negative charges on the P2−P5 residues.

However, even this low rate of catalysis permits practical applications of the variant enzyme given a sufficient supply of protease. Indeed, the more critical consideration in utilizing a selective protease is its discrimination between targeted and untargeted catalysis during the desired period

required to process substrate. After extended incubation, the only observed hydrolysis of LODC by trypsin H57A is at the target HRH at position 246−248. A similar level of discrimination was also observed for cleavage of the propeptide of trypsinogen. Potential trypsin H57A cleavage sites HR↓, HK↓, R↓H, and K↓H are not uncommon in proteins. However, many of these sites will occur within regions which are inaccessible to proteolysis when the protein is folded, so that sequences which are susceptible to trypsin H57A will be relatively rare. As has been done with factor Xa (Sahin-Tóth et al., 1995), trypsin H57A can be used to map sites that are proteolytically labile and, therefore, solvent accessible. Trypsin H57A has the advantage over factor Xa in that it requires less significant alteration in the target protein to create a cleavage site.

Substrate-assisted catalysis by serine proteases has now been shown for members of both the chymotrypsin and subtilisin families in spite of their structural divergence, and may prove to be a general property accessible to many serine proteases. Further structural and kinetic studies are being performed to characterize the substrate/enzyme interactions involved so that activity may be increased through additional mutagenesis. Subsequent modifications may be used not only to optimize the efficiency of this enzyme but also to refine its selectivity. It should be possible to extend substrate-assisted catalysis to serine proteases with chymotrypsin or elastase specificities for large hydrophobic or small hydrophobic amino acid side chains at P1. It may also be possible to narrow the existing specificities of selective proteases such as tissue plasminogen activator or thrombin, to create unique selectivities for engineered or native protein substrates. Such control of specificity through the application of protein engineering principles represents an important step toward the rational design of proteases capable of recognizing specific target proteins.

ACKNOWLEDGMENT

We wish to thank L. DeOgny for performing peptide synthesis and Dr. Paul Carter for providing sucAHR-pNA. We also wish to express our thanks to Dr. Margaret Phillips and Dr. Andrei Ostermann for providing ornithine decarboxylase and for suggesting its use as a protein substrate. We thank Dr. J. P. Quigley for his molecular biology expertise and for performing active-site titrations. Figure 1 was created using the program MAXIMAGE, written by Mark Rould.

REFERENCES

Bauer, C. A., Brayer, G. D., Sielecki, A. R., & James, M. N. (1981) *Eur. J. Biochem. 120*, 289−94.

Bayer, E. A., Grootjans, J. J., Alon, R., & Wilchek, M. (1990) *Biochemistry 29*, 11274−9.

Bizzozero, S. A., & Dutler, H. (1987) *Arch. Biochem. Biophys. 256*, 662−76.

Carter, P., & Wells, J. A. (1987) *Science 237*, 394−9.

Carter, P., Abrahmsen, L., & Wells, J. A. (1991) *Biochemistry 30*, 6142−8.

Corey, D. R., & Craik, C. S. (1992) *J. Am. Chem. Soc. 114*, 1784−90.

Corey, D. R., & Craik, C. S. (1993) in *Proteases and Their Inhibitors; Fundamental and Applied Aspects* (Aviles, F. X., Ed.) pp 425−44, Walter DeGruyer Publishers, New York.

Corey, D. R., McGrath, M. E., Vasquez, J. R., Fletterick, R. J., & Craik, C. S. (1992) *J. Am. Chem. Soc. 114*, 4905−7.

Ding, L., Coombs, G. S., Strandberg, L., Navre, M., Corey, D. R., & Madison, E. L. (1995) *Proc. Natl. Acad. Sci. U.S.A. 92*, 7627−7631.

Forsberg, G., Brobjer, M., Holmgren, E., Bergdahl, K., Persson, P., Gautvik, K. M., & Hartmanis, M. (1991) *J. Protein Chem. 10*, 517−26.

Haber, E., Quertermous, T., Matsueda, G. R., & Runge, M. S. (1989) *Science 243*, 51−6.

Hanson, S., Adelman, J., & Ullman, B. (1992) *J. Biol. Chem. 267*, 2350−9.

Hedstrom, L., Szilagyi, L., & Rutter, W. J. (1992) *Science 255*, 1249−53.

Kitamoto, Y., Yuan, X., Wu, Q., McCourt, D. W., & Sadler, J. E. (1994) *Proc. Natl. Acad. Sci. U.S.A. 91*, 7588−92.

Light, A., Savithri, H. S., & Liepnieks, J. J. (1980) *Anal. Biochem. 106*, 199−206.

Madison, E. L., Coombs, G. S., Corey, D. R. (1995) *J. Biol. Chem. 270*, 7558−62.

Matthews, D. J., & Wells, J. A. (1993) *Science 260*, 1113−7.

McGrath, M. E., Erpel, T., Bystroff, C., & Fletterick, R. J. (1994) *EMBO J. 13*, 1502−1507.

Nagai, K., & Thogersen, H. C. (1987) *Methods Enzymol. 153*, 461−81.

Narayana, S. V., Carson, M., El-Kabbani, O., Kilpatrick, J. M., Moore, D., Chen, X., Bugg, C. E., Volanakis, J. E., & DeLucas, L. J. (1994) *J. Mol. Biol. 235*, 695−708.

Osterman, A., Grishin, N. V., Kinch, L. N., & Phillips, M. A. (1995) *Biochemistry 33*, 13662−7.

Parks, T. D., Leuther, K. K., Howard, E. D., Johnston, S. A., & Dougherty, W. G. (1994) *Anal. Biochem. 216*, 413−7.

Perona, J. J., & Craik, C. S. (1995) *Protein Sci. 4*, 337−360.

Ponder, J. W., & Richards, F. M. (1987) *J. Mol. Biol. 193*, 775−91.

Sahin-Tóth, M., Dunten, R. L., & Kaback, H. R. (1995) *Biochemistry 34*, 1107−12.

Schechter, I., & Berger, A. (1968) *Biochem. Biophys. Res. Commun. 27*, 157−162.

Schellenberger, V., Turck, C. W., Hedstrom, L., & Rutter, W. J. (1993) *Biochemistry 32*, 4349−53.

Vu, T. K., Wheaton, V. I., Hung, D. T., Charo, I., & Coughlin, S. R. (1991) *Nature 353*, 674−7.

Walker, P. A., Leong, L. E.-C., Ng, P. W. P., Tan, S.-H., Waller, S., Murphy, D., & Porter, A. G. (1994) *Biotechnology 12*, 601−5.

Willett, W. S., Gillmor, S. A., Perona, J. J., Fletterick, R. J., & Craik, C. S. (1995) *Biochemistry 34*, 2172−80.

Yang, W. P., Goldstein, J., Procyk, R., Matsueda, G. R., & Shaw, S. Y. (1994) *Biochemistry 33*, 606−12.

BI9505171