Proc. Natl. Acad. Sci. USA
Vol. 87, pp. 6659–6663, September 1990
Biochemistry

Substrate specificity of trypsin investigated by using a genetic selection

(catalysis/mutagenesis/enzyme kinetics/proteases)

LUKE B. EVNIN*, JOHN R. VÁSQUEZ†, AND CHARLES S. CRAIK*†‡

Departments of *Biochemistry and Biophysics and †Pharmaceutical Chemistry, University of California, San Francisco, CA 94143

Communicated by Stephen J. Benkovic, June 1, 1990 (received for review March 20, 1990)

ABSTRACT **The structural determinants of the primary substrate specificity of rat anionic trypsin were examined by using oligonucleotide-directed mutagenesis coupled to a genetic selection. A library was created that encoded trypsins substituted at amino acid positions 189 and 190 at the base of the substrate binding pocket. A genetic selection, with a dynamic range of 5 orders of proteolytic activity, was used to search 90,000 transformants of the library. Rapid screening for arginyl amidolysis and esterolysis confirmed the activity of the purified isolates. Trypsin and 15 mutant trypsins with partially preserved function were identified and characterized kinetically on arginyl and lysyl peptide substrates. Alternative arrangements of amino acids in the substrate binding pocket sustained efficient catalysis. A negative charge at amino acid position 189 or 190 was shown to be essential for high-level catalysis. With the favored aspartic acid residue at position 189, several amino acids could replace serine at position 190. Modulation of the specificity for arginine and lysine substrates was shown to depend on the amino acid at position 190. The regulatory effect of the amino acid side chain at position 190 on the substrate specificity is also reflected in substrate binding pockets of naturally occurring trypsin homologs.**

The substrate specificity of an enzyme is determined by its capacity to form a stable complex with a particular ligand in both the ground state and the transition state. Steric constraints, intermolecular forces, and the hydrophobic effects each contribute to the stability of a given complex. Examination of three-dimensional structures of the enzyme–inhibitor and enzyme–pseudosubstrate complexes can reveal important hydrogen bonds, van der Waals contacts, and electrostatic interactions without indicating the magnitude of these effects in a specific reaction. Furthermore, determination of how and why a particular configuration of amino acids specifies a function is not currently possible from three-dimensional structure analysis alone. Critical components of substrate specificity in several enzymes have been studied by making amino acid substitutions in the structure and characterizing the functional consequences of the alterations (1–6). Though powerful, this approach generally requires a prior assessment of the importance of the substituted amino acid. A genetic selection can significantly expedite analysis as it provides a means to search a large number of protein structures for those that satisfy a functional requirement, making it unnecessary to economize on the number of positions or the number of mutants that may be sampled (7, 8).

Trypsin is a paradigm for the family of serine proteases that have evolved to cleave peptide bonds after Arg and Lys amino acid residues (9–11). The substrate specificity exhibited by this enzyme family toward these two structurally disparate, positively charged amino acids is defined by the topography of the enzyme, particularly the primary specificity pocket. Crystallographic analysis of trypsin–inhibitor complexes (12–14) and mutagenesis of amino acids that comprise the substrate binding pocket (1, 15, 16) indicated that two positions, 189 and 190 (trypsin amino acid numbering is based on the chymotrypsinogen amino acid numbering described in ref. 17), are critical in defining the substrate specificity of the enzyme. It was anticipated that substitution at these positions would alter function, with a minimum disruption of the structure of the binding pocket, because the remainder of the interactions between the enzyme and the substrate side chain are mediated by main-chain atoms (1).

We wished to examine the functional contribution to Arg and Lys specificity of the two amino acids at the base of the substrate binding pocket. A library encoding trypsins substituted at positions 189 and 190 was constructed, and a genetic selection was developed to test the activity of these proteins. A set of mutant trypsins with partially preserved function was isolated and kinetically characterized to investigate the components of Arg and Lys specificity.

MATERIALS AND METHODS

Materials. Restriction enzymes, T4 DNA ligase, and T4 DNA polymerase were purchased from New England Biolabs or Promega. Modified T7 DNA polymerase was purchased from United States Biochemical, isopropyl β-D-thiogalactopyranoside (IPTG), 4-methylumbelliferyl *p*-guanidinobenzoate, and 4-methylumbelliferyl *p*-(*N*,*N*,*N*-trimethylammonium) cinnamate chloride were purchased from Sigma. 7-Amino-4-methylcoumarin (AMC) was purchased from Enzyme Systems Products (Livermore, CA). Arginine β-naphthylamide, N^α-benzyloxycarbonyl-Gly-Pro-Arg-AMC, N^α-*p*-tosyl-Gly-Pro-Arg-AMC (Tos-GPR-AMC), N^α-*p*-Tos-Gly-Pro-Lys-AMC (Tos-GPK-AMC) and *t*-Boc-Leu-Gly-Arg *p*-nitroanilide were purchased from Bachem. Purified *Streptomyces griseus* trypsin was a generous gift of M. N. G. James (University of Alberta, Canada).

Strains and Plasmids. *Escherichia coli* X90 [F′ *lacI^q*, *lacZY*, *proAB*/Δ(*lac-pro*), *ara*, *nalA*, *argE* (am), *thi*, rif^r] (18) and *E. coli* SY903 (F′ *lacI^q*, *lacZ*::Tn*5*, *proAB*/*recA1*, *srl*::Tn*10*, Δ*lac*–pro, rif^r, *nalA*) (19) were gifts from A. Vershon (University of California, San Francisco). *E. coli* CJ236 (F′ *CJ105*/*dut1*, *ung1*, *relA1*) was a gift of Bio-Rad. *E. coli* LE112 (F′ *lacI^q*, *lacZ*::Tn*5*, *proAB*/*dut1*, *ung1*, *relA1*) contained the episome from strain SY903 and the chromosome from strain CJ236.

Three vectors were used in this study: pT3, p^mT3, and pLIB. Overexpression of rat anionic trypsin (20) was achieved by using a *tac* promoter and the *hisJ* signal peptide in pT3 (21). p^mT3 was created by replacing the 702-base-pair *Sca* I–*Eco*RV DNA fragment from pT3 with the 956-base-pair

The publication costs of this article were defrayed in part by page charge payment. This article must therefore be hereby marked "*advertisement*" in accordance with 18 U.S.C. §1734 solely to indicate this fact.

Abbreviations: IPTG, isopropyl β-D-thiogalactopyranoside; AMC, 7-amino-4-methylcoumarin; Tos-GPK-AMC, N^α-*p*-tosyl-Gly-Pro-Lys-AMC; Tos-GPR-AMC, N^α-*p*-tosyl-Gly-Pro-Arg-AMC.

‡To whom reprint requests should be addressed.

Sca I–*Pvu* II DNA fragment from pBluescript(−) (Stratagene) containing the M13 origin of replication. pLIB was constructed from p^{m}T3 by eliminating the unique *Sph* I site and deleting eight bases encoding the terminal two bases of the position 188 codon and all six bases encoding the amino acids at positions 189 and 190 in trypsin. This step incorporated a unique *Sph* I site at the location of the deletion. pLIB does not express trypsin as determined by immunoblot analysis.

Library Construction. Mutagenesis of pLIB with a pool of oligonucleotides comprising all possible nucleotide combinations at the positions encoding amino acids 189 and 190 (5′-CTAGAGGGAGGCAAGNNNNNNTGCCAGGGC-GACTCTGGTGGGCC-3′; N = A, C, G, or T) was performed as described (22) except that uracil-laden single-stranded DNA was isolated from strain LE112. It was anticipated that each of the oligonucleotides in the pool would anneal with roughly equal efficiency to the pLIB template since the oligonucleotides had an equal number of bases complementary to the template. Unexpectedly, a skewed library was recovered from this mutagenesis reaction. DNA sequence analysis of 25 randomly selected isolates showed that guanosine and adenosine are overrepresented by as much as 2-fold in each of the mutated positions (L.B.E. and C.S.C., unpublished data). The loop formed by the annealed oligonucleotide may be stabilized by the stacking of the purine bases.

Electrotransformation of the DNA into a concentrate of *E. coli* X90 cells was achieved with narrow-gapped cuvettes (0.2 cm) and the Bio-Rad Gene Pulser equipped with a pulse controller (23). The transformed cells were plated onto nitrocellulose discs overlaid on LB plates containing ampicillin at 50 mg/liter. Plasmid DNA was harvested from 2.5×10^5 colonies and subjected to *Sph* I digestion. This DNA encoded the library of trypsin mutants. A clonal population of each mutant was ensured by retransformation into X90 cells prior to selection for active trypsin.

Selection. Selection for trypsin activity was optimal on Mops minimal media (24) using 0.4% glucose as a carbon source, 0.01% vitamin B_1, 1 mM IPTG, and 300 μM arginine β-naphthylamide at 32°C. Addition of ampicillin at 50 mg/liter was not required for selection but was included to prevent the growth of untransformed *ArgE* revertants. Optimal discrimination between background colonies expressing inactive trypsin and survivors expressing active trypsin was achieved in the absence of the free Arg supplementation, although the time required to achieve the discrimination was reduced from roughly 72 to 48 hr by addition of 1 μM Arg to the selection medium.

Screening. Trypsin was harvested from 1.5 ml of an *E. coli* X90 culture grown to saturation in LB/ampicillin (50 mg/liter)/1 mM IPTG (22). For the microplate assay, 20 μl of the 40 μl of extract was added to 180 μl of 100 mM Tris·HCl, pH 8.0/0.4 mM *t*-Boc-Leu-Gly-Arg *p*-nitroanilide in 1 well of a 96-well microplate. Trypsin activity was measured in duplicate by following the release of *p*-nitroaniline at 405 nm. For the single substrate gel overlay screen, the activity was assayed as described (22).

Enzyme Purification. *E. coli* X90 cells containing the desired mutant plasmid were inoculated into 500 ml of LB/ampicillin (50 mg/liter)/1 mM IPTG and cultured for 12–16 hr. The periplasmic proteins of these cells were harvested in 25% sucrose/10 mM Tris·HCl, pH 8.0/5 mM EDTA as described (25) and were dialyzed exhaustively at 4°C in 10 mM sodium citrate (pH 2.8). The acid precipitate was separated by centrifugation for 20 min at 12,000 × *g*, and the supernatant containing trypsin was concentrated by ultrafiltration in an Amicon Centriprep 10 (10,000 M_r cutoff). From ≈20 ml of dialysate, 0.5 ml of concentrate was obtained. This trypsin solution was centrifuged at 14,000 × *g* for 5 min, and the precipitate was discarded.

The concentration of active sites was determined by following the release of 4-methylumbelliferone from 4-methylumbelliferyl *p*-guanidinobenzoate (26). No burst was observed with an extract from cells carrying pLIB. Active trypsin constituted ≈0.5% of the total protein present in the purified extract as determined by measurement of the concentration of total protein (27) and the concentration of trypsin active sites.

Michaelis–Menten parameters were obtained from a kinetic analysis of acid-precipitated extract and of affinity-purified trypsin and trypsin S190G [trypsin mutants are designated by the wild-type amino acid residue (single-letter code) and its position number followed by the substituted amino acid] (28). The k_{cat}, K_m, and Arg/Lys specificity constants of the two purified forms were comparable. The hydrolysis rate of N^{α}-benzyloxycarbonyl-Gly-Pro-Arg-AMC by extracts from cells containing pLIB was negligible.

RESULTS

Selection. The genetic selection relies on the demonstrated ability of *E. coli* to secrete active trypsin into the periplasm (22) and on the inability of *E. coli* to metabolize a β-naphthylamide derivative of an amino acid (29). *E. coli* X90 cells are auxotrophic for Arg since they lack one of the enzymes required for Arg biosynthesis. Consequently, arginine β-naphthylamide cannot be an anabolic source of Arg. Trypsin readily cleaves the arginyl naphthylamide bond. Thus, when X90 cells are grown on a medium containing arginine β-naphthylamide as the only source of Arg, they will survive only if they express active, Arg-specific trypsin.

Trypsin is expressed from vector p^{m}T3 with an amino-terminal signal peptide that directs trypsin into the periplasm and serves as a propeptide (21, 22, 25). To assess the level of trypsin activity required to confer survival, *E. coli* expressing trypsin or a mutant trypsin were plated on selective media. In addition to trypsin, the trypsin mutants D189E, D189C, D189S, D189N, and D189A were sufficiently active to scavenge enough Arg from the arginine β-naphthylamide to allow growth of the host. Trypsin D189S is reduced ≈5 orders of magnitude in k_{cat}/K_m (16). The mutants with His, Ile, Thr, or Gln at position 189 are less active than trypsin D189S as determined by activity gel analysis and failed to support growth. These tests indicated that the dynamic range of the selection spans 5 orders of magnitude of proteolytic activity.

Identification of Active Trypsin Mutants. A library encoding the set of trypsins substituted at positions 189 and 190 was constructed by region-specific mutagenesis. Three separate searches, encompassing 90,000 transformants (exclusive of transformants that contained the unmutated, parental plasmid) were conducted. The nucleotide sequence of the muta-

Table 1. Specificity of trypsin mutants isolated by selection

Position		Arg/Lys	Position		Arg/Lys
189	190	preference	189	190	preference
Asp	Ser	4.2	Ser	Asp	2.28
Asp	Cys	4.7	Ser	Glu	1.00
Asp	Thr	3.8	Gly	Asp	0.08
Asp	Val	2.2	Gly	Glu	0.20
Asp	Gly	40.4	Asn	Ser	—
Asp	Pro	135.5	Asn	Thr	—
			Ala	Ser	—
Glu	Ser	0.12			
Glu	Thr	0.10			
Glu	Ala	1.3			

Specificity of the 16 trypsins isolated from the library. A kinetic characterization of the 15 mutants and trypsin was completed on Tos-GPR-AMC and Tos-GPK-AMC. The Arg/Lys preference is the k_{cat}/K_m measured for hydrolysis of the Tos-GPR-AMC substrate divided by the k_{cat}/K_m measured for the Tos-GPK-AMC substrate.

tions was determined for 55 of 68 mutants in which the trypsin activity was confirmed; 16 trypsins were identified (Table 1).

Of the 68 active mutants, 38 were obtained in the third of the three searches in which conditions had been optimized. In this particular search, 11,000 transformants were initially searched to generate 195 potential positives. Of these, 121 isolates were confirmed by microplate screening analysis, and 38 were further confirmed by activity gel analysis. Sequence analysis of 28 of the 38 positives revealed 13 trypsins. In the two previous searches, a total of 10 trypsins had been identified. Among these 23 trypsins, there were 16 different enzymes.

A trend is that a negatively charged amino acid at position 189 or 190 is important for activity (Table 1). The preferred amino acid is Asp. The partner to Asp or Glu is from a limited set of amino acids. The principal limitation is probably the volume available at the base of the pocket because Asp-189 and Ser-190 are close packed in trypsin (13). Three of the mutants lack a negative charge at either position, trypsin D189A, D189N, and the double mutant D189N/S190T. The activity of these enzymes is at the level of detection of the selection (see below).

Kinetic Analysis of Mutant Trypsins. The trypsin mutants were characterized by analysis of the kinetic parameters k_{cat}, K_m, and k_{cat}/K_m for comparable tripeptide Arg and Lys substrates (Fig. 1). Relative to trypsin, all mutants except trypsin S190G have diminished k_{cat} values (Fig. 1*A*). Fig. 1*B* shows that except for trypsin S190T, all mutants exhibit elevated K_m values (Fig. 1*B*). All mutants exhibit a reduction in k_{cat}/K_m for both Arg and Lys substrates (Fig. 1*C*). Only the second-order rate constant k_{cat}/K_m could be obtained for the weakly active trypsins D189N, D189N/S190T, and D189A.

A Negative Charge Is Critical for Efficient Catalysis. Trypsins D189E and D189E/S190T preferred the Lys to the Arg substrate by ≈40-fold relative to trypsin (Table 1). Apparently, the additional methylene group separating the D189E carboxylate from the protein backbone was a handicap in the catalysis of Arg, which has a longer and larger side chain compared with the Lys side chain, leading to the reversal of the relative preference for Arg exhibited by

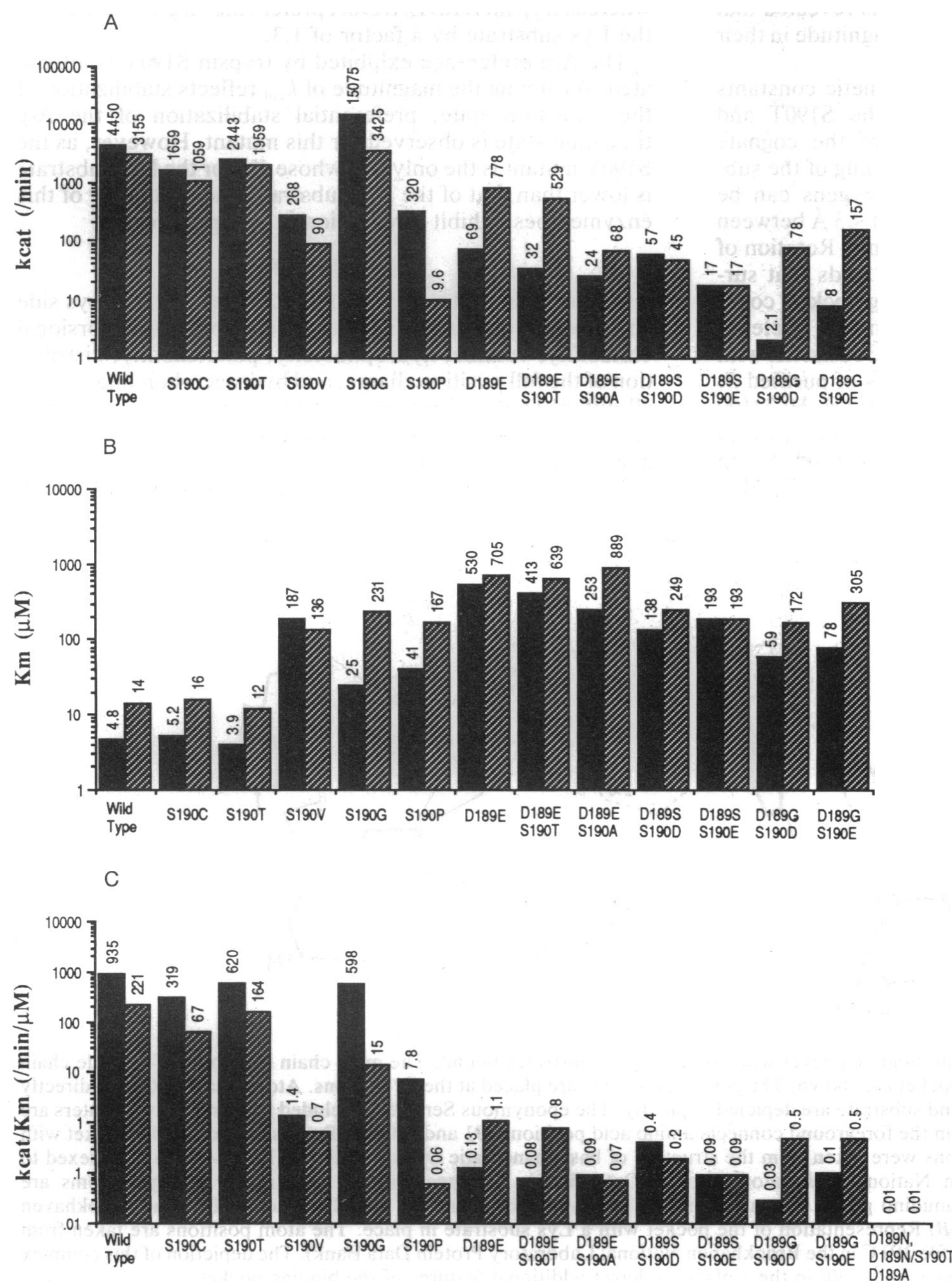

FIG. 1. Kinetic parameters of trypsin and trypsin mutants. Bars represent data from duplicate determinations on Tos-GPR-AMC (solid bars) or Tos-GPK-AMC (hatched bars). Five substrate concentrations were employed per determination: 200, 100, 50, 25, and either 12.5 or 500 μM. Substrates were dissolved in dimethyl formamide. AMC release was followed in a Perkin–Elmer LS-5B luminesence spectrometer, which was calibrated with a standard solution of AMC, at 460 nm with excitation at 380 nm. Enzyme was added to a final concentration of 2.5–150 nM. The final concentration of buffer was 50 mM Tris·HCl, pH 8.0/100 mM NaCl, 1% dimethyl formamide (26°C). Velocities were obtained from the linear portion of the free AMC vs. time curve. The kinetic parameters K_m and k_{cat} were obtained by using a linear regression analysis of the Eadie–Hofstee plot of v vs. v/[S]. Numerical values are noted above each bar. The error in these determinations is <20% and in most cases <10%. Only the k_{cat}/K_m values could be obtained for trypsin D189N, D189A, and D189N/S190T because of their reduced activity; determination of the concentration of active sites for these enzymes could not be obtained with either 4-methylumbelliferyl *p*-guanidinobenzoate or the chymotrypsin titrant 4-methylumbelliferyl *p*-(*N*,*N*,*N*-trimethylammonium)cinnamate chloride (26). For purposes of comparison, immunoblot analysis was used to determine the concentration of trypsin in these samples. (*A*) k_{cat} values (in units of min^{-1}). (*B*) K_m values (in units of μM). (*C*) k_{cat}/K_m values (in units of min$^{-1}\cdot\mu$M^{-1}).

trypsin. The increases in relative Lys specificity for the D189E and D189E/S190T mutants are largely a result of the differential effects on k_{cat} and presumably reflect selective stabilization of the transition state involving the Lys substrate. Our results are consistent with the investigation of trypsin's kinetic parameters with homoarginine (30). Our results differ from those previously obtained for trypsin D189E (31) and may reflect the different substrates and conditions used.

In contrast, Asp and Glu functioned similarly at position 190. The K_m, k_{cat}, and k_{cat}/K_m of the D189S mutants were very similar with Asp or Glu at position 190. The D189G/S190D and D189G/S190E mutants also exhibited very similar kinetic parameters (Fig. 1). Presumably, the hydrogen-bonding arrangement between enzyme and substrate has been modified, although the electrostatic properties of the position 189 and 190 duo were retained.

The substrate binding pocket can operate at a minimal efficiency on positively charged substrates without any negative charge at its base. Trypsins D189N, D189N/S190T, and D189A retain sufficient activity to be identified in searches through the library. However, kinetic analysis revealed that these mutants are reduced by ≈5 orders of magnitude in their k_{cat}/K_m values (Table 1).

Substitution for Ser at Position 190. The kinetic constants on both the Lys and Arg substrate for the S190T and D189E/S190T enzymes are within 2-fold of the cognate enzymes containing Ser-190. Computer modeling of the substitution of Thr for Ser indicates the O^{γ} oxygens can be superimposed by introducing only one contact <3 Å between the Thr C^{γ} methyl group and neighboring atoms. Rotation of Ile-138, which resides in the shell of amino acids that surrounds the amino acids of the primary binding pocket, could relieve this short contact. *S. griseus* trypsin is a bacterial homolog of trypsin with Thr at position 190. Consistent with our analysis of trypsin S190T, kinetic analysis of purified *S. griseus* trypsin at pH 7.0 and 26°C in 50 mM Tris·HCl/100 mM NaCl showed that this enzyme prefers Tos-GPR-AMC to Tos-GPK-AMC by a factor of 3.8 (Arg k_{cat} = 4500 min^{-1}, Arg K_m = 1.8 μM; Lys k_{cat} = 3100 min^{-1}, Lys K_m = 4.7 μM).

No known serine protease has Cys at position 190. However, the activity of trypsin S190C is reduced only 2.5-fold relative to trypsin. The K_m is essentially unchanged. Apparently, Cys can replace this Ser with impunity in spite of the difference in hydrophobicity and the larger atomic radius of the sulfur atom (1.85 Å versus 1.4 Å). Interaction between the position 190 sulfhydryl and the sulfhydryls comprising the adjacent disulfide bond (positions 191 and 220; Fig. 2) were not detected.

Differential Arg/Lys Substrate Specificity. With the tripeptide substrates Tos-GPR-AMC and Tos-GPK-AMC, trypsin maintains a 4-fold Arg/Lys preference (Table 1). For trypsins S190G and S190P, the Arg substrate is preferred 40- and 135-fold over the Lys substrate. For the latter, 1% and 0.03% of the Arg and Lys k_{cat} are retained. The general decrease in trypsin S190P activity may be due to steric clashes between the proline ring and the side chains of Tyr-228 and Asp-189. Apparently, deletion of the hydroxyl group from the side chain of position 190 has caused a substantial modulation of the Arg and Lys substrate selectivity. Similarly, trypsin D189E and D189E/S190T strongly prefer the Lys substrate, whereas trypsin D189E/S190A prefers the Arg substrate over the Lys substrate by a factor of 1.3.

The Arg preference exhibited by trypsin S190V is attenuated. Assuming the magnitude of k_{cat} reflects stabilization of the transition state, preferential stabilization of the Arg transition state is observed for this mutant. However, as the S190V mutant is the only one whose K_m for the Lys substrate is lower than that of the Arg substrate, the specificity of this enzyme does exhibit a reduction in Arg preference.

DISCUSSION

Aside from their identity in formal charge, Arg and Lys side chains are structurally disparate. The shape, length, torsional variability, volume, hydrophobicity, polarizability, distribution of the full positive charge, and hydrogen-bonding pattern all differ. Thus, an enzyme that has evolved to cleave both Arg and Lys substrates must take advantage of the similarity and yet have additional features that allow it to be specific for each side chain. Trypsin has a negative electrostatic field in

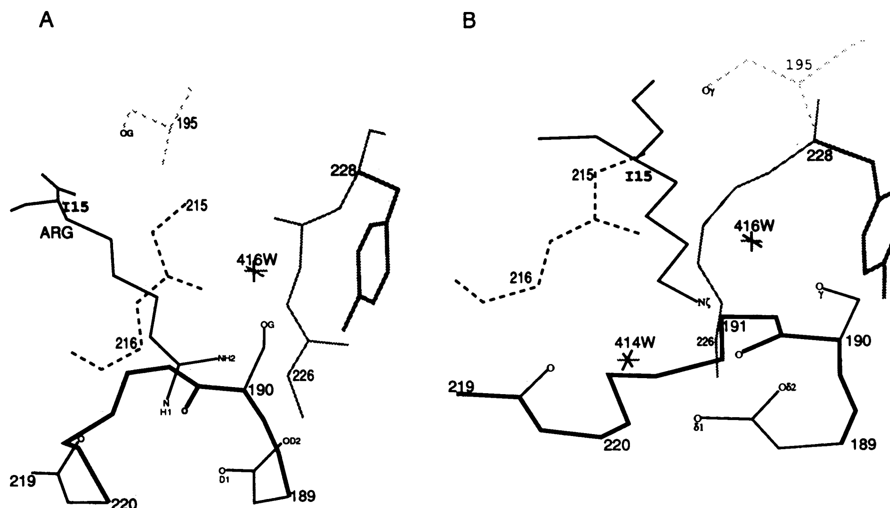

FIG. 2. Depiction of the trypsin substrate binding pocket with Arg and Lys substrates bound. The main chain and some of the side chain atoms that comprise the substrate binding pocket are shown. The position numbers are placed at the C^{α} carbons. Atoms that contribute directly to the hydrogen bonding between enzyme and substrate are depicted explicitly. The eponymous Ser-195 is included for perspective. Waters are denoted as W. The disulfide bond pictured in the foreground connects amino acid positions 191 and 220. (*A*) Representation of the pocket with an Arg substrate in place. The atom positions were taken from the structure of basic pancreatic trypsin inhibitor (BPTI)-Arg15 complexed to trypsinogen (file 4TPI in the Brookhaven National Laboratory Protein Data Bank). The positions of the Arg N^{η} nitrogen atoms are superimposable with the nitrogens of the amidino group of benzamidine in the trypsin–benzamidine complex (file 3PTB in the Brookhaven National Laboratory Protein Data Bank). (*B*) Representation of the pocket with a Lys substrate in place. The atom positions are taken from the structure of BPTI complexed to trypsin (file 2PTC in the Brookhaven National Laboratory Protein Data Bank). The depiction of this complex relative to Fig. 2*A* is rotated 30° in the *y* plane and −30° in the *x* plane to depict additional features of the binding pocket.

the substrate binding pocket (32). This negative field attracts the solvated Arg and Lys substrate side chains and helps stabilize the positive charge in the enzyme–substrate complex in which solvent is excluded. The importance of this charge is underscored by the impotency of the D189N and D189N/S190T trypsins.

Trypsin is responsible for a rate enhancement for proteolysis of roughly ten billionfold over the rate observed in water (33). Since the sole constraint on substitution is volume at position 190, it must be concluded that the chemical characteristics of the side chain at position 190 are unimportant for the bulk of the rate enhancement observed; Gly, Ala, Thr, Cys, Val, and Pro can replace Ser. However, a higher resolution analysis of the kinetic parameters of the trypsin Ser-190 substitutions revealed that the absence of a substantial prejudice for Arg or Lys is a carefully crafted feature of trypsin. Presumably, the substantial decreases in Lys specific activity of mutant S190G and S190P (and of mutant D189E/S190A relative to D189E) result from destruction of the critical hydrogen bond triad among the water molecule (water 416 in Fig. 2*B*), Ser-190 O^{γ}, and substrate Lys N^{ζ} when there is no γ hydroxyl group at position 190. Arg specificity is unperturbed because the critical cyclic network of hydrogen bonds between the guanidinium group of Arg and the carboxylate of Asp-190 is maintained and is the dominant feature of Arg substrate specificity. The differential effects on Lys and Arg specificity can be observed in both K_m and k_{cat}. However, unappreciated conformational changes elsewhere in the molecule may play a role in these effects. Clearly, additional structural and functional characterization of these enzymes is required for a more complete understanding of the functional consequences of these single and double mutations.

Our conclusions are supported by an analysis of the preference for Arg or Lys substrates exhibited by five trypsin homologs. *S. griseus* trypsin exhibited a 3.8-fold preference for Tos-GPR-AMC relative to Tos-GPK-AMC, which is nearly identical to that exhibited by rat trypsin. Thrombin and plasma kallikrein exhibit a substantial Arg/Lys substrate preference relative to trypsin (34). On a set of single residue ester and amide substrates, thrombin preferred Arg substrates by a factor of 7–10; kallikrein preferred Arg substrates by a factor of 15–16. As with the S190P, S190G and D189E/S190A, and S190V mutants, both thrombin and kallikrein do not have a γ position hydroxyl group at position 190; both have Ala at position 190. The x-ray crystal structure solutions for trypsin, thrombin, and a kallikrein show that the main-chain atoms of the substrate binding pocket are virtually superimposable. Furthermore, except for the change noted, all the amino acids in the first shell of the binding pocket are identical. The known Lys-specific trypsin homologs urokinase (34) and plasmin (35) have retained the Asp at position 189 and the Ser at position 190. Urokinase prefers Lys substrates by a factor of $\approx$100, although trypsin has 1000-fold the activity of this homolog on Lys substrates (34). The kinetic parameters for urokinase more closely match those of trypsin D189E than trypsin. The Michaelis–Menten constants for plasmin have yet to be described. Currently, no three-dimensional structure is available for either Lys-specific homolog.

This genetic selection provides a powerful tool for dissecting the elements of substrate specificity and, also, an avenue for the isolation of enzymes with a prescribed function without prior description of the critical structural elements. Ultimately, enzymes of a desired specificity differing from any known natural specificity could be designed using existing enzyme scaffolds if a complete understanding of the structural determinants of the function was extant.

We thank Professors Robert Fletterick, Robert Stroud, and Ira Herskowitz and Dr. Jeffrey Higaki for helpful discussions. This work was supported by National Science Foundation Grants DMB 8904956 and EET 8807179 to C.S.C. L.B.E. was also supported by National Institutes of Health Training Grant T32-GM 07810 and the Lucille P. Markey Charitable Trust, Miami, FL. J.R.V. was supported by the President's Dissertion Year Fellowship Program of University of California, San Francisco.

1. Craik, C. S., Largman, C., Fletcher, T., Roczniak, S., Barr, P. J., Fletterick, R. & Rutter, W. J. (1985) *Science* **228,** 291–297.
2. Cronin, C. N., Malcolm, B. A. & Kirsch, J. F. (1987) *J. Am. Chem. Soc.* **109,** 2222–2223.
3. Estell, D. A., Graycar, T. P., Miller, J. V., Powers, D. B., Burnier, J. P., Ng, P. N. & Wells, J. A. (1986) *Science* **233,** 659–663.
4. Wilks, H. M., Hart, K. W., Feeney, R., Dunn, C. R., Muirhead, H., Chia, W. N., Barstow, D. A., Atkinson, T., Clarke, A. R. & Holbrook, J. J. (1988) *Science* **242,** 1541–1544.
5. Bone, R., Silen, J. L. & Agard, D. A. (1989) *Nature (London)* **339,** 191–195.
6. Scrutton, N. S., Berry, A. & Perham, R. N. (1990) *Nature (London)* **343,** 38–43.
7. Reidharr-Olson, J. F. & Sauer, R. T. (1988) *Science* **241,** 53–57.
8. Hermes, J. D., Blacklow, S. C. & Knowles, J. R. (1990) *Proc. Natl. Acad. Sci. USA* **87,** 696–700.
9. Stroud, R. M. (1974) *Sci. Am.* **231,** 74–88.
10. Kraut, J. (1977) *Annu. Rev. Biochem.* **46,** 331–358.
11. Steitz, T. & Shulman, R. (1982) *Annu. Rev. Biophys. Bioeng.* **11,** 419–444.
12. Ruhlmann, A., Kukla, D., Schwager, P., Bartels, K. & Huber, R. (1973) *J. Mol. Biol.* **77,** 417–436.
13. Kreiger, M., Kay, L. M. & Stroud, R. M. (1974) *J. Mol. Biol.* **83,** 209–230.
14. Bode, W., Schwager, P. & Huber, R. (1978) *J. Mol. Biol.* **118,** 99–112.
15. Gráf, L., Craik, C. S., Patthy, A., Roczniak, S., Fletterick, R. J. & Rutter, W. J. (1987) *Biochemistry* **26,** 2616–2623.
16. Gráf, L., Jancsó, A., Szilágyi, L., Hegyi, G., Pintér, K., Náray-Szabó, G., Hepp, J., Medzihradszky, K. & Rutter, W. J. (1988) *Proc. Natl. Acad. Sci. USA* **85,** 4961–4965.
17. Hartley, B. S., Brown, J. R., Kauffman, D. L. & Smillie, L. B. (1965) *Nature (London)* **207,** 1157–1159.
18. Amann, E., Brosius, J. & Ptashne, M. (1983) *Gene* **25,** 167–178.
19. Sauer, R. T., Smith, D. L. & Johnson, A. D. (1988) *Genes Dev.* **2,** 807–816.
20. Craik, C. S., Choo, Q.-L., Swift, G., Quinto, C., MacDonald, R. & Rutter, W. (1984) *J. Biol. Chem.* **239,** 14255–14264.
21. Higaki, J. N., Evnin, L. B. & Craik, C. S. (1989) *Biochemistry* **28,** 9256–9263.
22. Evnin, L. B. & Craik, C. S. (1988) *Ann. N.Y. Acad. Sci.* **542,** 61–74.
23. Dower, W. J., Miller, J. F. & Ragsdale, C. W. (1988) *Nucleic Acids Res.* **16,** 6127–6145.
24. Neidhart, F. C., Bloch, P. L. & Smith, D. F. (1974) *J. Bacteriol.* **119,** 736–747.
25. Vásquez, J. R., Evnin, L. B., Higaki, J. N. & Craik, C. S. (1989) *J. Cell. Biochem.* **39,** 265–276.
26. Jameson, G. W., Adams, D. V. R. R. W., Kyle, W. S. A. & Elmore, D. T. (1973) *Biochem. J.* **131,** 107–117.
27. Smith, P. K., Krohn, R. I., Hermanson, G. T., Mallia, A. F., Gartner, F. H., Provenzano, M. D., Fujimoto, E. K., Goeke, N. M., Olson, B. J. & Klenk, D. C. (1985) *Anal. Biochem.* **150,** 76–85.
28. Knowles, J. R. (1987) *Science* **236,** 1252–1258.
29. Miller, C. G., Green, L. & Schultz, R. (1982) *Mol. Gen. Genet.* **186,** 228–234.
30. Baines, N. J., Baird, J. B. & Elmore, D. T. (1964) *Biochem. J.* **90,** 470–476.
31. Gráf, L., Hegyi, G., Likó, I., Hepp, J., Medzihradszky, K., Craik, C. S. & Rutter, W. J. (1988) *Int. J. Pept. Protein Res.* **32,** 512–518.
32. Soman, K., Yang, A.-S., Honig, B. & Fletterick, R. (1989) *Biochemistry* **28,** 9918–9923.
33. Kahne, D. & Still, W. C. (1988) *J. Am. Chem. Soc.* **110,** 7529–7534.
34. Ascenzi, P., Menegatti, E., Guarneri, M., Bortolotti, F. & Antonini, E. (1982) *Biochemistry* **21,** 2483–2490.
35. Weinstein, M. J. & Doolittle, R. F. (1972) *Biochim. Biophys. Acta* **258,** 577–590.