Binding of Nonphysiological Protein and Peptide Substrates to Proteases: Differences between Urokinase-Type Plasminogen Activator and Trypsin and Contributions to the Evolution of Regulated Proteolysis†

Robert C. Bergstrom,‡ Gary S. Coombs,§ Sheng Ye,‡ Edwin L. Madison,§ Elizabeth J. Goldsmith,§ and David R. Corey*,§

Departments of Pharmacology and Biochemistry, University of Texas Southwestern Medical Center at Dallas, 5323 Harry Hines Boulevard, Dallas, Texas 75390-9041, and Department of Molecular Biology, Corvas International, 3030 Science Park Road, San Diego, California 92121

Received December 23, 2002; Revised Manuscript Received March 21, 2003

ABSTRACT: Understanding the regulation of physiological processes requires detailed knowledge of the recognition of substrates by enzymes. One of the most productive model systems for the study of enzyme−substrate interactions is the serine protease family; however, most studies of protease action have used small substrates that contain an activated, non-natural scissile bond. Because few kinetic or structural studies have used protein substrates, the physiologically relevant target of most proteases, it seems likely that important mechanisms of substrate recognition and processing by proteases have not yet been fully elucidated. Consistent with this hypothesis, we have observed that K_m values for protein substrates are reduced as much as 200−15000-fold relative to those of analogous peptide substrates. Here we examine the thermodynamic consequences of interactions between proteases and their substrates using staphylococcal nuclease (SNase) and SNase variants as model protein substrates. We have obtained values for enthalpy, entropy, and K_d for binding of proteins and peptides by the nonspecific protease trypsin and the highly specific protease urokinase-type plasminogen activator (u-PA). To avoid cleavage of substrates during these measurements, we used inactive variants of trypsin and u-PA whose catalytic serine S195 had been replaced by alanine. Differences in the K_d values for binding of protein and peptide substrates closely approximate the large differences observed in the corresponding K_m values. Improved binding of protein substrates is due to decreased enthalpy, and this effect is pronounced for the selective protease u-PA. Fundamental differences in recognition of analogous protein and peptide substrates may have influenced the evolution of protease specificity.

Serine proteases are an important model system for understanding how enzymes recognize and process protein substrates (*1*, *2*). It is reasonable to expect that the recognition of protein substrates will differ from recognition of small molecules or peptides, but the nature and extent of any differences have not been fully elucidated.

One reason for this gap in our knowledge of such a closely examined protein family is that little is known about how proteases interact with protein substrates. There is a large and important body of work on interactions of proteases with protein inhibitors (*3*−*6*). Inferences into modes of substrate binding can be made from these studies (*7*), but data from substrates containing chemically altered backbone atoms suggest that there are significant differences between the binding of substrates and the binding of inhibitors (*8*).

Our knowledge of protease−substrate interactions is not as complete as our knowledge of protease−inhibitor interactions because substrates are unstable in the presence of proteases, making it difficult to conduct structural studies or thermodynamic measurements. There have been many rigorous kinetic studies of proteolysis, but these have usually employed small peptide substrates, many with chemically activated chromogenic or fluorogenic leaving groups. Analysis of the cleavage of these small substrates is straightforward, and these studies have provided much valuable data on specificity and mechanism, but they cannot mimic the extensive interactions that likely occur between proteases and protein substrates.

Improvements in the understanding of the recognition of protein substrates by proteases would increase our ability to predict the location of cleavage sites and to engineer proteases for new substrate specificities. It is also likely that interactions between proteases and potential substrates proteins have contributed to the evolution of regulated proteolysis. More broadly, knowledge of how proteases interact with proteins may provide general insights into how other classes of enzymes, such as kinases, recognize peptide sequences within protein targets.

Our approach to understanding the recognition of substrate proteins by serine proteases initially focused on kinetic studies of hydrolysis of protein substrates (*9*−*11*). During

† This work was supported by grants from the American Heart Association (0140018N) to D.R.C. and from the Robert A. Welch Foundation (I-1128) to E.J.G.

* To whom correspondence should be addressed. Phone: (214) 648-5096. Fax: (214) 648-5095. E-mail: david.corey@utsouthwestern.edu.

§ University of Texas Southwestern Medical Center at Dallas.

‡ Corvas International.

10.1021/bi027417x CCC: $25.00
Published on Web 04/15/2003

Table 1: Michaelis Constants of u-PA, t-PA, Plasmin, and Trypsin for Analogous Peptide and Protein Substrates

	K_m (μM)			
sequence P4 to P2′ [a]	peptide	protein	ratio	ref
	t-PA			
HYGR↓SG	4010	18	220	*9*
QRGR↓SA	2300	18	260	*9*
GSGR↓SA	4080	17	240	*9*
PFGR↓SA	2200	9.9	220	*9*
PFGR↓SA [b]	3500	3.7	980	*10*
SPGR↓VV [b]	15000	1	15000	*10*
	u-PA			
HYGR↓SA	3800	30	127	*9*
QRGR↓SA	2180	≥130	≤17	*9*
GSGR↓SA	600	52	11	*9*
PFGR↓SA	2200	14	160	*9*
YGAK↓AY	2300	≥130	≤18	*c*
	Plasmin			
GIYR↓SR	100	5.8	17	*11*
	Trypsin			
HYGR↓SG	210	57	3.7	*c*
QRGR↓SA	150	95	1.6	*c*
GSGR↓SA	140	18	7.8	*c*
PFGR↓SA	260	39	6.7	*c*
YGAK↓AY	150	95	1.5	*c*

[a] All peptides were 14 amino acids long; only the P4 to P2′ residues are shown. The P4′ to P2′ residues are the same in both protein and peptide substrates; the flanking residues differ. [b] Sequences were introduced into ornithine decarboxylase. All other sequences were introduced into SNase. [c] Determined for the current study.

these studies, we noted that the K_m values for cleavage of protein substrates by serine proteases trypsin (*9*), urokinase-type plasminogen activator, (u-PA)[1] (*9*), tissue-type plasminogen activator (t-PA) (*9*, *10*), and plasmin (*11*) were from 200- to 15000-fold lower than the K_m values for hydrolysis of short peptides that share the same amino acid sequence surrounding the scissile bond (Table 1).

Our findings were surprising, since we had expected that the three-dimensional structure around the scissile bond would complicate protease recognition and increase K_m for the cleavage of protein substrates relative to analogous peptides. Rudolph and colleagues have also observed low K_m values for protein substrates relative to peptides for human Cdc 25, a protein phosphatase (*12*). It is possible, therefore, that insights into the origin of low K_m values for cleavage of proteins by proteases might have relevance to other classes of enzyme that use proteins as substrates.

One explanation for the reduced K_m values that we observe is that binding of protein substrates by proteases is enhanced relative to binding of peptide substrates. If this were true, it would suggest that binding of protein substrates has evolved to compensate for the need to reorganize the substrate protein structure at the proteolytic site. Tighter binding cannot be assumed, however, because K_m is dependent on the rate of substrate acylation and is not necessarily similar to K_d.

Here, we determine K_d values for recognition of protein and peptide substrates by u-PA and trypsin. We demonstrate that substrate binding can be enhanced when amino acid sequences are displayed in the context of a protein scaffold and that u-PA uses binding interactions to discriminate between protein and peptide substrates to a greater extent than does trypsin.

[1] Abbreviations: t-PA, tissue-type plasminogen activator; u-PA, urokinase-type plasminogen activator; SNase, staphylococcal nuclease; ITC, isothermal calorimetry.

MATERIALS AND METHODS

Enzymes and Substrates. S195A trypsin was initially expressed as the zymogen trypsinogen in *Saccharomyces cerevisiae* strain DLM101α, followed by proteolytic cleavage to generate trypsin. Trypsinogen was expressed using the vector pYT (obtained from Dr. L. Hedstrom, Brandeis University) (*13*). Peptides were synthesized and purified as described previously (*14*). Plasmid pONF1, which expresses wild-type staphylococcal nuclease (SNase), was mutated and purified as described previously (*9*). The molecular weight of SNase is 16 807 and all mutations were made within the ω loop.

S195A u-PA was expressed in *Pichia pastoris* using the pPic9k vector containing cDNA encoding the u-PA protease domain fused to an α-factor leader sequence that directs secretion into the growth medium (*13*). In addition to the S195A mutation, the u-PA construct contains a C122S mutation to remove the unpaired cysteine in the protease domain and an N145D mutation to remove an N-linked glycosylation site.

Starter cultures were grown in 20-mL volumes of BMGY medium for 24 h at 30 °C. One liter of BMGY medium was inoculated with the 20-mL starter culture and shaken at 300 rpm for 32 h. The cells were pelleted and then resuspended in 200 mL of BMMY for 16 h. The supernatant was concentrated to 100 mL, dialyzed into 50 mM NaH_2PO_4, 300 mM NaCl, 10 mM imidazole (pH 8.0), and shaken with 10 mL of Nickel-NTA resin from Qiagen (Valencia, CA). This mixture was poured into an empty 25-mL column, and the beads were washed two times with NaH_2PO_4, 300 mM NaCl, 20 mM imidazole (pH 8.0). After washing, u-PA was eluted with 5 mL of NaH_2PO_4, 300 mM NaCl, 250 mM imidazole (pH 8.0). u-PA-containing fractions were identified by SDS−PAGE, dialyzed into 50 mM NaCl, 50 mM Tris-Cl (pH 6.2), and loaded onto a column packed with soybean trypsin inhibitor−sepharose conjugate (Sigma). This column was equilibrated with 50 mM Tris (pH 6.2), loaded with protein, washed with 0.5 M NaCl in 50 mM Tris (pH 6.2), and eluted with 30 mM formic acid.

Analytical Ultracentrifugation. K_d values were determined using a Beckman Optima XL-I ultracentrifuge. To obtain K_d values for the association of S195A trypsin and SNase or peptides, various stoichiometric ratios of the two components were run at rotor speeds of 13000−17000 rpm. Samples reached equilibrium after ∼16 h; equilibrium was verified by the overlay of two radial scan traces acquired 3 h apart. Baselines were obtained by overspeeding at the end of each experiment and recording the residual absorbance. The IDEAL1 version of Beckman data analysis software was used to estimate apparent weight-average molecular masses ($M_{w.app}$) and to indicate any nonideality or associative/dissociative behavior of the trypsin and SNase and complexes of the two. K_d values were determined using the program NONLIN to obtain a global fit of the weight-average molecular masses plotted versus ultracentrifuge cell loading concentrations expressed in molar concentration terms (*15*,

16). K_d values were determined with the assumption that the nonideality of the complex is due entirely to the complex and there are no contributions from the reactants or cross-terms involving the reactants and complex.

Isothermal Titration Calorimetry. Titrations were performed with a VP-ITC microcalorimeter (MicroCal Inc., Northampton, MA). The stirred cell contained protease at 30−50 μM, and the syringe contained 300−500 μM peptide or SNase substrate. Experiments were performed in 100 mM NaCl, 50 mM Hepes (pH 7.5), 20 mM $CaCl_2$ at 37 °C (experimental temperatures were within 0.02 °C of the nominal values). To prevent buffer effects during enthalpy measurements, the dialysis of the protease and "ligand" samples used the same buffer.

The instrument was calibrated for a baseline enthalpy, and the experiment was performed with 25−30 injections of 9 μL of substrate per injection at 180-s intervals, with continuous stirring at 400 rpm. Blank titrations of peptide or SNase into buffer were also performed in order to correct for heats of dilution and mixing. The resulting data were analyzed to determine the binding stoichiometry (n), the association constant K_a, and the enthalpy change (ΔH). ΔG and ΔS values were calculated from the fundamental equation of thermodynamics: $\Delta G = -RT \ln K_a$ and $\Delta S = (\Delta H - \Delta G)/T$.

Structure Determination. Initial conditions for screening were adopted from previous studies of the benzamidine−wild-type trypsin complex (*17*). Crystals were obtained by the hanging drop vapor diffusion method from a solution consisting of 14 mg/mL trypsin, 1.2 M $MgSO_4$, 30 mM sodium cacodylate (pH 6.8), 10 mM $CaCl_2$, 100 mM benzamidine (final concentrations) at 20 °C. Crystallization occurred within 2 weeks, producing large cubic crystals approximately 300 Å per side. Single crystals were soaked in 25% (v/v) glycerol and frozen in liquid propane, and diffraction data were collected. The structure was solved by molecular replacement using 1DPO (Protein Data Bank).

RESULTS

Staphylococcal Nuclease as a Model Protease Substrate. Our interest in examining the hydrolysis of protein substrates by serine proteases was prompted by the observation that t-PA cleaved an amino acid sequence added to the N-terminus of ornithine decarboxylase with a K_m value 15000-fold lower than that for the analogous sequence tested as a peptide (*10*). Inouye and co-workers reported a similar finding with SPase I, a signal peptidase from *Escherichia coli*, that exhibited a 50−100-fold decrease in K_m for cleavage of a protein, pro OmpA-nuclease A, relative to that for an analogous peptide substrate (*18*). These observations suggested that the protein context of a substrate sequence could have a profound effect on substrate recognition as measured by K_m. To extend our understanding of this phenomenon, we chose to perform more detailed studies using staphylococcal nuclease (SNase) as a substrate.

SNase has many advantages as a model substrate for protease recognition. Anfinsen demonstrated that SNase has a single primary site for cleavage by trypsin and that, once separated, the two fragments could be recombined to form active SNase (*19*, *20*). Cleavage within the ω loop (residues 43−52) generates two products that can be visualized upon gel electrophoresis and analyzed to obtain kinetic constants k_{cat} and K_m (*9*, *10*). The existence of a single cleavage site is important because multiple cleavage sites would greatly complicate or prevent quantitation and interpretation of cleavage experiments. Mutations, insertions, or deletions within the ω loop do not abolish expression or catalytic activity (*21*, *22*), suggesting that SNase variants with altered loop sequences are largely intact and can serve as useful substrates for studies that compare cleavage of different sequences.

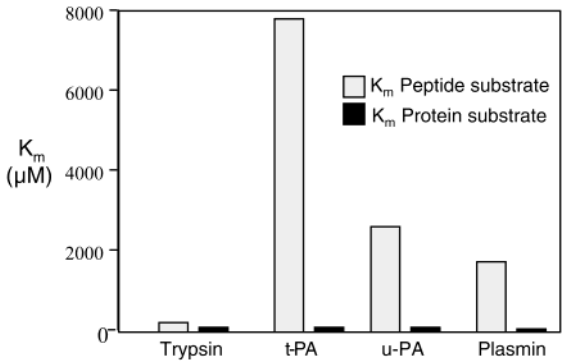

FIGURE 1: Average K_m values for cleavage of protein and peptide substrates by trypsin, t-PA, u-PA, and plasmin. Individual values are shown in Table 1. All peptides were 14 amino acids long; only the P4 to P2′ residues are shown. The P4′ to P2′ residues are the same in both protein and peptide substrates; the flanking residues differ.

In all figures and tables shown in this article, only the sequence immediately surrounding the scissile bond is shown. For peptide substrates, this sequence is within a 14-residue peptide; for protein substrates, it is within the ω loop of SNase.

Reduced K_m Values for Cleavage of Protein Substrates Relative to Peptides. We have examined the cleavage of SNase variants by u-PA, t-PA, plasmin, and trypsin and found that K_m values were reduced up to 15000-fold relative to those for peptide substrates containing analogous amino acids surrounding the scissile bond (Table 1, Figure 1). Decreased K_m values for cleavage of peptides were especially obvious in assays using t-PA, u-PA, and plasmin and were much less pronounced in assays that used trypsin.

For example, 34 of 36 peptides assayed with trypsin possessed K_m values of <1 mM, with an average value for all substrates of 0.18 mM (*9*−*11*, *23*−*25*). By contrast, 38 of 39 substrates were cleaved by t-PA with a K_m of >1 mM and an average value of 7.8 mM (*9*, *25*−*27*), 8 of 11 substrates were cleaved by u-PA with a K_m of >1 mM and an average value of 2.6 mM (*26*, *27*), and 8 of 13 substrates were cleaved by plasmin with a K_m of >1 mM and an average value of 1.7 mM (*11*).

Taken together, these 99 measurements of cleavage of peptide substrates by proteases demonstrate large differences for cleavage by selective proteases relative to cleavage by trypsin. The high K_m values for the cleavage of peptides by u-PA, t-PA, and plasmin are especially striking, because many of the peptides used in our assays were selected by substrate phage display (*28*) to be optimal sequences for cleavage by these enzymes. It is likely that a random selection of peptides would have yielded significantly higher K_m values.

In contrast to the high K_m values often observed for cleavage of peptides, the cleavage of protein substrates was characterized by K_m values <200 μM, regardless of the substrate or the protease used. Average K_m values were similar for trypsin (61 μM, $n = 5$ protein substrates), t-PA (34 μM, $n = 11$), u-PA (71 μM, $n = 6$), and plasmin (6.9 μM, $n = 3$). While SNase is not a physiologic substrate for t-PA, the K_m for cleavage by t-PA of its bona fide in vivo substrate, plasminogen, is 7.8 μM (*14*). This value, which was determined in the absence of the cofactor fibrin, is similar to the K_m values measured for cleavage of SNase.

The lower K_m values for the hydrolysis of protein substrates relative to peptide substrates suggest that there may exist protease−protein substrate contacts beyond the classic P5−P5′ binding loop subsites that result in higher binding affinities and that these interactions are especially necessary for efficient cleavage by t-PA, u-PA, and plasmin. However, before assuming that the widely disparate K_m values we observe reflect differences in substrate binding, it is necessary to demonstrate that these calculated K_m values approximate the corresponding, true dissociation constants. Serine proteases catalyze the reaction.

$$E + S \underset{k_{-1}}{\overset{k_1}{\rightarrow}} ES \xrightarrow{k_2} EP_2 + P_1 \xrightarrow{k_3} E + P_2$$

K_d is defined as k_{-1}/k_1 and K_m ($k_{-1}+k_2/k_1$) approximates K_d only when $k_{-1} \gg k_2$.

Trypsin and u-PA as Model Proteases for Characterizing the Binding of SNase Variants. Because K_m is not necessarily equivalent to K_d, we chose to experimentally determine K_d values for two different serine proteases, u-PA and trypsin. Trypsin was selected because it is a well-characterized example of a broadly specific protease. We chose u-PA for two reasons. The first was that kinetic studies had revealed a 160-fold reduction in K_m values with SNase substrates relative to analogous peptides, a large differential that seems likely to be biologically relevant. A second reason was that, in contrast to trypsin, u-PA is a highly specific protease. Comparison of data for these two enzymes, therefore, might provide new insights into the evolution of specificity by serine proteases and perhaps other enzymes that act on protein substrates.

Use of S195A Trypsin To Determine K_d Values. The measurement of K_d values of serine proteases for their substrates is complicated by the fact that proteases typically cleave their substrates so rapidly that techniques for directly measuring binding of wild-type enzymes cannot be used. To facilitate determination of K_d values for association of serine proteases with protein or peptide substrates, we obtained an inactive trypsin mutant (S195A) from Dr. Lizbeth Hedstrom (*13*).

The S195A trypsin variant is useful for these studies because its K_m values for cleavage of substrates containing *p*-nitroanilide leaving groups are similar to those of the wild-type enzyme (*29*). Significantly, k_{cat} values for hydrolysis of *p*-nitroanilide peptide substrates containing activated amide linkages are reduced by up to 41000-fold at pH 8.0, and cleavage of a peptide substrate that lacks activated amide linkages, the insulin β-chain, was not detected (*29*). On the basis of these data, we hypothesized that S195A proteases would mimic the substrate binding of the wild-type enzyme but would not degrade substrates during prolonged binding studies. S195A trypsin and thrombin have proven valuable for other biophysical studies, including investigations of the inhibitory mechanisms of serpins (*30*−*38*) and investigations of molecular recognition by thrombin (*39*−*41*).

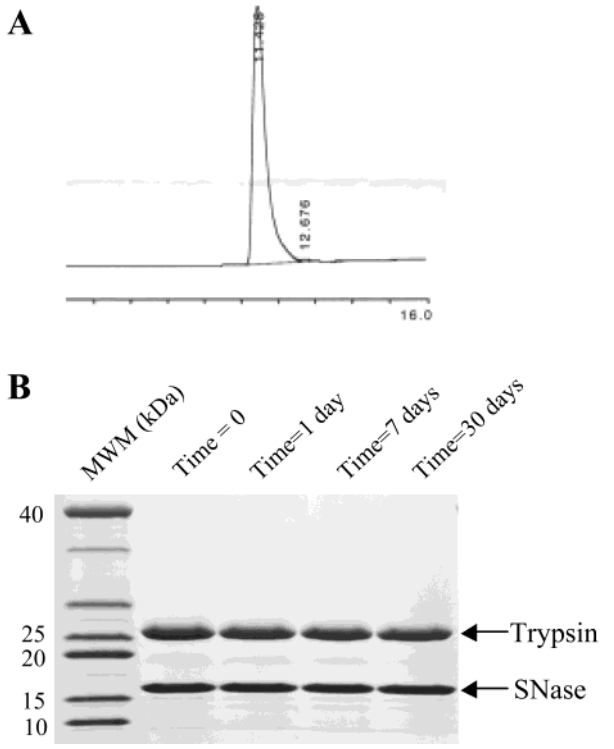

FIGURE 2: Stability of peptide and SNase variants to hydrolysis by S195A trypsin during extended incubations. (A) HPLC analysis of peptide LGGSGR↓SANAILE after 30-day incubation with trypsin S195A [500 μM trypsin, 500 μM peptide (pH 7.5), 30 days at 37 °C]. (B) SDS−PAGE analysis of SNase loop variant PGSGR↓SAG [500 μM trypsin, 500 μM SNase (pH 7.5), at 37 °C].

Table 2: Crystallographic Data for Rat Trypsin S195 Complexed with Benzamidine

space group	*I*23
cell dimensions	$a = 122.5$
resolution	2.0 Å
R_{free} value	0.31
R value	0.27
% complete	99

Binding assays require high enzyme concentrations and long incubations relative to the concentrations used in standard enzyme assays. To confirm that the S195A trypsin does not catalyze significant substrate turnover over the course of the various biochemical and biophysical assays needed to determine K_d values, we performed time-course incubations of peptide and protein substrates with trypsin. Evidence for substrate stability was especially important, because Tulinsky and colleagues had shown that S195A thrombin did possess very low but detectable residual activity toward peptides (*40*). Analysis of substrate peptides or SNase incubated with S195A trypsin did not reveal cleavage products (Figure 2A,B).

Structure of S195A Trypsin. To support the hypothesis that S195A trypsin could provide useful insights into substrate binding, we solved its structure in complex with the inhibitor benzamidine (Table 2). We observed that the structure of the S195A trypsin−benzamidine complex is nearly identical

Table 3: Determination of K_d for Binding by Trypsin Using Analytical Ultracentrifugation

substrate[a]	K_m (μM)	K_d (μM)
HYGR↓SG		
peptide	210 ± 31	200 ± 70
protein	57 ± 11	30 ± 8.8
YGAK↓AY		
peptide	150 ± 18	300 ± 56
protein	95 ± 21	100 ± 39
GSGR↓SA		
peptide	140 ± 18	300 ± 56
protein	18 ± 4.3	30 ± 12
PFGR↓SA		
peptide	260 ± 16	200 ± 61
protein	39 ± 8.7	40 ± 18

[a] All peptides were 14 amino acids long; only the P4 to P2′ residues are shown. The P4′ to P2′ residues are the same in both protein and peptide substrates; the flanking residues differ.

to that of the wild-type enzyme−inhibitor complex. The root-mean-square deviation of all backbone atoms of the two structures was calculated to be 0.47 Å, and a superimposition of the two structures showed no significant structural differences.

Importantly for specificity and catalysis, the positions of the remaining catalytic triad residues, the Asp 189 residue at the bottom of the specificity pocket, and the substrate binding loops exhibit little change in comparison to the wild-type structure. The structure of S195A has also been solved bound to the serine protease inhibitor *Manduca sexta* serpin K, and the structure of this complex was also observed to be similar to that of wild-type trypsin (*30*). Of course, while the structures appear similar, it is important to note that even small differences may have substantial effects on substrate binding.

Measurement of K_d Values by Analytical Ultracentrifugation. To test directly the similarity of measured K_m and K_d values, we used analytical ultracentrifugation to measure the binding of the S195A trypsin with peptides and SNase variants. In all but one case, the calculated K_d value for binding to protein substrates is smaller than the calculated K_d value for binding to peptide substrates, mirroring the data for K_m values (Table 3). The one exception is the peptide YYGAKAYRPDKM, which had been selected for its ability to bind tightly to trypsin (*23*, *42*). This sequence had similar K_d values in the peptide and protein contexts, a result that was also obtained for K_m determinations.

Measurement of K_d Values by Isothermal Titration Calorimetry. We used isothermal calorimetry (ITC) to confirm K_d values and to determine thermodynamic parameters ($\Delta G°$, $\Delta S°$, and $\Delta H°$) for enzyme−substrate binding. These studies were performed with S195A trypsin and S195A u-PA and both peptide and SNase substrates. Titrations of the trypsin and u-PA with the peptides and SNase variants fit a simple model with close to a 1:1 binding stoichiometry in all cases. All determinations were done at least three times, and the reported values are averages.

ITC measurements yielded K_d values that were similar to the K_m values reported above (Table 4). S195A trypsin and S195A u-PA bound protein substrates similarly, with K_d values ranging from 2 to 170 μM for trypsin and from 22 to 100 μM for u-PA. As with K_m values, the measured K_d values for the binding of peptides by S195A trypsin varied greatly from the analogous values with S195A u-PA, from 84 to 320 μM for trypsin to from 310 to 2600 for u-PA.

The calculated ΔH values for peptides and SNase variants were negative for all substrates (Table 4, Figure 3). In every case, the enthalpic contribution to binding of u-PA or trypsin to SNase variants is greater than the corresponding contribution for analogous peptides. This greater enthalpic contribution to protein relative to peptide binding supports the conclusion that extended substrate−protease contacts beyond

Table 4: Binding of Substrates by Trypsin and u-PA Measured Using Isothermal Titration Calorimetry

substrate[a]	K_m (μM)	K_d (μM)	N[b]	$\Delta H°$ (kcal/mol)	$\Delta S°$ (cal mol^{-1} K^{-1})	$\Delta G°$ (kcal/mol)
			Trypsin			
HYGR↓SG						
peptide	210 ± 31	270	1.1	−4.0 ± 0.16	3.3	−5.1
protein	57 ± 11	79	0.92	−6.1 ± 0.09	−0.9	−5.8
YGAK↓AY						
peptide	150 ± 18	110	0.99	−4.9 ± 0.21	2.1	−5.6
protein	95 ± 21	170	0.94	−7.6 ± 0.18	−7.2	−5.4
GSGR↓SA						
peptide	140 ± 18	84	1.2	−5.8 ± 0.04	−0.2	−5.8
protein	18 ± 4.3	2.0	0.90	−6.4 ± 0.10	5.5	−8.1
PFGR↓SA						
peptide	260 ± 16	320	0.84	−4.2 ± 0.07	2.1	−4.9
protein	39 ± 9	61	1.3	−5.9 ± 0.07	0.3	−6.0
			u-PA			
HYGR↓SG						
peptide	3800 ± 330	2600	1.1	−2.9 ± 0.14	2.6	−3.7
protein	30 ± 9	39	0.89	−7.1 ± 0.18	-3.5	−6.3
YGAK↓AY						
peptide	2300 ± 270	1700	1.2	−2.5 ± 0.03	4.1	−3.9
protein	≥130	100	0.96	−6.3 ± 0.22	-2.6	−5.6
GSGR↓SA						
peptide	600 ± 72	310	0.95	−3.8 ± 0.15	3.9	−5.0
protein	53 ± 21	29	1.0	−5.5 ± 0.08	2.9	−6.4
PFGR↓SA						
peptide	2200 ± 90	1800	1.0	−3.3 ± 0.17	1.9	−3.9
protein	14 ± 3	22	1.1	−7.2 ± 0.11	-1.9	−6.6

[a] All peptides were 14 amino acids long; only the P4 to P2′ residues are shown. The P4′ to P2′ residues are the same in both protein and peptide substrates; the flanking residues differ. [b] Calculated number of substrates bound to each protease molecule.

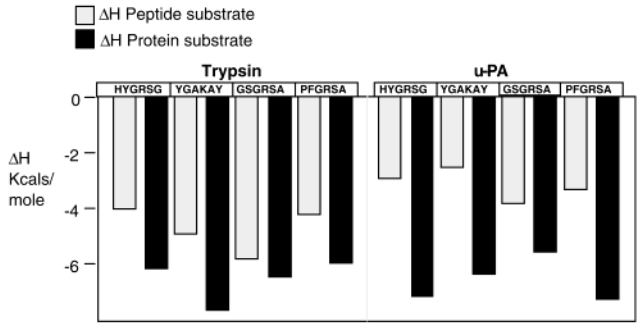

FIGURE 3: Δ*H* values for binding of protein or peptide substrates by trypsin and u-PA, determined by ITC. Numerical values are shown in Table 4. All peptides were 14 amino acids long; only the P4 to P2′ residues are shown. The P4′ to P2′ residues are the same in both protein and peptide substrates; the flanking residues differ.

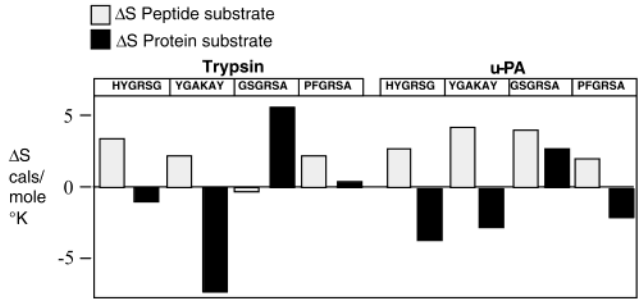

FIGURE 4: Δ*S* values for binding of protein or peptide substrates by trypsin and u-PA, determined by ITC. Numerical values are shown in Table 4. All peptides were 14 amino acids long; only the P4 to P2′ residues are shown. The P4′ to P2′ residues are the same in both protein and peptide substrates; the flanking residues differ.

the classic subsites contribute to binding and may explain our observation of higher affinity binding for protein relative to peptide substrates. For S195A u-PA, the average difference in Δ*H* values for protein and peptide substrates was 3.4 kcal/mol, nearly twice that of S195A trypsin, 1.75 kcal/mol, suggesting that enthalpic differences account for the larger difference in affinity of u-PA for protein relative to peptide substrates.

For the binding of peptides by u-PA and trypsin, Δ*S* values were positive in seven cases and neutral in one, ranging from −0.2 to 4.1 cal mol^{-1} K^{-1} (Table 4, Figure 4). The mostly positive entropic contributions favor peptide binding. This entropic contribution to peptide binding may appear counterintuitive because peptides are generally considered to be disordered in solution and become more ordered upon binding. However, these measurements reflect the entropy of the entire system, rather than that of the peptide alone, and other investigators have reported similar observations (*43*−*45*).

In contrast to the positive Δ*S* values for binding of peptides, the Δ*S* values for six of eight protein substrates were close to neutral or negative (from 0.3 to −7.2 cal mol^{-1} K^{-1}; Table 4, Figure 4), suggesting that an entropic penalty was paid for binding protein substrates. One sequence, GSGRSA, has a positive Δ*S* for protein binding by both u-PA and trypsin, 2.9 and 5.5 cal mol^{-1} K^{-1}, respectively. GSGRSA was selected by using substrate phage display to be an optimal sequence for cleavage by u-PA and is also a good substrate for trypsin (*24*, *25*), and this selection for productive binding may explain the favorable entropy of binding.

DISCUSSION

Association of Proteases with Proteins and Peptides. Thornton and co-workers have examined the three-dimensional structures of nine known sites for cleavage by trypsin in unrelated proteins, including SNase (*46*−*48*). In each of these proteins, the structure at the cleavage site is quite different from the crystallographically determined conformation of the reactive center loops of inhibitory proteins. For example, for SNase to form interactions similar to those observed in structures of trypsin in complex with bovine pancreatic trypsin inhibitor (BPTI), the crucial lysine P1 residue within the ω loop must be inverted nearly 180° to adopt an orientation similar to the that of P1 lysine of the BPTI (*48*).

Given that the ω loop of SNase must unfold to enter the protease active site, it might have been expected that the folded protein would bind to proteases more poorly than peptides containing the analogous sequence surrounding the scissile bond. In fact, we observe the opposite result, suggesting that interactions between SNase and trypsin distant from the scissile bond stabilize formation of the protein−protease complex. Interestingly, Rudolph and co-workers have observed low K_m values for protein versus peptide substrates for phosphatase Cdc25 and have also concluded that important interaction sites must exist beyond those surrounding the phosphate binding site (*12*).

One alternative explanation for the lower K_d and K_m values for SNase is the possibility that the substrate sequence is more constrained when it is in the context of a protein than when it is contained within a peptide, reducing the entropic penalty paid upon binding because far fewer conformational states need to be sampled. We cannot rule out this hypothesis, but three lines of evidence argue against it. The first is that we have observed that constraining a substrate peptide sequence with an intramolecular disulfide bond does not significantly reduce K_m for cleavage by t-PA, a protease that is closely related to u-PA (*14*). However, this disulfide loop was relatively large, nine amino acids, so this result cannot be considered conclusive. The second is our initial observation of reduced K_m with ornithine decarboxylase modified with a substrate sequence at the N-terminus, a region that should be highly flexible and that is not part of the protein's normal folded structure (*10*). Finally, ITC measurements indicate that reduced entropy is a relatively minor and inconsistent contributor to higher affinity binding of the proteins that we have examined (Table 4, Figure 4).

A second alternative explanation for reduced K_d values is that the proteases bind nonproductively to other sequences within SNase. SNase has many positively charged arginine and lysine residues on its surface that could fit into the S1 pocket of the protease. However, it is not likely that such interactions play a major role, since nonproductive binding would be expected to lead to competitive inhibition and apparent K_m values that are higher than the observed K_d values.

Previously Characterized Examples of Proteases That Use Secondary Sites for Binding Protein Substrates. It is important to note that it has been clearly established that specific proteases can evolve productive, secondary sites of interaction with physiological substrates. For example, thrombin possesses an exosite that has been characterized at atomic

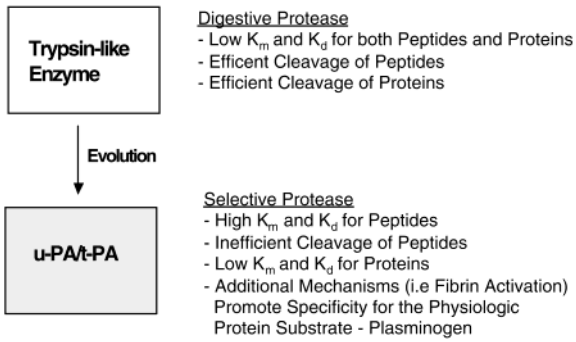

FIGURE 5: Possible scheme for the evolution of a nonselective trypsin-like enzyme into u-PA and t-PA. Mutations in the trypsin-like enzyme reduce catalysis of peptides and improve binding to proteins. Further mutations then lead to specificity for a physiologic protein substrate, which in the case of u-PA is plasminogen, relative to other proteins that might be encountered in vivo.

resolution and significantly increases the affinity and specificity of the enzyme for fibrinogen (*49*, *50*). In addition, mutagenesis experiments have shown that the binding affinity and specificity of prothrombinase for thrombin are also modulated by interactions distinct from those at the site of amide bond cleavage (*51*). Similarly, secondary binding sites have been shown to play key roles in the recognition of both the physiological substrate plasminogen and the physiological inhibitor PAI-1 by t-PA (*52*).

These important secondary binding sites, however, differ from the potential exosites we postulate for binding SNase because they have evolved to enhance enzyme affinity and specificity for a specific, physiological substrate such as plasminogen or thrombin. In contrast, the binding of a nonphysiological substrate such as SNase to u-PA or t-PA has never been subjected to selective pressure to enhance the protein−protein interaction. Therefore, if an exosite is responsible for enhanced binding of SNase variants relative to peptides, it would probably increase the binding of many different proteins.

Enhanced Binding and the Evolution of Enzyme Specificity for Protein Substrates. The physiological demands on trypsin and u-PA are quite different. u-PA is a highly selective protease, with one primary physiologic substrate, plasminogen, while the physiologic role of trypsin is the complete digestion of polypeptides. The reduced binding of peptides by u-PA relative to trypsin may be important for restricting the substrate range of u-PA by minimizing the cleavage of peptide hormones or other biologically important peptides (Figure 5). Trypsin, a broadly specific digestive protease, must cleave many different peptide sequences and therefore retains the ability to bind peptides with low K_m and K_d values.

Conversely, it is also possible to view our findings in terms of enhanced binding by specific serine proteases, such as u-PA, of proteins relative to peptides and to question why this feature has evolved. We have observed that k_{cat} values for cleavage of proteins versus peptides by t-PA, u-PA, and trypsin are reduced on average by 280-, 140-, and 24-fold, respectively (*9*−*11*, *23*−*27*). Catalytic efficiency values (k_{cat}/K_m) for t-PA, u-PA, and trypsin, however, are reduced on average by only 3.3-, 1.2-, and 1.5-fold, respectively (*9*−*11*, *23*−*27*). Therefore, enhanced protein binding may be one of the mechanisms used by u-PA and t-PA to compensate for reduced turnover rates and overcome the obstacle to successful catalysis posed by the native structure of the region surrounding the scissile bond. Other mechanisms, such as activation by the cofactor fibrin in the case of t-PA, are used to further enhance activity toward plasminogen, the physiological substrate for u-PA and t-PA, and to promote the high selectivity seen in vivo.

CONCLUSIONS

These studies demonstrate that K_m values for the cleavage of protein and peptide substrates by u-PA and trypsin are similar to the corresponding K_d values for substrate binding. K_d values for binding of protein substrates are lower than those for binding of peptide substrates, and this difference is much greater for the selective protease u-PA than for the nonselective protease trypsin. Enhanced binding to proteins is enthalpically driven, while entropy values can be positive or negative, depending on the particular substrate sequence.

These observations suggest that more extensive protease−substrate contacts occur with protein substrates in comparison with analogous peptide substrates. These contacts would have to be outside the P5−P5′ region typically thought to be responsible for substrate recognition and may represent an important factor contributing to substrate discrimination by serine proteases. Similar mechanisms for the evolution of phosphatase specificity have been proposed (*12*), and it may be useful to perform similar investigations with other enzymes that modify protein substrates.

ACKNOWLEDGMENT

The authors thank Dr. Lizbeth Hedstrom (Brandeis University) for generously providing the vectors and experimental protocols needed to obtain S195A trypsin and for advice on protein expression. We thank Brian Cannon for helping prepare the protein used in these studies, Dr. Joseph Albanesi for assistance with analysis of the ultracentrifugation data, Dr. Margaret Phillips for advice on kinetic analysis, and Dr. Celestine Thomas for assistance with ITC.

REFERENCES

1. Perona, J. J., and Craik, C. S. (1997) *J. Biol. Chem. 272*, 29987−29990.
2. Hedstrom, L. (2002) *Chem. Rev. 102*, 4501−4523.
3. Laskowski, M., and Kato, I. (1980) *Annu. Rev. Biochem. 49*, 593−626.
4. Lu, S. M., Lu, W., Qasim, M. A., Anderson, S., Apostol, I., Ardelt, W., Bigler, T., Chiang, Y. W., Cook, J., James, M. N., Kato, I., Kelly, C., Kohr, W., Komiyama, T., Lin, T. Y., Ogawa, M., Otlewski, J., Park, S. J., Qasim, S., Ranjbar, M., Tashiro, M., Warne, N., Whatley, H., Wieczorek, A., Wieczorek, M., Wilusz, T., Wynn, R., Zhang, W., and Laskowski, M. (2001) *Proc. Natl. Acad. Sci. U.S.A. 98*, 410−415.
5. Bode, W., and Huber, R. (1992) *Eur. J. Biochem. 204*, 433−451.
6. Gettins, P. G. W. (2002) *Chem. Rev. 102*, 4751−4804.
7. Laskowski, M., and Qasim M. A. (2000) *Biochim Biophys Acta 1477*, 324−337.
8. Coombs, G. S., Rao, M. S., Olson, A. J., Dawson, P. E., and Madison, E. L. (1999) *J. Biol. Chem. 274*, 24074−24079.
9. Coombs, G. S., Bergstrom, R. C., Madison, E. L., and Corey, D. R. (1998) *J. Biol. Chem. 273*, 4323−4328.
10. Coombs, G. S., Dang, A. T., Madison, E. L., and Corey, D. R. (1996) *J. Biol. Chem. 271*, 4461−4467.
11. Hervio, L., Bergstrom, R. C., Coombs, G. C., Trivedi, K., Corey, D. R., and Madison, E. L. (2000) *Chem. Biol. 7*, 443−453.
12. Rudolph, J., Epstein, D. M., Parker, L., and Eckstein, J. (2001) *Anal. Biochem. 289*, 43−51.
13. Hedstrom, L., Szilagyi, L., and Rutter, W. J. (1992) *Science 255*, 1249−1253.

14. Madison, E. L., Coombs, G. S., and Corey, D. R. (1995) *J. Biol. Chem. 270*, 7558−7562.
15. Yphantis, D. A. (1964) *Biochemistry 3*, 297−317.
16. Johnson, M. L., Correia, J. J., Yphantis, D. A., and Halvorson, H. R. (1981) *Biophys. J.* 36, 575−588.
17. Wilke, M. E., Higaki, J. N., Craik, C. S., and Fletterick, R. J. (1991) *J. Mol. Biol. 219*, 511−523.
18. Chatterjee, S., Suciu, D., Dalbey, R. E., Kahn, P. C., and Inouye, M. (1995) *J. Mol. Biol. 245*, 311−314.
19. Taniuchi, H., Anfinsen, C. B., and Sodja, A. (1967) *Proc. Natl. Acad. Sci. U.S.A. 58*, 1235−1242.
20. Taniuchi, H., and Anfinsen, C. B. (1968) *J. Biol. Chem. 243*, 4778−4786.
21. Baldisseri, D. M., and Torchia, D. A. (1991) *Biochemistry 30*, 3628−3633.
22. Poole, L. B., Loveys, D. A., Hale, S. P., Gerlt, J. A., Stanczyk, S. M., and Bolton, P. H. (1991) *Biochemistry 30*, 3621−3627.
23. Coombs, G. S., Hazzard, J., and Corey, D. R. (1995) *Bioorg. Med. Chem. Lett. 5*, 611−614.
24. Corey, D. R., Willett W. S., Coombs, G. C., and Craik, C. S. (1995) *Biochemistry 34*, 11521−11527.
25. Ding, L., Coombs, G. S., Strandberg, L., Navre, M., Corey, D. R., and Madison, E. L. (1995) *Proc. Natl. Acad. Sci. U.S.A. 92*, 7627−7631.
26. Ke, S.-H., Coombs, G. S., Tachias, K., Navre, M., Corey, D. R., and Madison, E. L. (1997) *J. Biol. Chem. 272*, 16603−16609.
27. Ke, S.-H., Coombs, G. S., Tachias, K., Corey, D. R., and Madison, E. L. (1997) *J. Biol. Chem. 272*, 20456−20462.
28. Matthews, D. J., and Wells, J. A. (1993) *Science 260*, 1113−1117.
29. Corey, D. R., and Craik, C. S. (1992) *J. Am. Chem. Soc. 114*, 1784−1790.
30. Ye, S., Cech, A., Belmares, R., Bergstrom, R. C., Corey, D. R., Kanost, M. R., and Goldsmith, E. J. (2001) *Nat. Struct. Biol. 8*, 979−983.
31. Peterson, F. C., Gordon, N. C., and Gettins, P. G. W. (2001) *Biochemistry 40*, 6275−6283.
32. Ludeman, J. P., Whisstock, J. C., Hopkins, P. C. R., Le Bonniec, B. F., and Bottomley, S. P. *Biophys. J. 80*, 491−497.
33. Dekker, R. J., Eichinger, A., Stoop, A. A., Bode, W., Pannekoek, H., and Horrevoets, A. J. G. (1999) *J. Mol. Biol. 293*, 613−627.
34. Picard, V., Marque, P. E., Paolucci, F., Aiach, M., and Le Bonniec B. F. (1999) *J. Biol. Chem. 274*, 4586−4593.
35. Stone, S. R., and Lebonniec, B. F. (1997) *J. Mol. Biol. 265*, 344−362.
36. Olson, S. T., Swanson, R., Day, D., Verhamme, I., Kvassman, J., and Shore, J. D. (2001) *Biochemistry 40*, 11472−11756.
37. Mellet, P., Mely, Y., Hedstrom, L., Cahoon, M., Belorgey, D., Rubin, H., and Bieth, J. G. (2002) *J. Biol. Chem. 277*, 38901−38914.
38. Rezaie, A. R. (1999) *Biochemistry 38*, 14592−14599.
39. Maurer, M. C., Trosset, J. Y., Lester, C. C., DiBella, E. E., and Scheraga, H. A. (1999) *Proteins 34*, 29−48.
40. Krishnan, R., Sadler, J. E., and Tulinsky, A. (2000) *Acta Crystallogr. D56*, 406−410.
41. Baglin, T. P., Carrell, R. W., Church, F. C., Esmon, C. T., and Huntington, J. A. (2002) *Proc. Natl. Acad. Sci. U.S.A. 99*, 11079−11084.
42. Eichler, J., and Houghton, R. A. (1993) *Biochemistry 32*, 11035.
43. Parker, M. H., Lunney E. A., Ortwine D. F., Pavlovsky A. G., Humblet C., and Brouillette C. G. (1999) *Biochemistry 38*, 13592−13601.
44. Velazquez-Campoy, A., Luque I., Todd M. J., Milutinovich M., Kiso Y., and Freire E. (2000) *Protein Sci. 9*, 1801−1809.
45. Velazquez-Campoy, A., Todd, M. J., and Freire, E. (2000) *Biochemistry 39*, 2201−2207.
46. Hubbard, S. J., Campbell, S. F., and Thornton J. M. (1991) *J. Mol. Biol. 220*, 507−530.
47. Hubbard, S. J., Eisenmenger, F., and Thornton, J. M. (1994) *Protein Sci. 3*, 757−768.
48. Hubbard, S. J., Beynon, R. J., and Thornton, J. M. (1998) *Protein Eng. 11*, 349−359.
49. Di Cera, E. (1998) *Chem. Rev. 98*, 1563−1591.
50. Rose, T., and Di Cera, E. (2002) *J. Biol. Chem. 277*, 18875−18880.
51. Orcutt, S. J., Concetta, P., and Krishnaswamy S. (2002) *J. Biol. Chem. 278*, 46191−46196.
52. Madison, E. L., Goldsmith, E. J., Gerard, R. D., Gething, M. J., and Sambrook, J. F. (1989) *Nature 339*, 721−724.

BI027417X