Substrate specificity of prostate-specific antigen (PSA)

Gary S Coombs[1], Robert C Bergstrom[2], Jean-Luc Pellequer[3], Scott I Baker[2], Marc Navre[4], Matthew M Smith[4], John A Tainer[3], Edwin L Madison[1] and David R Corey[2]

Background: The serine protease prostate-specific antigen (PSA) is a useful clinical marker for prostatic malignancy. PSA is a member of the kallikrein subgroup of the (chymo)trypsin serine protease family, but differs from the prototypical member of this subgroup, tissue kallikrein, in possessing a specificity more similar to that of chymotrypsin than trypsin. We report the use of two strategies, substrate phage display and iterative optimization of natural cleavage sites, to identify labile sequences for PSA cleavage.

Results: Iterative optimization and substrate phage display converged on the amino-acid sequence SS(Y/F)Y↓S(G/S) as preferred subsite occupancy for PSA. These sequences were cleaved by PSA with catalytic efficiencies as high as 2200–3100 M^{-1} s^{-1}, compared with values of 2–46 M^{-1} s^{-1} for peptides containing likely physiological target sequences of PSA from the protein semenogelin. Substrate residues that bind to secondary (non-S1) subsites have a critical role in defining labile substrates and can even cause otherwise disfavored amino acids to bind in the primary specificity (S1) pocket.

Conclusions: The importance of secondary subsites in defining both the specificity and efficiency of cleavage suggests that substrate recognition by PSA is mediated by an extended binding site. Elucidation of preferred subsite occupancy allowed refinement of the structural model of PSA and should facilitate the development of more sensitive activity-based assays and the design of potent inhibitors.

Addresses: [1]Corvas International, Department of Molecular Biology, 3030 Science Park Road, San Diego, CA 92121, USA. [2]Howard Hughes Medical Institute, Departments of Pharmacology and Biochemistry, University of Texas Southwestern Medical Center at Dallas, 5323 Harry Hines Boulevard, Dallas, TX 75235, USA. [3]The Scripps Research Institute, Department of Molecular Biology, La Jolla, CA 92037, USA. [4]Affymax Research Institute, Santa Clara, CA 95501, USA.

Correspondence: Edwin L Madison and David R Corey

E-mail: ed_madison@corvas.com
Corey@howie.swmed.edu

Key words: prostate-specific antigen, protease specificity, subsite occupancy, substrate phage display

Received: **18 June 1998**
Accepted: **15 July 1998**

Published: **24 August 1998**

Chemistry & Biology September 1998, **5**:475–488
http://biomednet.com/elecref/1074552100500475

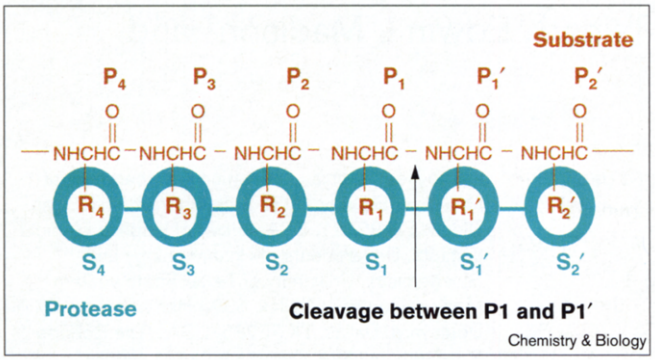
Introduction

Prostate specific antigen (PSA) is a (chymo)trypsin family serine protease that was first found in human semen [1,2]. PSA can also be detected in serum, where elevated levels have been correlated with benign and metastatic disease, making PSA a useful marker for the growth of prostatic tumors [3–5]. Clinical determination of PSA levels currently relies exclusively on immunoassays that do not discriminate between active PSA, PSA zymogen, or inactive complexes of PSA with circulating serpins such as α2-antichymotrypsin. Identification of easily detectable substrates for PSA would allow localization and quantitation of active forms of this enzyme, information that is likely to provide new insights into the biological and pathological roles of PSA.

Proposed native substrates for PSA include semenogelin [6–8], the predominant protein component of the coagulum that forms upon ejaculation and must be digested to permit sperm motility. PSA has also been observed to cleave insulin growth-factor-binding protein-3 [9,10], fibronectin and laminin [11], as well as parathyroid-hormone-related protein [12]. Cleavage of these substrates suggests that PSA could have a direct role in cancer cell proliferation and invasion. Selective inhibition of PSA, therefore, might have important physiological consequences, and development of efficient inhibitors would allow the hypothesis that improper PSA expression stimulates the growth of prostate tumors to be tested.

PSA shares approximately 78% identity with human glandular kallikrein (hK2) and 62% identity with human pancreatic kallikrein (hK1) [2,13–15]. Catalysis by PSA has been characterized qualitatively through the hydrolysis of peptides and complete digestion of semenogelin revealing that, unlike hK2 and tissue kallikrein, which possess trypsin-like specificity for arginine and lysine, PSA possesses a chymotrypsin-like preference for tyrosine and leucine [6,7,16,17]. More recently, Isaacs and coworkers [18] have synthesized fluorogenic substrates containing 7-amino-4-methyl-coumarin leaving groups that are cleaved with k_{cat}/K_m values of up to 270 M^{-1} s^{-1}. These investigators have also characterized less efficient substrates that display selectivity towards PSA relative to other proteases, and have demonstrated that PSA possesses a surprising ability to use glutamine as a P1 residue (Figure 1) during cleavage of activated substrates [18].

Figure 1



Nomenclature for the interaction of substrate amino acids with protease subsites [56].

In this study we define the preferred subsite occupancy for peptide cleavage by PSA. Using two independent methods, substrate phage display [19,20] and iterative optimization of native substrate sequences [21,22], we have identified and kinetically characterized reactive peptide substrates for PSA containing the P4 to P2′ residues SS(F/Y)Y↓S(G/S) (using single-letter amino acid code; the arrow denotes the position of peptide-bond hydrolysis). Definition of this specificity is a first step towards the design of readily detectable substrates and potent inhibitors, and offers insights into substrate discrimination by PSA. Similar to urokinase plasminogen activator (u-PA) [23] and tissue-type plasminogen activator (t-PA) [24], PSA appears to have a narrow specificity for optimal occupancy of non-S1 subsites that is in contrast to the broad specificity exhibited by related enzymes such as trypsin or chymotrypsin. Knowledge of preferred subsite occupancy for peptide substrates encouraged refinement of the three-dimensional model of PSA [25], primarily by remodeling the conformation of residues in PSA corresponding to the kallikrein loop.

Results and discussion

Kinetics of hydrolysis of peptides modeled after native cleavage sites and interchange of P1 residues

We initiated our studies by examining PSA-mediated cleavage of peptides **I** and **II** (Table 1) derived from cleavage sites within semenogelin identified by Lilja *et al.* [17]. For these studies we employed purified PSA from semen. Davie and coworkers [26] have noted that such preparations can contain proteolytic contaminants that might affect kinetic analysis, and kallikrein hK2, a protease specific for P1 arginine has been directly implicated as a contaminant in some PSA preparations [27].

Although the presence of contaminating enzyme activity can never be completely ruled out we do not believe that this affected our assays because, with the exception of one

Table 1

Iterative optimization of peptide hydrolysis. PSA-mediated hydrolysis of peptides derived from labile sequences within semenogelin.

	Substrate *(Pn,...P3, P2, P1,↓P1′, P2′, P3′,...Pn′)	k_{cat} (s^{-1})	K_m (μM)	k_{cat}/K_m ($M^{-1}\,s^{-1}$)
Sequences from semenogelin				
I	GSQQLL↓HNKQEGRD	0.012	5900	2.0
II	GISSQY↓SNTEERLW	1.30	28000	46
Exchange of P1 residues between I *and* II				
III	GSQQLY↓HNKQEGRD	0.037	14000	2.6
IV	GISSQL↓SNTEERLW	0.0035	2300	1.5
Substitution of P2 residues within II				
V	GISSGY↓SNTEERLW	0.13	1900	70
VI	GISSAY↓SNTEERLW	0.30	290	1000
VII	GISSPY↓SNTEERLW	1.1	10000	110
VIII	GISSFY↓SNTEERLW	2.3	1300	1800
Substitutions of P1′ and P2′ in VIII				
IX	GISSFY↓GNTEERLW	0.75	490	1500
X	GISSFY↓ANTEERLW	1.80	180	1000
XI	GISSFY↓FNTEERLW	ND		
XII	GISSFY↓QNTEERLW	ND		
XIII	GISSFY↓SGTEERLW	2.1	1100	1900
XIV	GISSFY↓SATEERLW	1.7	1100	1500
XV	GISSFY↓SSTEERLW	1.7	790	2200

*Positional nomenclature of subsite residues. Arrows denote the position of peptide-bond hydrolysis. The peptide bond is cleaved between P1 and P1′. ND, not determined due to insolubility of substrate peptides. Calculated error values were between 6 and 26%.

series of peptide substrates (noted below), cleavage was only observed adjacent to P1 leucine and tyrosine residues, a preference seen during extensive digestion of semenogelin [7]. In addition, Davie and coworkers [26] noted that recombinant PSA cannot readily cleave the Leu–Cys(SO_3) bond in insulin β chain, a finding that contrasts with an earlier report [16] that PSA derived from seminal plasma could rapidly cleave this bond. We found that the PSA used in our studies behaved similarly to recombinant PSA and did not cleave the Leu–Cys(SO_3) bond (data not shown).

We observed that peptide **I**, in which leucine was the P1 residue, was cleaved with a 23-fold lower catalytic efficiency, k_{cat}/K_m, than peptide **II** (Table 1), in which tyrosine was the P1 residue. We then exchanged tyrosine and leucine residues between **I** and **II** to yield peptides **III** and **IV** to test the relative importance of tyrosine versus leucine occupancy at the P1 position.

Introduction of tyrosine did not substantially alter catalysis of **III** relative to **I**, whereas introduction of leucine reduced catalysis of **IV** relative to **II** by approximately 30-fold. The 18-fold difference in catalysis between tyrosine-containing peptides **II** and **III** demonstrated that residues other than P1 can influence the rate of hydrolysis and led

us to search for more labile substrates by systematically examining the effect of subsite occupancy on catalysis.

Iterative substrate optimization

To identify more labile substrates for PSA and to begin to develop rules for preferred subsite occupancy we modified peptide **II**, the more reactive of the initial physiologically relevant substrates, and kinetically assayed the resultant peptides (Table 1). The P1 tyrosine was held constant, but residues were substituted at the P2, P1′, and P2′ positions. Substitution of the glutamine at P2 in peptide **II** by glycine, alanine, proline or phenylalanine improved catalysis by 1.5, 22, 2.4, and 39-fold, respectively, supporting an important role of P2/S2 interactions in promoting efficient catalysis. Because introduction of phenylalanine at P2 yielded the most labile substrate, it was held constant during subsequent modifications at P1′ and P2′. Our ability to measure the effect of substitutions at P1′ was limited by the insolubility of variant peptides **XI** and **XII**. Substitution of alanine or glycine in place of serine at P1′, and substitution of glycine, alanine or serine for asparagine at P2′ altered catalysis by less than twofold, with SSFY↓SS being the preferred P4–P2′ sequence.

Identification of potentially labile sequences using substrate phage display

Iterative optimization is a useful approach for identifying peptides with improved reactivity relative to peptides derived from native cleavage sequences. It is not feasible to test all combinations of residues, however, because the number of peptide substrates that can be synthesized and assayed kinetically is limited. To further delineate amino-acid preferences for catalysis by PSA we therefore used substrate phage display [19,20] to identify labile sequences through library screening. To avoid biasing results, we employed iterative substrate optimization and substrate phage display independently and did not compare data regarding labile sequences prior to choosing candidate peptides for assay.

For phage display, a polyvalent fd phage library that displayed random octapeptide sequences was prepared. Each member of this library displayed an amino-terminal extension from phage coat protein III that contained a four-residue flexible linker region (GGAG), a randomized region of eight amino acids, a six amino-acid linker sequence (GGAGSS), and the epitopes for mAbs 179 and 3-E7. The mAb epitopes serve as an affinity tag for removal of undigested phage and as a marker in a dot blot assay for monitoring phage reactivity (see below). Incubation of phage with PSA did not affect infectivity of the phage, nor did PSA appear to cleave phage protein III, the antibody epitopes or the flexible linker sequences. As a result, cleavage of the randomized eight amino-acid insert by PSA was required to remove mAb 179 and 3E-7 epitopes from an individual phage particle. Incubation of the library with PSA, followed by removal of phage retaining the antibody epitopes, therefore, accomplished a large enrichment of phage clones whose random octamer sequence was efficiently cleaved by PSA.

Analysis of selected phage clones and identification of a consensus sequence

Following five rounds of selection to enrich and amplify phage displaying sequences readily cleaved by PSA, 88 phage clones were identified as PSA substrates (Table 2; Figure 2). Of the 88 sequences, 60 contained at least one tyrosine residue, and 41 sequences contained leucine. Ten sequences contained neither tyrosine nor leucine; seven of these ten sequences contained phenylalanine, and the remaining three sequences contained isoleucine, methionine or tryptophan. The widespread availability of tyrosine or leucine for P1 occupancy is in agreement with data from digestion of semenogelin, which shows that tyrosine and leucine are primarily used as P1 residues during hydrolysis of this physiologically relevant substrate [7,8].

Taken together, the substrate phage clones suggest the following consensus sequence for peptide substrates of PSA: P5 (G), P4 (A, G, S), P3 (S > T, A, R), P2 (Y > X), P1 (Y > L), P1′ (A, S, T > Q), P2′ (A, G, S > R), P3′ (A, G) (Figure 2a) where X denotes similar preference for several other amino acids. Such analysis of subsite preference from phage display data is complicated by the fact that almost every putative P5 to P3′ sequence contains at least one residue from one of the two constant linker sequences (GGAG and GGAGSS), biasing the library towards containing serine, glycine or alanine. After removing these constant residues from the selected sequences, however, the consensus sequence that emerges is only slightly altered: P5 (R, L > X), P4 (S > A), P3 (S > A, R, T), P2 (Y > X), P1 (Y > L), P1′ (S, T, A > Q), P2′ (S > A, R), P3′ (A, S) (Figure 2b), with glycine significantly less prominent at P5, P4, P2′, and P3′ but with other preferences unchanged. At the P2, P1′, and P2′ positions, substitutions which produced highly active peptides during iterative optimization are well represented in the phage display data, whereas subsititutions that yield low activity substrates, such as glycine or proline at P2, are rarely found among the selected sequences.

Dot blot ranking and refinement of the consensus sequence

Dot blot analysis [23,24], in which phage are treated with PSA, blotted onto a nitrocellulose membrane, and detected by assaying the immunoreactivity of any uncleaved mAb epitopes, was performed to monitor digestion of all 88 clones by PSA and to obtain an approximate ranking of their susceptibility to proteolysis. This procedure identifies those phage that are cleaved most rapidly and allows subsequent detailed kinetic assays to be focused on the most promising potential substrates. Under increasingly stringent assay conditions, nine of the 88 substrate phage

Table 2

Alignment of the randomized regions of selected clones from substrate phage display.

Clone	P8	P7	P6	P5	P4	P3	P2	P1	P1′	P2′	P3′	P4′	P5′	P6′	P7′
1					S	I	L	Y	G	V	S	K			
2							H	Y	S	A	V	T	R	P	
3			P	S	S	T	F	Y	R	A					
4		L	R	L	A	S	T	Y	S						
5					A	S	A	F	M	R	A	V			
6						S	L	Y	T	A	L	E	M		
7			L	R	S	R	M	Y	T	P					
8							Y	Y	T	S	P	S	P	S	
9					R	S	V	L	H	S	S	H			
10		P	V	R	N	R	V	L	A						
11			R	R	A	S	S	L	H	A					
12	Y	P	E	R	S	R	F	Y							
13					A	Y	Y	Y	T	V	R	A			
14								Y	Q	S	A	R	P	Q	S
15						R	K	Y	L	G	F	L	R		
16		A	P	R	A	S	V	Y	S						
17							Y	Y	R	A	P	D	S	A	
18						S	Y	Y	V	N	T	A	T		
19								L	A	R	A	H	D	A	N
20			L	T	G	A	R	L	T	S					
21		T	S	T	S	K	L	Y	A						
22			P	I	S	T	F	F	T	S					
23						T	M	Y	M	S	P	F	P		
24			T	L	R	S	A	Y	A	T					
25		F	A	H	S	S	S	Y	A						
26						R	L	Y	S	M	E	V	A		
27		F	P	P	P	R	L	Y	S						
28				V	Q	S	R	Y	Q	L					
29			R	P	A	A	V	Y	S						
30				K	A	R	N	Y	M	D	A				
31								F	Q	S	S	T	M	R	T
32			A	F	R	S	S	L	R	S					
33		G	S	V	S	R	A	Y	R						
34			Y	S	Q	L	Y	Y	T	S					
35			P	L	M	K	P	L	W	A					
36							F	Y	S	S	T	G	S	E	
37			P	Y	R	T	Q	L	F	A					
38		R	S	T	S	T	V	Y	T						
39				T	S	S	F	Y	H	N	P				
40			G	L	T	A	Y	Y	G	T					
41						H	P	L	S	S	F	K			
42				N	F	P	P	F	R	R	A				
43		P	R	L	S	Q	R	Y	G						
44						N	S	L	S	R	L	M	S		

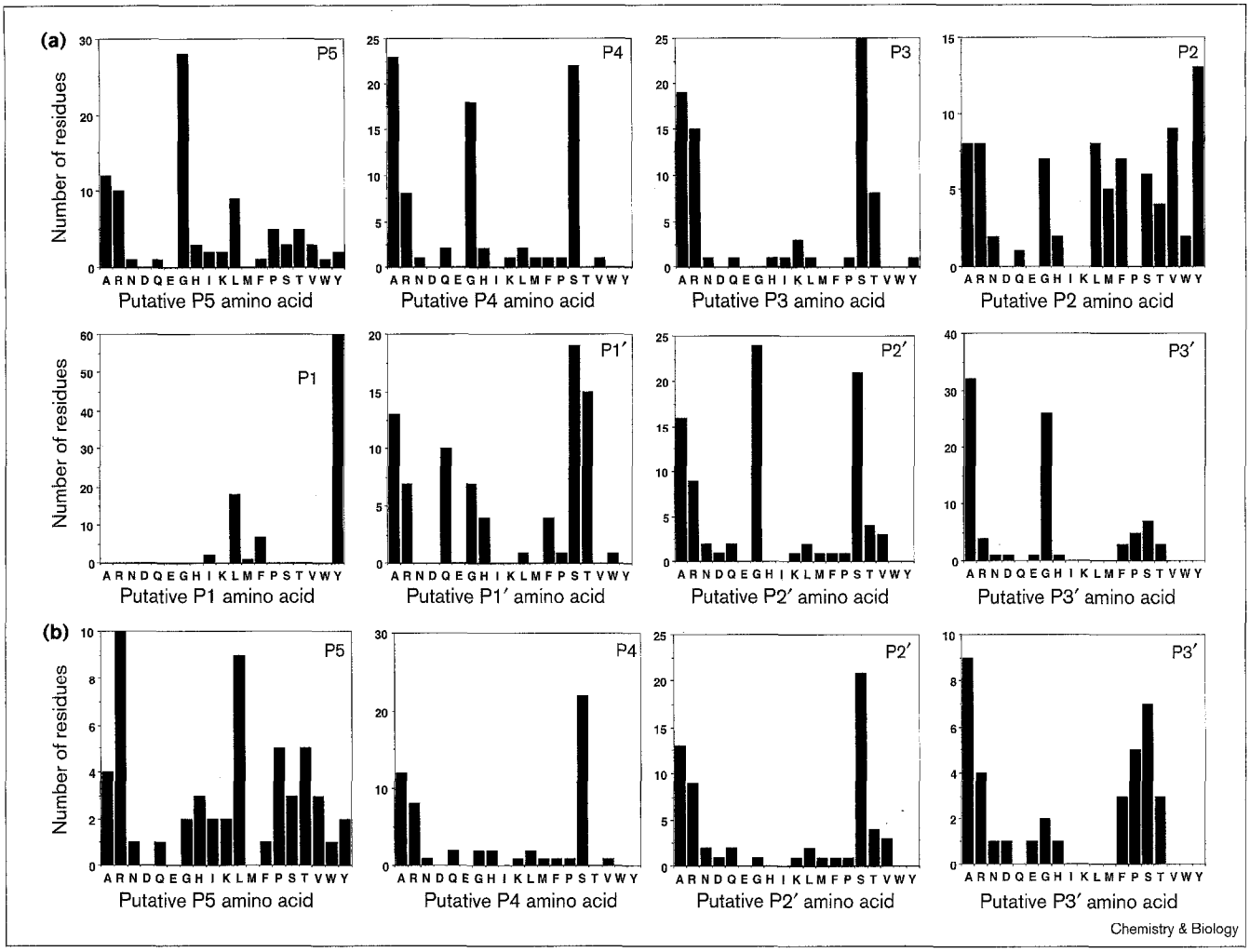
Clone	P8	P7	P6	P5	P4	P3	P2	P1	P1′	P2′	P3′	P4′	P5′	P6′	P7′
45								L	V	R	S	S	R	R	A
46			F	G	S	A	A	Y	Q	Q					
47	N	A	Q	R	S	T	Y	Y							
48							T	Y	R	A	S	H	Y	S	
49			S	I	R	S	V	Y	R	Q					
50					S	A	T	M	A	S	P	P			
51			R	S	R	S	R	L	F	R					
52		Y	T	V	T	K	V	Y	S						
53								Y	H	V	R	R	L	G	R
54				P	T	A	R	Y	S	A	A				
55								L	P	R	T	V	M	F	G
56	T	L	S	L	S	S	Y	Y							
57		A	L	R	G	A	M	Y	A						
58		S	R	F	S	S	Y	Y	T						
59		P	F	Q	T	S	L	Y	S						
60						S	S	Y	Q	S	R	R	L		
61		R	L	Y	S	T	L	Y	S						
62							P	L	S	R	V	Q	S	A	
63		T	R	L	A	R	P	Y	S						
64		W	T	R	S	S	N	Y	G						
65			S	T	R	A	S	Y	T	S					
66		S	P	R	S	A	V	I	S						
67							Y	Y	Q	A	F	S	R	R	
68		S	S	R	A	A	Y	A							
69					R	S	V	L	F	S	A	G			
70				W	S	A	M	Y	A	T	S				
71							Y	Y	A	L	G	Q	P	S	
72							F	Y	A	A	D	E	L	S	
73								L	S	R	H	Y	A	L	P
74					L	S	R	L	S	S	N	P			
75						R	W	I	R	K	A	Q	S		
76		T	T	K	H	R	F	Y	T						
77			S	A	A	A	W	F	T	A					
78				F	V	R	R	F	T	S	G				
79							Y	Y	Q	T	R	F	P	T	
80		S	G	A	T	S	M	Y	S						
81		V	F	L	S	R	A	Y	A						
82							Y	Y	Q	S	A	S	S	T	
83			P	H	A	R	R	Y	F	S					
84			L	H	H	S	A	Y	T	F					
85			R	A	A	R	L	Y	T	A					
86			S	G	L	S	A	L	Q	S					
87				A	T	S	T	F	Q	S	S				
88		K	K	L	S	T	H	Y	A						

Alignments were made assuming a preference for tyrosine > leucine > other hydrophobic residues at P1. Amino- and carboxyl-terminal residues flanking these sequences are contributed by the phage constant regions and are GGAG and GGAGSS respectively.

appeared to be particularly labile and were selected for further analysis (Table 3).

All nine sequences from the most labile phage contained a tyrosine residue that could act as a potential P1 residue. Assuming tyrosine as P1, seven of nine contained a hydrophobic residue at P5. Six of nine contained serine or threonine at P4, five of nine contained serine or threonine at P3, seven of nine contained a hydrophobic residue at P2, and five of nine contained serine at P1′. Only two of the nine clones had a randomized residue at P2′, in both cases alanine. For the other seven, this residue was either a glycine (for six sequences) or an alanine (for the other sequence) contributed by the constant phage linker sequence adjacent to the epitope tag. Similarly, at P3′ seven of nine clones contained either a glycine (one sequence) or an alanine (six sequences) contributed by the linker region. The observation that randomized residues do not make a strong contribution to P2′ or P3′ suggests that either glycine and alanine are preferred at P2′ and P3′ or there is no strong preference for any residue at these positions and glycine and alanine are appearing by default.

Figure 2



Elucidation of the occupancy of P5–P3′ sites through substrate phage display. **(a)** P5–P3′ occupancy including residues contributed by the constant phage linker region. **(b)** P5, P4, P2′ and P3′ occupancy excluding residues contributed by the phage linker region. Occupancy of P3–P1′ was similar regardless of whether residues contributed by the phage linker region were considered. P5–P3′ assignments were based on the assumption that occupancy of P1 by tyrosine or leucine was preferred.

Seven out of nine clones contained randomized residues at P6 and P7, but no clear consensus emerged for these positions, which are expected to be solvent exposed when peptide is bound to protease. It might, however, be significant that glycine was not found at P7 in any of these clones, and occurred at P6 in only two clones, even though its presence in the constant region greatly increased the potential for glycine to occupy these positions if it exerted a neutral effect. Thus, preferences for positions P7 to P3′ are P6/P7 (broad specificity, probably excluding glycine), P5 (hydrophobic), P4 (S > T, A), P3 (S), P2 (Y > V, L), P1 (Y), P1′ (S), P2′ (G > A), P3′ (A) with the caveat that P2′ and P3′ are biased by the identity of the constant linker region and might not exert a large effect on specificity. These results suggest a P4 to P2′ consensus of SSYYSG, a sequence strikingly similar to SSFYSS, the best sequence identified independently by alteration of a native sequence cleaved within semenogelin (Table 1) and consistent with the broader consensus obtained by analysis of all 88 substrate phage clones (Table 2).

Kinetic analysis of the cleavage of peptides containing sequences present in selected substrate phage

Six peptides containing amino-acid sequences present in the most labile phage and one peptide containing the consensus cleavage sequence SSYYSG were chosen for kinetic analysis to confirm the results obtained from phage display and to quantify catalysis of the putative cleavage sequences within selected phage (Table 4). All six selected peptides, **XV–XX**, were cleaved with efficiencies greater than peptides **I** and **II** derived from semenogelin,

Table 3

Alignment of selected clones showing the highest reactivity towards PSA as determined by monitoring cleavage of intact phage by a dot blot assay.

Clone	P8	P7	P6	P5	P4	P3	P2	P1	P1′	P2′	P3′	P4′	P5′	P6′
4		L	R	L	A	S	T	Y	S					
16		A	P	R	A	S	V	Y	S					
17							Y	Y	R	A	P	D	S	A
52		Y	T	V	T	K	V	Y	S					
54				P	T	A	R	Y	S	A	A			
56	T	L	S	L	S	S	Y	Y						
58		S	R	F	S	S	Y	Y	T					
61		R	L	Y	S	T	L	Y	S					
81		V	F	L	S	R	A	Y	A					

Alignments are made assuming preference for tyrosine > leucine at P1. Amino- and carboxyl-terminal residues flanking these sequences are contributed by the phage constant regions and are GGAG and GGAGSS respectively.

although, as will be discussed below, one peptide, **XXI**, was significantly less labile than the other selected sequences. Leaving aside **XXI**, the selected peptides were cleaved 10–33-fold more efficiently than tyrosine-containing peptide **II**. Increases in catalytic efficiency were reflected in improved values for k_{cat} for four of these five substrates, and in decreased values for K_m for all substrates. We also assayed cleavage of peptide **XXII** based on the consensus sequence from the selected peptides and observed that hydrolysis of **XXII** by PSA was more than twice as efficient as any one of the selected sequences. k_{cat}/K_m was 67-fold greater than that observed for peptide **II** despite the fact that **II** and **XXIII** share common P4, P3, P1, and P1′ residues.

Influence of surrounding subsites on P1 occupancy

As noted above, hydrolysis of peptide **XXI** containing the octamer sequence PTARYSAA was relatively inefficient with a catalytic efficiency of 76 $M^{-1}s^{-1}$, 6–18-fold less than that of the other selected sequences listed in Table 4. Surprisingly, high performance liquid chromatography (HPLC) purification of the cleaved peptide revealed two cleavage sites, and mass spectral analysis of the purified fragments indicated that the major product was due to cleavage after a P1 serine (PTARYS↓AA), with a smaller amount of cleavage after the tyrosine we had expected to be the predominant P1 residue (PTARY↓SAA) (Figure 3). We also observed a shift in specificity away from P1 tyrosine for peptides containing asparagine (PTARY↓NAA) or glutamine (PTARYQ↓AA) rather than serine (data not shown). Cleavage after P1 glutamine was more efficient than cleavage after asparagine and was similar to cleavage after serine. The altered P1 specificity is reminiscent of the observation by Isaacs and colleagues [18] that glutamine can function as a P1 residue in the context of substrates containing activated amide bonds with 7-amino-4-methyl coumarin leaving groups.

Table 4

PSA-mediated hydrolysis of peptides identified by substrate phage display and related variants.

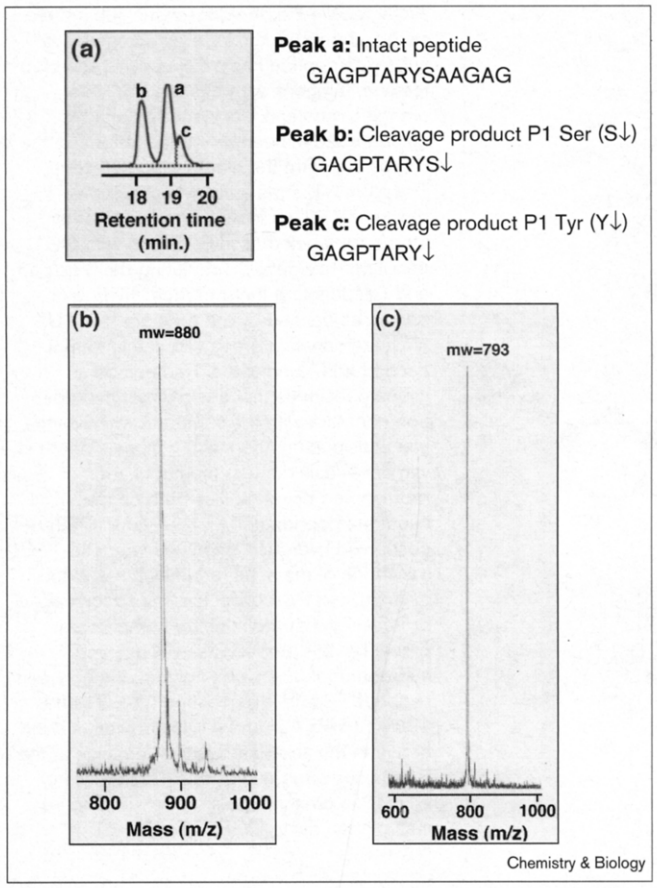
	Substrate *(Pn,...P3, P2, P1,↓P1′, P2′, P3′,...Pn′)	k_{cat} (s^{-1})	K_m (μM)	k_{cat}/K_m ($M^{-1}\,s^{-1}$)
XVI	GAGAPRASVY↓SGAG	3.3	7100	460
XVII	GAGYY↓RAPDSAGAG	14	10000	1400
XVIII	GAGSRFSSYY↓TGAG	5.1	4000	1300
XIX	GAGLRLASTY↓SGAG	2.0	4400	450
XX	GAGVFLSRAY↓AGAG	0.37	630	590
XXI	GAGPTARYS↓AAGAG	0.44	5800	76
XXII	GAGLRLSSYY↓SGAG	0.83	270	3100
XXIII	GAGLRLSSRY↓SGAG	<5.0	>38000	130
XXIV	GAGLRLSSRY↓SAAG	<3.8	>26000	150
XXVIII	GAGPTAKYS↓AAGAG	0.70	6500	110

*Positional nomenclature of subsite residues. Arrows denote the position of peptide bond hydrolysis as identified by mass spectral analysis. Peptide **XXI** is also cleaved after tyrosine, but the kinetics shown are obtained through monitoring initial rates of reaction after serine and prior to evolution of significant product from cleavage after tyrosine. Underlined amino acids in **XXII** and **XXIV** are changes relative to parent peptide **XXII**. Calculated error values were between 7 and 23%.

To investigate the origin of this unexpected shift in P1 specificity from tyrosine to serine for cleavage of **XXI**, we inserted residues found in peptide **XXI** into the P5, P2, and P2′ positions (based on cleavage at tyrosine at P1) of peptide **XXII** that contained the consensus sequence for efficient cleavage by PSA. Insertion of arginine adjacent to tyrosine to create peptide **XXIII** caused a greater than 140-fold increase in K_m, from 270 μM to greater than 38 mM (Table 4), and a reduction of overall catalytic efficiency, k_{cat}/K_m, of approximately 24-fold. As with parent peptide **XXII**, the predominant cleavage products for peptide **XXIII** were derived from cleavage after tyrosine. A measurable amount of products was also detected due to cleavage after serine, however (Figure 4a). Peptide **XXIV**, in which arginine and alanine were substituted for P2 tyrosine and P2′ glycine respectively, was also cleaved inefficiently by PSA and produced increased levels of cleavage after the serine. The additional substitution of proline for leucine to generate peptide **XXV** made serine the primary P1 residue (Figure 4a).

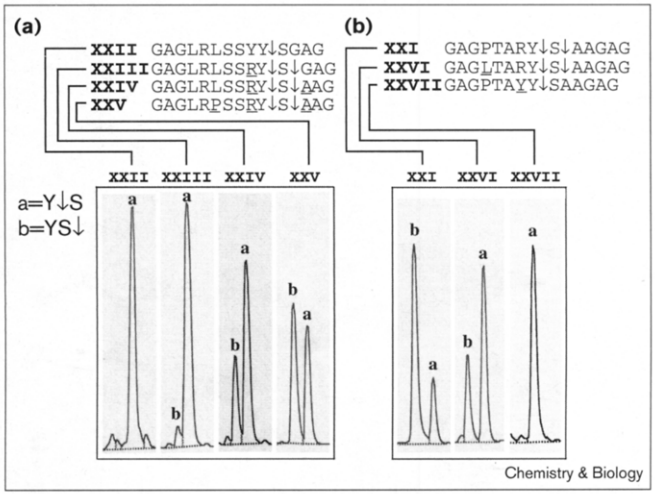
The gradual alteration of P1 specificity of peptides **XXII–XXV** from tyrosine to serine demonstrated that the introduction of non-P1 point mutations into consensus peptide **XXII** not only affected the rate at which PSA hydrolyzed variants of peptide **XXII** but also significantly influenced the selection of a P1 residue for this family of peptides. To probe further the molecular determinants that contributed to shifting the specificity of PSA for substrate P1 residues, we performed the converse experiment by attempting to shift the P1 specificity of PSA for peptide **XXI** containing the P6–P2′ sequence PTARYSAA from preferring serine as a P1 residue to preferring a P1

Figure 3



Shift of P1 selectivity for digestion of peptide **XXI**. **(a)** HPLC chromatogram of cleavage of peptide **XXI** containing the sequence PTARYSAA showing three peaks. Peak a corresponds to intact peptide, peak b to cleavage after P1 serine and peak c to cleavage after P1 tyrosine. **(b)** Mass spectral data for fragment b. **(c)** Mass spectral data for fragment c.

Figure 4



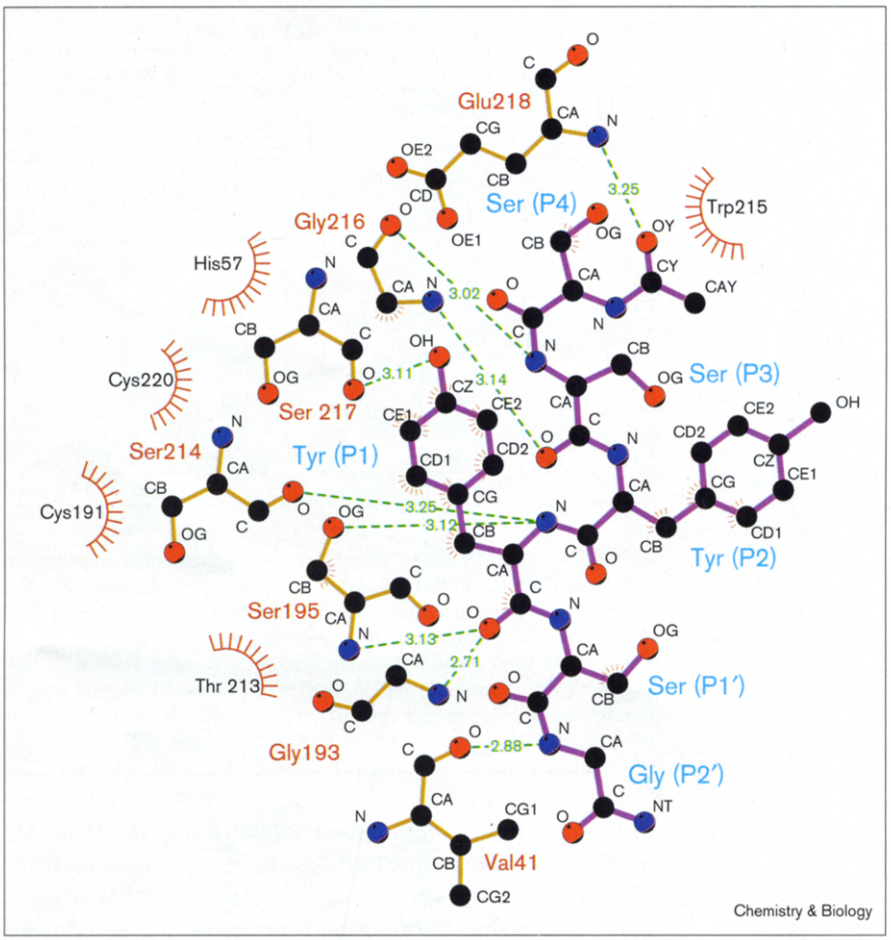
Shift of P1 selectivity during interchange of residues between peptides **XXI** and **XXII** and complete digestion by PSA. **(a)** Introduction of residues from peptide **XXI** into peptide **XXII** and HPLC monitoring of digestion by PSA. **(b)** Introduction of residues for peptide **XXII** into peptide **XXI**. Peak a denotes cleavage after P1 tyrosine. Peak b denotes cleavage after P1 serine. The identity of the P1 residue was determined by mass spectrometry. Underlined residues were introduced into the peptide **XXI** or **XXII**.

tyrosine. Two different point mutations in peptide **XXI** achieved this goal (Figure 4b). Replacement of the proline in peptide **XXI** with a leucine residue yielded peptide **XXVI**, which was preferentially cleaved by PSA using tyrosine as a P1 residue rather than serine. Substitution of a tyrosine for the arginine in **XXI** created peptide **XXVII**, which was also cleaved by PSA almost entirely by using tyrosine as a P1 residue.

These data reveal that occupancy of the S2 subsite by arginine is strongly disfavored by PSA. To determine if this was due to positive charge we synthesized peptide **XXVIII**, which contains a lysine at the analogous position of arginine in **XXI**. Similar to cleavage of **XXI**, peptide **XXVIII** is hydrolyzed slowly by PSA and exhibits a shift in preferred S1 occupancy from tyrosine to serine (Table 4), suggesting that the presence of positively charged residues adjacent to tyrosine encourages a substrate-binding mode in which tyrosine occupies the S2 subsite rather than the S1 subsite. We note that both iterative subsite optimization (Table 1) and substrate phage display (Table 2) suggest that S2 occupancy by tyrosine is preferred, a characteristic that probably works to promote productive occupancy of the S1 subsite by adjacent residues. Although disfavored occupancy of the S2 subsite by lysine or arginine is sufficient to produce some shifting of S1 preferences, the shift of S1 selectivity seen with peptides **XXIV** and **XXV** demonstrates that subsite interactions beyond S1 and S2 affect peptide bond cleavage. Both glycine to alanine and leucine to proline substitutions would tend to rigidify the peptide substrate and their presence might influence subsite occupancy by altering the conformation and/or restricting the flexibility of bound substrate.

The influence of S2 and other subsites on the specificity of PSA is reminiscent of the specificities of the closely related enzymes kallikrein and tonin. In porcine pancreatic kallikrein, the X-ray crystal structure suggests that the sidechain of a substrate P2 tyrosine can insert between Tyr99 and Trp215 (chymotrypsin numbering) [28], and kinetic studies reveal that bulky residues at P2 such as phenylalanine or tyrosine are preferred [29]. In addition, physiologically relevant cleavages of human kininogen by human tissue kallikrein occur at both Arg–Ser and Met–Lys bonds [30], revealing that permissive S1 specificity, similar to that observed for PSA, is required for proper processing of physiological substrates

Figure 5



Projected view of the consensus substrate peptide bound to the PSA. The consensus substrate peptide has purple bonds and blue residue numbers, whereas the PSA has orange bonds and black and red residue numbers. Atoms containing the letter Y in the P4 residue form the blocking acetyl group. The atom NT is the amidated group that blocks the P2′ carboxyl terminus. All other atoms are labeled as in PDB files with OG denoting γ oxygen, CB denoting the β carbon, and CA denoting the α carbon. Hydrogen bonds, as defined by the program HBPLUS [57], are drawn in green with the length of each bond (Å) indicated. Red numbers indicate residues that are in direct hydrogen-bond contact with the substrate whereas the black residue numbers show those that are in van der Waals contact (contacts are represented by red spikes both on the substrate peptide and on the PSA). Highly conserved hydrogen bonds between the backbone of the substrate/inhibitor and the mainchain of PSA expected in a canonical protease–substrate/inhibitor complex are present in this model. These conserved, hydrogen bonds include interactions between residues P3 and 216, P2 and 214, P1 and 193 and 195, P2′ and 41. In addition, a bond between the acetylated amino terminus of the P4 residue and the mainchain amide-NH of Glu218 is present in our model. This figure was plotted using LIGPLOT [58].

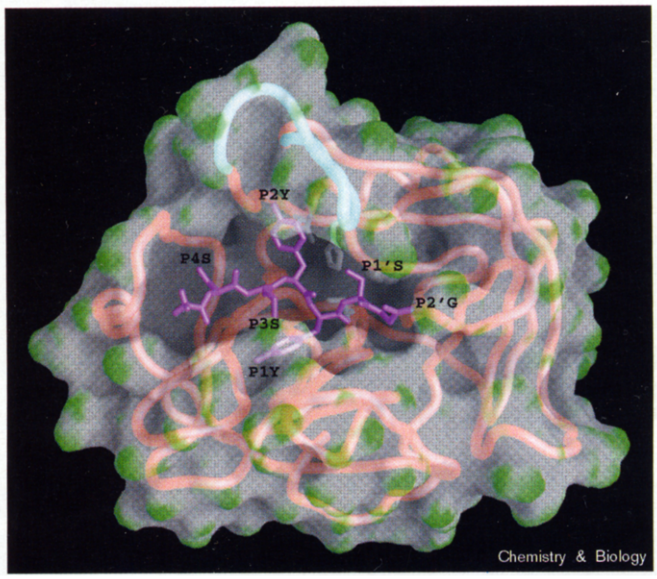
by tissue kallikrein. Similarly, tonin, which shares 53% homology with PSA, also has a broad S1 specificity. Tonin can cleave after P1 arginine, alanine, or phenylalanine residues [31] and is capable of releasing angiotensin II from angiotensinogen by cleaving a Phe–His bond [32]. An intriguing question is whether PSA, like kininogen or kallikrein, requires flexible recognition of P1 amino acids for proper physiological function or whether this trait is a remnant of its relatively recent divergence from kallikrein and tonin.

The dependence of the identity of the P1 residue on amino acid context demonstrates an important caveat for studies of protease specificity that use substrate phage display technology. We would not have appreciated the possibility of shifting P1 specificity for individual substrates without performing detailed analytical and kinetic characterization of hydrolysis by PSA of peptide substrates containing amino acid sequences present in selected substrate phage clones. Although certainly informative, sequence alignments of substrate phage alone cannot directly provide definitive qualitative, much less quantitative, information on protease selectivity. On the other hand, kinetic analysis of individual peptide substrates is severely limited by the small fraction of potential substrates that can be examined and by the solubility of individual peptides. Furthermore, in our studies it was the consensus peptide derived from phage display, not one of the individual selected sequences, that afforded the most efficiently hydrolyzed substrate. Only by using the two strategies together, then, is it possible to examine protease specificity both comprehensively and quantitatively.

Use of sequence preferences to refine a structural model for PSA–substrate recognition

Several investigators have previously reported models for PSA [25,33,34], and one of these models has been deposited in the Protein Data Bank (1PFA.PDB) [25]. Attempts to dock the optimized, P4–P2′ consensus substrate SSYY↓SG into this model, however, were unsuccessful, primarily due to significant steric clashes between the P2 tyrosine residue of the substrate and the large insertion loop at position 95 (chymotrypsin numbering system) of PSA.

Figure 6



Docking of the consensus substrate peptide to the PSA. The substrate peptide is drawn in purple from position P4 to P2′. PSA is represented as a red trace from all its Cα atoms. The molecular surface of PSA is colored according to the curvature of all the heavy atoms present in the model. The blue region of the PSA trace corresponds to the 11 residue insertion, compared with chymotrypsin, between residues 95 and 96 (chymotrypsin numbering system). Sidechains of residues that form the catalytic triad (Asp102, His57 and Ser195) are white.

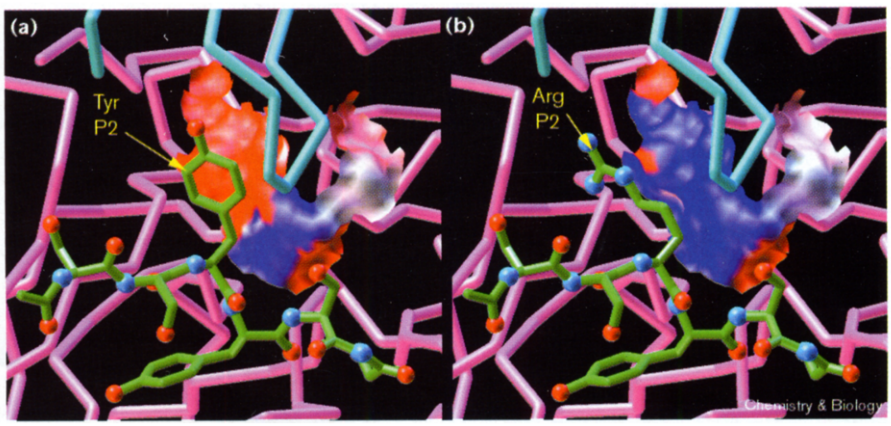
To gain additional insight into the molecular basis of substrate specificity for PSA, we constructed a new model of PSA bound to substrate (Figures 5–7). The known structure of the closely related serine protease porcine tissue kallikrein [28], which exhibits 60% amino-acid identity to PSA, was chosen to begin homology modeling of PSA. It is generally believed that substrates and inhibitors bind in the active-site cleft of chymotrypsin family serine proteases in a single 'canonical' conformation characterized by formation of a short anti-parallel β sheet between the P3–P1 residues of the substrate or inhibitor and residues 214–216 (chymotrypsin numbering) of the protease. This short anti-parallel β sheet contains three highly conserved hydrogen bonds: the amide-NH group of the P1 residue to the carbonyl group of Ser214, the carbonyl group of the P3 residue to the amide-NH group of Gly216 and the amide-NH group of the P3 residue to the carbonyl group of Gly216. All three hydrogen bonds of the canonical anti-parallel β sheet are present in our PSA model, and all are within optimal distances. In addition, the carbonyl oxygen atom of the P1 residue of the consensus substrate extends into the oxyanion hole of PSA where, as expected, it forms hydrogen bonds with both the amide-NH of Gly193 and the amide-NH of Ser195. Because the amino terminus of the consensus substrate in the model is acetylated, a new hydrogen bond occurs between the carbonyl oxygen atom of the acetyl group and the amide-NH group of Glu218. This new hydrogen bond suggests the pathway that the next residue (P5) of a proteinacious substrate might take.

The backbone conformation of the docked peptide substrate makes a type VIa2 turn [35] at the P2–P1 position. A type VIa2 turn is characterized by a pair of dihedral angles β–α_r for residue i + 1 and i + 2. Every other residue of the substrate has an extended conformation (β region). Sidechains of all substrate residues are within the known rotamer library. The Cβ atom of the P4 serine residue is in direct van der Waals contact with Trp215, a highly conserved residue, which suggests that peptides containing a β-branched amino acid at P4 might be poor substrates of PSA. The serine residue at position P3 is in hydrogen-bonding distance with the atoms Oε1 and Oε2 of Glu218 of PSA, although the geometry is not ideal. The sidechain of the P3 serine residue is partially buried by this glutamate residue. The presence of an acidic amino acid at position 218 is unusual in the (chymo)trypsin serine protease family [25]; consequently, these interactions might contribute to the specificity of PSA.

The tyrosine residue at position P2 of the docked substrate fits into a pocket created by the large 95–96 loop of PSA. The phenol ring of the tyrosine lies against the imidazole ring of His57, a member of the catalytic triad. The P2 tyrosine is covered by the sidechain of Arg95, and the hydroxyl group of this tyrosine forms a hydrogen bond with the Nη2 group of Arg95G. Although Arg95G is expected to exhibit significant mobility due to the intrinsic flexibility of the 95–96 loop of PSA, the location of this basic residue suggests that, as observed in this study, its affect on the position of a P2 basic residue would decrease catalysis, probably by perturbing the electrostatic potential in the vicinity of the catalytic triad (Figure 7).

The tyrosine at position P1 of the peptide substrate binds to the specificity pocket nicely. The S1 pocket of PSA is formed by residues 213–220, 189–192, and 225–228, and is closed by the P2 and P3 residues of the substrate itself. The sides of the S1 pocket are hydrophobic but the bottom of this pocket is polar, containing Oγ atoms from Ser189 and Ser226 and a carbonyl group from Ser217. The tyrosine-OH group is within hydrogen-bonding distance of both the Oγ of Ser226 and the mainchain carbonyl oxygen of Ser217. The inward side of the pocket is locked by a hydrogen bond between the Oγ atom of Ser192 and the Oγ atom of Ser226. It was postulated previously [25] that the S1 binding pocket would prefer hydrophobic residues. Like hydrophobic residues such as phenylalanine or leucine, a tyrosine residue at P1 can form a large number of van der Waals contacts with surrounding PSA residues along the sides of the specificity pocket. Unlike these hydrophobic residues, however, a P1 tyrosine residue can also form the hydrogen-bonding

Figure 7



Perturbation of the electrostatic potential of the Asp102–His57 interaction by a basic residue at the P2 position of a PSA substrate. Electrostatic potentials were calculated using Delphi (Version 3.0) [59] and PARSE parameters [60] with a 129 cubes grid, allowing a final resolution of 2.4 grid point/Å. Parameters used in these calculations were as follows: internal and external dielectric constants were 2 and 80, respectively, ionic strength was 150 mM, ionic exclusion radius was 2 Å, and the probe radius for surface calculation was 1.4 Å. One focusing step was used (initial was 49%, final was 98%). The electrostatic potential was read into AVS (AVS, Waltham, MA) and mapped onto the molecular surface of residues 57 and 102. The PSA is drawn as a Cα trace in pink with the 95–96 loop highlighted in cyan. The substrate peptide is drawn in green. In panel **(a)** the consensus substrate is shown with a tyrosine at position P2 whereas in **(b)** we replaced the residue P2 by an arginine. Positive potential is indicated by blue and negative potential is indicated by red. In (b) the electrostatic potential in the vicinity of Asp102 is altered due to juxtaposition of the buried P2 arginine from the substrate.

interactions with residues at the bottom of the specificity pocket as described above. Because they are buried in the hydrophobic S1 pocket, these hydrogen-bonding interactions are expected to be strong.

To test this hypothesis computationally, we mutated the consensus substrate peptide by replacing the tyrosine in P1 by either leucine, phenylalanine, isoleucine or methionine. We calculated the difference in the gas phase energy between the mutated and the consensus substrate using the CHARMM22 all atom force field. We obtained the following positive differences in energy: 1.0, 12.3, 30.9, 46.4 kcal/mol, for leucine, phenylalanine, isoleucine, and methionine mutants, respectively. Consistent with the observations of this study, the calculations predict that the consensus substrate is more energetically favorable than the mutated substrates.

These calculations also suggest that, compared with tyrosine, phenylalanine is strongly disfavored at P1 primarily due a loss of 10 kcal/mol from the electrostatic energy term. Finally, our calculations suggest that the large hydrophobic sidechains (isoleucine, phenylalanine, methionine) are more disfavored than the shorter leucine sidechain due to reductions in both the van der Waals and electrostatic energy terms. These results imply that the bottom of the S1 binding pocket is polar and, consequently, that large hydrophobic sidechains provide poor electrostatic complementarity in this pocket. The rank order of preference for the five amino acid residues observed at the P1 position of the selected, substrate phage is identical to that calculated using our new model, a result that provides support for the computational value of the model.

The serine at position P1′ of the docked, consensus substrate is in van der Waals contact with His57, suggesting that, as observed in the selected substrate sequences, small residues might be preferred at this position. The glycine residue at position P2′ in the consensus substrate is bordered by Ser192, Gly193, Val41, Cys42, and Arg39. The carbonyl group of the P2′ glycine is within hydrogen-bonding distance of both the Nε and Nη1 groups of Arg39. The size of the S2′ pocket suggests that large residues would be poorly accommodated, an observation that is consistent with the substrate phage display results described above.

Comparison of peptide digestion by PSA and chymotrypsin

PSA shares with chymotrypsin a preference for recognition of hydrophobic amino acids at the S1 subsite. To put catalysis by PSA into better perspective relative to an enzyme known for its high activity and lack of sequence selectivity, we assayed hydrolysis of three substrate peptides **III**, **XVI**, and **XXII** by chymotrypsin and compared the kinetics of digestion to those observed using PSA (Table 5). These peptides were chosen because they varied from poor, **III**, to moderate, **XVI**, to highly labile, **XXII**, substrates for PSA.

Hydrolysis of the peptides by PSA was more sequence dependent than hydrolysis by chymotrypsin, as k_{cat}/K_m values for cleavage by PSA varied 1200-fold, whereas k_{cat}/K_m values for cleavage by chymotrypsin varied by only 3.6-fold. The cleavage of each peptide by chymotrypsin was more efficient than catalysis by PSA, with k_{cat}/K_m values 24–4300-fold greater for chymotrypsin than for PSA. Relatively inefficient catalysis by PSA was reflected in both K_m and k_{cat} values, with K_m values for cleavage by

Table 5

Comparison of hydrolysis of peptide substrates by PSA and chymotrypsin.

		PSA			Chymotrypsin			
	Substrate *(Pn,...P3, P2, P1,↓P1′, P2′, P3′,...Pn′)	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 $(M^{-1}\ s^{-1})$	k_{cat}/K_m Chymo/PSA
III	GSQQLY↓HNKQEGRD	0.037	14000	2.6	24	230	1.1 $(10)^5$	4300
XVI	GAGAPRASVY↓SGAG	3.3	7100	460	36	130	2.7 $(10)^5$	590
XXII	GAGLRLSSYY↓SGAG	0.83	270	3100	9.8	130	7.5 $(10)^4$	24

*Positional nomenclature of subsite residues. Arrows denote the position of peptide bond hydrolysis as identified by mass spectral analysis. Calculated error values were between 13–31%.

PSA increased from 2–61-fold relative to values for chymotrypsin, and k_{cat} values decreased 12–650-fold. The lower efficiency of PSA-mediated catalysis suggests that PSA is generally less able than chymotrypsin to bind substrate in a catalytically productive orientation, although the magnitude of the difference between PSA and chymotrypsin is dependent on the precise amino-acid sequence of the substrate.

The observation that, compared with chymotrypsin, PSA is a less efficient catalyst for peptide hydrolysis but possesses a more sharply defined specificity is similar to results previously obtained by comparing peptide hydrolysis by trypsin, tissue-type plasminogen activator (t-PA) [23], and urokinase-type plasminogen activator (u-PA) [24]. Like chymotrypsin, trypsin is a highly active protease with little preference for sequence beyond the S1 subsite, whereas t-PA and u-PA are highly selective enzymes that achieve selectivity in spite of substantial sequence identity and overall structural similarity [36,37] with trypsin. Like PSA relative to chymotrypsin, t-PA and u-PA are much less active and much more selective towards peptide substrates than is trypsin. Furthermore, similar to correlations we observe for PSA, cleavage of peptides by u-PA or t-PA is characterized by higher K_m and lower k_{cat} values than observed with trypsin. These parallels suggest, like u-PA and t-PA, PSA might possess restricted substrate specificity *in vivo*, with the low activity of PSA towards most potential substrates being a physiologically important mechanism for controlling proteolysis.

Our previous studies of optimal subsite occupancy for t-PA demonstrated that the most labile substrate sequences for this protease were not particularly selective and were also cleaved efficiently by u-PA [24]. To isolate substrates that could discriminate between these two closely related mammalian plasminogen activators, we developed a new technique, substrate subtraction phage display, which included a subtraction step during the phage selection that removed substrates that were cleaved efficiently by u-PA [38]. Our observation that hydrolysis of selected peptides **III**, **XVI**, and **XXII** by chymotrypsin was more efficient than hydrolysis by PSA suggests that a similar approach may be required to isolate peptide substrates that are specific for PSA. An alternative approach to achieving PSA/chymotrypsin selectivity could take advantage of the relatively permissive P1 requirements of PSA, and Isaacs and coworkers [18] have previously demonstrated that molecules that include glutamine as a P1 residue are specific for hydrolysis by PSA relative to hydrolysis by chymotrypsin.

Significance

Prostate-specific antigen (PSA), a (chymo)trypsin family protease is an important marker for prostatic malignancy. Two independent approaches to defining protease specificity, substrate phage display and iterative optimization of native substrates, converge on SS(Y/F)Y↓SG as a consensus sequence for efficient peptide hydrolysis by PSA. Although tyrosine is preferred as a P1 residue, non-P1 residues also play an important role in defining labile substrates and can even cause otherwise disfavored amino acids to bind in the S1 pocket. Thus, the common characterization of the specificity of PSA as 'chymotrypsin-like' is insufficient and potentially misleading. Understanding of substrate preferences has allowed refinement of existing structural models for PSA. The new model provides insight into the molecular determinants of the preference for tyrosine as a P1 residue and suggests that the P2 residue of a substrate or inhibitor fits into a pocket formed by a large insertion loop that is not present in most chymotrypsin family proteases.

Elucidation of substrate sequence specificity for PSA will facilitate its use as a scaffold for engineering selective proteolysis and investigating the evolution of protease specificity. Such studies are particularly intriguing because the 78% identity between PSA and human glandular kallikrein [39], an enzyme specific for arginine at P1, suggests that the kallikrein scaffold could be relatively adaptable and that it might be easier to alter the P1 specificity of PSA than has been the case with trypsin or chymotrypsin [40,41], where such efforts have required a relatively large number of point mutations and have com-

promised catalysis. In addition, the importance of recognition of the S2 and other subsites suggests that PSA has an extended substrate-binding site that provides multiple targets for the alteration of specificity by mutagenesis, further facilitating use of PSA as a starting point for the development of novel specificities by rational design or selection.

Knowledge of substrate preferences for PSA should facilitate development of more sensitive assays for catalysis to complement existing immunoassays and enhance the diagnostic information that can be gained from analysis of PSA. Highly specific substrates for PSA should also guide design of inhibitors useful for probing the importance of PSA levels during tumor progression and for the investigation of additional physiological functions of the enzyme in both cancerous and normal prostatic tissue. Inhibitors could also act as tight-binding targeting agents for delivering immunotoxins, or radioimaging agents to prostate tumors; additionally, PSA-directed prodrugs have recently been shown to be cytotoxic to PSA-producing cells [42].

Materials and methods

Determination of the concentrations of active proteases

Purified PSA was obtained from Calbiochem (La Jolla, CA). Purified chymotrypsin and insulin β chain were obtained from Sigma (St. Louis, MO). The concentration of active PSA or chymotrypsin was determined by titration with 4-methylumbelliferyl p-(NNN-trimethylammonium)-cinnamate chloride (MUTMAC) (Sigma, St. Louis, MO) [43]. The reaction of PSA with MUTMAC was too slow to be accurately quantitated by continuous measurement of the release of fluorescent material, so we measured the difference in fluorescence between the initial MUTMAC solution and the solution after complete labeling of PSA. Titration reactions were performed in buffer containing 50 mM Tris, pH 8.0, 10 mM $CaCl_2$, and 0.1% Tween-20. Titration reactions were allowed to proceed to completion and fluorescent measurements were taken with a Perkin-Elmer LS50B Luminescence fluorometer with excitation and emission wavelengths of 365 nm and 445 nm, respectively. Titration reactions with PSA were performed in a 30 µl volume for 10 min to ensure that all the enzyme had reacted and then diluted to 1 ml for fluorescent measurement. Calculated concentrations were similar regardless of the time allowed for labeling of PSA by MUTMAC and approximated the antigenic concentration of PSA determined by antibody labeling.

Substrate phage display selection using PSA

Substrate phage panning of polyvalent phage has been described previously [20,23,24]. For PSA, selection of positive phage was performed as follows: approximately $3(10)^{10}$ phage (33–45 µl of the fTC-LIB-N8 library) in a PSA compatible assay buffer (50 mM Tris pH 7.4, 100 mM NaCl, 10 mM $CaCl_2$, 0.01% Tween 20) was digested for 4 h at 37°C with concentrations of PSA ranging from 10–100 µg/ml. Total volume of each digest was 250 µl. Reactions were then placed on ice, bovine serum albumin was added to 0.1%, and 100 µg mAb 3-E7 was added. After 30 min on ice, 300 µl of Pansorbin cells (protein A-bearing *Staphylococcus aureus*, Calbiochem) were added and the resultant slurries were rocked at 4°C for 1 h. The mixtures were then centrifuged for 2 min and supernatants were recovered to repeat the Pansorbin cell adsorption step. Final supernatants were amplified overnight in *Escherichia coli* K91 cells. A small aliquot of each final supernatant was used for titering on K91.

Individual clones were selected from the titer plates, grown in 4 ml cultures and tested as PSA substrates by dot blot analysis as described previously [23,24]. Four subsequent rounds of selection were performed. In round 2, phage were treated for one or four hours with 5 or 20 µg/ml PSA. Round 3 libraries were treated for 1 h with 2.5, 5, or 10 µg/ml PSA. Round 4 libraries were treated with one or two µg/ml PSA for 30 min or 1 h. In round 5 phage were digested for 5 or 10 min with 0.5 or 1 µg/ml PSA, and reactions were stopped by adding 15 µl of 128 mM PMSF.

One additional modification was added to the protocol for rounds 4 and 5. Phage libraries were initially treated with the 3-E7 antibody mAb 3-E7 (Gramsch Laboratories, Schwabhausen, Germany) and adsorbed to Pansorbin cells, and substrate phage were released by proteolytic cleavage. The resultant supernatants were then resubjected to Pansorbin cell adsorption. This technique was adopted to prevent accumulation of 'nonreactive' phage, which do not interact with mAb 3E-7 even before treatment with protease. These phage are normally selected along with substrate phage, and their accumulation typically limits the number of rounds of selection that can be performed. 'Nonreactive' phage have not yet been characterized comprehensively; it seems likely, however, that the randomized octamer in many of these phage is either efficiently digested by *E. coli* proteases or interacts with the 3E-7 epitope. Alternatively, nonreactive phage could have deleted or mutated the antibody epitope. Regardless of the mechanism by which the nonreactive phage escape recognition by mAb 3E-7, these phage were eliminated by the modification described above for the round 4 and 5 selections.

Kinetics of peptide cleavage

Peptides were synthesized and purified as described previously [23,24]. Kinetic data were obtained by incubating varying concentrations of peptide with a constant enzyme concentration to achieve between 5 and 20% cleavage of peptide during each reaction. Reactions were performed in 50 mM Tris, pH 8.0, 10 mM $CaCl_2$, 0.1% Tween-20 as described previously [23,24]. The percentage of cleavage was correlated with the time of digestion to ensure that conditions were within the range of linear responsiveness for the assay. PSA concentrations varied between 50 nM and 200 nM. Peptide concentrations varied between 100 µM and 12 mM and in all cases were chosen to surround K_m. Both parent peptides and cleavage products were collected and characterized by mass spectral analysis using a Voyager MALDI–TOF spectrometer (PerSeptive Biosystems). Data were interpreted by Eadie-Hofstee analysis. Statistical errors were calculated as described previously [44].

Modeling of PSA

Sidechains of tissue kallikrein were replaced by those of PSA using the most favored rotamer conformation according to the library of Tuffery *et al.* [45] (S.W.W. Chen and J.-L.P., unpublished observations). Sidechain conformations were minimized using the all atom force field CHARMM22 [46] in the X-PLOR package [47]. We used the conjugate gradient method to minimize the resulting coordinates [48], and we selected shifted electrostatic and switched van der Waals functions for nonbonded terms. The dielectric constant was set at 1.0 and the nonbonded cut off was 12 Å with a cut on and a cut off of 8.5 Å and 11 Å, respectively. Hydrogens were systematically rebuilt before each minimization step using the program HBUILD available in X-PLOR [49].

The single deletion in PSA, compared with tissue kallikrein, occurs at position 146A–147 (chymotrypsin numbering) in the 'autolysis loop'. Consequently, the autolysis loop of the PSA model was built by starting with the corresponding region of tissue kallikrein (2kai.pdb, 2pka.pdb) and manually resolving the two-residue deletion using Turbo-Frodo. All sidechains were minimized, excluding the electrostatic term, until the norm of the gradient reached 0.5.

Compared with the archetypal serine protease trypsin, PSA, kallikrein and tonin have an insertion of 11 residues between positions 95 and 96 (chymotrypsin numbering system). We extracted a portion of this loop (residues 95C–96) from the model of Villoutreix and colleagues [25] and inserted it into our current model. We then inserted residues 95–95B

from the corresponding loop in the tonin structure into our model and manually adjusted the dihedral angles in this region to connect the two loop segments. Following manual adjustment of these dihedral angles, the 95–96 loop was minimized using parameters described above.

We used the new PSA model to dock the consensus substrate sequence SSYY↓SG in the 'canonical' conformation that has been observed in many protease/inhibitor complexes, including the complex between hirustatin and kallikrein [50]. The sidechains of residues 27–32 of hirustatin (VHCRIR) from the hirustatin/kallikrein complex (1hia.pdb) were mutated to match the consensus PSA substrate SSYY↓SG using the program Xfit [51], and this six amino-acid sequence was then docked into the PSA model. We used the reactive center loop of hirustatin for this procedure for two reasons. First, unlike many 'standard' protease inhibitors, hirustatin does not contain proline in its reactive center loop and, second, hirustatin inhibits kallikrein, the serine protease that is most closely related to PSA. Chi dihedral angles of the tyrosines at position P1 and P2 were manually adjusted. All sidechains of PSA and the docked substrate were optimized, excluding the electrostatic term, until the norm of the gradient reached 0.5. The chirality of residues 20, 91 and 221 were adjusted in Turbo-Frodo as recommended by a PROCHECK output [52].

To optimize the geometry of the large 95–96 loop of PSA in the docked complex, we performed a molecular dynamic simulation at 300K for 50 ps to explore more favorable conformational states. We used the Verlet algorithm [53] with a temperature coupling method according to Berendsen *et al.* [54]. Initial velocities were obtained from a Maxwellian distribution. During the simulation, the integration step was 1 fs, bond distances were restrained by the shake method [55], the electrostatic term was turned off, and other nonbonded parameters were identical to those used during previous minimizations. Examination of the Ramachandran map produced by the PROCHECK program [52] indicated that this procedure substantially improved the geometry of the final conformation of the new PSA model. Finally, we minimized all atoms of the consensus substrate peptide and obtained the final docking conformation shown in Figure 4.

Acknowledgements

The authors wish to thank Steve Madden for peptide synthesis. D.R.C. is an Assistant Investigator with the Howard Hughes Medical Institute. This work was supported in part by National Institute of Heath Grants RO1 HL52475 and PO1 to E.L.M. and by a grant from the Robert A. Welch Foundation (I-1244) to D.R.C.

References

1. Seregni, E., Botti, C., Ballabio, G. & Bombardieri, E. (1996). Biochemical characteristics and recent biological knowledge of prostate-specific antigen. *Tumori* **82**, 72-77.
2. Sokoll, L.J. & Chan, D.W. (1997). Prostate specific antigen. Its discovery and biochemical characteristics. *Urologic Clinics North America* **24**, 253-259.
3. Abrahamsson, P-E., Lilja, H. & Oesterling, J.E. (1997). Molecular forms of serum prostate-specific antigen. *Urologic Clinics North America* **24**, 353-365.
4. Crawford, E.D., DeAntoni, E.P. & Ross, C.A. (1996). The role of prostate-specific antigen in the chemoprevention of prostate cancer. *J. Cell. Biochem.* **25S**, 149-155.
5. Montie, J.E. & Meyers, S.E. (1997). Defining the ideal tumor marker for prostate cancer. *Urologic Clinics North America* **24**, 247-252.
6. Christiansen, A., Laurell, C-B. & Lilja, H. (1990). Enzymatic activity of prostate-specific antigen and its reactions with extracellular serine proteinase inhibitors. *Eur. J. Biochem.* **194**, 755-763.
7. Robert, M., Gibbs, B.F., Jacobson, E. & Gagnon, C. (1997). Characterization of prostate-specific antigen proteolytic activity on its major physiologic substrate, the sperm motility inhibitor precursor/semenogelin I. *Biochemistry* **36**, 3811-3819.
8. Peter, A., Lilja, H. & Malm, J. (1998). Semenogelin I and semenogelin II, the major gel-forming proteins in human semen are substrate for transglutaminase. *Eur. J. Biochem.* **252**, 216-221.
9. Cohen, P., *et al.* & Rosenfled, R.G. (1992). Prostate-specific antigen (PSA) is an insulin-like growth factor binding protein-3 protease found in seminal plasma. *J. Clin. Endocrinol. Metab.* **5**, 1046-1053.
10. Kanety, H., *et al.* & Karasik, A. (1993). Serum insulin-like growth factor-binding protein-2 (IGFBP-2) is increased and IGFBP-3 is decreased in patients with prostate cancer: correlation with serum prostate specific antigen. *J. Clin. Endocrinol. Metab.* **77**, 229-233.
11. Webber, M.M., Waghray, A. & Bello, D. (1995). Prostate-specific antigen, a serine protease, facilitates human prostate cancer cell invasion. *Clin. Cancer Res.* **1**, 1089-1-94.
12. Iwamura, M., Hellman, J., Cockett, A.T., K., Lilja, H. & Gershagen, S. (1996). Alteration of the hormonal bioactivity of parathyroid-related hormone protein (PTHrP) as a result of limited proteolysis by prostate-specific antigen. *Urology* **48**, 317-325.
13. Watt, K.W. K., Lee, P-J., M'Timkulu, T., Chan, W-P. & Loor, R. (1986). Human prostate-specific antigen: structural and functional similarity with serine proteases. *Proc. Natl Acad. Sci. USA* **83**, 3166-3170.
14. Schaller, J., Akiyama, K., Tsuda, R., Hara, R., Marti, M. & Rickli, E.E. (1987). Isolation, characterization and amino acid sequence of γ-semenoprotein, a glycoprotein from human seminal plasma. *Eur. J. Biochem.* **170**, 111-120.
15. Clements, J.A. (1994) The human kallikrein gene family: a diversity of expression and function. *Mol. Cell. Endocrinol.* **99**, C1-C6.
16. Akiyama, K., Nakamura, T., Iwanaga, S. & Hara, M. (1987). The chymotrypsin-like activity of human prostate-specific antigen, γ-semenoprotein. *FEBS Lett.* **225**, 168-172.
17. Lilja, H., Abrahamsson P. & Lundwall, A. (1989). Semenogelin, the predominant protein in human semen. *J. Biol. Chem.* **264**, 1894-1900.
18. Denmeade, S.R., Wei, L., Lovgren, J., Malm, J., Lilja, J. & Isaacs, J.T. (1997). Specific and efficient substrates for assaying the proteolytic activity of prostate specific antigen. *Cancer Res.* **57**, 4924-4930.
19. Mathews, D.J. & Wells, J.A. (1993). Substrate phage: selection of protease substrates by monovalent phage display. *Science* **260**, 1113-1117.
20. Smith, M.M., Shi, L. & Navre, M. (1995). Rapid identification of highly active and selective substrates for stromelysin and matrilysin using bacteriophage peptide display libraries. *J. Biol. Chem.* **270**, 6440-6449.
21. Scarborough, P.E., *et al.*, & Dunn, B.M. (1993). Exploration of subsite binding specificity of human cathepsin D through kinetics and rule-based molecular modeling. *Protein Sci.* **2**, 264-272.
22. Kinoshita, A., Urata, H., Bumpus, F.M. & Husain, A. (1991). Multiple determinants for the high substrate specificity of an angiotensin II-forming chymase from the human heart. *J. Biol. Chem.* **266**, 19192-19197.
23. Ke S-H., Coombs, G.S., Tachias, K., Corey, D.R. & Madison, E.L. (1997). Optimal subsite occupancy and design of a selective inhibitor of Urokinase. *J. Biol. Chem.* **272**, 20456-20462.
24. Ding, L., Coombs, G.S., Strandberg, L., Navre, M., Corey, D.R. & Madison, E.L. (1995). Origins of the specificity of tissue-type plasminogen activator. *Proc. Natl Acad. Sci. USA* **92**, 7627-7631.
25. Villoutreix, B.O., Getzoff, E.D. & Griffin, J.H. (1994). A structural model for the prostate disease marker, human prostate-specific antigen. *Protein Sci.* **3**, 2033-2044.
26. Takayama, T.K., Fujikawa, K. & Davie, E.W. (1997). Characterization of the precursor of prostate-specific antigen. *J. Biol. Chem.* **272**, 21582-21588.
27. Frenette, G., Gervais, Y., Tremblay, R.R. & Dube, J.Y. (1998). Contamination of purified prostate-specific antigen by kallikrein hK2. *J. Urology* **159**, 1375-1378.
28. Bode, W., Chen, Z. & Bartels, W. (1983). Refined 2 Å X-ray crystal structure of porcine pancreatic kallikrein A, a specific trypsin-like serine protease. *J. Mol. Biol.* **164**, 237-282.
29. Fiedler, F. (1987). Effects of secondary interactions on the kinetics of peptide and peptide ester hydrolysis by tissue kallikrein and trypsin. *Eur. J. Biochem.* **163**, 303-312
30. Chagas, J.R., *et al.* & Prado, E.S. (1995). Determinants of the unusual cleavage specificity of lysyl-bradykinin-releasing kallikreins. *Biochem. J.* **306**, 63-69.
31. Fujinaga, M. & James, M.N.G. (1987). Rat submaxillary gland serine protease, Tonin. Solution structure and refinement at 1.8 Å resolution. *J. Mol. Biol.* **195**, 373-396.
32. Grise, C., Boucher, R., Thibault, G. & Genest, J. (1981). Formation of angiotensin II by tonin from partially purified human angiotensinogen. *Can. J. Biochem.* **59**, 250-255.
33. Bridon, D.P. & Dowell, B.L. (1995). Structural comparison of prostate-specific antigen and human glandular kallikrein using molecular modeling. *Adult Urol.* **45**, 801-806.
34. Vihinen, M. (1994). Modeling of prostate specific antigen and human glandular kallikrein structures. *Biochem. Biophys. Res. Commun.* **204**, 1251-1256.

35. Hutschinson, E.G. & Thornton, J.M. (1994). A revised set of potentials for β-turn formation in proteins. *Protein Sci.* **3**, 2207-2216.
36. Lamba, D., *et al.* & Bode, W. (1996). The 2.3 Å crystal structure of the catalytic domain of recombinant two chain human tissue-type plasminogen activator. *J. Mol. Biol.* **258**, 117-135.
37. Spraggon, G., *et al.*, & Jones, E.Y. (1995). The crystal structure of the catalytic domain of human urokinase-type plasminogen activator. *Structure* **3**, 681-691.
38. Ke, S.-H., Coombs, G.S., Tachias, K., Navre, M., Corey, D.R. & Madison, E.L. (1997). Distinguishing the specificity of closely related proteases. *J. Biol. Chem.* **272**, 16603-16609.
39. Mikolajczyk, S.D., Millar, L.S., Kumar, A. & Saedi, M.S. (1998). Human glandular kallikrein, hK2, shows arginine-restricted specificity and forms complexes with plasma protease inhibitors. *Prostate* **34**, 44-50.
40. Hedstrom, L., Szilagyi, L. & Rutter, W.J. (1992). Converting trypsin to chymotrypsin the role of surface loops. *Science* **255**, 1249-1253.
41. Venekei, I., Szilagyi, L., Graf, L. & Rutter, W.J. (1996). Attempts to convert chymotrypsin to trypsin. *FEBS Lett.* **379**, 143-147.
42. Denmeade, S.R., Nagy, A., Gao, J., Lilja, H., Schally, A.V. & Isaacs, J.T. (1998). Enzymatic activation of a doxorubicin-peptide prodrug by prostate-specific antigen. *Cancer Res.* **58**, 2537-2540.
43. Jameson, G.W., Roberts, D.V., Adams, R.W., Kyle, S.A. & Ellmore, D.T. (1973). Determination of the operational molarity of solutions of bovine α-chymotrypsin, trypsin, thrombin, and factor Xa by spectrofluorometric titration. *Biochem. J.* **131**, 107-117.
44. Taylor, J.R. (1982). *An Introduction to Error Analysis: The Study of Uncertainties in Physical Measurements.* University Science Books, Mill Valley, CA.
45. Tuffery, P., Etchebest, C., Hazout, C. & Lavery, R. (1991). A new approach to the rapid determination of protein sidechain conformations. *J. Biomol. Struct. Dyn.* **8**, 1267-1289.
46. Brooks, B.R., Bruccoleri, R.E., Olafson, B., D., States, D.J., Swaminathan, S. & Karplus, M. (1983). CHARMM: A program for macromolecular energy, minimization, and molecular graphics. *J. Comput. Chem.* **4**, 187-217.
47. Brünger, A.T. (1992). X-PLOR Manual, Version 3.0. New Haven, Yale University.
48. Powell, M.J.D. (1977). Restart procedures for the conjugate gradient method. *Math. Program.* **12**, 241-254.
49. Brunger, A.T. & Karplus, M. (1988). Polar hydrogen positions in proteins: empirical energy function placement and neutron diffraction comparison *Proteins* **4**, 148-156.
50. Mittl, P.R.E., *et al.* & Grutter, M.G. (1997). A new structural class of serine protease inhibitors revealed by the structure of the hirustatin-kallikrein complex. *Structure* **5**, 253-264.
51. McRee, D.E. (1992). XtalView: a visual protein crystallographic software system for X11/XView *J. Mol. Graph.* **10**, 44-47.
52. Laskowski, R.A., MacArthur, M.W., Moss, D.S. & Thornton, J.M. (1993). PROCHECK: A program to check the stereochemical quality of protein structures. *J. Appl. Crystallogr.* **26**, 283-291.
53. Verlet, L. (1967). Computer experiments on classical fluids. I. Thermodynamical properties of Lennard-Jones molecules. *Phys. Rev.* **159**, 89-105.
54. Berendsen, H.J.C., Postma, J.P.M., Van Gunsteren, W.F., DiNola, A. & Haak, J.R. (1984). Molecular dynamics with coupling to an external bath. *J. Chem. Phys.* **81**, 3684-3690.
55. Ryckaert, J.P., Ciccotti, G. & Berendsen, H. (1977). Numerical integration of cartesian equations of motion of a system with constraints. Molecular dynamics of N-alkanes. *J. Comput. Phys.* **23**, 327-341.
56. Schechter, I. & Berger, A. (1967). On the size of the active site in proteases. I. Papain. *Biochem. Biophys. Res. Commun.* **27**, 157-162.
57. McDonald, I.K. & Thornton, J.M. (1994). Satisfying hydrogen bonding potential in proteins. *J. Mol. Biol.* **238**, 777-793.
58. Wallace, A.C., Laskowski, R.A. & Thornton, J.M. (1995). LIGPLOT: a program to generate schematic diagrams of protein-ligand interactions. *Protein Eng.* **8**, 127-134.
59. Nicholls, A., Sharp, K.A. & Honig, B. (1991). Protein folding and association: Insights from the interfacial and thermodynamic properties of hydrocarbons. *Proteins* **11**, 281-296.
60. Sitkoff, D., Lockhart, D.J., Sharp, K.A. & Honig, B. (1994). Calculation of the electrostatic effects at the amino terminus of a helix. *Biophys. J.* **67**, 2251-2260.

Because *Chemistry & Biology* operates a 'Continuous Publication System' for Research Papers, this paper has been published via the internet before being printed. The paper can be accessed from http://biomednet.com/cbiology/cmb – for further information, see the explanation on the contents pages.