Negative selectivity and the evolution of protease cascades: the specificity of plasmin for peptide and protein substrates

Laurence S Hervio[1], Gary S Coombs[2], Robert C Bergstrom[3], Kavita Trivedi[3], David R Corey[3] and Edwin L Madison[1,2,4]

Background: Understanding the networks of selective proteolysis that regulate complex biological systems requires an appreciation of the molecular mechanisms used to maintain substrate specificity. Human plasmin, a serine protease that promotes the dissolution of blood clots and is essential in maintaining normal hemostasis, is usually described as having broad substrate specificity. Recent evidence that plasmin also plays a key role in a variety of other important biological and pathological processes, however, has suggested that this description might need to be re-evaluated.

Results: We used substrate phage display to elucidate optimal subsite occupancy for substrates of plasmin. We identified a peptide substrate that is cleaved 710,000-fold more efficiently by plasmin than a peptide containing the activation sequence of plasminogen. Plasmin achieves this unexpected, large differential activity even though both target sequences possess an arginine residue in the P1 position. We also demonstrate that proteolysis by plasmin can be targeted to an engineered protein substrate and that introduction of substrate sequences identified by phage display into plasminogen increases plasmin-mediated cleavage of the mutant 2000-fold.

Conclusions: The specificity of plasmin is more tightly controlled than previously recognized; interactions with substrates at all subsites between S4 and S2′ contribute to catalysis. Furthermore, in contrast to most enzymes that exhibit positive selectivity for substrate, the evolution of substrate specificity by plasmin has apparently been dominated by a strong negative selection against development of autoactivation activity. This 'negative selectivity' avoids short-circuiting regulation of the fibrinolytic system and other important biological processes, and might be an important general mechanism for controlling protease cascades.

Addresses: [1]The Scripps Research Institute, Department of Vascular Biology, La Jolla, CA 92037, USA. [2]Corvas International, Department of Molecular Biology, 3030 Science Park Road, San Diego, CA 92121, USA. [3]Departments of Pharmacology and Biochemistry, University of Texas Southwestern Medical Center at Dallas, 5323 Harry Hines Boulevard, Dallas, TX 75390, USA. [4]Department of Vascular Biology, The Torrey Pines Institute for Molecular Studies, 3550 General Atomics Court, San Diego, CA 92121, USA.

Correspondence: David R Corey; Edwin L Madison
E-mail: Corey@howie.swmed.edu;
ed_madison@corvas.com

Key words: plasmin, plasminogen, protease specificity, subsite occupancy, substrate phage display

Received: **4 February 2000**
Revisions requested: **16 March 2000**
Revisions received: **27 March 2000**
Accepted: **5 April 2000**

Published: **30 May 2000**

Chemistry & Biology 2000, **7**:443–453

1074-5521/00/$ – see front matter

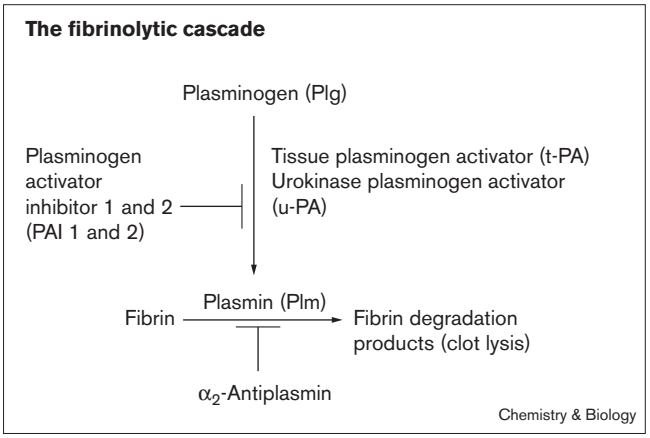
Introduction

The use of selective proteolysis to modulate protein activity has been a recurring theme during the evolution of a variety of complex biological systems such as fibrinolysis, blood coagulation, complement activation and embryonic development. To achieve this regulation, which requires stringent control of proteolysis to avoid indiscriminate, general destruction of proteins, specific proteases have evolved several distinct mechanisms to promote efficient cleavage of target proteases while leaving other proteins intact [1–5]. An important first step towards understanding the biochemistry, evolution and regulation of many complex biological processes, therefore, is identification of the mechanism(s) by which specific proteolysis is accomplished. Detailed knowledge of the molecular basis of these mechanisms can, in turn, provide new insights into how structure regulates function during a particular biological process and how mutation of proteins participating in this process can produce pathological consequences.

The endogenous, human fibrinolytic pathway promotes the dissolution of blood clots and therefore serves as a critical counterbalance *in vivo* to the blood coagulation cascade. Initiation of this pathway is mediated by two closely related serine proteases, tissue-type plasminogen activator (t-PA) and urokinase-type plasminogen activator (u-PA), which cleave the circulating zymogen plasminogen to produce the mature protease plasmin (Figure 1). Plasmin then directly catalyzes clot dissolution by efficiently degrading the fibrin meshwork of thrombi. Appreciation that thrombosis is the focal event in the initiation of acute myocardial infarction, the major cause of morbidity and mortality in many countries, sparked great interest in plasmin and the endogenous fibrinolytic cascade. In fact, administration of plasminogen activators to convert endogenous plasminogen into plasmin has now become the standard of care for heart-attack patients.

The human plasminogen activators t-PA and u-PA cleave plasminogen efficiently but hydrolyze other substrates

Figure 1



Control of fibrinolysis.

poorly [6,7]. These proteases, therefore, are highly specific enzymes in the traditional sense — they have evolved positive selectivity, which enables them to process one substrate while leaving substrates with similar chemical properties intact. Plasmin, by contrast, has been thought to be a relatively nonselective enzyme, like trypsin. Plasmin cleaves fibrin efficiently at multiple sites [8], and plasmin's catalytic activity is ultimately controlled by binding to the serpin α_2-antiplasmin [9].

Recent studies indicate that plasmin plays a key role in a variety of biological processes in addition to normal hemostasis. Mice deficient in plasmin(ogen), exhibit not only extensive thrombosis [10,11] but also severe and generalized defects in wound-healing and tissue repair [12], defects in smooth muscle cell migration and neointima formation following vascular injury [13], and defective recruitment of monocytes during an inflammatory response [14]. Furthermore, plasminogen deficiency appears to cause ligneous conjunctivitis in both mice [15] and humans [16,17] and accelerate the development of vessel-wall disease in mice predisposed to atherosclerosis [18]. Host plasminogen has also been shown to play an important role in the life cycle of *Borrelia burgdorferi*, the causative agent of Lyme disease [19]. Finally, strong evidence has been presented recently that plasmin mediates experimental neuronal degeneration by acting on substrates other than fibrin [20]. Additional insights into the molecular mechanisms that contribute to the substrate specificity of plasmin, therefore, would be expected to increase understanding of all of these important biological and pathological processes.

In this study, we combine substrate phage display [21,22] with detailed kinetic analysis of the cleavage of peptides and engineered protein substrates to elucidate the relationship between the substrate specificity of plasmin and its biological function. We find that non-S1 subsites make a surprisingly large contribution to the substrate selectivity of plasmin; optimal non-S1 subsite occupancy can produce a 710,000-fold increase in the efficiency of substrate catalysis by plasmin. In addition, our findings demonstrate that, in contrast to the evolution of stringent positive selectivity for substrate by t-PA and u-PA, the evolution of plasmin has been dominated by a strong negative selection against the development of autoactivation activity (i.e. activation of plasminogen by plasmin), an activity that would short-circuit the regulation of fibrinolysis and other key biological processes.

Results and discussion

Construction and screening of substrate phage libraries for cleavage by plasmin

A polyvalent fd phage library that displayed random hexapeptide sequences and contained 2×10^8 independent recombinants was prepared [22]. Each member of this library displayed an amino-terminal extension from phage coat protein III (pIII) that contained a randomized region of six amino acids, a six-residue linker sequence (SSGGSG, using single-letter amino acid code), and the epitopes for monocloncal antibody (mAb) 179 and mAb 3-E7. Because plasmin did not digest the pIII sequence, the antibody epitopes or the flexible linker sequence, the loss of antibody epitopes from the phage surface upon incubation with plasmin required cleavage of the randomized peptide insert. Thus, incubation of the library with plasmin, followed by removal of phage retaining the antibody epitopes; accomplished a large enrichment of phage clones whose randomized hexamer sequence could be cleaved by plasmin.

Identification of a consensus sequence for hydrolysis of substrate phage

Following five rounds of selection to enrich and amplify phage which display sequences that are readily cleaved by plasmin, 111 phage clones were identified as plasmin substrates. DNA sequencing of these clones revealed the presence of 69 distinct hexamer sequences among the selected phage (Table 1). As expected from the trypsin-like primary specificity of plasmin, each hexamer contained at least one basic residue. Out of the 69 distinct, randomized hexamers present in the selected substrate phage 15 contained a single basic residue, and, in striking contrast to the other two human fibrinolytic serine proteases, t-PA and u-PA, plasmin did not demonstrate a strong preference for arginine versus lysine. In eight of these 15 hexamers, the single basic residue was arginine, whereas the remaining seven sequences contained lysine. Examination of these 15 hexamers also indicated that plasmin exhibits a strong preference for aromatic residues at the P2 position of substrates. Eleven of the 15 substrate hexamers contained phenylalanine, tyrosine or tryptophan at the P2 position.

This preference for large P2 residues is in striking contrast to the specificity of t-PA and u-PA, which strongly prefer

Table 1

Alignment of the amino acid sequences of the randomized regions of phage clones selected by substrate phage display.

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

Clone #	P5	P4	P3	P2	P1	P1′	P2′	P3′	P4′	P5′	P6′	P7′
36			F	V	K	R	A	R				
37		K	I	G	R	N	T					
38			V	L	R	K	R	A				
39			T	Q	R	R	V	R				
40		R	Y	T	K	S	T					
41				R	R	Y	L	D	R			
42				R	R	D	R	S	L			
43				V	K	M	R	R	D			
44		G	T	S	R	N	K					
45		G	S	L	R	S	R					
46	H	G	K	S	R	P						
47		M	H	V	R	R	R					
48		T	K	L	R	G	R					
49		R	R	V	K	Q	T					
50		G	V	N	K	N	K					
51			W	T	R	R	N	G				
52			T	Q	R	R	L	R				
53	M	F	L	A	K	R						
54			K	S	K	S	R	D				
55		K	R	S	R	N	G					
56		T	D	L	K	S	K					
57		A	H	N	K	R	L					
58		S	K	M	R	M	G					
59		R	E	H	R	V	N					
60				M	R	R	K	R	V			
61				A	K	K	S	K	L			
62		V	A	S	R	S	K					
63		D	L	S	R	L	K					
64			I	N	R	S	R	Q				
65				L	K	R	R	N	L			
66			K	S	K	S	R	D				
67			G	Y	K	N	S	A				
68		I	R	Y	K	G	K					
69		K	R	S	R	S	G					

Peptide sequences have been shifted to the left or right to align the clones to a consensus sequence. The amino- and carboxy-terminal residues flanking these sequences are contributed by the phage constant regions and are GGAG and GGAGSS, respectively.

glycine or alanine at the P2 position. On the basis of the 15 substrate hexamers that contained a single basic residue, and therefore did not present the theoretical possibility of more than one potential cleavage site, a preliminary consensus sequence for optimal subsite occupancy for substrates of plasmin, from positions P3–P1′ emerges as X_1(Y/F)(R/K)X_2, where X_1 and X_2 can be several different amino acids but are most often valine and serine, respectively. The preferences of plasmin at the P3 and P1′ subsites are clearly less stringent than those at the P1 and P2 positions.

On the basis of analysis of selected sequences that contained a single basic residue, we aligned and analyzed all 69 randomized hexamer regions present in the selected substrate phage clones. The predicted P1 residue was arginine in 40 of the substrate sequences and lysine in the remaining 29. 46% of the selected substrate phage contained an aromatic residue at the P2 position, with tyrosine present at this position in 20% of the phage clones, phenylananine in 17% and tryptophan in 9% of the sequences. The P1′ residue of 30% of the selected hexamer sequences was serine, whereas glycine (16 clones) and valine (11 clones) were the most common occupants of the P3 position of the substrate phage. In 11 of 16 clones containing glycine at P3, however, the glycine residue was contributed by fixed flanking sequences in the substrate phage rather than the randomized hexamer region.

On the basis of our experience with several related proteases [6,7,23], the catalytic efficiency of plasmin toward the most and least labile of the selected substrate sequences would be expected to vary by at least three orders of magnitude. Acquisition of the clearest and most accurate information regarding optimal subsite occupancy for substrates of plasmin, therefore, requires identification and examination of the most labile substrate sequences

Table 2

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

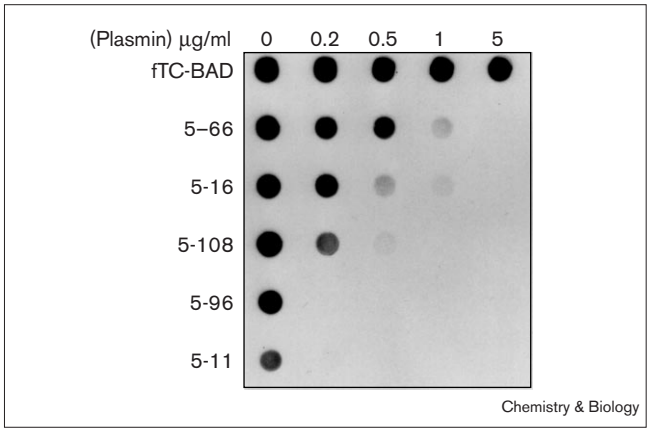
Clone #	P5	P4	P3	P2	P1	P1′	P2′	P3′	P4′	P5
6				Y	K	I	G	G	S	
13		A	V	Y	K	A	G			
23			R	F	K	K	G	K		
27				Y	K	S	R	K	Y	
28			I	Y	R	S	R	S		
29		A	N	F	R	G	K			
30		E	K	F	R	K	K			
38			V	L	R	K	R	A		
39			T	Q	R	R	V	R		
45		G	S	L	R	S	R			
47		M	H	V	R	R	R			
52			T	Q	R	R	L	R		
68		I	R	Y	K	G	K			

Peptide sequences have been shifted to the left or right to align the clones to a consensus sequence. The amino- and carboxy-terminal residues flanking these sequences are contributed by the phage constant regions and are GGAG and GGAGSS, respectively.

among the selected substrate phage clones. Consequently, we performed digestions of all 69 substrate phage, utilizing a series of increasingly stringent conditions, and we analyzed the resulting reaction products using a sensitive dot-blot assay [7] (Figure 2).

Under the most stringent conditions adopted in these experiments, 13 of the 69 selected phage were particularly labile plasmin substrates (Table 2). Eight of these highly labile substrate hexamers contained an arginine residue at P1, while five hexamers contained a P1 lysine residue, again suggesting that plasmin shows little preference between these two basic P1 residues. Eight of the thirteen substrate clones contained either tyrosine (five clones) or phenylalanine (three clones) at P2. These data provide additional support for tyrosine or phenylalanine at P2 and arginine or lysine at P1 as optimal subsite occupancy for peptide substrates of plasmin. Plasmin does not appear to possess stringent specificity at the P3 position of substrates; arginine, valine, glycine, and threonine are all present at the P3 position of two of the 13 clones. At the P1′ position of peptide substrates, plasmin appears to prefer arginine (three clones), lysine (three clones), or serine (three clones). Glycine is found at the P1′ position in two clones. Plasmin also shows clear preferences at the P2′ position of substrates; eleven of the 13 highly labile substrate hexamers contain arginine (five clones), lysine

Figure 2



Dot-blot proteolytic analysis of control and substrate phage clones. Phage from individual clones were prepared, digested with the indicated concentrations of plasmin for 15 min at 37°C, spotted onto a nitrocellulose filter, and detected as described in the Materials and methods section. Loss of positive staining indicates cleavage in the randomized hexamer region of the gene III fusion protein to remove the antibody epitope. Phage fTC-BAD is a negative control phage whose randomized hexamer region does not contain a basic residue. Phage 5-66, 5-16, 5-108, 5-96 and 5-11 are individual phage clones isolated during the fifth round of selection.

(three clones) or glycine (three clones) at this position. On the basis of analysis of the 13 most rapidly cleaved hexamer substrates, therefore, optimal subsite occupancy for plasmin substrates, from P3–P2′, appears to be (V, G, T, R > X) (Y, F) (R, K) ⇓ (R, K, S > G) (R > K, G), where X represents a variety of amino acids and the arrow represents the site of peptide-bond hydrolysis.

Confirmation of specificity by kinetic analysis of peptide substrates

Six peptides containing amino acid sequences present in the most labile substrate phage were chosen for kinetic analysis. In all cases, the predicted site of peptide-bond hydrolysis was confirmed by mass spectral analysis of reaction products. Plasmin cleaved these peptides with catalytic efficiencies (k_{cat}/K_m) varying from 4100 $M^{-1}s^{-1}$ for peptide **1**, to $1.2(10)^6$ $M^{-1}s^{-1}$ for peptide **6** (Table 3). By contrast to the ability of plasmin to efficiently cleave the peptides identified using substrate phage display, plasmin cleaved peptide **7** (modeled after the plasminogen activation sequence) very poorly, exhibiting a k_{cat}/K_m value for this reaction of only 1.7 $M^{-1}s^{-1}$. Thus, there is a 710,000-fold difference between the efficiency of cleavage of the optimal substrate identified by substrate phage display and the plasminogen activation sequence, even though both substrates possess arginine as a P1 residue.

The primary physiological inhibitor of plasmin is α_2-antiplasmin [9], a serpin whose reactive center loop has presumably evolved to match the substrate specificity of

Table 3

Comparison of k_{cat}, K_m, and k_{cat}/K_m for peptide hydrolysis by plasmin.

	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}/K_m relative to **7**
Plasmin-selective peptides from phage display screen					
1	LGGSGTQR↓RLRLE	12	2900	4.1 (10)^3	2400
2	LGGSGANFR↓GKLE	68	2400	2.8 (10)^4	16,000
3	LGGSGYK↓IGGSLE	12	210	5.7 (10)^4	34,000
4	LGGSIRYK↓GKSLE	44	750	5.9 (10)^4	35,000
5	LGGSGAVYK↓AGLE	40	420	9.5 (10)^4	56,000
6	LGGSGIYR↓SRSLE	120	100	1.2 (10)^6	710,000
Activation sequence from plasminogen†					
7	KKSPGR↓VVGGSVAH	0.0086	5100	1.7	1
Reactive center sequence from α 2-antiplasmin					
8	LGGSAMSR↓MSSLE	3.2	1600	2.0 (10)^3	1200
Introduction of residues from peptide **7** into peptide **6**					
9	LGGS**S**IYR↓SRSLE	35	6800	5.1 (10)^3	3000
10	LGGSG**P**YR↓SRSLE	92	1700	5.4 (10)^4	31,000
11	LGGSGI**G**R↓SRSLE	53	1850	2.9 (10)^4	17,000
12	LGGSGIYR↓**V**RSLE	16	470	3.4 (10)^4	20,000
13	LGGSGIYR↓S**V**SLE	39	3700	1.1 (10)^4	6500

*Positional nomenclature of subsite residues. Arrows denote the position of peptide bond hydrolysis. The peptide bond is cleaved between P1 and P1′. Residues in bold are those introduced from peptide **7** into peptide **6**. Standard errors calculated for these measurements varied from 4–30%. †To prevent cyclization of this peptide substrate, serine residues have been substituted for the cysteine residues found in plasminogen at positions P4 and P5′.

plasmin very well. Surprisingly, however, when we assayed peptide substrate **8** (modeled after the reactive site within α_2-antiplasmin) we discovered that it was cleaved less efficiently than all six of peptides (**1**–**6**) identified by substrate phage display. The observation that peptide **6** is cleaved 600 times more efficiently than the α_2-antiplasmin-based peptide **8** indicates the impressive ability of substrate phage technology to elucidate optimal subsite occupancy for proteases that could not have been predicted from examining the sequence of physiological substrates or inhibitors.

The 710,000-fold difference in the catalytic efficiency of plasmin for hydrolysis of peptides **6** and **7** occurs despite the fact that peptides **6** and **7** share the same P1 residue, arginine. The specificity of plasmin, therefore, is governed to an unexpectedly large extent by interactions beyond the S1 specificity pocket. To evaluate the role of non-S1 interactions in defining the specificity of plasmin, we systematically introduced the P4, P3, P2, P1′ and P2′ residues of peptide **7** into **6** to create peptides **9**–**13**. Each of these peptides was a much better substrate for cleavage by plasmin than peptide **7**, indicating that no single interaction accounts for the extremely inefficient cleavage of **7** by plasmin. Furthermore, introduction into peptide **6** of a serine residue at P4, proline at P3, glycine at P2, or valine at P1′ and P2′ reduced k_{cat}/K_m by 240, 22, 43, 35 and 110-fold, respectively, relative to cleavage of **6**. The finding that every alteration reduces catalysis suggests that inefficient catalysis of peptide **7** is due to the combination of suboptimal interactions at each non-S1 subsite.

Site-selective cleavage of an engineered protein by plasmin

Our results with peptide substrates strongly suggested that interactions of the non-S1 subsites are critical for defining the specificity of plasmin. The physiological substrates of plasmin, however, are folded proteins rather than peptides. Consequently, to gain insights into whether the preferences for substrate hydrolysis observed for peptide cleavage would also apply within the framework of a model protein, we engineered the selected hexapeptide sequence from substrate **6** into staphyloccocal nuclease (SNase) [24]. SNase is an ideal model protein substrate for investigating proteolysis because it has been well characterized by extensive structural and mutagenesis studies. SNase contains a loop at residues 43–52 which is accessible to and preferentially cleaved by trypsin and which can be mutated without altering the structure of other regions of the protein. Cleavage of SNase within this surface loop yields two stable products that remain associated noncovalently. The kinetic constants for hydrolysis of SNase variants by trypsin family proteases can therefore be calculated by measuring the time-dependent appearance of the two cleavage products using sodium dodecyl sulfate–polyacrylamide gel electrophoresis (SDS–PAGE) and densitometry.

We engineered four variants of SNase. One of these, **14**, contained the optimal P4–P2′ sequence for cleavage by

Table 4

Kinetic parameters k_{cat}, K_m, and k_{cat}/K_m for the hydrolysis by plasmin of SNase variants containing various target sequences.

	Substrate (P4,P3, P2, P1,↓P1′, P2′,)*	k_{cat} s^{-1}	K_m (μM)	k_{cat}/K_m (M^{-1} s^{-1})	K_{cat}/K_m relative to **14**
14	GIYR↓SR	0.6	5.8	1 (10)^5	1
15	GI**V**R↓SR	0.039	12	3.3 (10)^3	0.03
16	**A**IYR↓SR	0.036	3	1.2 (10)^4	0.12
17	SPGR↓VV	nd	nd	nd	–

*Positional nomenclature of subsite residues. Arrows denote the position of peptide-bond hydrolysis. The peptide bond is cleaved between P1 and P1′. Residues in bold are altered from parent selected sequence **14**. nd, no selective cleavage detected. Standard errors calculated for these measurements varied from 4–30%.

plasmin, GIYRSR. Cleavage of this variant by plasmin was characterized by a k_{cat}/K_m value of ~100 000 $M^{-1}s^{-1}$ (Table 4), a value similar to that previously observed for cleavage of engineered SNase variants by trypsin [24]. Two variants, **15** and **16**, contained altered residues at P2 and P4 and were both less efficient substrates than **14**. Especially striking is the finding that introduction of a single methyl group at the P4 position through a glycine to alanine mutation reduces catalysis of engineered SNase variant **16** by almost tenfold relative to **14**. The preference for cleavage of engineered SNase **14** relative to similar substrates **15** and **16** suggests that the selected sequence from phage display that was used to create SNase **14** is an ideal sequence for designing proteins to be targets for cleavage by plasmin.

We also assayed a SNase variant (**17**) that contained the hexapeptide sequence from the activation site of plasminogen. In contrast to efficient cleavage of substrate **14**, we could not detect any selective cleavage with the 43–52 loop of **17**. As with peptide substrates, therefore, non-P1 subsite occupancy can dramatically modulate the specificity of plasmin toward protein substrates.

Engineering plasminogen for activation by plasmin: removing a constraint on protease cascades

The extremely low activity of plasmin towards peptides containing the plasminogen activation cleavage-site sequence suggested that plasmin would exhibit poor autoactivation activity (i.e. be unable to convert plasminogen into plasmin). To test this hypothesis, we examined cleavage of microplasminogen [25,26], a readily expressed variant of plasminogen, by catalytic amounts of plasmin. We constructed variants of microplasminogen with the active-site serine residue replaced by alanine so that any plasmin formed during the reaction would be catalytically inactive. This was an important consideration because if active plasmin was formed, it would increase the concentration of enzyme and prevent kinetic analysis of cleavage. As expected, we observed that the activity of wild-type (wt) plasmin towards microplasminogen (micro-plg) **18** containing the wt activation sequence was low. The k_{cat}/K_m value measured for this autoactivation reaction was ~340 $M^{-1}s^{-1}$ (Table 5).

The plasminogen cleavage site is located within a nine-residue, disulfide-bonded surface loop in native plasminogen, and constraints imposed upon the conformation of the cleavage sequence by this location might contribute to the poor autoactivation activity of plasmin. To investigate this possibility, we constructed a variant of micro-plg in which these two cysteine residues (at positions 558 and 566 in the plasminogen numbering system) had been replaced by serine residues. This variant, micro-plg **19**, also carried the Ser741→Ala mutation to facilitate kinetic analysis of its activation by wt plasmin. The catalytic efficiency of plasmin toward micro-plg **19** was ~1.1×10^3 $M^{-1}s^{-1}$, which is only about threefold higher than the enzyme's activity toward wt micro-plg. Any constraints imposed by the Cys558–Cys566 disulfide bond, therefore, have little influence on the autoactivation reaction.

The kinetic analysis of the activation of micro-plgs **18** and **19** described above indicated that the wt plasminogen activation cleavage sequence, whether present in the zymogen in a cyclic (micro-plg **18**) or 'linear' configuration (micro-plg **19**), is a poor substrate for plasmin. We therefore constructed two additional variants of micro-plg to test the hypothesis that specific subsite occupancy that enhanced the lability of peptide substrates toward plasmin could also enhance the lability of plasminogen toward plasmin. In the variant micro-plg **20**, the P3, P2, P1′, and P2′ residues of the cyclic activation site of micro-plg **18** are replaced with the corresponding residues present in the most labile, selected peptide (**6**). By contrast to wt micro-plg, this new micro-plg variant was efficiently cleaved by plasmin, with a k_{cat}/K_m value of ~6.8×10^5 $M^{-1}s^{-1}$, increasing the 'autoactivation activity' of plasmin by a factor of ~2000. Similar results were obtained with micro-plg variants containing the activation sequence in a linear configuration. Micro-plg **21**, which contains the optimal P4–P2′ sequence identified by substrate phage display, was cleaved by plasmin with a catalytic efficiency of ~7.1×10^5 $M^{-1}s^{-1}$.

Table 5

Kinetic parameters k_{cat}, K_m and k_{cat}/K_m for the hydrolysis by plasmin of microplasminogen variants containing various target sequences.

	Substrate (P4, P3, P2, P1,↓P1′, P2′,)*	k_{cat} s^{-1}	K_m (μM)	k_{cat}/K_m (M^{-1} s^{-1})	K_{cat}/K_m relative to **18**
18	CPGR↓VV	0.0016	4.7	$3.4(10)^2$	1.0
19	**S**PGR↓VV†	0.039	36.0	$1.1(10)^3$	3.2
20	C**IYR↓SR**	2.7	4.0	$6.8(10)^5$	2000
21	**GIYR↓SR**†	1.5	2.0	$7.1(10)^5$	2100

*Positional nomenclature of subsite residues. Arrows denote the position of peptide-bond hydrolysis. The peptide bond is cleaved between P1 and P1′. †P5′ (not shown) is mutated from cysteine to serine. In all four substrates, the catalytic serine (Ser741) has been mutated to alanine. Residues in bold are altered from wt plasminogen sequence **18**. Standard errors calculated for these measurements varied from 4% to 30%.

Interplay of substrate specificity and *in vivo* physiology for plasmin

This study reveals that the sequence (V, G, T, R > X) (Y, F) (R, K) ⇓ (R, S, K > G) (R > K, G) (where X indicates a variety of amino acids) represents optimal subsite occupancy, from P3–P2′, for peptide substrates of plasmin. Optimal subsite occupancy is similar for peptide and protein substrates of plasmin. The substrate specificity of plasmin is both less stringent and significantly different than that of the most closely related homologs, the plasminogen activators t-PA and u-PA. This divergence is most obvious at the P2 position, where plasmin strongly prefers tyrosine or phenylalanine and t-PA and u-PA strongly prefer glycine or alanine. The structural basis of this distinction is almost certainly the six-residue deletion, compared with trypsin and chymotrypsin, within the 90-loop of plasmin. This deletion, which eliminates the conserved '99-β-hairpin' present in most (chymo)trypsin-like proteases and creates the '94-shunt', substantially opens and enlarges the P2 pocket of plasmin compared with t-PA or u-PA [27,28] (Figure 3). This deletion also exposes the catalytic triad to solvent, which would be expected to significantly weaken the hydrogen-bonding network formed by these three residues and therefore markedly reduce catalysis by decreasing the nucleophilicity of the active-site serine. Proper placement of a large P2 residue in the plasmin–substrate complex, however, could eliminate this barrier to efficient catalysis by shielding the plasmin catalytic triad from solvent.

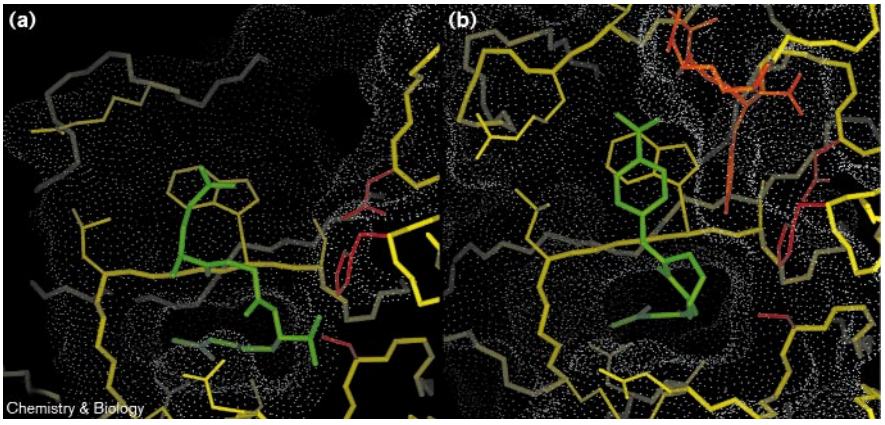
Textbook models of serine protease specificity often emphasize the role of the S1, or 'primary', specificity pocket in determining substrate specificity. Our data for plasmin, however, reveal that occupancy of the surrounding non-S1 subsites can modulate the activity of plasmin towards peptide substrates by approximately six orders of magnitude, making them critical determinants for substrate selectivity. This property of plasmin is surprising. Like trypsin, plasmin is a very active protease, and the specificity of highly active, trypsin-like enzymes has been generally considered to be dictated predominantly by the P1 residue.

Negative selectivity: a mechanism to avoid short-circuiting regulation of protease cascades

This extreme dependence on non-P1 subsite occupancy for efficient catalysis is physiologically important because

Figure 3

View of the active site region of **(a)** plasmin [28] and **(b)** t-PA [31]. Residues forming the catalytic triad (Asp102, His57 and Ser195) are colored red. Small-molecule inhibitors [Glu–Gly–Arg–chloromethyl ketone for plasmin and 2,7-*bis*-(4-amidinobenzylidene)-cycloheptan-1-one for t-PA] are shown in green. The additional six residues in the 90-loop of t-PA, compared with plasmin, are shown in orange. The shorter 90-loop in plasmin exposes the catalytic triad to solvent and substantially enlarges the S2 pocket.



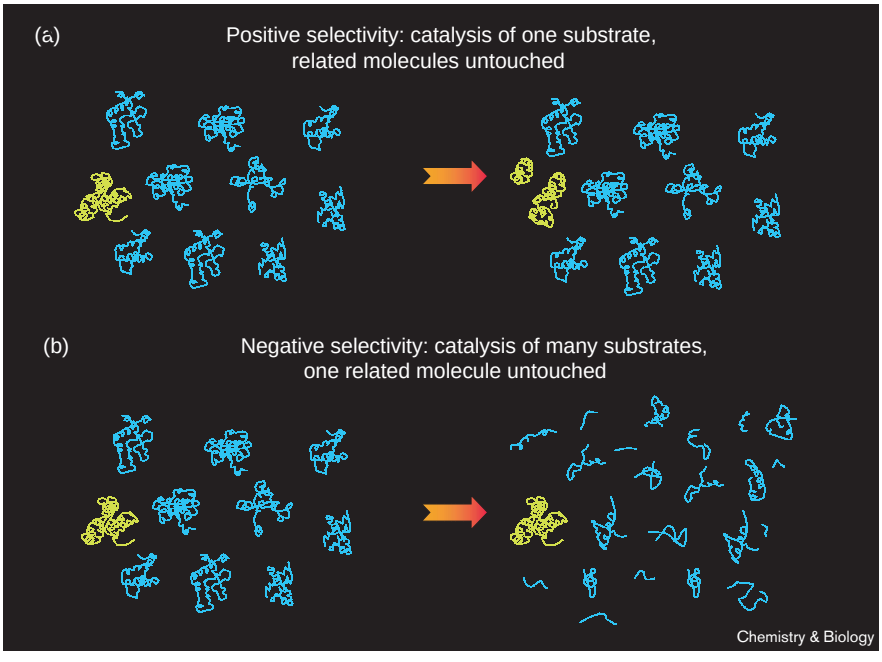
it accounts for the poor recognition of the plasminogen cleavage site by plasmin. Presumably, this poor recognition occurs because plasmin has been subjected to a very strong, negative selection against autoactivation activity during its evolution from a trypsin-like progenitor. This stringent negative selection might have driven both the development of a catalytic machinery that is extraordinarily responsive to suboptimal non-P1 subsite occupancy, and the creation of an activation cleavage sequence in the zymogen plasminogen that contains suboptimal non-P1 subsite occupancy for the mature enzyme plasmin. Strong negative selection against autoactivation activity might be especially critical for proteases such as plasmin that participate in proteolytic cascades.

As all three proteases that participate in the fibrinolytic system, t-PA, u-PA and plasmin, recognize the basic residues arginine or lysine at P1, the creation of an activation cleavage sequence in plasminogen that contained a nonbasic P1 residue would prevent not only autoactivation of plasmin(ogen) but also the activation of plasminogen by the physiological plasminogen activators t-PA and u-PA. Consequently, the evolution of plasmin(ogen) to utilize exclusively suboptimal non-P1 subsite occupancy to prevent autoactivation, and therefore maintain a basic P1 residue in the activation cleavage site, appears to have played an important role in preserving the integrity of the fibrinolytic system.

Strong selection against cleavage of a particular sequence (i.e. negative selection) contrasts sharply with the positive selection that normally creates the restricted substrate specificity of enzymes (Figure 4). In most cases, evolution leads to optimal catalysis of a specific substrate relative to chemically related molecules. For plasmin, however, the opposite has been true. The evolution of specificity has been driven by the necessity not to cleave the plasminogen activation sequence while retaining high activity against many other substrates. It would not be surprising if particular enzymes that participate in other proteolytic cascades had been subjected to similar selective pressures.

These observations might also provide additional insight into the evolution of substrate specificity by t-PA and u-PA. Like plasmin, t-PA and u-PA exhibit poor activity towards peptides containing the plasminogen activation cleavage-site sequence. This low activity appeared paradoxical because the plasminogen cleavage site is the physiological target sequence for t-PA and u-PA. Moreover, we used substrate phage display to isolate peptides that were efficiently cleaved by both t-PA and u-PA, which eliminated the possibility that this low activity against plasminogen-derived peptides was due to an intrinsic low activity towards peptides. Based on the current study and the high homology between t-PA, u-PA and plasmin, we speculate that the poor activity of t-PA and u-PA towards plasminogen-derived peptides might reflect the strong negative selection against cleavage of this sequence by plasmin. Consequently, the development of high activity towards plasminogen by t-PA and u-PA has relied on the evolution of productive, secondary enzyme–substrate interactions and/or the participation of macromolecular cofactor(s) to overcome suboptimal interactions surrounding the scissile bond.

Figure 4



Positive-versus-negative selectivity. **(a)** Proteases that exhibit positive selectivity for substrate cleave a target protein while leaving other proteins intact. **(b)** Plasmin, which exhibits negative selectivity, cleaves a wide range of proteins but leaves plasminogen intact.

Significance

Identification of the genes that constitute the human genome will almost certainly lead to a greater understanding of biological processes as interdependent systems. A prerequisite for this understanding is appreciation of the sophisticated regulatory mechanisms that control enzyme activities. One effective strategy of enzyme regulation that is frequently encountered in complex biological systems is the use of protease cascades for efficient signal transduction. In this study, we investigated how the human fibrinolytic cascade balances the competing demands of stringent specificity and highly efficient proteolysis.

We used substrate phage display to define the optimal selectivity of plasmin at subsites P3 to P2′ to be (V, G, T, R > X) (Y, F) (R, K) ⇓ (R, S > K, G) (R, K, G) for cleavage of peptide substrates. Knowledge of these preferences will facilitate the design of selective inhibitors of plasmin and might aid prediction of plasmin-sensitive sites within potential physiological substrates. These data also demonstrate that the substrate specificity of plasmin is strikingly different from that of tissue-type plasminogen activator (t-PA) and urokinase-type plasminogen activator (u-PA), the other two serine proteases that participate in the fibrinolytic cascade, clearly reflecting the distinct biological roles of these enzymes.

An important finding of this study is that plasmin cleaves an optimized substrate 710,000-fold more efficiently than a substrate containing the sequence that is cleaved to convert plasminogen into plasmin. This 710,000-fold difference in catalytic activity is achieved even though both target sequences contain the same P1 residue (i.e. arginine), demonstrating that the substrate specificity of plasmin exhibits an unexpectedly large dependence upon interactions at non-S1 subsites. We conclude that plasmin has evolved to be a highly active protease that uses non-S1 interactions to discriminate stringently against cleavage of its own activation sequence. This 'negative selectivity' against autoactivation prevents short-circuiting of the fibrinolytic cascade and might represent a general strategy for maintaining the integrity of protease cascades in complex biological systems.

Materials and methods

Reagents

Competent MC1061 (F^-) *Escherichia coli* and nitrocellulose were purchased from Bio-Rad Laboratories, (Hercules, CA). Pansorbin (Protein A-bearing *Staphyloccus aureus*) cells were obtained from Calbiochem (San Diego, CA). K91 (F^+) and MC1061 (F^-) strains of *E. coli* were provided by Steve Cwirla (Affymax, Santa Clara, CA). MAb 3-E7 was purchased from Gramsch Laboratories (Schwabhausen, Germany). Plasmin was purchased from American Diagnostica (Greenwich, CT, USA). The substrate phage library fTC-LIB-N6 was generously provided by Marc Navre (Affymax).

Substrate phage display

We and others have previously described protocols for panning poly valent substrate phage [6,7,21,22]. For plasmin, selection of substrate phage was performed as follows: ~2×10^{10} phage (fTC-LIB-N6 library) were digested in a plasmin-compatible assay buffer (50 mM Tris-HCl (pH 7.4),100 mM NaCl, 1 mM EDTA, 0.01% Tween 20) for between 30 min and 2 h at 37°C using 1–5 μg/ml of plasmin. The total volume of each digest was 250 μl. Digestion of the phage by plasmin was terminated by the addition of 20 μg/ml of aprotinin. Reaction mixtures were then placed on ice, bovine serum albumin (0.1%) and mAb 3-E7 (10 μg) were added, and incubation was continued for 30 min. Following addition of 100 μl of pansorbin cells, the reaction mixtures were rocked at 4°C for 1 h. Samples were centrifuged for 2 min, the supernatants were recovered and the pansorbin adsorption was repeated. A small aliquot of the phage present in the final supernatant was used for titering, and the remaining phage were amplified overnight in K91 cells. Individual clones were selected from the titer plates and grown overnight in 2 ml LB media for dot-blot analysis.

Dot-blot assay

Phage were precipitated from 100 μl of culture supernatant by adding 20 μl of 20% polyethylene glycol, 2.5 M NaCl, incubating the mixture for 30 min on ice, and spinning the mixture for 5 min in a microcentrifuge. The resulting phage pellets were resuspended in 10 μl of TBS (50 mM Tris-HCl (pH 7.4), 150 mM NaCl) and transferred into individual wells of a 96-well microtiter plate. 90 μl of buffer containing varying concentrations of plasmin was then added to each well and the plate was incubated between 15 min and 2 h at 37°C. Following this incubation, 50 μl of the reaction mixture was added to 50 μl of TBS buffer containing 20 μg/ml of aprotinin. These samples were spotted onto a nitrocellulose filter using a dot-blot apparatus (BioRad). The nitrocellulose filter was blocked with 7% nonfat milk in TBS-T (TBS + 0.05% Tween 20) for 1 h, washed three times with TBS-T, incubated for 1 h with mAb 3-E7 at 1 μg/ml, washed several times, and incubated 1 h with a 1:5000 dilution of a goat anti-mouse IgG horseradish peroxidase conjugate (Amersham Pharmacia Biotech AB, Uppsala, Sweden). Finally, the filter was washed several times with TBS-T and stained using the Amersham Western Enhanced Chemiluminescent kit.

Kinetics of peptide cleavage

Peptides were synthesized and purified as described [6,7]. 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 [6,7]. 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. Both parent peptides and cleavage products were collected and characterized by mass spectral analysis using a Voyager MALDI–TOF spectrometer (PerSeptive Biosystems, Foster City, CA). Data were interpreted by Eadie–Hofstee analysis.

Kinetics of Staphylococcal nuclease cleavage

Potential protein substrates were incubated with plasmin for various time periods at 37°C. Substrate protein concentrations varied from 1–20 μM. For kinetic assays, plasmin concentrations were 50 nM, and five to nine substrate concentrations were used within the range listed above. Proteolytic digests were terminated between 10% and 20% completion by addition of loading buffer and heating to 100°C for 10–20 min. Samples of each digest were separated on 15% SDS-polyacrylamide gels and stained with Coomassie brilliant blue. After destaining, substrate to proteolytic product ratios were determined by densitometric scanning on a model 300A scanning densitometer (Molecular Dynamics, Sunnyvale, CA, USA) operating with Image Quant 3.0 software. The ratios obtained were used to determine initial velocities of cleavage. The site of cleavage was confirmed by amino-terminal amino acid sequencing of the carboxyl-terminal proteolytic product. Values for V_{max} and K_m were derived from Eadie–Hofstee analysis.

Expression and purification of recombinant plasminogen

Pichia pastoris strain SMD1168 and the expression plasmid pPic9KX (pPic9K with an *Xho*I site in the kanamycin-resistance gene deleted by silent mutagenesis) were generously supplied by Elizabeth Komives (University of California, San Diego, CA). A truncated form of the plasminogen cDNA encoding residues 546–791 which contain only the protease domain and enough of the activation peptide to allow formation of the interdomain C548–C666 disulfide bond was fused directly to coding sequence for an amino-terminal six-residue histidine tag by polymerase chain reaction (PCR) amplification. This construct was then ligated into the *Xho*I and *Not*I unique restriction sites in the polylinker of vector pPic9KX so that the histidine tag is directly downstream of, and in frame with, the AAAAGA sequence that terminates the coding region for the matα leader sequence responsible for secretion of each fusion protein, and encodes a KEX2 protease cleavage site.

Mutants were constructed by the PCR gene splicing by overlap extension (SOE) technique [29] using this vector as template. These plasmids were then linearized at *Sac*I or *Bgl*II restriction sites and transformed into *P. pastoris* by electroporation [30]. Expressing clones were screened for by batching 0.5 ml induced media onto 20 µl Ni^{2+} resin, eluting in 0.5 M Imidazole, 50 mM sodium acetate, pH 7.5, 50 mM NaCl and running eluents on 10% polyacrylamide gels. Proteins were expressed by growing each expressing clone in 1l of BMGY media (1.0% yeast extract, 2.0% peptone, 100 mM potassium phosphate buffer pH 6.0, 2% glycerol, 3.4 g yeast nitrogen base, 10 g $(NH_4)_2SO_4$, and 0.4 µg/ml biotin) for 48 h, starting with a fresh 20 ml inoculum grown in 1% yeast extract, 2% peptone, 2% D-glucose (YPD) for 24 h. Cultures were induced by spinning down cells 10 min at 3000 rpm, discarding spent BMGY and resuspending cells in 200 ml of BMMY (as BMGY, except that 2% glycerol is replaced with 0.5% methanol). After 24 h in BMMY, cultures were harvested and plasminogens were purified using Ni^{2+} affinity chromatography. Purified protein was aliquotted and stored at −80°C. Protein stock concentrations were determined by BCA assay using a kit supplied by Pierce (Rockford, IL).

Kinetics of plasminogen activation

Concentrations of each plasminogen variant ranging from 1 µM to 35 µM were assayed kinetically for rates of activation by 1 nM to 100 nM plasmin. Because each of the mutants assayed lacks its catalytic serine residue, its cleavage by wt plasmin was monitored by densitometric scanning of substrate and product bands in Coommassie-stained polyacrylamide gels. Quantitation of band density was accomplished using the shareware program NIH Image 1.61. Velocities were determined from the percentages of cleavage in each reaction. Kinetic constants were obtained from Eadie-Hofstee plots. Values reported are an average of two to three assays.

Acknowledgements

This work was supported in part by National Institute of Heath Grants RO1 HL52475 to E.L.M and by a grant from the Robert A. Welch Foundation (I-1244) to D.R.C. L.S.H. received supplemental support from the Philippe Foundation, Inc.

References

1. Neurath, H. (1984). Evolution of proteolytic enzymes. *Science* **224**, 350-356.
2. Khan, A.R. & James, M.N. (1998). Molecular mechanisms for the conversion of zymogens to active proteolytic enzymes. *Protein Sci.* **7**, 815-836.
3. Ruf, W. & Dickinson, C.D. (1998). Allosteric regulation of the cofactor dependent serine protease coagulation factor VIIa. *Trends Cadiovasc. Med.* **8**, 350-356.
4. Jing, H., *et al.*, & Narayan, S.V.L. (1998). Structures of native and complexed complement factor-D – implications of the atypical His57 conformation and self-inhibitory loop in the regulation of specific serine protease activity. *J. Mol. Biol.* **282**, 1061-1081.
5. Perona, J.P. & Craik, C.S. (1995). Structural basis of substrate specificity in the serine proteases. *Protein Sci.* **4**, 337-360.
6. 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.
7. 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.
8. Takagi, T. & Doolittle, R.F. (1975). Amino acid sequence studies on plasmin-derived fragments of human fibrinogen: amino-terminal sequences of intermediate and terminal fragments. *Biochemistry* **14,** 940-946.
9. Potempa, J., Shieh, B.H. & Travis, J. (1988). Alpha-2-antiplasmin: a serpin with two separate but overlapping reactive sites. *Science* **241**, 699-700.
10. Bugge, T.H., Flick, M.J., Daugherty, C.C. & Degen, J.L. (1995). Plasminogen deficiency causes severe thrombosis but is compatible with development and reproduction. *Genes Dev.* **9**, 794-807.
11. Ploplis, V.A., *et al.*, & Collen, D. (1995). Effects of disruption of the plasminogen gene on thrombosis, growth, and health in mice. *Circ. Res.* **92**, 2585-2593.
12. Romer, J., Bugge, T., Pyke, C., Lund, L., Flick, M., Degen, J. & Dano, K. (1996). Impaired wound healing in mice with a disrupted plasminogen gene. *Nat. Med.* **2**, 287-292.
13. Lijnen, R.H., Van Hoef, B., Lupu, F., Moons, L., Carmeliet, P. & Collen, D. (1998). Function of the plasminogen/plasmin and matrix metalloproteinase systems after vascular injury in mice with targeted inactivation of fibrinolytic system genes. *Arterioscler. Thromb. Vasc. Biol.* **18,** 1035-1045.
14. Ploplis, V.A., French, E.L., Carmeliet, P., Collen, D. & Plow, E.F. (1998). Plasminogen deficiency differentially affects recruitment of inflammatory cell populations in mice. *Blood* **91**, 2005-2009.
15. Drew, A.F., *et al*, & Bugge, T.H. (1998). Ligneous conjunctivitis in plasminogen-deficient mice. *Blood* **91**, 1616-1624.
16. Mingers, A., Heimburger, N., Zeitler, P., Kreth, H. & Schuster, V. (1997). Homozygous type 1 plasminogen deficiency. *Semin. Thromb.Hemost.* **23**, 259-269.
17. Schuster, V., Mingers A., Seidenspinner, S., Nussgens, Z., Pukrop, T. & Kreth, H. (1997). Homozygous mutations in the plasminogen gene of two unrelated girls with ligneous conjunctivitis. *Blood* **90**, 958-966.
18. Xiao, Q., *et al.*, & Degen, J.L. (1997). Plasminogen deficiency accelerates vessel wall disease in mice predisposed to atherosclerosis. *Proc. Nat. Acad. Sci.* **94,** 10335-10340.
19. Coleman, J.L., Gebbia, J.A., Piesman, J., Degen, J.L., Bugge, T.H. & Benach, J.L. (1997). Plasminogen is required for efficient dissemination of *B. burgdorferi* in ticks and for enhancement of spirochetemia in mice. *Cell* **89**, 1111-1119.
20. Tsirka, S.E., Bugge, T.H., Degen, J.L. & Strickland, S. (1997). Neuronal death in the central nervous system demonstrates a non-fibrin substrate for plasmin. *Proc. Nat. Acad. Sci. USA* **94**, 9779-9781.
21. Mathews, D.J. & Wells, J.A. (1993). Substrate phage: selection of protease substrates by monovalent phage display. *Science* **260**, 1113-1117.
22. 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.
23. Coombs, G.S., *et al.*, & Corey, D. R. (1998). Substrate specificity of prostate-specific antigen (PSA). *Chem. Biol.* **5**, 475-488.
24. Coombs, G.S., Bergstrom, R.C., Madison, E.L. & Corey, D.R. (1998). Directing sequence-specific proteolysis to new targets. The influence of loop size and target sequence on sequence-selective proteolysis by tissue-type plasminogen activator and urokinase-type plasminogen activator. *J. Biol. Chem.* **273**, 4323-4328.
25. Shi G.Y. & Wu H.L. (1988). Isolation and characterization of microplasminogen. a low molecular weight form of plasminogen. *J. Biol. Chem.* **263**, 17071-17075.
26. Wu H.L., Shi G.Y., Wohl R.C. & Bender M.L. (1987). Structure and formation of microplasmin. *Proc. Natl. Acad. of Sci. USA* **84**, 8793-8799.
27. Wang, X., Lin, X., Loy, J., Tang, J. & Zhang, X. (1998). Crystal structure of the catalytic domain of human plasmin complexed with streptokinase. *Science* **281**, 1662-1665.
28. Parry, M., *et al.*, & Bode, W. (1998). The ternary microplasmin–staphylokinase–microplasmin complex is a proteinase-cofactor-substrate complex in action *Nat. Struct. Biol.* **5**, 917-923.
29. Higuchi, R., Krummel, B. & Saiki, R.K. (1988). A general method of *in vitro* preparation and specific mutagenesis of DNA fragments: study of protein and DNA interactions. *Nucleic Acids Res.* **16**, 7351-7367.
30. Fairlie, W.D., Russell, P.K., Zhang, H.P. & Breit, S.N. (1999). Screening procedure for *Pichia pastoris* clones containing multiple copy gene inserts. *BioTechniques* **26**, 1042-1044.