High Throughput Substrate Specificity Profiling of Serine and Cysteine Proteases Using Solution-phase Fluorogenic Peptide Microarrays*

Dhaval N. Gosalia‡§, Cleo M. Salisbury¶§, Jonathan A. Ellman¶‖,
and Scott L. Diamond‡**‖

Proteases regulate numerous biological processes with a degree of specificity often dictated by the amino acid sequence of the substrate cleavage site. To map protease/substrate interactions, a 722-member library of fluorogenic protease substrates of the general format Ac-Ala-*X*-*X*-(Arg/Lys)-coumarin was synthesized (*X* = all natural amino acids except cysteine) and microarrayed with fluorescent calibration standards in glycerol nanodroplets on glass slides. Specificities of 13 serine proteases (activated protein C, plasma kallikrein, factor VIIa, factor IXaβ, factor XIa and factor α XIIa, activated complement C1s, C1r, and D, tryptase, trypsin, subtilisin Carlsberg, and cathepsin G) and 11 papain-like cysteine proteases (cathepsin B, H, K, L, S, and V, rhodesain, papain, chymopapain, ficin, and stem bromelain) were obtained from 103,968 separate microarray fluorogenic reactions (722 substrates × 24 different proteases × 6 replicates). This is the first comprehensive study to report the substrate specificity of rhodesain, a papain-like cysteine protease expressed by *Trypanasoma brucei rhodesiense*, a parasitic protozoa responsible for causing sleeping sickness. Rhodesain displayed a strong P_2 preference for Leu, Val, Phe, and Tyr in both the P_1 = Lys and Arg libraries. Solution-phase microarrays facilitate protease/substrate specificity profiling in a rapid manner with minimal peptide library or enzyme usage. *Molecular & Cellular Proteomics 4:626–636, 2005.*

Because of their critical roles in biological pathways like hormone activation, proteasomal degradation, and apoptosis, proteases are essential for cellular function and viability. Proteases regulate hormonal activation, cellular homeostasis, apoptosis, and coagulation and play an important role in the pathogenicity and progression of many diseases (1). Proteases comprise one of the largest protein families in organisms from *Escherichia coli* to humans (2–4). Improved understanding of proteases will provide insight into biological systems and will likely provide a number of important new therapeutic targets (1).

To properly function, proteases must preferentially cleave their target substrates in the presence of other proteins. While many factors impact protease substrate selection, one of the key aspects is the complementary nature of the enzyme-active site with the residues surrounding the cleaved bond in the substrate. As such, determination of the residues that comprise the preferred cleavage site of a protease provides critical information regarding substrate selection. Furthermore, determination of substrate specificity also provides a framework for the design of potent and selective inhibitors.

Here we exploit solution-phase substrate nanodroplet microarrays (5), in which fluorogenic substrates suspended in glycerol droplets are treated with aerosolized aqueous enzyme solutions, to provide protease substrate specificity profiles (6). These arrays allow high throughput characterization of the preferred residues on the P side (7) of the substrate in a highly parallel and miniaturized format. We report the use of these arrays here to map the substrate specificity of 24 serine and cysteine proteases in a rapid and efficient manner.

EXPERIMENTAL PROCEDURES

Materials—Purified human activated protein C (APC),[1] human plasma kallikrein, human factor VIIa, human factor IXa β, human factor XIa, and human factor α XIIa were purchased from Enzyme Research Laboratories (South Bend, IN). Human two-chain activated complement C1r, human two-chain activated complement C1s, human complement factor D, human tryptase (lung), bovine trypsin (pancreas, high-purity endotoxin-free), subtilisin Carlsberg (*Bacillus licheniformis*), human cathepsin G (neutrophil), human cathepsin B (liver), human cathepsin H (liver), human cathepsin K (recombinant, *E. coli*), human cathepsin L (liver), human cathepsin S (spleen), human

---

From the ‡Department of Bioengineering, Institute for Medicine and Engineering, University of Pennsylvania, Philadelphia, Pennsylvania 19104; ¶Center for New Directions in Organic Synthesis, Department of Chemistry, University of California, Berkeley, California 94720; and **Department of Chemical and Biomolecular Engineering, Institute for Medicine and Engineering, University of Pennsylvania, Philadelphia, Pennsylvania 19104

Received, January 5, 2005, and in revised form, February 10, 2005
Published, MCP Papers in Press, February 10, 2005, DOI 10.1074/mcp.M500004-MCP200

[1] The abbreviations used are: APC, activated protein C; ACC, 7-amino-4-carbamoylmethylcoumarin; FRET, fluorescence resonance energy transfer; PS-SCL, positional scanning-synthetic combinatorial library.

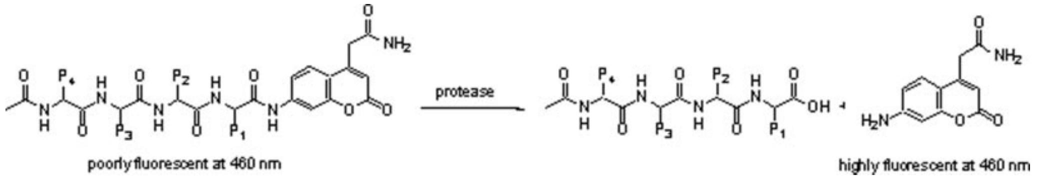

FIG. 1. **Overall design of fluorogenic substrates.**

cathepsin V (recombinant, NSO cells), papain (carica papaya), and stem bromelain (pineapple) were purchased from Calbiochem (La Jolla, CA). Chymopapain (papaya latex) and ficain (fig tree latex) were purchased from Sigma Aldrich (St. Louis, MO). The rhodesain, a gift from the L. S. Brinen laboratory at University of California, San Francisco, was expressed from *Pichia pastoris* as previously reported (8). Lipidated recombinant human tissue factor was purchased from American Diagnostica (Stamford, CT). All enzymes were stored according to manufacturers' instructions.

Methods—The fluorogenic substrate library was synthesized and printed according to protocols previously described (5, 6, 9). The P_1 = Arg and P_1 = Lys sublibraries, along with calibration standards (unacylated 7-amino-4-carbamoylmethylcoumarin (ACC), acetyl-capped ACC, and blanks), were printed at either 50 or 100 μM in a 16 × 24 format equivalent to a 384-well plate on polylysine-coated glass slides (Erie Scientific, Portsmouth, NH) using a 1 × 1 pin (Telechem, Sunnyvale, CA) protocol on an OmniGrid Accent (Gene Machines, San Carlos, CA) microarrayer. Calibration standards were printed on each array, to enable quantification and normalization of fluorescence intensity between slides.

The proteases were reconstituted and diluted in buffers recommended by the manufacturers. The enzyme solutions were delivered to the microarrays at the following concentrations: 10 μM APC in 20 mM Tris-HCl, 100 mM NaCl (pH 7.4); 1 μM plasma kallikrein in 4 mM NaOAc-HCl, 150 mM NaCl (pH 5.3); 5 μM factor VIIa in 20 mM Tris-HCl, 100 mM NaCl, 1 mM $CaCl_2$ (pH 7.4), and 15 nM lipidated recombinant human tissue factor; 10 μM factor IXa β in 20 mM Tris-HCl, 100 mM NaCl (pH 7.4); 100 nM factor XIa in 4 mM NaOAc-HCl, 150 mM NaCl (pH 5.3); 10 μM factor α XIIa in 4 mM NaOAc-HCl, 150 mM NaCl (pH 5.3); 10 μM two-chain activated complement C1r in 47 mM NaH_2PO_4 buffer, 5 mM EDTA, 1 mM benzamidine (pH 7.4); 10 μM two-chain activated complement C1s in 47 mM NaH_2PO_4 buffer, 5 mM EDTA, 1 mM benzamidine (pH 7.4); 4 μM complement factor D in 15 mM NaH_2PO_4, 135 mM NaCl (pH 7.2); 1 μM tryptase in 50 mM NaOAc, 1 M NaCl, 50 μM heparin, 0.01% NaN_3 (pH 5.0); 1 μM trypsin in a 0.9% NaCl aqueous solution; 1 μM subtilisin Carlsberg in water; 10 μM cathepsin G in 50 mM NaOAc, 150 mM NaCl (pH 5.5); 10 μM cathepsin B in 20 mM NaOAc buffer, 1 mM EDTA (pH 5.0); 5 μM cathepsin H in 100 mM Na_2HPO_4, 2 mM EDTA, 2 mM DTT (pH 6.0); 500 nM cathepsin K in 50 mM NaOAc, 100 mM NaCl, 2.5 mM EDTA, 2.5 mM DTT (pH 5.5); 5 μM cathepsin L in 20 mM malonate buffer, 400 mM NaCl, 1 mM EDTA (pH 5.5); 5 μM cathepsin S in 100 mM Na_2HPO_4, 2 mM EDTA, 2 mM DTT, 0.4% Triton X-100 (pH 6.0); 500 nM cathepsin V in 50 mM NaOAc, 100 mM NaCl, 2.5 mM EDTA, 2.5 mM DTT (pH 5.5); 5 μM rhodesain in 100 mM NaOAc, 2.5 mM DTT (pH 5.5); 10 μM papain in PBS (pH 6.0); 10 μM chymopapain in 50 mM NaOAc, 2.5 mM cysteine, 0.5 mM EDTA (pH 6.2); 10 μM ficain in 1 M KH_2PO_4 buffer (pH 7.0) and 10 μM stem bromelain in PBS (pH 6.0).

The 100 μM substrate arrays were used for the plasma kallikrein and factor XIa β assays; all other assays employed the 50 μM substrate arrays. The proteases were delivered to the substrate arrays as described previously and underwent ~30-fold dilution from the initial delivery concentration after mixing in the microspot (5, 6). The slides (n = 6 per enzyme) were incubated at 37 °C for 30 min to 6 h. This incubation time resulted in ~5–25% cleavage of the best substrate on each array, assuring that the assays were run within the linear range. The activated slides were scanned using a cooled CCD-based imager (NovaRay; Alpha Innotech, San Leandro, CA), at Ex λ = 405/40 nm and Em λ = 475/40 nm with integration times of 2,500 msec and 15-μm pixel resolution. Images were acquired in a 16-bit format, and the analysis and presentation of the data was performed using Array Vision (Imaging Research, Ontario, Canada) and Cluster and Treeview (10). For the most preferred substrates, the signal/background ratio was typically >60 with coefficient of variance <5%.

RESULTS AND DISCUSSION

A 722-member, spatially separated ACC library of the format Ac-P_4-P_3-P_2-P_1-ACC-NH_2 was prepared with Ala at the P_4 site, all combinations of proteinogenic amino acids (except Cys) at the P_2 and P_3 sites, and a Lys or Arg residue at the P_1 site (Fig. 1) (6). The substrate specificities of 13 serine proteases and 11 cysteine proteases (Table I, Figs. 2 and 3) were profiled using these substrates in a microarray-based format. The proteases profiled here span three evolutionarily unrelated clans (PA, SB, and C1). Multiple statistically homologous proteases from the PA clan (the S1 family) and the C1 clan (the CA family) were chosen to show that sequence homology need not correlate with substrate specificity.

Serine Proteases—All the serine proteases profiled are from the S1 family of clan PA except subtilisin Carlsberg, which is from the S8 family of clan SB. The two clans differ in the order of their overall fold and catalytic residues, with clan PA having double β barrels with His-Asp-Ser as the catalytic triad and clan SB having parallel β sheets and Asp-His-Ser as the catalytic triad (3, 11).

The proteases profiled from the S1 family showed a preference for P_1 = Lys or Arg due to the presence of the highly conserved negatively charged Asp^{189} (chymotrypsinogen numbering) present at the bottom of the S_1 pocket (12, 13). P_1 Arg residues are able to form a direct ionic bond with Asp^{189}, whereas the corresponding Lys interaction is water-bridged because of the shorter side-chain of Lys. The similarities in active site composition and geometry of the serine proteases profiled can be attributed to convergent evolution, whereas the differences in the substrate specificities can be attributed to changes in the identities of residues present in the substrate binding clefts (14).

Blood Coagulation Proteases—APC is responsible for regulating the blood coagulation pathway by proteolytic inactivation of coagulation cofactors Va and VIIIa, thereby inhibiting the generation of thrombin. Thrombotic risk is associated with the inherited Leiden mutation, which results in APC-resistant

Table I
Protease substrate specificities

Enzyme	Observed specificity			Natural action	EC number	Clan	Family
	P$_1$	P$_2$	P$_3$				
Human APC	K	K>Q>T,V,P	Q>I,K,M	Inhibits factors Va and VIIIa	EC 3.4.21.69	PA	S1
	R	K,T>P,Q,V	Q>M,R,S,T				
Human plasma kallikrein	K	Y,F,K>N,T	K>E,M	Liberates bradykinin from kininogen; also activates plasminogen, factor XII and VII	EC 3.4.21.34	PA	S1
	R	K>F,L>N>T,Y	R>H>A,E,L,G				
Factor VIIa + TF	K	Low cleavage		Initiates extrinsic coagulation cascade; activates factors IX and X	EC 3.4.21.21	PA	S1
	R	V>T	Q,R>N,P				
Factor IXa	K	Low cleavage		Activates factor X; may also activate factor VII	EC 3.4.21.22	PA	S1
	R	F,P,K,Y>L,N	L,M,Q,R				
Factor XIa	K	N>T>S	F>K,M>E,H,L	Activates factor IX in contact phase of activation	EC 3.4.21.27	PA	S1
	R	N>S,T>A	F>E,W				
Factor XIIa	K	Low cleavage		Activates factor XI and plasma kallikrein precursor	EC 3.4.21.38	PA	S1
	R	F,T>G,N,S	M>Q				
Complement C1r	K	A,S,V,I	K,R>Y	Activates complement C1s	EC 3.4.21.41	PA	S1
	R	K	R				
Complement C1s	K	S>A,G	Broad	Complement activation through cleavage of C2 and C4	EC 3.4.21.42	PA	S1
	R	S>G>A	Broad				
Complement factor D	K	A,S,T,V	R>K	Hydrolyzes factor B to cleave complement C3	EC 3.4.21.46	PA	S1
	R	K	R				
Trypsin	K	A,S,N>H	K>Q,R,M	Intestinal digestion of food proteins and activation of other digestive peptidase zymogens	EC 3.4.21.4	PA	S1
	R	S>N>H,T	K,M>N>Q,R,S,T				
Tryptase	K	S>T,N	K>R	Cleaves variety of extracellular proteins	EC 3.4.21.59	PA	S1
	R	N>S,T	R>K				
Subtilisin Carlsberg	K	A>I,S,V>T	Q,R,S,T>E,M	Serine endopeptidase with no cysteine residues	EC 3.4.21.62	SB	S8
	R	I,V>A,S,T	R≫M,Q,S,T				
Cathepsin G	K	T>V>A	I,E>A,V	Protein and bactericidal degradation during neutrophil phagocytosis	EC 3.4.21.20	PA	S1
	R	T>V>L	V>E,Q				
Cathepsin B	K	V>A,I,K,T	K,P > L	Intracellular lysosomal exopeptidase showing peptidyl dipeptidase activity	EC 3.4.22.1	CA	C1
	R	T,V>K,I	K,L,P,R				
Cathepsin H	K	L≫F,I,V	Broad	Lysosomal exopeptidase contributing to antigen processing	EC 3.4.22.16	CA	C1
	R	L>F,I,K>V	Broad				
Cathepsin K	K	L>V>P	Broad	Degrades the organic matrix of bone in osteoclastic bone resorption	EC 3.4.22.38	CA	C1
	R	L≫K	Broad				
Cathepsin L	K	L>F>V	K≫R	Endopeptidase activity in lysosomal proteolysis	EC 3.4.22.15	CA	C1
	R	F,L,V	R				
Cathepsin S	K	L>V	Broad	Lysosomal proteolysis, elastin degradation, MHC immune response	EC 3.4.22.27	CA	C1
	R	L	Broad				
Cathepsin V	K	L>V	Broad	Degradation of the invariant chain of the MHC class II complex	EC 3.4.22.43	CA	C1
	R	L>F,V	Broad				
Rhodesain	K	F,L,Y>V,W>I>M	Broad	Active component in parasitic protozoa *T.b. rhodesiense* causing sleeping sickness	EC 3.4.22.XX	CA	C1
	R	F,L,Y>V>I>W>M	Broad				
Papain	K	V	P	Proteolytic component of the latex of papaya, may provide resistance to pests in host plants	EC 3.4.22.2	CA	C1
	R	V	P				
Chymopapain	K	V>S,T	K>R	Proteolytic component of the latex of papaya, may provide resistance to pests in host plants	EC 3.4.22.6	CA	C1
	R	V>I>T	A>V,R>P				
Ficain	K	L,V>T,Y	P	Proteolytic component of the latex of fig, may provide resistance to pests in host plants	EC 3.4.22.3	CA	C1
	R	L>V,T	P				
Stem bromelain	K	R	P	Most abundant cysteine endopeptidase from the stem of pineapple, may provide resistance to pests in host plants	EC 3.4.22.32	CA	C1
	R	R	P				

factor V (15). APC demonstrated a preference for Thr, Lys, Gln, Val, and Pro in the P$_2$ position (Fig. 2*A*). These data are consistent with the known physiologic substrates of APC: factor Va (cleaved at multiple sites, including Thr^{305}Arg^{306}∼Asn^{307} and Thr^{678}Arg^{679}∼Lys^{680}, where ∼ represents the cleaved bond) and factor VIIIa (Gln^{562}Arg^{563}∼Gly^{564} and Pro^{739}Arg^{740}∼Ser^{741}) (3). Furthermore, in a study with 25 synthetic peptidyl coumarin substrates, Ohno *et al.* found Boc-Leu-Ser-Thr-Arg-coumarin to be the most efficient APC substrate (16). At the P$_3$ position, both the Arg and Lys sublibraries revealed a prefer-

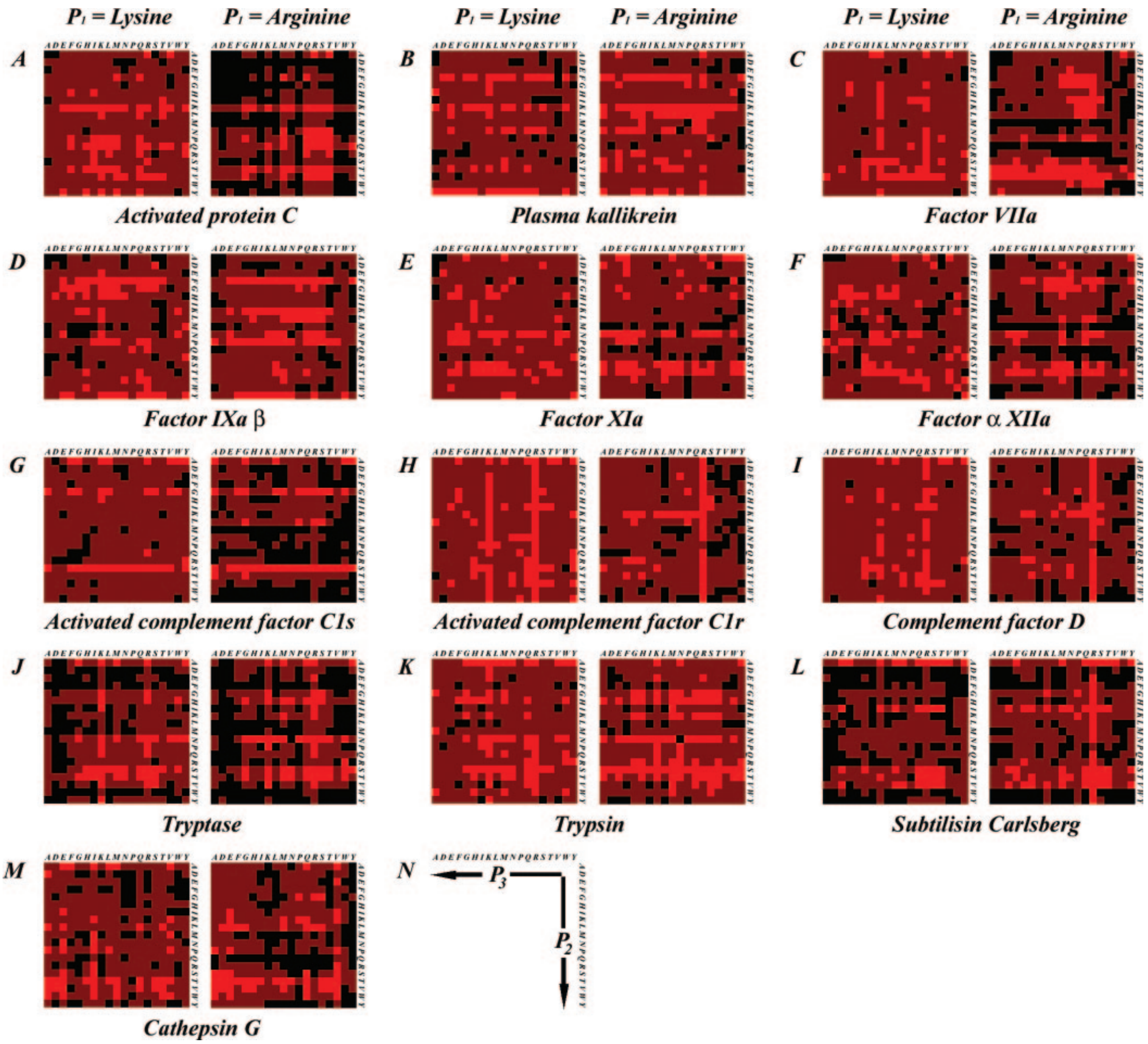

FIG. 2. **Serine proteases.** Characterization of the specificity of serine proteases using the Ac-Ala-P_3-P_2-Lys-ACC-NH_2 and Ac-Ala-P_3-P_2-Arg-ACC-NH_2 substrate microarrays. Each square in the grid is colored in proportion to the average ($n = 6$) quantitated fluorescence intensity of the corresponding substrates after treatment with the enzyme, indicating the relative amount of cleavage. The vertical axis indicates the P_2 residue and the horizontal axis indicates the P_3 residues, as shown in *N*.

ence for Gln and Met. In addition to these shared preferences, the P_1 = Lys sublibrary also revealed efficient cleavage of substrates with P_3 Lys and Ile residues, whereas the P_1 = Arg sublibrary showed a preference for P_3 Ser and Thr residues.

Human plasma kallikrein plays an important role in contact activation of clotting, fibrinolysis, blood pressure regulation, and neutrophil stimulation. Deficiency of the zymogen, plasma prekallikrein, results in Fletcher trait, an autosomal recessive disorder associated with myocardial infarction, thromboembolism, and other vascular diseases (11). Human plasma kallikrein showed a preference for the aromatic residues Phe and Tyr, along with Lys, Asn, Thr (and, for the P_1 = Arg sublibrary, Leu) in the P_2 position (Fig. 2*B*). The enzyme showed weaker P_3 specificity (Lys, Glu, and Met for P_1 = Lys sublibrary and Arg, His, Ala, Glu, Leu, and Gly for P_1 = Arg sublibrary). The data are consistent with the physiological substrates of plasma kallikrein: factor XII ($Thr^{352}Arg^{353}{\sim}Val^{354}$), kininogen ($Phe^{388}Arg^{389}{\sim}Ser^{390}$), the urokinase plasminogen activator precursor ($Phe^{157}Lys^{158}{\sim}Ile^{159}$), and human prorenin ($Lys^{42}Arg^{43}{\sim}Leu^{44}$). These findings are also consistent with previous studies of plasma kallikrein specificity (3, 17–20). In a panel of 20 coumarin substrates, Iwanaga *et al.* found

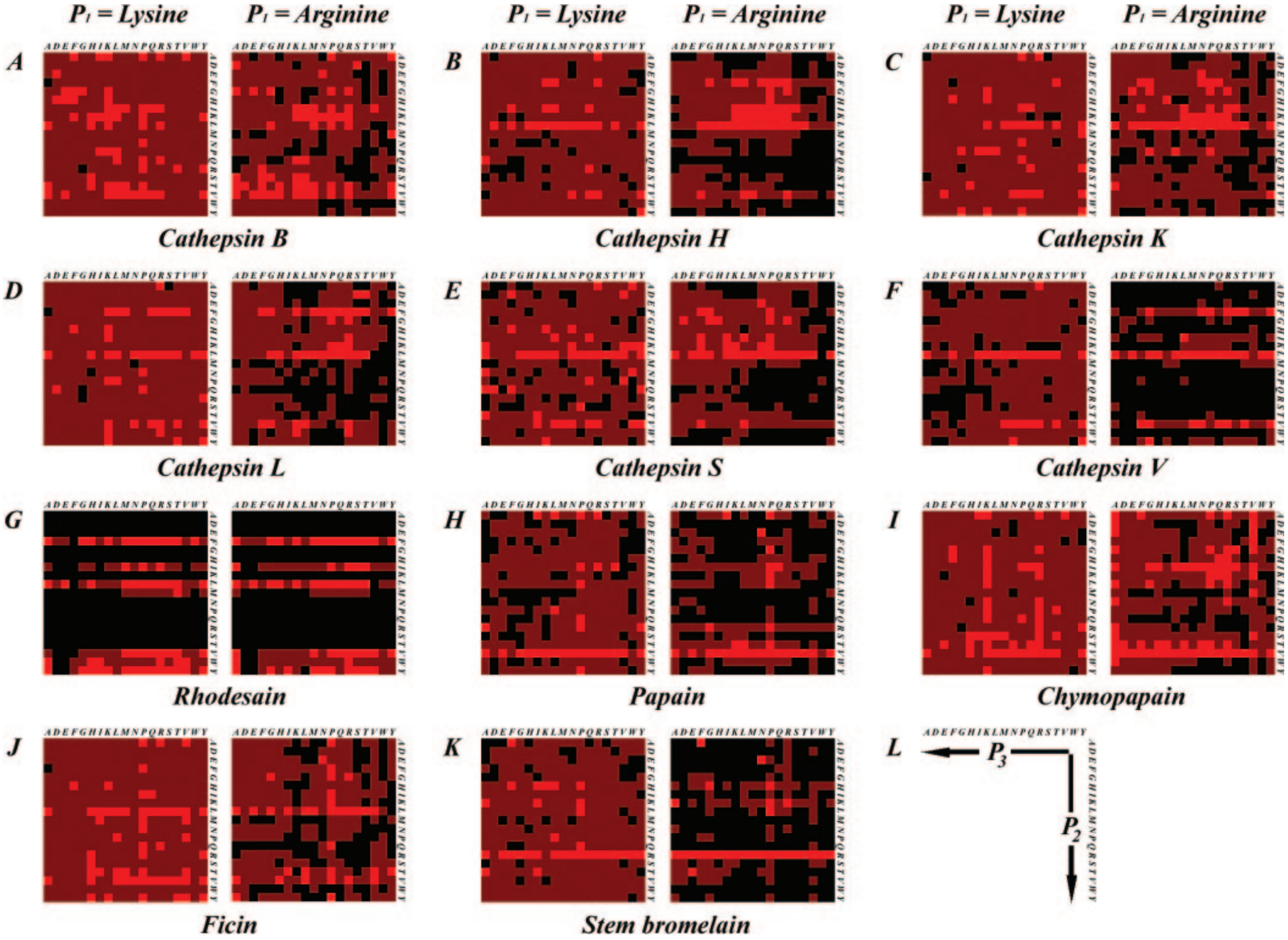

FIG. 3. **Cysteine proteases.** Characterization of the specificity of cysteine proteases using the Ac-Ala-P_3-P_2-Lys-ACC-NH_2 and Ac-Ala-P_3-P_2-Arg-ACC-NH_2 substrate microarrays. Each square in the grid is colored in proportion to the average (n = 6) quantitated fluorescence intensity of the corresponding substrates after treatment with the enzyme, indicating the relative amount of cleavage. The vertical axis indicates the P_2 residue, and the horizontal axis indicates the P_3 residues, as shown in *L*.

substrates with a P_2 Phe residue to be more efficiently cleaved than those with P_2 Gly or Pro residues (21). Whereas it has been noted that extending substrate length beyond the P_2 site has little effect on turnover efficiency for some synthetic substrates (20), extension to at least the P_4 site is critical for the physiological functioning of plasma kallikrein (17).

The tissue factor/factor VIIa complex initiates the extrinsic pathway of the blood coagulation cascade. Elevated levels of VIIa have been implicated in heart disease, although deficiency results in severe bleeding (22). Factor VIIa functions physiologically in complex with Ca^{2+} and tissue factor. Ca^{2+} is required by factor VIIa to become conformationally active (23), and tissue factor enhances the amidolytic efficiency of factor VIIa by 60- to 100-fold (23, 24) by causing a conformational change in factor VIIa (23, 25). As expected, we found that in the absence of tissue factor and $CaCl_2$, factor VIIa showed essentially no cleavage with our sublibraries (data not shown), consistent with literature reports of poor activity in the absence of these activators (11, 20, 23, 26, 27). With addition of tissue factor and $CaCl_2$, factor VIIa showed enhanced cleavage of the P_1 = Arg sublibrary (very low cleavage of the P_1 = Lys sublibrary was observed). A strong preference was observed for the branched amino acids Val and Thr in the P_2 position and Gln, Arg, Asn, and Pro in the P_3 position for the P_1 = Arg library (Fig. 2*C*). These data are consistent with the physiologic substrates of factor VIIa: factor IX ($Thr^{179}Arg^{180}\sim Val^{181}$ and $Thr^{144}Arg^{145}\sim Ala^{146}$) and factor X ($Thr^{150}Arg^{151}\sim Ile^{152}$) (3).

Factor IXaβ is involved in thrombin generation, and deficiency results in hemophilia (11). Factor IXaβ has low proteolytic activity in the absence of its physiological cofactor, coagulation factor VIIIa (13, 19, 20, 26). In the absence of factor VIIIa, we observed cleavage of the P_1 = Arg sublibrary (very low cleavage of the P_1 = Lys sublibrary was observed). This difference in activity was likely due to the increased stability of the enzyme active site when binding substrates containing a P_1 Arg residue. Within the S_1 subsite, the shorter side-chain of the P_1 Lys forms a water-bridged interaction with Asp^{189}, whereas the longer side-chain of the P_1 Arg stabilizes the S_1 pocket via a salt bridge formed between the Arg guanidium group and the carboxylate of Asp^{189}. The stronger salt bridge interaction results in a 30- to 40-fold preference for Arg over Lys at the P_1 position (13). Factor IXaβ showed a preference for Phe, Pro, Tyr, Leu, Lys, and Asn in the P_2 position and fairly broad specificity at the P_3 position for the P_1 = Arg sublibrary (Fig. 2*D*). These results are consistent with a previous report by Iwanaga *et al.*, which found a preference for substrates with a P_2 Pro residue over those with a P_2 Ser, Thr, or Gly residue when studying a panel of 15 peptidyl coumarin substrates (21).

Human factor XIa showed a strong preference for Thr, Ser, and Asn in the P_2 position for both the Arg and Lys sublibraries (Fig. 2*E*). A diverse set of residues was accepted at the P_3 site. These results are consistent with the physiologically relevant substrate coagulation factor IX (Thr^{144}Arg^{145}~Ala^{146} and Thr^{179}Arg^{180}~Val^{181}) (3) and with previously reported synthetic substrate studies: Iwanaga *et al.* reported highly efficient cleavage of P_2 Ser and Thr substrates from a panel of 15 peptidyl coumarin substrates (21).

As with factor VIIa and IXa β, human factor α XIIa showed poor proteolytic activity toward the P_1 Lys substrates, as had been observed previously (28). The P_1 = Arg sublibrary showed preference for Phe, Thr, Ser, Gly, and Asn in the P_2 position and Met and Gln in the P_3 position (Fig. 2*F*). These data correlate well with the physiologic substrates of factor XIIa: coagulation factors VII (Gly^{193}Arg^{194}~Ile^{195}) and XII (Thr^{352}Arg^{353}~Val^{354}) and the precursor of plasma kallikrein (Thr^{370}Arg^{371}~Ile^{372}) (3) and with single substrate studies that report a preference for Phe and Gly in the P_2 position and disfavor Glu in the P_3 position (19, 20, 29).

Complement Pathway Proteases—The complement pathway plays an important role in host defense against pathogens and infections. The complement pathway consists of multiple functionally linked proteins that mediate inflammatory responses, clear foreign pathogens, and kill susceptible cells (30, 31). These enzymes have also been implicated in prion aggregation (32) and Alzheimer disease (33).

Activated human complement factor C1s preferred the small polar residues Ser, Ala, and Gly in the P_2 position and had broad specificity at the P_3 position (Fig. 2*G*). This is consistent with the physiological substrate of C1s, C2 (Gly^{222}-Arg^{223}~Lys^{224})^3 and with previous studies employing peptidyl substrates. McRae *et al.* report the fastest cleavage of substrates of the format Cbz-P_2-Arg-thioester when P_2 = Ser, Ala, or Gly, and Keough *et al.* report efficient cleavage of P_2 = Ser and Gly *p*-nitroanilide substrates (34–36).

Compared with complement C1s, complements factor D and C1r were relatively inactive against the peptide libraries. This disparity is due to the fact that these enzymes become catalytically active through an induced fit mechanism in the presence of the physiologic substrate (for factor D) (37) or in a larger protein complex (for C1r) (38). As a consequence, peptide substrates are not sufficient to reveal the full catalytic potential of the enzymes (35). Interestingly, while cleavage of tri-, tetra-, and pentapeptides have been reported to be slow, the dipeptidase activity of these two enzymes is relatively efficient (35, 37, 39). For both factor D and C1r, we observed efficient cleavage of substrates with P_3 = Arg or Lys (Fig. 2, *H* and *I*). While P_3 basic residues are found in the natural substrates for activated C1r (C1r and C1s precursor) (3), this apparent specificity would also be expected if the cleavage efficiency of dipeptides is faster than that of longer substrates. In such a case, the efficient dipeptidase activity of the enzyme would result in fast cleavage of peptides containing a basic amino acid two residues from the N terminus (for the tetrapeptidyl substrates as used in this study, this would result in cleavage between a basic P_3 residue and the P_2 site). A second dipeptide cleavage could then occur between the P_1 basic residue and the coumarin moiety, resulting in an increase in fluorescence intensity and an apparent P_3 basic preference. Given the dipeptidase activity for these two enzymes, we cannot discount that the observed specificity results from off target cleavage.

Other Serine Proteases—Tryptases are serine proteases stored in mast cell granules and are known to mediate inflammatory and allergic responses. They have been implicated in asthma and other allergy disorders. Secreted as catalytically active, noncovalently bound tetramers, there are at least four closely related tryptases: α, βI, II, and III, sharing at least 93% identical amino acid sequences (40). Human lung tryptase (α) showed a preference for the polar amino acids Asn, Ser, and Thr in the P_2 position for both the P_1 = Arg and Lys sublibraries (Fig. 2*J*). A strong preference for basic residues Arg and Lys was seen in the P_3 position. These results are in agreement with data from positional scanning libraries (PS-SCLs) of peptidyl coumarin substrates for α tryptase (41, 42).

The porcine pancreatic trypsin used in our study preferred polar residues such as Ser, Asn, and His at the P_2 position (Fig. 2*K*). Additionally, we observed cleavage of P_2 Thr substrates for the P_1 = Arg sublibrary and P_2 Ala substrates for P_1 = Lys sublibrary. This is consistent with findings in single substrate experiments reported in literature and natural substrates: chymotrypsinogen A (Ser^{14}Arg^{15}~Ile^{16}), pancreatic elastase 2A precursor (Thr^{11}Arg^{12}~Val^{13}) and pancreatic elastase 2B precursor (Ser^{11}Arg^{12}~Met^{13}) (3, 19, 20, 41, 43). Broader specificity was seen in the P_3 position, with Lys, Gln, Arg, Met, Asn, Ser, and Thr all accepted for the P_1 = Arg sublibrary. In general, Pro was least preferred at the P_3 site. The broad acceptance of substrates has also been recently observed in PS-SCL studies described by Furlong *et al* (41).

While the serine protease subtilisin Carlsberg from *B. licheniformis* reportedly does not prefer substrates with P_1 basic residues (more commonly reported as accepting Glu, Gln, Asp, Phe, and Leu in the P_1 position) (44–46), we show that both the P_1 = Arg and Lys sublibraries are cleaved by the enzyme. We see that the branched residues Ile, Thr, and Val in addition to Ala and Ser are preferred in the P_2 position and Gln, Arg, Ser, Thr, and Met in the P_3 position (Fig. 2*L*). Cleavage of tetrapeptide substrates with P_1 = Lys and Arg were confirmed in conventional solution-phase well-plate assays (data not shown).

Human cathepsin G, along with human neutrophil elastase and proteinase 3, is found in the azurophilic granules of neutrophils and is responsible for the degradation of foreign organisms, activation of CD IIb/18-bound Xa, and platelet activation (11, 47–49). Human neutrophil cathepsin G is a unique serine protease because of its chymotrypsin-like (P_1 hydrophobic) and trypsin-like (P_1 basic) substrate specificities (47, 48, 50–52). This dual specificity can be attributed to Glu^{226} residing at the bottom of the S_1 pocket capable of interacting with both basic and aromatic/hydrophobic residues (47, 48, 53). For the trypsin-like activity, we observed a preference for Thr and Val at the P_2 position. This is different from the reported preference for Pro or Met in the P_2 position when the P_1 is a hydrophobic residue like Phe, Leu, or Met (50–52). In the P_3 position, Ile, Glu, Ala, and Val were preferred for the P_1 = Lys sublibrary, and Val, Glu, and Gln were preferred for the P_1 = Arg sublibrary were preferred (Fig. 2*M*). Similar P_3 specificity has been reported with cathepsin G substrates having P_1 = Phe, Leu, and Met (50–52).

Cysteine Proteases—All the cysteine proteases profiled are from the C1 papain family of clan CA and have a catalytic triad of Cys, His, and Asn (3, 11). A characteristic of the C1 family of peptidases is its strong dependence on the S_2 subsite with the nature of the P_2 residue of the substrate being the main factor in determining the specificity of the protease (54). Most C1 family cysteine proteases prefer hydrophobic aliphatic or aromatic residues in the P_2 position (55–59).

Cysteine Cathepsins—Cathepsins are papain-like cysteine proteases that play an important role in cellular protein degradation through the lysosomal pathway (60). They represent a significant class of drug targets having been implicated in a variety of degenerative and invasive processes such as cancer, inflammation, rheumatoid and osteoarthritis, Alzheimer disease, multiple sclerosis, pancreatitis, liver disorders, lung disorders, lysosomal disorders, myocardial disorders, diabetes, muscular dystrophy, tumor invasion, and metastasis (60). Recently, some of the cathepsins have been shown to play an important role in specific physiological functions associated with their tissue localization (60, 61). The cathepsins show either endopeptidase (F, K, L, S, and V) or exopeptidase (B, C, H, and X) activity, with the primary determinant of specificity for the cathepsins reported to be the S_2 subsite.

For cathepsin B, we observed a preference for branched residues like Val, Thr, and Ile in the P_2 position for both sublibraries (Fig. 3*A*). The preference for bulky aliphatic residues at the P_2 site is consistent with typical C1 family member specificity and with a single substrate study that reported preferential cleavage of the substrate Bz-Phe-Val-Arg-coumarin (62). We also observed cleavage of substrates containing a P_2 Lys residue. The efficient cleavage of P_2 = Lys substrates can be attributed to the presence of Glu^{205} in the P_2 pocket (papain numbering), which can mediate the charge of the basic residues via salt bridging (63–65). Cathepsin B demonstrated specificity for Lys, Leu, and Pro in the P_3 position, which is consistent with literature reports (59, 66).

Cathepsin H, the only other exopeptidase profiled, showed a clear preference for the aliphatic hydrophobic residue Leu in the P_2 position (Fig. 3*B*). The other residues showing a minor preference at this position were Phe, Ile, and Val (and Lys for the P_1 = Arg sublibrary). Single substrate experiments with bovine cathepsin H showed a preference for aliphatic hydrophobic residues over aromatic residues (67). A broad specificity was observed for the S_3 subsite.

Cathepsin K also revealed a clear preference for Leu in the P_2 position and a broad specificity in the P_3 position (Fig. 3*C*), consistent with literature reports (68, 69). We observed a minor preference for P_2 Lys (for the P_1 = Arg sublibrary), consistent with a previous report of cathepsin K accepting P_2 basic residues (69). Unlike the homologous cathepsins V, L, and H, cathepsin K also accepted Pro at the P_2 position. This P_2 Pro acceptance is required for the unique collagenase activity of cathepsin K (68) and has been noted previously in both PS-SCL (68) and fluorescence resonance energy transfer (FRET) studies (69). While cooperative interactions between substrate subsites cannot be detected using PS-SCLs, the study by Alves *et al.* using FRET substrates had noted that the acceptance of P_2 Pro residues was dependent on the neighboring substrate residues. Of the six P_2 Pro substrates tested in the FRET study, the most efficient cleavage was observed with substrates containing a P_3 Lys residue (69). We found that for our Ac-Ala-P_3-Pro-(Arg/Lys)-ACC-NH_2 substrates, P_3 Lys substrates were among the most efficient (along with P_3 = Pro and Gly for the P_1 = Arg sublibrary and P_3 = Gly, Ile, and Leu for the P_1 = Lys sublibrary).

Cathepsin L showed a preference for the bulky hydrophobic residues Val, Phe, and Leu in the P_2 site (Fig. 3*D*), with considerably less acceptance of P_2 Pro substrates. Prior studies have attributed the preference for large aromatic and bulky residues at the P_2 site to the presence of the small residue Ala^{205} (papain numbering) at the bottom of the S_2 pocket, which would suggest the presence of a large open-ended pocket (54, 58). In cathepsin K, this residue is a Leu, a change that is implicated in the differential acceptance of P_2 Pro residues between the two enzymes (68). The P_2 specificity observed in the microarray assay is in accordance with Maciewicz *et al.* who found cathepsin L preferred the substrate Cbz-Phe-Arg-coumarin (62). Similarly, using a panel of FRET substrates to study cathepsin L, Puzer *et al.* report a preference for substrates with P_2 Leu and Phe (70). At the P_3 position, basic residues were preferred, consistent with literature reports (68, 70).

Cathepsin S is homologous to the other endopeptidase cathepsins as demonstrated by the superposition of its structure onto those of cathepsins K, V, and L with root mean square deviation values of <1 Å (54). Cathepsin S showed a preference for Leu in the P_2 position in both sublibraries with a broad specificity for the P_3 position (Fig. 3*E*). Similar P_2 specificity for hydrophobic residues has been observed in single substrate studies for bovine and human cathepsin S (58, 71). The diminished acceptance of P_2 Val residues relative to cathepsin V and cathepsin L may be attributed to the presence of a Val residue that more greatly restricts the entrance to the S_2 pocket in cathepsin S relative to cathepsin V (which contains a Leu residue at this position) and Cat L (which has a Met residue at this position).

Cathepsin V, like cathepsin L, shows preference for Leu and Val in the P_2 position with broad specificity in the P_3 position (Fig. 3*F*), consistent with literature reports (70, 72). This similarity can be attributed to the fact that both cathepsins V and L contain an Ala^{205} in the S_2 pocket, allowing for accommodation of large residues at position 205, resulting in a deeper S_2 subsite pocket (72). Puzer *et al.* reported that cathepsin V preferred Leu over Phe in the P_2 position in studies with a panel of dipeptidyl-coumarin substrate studies (70, 73), showing that cathepsin V has an intermediate specificity between cathepsin L and S.

Rhodesain—Rhodesain is a papain-like cysteine protease expressed by *Trypanasoma brucei rhodesiense*, a parasitic protozoa. The parasite is the cause of sleeping sickness, which infects nearly 25,000∼50,000 people yearly, and if left untreated, is fatal. The current chemotherapeutic treatments are poorly effective for the disease. In fact, the only treatment effective against *T. b. rhodesiense*, melarsopol, results in potentially fatal encephalopathy in 5–10% of treated patients (8). As new inhibitors are required to combat the disease, elucidating the substrate specificity is of prime importance. While a comprehensive study of the substrate specificity of rhodesain has not previously been reported, an initial study by Caffrey *et al.* employing four dipeptidyl coumarin substrates found efficient cleavage of Cbz-Val-Arg-coumarin and Cbz-Phe-Arg-coumarin (8). We show here that the protease shows a very strong preference for hydrophobic and aromatic residues (Leu, Val, Phe, Tyr, and a minor preference for Ile, Trp, Met) in the P_2 position and has a broad specificity for non-acidic residues in the P_3 position (Fig. 3*G*).

Plant Cysteine Proteases—Papain, a plant enzyme from the papaya fruit, is the archetypal cysteine protease. Papain showed a strong preference for Val in the P_2 position and Pro in the P_3 position (Fig. 3*H*). These data are consistent with PS-SCL and quenched combinatorial library studies reported in the literature (74, 75).

Chymopapain from papaya latex, ficain from fig tree latex, and stem bromelain from pineapple are fruit enzymes belonging to the papain family of cysteine proteases. Despite their homology, each displayed unique substrate specificity in the microarray assay. Chymopapain showed strong P_2 specificity for the branched amino acids Val, Thr, and Ile (Fig. 3*I*). Ficain preferred Leu, Val, and Thr in the P_2 position (Fig. 3*J*), and stem bromelain showed a strict preference for Arg in the P_2 position (Fig. 3*K*). This strong preference of stem bromelain for Arg at both the P_1 and P_2 sites has previously been noted for other synthetic substrates; Cbz-Arg-Arg-coumarin and Cbz-Arg-Arg-*p*-nitroanilide substrates were found to be the most efficiently cleaved substrates in a panel of 18 potential substrates by Rowan *et al* (76). In the P_3 position, both ficain and stem bromelain, like papain, showed a preference for substrates with Pro at this site.

CONCLUSIONS

Microarray-based methodologies are becoming increasingly important proteomic tools. However, to date most microarray-based methodologies have relied on covalent attachment of the library members (compound, peptides, proteins) under investigation to the array surface. Direct attachment introduces the potential for solid-phase effects, such as inaccessibility and nonspecific binding (77, 78). The solution-phase microarray approach discussed here minimizes these risks because the glycerol used to form the array microspots is a water mimic capable of forming multiple hydrogen bonds. As such, the risk of protein aggregation/denaturation, compound precipitation, and nonspecific interactions is reduced (79–81). Furthermore, the glycerol provides the necessary lubrication for enzymatic catalysis (79–81). While glycerol has been noted to result in increased proteolytic efficiency through oligomerization and structural stabilization for some proteases (82–84), correlation studies between microarray and well-plate data indicate no dramatic variations in protease specificity (6). Here we have demonstrated the broad generality of this solution-phase microarray approach by efficiently profiling the specificity of 13 serine proteases and 11 cysteine proteases with a spatially separated, 722-member fluorogenic substrate microarray.

The high throughput nature of the solution-phase microarray enabled us to map substrate specificity in a highly parallel and miniaturized format with minimal consumption of biological samples (5, 6). At the high density that the fluorogenic substrate library was arrayed, a complete Ac-P_4-P_3-P_2-P_1-ACC-NH_2 library with one member held fixed (19^3 compounds) can be accommodated on one or two microarrays, allowing rapid elucidation of specificities of hundreds of distinct enzymes. Also, differential proteolytic activities of cellular lysates or other complex biological fluids (*e.g.* blood) can be characterized or phenotyped with potential for diagnostic applications.[2] The method enables us to decipher the cooperative interactions between substrate subsites and provides access to functional evolutionary data (6). The assay method is amenable to drug discovery, metabolic/toxicity profiling,

---

[2] D. N. Gosalia and S. L. Diamond, unpublished results.

and may be useful for engineering potent and specific substrates and inhibitors (5, 6).

Acknowledgment—We thank Dustin J. Maly for assistance with the preparation of the P_1 = Lys library.

* This work was supported by the National Institutes of Health Grants GM54051 (to J. A. E.) and HL56621 (to S. L. D.) and the University of California Cancer Research Coordination Committee. The costs of publication of this article were defrayed in part by the payment of page charges. This article must therefore be hereby marked "*advertisement*" in accordance with 18 U.S.C. Section 1734 solely to indicate this fact.

§ These authors contributed equally to this work.

‖ To whom correspondence should be addressed: Jonathan A. Ellman, Dept. of Chemistry, University of California, Berkeley, CA 94720-1460. Tel.: 510-642-4488; Fax: 510-642-8369; E-mail: jellman@uclink.berkeley.edu. Scott L. Diamond, Institute for Medicine and Engineering, University of Pennsylvania, 1024 Vagelos Research Laboratory, 3340 Smith Walk, Philadelphia, PA 19104. Tel.: 215-573-5702; Fax: 215-573-7227; E-mail: sld@seas.upenn.edu.

REFERENCES

1. Leung, D., Abbenante, G., and Fairlie, D. P. (2000) Protease inhibitors: Current status and future prospects. *J. Med. Chem.* **43,** 305–341
2. Southan, C. (2000) Assessing the protease and protease inhibitor content of the human genome. *J. Pept. Sci.* **6,** 453–458
3. Rawlings, N. D., Tolle, D. P., and Barrett, A. J (2004). MEROPS: The peptidase database. *Nucleic Acids Res.* **32,** D160–D164
4. Puente, X. S., Sanchez, L. M., Overall, C. M., and Lopez-Otin, C. (2003) Human and mouse proteases: A comparative genomic approach. *Nat. Rev. Genet* **4,** 544–558
5. Gosalia, D. N., and Diamond, S. L. (2003) Printing chemical libraries on microarrays for fluid phase nanoliter reactions. *Proc. Natl. Acad. Sci. U. S. A.* **100,** 8721–8726
6. Gosalia, D. N., Salisbury, C. M., Maly, D. J., Ellman, J. A., and Diamond, S. L (2005). Profiling serine protease substrate specificity with solution phase fluorogenic peptide microarrays. *Proteomics* **5,** 1292–1298
7. Schechter, I., and Berger, A. (1967) On the size of the active site in proteases. I. Papain. *Biochem. Biophys. Res. Commun.* **27,** 157–162
8. Caffrey, C. R., Hansell, E., Lucas, K. D., Brinen, L. S., Alvarez Hernandez, A., Cheng, J., Gwaltney, S. L., 2nd, Roush, W. R., Stierhof, Y. D., Bogyo, M., Steverding, D., and McKerrow, J. H. (2001) Active site mapping, biochemical properties and subcellular localization of rhodesain, the major cysteine protease of *Trypanosoma brucei rhodesiense*. *Mol. Biochem. Parasitol.* **118,** 61–73
9. Salisbury, C. M., Maly, D. J., and Ellman, J. A. (2002) Peptide microarrays for the determination of protease substrate specificity. *J. Am. Chem. Soc.* **124,** 14868–14870
10. Eisen, M. B., Spellman, P. T., Brown, P. O., and Botstein, D. (1998) Cluster analysis and display of genome-wide expression patterns. *Proc. Natl. Acad. Sci. U. S. A.* **95,** 14863–14868
11. Barrett, A. J., Rawlings, N. D., and Woessner, J. F (1998). *Handbook of Proteolytic Enzymes.* Academic Press, London.
12. Krem, M. M., and Di Cera, E. (1998) Conserved water molecules in the specificity pocket of serine proteases and the molecular mechanism of Na^+ binding. *Proteins* **30,** 34–42
13. Sichler, K., Hopfner, K. P., Kopetzki, E., Huber, R., Bode, W., and Brandstetter, H. (2002) The influence of residue 190 in the S1 site of trypsin-like serine proteases on substrate selectivity is universally conserved. *FEBS Lett.* **530,** 220–224
14. Perona, J. J., and Craik, C. S. (1995) Structural basis of substrate specificity in the serine proteases. *Protein Sci.* **4,** 337–360
15. Castoldi, E., and Rosing, J. (2004) Factor V Leiden: A disorder of factor V anticoagulant function. *Curr. Opin. Hematol.* **11,** 176–181
16. Ohno, Y., Kato, H., Morita, T., Iwanaga, S., Takada, K., Sakakibara, S., and Stenflo, J. (1981) A new fluorogenic peptide substrate for vitamin K-dependent blood coagulation factor, bovine protein C. *J. Biochem.* **90,** 1387–1395
17. Almeida, P. C., Chagas, J. R., Cezari, M. H., Juliano, M. A., and Juliano, R. (2000) Hydrolysis by plasma kallikrein of fluorogenic peptides derived from prorenin processing site. *Biochim. Biophys. Acta* **1479,** 83–90
18. Morita, T., Kato, H., Iwanaga, S., Takada, K., and Kimura, T. (1977) New fluorogenic substrates for α-thrombin, factor Xa, kallikreins, and urokinase. *J. Biochem.* **82,** 1495–1498
19. McRae, B. J., Kurachi, K., Heimark, R. L., Fujikawa, K., Davie, E. W., and Powers, J. C. (1981) Mapping the active sites of bovine thrombin, factor IXa, factor Xa, factor XIa, factor XIIa, plasma kallikrein, and trypsin with amino acid and peptide thioesters: Development of new sensitive substrates. *Biochemistry* **20,** 7196–7206
20. Kawabata, S., Miura, T., Morita, T., Kato, H., Fujikawa, K., Iwanaga, S., Takada, K., Kimura, T., and Sakakibara, S. (1988) Highly sensitive peptide-4-methylcoumaryl-7-amide substrates for blood-clotting proteases and trypsin. *Eur. J. Biochem.* **172,** 17–25
21. Iwanaga, S., Morita, T., Kato, H., Harada, T., Adachi, N., Sugo, T., Maruyama, I., Takada, K., Kimura, T., and Sakakibara, S. (1979) Fluorogenic peptide substrates for proteases in blood coagulation, kallikrein-kinin and fibrinolysis systems. *Adv. Exp. Med. Biol.* **120A,** 147–163
22. Tuddenham, E. G., Pemberton, S., and Cooper, D. N. (1995) Inherited factor VII deficiency: Genetics and molecular pathology. *Thromb. Haemost.* **74,** 313–321
23. Lawson, J. H., Butenas, S., and Mann, K. G. (1992) The evaluation of complex-dependent alterations in human factor VIIa. *J. Biol. Chem.* **267,** 4834–4843
24. Butenas, S., Ribarik, N., and Mann, K. G. (1993) Synthetic substrates for human factor VIIa and factor VIIa-tissue factor. *Biochemistry* **32,** 6531–6538
25. Nemerson, Y. (1988) Tissue factor and hemostasis. *Blood* **71,** 1–8
26. Butenas, S., Orfeo, T., Lawson, J. H., and Mann, K. G. (1992) Aminonaphthalenesulfonamides, a new class of modifiable fluorescent detecting groups and their use in substrates for serine protease enzymes. *Biochemistry* **31,** 5399–5411
27. Butenas, S., and Mann, K. G. (1996) Kinetics of human factor VII activation. *Biochemistry* **35,** 1904–1910
28. Kato, H., Adachi, N., Ohno, Y., Iwanaga, S., Takada, K., and Sakakibara, S. (1980) New fluorogenic peptide substrates for plasmin. *J. Biochem.* **88,** 183–190
29. Cho, K., Tanaka, T., Cook, R. R., Kisiel, W., Fujikawa, K., Kurachi, K., and Powers, J. C. (1984) Active-site mapping of bovine and human blood coagulation serine proteases using synthetic peptide 4-nitroanilide and thio ester substrates. *Biochemistry* **23,** 644–650
30. Campbell, R. D., Law, S. K., Reid, K. B., and Sim, R. B. (1988) Structure, organization, and regulation of the complement genes. *Annu. Rev. Immunol.* **6,** 161–195
31. Muller-Eberhard, H. J. (1988) Molecular organization and function of the complement system. *Annu. Rev. Biochem.* **57,** 321–347
32. Klein, M. A., Kaeser, P. S., Schwarz, P., Weyd, H., Xenarios, I., Zinkernagel, R. M., Carroll, M. C., Verbeek, J. S., Botto, M., Walport, M. J., Molina, H., Kalinke, U., Acha-Orbea, H., and Aguzzi, A. (2001) Complement facilitates early prion pathogenesis. *Nat. Med.* **7,** 488–492
33. Rogers, J., Cooper, N. R., Webster, S., Schultz, J., McGeer, P. L., Styren, S. D., Civin, W. H., Brachova, L., Bradt, B., Ward, P., *et al.* (1992) Complement activation by β-amyloid in Alzheimer disease. *Proc. Natl. Acad. Sci. U. S. A.* 89, 10016–10020
34. Keogh, S. J., Harding, D. R., and Hardman, M. J. (1987) The complement component C1s catalysed hydrolysis of peptide 4-nitroanilide substrates. *Biochim. Biophys. Acta* **913,** 39–44
35. Kam, C. M., McRae, B. J., Harper, J. W., Niemann, M. A., Volanakis, J. E., and Powers, J. C. (1987) Human complement proteins D, C2, and B. Active site mapping with peptide thioester substrates. *J. Biol. Chem.* **262,** 3444–3451
36. McRae, B. J., Lin, T. Y., and Powers, J. C. (1981) Mapping the substrate binding site of human C1r and C1s with peptide thioesters. Development of new sensitive substrates. *J. Biol. Chem.* **256,** 12362–12366
37. Volanakis, J. E., and Narayana, S. V. (1996) Complement factor D, a novel serine protease. *Protein Sci.* **5,** 553–564
38. Budayova-Spano, M., Lacroix, M., Thielens, N. M., Arlaud, G. J., Fontecilla-Camps, J. C., and Gaboriaud, C. (2002) The crystal structure of the zymogen catalytic domain of complement protease C1r reveals that a disruptive mechanical stress is required to trigger activation of the C1

complex. *EMBO J.* **21,** 231–239

39. Taylor, F. R., Bixler, S. A., Budman, J. I., Wen, D., Karpusas, M., Ryan, S. T., Jaworski, G. J., Safari-Fard, A., Pollard, S., and Whitty, A. (1999) Induced fit activation mechanism of the exceptionally specific serine protease, complement factor D. *Biochemistry* **38,** 2849–2859
40. Huang, C., Li, L., Krilis, S. A., Chanasyk, K., Tang, Y., Li, Z., Hunt, J. E., and Stevens, R. L. (1999) Human tryptases α and β/II are functionally distinct due, in part, to a single amino acid difference in one of the surface loops that forms the substrate-binding cleft. *J. Biol. Chem.* **274,** 19670–19676
41. Furlong, S. T., Mauger, R. C., Strimpler, A. M., Liu, Y. P., Morris, F. X., and Edwards, P. D. (2002) Synthesis and physical characterization of a P1 arginine combinatorial library, and its application to the determination of the substrate specificity of serine peptidases. *Bioorg. Med. Chem.* **10,** 3637–3647
42. Harris, J. L., Niles, A., Burdick, K., Maffitt, M., Backes, B. J., Ellman, J. A., Kuntz, I., Haak-Frendscho, M., and Craik, C. S. (2001) Definition of the extended substrate specificity determinants for β-tryptases I and II. *J. Biol. Chem.* **276,** 34941–34947
43. Tanaka, T., McRae, B. J., Cho, K., Cook, R., Fraki, J. E., Johnson, D. A., and Powers, J. C. (1983) Mammalian tissue trypsin-like enzymes. Comparative reactivities of human skin tryptase, human lung tryptase, and bovine trypsin with peptide 4-nitroanilide and thioester substrates. *J. Biol. Chem.* **258,** 13552–13557
44. Kaspari, A., Schierhorn, A., and Schutkowski, M. (1996) Solid-phase synthesis of peptide-4-nitroanilides. *Int. J. Pept. Protein Res.* **48,** 486–494
45. Milgotina, E. I., Shcheglov, A. S., Lapa, G. B., Chestukhina, G. G., and Voyushina, T. L. (2001) Enzymatic synthesis of chromogenic substrates for Glu, Asp-specific proteinases. *J. Pept. Res.* **58,** 12–16
46. Meldal, M., Svendsen, I., Breddam, K., and Auzanneau, F. I (1994). Portion-mixing peptide libraries of quenched fluorogenic substrates for complete subsite mapping of endoprotease specificity. *Proc. Natl. Acad. Sci. U. S. A.* **91,** 3314–3318
47. Polanowska, J., Krokoszynska, I., Czapinska, H., Watorek, W., Dadlez, M., and Otlewski, J. (1998) Specificity of human cathepsin G. *Biochim. Biophys. Acta* **1386,** 189–198
48. Rehault, S., Brillard-Bourdet, M., Juliano, M. A., Juliano, L., Gauthier, F., and Moreau, T. (1999) New, sensitive fluorogenic substrates for human cathepsin G based on the sequence of serpin-reactive site loops. *J. Biol. Chem.* **274,** 13810–13817
49. Goel, M. S., and Diamond, S. L. (2003) Neutrophil cathepsin G promotes prothrombinase and fibrin formation under flow conditions by activating fibrinogen-adherent platelets. *J. Biol. Chem.* **278,** 9458–9463
50. Tanaka, T., Minematsu, Y., Reilly, C. F., Travis, J., and Powers, J. C. (1985) Human leukocyte cathepsin G. Subsite mapping with 4-nitroanilides, chemical modification, and effect of possible cofactors. *Biochemistry* **24,** 2040–2047
51. McRae, B., Nakajima, K., Travis, J., and Powers, J. C. (1980) Studies on reactivity of human leukocyte elastase, cathepsin G, and porcine pancreatic elastase toward peptides including sequences related to the reactive site of α 1-protease inhibitor (α1-antitrypsin). *Biochemistry* **19,** 3973–3978
52. Nakajima, K., Powers, J. C., Ashe, B. M., and Zimmerman, M. (1979) Mapping the extended substrate binding site of cathepsin G and human leukocyte elastase. Studies with peptide substrates related to the α1-protease inhibitor reactive site. *J. Biol. Chem.* **254,** 4027–4032
53. Hof, P., Mayr, I., Huber, R., Korzus, E., Potempa, J., Travis, J., Powers, J. C., and Bode, W. (1996) The 1.8 A crystal structure of human cathepsin G in complex with Suc-Val-Pro-PheP-(OPh)2: A Janus-faced proteinase with two opposite specificities. *EMBO J.* **15,** 5481–5491
54. Pauly, T. A., Sulea, T., Ammirati, M., Sivaraman, J., Danley, D. E., Griffor, M. C., Kamath, A. V., Wang, I. K., Laird, E. R., Seddon, A. P., Menard, R., Cygler, M., and Rath, V. L. (2003) Specificity determinants of human cathepsin s revealed by crystal structures of complexes. *Biochemistry* **42,** 3203–3213
55. Berger, A., and Schechter, I. (1970) Mapping the active site of papain with the aid of peptide substrates and inhibitors. *Philos. Trans R. Soc. Lond. B. Biol. Sci.* **257,** 249–264
56. Barrett, A. J. (1980) Fluorimetric assays for cathepsin B and cathepsin H with methylcoumarylamide substrates. *Biochem. J.* **187,** 909–912
57. Menard, R., Carmona, E., Plouffe, C., Bromme, D., Konishi, Y., Lefebvre, J., and Storer, A. C. (1993) The specificity of the S1′ subsite of cysteine proteases. *FEBS Lett.* **328,** 107–110
58. Bromme, D., Bonneau, P. R., Lachance, P., and Storer, A. C. (1994) Engineering the S2 subsite specificity of human cathepsin S to a cathepsin L- and cathepsin B-like specificity. *J. Biol. Chem.* **269,** 30238–30242
59. Portaro, F. C., Santos, A. B., Cezari, M. H., Juliano, M. A., Juliano, L., and Carmona, E. (2000) Probing the specificity of cysteine proteinases at subsites remote from the active site: Analysis of P4, P3, P2′ and P3′ variations in extended substrates. *Biochem. J.* **347 (Pt 1),** 123–129
60. Turk, D., and Guncar, G. (2003) Lysosomal cysteine proteases (cathepsins): Promising drug targets. *Acta Crystallogr. D. Biol. Crystallogr.* **59,** 203–213
61. Lecaille, F., Kaleta, J., and Bromme, D. (2002) Human and parasitic papain-like cysteine proteases: Their role in physiology and pathology and recent developments in inhibitor design. *Chem. Rev.* **102,** 4459–4488
62. Maciewicz, R. A., and Etherington, D. J. (1988) A comparison of four cathepsins (B, L, N and S) with collagenolytic activity from rabbit spleen. *Biochem. J.* **256,** 433–440
63. Hasnain, S., Hirama, T., Huber, C. P., Mason, P., and Mort, J. S. (1993) Characterization of cathepsin B specificity by site-directed mutagenesis. Importance of Glu^{245} in the S2-P2 specificity for arginine and its role in transition state stabilization. *J. Biol. Chem.* **268,** 235–240
64. Jia, Z., Hasnain, S., Hirama, T., Lee, X., Mort, J. S., To, R., and Huber, C. P. (1995) Crystal structures of recombinant rat cathepsin B and a cathepsin B-inhibitor complex. Implications for structure-based inhibitor design. *J. Biol. Chem.* **270,** 5527–5533
65. Musil, D., Zucic, D., Turk, D., Engh, R. A., Mayr, I., Huber, R., Popovic, T., Turk, V., Towatari, T., Katunuma, N., et al. (1991). The refined 2.15 A X-ray crystal structure of human liver cathepsin B: The structural basis for its specificity. *EMBO J.* 10, 2321–2330
66. Taralp, A., Kaplan, H., Sytwu, II, Vlattas, I., Bohacek, R., Knap, A. K., Hirama, T., Huber, C. P., and Hasnain, S. (1995) Characterization of the S3 subsite specificity of cathepsin B. *J. Biol. Chem.* **270,** 18036–18043
67. Xin, X. Q., Gunesekera, B., and Mason, R. W. (1992) The specificity and elastinolytic activities of bovine cathepsins S and H. *Arch Biochem. Biophys.* **299,** 334–339
68. Lecaille, F., Choe, Y., Brandt, W., Li, Z., Craik, C. S., and Bromme, D. (2002) Selective inhibition of the collagenolytic activity of human cathepsin K by altering its S2 subsite specificity. *Biochemistry* **41,** 8447–8454
69. Alves, M. F., Puzer, L., Cotrin, S. S., Juliano, M. A., Juliano, L., Bromme, D., and Carmona, A. K. (2003) S3 to S3′ subsite specificity of recombinant human cathepsin K and development of selective internally quenched fluorescent substrates. *Biochem. J.* **373,** 981–986
70. Puzer, L., Cotrin, S. S., Alves, M. F., Egborge, T., Araujo, M. S., Juliano, M. A., Juliano, L., Bromme, D., and Carmona, A. K. (2004) Comparative substrate specificity analysis of recombinant human cathepsin V and cathepsin L. *Arch. Biochem. Biophys.* **430,** 274–283
71. Bromme, D., Steinert, A., Friebe, S., Fittkau, S., Wiederanders, B., and Kirschke, H. (1989) The specificity of bovine spleen cathepsin S. A comparison with rat liver cathepsins L and B. *Biochem. J.* **264,** 475–481
72. Somoza, J. R., Zhan, H., Bowman, K. K., Yu, L., Mortara, K. D., Palmer, J. T., Clark, J. M., and McGrath, M. E. (2000) Crystal structure of human cathepsin V. *Biochemistry* **39,** 12543–12551
73. Bromme, D., Li, Z., Barnes, M., and Mehler, E. (1999) Human cathepsin V functional expression, tissue distribution, electrostatic surface potential, enzymatic characterization, and chromosomal localization. *Biochemistry* **38,** 2377–2385
74. Harris, J. L., Backes, B. J., Leonetti, F., Mahrus, S., Ellman, J. A., and Craik, C. S. (2000) Rapid and general profiling of protease specificity by using combinatorial fluorogenic substrate libraries. *Proc. Natl. Acad. Sci. U. S. A.* **97,** 7754–7759
75. St Hilaire, P. M., Willert, M., Juliano, M. A., Juliano, L., and Meldal, M. (1999) Fluorescence-quenched solid phase combinatorial libraries in the characterization of cysteine protease substrate specificity. *J. Comb Chem.* **1,** 509–523
76. Rowan, A. D., Buttle, D. J., and Barrett, A. J. (1990) The cysteine proteinases of the pineapple plant. *Biochem. J.* **266,** 869–875
77. Vijayendran, R. A., and Leckband, D. E. (2001) A quantitative assessment of heterogeneity for surface-immobilized proteins. *Anal Chem.* **73,** 471–480
78. Johnson, C. P., Jensen, I. E., Prakasam, A., Vijayendran, R., and Leckband, D. (2003) Engineered protein a for the orientational control of immobilized

proteins. *Bioconjug. Chem.* **14,** 974–978

79. Zaks, A., and Klibanov, A. M. (1988) The effect of water on enzyme action in organic media. *J. Biol. Chem.* **263,** 8017–8021
80. Xavier, K. A., Shick, K. A., Smith-Gil, S. J., and Willson, R. C. (1997) Involvement of water molecules in the association of monoclonal antibody HyHEL-5 with bobwhite quail lysozyme. *Biophys. J.* **73,** 2116–2125
81. Klibanov, A. M. (1997) Why are enzymes less active in organic solvents than in water? *Trends Biotechnol.* **15,** 97–101
82. Gekko, K., and Timasheff, S. N. (1981) Mechanism of protein stabilization by glycerol: Preferential hydration in glycerol-water mixtures. *Biochemistry* **20,** 4667–4676
83. Sichler, K., Banner, D. W., D'Arcy, A., Hopfner, K. P., Huber, R., Bode, W., Kresse, G. B., Kopetzki, E., and Brandstetter, H. (2002) Crystal structures of uninhibited factor VIIa link its cofactor and substrate-assisted activation to specific interactions. *J. Mol. Biol.* **322,** 591–603
84. Sturzebecher, J., Kopetzki, E., Bode, W., and Hopfner, K. P. (1997) Dramatic enhancement of the catalytic activity of coagulation factor IXa by alcohols. *FEBS Lett.* **412,** 295–300