DOI 10.1002/pmic.200401011 *Proteomics* 2005, *5,* 1292–1298

REGULAR ARTICLE

Profiling serine protease substrate specificity with solution phase fluorogenic peptide microarrays

Dhaval N. Gosalia[1], Cleo M. Salisbury[2]*, Dustin J. Maly[2], Jonathan A. Ellman[2]* and Scott L. Diamond[1, 3]***

[1] Department of Bioengineering, Institute for Medicine and Engineering, University of Pennsylvania, PA, USA
[2] Center for New Directions in Organic Synthesis, Department of Chemistry, University of California, Berkeley, CA, USA
[3] Department of Chemical and Biomolecular Engineering, Institute for Medicine and Engineering, University of Pennsylvania, PA, USA

Received: June 1, 2004
Accepted: September 3, 2004

A novel microarray-based proteolytic profiling assay enabled the rapid determination of protease substrate specificities with minimal sample and enzyme usage. A 722-member library of fluorogenic protease substrates of the general format Ac-Ala-X-X-(Arg/Lys)-coumarin was synthesized and microarrayed, along with fluorescent calibration standards, in glycerol nanodroplets on microscope slides. The arrays were then activated by deposition of an aerosolized enzyme solution, followed by incubation and fluorometric scanning. The specificities of human blood serine proteases (human thrombin, factor Xa, plasmin, and urokinase plasminogen activator) were examined. The arrays provided complete maps of protease specificity for all of the substrates tested and allowed for detection of cooperative interactions between substrate subsites. The arrays were further utilized to explore the conservation of thrombin specificity across species by comparing the proteolytic fingerprints of human, bovine, and salmon thrombin. These enzymes share nearly identical specificity profiles despite ~390 million years of divergent evolution. Fluorogenic substrate microarrays provide a rapid way to determine protease substrate specificity information that can be used for the design of selective inhibitors and substrates, the study of evolutionary divergence, and potentially, for diagnostic applications.

Keywords:
Combinatorial libraries / Microarray / Microfluidics / Protease profiling / Substrate specificity

1 Introduction

Proteases are one of the largest protein families in organisms from *Escherichia coli* to humans [1–3]. Due to their critical roles in hormone activation, proteasomal degradation, apoptosis, and other biological pathways, proteases are essential for cellular function and viability. Furthermore, the function of proteases is important in diverse diseases, including viral (*e.g.*, human immunodeficiency virus, severe acute respiratory syndrome) and parasitic (*e.g.*, malaria) infections, as well as cardiovascular disease, cancer, and Alzheimer's disease [4, 5]. Understanding these proteases and the many new proteases that have been identified through recent genomic and proteomic efforts will provide insight into biological systems and will likely provide a number of important new therapeutic targets [4]. A key aspect of proteolytic pathways is the ability of

Correspondence: Dr. Jonathan A. Ellman, Department of Chemistry, Berkeley CA 94720–1460, USA
E-mail: jellman@uclink.berkeley.edu
Fax: +1-510-642-8369

Abbreviations: ACC, 7-amino-4-carbamoylmethylcoumarin; **Fmoc**, 9-fluorenylmethylcarbonyl; **PS-SCL**, positional scanning-synthetic combinatorial library; **uPA**, urokinase plasminogen activator

* These authors contributed equally.
** Additional corresponding author: Dr. Scott Diamond, E-mail: sld@seas.upenn.edu

proteases to preferentially cleave target substrates in the presence of other proteins. Substrate and cation exosite binding, protease localization, and temporal expression can all contribute to substrate specificity. However, one of the key factors is the complementarity of the enzyme active site with the residues of the substrate. Determining the residues that represent the preferred cleavage site provides a functional mapping of the active site and facilitates the identification of the physiological substrates of the protease. Furthermore, determination of substrate specificity also provides a framework for the design of potent and selective inhibitors [6].

One extensively used method for rapidly accessing protease specificity is positional scanning-synthetic combinatorial libraries (PS-SCLs) of fluorogenic peptidyl coumarin substrates [7–16]. Proteolytic cleavage of these substrates at the peptide-anilide bond liberates the highly fluorescent coumarin leaving group, allowing for the determination of cleavage rates through the increase in fluorescence (Fig. 1) and determination of preferred residues on the P side of the substrate (the standard nomenclature is to refer to the substrate residues as P_n, $P_{n-1}\ldots P_2$, P_1, $P_{1'}$, $P_{2'}\ldots P_{m-1'}$, $P_{m'}$ (where amide bond hydrolysis occurs between P_1 and $P_{1'}$) and to denote the corresponding enzyme binding sites as S_n, $S_{n-1}\ldots S_2$, S_1, $S_{1'}$, $S_{2'}\ldots S_{m-1'}$, $S_{m'}$ [17]). PS-SCLs have been widely used to determine the preferred cleavage motifs of numerous serine and cysteine proteases, including the caspase family [7], the human proteasome [14], and proline-specific dipeptidyl peptidases [15]. However, positional scanning libraries are limited by the fact that each library member consists of a mixture of substrates. This limitation was underscored in a study of the serine protease MTSP-1 by Takeuchi *et al*. [8]. They found that PS-SCLs for the P_3 and P_4 sites suggested that Arg and Lys were the optimal residues, but the enzyme did not efficiently cleave single substrates with basic residues at both P_3 and P_4. Using phage display, it was determined that basic residues are preferred at P_3 *or* P_4, but not at both sites. The cooperative interactions between residues in the substrate cannot be assessed with PS-SCL because of its mixture-based nature.

To fully access cooperativity while maintaining speed and efficiency in assaying large substrate libraries, we previously developed spatially addressed microarrays where substrates were chemoselectively and covalently bound to the slide surface [18]. This strategy takes advantage of the miniaturized nature of the array to minimize substrate and enzyme consumption. The arrays have been used to provide a complete fingerprint of the P_2–P_3 specificity of the serine protease thrombin. A solution-phase version of the microarrays could further increase the efficiency of the assay by eliminating the synthetic steps required for the derivatization of the substrates and slides, and could also reduce any enzyme accessibility issues caused by linking the substrate to the surface. Here we report the development of nanodroplet microarrays [19], where substrates suspended in glycerol droplets were treated with aerosolized aqueous enzyme solutions to allow profiling of substrate specificity.

2 Materials and methods

2.1 Proteases

Purified human thrombin, human plasmin, human factor Xa, bovine thrombin (Enzyme Research Laboratories, South Bend, IN, USA) and human urokinase plasminogen activator (uPA; American Diagnostica, Stamford, CT, USA) were stored according to the manufacturers' instructions. Salmon thrombin was provided by Dr. Paul Janmey (University of Pennsylvania, PA, USA).

2.2 Fluorogenic substrate library

A complete procedure for library synthesis appears in the supplementary material. Briefly, Rink amide polystyrene resin was deprotected, and then acylated with 9-fluorenylmethylcarbonyl-protected (Fmoc) 7-amino-4-carbamoylmethylcoumarin (ACC-OH) [20]. The Fmoc group was removed, and the P_1 amino acid was loaded under highly activating conditions. Following a capping step, the resin was distributed into the wells of 96-well Robbins blocks (Flexchem System; Robbins Scientific, Sunnyvale, CA, USA), and the substrate synthesis was completed by standard Fmoc-based solid-phase peptide synthesis techniques [21]. After cleavage from the resin, the substrates were analyzed by HPLC-MS to demonstrate acceptable purity and reconstituted with DMSO.

2.3 Single substrate synthesis

Single substrates were prepared as previously described [11] and purified by reverse-phase preparatory HPLC, followed by lyophilization. To assure purity, the substrates were analyzed

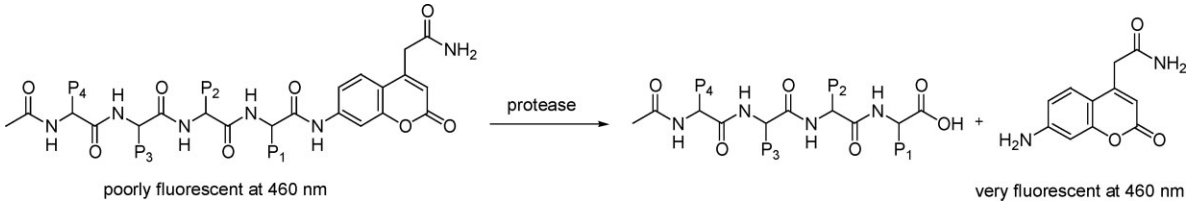

Figure 1. Overall design of fluorogenic substrates.

by HPLC-MS (column: C18 (4.6 × 100 mm), conditions: CH_3CN/H_2O-0.1% TFA, 5–95% over 14 min, 0.4 mL/min, detection at 220/254/280 nm for 22 min). The analytical results are as follows: Ac-FFGR-ACC-NH_2 (t_R = 8.4 min, *m/z* = 768.5), Ac-AFGK-ACC-NH_2 (t_R = 7.0 min, *m/z* = 664.4), Ac-AFFK-ACC-NH_2 (t_R = 8.4 min, *m/z* = 754.5), Ac-FFFR-ACC-NH_2 (t_R = 9.5 min, *m/z* = 858.5), Ac-AFGR-ACC-NH_2 (t_R = 7.3 min, *m/z* = 692.4) Ac-FFFK-ACC-NH_2 (t_R = 9.3 min, *m/z* = 830.5), Ac-FFGK-ACC-NH_2 (t_R = 8.4 min, *m/z* = 740.5), Ac-AFFR-ACC-NH_2 (t_R = 8.8 min, *m/z* = 782.5), Ac-AEFK-ACC-NH_2 (t_R = 6.8 min, *m/z* = 736.2), Ac-AEFR-ACC-NH_2 (t_R = 7.1 min, *m/z* = 764.2).

2.4 Microarray printing

An OmniGrid Accent (Gene Machines, San Carlos, CA, USA) was used for arraying. The P_1 = Arg and P_1 = Lys sublibraries, along with calibration standards, were printed at either 50 or 100 μM concentration in a 16 × 24 format equivalent to a 384-well plate using a 1 × 1 pin protocol on polylysine coated glass slides (Erie Scientific, Portsmouth, NH, USA). All slides were washed in dry ethanol and vacuum-dried before arraying. The average size of the spots with SMP-4 Stealth pins (Telechem, Sunnyvale, CA, USA) using 50/50 by volume glycerol/DMSO was 200 μm in diameter, with a corresponding volume of 1.6 nL *per* spot as determined by differential interference contrast microscopy of the height of the droplet. Calibration standards (unacylated ACC, acetyl-capped ACC, and blanks) were printed on each array, to enable quantification and normalization of fluorescence intensity between slides. The center-to-center spacing for each spot was 500 μm with a resulting spot density of 400 spots/cm^2. Arraying was performed in a dark room at 45% humidity and the slides were stored at −20°C in the dark until use. Slides were stored for up to three months with no noticeable loss in activity.

2.5 Protease assay using fluorogenic substrate microarrays

Proteases were reconstituted and diluted in buffer as suggested by the vendors (human thrombin (10 μM in pH 6.5, 50 mM sodium citrate, 200 mM NaCl, 0.1% PEG-8000); bovine thrombin (10 μM in pH 6.5, 50 mM sodium citrate, 200 mM NaCl, 0.1% PEG-8000); salmon thrombin (5 μM in pH 6.5, 50 mM sodium citrate, 200 mM NaCl, 0.1% PEG-8000); factor Xa (20 μM in pH 7.4, 20 mM Tris, 700 mM NaCl); plasmin (10 μM in pH 8.5, 100 mM HEPES, 100 mM NaOAc); uPA (10 μM in pH 7.2, 150 mM PBS)). The 100 μM substrate concentration arrays were used for the uPA assays; all other assays employed the 50 μM arrays. Substrate concentrations were determined to be below the K_m for efficient substrates (see supplementary material). The proteases were delivered to the array as previously described [19] *via* a 120 kHz ultrasonic nozzle (Sonotek, Milton, NY, USA). The liquid samples were aerosolized at a liquid flow rate of 400 nL/s into the nozzle using a UMPII flow pump (World Precision Instruments, Saratoga, FL, USA) to achieve an addition of ~0.05 nL of enzyme solution to each spot and a ~30–fold dilution of the enzyme concentration. The slides (six replicates *per* enzyme) were incubated at 37°C for 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. Slides were scanned using Alpha Array from Alpha Innotech (San Leandro, CA, USA), with exposure times for all slides of 2500 ms (Ex: 405/40 nm, Em: 475/40 nm). Images were acquired in a 16–bit format and the analysis and presentation of the data were performed by using Array Vision (Imaging Research, Ontario, Canada) and Cluster and Treeview [22].

2.6 Single-substrate kinetic assays in well plates

Substrate stock solutions were prepared in DMSO and diluted in assay buffer (25 mM Tris pH 7.2, 150 mM NaCl, 5 mM $CaCl_2$, and 0.01% Tween) to a final concentration determined to be below the K_m for efficient substrates (see supplementary material). The final concentration of substrate was 75 μM and the concentration of DMSO in the assay was less than 5%. The factor Xa concentration was 12 nM and the assays were performed in triplicate. Hydrolysis of ACC substrates was monitored fluorometrically (Ex: 380 nm, Em: 460nm) on a Fluoromax-2 spectrofluorimeter (Molecular Devices, Sunnyvale, CA, USA).

3 Results

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). The substrate specificities of the serine proteases thrombin, factor Xa, plasmin, and uPA were profiled using the microarray-based format.

The P_1 = Lys sublibrary was treated with human, bovine, and salmon thrombin. The enzymes all show very strong substrate specificity for Pro at the P_2 position and a broader specificity at the P_3 position (Fig. 2), which is consistent with previous peptidyl-coumarin microarray data [18], as well as with PS-SCL [11, 16] and single substrate data [23].

Treatment of the P_1 = Lys and P_1 = Arg sublibraries with Factor Xa showed a strong preference for Phe, Gly, and Ser at the P_2 position (Fig. 3). Because previous studies [9, 16, 24, 25] have reported varying results regarding the P_2 specificity of Factor Xa, we further investigated the substrate specificity of the enzyme to see whether cooperative interactions between substrate subsites could be responsible for these dis-

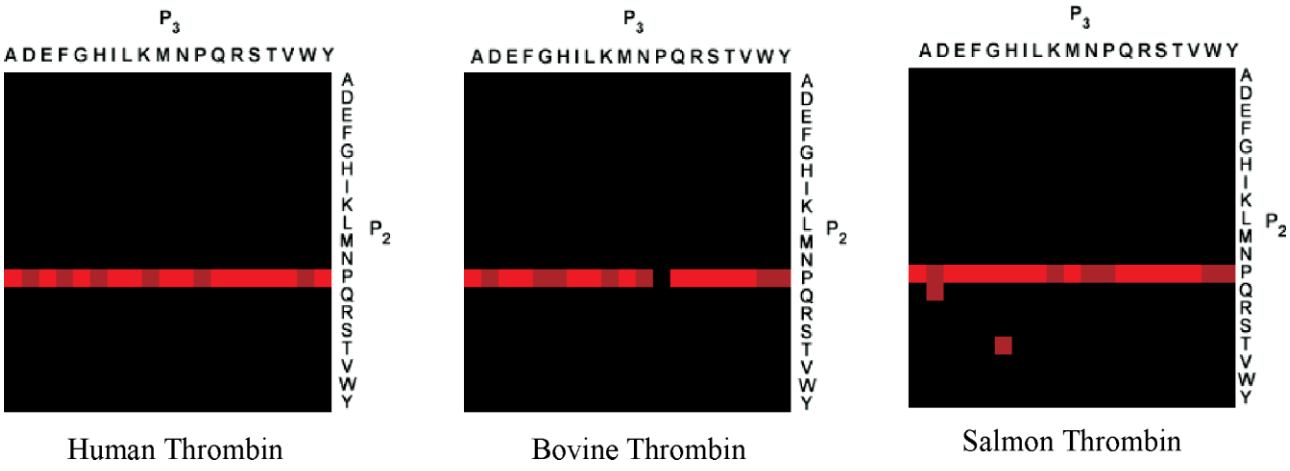

Figure 2. Characterization of the substrate specificity of human, bovine, and salmon thrombin, using the Ac-Ala-P_3-P_2-Lys-ACC-NH_2 sublibrary microarrays. Each square is colored in proportion to the average quantitated fluorescence intensity of the corresponding substrates of six replicate arrays after treatment with the enzyme, indicating the relative amount of cleavage.

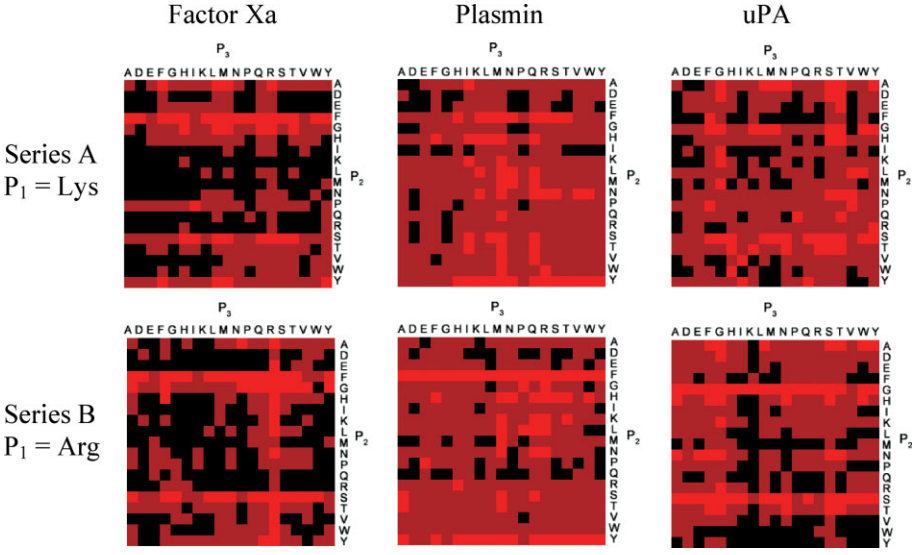

Figure 3. Characterization of the specificity of factor Xa, plasmin, and uPA using the Ac-Ala-P_3-P_2-Lys-ACC-NH_2 and Ac-Ala-P_3-P_2-Arg-ACC-NH_2 substrate microarrays. Each square in the grid is colored in proportion to the average quantitated fluorescence intensity of the corresponding substrates of six replicate arrays, after treatment with the enzyme, indicating the relative amount of cleavage.

parities. A series of substrates was prepared, purified, and screened in a standard microtiter plate assay. The data for the well plate assays (Table 1) confirmed both the presence of extensive cooperative interactions between substrate residues and the preferences observed in the microarray assay. The P_1 = Arg arrays also showed a strong preference for a P_3 Arg residue, consistent with the observation from a phage display library, which reported isolation of multiple clones containing Arg at the P_1 and P_3 sites [26].

Treatment of both the P_1 = Lys and P_1 = Arg sublibraries with plasmin showed a strong preference at the P_2 position for the aromatic residues Phe and Tyr, in addition to His (and Asn for the P_1 = Lys sublibrary). Met and Gln (and Arg for the P_1 = Arg sublibrary) are the preferred residues in the P_3 position (Fig. 3). The data is consistent with the previously published PS-SCL data [9, 11] and the known physiologic substrates of plasmin [11]. For both substrate sublibraries, uPA showed a preference for small amino acids at both the P_2 and P_3 positions (P_2 = Ala, Gly, Ser, and Thr and P_3 = Gly, Ser, and Thr). The results are consistent with previously reported phage display libraries [27] and PS-SCLs [9].

4 Discussion

4.1 Peptidyl-coumarin microarrays rapidly provide the substrate specificity of serine proteases

Proteases play an important role in all organisms. To completely understand their functional roles and biological importance, a better understanding of their substrate specificity is necessary. Elucidating the substrates of hundreds of distinct enzymes is a challenging task, which

Table 1. Relative k_{cat}/K_m values for the hydrolysis of solution phase peptidyl coumarin substrates by factor Xa

Substrate	Normalized k_{cat}/K_m
Ac-AFFR-ACC-NH_2	1.00 ± 0.05
Ac-AFGR-ACC-NH_2	0.23 ± 0.01
Ac-FFGR-ACC-NH_2	0.27 ± 0.01
Ac-FFFR-ACC-NH_2	0.63 ± 0.03
Ac-AEFR-ACC-NH_2	0.11 ± 0.01
Ac-AFFK-ACC-NH_2	0.13 ± 0.01
Ac-AFGK-ACC-NH_2	No cleavage
Ac-FFGK-ACC-NH_2	No cleavage
Ac-FFFK-ACC-NH_2	0.08 ± 0.01
Ac-AEFR-ACC-NH_2	No cleavage

Purified single substrates were assayed in solution in triplicate. Assays were performed with 75 μM substrate and 12 nM enzyme. The k_{cat}/K_m value for the hydrolysis of the most efficiently cleaved substrate, Ac-AFFR-ACC-NH_2, was set to 1.00.

may be expedited by microarray technology. Microarray-based substrate specificity determination allows for rapid profiling of different proteases in a high-density/high-throughput format. Large libraries containing thousands of substrates can be routinely spatially arrayed on a microscope slide and hundreds of slides can be simultaneously printed and stored for future use. Furthermore, since this approach uses < 4 μL of enzyme solution *per* slide, the assay requires only minimal enzyme consumption. The nanoliter volume reaction assays allow profiling of numerous proteases with minimal sample usage, thus conserving both enzymes and substrates.

The serine proteases profiled here were chosen because they play an important role in a variety of physiologic processes including blood coagulation, fibrinolysis and tissue remodeling. Furthermore, while the specificity of these enzymes has been well studied [9, 11, 16, 23–31], a complete mapping of all cooperative interactions between the P_2 and P_3 sites has only recently been established for human thrombin [18].

4.2 Thrombin substrate specificity shows functional conservation across species

Thrombin plays a critical role in blood coagulation and vascular integrity. The importance of the enzyme is highlighted by the conservation of the enzyme across multiple species. Sequencing studies report 87.3% homology between the active B chain of human and bovine thrombin [32]. While salmon thrombin has not yet been sequenced, thrombin from the slightly more evolved fish, rainbow trout, shares 68.6% homology to that of humans, and 68.8% to that of cows [32]. The data from the microarrays provides an additional means to study this evolutionary conservation by focusing on functional comparisons of the proteases. The profiles of the three thrombin variants show stringency in their substrate specificities for Pro at the P_2 position, giving evidence of a highly conserved active site, despite the nearly 390 million years of evolutionary divergence between salmon and humans. This functional conservation observed in the microarray studies is supported by a previous report by Salte *et al.* [33] which demonstrated that salmon antithrombin inhibits human and salmon thrombin with equal efficiency.

The observed preference for P_2 proline is consistent with the physiologic substrates of thrombin including factors V, VII, XI, XII, Protein C and uPA [11]. Crystallographic studies of the active site of thrombin have provided an explanation for the strict P_2 Pro preference. Bode *et al.* [34] showed that an insertion of seven amino acids from the 60s loop at the S_2 site creates a rigid pocket into which P_2 proline can be modeled. The functional and sequence homology can provide valuable insight into the molecular evolution of complex biological processes.

4.3 Factor Xa demonstrates extensive cooperative interactions between substrate subsites

Factor Xa is a blood coagulation enzyme that serves to activate prothrombin and factor VII. While this enzyme has been extensively studied [9, 16, 24–26], the inherent specificity of the enzyme remains controversial. The physiologic substrates of factor Xa (prothrombin, factor VII, and the factor Xa autolysis loop) have a P_1 = Arg, P_2 = Gly motif [35], and several studies using synthetic substrates, including a PS-SCL [9] and a panel of *p*-nitroanilide substrates [24], have shown a distinct preference for Gly at the P_2 position. However, a tripeptide PS-SCL [16] and a recent report using a series of substrates based on the sequences donor-VFGRSLEDQ-quencher and Ac-VQXR-*p*-nitroanilide show a preference for Phe at the P_2 position [25]. Our microarray data suggests that there are considerable cooperative interactions between substrate subsites for Factor Xa, and that the variations observed in the previous studies may be due to this cooperativity. To further investigate this possibility, a series of substrates containing Arg and Lys at the P1 site and Gly and Phe at the P_2 site were synthesized, purified, and assayed in solution (Table 1). The most efficiently cleaved substrates contained Phe at the P_2 position, but the presence of a P_2 Phe was neither sufficient nor necessary for high cleavage efficiencies (*e.g.*, no cleavage of Ac-AEFK-ACC-NH_2 was observed, while Ac-AFGR-ACC-NH_2 was efficiently proteolyzed). The presence of cooperative interactions between substrate subsites was also observed with these substrates. Switching the P_4 Ala in the sequence Ac-AFFR-ACC to Phe resulted in a decrease in activity of 40%, while the same switch for the sequence Ac-AFGR-ACC resulted in a slight increase in cleavage efficiency, indicating cooperative effects between the P_2 and P_4 sites. Similarly, changing the P_2 Phe in the sequence Ac-AFFR-ACC to Gly resulted in a four-fold decrease in cleavage efficiency while the same switch in the sequence Ac-FFFR-ACC resulted in a decrease of only two-fold. Overall, the solution phase assays confirmed the array

results: for substrates of the format Ac-Ala-P_3-P_2-(Arg/Lys)-ACC-NH_2, P_2 Phe substrates were typically preferred over the corresponding P_2 Gly sequences, but peptides with the physiologically observed P_1 = Arg, P_2 = Gly sequence motif can serve as substrates, depending on the complete sequence.

The cooperativity seen here between substrate subsites may be an effect of ligand-induced motion at the active site of the enzyme. Crystallographic studies of factor Xa show significant changes in the backbone and side chain geometries at the S_1 site upon ligand binding [35]. Furthermore, the studies indicate that there is also a degree of flexibility at the S_2 site. In the low energy conformation of factor Xa, Tyr99 blocks the S_2 site, explaining the observed preference at the P_2 site for Gly, which has no side chain, in the natural substrates. However, Engh *et al.* [35] noted that the Tyr99 side chain and the conformation at the S_2 site are not absolutely rigid, thus potentially explaining the efficient cleavage of P_2 Phe substrates. The observed P_3 specificity in both physiological substrates and the microarray data show a preference for hydrophilic residues, and a phage display study reported the isolation of multiple clones with Arg in the P_3 site [26]. The crystal structure, showing a solvent-exposed S_3 site, also supports the observed preference for charged and hydrophobic residues at the P_3 position. While the crystal structures only suggest the potential for cooperativity between the subsites, the microarray, confirmed by well plate assays of purified substrates, showed the critical nature of obtaining complete specificity profiles.

4.4 Plasmin has a strong preference for P_2 aromatic residues

Plasmin mediates fibrinogenolysis and fibrinolysis. The microarray data shows a preference for aromatic residues at the P_2 site. This is consistent with the known and postulated physiologic substrates of plasmin: osteocalcin (P_2 = Tyr), vitronectin (P_2 = Tyr), PAR1 (P_2 = Tyr), and factor Xa (P_2 = Phe) [11]. The preference for a P_2 aromatic side chain was also observed in the PS-SCL data [11], which show a P_2 preference for Phe and Tyr, followed by Trp at this position. These observations have been explained by the existence of a pocket, seen in the crystal structure, specific for an aromatic residue [11]. Takeuchi *et al.* [8] suggest that the δ+ aryl ring protons of the substrate interact with a Glu residue that extends from above the S_2 site toward the P_2 residue. They further suggest that the δ+ amide group of a nearby Gln residue can make contact with the δ- electrons of the aryl system to provide exquisite selectivity for aromatic residues at the P_2 site.

4.5 uPA shows stringency for small polar residues at P_2 and P_3

The physiological role of uPA is to activate plasmin by cleaving the zymogen plasminogen, at the Arg-Val bond within the sequence CPGR-VVGG. The presence of small polar residues at the P_2 and P_3 sites on the natural substrate correlates with the observed specificity for uPA on the microarray. The molecular basis for this preference for small amino acids can be understood from the crystal structure [28]. The size of the S_2 site is restricted by the bulky His99 residues (chymotrypsin numbering) and the site that corresponds to the aryl binding site of thrombin and is hypothesized to serve as the S_3 site, is also obstructed, being partially filled by the Leu97B and Thr97A side chains. In recent years, uPA has served as a potential drug target due to its association with tumor progression. Mice deficient in uPA are resistant to tumor progression in several tumor types [36] and uPA is known to be up-regulated in a number of cancers including squamous cell carcinomas [37] and colon cancers [38]. Furthermore, uPA has been implicated as a useful diagnostic for disease characterization in liver cancer [39] and as a prognostic indicator for survival [40] and treatment success [41] in breast cancer. As such, determining the complete substrate specificity of the enzyme could provide useful information for the design of selective substrates for uPA detection and quantitation and, potentially, for small molecule drug design.

5 Concluding remarks

The data shown here demonstrate that nanodroplet microarrays can be used to determine substrate specificity. At the high density that this 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) could be accommodated on only two slides. The advantage of this approach over the covalent microarray methodology is that neither the substrates nor the slides require derivatization, and that the risk of protein aggregation on the surface is minimized. While the use of glycerol in proteolytic assays could potentially alter substrate specificity [42] or slow reaction rates [43], the correlations seen in this study between the two microarrays approaches, and between the factor Xa array and well plate data indicates that glycerol does not dramatically alter the observed protease specificities. In conclusion, we have presented a solution phase microarray approach that could be applied to the elucidation of the protease–substrate degradome [44] to advance the understanding of the functional and systemic importance of hundreds of distinct enzymes. This protease profiling method allows determination of important cooperative interactions between substrate subsites, provides a fast means to access functional evolutionary data, and can be applied to the design of potent and specific substrates and inhibitors. Furthermore, the assay could be amenable to drug discovery, metabolic/toxicity testing, and, potentially, diagnostic applications.

D. N. G. would like to acknowledge Dr. Anthony G. Passerini for helpful conversations on microarray data analysis. C. M. S. would like to acknowledge Dr. Peter L. Steck for developing the synthesis of Fmoc-Arg(Pbf)-Cl. J. A. E. acknowledges support

from the NIH (grant #GM54051) and S. L. D. acknowledges support from the NIH (grant #56621). The Center for New Directions in Organic Synthesis is supported by Bristol-Myers Squibb as a Sponsoring Member and Novartis as a Supporting Member.

6 References

[1] Southan, C., *J. Pept. Sci.* 2000, *6*, 453–458.

[2] Rawlings, N. D., O'Brien, E. A., Barrett, A. J., *Nucleic Acids Res.* 2002, *30*, 343–346.

[3] Puente, X. S., Sanchez, L. M., Overall, C. M., Lopez-Otin, C., *Nat. Rev. Genet.* 2003, *4*, 544–557.

[4] Leung, D., Abbenante, G., Fairlie, D. P., *J. Med. Chem.* 2000, *43*, 305–341.

[5] Anand, K., Ziebuhr, J., Wadhwani, P., Mesters, J. R., Holgenfeld, R., *Science* 2003, *300*, 1763–1767.

[6] Maly, D. J., Huang, L., Ellman, J. A. *Chembiochem* 2002, *3*, 16–37.

[7] Thornberry, N. A., Rano, T. A., Peterson, E. P., Rasper, D. M. *et al.*, *J. Biol. Chem.* 1997, *272*, 17907–17911.

[8] Takeuchi, T., Harris, J. L., Huang, W., Yan, K. W. *et al.*, *J. Biol. Chem.* 2000, *275*, 26333–26342.

[9] Harris, J. L., Backes, B. J., Leonetti, F., Mahrus, S. *et al.*, *Proc. Natl. Acad. Sci. USA* 2000, *97*, 7754–7759.

[10] Harris, J. L., Niles, A., Burdick, K., Maffitt, M. *et al.*, *J. Biol. Chem.* 2001, *276*, 34941–34947.

[11] Backes, B. J., Harris, J. L., Leonetti, F., Craik, C. S., Ellman, J. A., *Nat. Biotechnol.* 2000, *18*, 187–193.

[12] Dauber, D. S., Ziermann, R., Parkin, N., Maly, D. J. *et al.*, *J. Virol.* 2002, *76*, 1359–1368.

[13] Harris, J. L., Peterson, E. P., Hudig, D., Thornberry, N. A., Craik, C. S., *J. Biol. Chem.* 1998, *273*, 27364–27373.

[14] Harris, J. L., Alper, P. B., Li, J., Rechsteiner, M., Backes, B. J., *Chem. Biol.* 2001, *8*, 1131–1141.

[15] Leiting, B., Pryor, K. D., Wu, J. K., Marsilio, F. *et al.*, *Biochem. J.* 2003, *371*, 525–532.

[16] Furlong, S. T., Mauger, R. C., Strimpler, A. M., Liu, Y.-P. *et al.*, *Bioorg. Med. Chem.* 2002, *10*, 3637–3647.

[17] Schechter, I., Berger, A., *Biochem. Biophys. Res. Commun.* 1967, *27*, 157–162.

[18] Salisbury, C. M., Maly, D. J., Ellman, J. A., *J. Am. Chem. Soc.* 2002, *124*, 14868–14870.

[19] Gosalia, D. N., Diamond, S. L., *Proc. Natl. Acad. Sci. USA* 2003, *100*, 8721–8726.

[20] Maly, D. J., Leonetti, F., Backes, B. J., Dauber, D. S. *et al.*, *J. Org. Chem.* 2002, *67*, 910–915.

[21] Fields, G. B., Noble, R. N., *Int. J. Pep. Protein Res.* 1990, *35*, 161–214.

[22] Eisen, M. B., Spellman, P. T., Brown, P. O., Botstein, D., *Proc. Natl. Acad. Sci. USA* 1998, *95*, 14863–14868.

[23] Vindigni, A., Dang, Q. D., DiCera, E., *Nat. Biotechnol.* 1997, *15*, 891–895.

[24] Cho, K., Tanaka, T., Cook, R. R., Kisiel, W. *et al.*, *Biochemistry* 1984, *23*, 644–650.

[25] Bianchini, E. P., Louvain, V. B., Marque, P.-E., Juliano, M. A. *et al.*, *J. Biol. Chem.* 2002, *277*, 20527–20534.

[26] Matthews, D. J., Wells, J. A., *Science* 1993, *260*, 1113–1117.

[27] Ke, S.-H., Coombs, G. S., Tachias, K., Corey, D. R. *et al.*, *J. Biol. Chem.* 1997, *272*, 20456–20462.

[28] Ke, S.-H., Coombs, G. S., Tachias, K., Navre, M. *et al.*, *J. Biol. Chem.* 1997, *272*, 16603–16609.

[29] Le Bonniec, B. F., Myles, T., Johnson, T., Knight, C. G. *et al.*, *Biochemistry* 1996, *35*, 7114–7122.

[30] Ludeman, J. P., Pike, R. N., Bromfield, K. M., Duggan, P. J. *et al.*, *Int. J. Biochem. Cell Biol.* 2003, *35*, 221–225.

[31] Hervio, L. S., Coombs, G. S., Bergstrom, R. C., Trivedi, K. *et al.*, *Chem. Biol.* 2000, *7*, 443–452.

[32] Banfield, D. K., MacGillivray, R. T. A., *Proc. Natl. Acad. Sci. USA* 1992, *89*, 2779–2783.

[33] Salte, R., Norberg, K., Oedegaard, O. R., *Thromb. Res.* 1995, *80*, 193–200.

[34] Bode, W., Mayr, I., Baumann, U., Huber, R. *et al.*, *EMBO J.* 1989, *8*, 3467–75.

[35] Brandstetter, H., Kuehne, A., Bode, W., Huber, R. *et al.*, *J. Biol. Chem*, 1996, *271*, 29988–29992.

[36] Shapiro, R. L., Duquette, J. G., Roses, D. F., Nunes, I. *et al.*, *Cancer Res.* 1996, *56*, 3597–3604.

[37] Parolini, S., Flagiello, D., Cinquetti, A., Gozzi, R. *et al.*, *Br. J. Cancer* 1996, *74*, 1168–1174.

[38] Skelly, M. M., Troy, A., Duffy, M. J., Mulcahy, H. E. *et al.*, *Clin. Cancer Res.* 1997, *3*, 1837–1840.

[39] Huber, K., Kirchheimer, J. C., Ermler, D., Bell, C., Binder, B. R., *Cancer Res.* 1992, *52*, 1717–1720.

[40] Foekens, J. A., Schmitt, M., van Putten, W. L., Peters, H. A. *et al.*, *Cancer Res.* 1992, *52*, 6101–6105.

[41] Foekens, J. A., Look, M. P., Peters, H. A., van Putten, W. L. *et al.*, *J. Natl. Cancer Inst.* 1995, *87*, 751–756.

[42] Hertmanni, P., Picque, E., Thomas, D., Larreta-Garde, V., *FEBS Lett.* 1991, *279*, 123–131.

[43] Castro, G. R., *Enzyme Microb. Technol.* 1999, *25*, 689–694.

[44] Lopez-Otin, C., Overall, C. M., *Nat. Rev. Mol. Cell Biol.* 2002, *3*, 509–519.

