BIOCHEMISTRY
including biophysical chemistry & molecular biology

Article

pubs.acs.org/biochemistry

# Discovery of Amino Acid Motifs for Thrombin Cleavage and Validation Using a Model Substrate

Natasha M. Ng,[†] James D. Pierce,[‡] Geoffrey I. Webb,[§] Boris I. Ratnikov,[‡] Lakshmi C. Wijeyewickrema,[†] Renee C. Duncan,[†] Amy L. Robertson,[†] Stephen P. Bottomley,[†] Sarah E. Boyd,[∥] and Robert N. Pike*[,†]

[†]Department of Biochemistry and Molecular Biology, Monash University, Clayton, Victoria 3800, Australia

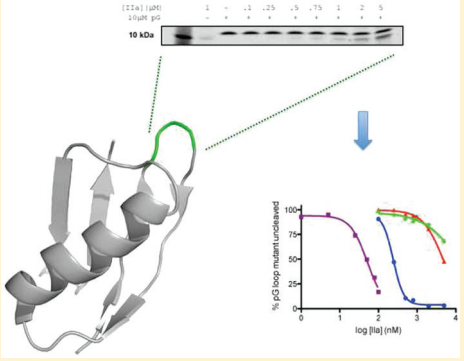
[‡]Sanford-Burnham Medical Research Institute, La Jolla, California 92037-1062, United States

[§]Clayton School of Information Technology, Monash University, Clayton, Victoria 3800, Australia

[∥]School of Mathematical Sciences, Monash University, Clayton, Victoria 3800, Australia

Supporting Information

**ABSTRACT:** Understanding the active site preferences of an enzyme is critical to the design of effective inhibitors and to gaining insights into its mechanisms of action on substrates. While the subsite specificity of thrombin is understood, it is not clear whether the enzyme prefers individual amino acids at each subsite in isolation or prefers to cleave combinations of amino acids as a motif. To investigate whether preferred peptide motifs for cleavage could be identified for thrombin, we exposed a phage-displayed peptide library to thrombin. The resulting preferentially cleaved substrates were analyzed using the technique of association rule discovery. The results revealed that thrombin selected for amino acid motifs in cleavage sites. The contribution of these hypothetical motifs to substrate cleavage efficiency was further investigated using the B1 IgG-binding domain of streptococcal protein G as a model substrate. Introduction of a $P_2$−$P_1'$ LRS thrombin cleavage sequence within a major loop of the protein led to cleavage of the protein by thrombin, with the cleavage efficiency increasing with the length of the loop. Introduction of further $P_3$−$P_1$ and $P_1$−$P_1'$−$P_3'$ amino acid motifs into the loop region yielded greater cleavage efficiencies, suggesting that the susceptibility of a protein substrate to cleavage by thrombin is influenced by these motifs, perhaps because of cooperative effects between subsites closest to the scissile peptide bond.




Thrombin (EC 3.4.4.13) is a trypsin-like serine protease most widely known for its role in the blood coagulation system, where it interacts with several molecules to maintain hemostasis. Thrombin acts as a procoagulant enzyme in the conversion of fibrinogen to fibrin in the final stages of blood coagulation but also has regulatory activities that both amplify and attenuate the hemostatic response.[1]

To understand thrombin function and design selective, potent inhibitors of the enzyme, we must understand how the protease selects its substrates for cleavage. Thrombin has relatively high specificity with respect to the limited number of peptide bonds that it cleaves within a variety of substrates.[2] Previous studies have shown that this relatively narrow specificity can be attributed to a number of molecular mechanisms. The active site is partially occluded by two insertion loops;[3] two exosites affect the conformation of the protease and facilitate the binding of substrates to the active site,[3,4] and thrombomodulin[5,6] can associate with thrombin and overturn the protease's activity from procoagulant to anticoagulant.

The substrate specificity of thrombin has been well-studied, but these efforts have focused on individual subsites, rather than on all of the subsites as a coordinated unit:[7] this potentially overlooks an important aspect of how thrombin interacts with its substrates. The substrates used to examine specificity have usually been based on known thrombin substrates or inhibitors or examples of these molecules with mutations at specific positions; this nonrandomized approach discounts the possibility of cooperative effects occurring between subsites, which has been previously observed for a number of proteases.[8] Indeed, a previous study has shown that thrombin shows cooperativity between the $S_2$ and $S_1$ subsites.[9] Here, we have examined whether thrombin selects for amino acid motifs in substrates, which would be indicative of the occurrence of subsite cooperativity, and whether these motifs better account for cleavage of known protein substrates than position-specific scoring matrices. We have then further characterized the effects of the elucidated cleavage motifs on the cleavage of an engineered protein substrate of the enzyme. The results of this study indicate that several of the discovered motifs have a strong impact on how thrombin selects its substrates for cleavage.

**Received:** August 24, 2011
**Revised:** October 31, 2011
**Published:** November 3, 2011

## EXPERIMENTAL PROCEDURES

**Materials.** Construction of the modified fUSE5 phage library has been described previously.[10] Dynabeads M-450 Epoxy was purchased from Invitrogen (St. Louis, MO). Monoclonal anti-FLAG M2 antibody produced in mouse was purchased from Sigma-Aldrich (Madison, WI). D-Phenylalanyl-L-prolyl-L-arginyl-chloromethyl ketone (PPACK) dihydrochloride was purchased from Calbiochem (San Diego, CA). Thrombin was purchased from Haematologic Technologies and was also kindly donated by J. Huntington (University of Cambridge, Cambridge, U.K.). Platinum PCR Supermix was purchased from Invitrogen. fUSE5 forward primer (5′-TAA TAC GAC TCA CTA TAG GGC AAG CTG ATA AAC CGA TAC AAT T-3′) and fUSE5 Super Reverse primer (5′-CCG TAA CAC TGA GTT TCG TC-3′) were used for phage substrate sequencing. Genes for the protein G mutants were synthesized by GenScript. NdeI and EcoRI restriction enzymes were purchased from Promega (Madison, WI). Tris-Tricine mini gels (10 to 20%) were purchased from NuSep (Sydney, Australia). Novex 10 to 20% Tricine gels were purchased from Invitrogen. Complementary primer pairs for site-directed mutagenesis of protein G loop mutants were synthesized by GeneWorks (Adelaide, Australia). Monoclonal HRP-conjugated anti-FLAG M2 antibody was purchased from Sigma-Aldrich (Madison, WI). Monoclonal anti-M13 antibody was purchased from Pharmacia. SIGMA*FAST* OPD (*o*-phenylenediamine dihydrochloride) substrate tablets were purchased from Sigma-Aldrich.

**Discovery of Amino Acid Sequence Motifs by the Phage Display Method.** The substrate phage library was composed of randomized hexapeptides displayed on the gene 3 protein (g3p) of filamentous M13 bacteriophage. The hexapeptide was also flanked by a FLAG epitope engineered at the N-terminus of g3p.[11] The substrate phage library was generated using a modified version of the fUSE5 phagemid. The phage display method used for biopanning has been described previously.[12] An aliquot of the substrate phage library ($4 \times 10^7$ phage) was incubated with 200 nM thrombin in PBS and 1 mM $MgSO_4$ for 2 h at 37 °C. A control selection contained no protease. Incubation was stopped with the addition of 10 $\mu$M PPACK. Phage that were cleaved by thrombin were separated from the noncleaved phage by immunodepletion by the addition of M2 anti-FLAG monoclonal antibody bound to M-450 Dynabeads. The cleaved phage remaining in the supernatant were amplified using *Escherichia coli* K91 as described previously[13,14] and were then used for subsequent rounds of substrate selection.

**Analysis of Substrate Hydrolysis by a Phage Enzyme-Linked Immunosorbent Assay (ELISA).** Hydrolysis of individual phage substrates by thrombin was assessed using a modified ELISA. The wells of a 96-well microtiter plate were coated with anti-M13 antibody (2.5 $\mu$g/mL) in PBS overnight at 4 °C. After being coated, the wells were blocked for 1 h at 37 °C in TBST [50 mM Tris (pH 7.8), 150 mM NaCl, and 0.1% (v/v) Tween 20] containing 100 mg/mL BSA. An overnight phage culture (100 $\mu$L) was added to each well and incubated for 2 h at 4 °C, after which unbound phage were removed with four washes of ice-cold TBST. To assess hydrolysis, 50 nM thrombin was added to the appropriate wells in PBS containing 1 mM $MgSO_4$ for 2 h at 37 °C. Control wells did not contain protease. The protease solution was removed via four washes with ice-cold TBST. To measure thrombin hydrolysis of the peptides encoded by the phage, horseradish peroxidase (HRP)-conjugated monoclonal M2 antibody was added to each well, and the plates were incubated at 37 °C for 1 h. Binding of the HRP-conjugated M2 antibody to the FLAG epitope was assessed by the addition of the OPD HRP substrate, followed by detection at 490 nm. The extent of hydrolysis, taken as a measure of substrate efficiency, was calculated by the ratio of the OD at 490 nm of the thrombin-treated samples versus that of matched samples lacking protease. Each clone that was hydrolyzed more than 20% by thrombin was subsequently sequenced to yield a set of 631 unique sequences (Table 1 of the Supporting Information).

**Computational Identification of Significant Amino Acid Motifs.** The hexamer inserts from each of the 631 phage sequences were used as input to the Magnum Opus association discovery software program[15] to find statistically significant amino acid motifs (Table 1). Included in the output are the measures of observed occurrences, expected occurrences, lift, and leverage. Observed occurrence is the number of times a given motif appeared in the input data set, while expected occurrence is the number of times a motif would be expected to appear in the input data set if the amino acids within the motif occurred independently of each other. Lift and leverage both measure how often the amino acids of each motif co-occur in the substrates, as compared to the situation if their occurrence in the substrates was statistically independent. Lift is a measure of how much the motif increases the probability of cleavage, relative to the default rate of cleavage in the data set. By using a lift value of >1.0, only motifs that increase the probability of cleavage over the default are reported. Lift is susceptible to noise in small data sets, but Magnum Opus reduces the risk of overestimating the value of lift by applying the Bayesian smoothing mechanism called the m-estimate. Leverage is the difference between co-occurrences of the amino acids in motifs compared to the expected number of times those amino acids would occur if they were statistically independent. Leverage complements the measure of lift but is susceptible to the rare item problem and therefore has the potential to miss rare motifs.

**Table 1. Significant Amino Acid Motifs Discovered Using Magnum Opus**

| motif | observed[b] | expected[c] | lift[d] | leverage[e] |
|---|---|---|---|---|
| [VLMP][RK][SAG][a] | 119 | 18.8 | 6.3 | 100.2 |
| [RK][SAG]X[SAG] | 63 | 12.4 | 5.1 | 50.6 |
| R[SAG]X[VLP] | 53 | 9.3 | 5.7 | 43.7 |
| [VL][LMP][RK] | 35 | 10.1 | 3.45 | 24.9 |
| R[SAG]XXX[VLM] | 35 | 2.7 | 13.1 | 32.3 |

[a]Brackets are used to denote where multiple amino acids are found at the same position within the motif. X implies that any amino acid can be accepted at the given position. [b]Observed is the total number of times a motif occurred. [c]Expected is the number of times motif should occur randomly. [d]Lift is observed/expected. [e]Leverage is observed − expected.

**Prediction of Protease Specificity (PoPS) Computational Models of Thrombin Specificity.** Specificity data previously generated for thrombin[16] were used to develop a PoPS model of specificity.[17] Those data were used in preference to the high-throughput data described here, because the cleavage site (after an arginine residue) was reproducibly found at a similar position, which is required for the generation

Table 2. Rankings of the Cleavage Sites in Thrombin Physiological Substrates Using PoPS Models with or without Dependency Rules Created from the Discovered Amino Acid Motifs

| physiological substrate | cleavage site ($P_5$–$P_5'$) | motif(s) | PoPS (without rules) | PoPS (with rules) |
|---|---|---|---|---|
| fibrinogen A | GGGVR$^{16}$–GPRVV[a] | [VLMP][RK][SAG] R[SAG]XXX[VLM] | 26 | 1 |
| fibrinogen B | FFSAR$^{14}$–GHRPL | R[SAG]XXX[VLM] | 6 | 1 |
| factor VIII | NQSPR$^{1689}$–SFQKK | [VLMP][RK][SAG] | 28 | 1 |
| factor VIII | AIGPR$^{740}$–SFSQN | [VLMP][RK][SAG] [RK][SAG]X[SAG] | 23 | 1 |
| factor VIII | FIQIR$^{372}$–SVAKK | [RK][SAG]X[SAG] | 10 | 12 |
| PAR-1 | TLDPR$^{41}$–SFLLR | [VLMP][RK][SAG] R[SAG]X[VLP] | 5 | 1 |
| PAR-4 | LPAPR$^{47}$–GYPGQ | [VLMP][RK][SAG] R[SAG]X[VLP] | 2 | 1 |
| factor XIII | EGVPR$^{36}$–GVNLQ | [VLMP][RK][SAG] [VL][LMP][RK] | 9 | 1 |
| antithrombin | IAGR$^{393}$–SFLNP | R[SAG]X[VLP] | 8 | 8 |
| protein C | QVDPR$^{169}$–LIDGK | none | 2 | 2 |
| factor V | AWYLR$^{1573}$–SNNGN | [VLMP][RK][SAG] | 35 | 1 |
| factor V | PLSPR$^{1046}$–TFHPL | none | 4 | 7 |
| factor V | ALGIR$^{709}$–SFRNS | none | 8 | 11 |

[a]Superscript numbers indicate the amino acid residues at which cleavage occurs. The amino acid residue numbering is based on that of ref 2.

of the PoPS specificity matrix. The values for each subsite were scaled to the range of −5.0 to 5.0, as required by the software, and subsite weights were set to be the scale factors derived from the scaling of each subsite, making matrix values of the PoPS model mathematically equivalent to the originally measured values.[16]

A secondary model was developed on the basis of the amino acid motifs discovered by Magnum Opus. These motifs were built into the thrombin PoPS model as "dependency rules" and were all assigned the same arbitrary score of 30, using the maximum allowable score of 5.0 for each subsite. In this instance, the score has no impact other than to discriminate instances of the specificity motifs within substrate sequences.

**Cloning and Site-Directed Mutagenesis of Protein G Mutants.** The genes for four mutants of the B1 IgG-binding domain of streptococcal protein G (pG) were synthesized by GenScript and cloned into pET- 21b(+) (Novagen) for expression. To create the pG loop mutants, we designed complementary primer pairs to introduce one, two, or three Gly residues immediately before the LRS thrombin cleavage sequence, or to remove one Leu residue after the LRS sequence using site-directed mutagenesis (Table 3 of the Supporting Information). Resulting DNA products were transformed into DH5$\alpha$ cells and plated onto LB-agar plates containing 100 $\mu$g/mL ampicillin. Random colonies were picked from these plates, and the DNA was sequenced using T7 promoter primers.

**Expression and Purification of pG Mutants.** Following sequence verification, pG mutant DNA was transformed into BL21(DE3) cells, which were grown in 2YT broth containing 100 $\mu$g/mL ampicillin to an $OD_{600}$ of 0.8. BL21(DE3) cells were induced with 1 mM IPTG for 4 h. The cell culture pellet was collected following centrifugation at 5000$g$ for 20 min at 4 °C, resuspended in 30 mL of lysis buffer [50 mM $NaH_2PO_4$, 300 mM NaCl, and 10 mM imidazole (pH 8.0)], and stored at −80 °C. The cell pellet was thawed, sonicated for 6 × 10 s intervals, and centrifuged at 27000$g$ for 20 min at 4 °C. The supernatant was collected and diluted to 50 mL in 20 mM imidazole buffer [50 mM $NaH_2PO_4$, 300 mM NaCl, and 20 mM imidazole (pH 8.0)]. The supernatant was loaded onto a 5 mL HisTrap nickel column (GE Healthcare) equilibrated with 20 mM imidazole buffer, and unbound material was removed by washing with 2 column volumes of the same buffer. Protein was eluted with a stepwise gradient from 20 to 250 mM imidazole buffer [50 mM $NaH_2PO_4$, 300 mM NaCl, and 250 mM imidazole (pH 8.0)]. Fractions containing pG mutants were pooled, loaded onto a Superdex 75 16/60 gel filtration column, and chromatographed in 20 mM Tris (pH 7.4) and 50 mM NaCl. On average, 10 mg of pure protein (per liter of original culture) eluted from the gel filtration column at a position expected for fully folded protein G. Analysis of wild-type and mutant protein G molecules using circular dichroism spectroscopy revealed no changes to the secondary structure of proteins due to any of the mutations made.

**Hydrolysis of pG Mutants.** Hydrolysis of the pG mutants by thrombin was assessed via a cleavage assay. For each mutant, 10 $\mu$M pG mutant was incubated with increasing concentrations of thrombin (from 100 nM to 5 $\mu$M) in Tris-buffered saline [20 mM Tris and 150 mM NaCl (pH 7.4)] for either 2 h or overnight at 37 °C. Incubation was stopped in each assay by the addition of 10 $\mu$M PPACK. pG samples were run on reduced 10–20% Tris-Tricine gels and protein bands visualized after gels had been stained with Coomassie brilliant blue R-250. $EC_{50}$ values based on the disappearance of the full-length pG substrate were calculated by densitometry using ImageQuant TL (GE Healthcare) and GraphPad Prism version 4.0. The pG mutants were unable to inhibit cleavage of a fluorescent peptide substrate, tosyl-Gly-Pro-Arg-NHMec, even at concentrations up to 50 $\mu$M. This indicated that the concentrations of pG being used in the assays were at least 10-fold lower than the $K_m$ value for the substrate, and this allowed the conversion of the $EC_{50}$ values obtained into $k_{cat}/K_m$ values using the equation[18] $k_{cat}/K_m = \ln 2/tEC_{50}$. It should be noted that $t$ is the time of the assay used to yield the $EC_{50}$ value.

## RESULTS

**Discovery of Amino Acid Sequence Motifs by the Phage Display Method.** The phage display method was used to find commonly occurring amino acid motifs within

thrombin-selected phage peptides. Two successive rounds of selection using thrombin led to the isolation of optimal substrates from a phage library of $4 \times 10^7$ independent clones; 1152 clones were screened for their ability to be cleaved by thrombin using the substrate phage ELISA as reported previously,[10,11] and 672 clones were cleaved efficiently by thrombin where >20% hydrolysis of susceptible phage was observed in comparison to nonsusceptible phage. The amino acid sequence of each phage peptide verified to be cleaved by thrombin was determined by sequencing the corresponding DNA inserts of the selected phage, yielding a final set of 631 unique phage sequences (Table 1 of the Supporting Information).

**Computational Identification of Significant Amino Acid Motifs.** The set of 631 hexamer inserts was used as input into the Magnum Opus association discovery software program.[15] Magnum Opus searched for statistically significant amino acid motifs. In addition to the search for motifs that contained any of the 20 individual amino acids, a more refined search was also conducted using specific subsets of amino acids that were known to be collectively favorable to thrombin. The first grouping contained residues V, L, M, and P; the second grouping contained R and K, and the third grouping contained S, A, G, and T. Magnum Opus was used to find statistically significant motifs derived from any of the possible subsets of these groupings (Table 2 of the Supporting Information). Motifs were considered significant if they had a lift (observed/expected) value greater than 1 (Table 1). These amino acids within the motif appeared together more frequently than would be expected if they co-occurred independently of each other, suggesting that cooperative mechanisms between subsites had positively influenced the selection of these particular phage sequences by thrombin. Assuming that thrombin cleaves after either an Arg or a Lys residue at position $P_1$, the [VL][LMP]-[RK] motif implies positive cooperativity among the $S_3$, $S_2$, and $S_1$ subsites, the [VLMP][RK][SAG] motif implies positive cooperativity among the $S_2$, $S_1$, and $S_1'$ subsites, the [RK][SAG]X[SAG] and R[SAG]X[VLP] motifs imply positive cooperativity among the $S_1$, $S_1'$, and $S_3'$ subsites, and the R[SAG]XXX[VLM] motif implies positive cooperativity among the $S_1$, $S_1'$, and $S_5'$ subsites (Table 1).

**Computational Validation of Amino Acid Motifs.** A position-specific scoring matrix (PSSM) model of substrate specificity was created in PoPS,[17] utilizing previously described thrombin phage display data.[16] This model was subsequently used to analyze known cleavage sites in physiological thrombin substrates.[2] Table 2 illustrates that the known cleavage sites were generally not ranked highly by this model. For example, the known cleavage site of factor V after amino acid residue 1573 was ranked only the 35th best cleavage site within the factor V protein sequence, and two thrombin substrates for which there is no involvement of the exosite in cleavage, PAR-4 and factor XIII, had cleavage sites ranked second and ninth, respectively.

The amino acid motifs identified by Magnum Opus from the thrombin-selected phage sequences (Table 1) were then integrated into a PoPS model as dependency rules and used to analyze the same physiological thrombin substrates. When a motif occurred within a potential substrate, an arbitrary score of 30 was assigned. The PoPS dependency rules outperformed the PSSM model, with the majority of the thrombin cleavage sites improving their ranking within the substrate sequence. In particular, the known cleavage sites in fibrinogen A$\alpha$ and B$\beta$ chains, PAR-1, PAR-4, factor XIII, and two of the three sites in factor VIII improved to be predicted as the most likely to be cleaved in the protein (Table 2). Most of the sites that greatly improved their predicted cleavage ranking actually contained two of the motifs, indicating that the motifs are highly represented at physiological cleavage sites within proteins. The greatest improvement in ranking was for the known cleavage site of factor V after amino acid residue 1573, rising from the 35th ranked site to be equally top-ranked: this site only contained one motif, but it was the motif with the strongest leverage value of all motifs (Table 1). This suggests that for most thrombin physiological substrates, the amino acid motifs discovered positively influence the selection of sites of cleavage within the proteins.

**Hydrolysis of Protein G Mutants by Thrombin.** To test whether the discovered motifs contribute to cleavage efficiency by thrombin, we initially created a thrombin protein substrate by substituting residues in the major loop of the B1 IgG-binding domain of streptococcal protein G (pG) with the $P_2$–$P_1'$ LRS thrombin cleavage sequence, derived from the motif with highest leverage value. pG was chosen as an ideal candidate for engineering to be an optimal thrombin substrate for a number of reasons. First, pG is a highly stable and extensively studied protein,[19] and previous studies allow the relatively accurate prediction of how specific amino acids at selected positions influence the stability of the molecule.[20−22] This allowed us to select existing residues in the major loop of pG for substitution with the $P_2$–$P_1'$ LRS thrombin cleavage sequence (Figure 1), with minimal effects on the overall conformation of pG. Also, importantly, wild-type pG was not cleaved by thrombin (Figure 2A).

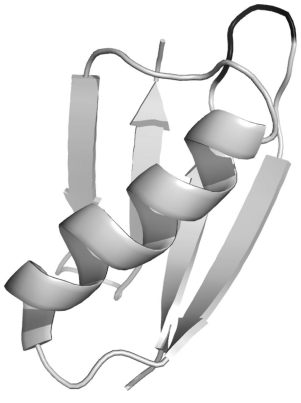



**Figure 1.** (A) X-ray crystal structure of the B1 IgG-binding domain of streptococcal protein G (pG) (Protein Data Bank entry 1GB1). The location for insertion of the $P_2$–$P_1'$ LRS thrombin cleavage sequence into the loop region is highlighted in darker gray. (B) Sequence of wild-type pG compared to those for the mutant in which the $P_2$–$P_1'$ LRS thrombin cleavage sequence was inserted into the loop region within the pG sequence (bold) and the pG PAR-1 sequence mutant, where the PAR-1 cleavage sequence (bold) was inserted at the N-terminus.

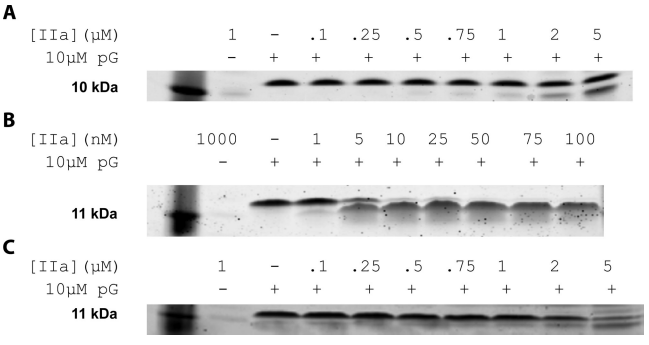



**Figure 2.** Thrombin cleavage of wild-type and mutant pG. (A) Thrombin (100 nM to 5 $\mu$M) did not cleave wild-type pG. The band that appears underneath the wild-type pG band is from the thrombin preparation. (B) A pG mutant containing a PAR-1 cleavage sequence at the N-terminus of the protein is cleaved by thrombin. (V) Cleavage of a mutant of pG in which the LRS sequence had been substituted into the major loop of the protein.

The genes for two mutants of pG were initially synthesized, one in which the existing residues within the loop region connecting $\beta$-strands 1 and 2 (residues 9−11) were substituted with the $P_2$−$P_1$′ LRS thrombin cleavage sequence, designated loop in Table 3, and hereafter termed the pG loop mutant. The other mutant was synthesized as a positive control, in which the $P_5$−$P_4$′ PAR-1 thrombin cleavage sequence (TLDPR−SFLL) was inserted at the N-terminus of pG, hereafter termed the PAR-1 pG mutant. Following cloning, expression, and purification, cleavage assays were performed with both pG mutants using a range of thrombin concentrations. The PAR-1 pG mutant was very efficiently cleaved, with distinct cleavage bands appearing after treatment with as little as 5 nM thrombin (Figure 2B). The pG loop mutant was also cleaved by thrombin but required much higher concentrations (1−5 $\mu$M) before cleavage bands appeared (Figure 2C). To facilitate a subsequent comparison with other mutants, the cleavage efficiency of the pG loop mutant was quantified. Densitometry was performed on the basis of the disappearance of the full-length substrate, and dose−response curves were plotted to calculate the log $EC_{50}$ values for each mutant, which in turn were converted into $k_{cat}/K_m$ values as it could be established that assays were conducted with substrate concentrations at least 10-fold below the $K_m$ value (Table 3). The $k_{cat}/K_m$ value for the cleavage of the pG loop mutant by thrombin was estimated to be 20 $M^{-1}$ $s^{-1}$, suggesting relatively inefficient cleavage.

The relatively inefficient cleavage of the pG loop mutant led us to suspect that the loop length might not be optimal for access to the relatively "deep" active site of thrombin. To date, few studies have examined the effect of substrate loop length on protease cleavage, but it has anecdotally been assumed that an increased loop length will increase loop flexibility and accessibility to a protease for cleavage, especially for proteases such as thrombin that have an active site surrounded by extensive loops, rendering the subsites less accessible. For relatively selective proteases, such as caspase-3 and tissue-type plasminogen activator (t-PA), it has been found that a longer loop length enhances the catalytic efficiency of cleavage.[18,23] In addition, in a previous study of the serpin, plasminogen activator inhibitor-1 (PAI-1), the rate of inhibition of t-PA was dependent on the length of the reactive center loop.[24] Given that the pG loop mutant is cleaved by thrombin, albeit at high thrombin concentrations, it would be interesting to determine whether there is a similar dependence of substrate loop length on cleavage susceptibility by thrombin.

Four extra mutants were created by site-directed mutagenesis; three mutants contained one, two, and three additional Gly residues immediately preceding the $P_2$−$P_1$′ LRS thrombin cleavage site, designated loop +1, loop +2, and loop +3, respectively, in Table 3. The fourth mutant had a Leu residue deleted at the $P_2$′ position, designated loop −1 in Table 3. Following expression and purification, cleavage assays were performed on all four insertion/deletion mutants to compare the extent of cleavage with the original pG loop mutant (Figure 3 and Table 3). All pG loop extension mutants had $k_{cat}/K_m$ values that were lower than that of the original pG loop mutant (20 $M^{-1}$ $s^{-1}$). Interestingly, there does not appear to be a linear correlation between an increased loop length and an increased susceptibility to thrombin cleavage, because one additional Gly residue in the pG loop caused a 12.6-fold increase in the $k_{cat}/K_m$ value for thrombin (253 $M^{-1}$ $s^{-1}$), but the addition of a further residue caused an only minor further increase for the loop +2 mutant (260 $M^{-1}$ $s^{-1}$), and the loop +3 mutant had a decreased $k_{cat}/K_m$ value of 209 $M^{-1}$ $s^{-1}$ (Table 3). This strongly suggests that the addition of one residue to the pG loop was sufficient to have a major effect on its accessibility for thrombin cleavage. Given the strong effect of adding a residue, it was very surprising to note that the $k_{cat}/K_m$ value for the loop −1 mutant was the highest of all the pG loop mutants (401 $M^{-1}$ $s^{-1}$), suggesting it was the most efficiently cleaved.

**Table 3. Effects of Loop Length on Thrombin Cleavage Efficiency**

| pG mutant | $P_5$−$P_5$′ amino acid sequence | $k_{cat}/K_m$ ($M^{-1}$ $s^{-1}$)[a] |
|---|---|---|
| loop | ILNLR−SLKGE | 20 |
| loop +1 | LNGLR−SLKGE | 253 |
| loop +2 | NGGLR−SLKGE | 260 |
| loop +3 | GGGLR−SLKGE | 209 |
| loop −1 | ILNLR−SKGET | 401 |

[a]Standard errors for all log $EC_{50}$ values, used to derive $k_{cat}/K_m$ values, were <15% of the value.

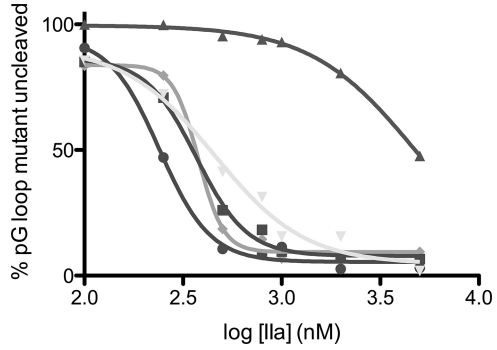



**Figure 3.** Effect of pG loop length on thrombin cleavage. Densitometric analysis of the disappearance of the uncleaved mutant yields a dose−response curve highlighting the percentage of each uncleaved pG loop mutant after a 2 h incubation at 37 °C, where thrombin concentrations were increased from 100 nM to 5 $\mu$M. $EC_{50}$ values were derived from nonlinear regression fits of these curves. The following mutants are shown: pG loop (▲), pG loop +1 (◆), pG loop +2 (■), pG loop +3 (▼), and pG loop −1 (●).


dx.doi.org/10.1021/bi201333g | *Biochemistry* 2011, 50, 10499−10507

**Effect of Subsite Cooperativity on Hydrolysis of pG by Thrombin.** The relatively high $k_{cat}/K_m$ value for cleavage of the pG loop −1 mutant by thrombin can be explained by examining the primary sequence of the loop in the context of the previously identified cleavage motifs. The pG loop −1 mutant $P_5$−$P_5'$ sequence (ILNLR−SKGET) matches the [RK][SAG]X[SAG] cooperativity motif identified from the phage display results for the $P_1$−$P_3'$ substrate positions, as well as the [VLMP][RK][SAG] motif for the $P_2$−$P_1'$ positions. This suggests that subsite cooperative effects might be sufficiently substantial to overcome the structural constraints of having a shorter loop, which should otherwise have been less susceptible to cleavage.

To investigate and validate the identified motifs and their effects on cleavage susceptibility, we synthesized a new series of pG loop mutants representing these motifs by site-directed mutagenesis (Table 4). Cleavage assays were performed on each mutant with a 2 h incubation period (Figure 4A). Interestingly, the LILLR−SLKGE pG mutant had a $k_{cat}/K_m$ value of 1853 $M^{-1}$ $s^{-1}$, the highest of all the values for the pG loop mutants (Table 4). This mutant was created by the deletion of the $P_3$ Asn residue from the initial ILNLR−SLKGE pG loop mutant and demonstrates that the $P_3$−$P_1$ [VL]-[LMP][RK] cooperativity motif has a strong influence on cleavage susceptibility. As the pG loop −1 mutant also had a significantly higher $k_{cat}/K_m$ value than the initial pG loop mutant, the effect of decreasing the loop length on increased cleavage susceptibility by thrombin was investigated. Another pG loop mutant, ILNGR−SLKGE, was synthesized to investigate how removal of the $P_2$−$P_1'$ [VLMP][RK][SAG] motif affected cleavage susceptibility. The resulting $k_{cat}/K_m$ value for thrombin cleavage decreased to 8.6 $M^{-1}$ $s^{-1}$ (Table 4), compared to that of the initial pG loop mutant (20 $M^{-1}$ $s^{-1}$). This shows that the decreased loop length of the LILLR−SLKGE and pG loop −1 mutants was not the primary factor contributing to the increased cleavage susceptibility. In addition, the decreased $k_{cat}/K_m$ value for the ILNGR−SKGET mutant suggests that the [VLMP][RK][SAG] cooperativity motif has an effect on cleavage susceptibility.

Two pG mutants with $P_5$−$P_5'$ sequences of ILNGR−SLKGL and ILNGR−SLVGE were designed to test the influence of the $P_1$−$P_1'$−$P_5'$ R[SAG]XXX[VLM] and $P_1$−$P_1'$−$P_3'$ R[SAG]X-[VLP] motifs on cleavage susceptibility. These results were compared to those for the ILNGR−SLKGE pG mutant that contained no motifs (Figure 4B and Table 4). All three pG mutants exhibited a negligible amount of substrate cleavage after a 2 h incubation period; thus, to compare the effects of these motifs on cleavage susceptibility, we employed an overnight incubation time. Figure 4B shows that in comparison to the motif-less ILNGR−SLKGE pG mutant, the $P_1$−$P_1'$−$P_5'$ ILNGR−SLKGL pG mutant shows a greater degree of cleavage, whereas the $P_1$−$P_1'$−$P_3'$ ILNGR−SLVGE mutant showed that the substitution of $P_3'$ Lys with Val decreased the amount of cleavage slightly (Table 4). The increased incubation time of all three pG mutants highlights the importance of the [VLMP][RK][SAG] motif, where mutants containing this motif showed larger amounts of cleavage after a 2 h incubation. Although the R[SAG]XXX[VLM] motif improved the efficiency of cleavage 2-fold (Table 4), the long incubation time indicates that this motif makes relatively minor contributions to cleavage susceptibility.

**Table 4. Effects of the Amino Acid Motifs on Thrombin Cleavage Efficiency**

| pG mutant | subsite cooperativity motifs | $k_{cat}/K_m$ ($M^{-1}$ $s^{-1}$)[a] |
|---|---|---|
| ILNGR−SLKGE | none | 8.6 |
| ILN**LR**−SLKGE | $P_2$−$P_1$−$P_1'$ [VLMP][RK][SAG] | 20 |
| ILN**LR**−SK**G**ET | $P_2$−$P_1$−$P_1'$ [VLMP][RK][SAG] | 401 |
| | $P_1$−$P_1'$−$P_3'$ R[SAG]X[SAG] | |
| ILNG**R**−SK**G**ET | $P_1$−$P_1'$−$P_3'$ R[SAG]X[SAG] | 7.8 |
| ILNG**R**−SLKG**L** | $P_1$−$P_1'$−$P_5'$ R[SAG]XXX[VLM] | 18.8 |
| ILNG**R**−SL**V**GE | $P_1$−$P_1'$−$P_3'$ R[SAG]X[VLP] | 6.2 |
| LI**LLR**−SLKGE | $P_3$−$P_2$−$P_1$ [VL][LMP][RK] | 1853 |
| | $P_2$−$P_1$−$P_1'$ [VLMP][RK][SAG] | |

[a]Standard errors for all log $EC_{50}$ values, used to derive $k_{cat}/K_m$ values, were <15% of the value.

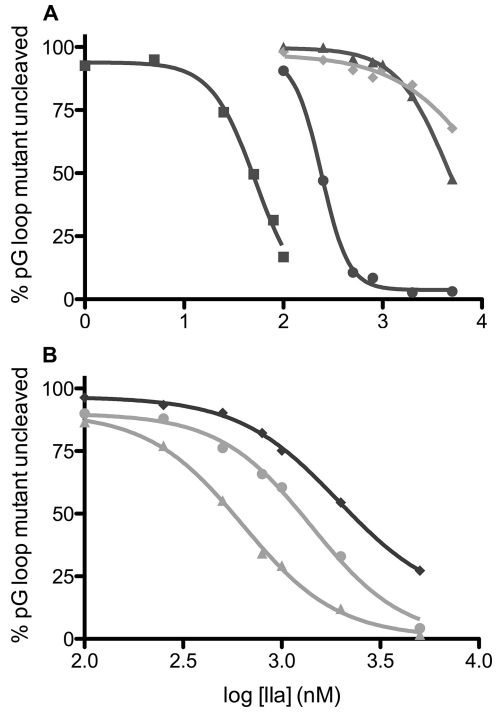



**Figure 4.** Effects of motifs on the thrombin cleavage efficiency for pG loop mutants. (A) Dose−response curve of four pG mutants: ILNLR−SLKGE (▲), ILNLR−SKGET (●), ILNGR−SKGET (◆), and LILLR−SKGET (■). The curves highlight the percentage of uncleaved pG loop mutant after a 2 h incubation at 37 °C, where thrombin concentrations increased from 100 nM to 5 μM. The $EC_{50}$ values were derived from these curves. (B) Dose−response curve highlighting the percentage of uncleaved substrate for three pG mutants: ILNGR−SLKGL (▲), ILNGR−SLKGE (●), and ILNGR−SLVGE (◆). pG mutants were incubated overnight using a thrombin concentration range from 100 nM to 5 μM.

## DISCUSSION

In-depth characterization of protease specificity is used to understand the biological and physiological roles of proteases, as well as to develop specific substrates and inhibitors for the enzymes. Specificity has been determined for many proteases, but only very rarely has cooperativity between subsites in the active site of the enzymes been taken into account.[8] Subsite cooperativity has been found to strongly influence the binding of residues in adjoining or even widely spaced subsites, and thus, it is clearly a factor that should be taken into account when gaining an understanding of the substrate selectivity of a


dx.doi.org/10.1021/bi201333g | Biochemistry 2011, 50, 10499−10507

protease. In the case of thrombin, subsite cooperativity has previously been identified between $S_2$ and $S_1$, using a set of chromogenic substrates.[9]

In this study, a randomized phage display library was used to probe the specificity of thrombin, to investigate whether substrate selection and efficiency of cleavage by thrombin were affected by cooperativity between subsites in the active site of the enzyme. Computational analysis of the hydrolyzed phage sequences identified several motifs in which the amino acids co-occurred at a higher frequency than would be expected if thrombin selected those amino acids independently of each other. In particular, the motifs implied that positive cooperativity was occurring across the following subsites: $S_3$−$S_1$, $S_2$−$S_1'$, $S_1$−$S_1'$−$S_3'$, and $S_1$−$S_1'$−$S_5'$. Cleavage sites in thrombin physiological substrates were analyzed using PoPS,[17] in which one model assumed independence between subsites and a second model allowed for interdependency between the subsites. The computational modeling showed marked improvement for the model that allowed for interdependency. These results provided evidence that subsite cooperativity in thrombin is much more extensive than has been reported.

It should be noted that many of the motifs we have identified involve the prime side residues of the substrate. A major way in which the influence of such prime side amino acid motifs on thrombin cleavage could have been tested would have been to use fluorescence-quenched substrates (FQS) containing $P_1'$−$P_5'$ residues, for example. Using X-ray crystallographic studies, we have recently found that the binding of FQS to thrombin strongly involves the quenching group, resulting in a conformation of the substrate that allows little contact with prime side residues of the substrate (N. M. Ng et al., unpublished data). This clearly would make it impossible to fully investigate the prime side motifs that we have uncovered, and therefore, we wished to investigate the effects of subsite cooperativity patterns on thrombin cleavage efficiency using a protein substrate. Ideally, it would be preferable to use a substrate in which mutations would be unlikely to affect the structure of the protein. Protein G (pG) was therefore an excellent candidate, as the structural effect of substituting amino acids at different positions within this protein has been extensively studied. Furthermore, previous work has shown that proteases preferentially cleave substrates within extended loop regions,[23,25−28] and pG contains a major loop in the B1 IgG-binding domain. Finally, pG is an excellent model for studying thrombin specificity because the wild-type protein is not cleaved by thrombin and does not contain determinants for exosite binding; thus, any observed effects can be attributed to interactions with thrombin's active site, the focus of this investigation.

A mutant containing the $P_2$−$P_1'$ pattern of LRS inserted into the loop region was cleaved by thrombin, albeit with a relatively low efficiency (Figure 2). It is possible that the low efficiency of cleavage in the major loop of pG by thrombin is the result of the loop length of the protein being limiting for cleavage by the enzyme. To test this hypothesis, we created several pG mutants that successively extended the length of the loop by one residue. Increasing the loop length by one residue improved cleavage susceptibility 12.6-fold, indicating that loop length was indeed having an effect on cleavage by thrombin. Further increases in loop length did not yield further improvement in cleavage efficiency, indicating that addition of the one extra amino acid was all that was required to increase the efficiency of cleavage by the enzyme. Interestingly, when a Leu or Asn residue was removed from the loop at the $P_2'$ or $P_3$ position, respectively, thrombin was able to cleave the substrate more efficiently than any of the pG loop +1, +2, or +3 mutants (Tables 3 and 4). This suggests that the primary sequence of the loop is the more important determinant of its susceptibility to cleavage, overriding the demonstrably positive effect of increasing the loop length. As the primary sequence of the two deletion mutants matched the $P_1$−$P_3'$ [RK][SAG]X[SAG] and $P_3$−$P_1$ [VL][LMP][RK] cooperativity motifs, it also demonstrated that the subsite cooperative effects have a more significant influence on cleavage susceptibility than loop length.

A possible explanation for subsite cooperativity, at least in some instances, is that those subsites occupy a shared region within the active site. Substrate residues would therefore be competing for access to the same region of the active site and would necessarily have complementary physicochemical properties. A review of X-ray crystal structures of thrombin−inhibitor complexes found that thrombin recognizes inhibitor residues in an open $\beta$-strand conformation at the active site.[28] Thus, the side chains of every second residue, such as $P_4$, $P_2$, $P_1'$, $P_3'$, and $P_5'$, would be expected to be oriented in one direction, with the alternate residues ($P_3$, $P_1$, $P_2'$, and $P_4'$) oriented in the opposite direction. Interestingly, even though cooperative effects might therefore be expected to occur between alternate subsites (i.e., separated by one subsite), the cooperativity motifs reported here (Table 1) imply that some of the interdependent subsites are immediately adjacent to each other or separated by more than a single subsite. This is consistent with our previous results showing that the complement serine protease, C1s, displays cooperativity between widely separated subsites.[29]

The [VLMP][RK][SAG] motif was shown to have a major influence on cleavage susceptibility, with the ILNGR−SKGET, ILNGR−SLKGE, ILNGR−SLKGL, and ILNGR−SLVGE pG mutants being poorly cleaved by thrombin in comparison to each of the loop mutants and the LILLR−SLKGE mutant, which all contain the $P_2$−$P_1'$ LRS cleavage sequence. This suggests that, at least for the positions and residues tested in this study, the greatest cooperative effects occur between subsites closest to the scissile peptide bond. These subsites also display the narrowest substrate specificity, because of the 148 and 60 loops of thrombin that protrude into the active site cleft.[30] The proximity of these loops, especially the 60 loop, to the active site cleft means that there is an increased likelihood for interactions with substrates. In particular, there is the potential for ring-stacking interactions with loop residues, particularly with the phenolic side chain of Tyr60A in the 60 loop, as well as steric hindrance.[30] As a result, the range of substrate residues able to bind at the active site is more limited. It is plausible that the subsites with the narrowest specificity display the strongest cooperative effects to ensure the best possible fit of the substrate at the active site.

The [VL][LMP][RK] cooperativity motif, found in the LILLR−SLKGE mutant, affects the $S_3$−$S_1$ subsites. This motif matches the known cleavage sites in two physiological substrates: coagulation factor XIII and insulin-like growth factor binding protein 5, with $P_4$−$P_4'$ VVPR−GVNL and MVPR−AVYL sequences, respectively. Cooperativity between the $S_3$ and $S_2$ subsites is perhaps unsurprising, as these two subsites share a hydrophobic cavity that is able to accommodate two medium-sized hydrophobic residues.[31] A study found that the optimal binding of a substrate at the active site of $Na^+$-

bound thrombin requires a rigid bond between the $P_2$ and $P_3$ substrate residues, coupled with a strong interaction at $P_1$ and a hydrophobic residue at $P_3$.[9] A substrate that matches the [VL][LMP][RK] motif would satisfy these requirements. The rigid bond between $P_2$ and $P_3$ is often ensured by a Pro residue at $P_2$, as seen in a number of physiological substrates such as PAR-1 and coagulation factors V and XI. The interactions with the $P_2$ residue and the 60 loop of thrombin constrain the type of residues able to occupy the $S_2$ subsite to hydrophobic residues, such as Pro and Val.[31] The occurrence and significance of the [VLMP][RK][SAG] motif, as demonstrated here, also highlight the narrow specificity at the $S_2$ subsite. Although hydrophobic residues are preferred at $P_3$, there are a wider variety of residues in physiological thrombin substrates at this position than at $P_2$, suggesting that preferences at this position are less stringent.

This study is aimed to identify mechanisms whereby the active site of thrombin interacts with substrates. While the enzyme contains two exosites that influence the kinetics and position of cleavage in many of the enzyme's substrates,[2] the active site of the enzyme possesses considerable specificity for its substrates relative to other related serine proteases[7] and thus must play an important role in the interaction with substrates. The findings of this study indeed indicate that in the absence of any interactions with the exosite, the presence of the correct amino acid motifs for cleavage by thrombin can influence the efficiency of cleavage by factors of up to 300-fold (calculated by determining the ratio of the efficiency value for the worst vs the best substrate determined in this study). In support of the importance of the active site in assessing the interaction with physiological substrates, our findings indicate that taking into account the motifs for cleavage discovered here greatly improved the prediction of the correct cleavage site within the substrates for the protease.

The existence of cooperativity between subsites, as indicated by the motifs for cleavage by thrombin discovered here, has significant implications for protease specificity research, and for bioinformatics approaches to the discovery of novel substrates. In this study, the use of high-throughput phage display provided a data set that could be analyzed with a machine learning approach to discover motifs that represent instances of subsite cooperativity. Incorporating the amino acid motifs as dependency rules in a PoPS model greatly improved the modeling of thrombin cleavage of known substrates, relative to a model that assumed subsite independence. Experimental determination of the effect of the cooperativity motifs on cleavage by thrombin shows that they have a strong effect on specificity, enough to overcome a reduced loop length that would otherwise have a negative effect on thrombin activity. Together, these results provide strong evidence that cooperativity between subsites in determining thrombin specificity is more extensive than previously described. Because we have also previously found that subsite cooperativity exists for the complement serine protease, C1s,[29] it appears that this might be a general phenomenon in serine proteases that needs to be taken into consideration in future studies on the mechanisms of interaction between these enzymes and their substrates.

## ■ ASSOCIATED CONTENT

### Supporting Information

Sequences obtained using phage display analysis (Table 1), amino acid groupings used in Magnum Opus searches (Table 2), and oligonucelotides used for site-directed mutagenesis of pG loop mutants (Table 3). This material is available free of charge via the Internet at http://pubs.acs.org.

## ■ AUTHOR INFORMATION


### Corresponding Author

*Department of Biochemistry and Molecular Biology, Monash University, Clayton, Victoria 3800, Australia. E-mail: rob.pike@monash.edu. Phone: 61-3-99029300. Fax: 61-3-99029500.


### Author Contributions

S.E.B. and R.N.P. contributed equally to this work.


### Funding

This work was supported by Program Grant 490900 from the National Health and Medical Research Council of Australia.


## ■ ACKNOWLEDGMENTS


We acknowledge the environment and assistance provided by Prof. Jeffrey Smith (Sanford-Burnham Medical Research Institute).


## ■ ABBREVIATIONS

pG, B1 IgG-binding domain of streptococcal protein G; PoPS, Prediction of Protease Specificity; HRP, horseradish peroxidase; OPD, *o*-phenylenediamine dihydrochloride; PBS, phosphate-buffered saline; TBST, Tris-buffered saline containing Tween 20; PAR, protease-activated receptor; PPACK, D-phenylalanyl-L-prolyl-L-arginyl-chloromethyl ketone dihydrochloride; Tyr60A, tyrosine on the 60 loop of thrombin. The residues on loops of thrombin are classed as "insertions" relative to the sequence of chymotrypsin, the archetypal serine protease; thus, residues on the loops are given alphabetical designations rather than numbers in the nomenclature for describing the structure of thrombin.

## ■ REFERENCES


(1) Walsh, P. N., and Ahmad, S. S. (2002) Proteases in Blood Clotting. In *Essays in Biochemistry: Proteases in Biology and Medicine* (Hooper, N. M., Ed.) pp 95−111, Portland Press, London.

(2) Swords Jenny, N., Lundblad, R. L., and Mann, K. G. (2006) Thrombin. In *Hemostasis and Thrombosis: Basic Principles and Clinical Practice* (Colman, R. W., Marder, V. J., Clowes, A. W., George, J. N., and Goldhaber, S. Z., Eds.) 5th ed., pp 193−213, Lippincott Williams and Wilkins, Philadelphia.

(3) Stubbs, M. T., and Bode, W. (1993) A player of many parts: The spotlight falls on thrombin's structure. *Thromb. Res. 69*, 1−58.

(4) Lechtenberg, B. C., Johnson, D. J., Freund, S. M., and Huntington, J. A. (2010) NMR resonance assignments of thrombin reveal the conformational and dynamic effects of ligation. *Proc. Natl. Acad. Sci. U.S.A. 107*, 14087−14092.

(5) Esmon, C. T., and Owen, W. G. (1981) Identification of an endothelial cell cofactor for thrombin-catalyzed activation of protein C. *Proc. Natl. Acad. Sci. U.S.A. 78*, 2249−2252.

(6) Hofsteenge, J., Taguchi, H., and Stone, S. R. (1986) Effect of thrombomodulin on the kinetics of the interaction of thrombin with substrates and inhibitors. *Biochem. J. 237*, 243−251.

(7) Bianchini, E. P., Louvain, V. B., Marque, P. E., Juliano, M. A., Juliano, L., and Le Bonniec, B. F. (2002) Mapping of the catalytic groove preferences of factor Xa reveals an inadequate selectivity for its macromolecule substrates. *J. Biol. Chem. 277*, 20527−20534.

(8) Ng, N. M., Pike, R. N., and Boyd, S. E. (2009) Subsite cooperativity in protease specificity. *Biol. Chem. 390*, 401−407.

(9) Vindigni, A., Dang, Q. D., and Di Cera, E. (1997) Site-specific dissection of substrate recognition by thrombin. *Nat. Biotechnol. 15*, 891−895.

(10) Kridel, S. J., Chen, E., and Smith, J. W. (2001) A substrate phage enzyme-linked immunosorbent assay to profile panels of proteases. *Anal. Biochem. 294*, 176−184.
(11) Kridel, S. J., Chen, E., Kotra, L. P., Howard, E. W., Mobashery, S., and Smith, J. W. (2001) Substrate hydrolysis by matrix metalloproteinase-9. *J. Biol. Chem. 276*, 20572−20578.
(12) Ratnikov, B., Cieplak, P., and Smith, J. W. (2009) High throughput substrate phage display for protease profiling. *Methods Mol. Biol. 539*, 93−114.
(13) Madison, E. L., Coombs, G. S., and Corey, D. R. (1995) Substrate specificity of tissue type plasminogen activator. Characterization of the fibrin independent specificity of t-PA for plasminogen. *J. Biol. Chem. 270*, 7558−7562.
(14) Ding, L., Coombs, G. S., Strandberg, L., Navre, M., Corey, D. R., and Madison, E. L. (1995) Origins of the specificity of tissue-type plasminogen activator. *Proc. Natl. Acad. Sci. U.S.A. 92*, 7627−7631.
(15) Webb, G. I., and Zhang, S. (2005) k-Optimal-Rule-Discovery. *Data Mining Knowledge Discovery 10*, 39−79.
(16) Ng, N. M., Quinsey, N. S., Matthews, A. Y., Kaiserman, D., Wijeyewickrema, L. C., Bird, P. I., Thompson, P. E., and Pike, R. N. (2009) The effects of exosite occupancy on the substrate specificity of thrombin. *Arch. Biochem. Biophys. 489*, 48−54.
(17) Boyd, S. E., Pike, R. N., Rudy, R. B., Whisstock, J. C., and Garcia de la Banda, M. (2005) POPS: A computational tool for modeling and predicting protease specificity. *J. Bioinf. Comput. Biol. 3*, 551−585.
(18) Timmer, J. C., Zhu, W., Pop, C., Regan, T., Snipas, S. J., Eroshkin, A. M., Riedl, S. J., and Salvesen, G. S. (2009) Structural and kinetic determinants of protease substrates. *Nat. Struct. Mol. Biol. 16*, 1101−1108.
(19) Gronenborn, A. M., Filpula, D. R., Essig, N. Z., Achari, A., Whitlow, M., Wingfield, P. T., and Clore, G. M. (1991) A novel, highly stable fold of the immunoglobulin binding domain of streptococcal protein G. *Science 253*, 657−661.
(20) Minor, D. L. Jr., and Kim, P. S. (1994) Measurement of the $\beta$-sheet-forming propensities of amino acids. *Nature 367*, 660−663.
(21) Cregut, D., and Serrano, L. (1999) Molecular dynamics as a tool to detect protein foldability. A mutant of domain B1 of protein G with non-native secondary structure propensities. *Protein Sci. 8*, 271−282.
(22) Li, H., Wang, H. C., Cao, Y., Sharma, D., and Wang, M. (2008) Configurational entropy modulates the mechanical stability of protein GB1. *J. Mol. Biol. 379*, 871−880.
(23) Coombs, G. S., Bergstrom, R. C., Madison, E. L., and Corey, D. R. (1998) Directing sequence-specific proteolysis to new targets. The influence of loop size and target sequence on selective proteolysis by tissue-type plasminogen activator and urokinase-type plasminogen activator. *J. Biol. Chem. 273*, 4323−4328.
(24) Tucker, H. M., and Gerard, R. D. (1996) Sequence requirements in the reactive-center loop of plasminogen-activator inhibitor-1 for recognition of plasminogen activators. *Eur. J. Biochem. 237*, 180−187.
(25) Hubbard, S. J., Campbell, S. F., and Thornton, J. M. (1991) Molecular recognition. Conformational analysis of limited proteolytic sites and serine proteinase protein inhibitors. *J. Mol. Biol. 220*, 507−530.
(26) Rote, K. V., and Rechsteiner, M. (1986) Degradation of proteins microinjected into HeLa cells. The role of substrate flexibility. *J. Biol. Chem. 261*, 15430−15436.
(27) Fairlie, D. P., Tyndall, J. D., Reid, R. C., Wong, A. K., Abbenante, G., Scanlon, M. J., March, D. R., Bergman, D. A., Chai, C. L., and Burkett, B. A. (2000) Conformational selection of inhibitors and substrates by proteolytic enzymes: Implications for drug design and polypeptide processing. *J. Med. Chem. 43*, 1271−1281.
(28) Tyndall, J. D., Nall, T., and Fairlie, D. P. (2005) Proteases universally recognize $\beta$ strands in their active sites. *Chem. Rev. 105*, 973−999.
(29) Boyd, S. E., Kerr, F. K., Albrecht, D. W., de la Banda, M. G., Ng, N., and Pike, R. N. (2009) Cooperative effects in the substrate specificity of the complement protease, C1s. *Biol. Chem. 390*, 503−507.
(30) Bode, W., Turk, D., and Karshikov, A. (1992) The refined 1.9-Å crystal structure of D-Phe-Pro-Arg chloromethylketone-inhibited human $\alpha$-thrombin: Structure analysis, overall structure, electrostatic properties, detailed active-site geometry, and structure-function relationships. *Protein Sci. 1*, 426−471.
(31) Bode, W., Mayr, I., Baumann, U., Huber, R., Stone, S. R., and Hofsteenge, J. (1989) The refined 1.9 Å crystal structure of human $\alpha$-thrombin: Interaction with D-Phe-Pro-Arg chloromethylketone and significance of the Tyr-Pro-Pro-Trp insertion segment. *EMBO J. 8*, 3467−3475.