

SCIENTIFIC REPORTS



OPEN

High throughput protease profiling comprehensively defines active site specificity for thrombin and ADAMTS13

Colin A. Kretz¹, Kärt Tomberg^{1,3}, Alexander Van Esbroeck⁴, Andrew Yee² & David Ginsburg^{2,3,5}

We have combined random 6 amino acid substrate phage display with high throughput sequencing to comprehensively define the active site specificity of the serine protease thrombin and the metalloprotease ADAMTS13. The substrate motif for thrombin was determined by >6,700 cleaved peptides, and was highly concordant with previous studies. In contrast, ADAMTS13 cleaved only 96 peptides (out of >10⁷ sequences), with no apparent consensus motif. However, when the hexapeptide library was substituted into the P3-P3' interval of VWF73, an exosite-engaging substrate of ADAMTS13, 1670 unique peptides were cleaved. ADAMTS13 exhibited a general preference for aliphatic amino acids throughout the P3-P3' interval, except at P2 where Arg was tolerated. The cleaved peptides assembled into a motif dominated by P3 Leu, and bulky aliphatic residues at P1 and P1'. Overall, the P3-P2' amino acid sequence of von Willebrand Factor appears optimally evolved for ADAMTS13 recognition. These data confirm the critical role of exosite engagement for substrates to gain access to the active site of ADAMTS13, and define the substrate recognition motif for ADAMTS13. Combining substrate phage display with high throughput sequencing is a powerful approach for comprehensively defining the active site specificity of proteases.

The specificity of a protease for its substrate(s) is dictated by complex interactions of exosites to capture and appropriately orient the substrate with the active site, which catalyzes peptide bond hydrolysis¹. While some proteases are highly selective for residues surrounding the P1-P1' scissile bond², others are more promiscuous³⁻⁵. For serine proteases, the fit of a substrate into the active site is largely dictated by the interaction of the P1 residue of the substrate with the S1-specificity pocket of the protease⁶. Thrombin, the final effector serine protease in the coagulation system, exhibits strong preference for Arg at position P1, although Lys can substitute for some substrates⁷. In contrast, metalloproteases are generally considered to be less-selective for amino acid content near the cleavage site^{8,9}. However, recent studies suggest that the matrix metalloprotease family exhibits a preference for P3 proline and aliphatic residues at P1'¹⁰. Understanding the amino acid sequences recognized by proteases is critical because it can lead to novel diagnostic tools and may contribute to the development pharmaceutical agents¹.

ADAMTS13, a member of the metzincin family of metalloproteases, regulates the platelet-binding capacity of von Willebrand Factor (VWF) by proteolytic processing¹¹. ADAMTS13 cleaves VWF when sufficient shear forces unfold the A2 domain, exposing the cryptic Tyr1605-Met1606 scissile bond and a number of exosite-binding domains¹²⁻¹⁴. Deficiency in ADAMTS13 causes thrombotic thrombocytopenia purpura (TTP), a disorder characterized by thrombocytopenia and hemolytic anemia caused by deposition of VWF-rich thrombi in the micro-circulation¹⁵. Fragments of VWF, such as VWF73 (comprising Asp1596-Arg1668), have been used as biochemical tools to study ADAMTS13 in an *in vitro* setting and form the basis for clinical assays of ADAMTS13 activity¹⁶. However, the efficiency of cleavage declines rapidly with shorter VWF fragments¹⁷, suggesting an important role for exosite interactions in VWF cleavage by ADAMTS13¹⁷⁻²¹.

¹Department of Medicine, McMaster University and the Thrombosis and Atherosclerosis Research Institute, Hamilton, Ontario, Canada. ²Life Sciences Institute, University of Michigan, Ann Arbor, MI, USA. ³Department of Human Genetics, University of Michigan, Ann Arbor, MI, USA. ⁴Department of Electrical Engineering and Computer Science, University of Michigan, Ann Arbor, MI, USA. ⁵Howard Hughes Medical Institute and Departments of Internal Medicine and Pediatrics, University of Michigan, Ann Arbor, MI, USA. Colin A. Kretz and Kärt Tomberg contributed equally to this work. Correspondence and requests for materials should be addressed to C.A.K. (email: colin.kretz@taari.ca)

M13 filamentous substrate phage display is a useful technique for probing the substrate recognition determinants of proteases^{7,22}. However, after several rounds of selection²³, biases in phage amplification, infectivity, and prokaryotic protein expression can limit the number of informative clones isolated with this technique. Recent advances in high throughput DNA sequencing technology²⁴ have enabled comprehensive analysis of every clone in the library following a single round of selection^{25–29}. By coupling substrate phage display with high throughput sequencing, we recently characterized a comprehensive VWF73 mutagenesis library, and showed that substitutions within the P3-P2' interval were among the most deleterious to proteolysis by ADAMTS13³⁰.

To further characterize the active site specificity of ADAMTS13, we now report comprehensive protease specificity profiling by combining random 6 amino acid substrate phage display and high throughput sequencing. As proof-of-concept, we define the most comprehensive substrate specificity profile for thrombin to-date, confirming known requirements for Arg at P1, and revealing both positive and negative regulators of thrombin substrate recognition. The poor recognition of peptides by ADAMTS13 was expanded 17-fold when the library was inserted into the P3-P3' residues of VWF73, revealing a broader substrate recognition potential for ADAMTS13 than previously appreciated. These data confirm the importance of exosite engagement for ADAMTS13 substrate recognition, and provide a detailed substrate recognition profile that may guide identification of novel substrates.

Results

Characterization of substrate phage display library. A random 6 amino acid substrate phage display library consisting of 2.3×10^8 independent clones was constructed, which represents 3.5 X of the 20^6 possible peptide sequences. High throughput sequencing of the unselected library confirmed the broad representation of sequences in the library (Figure S1, Fig. 1A) and revealed >5.5 million unique peptides (Table 1). More than 1 million peptides were identified by only a single sequencing read, likely a consequence of the library depth exceeding sequencing read depth. Each amino acid was comparably distributed across all 6 positions (Fig. 1B) with only modest deviation from expected frequencies (Fig. 1C). Stop codons should be limited in the FUSE55 phage display system because premature termination of the bacteriophage PIII protein abolishes phage assembly. Consistent with this prediction, only 0.04% of sequencing reads contained a stop codon, substantially lower than the 17% expected within the synthesized oligonucleotide.

Thrombin Selection. To confirm the utility of high throughput sequencing to identify phage displaying cleavable peptides from a single round of selection, we screened the serine protease thrombin (Figure S2). Thrombin is a well-characterized serine protease, with known substrate recognition determinants. Out of 5.3×10^6 unique peptide sequences identified following thrombin selection (Table 1, Figure S3), 6722 peptides were significantly enriched, and identified as cleaved ($p_{\text{FDR}} < 0.05$, Figure S4) (see Supplementary data 1). Analysis of selected phage sequences confirms a general preference for Arg and exclusion of acidic amino acids in the cleaved peptides (Fig. 2A). Arg was the dominant amino acid within the most significantly cleaved peptides (Fig. 2B), consistent with the known requirement at P1 of thrombin substrates⁷. Of the 18 cleaved peptides lacking Arg, 14 contained Lys.

Although thrombin shows preference for Arg, 906/1992 significantly depleted peptides (uncleaved) contained at least one Arg residue. To determine amino acid motifs that promote or antagonize thrombin cleavage at Arg, all peptides containing an Arg in the cleaved and uncleaved peptide pools were aligned by assigning the Arg as P1 (see Methods) and compared (Fig. 3A). Low molecular weight amino acids at the presumptive P2 and P1' positions promoted cleavage, with P2 Pro and P1' Ser the dominant residues (Fig. 3B). In contrast, bulky aliphatic amino acids at P2 or P1' antagonized cleavage, but promoted cleavage when present at more distal sites (Fig. 3C). By contrast, acidic and basic amino acids throughout the peptide antagonized thrombin cleavage. Analysis of peptides containing multiple Arg residues indicated that Arg at presumptive position P2 and/or P1' antagonize thrombin substrate recognition (Fig. 3B). Analysis of cleaved and uncleaved peptides containing only a single Arg residue (Figure S5) yielded comparable results, suggesting that multiple Arg residues within a single peptide did not appreciably confound data analysis.

ADAMTS13 Selection. Compared with thrombin, ADAMTS13 appears to exhibit narrow substrate specificity, since VWF is its only known substrate¹¹. Consistent with this observation, only 96 cleaved peptides were identified from the random peptide library following overnight selection by ADAMTS13 ($p_{\text{FDR}} < 0.05$, Figure S4). Although cleaved peptides preferentially contained bulky hydrophobic amino acids (Fig. 2C), no obvious motif was observed (Figure S6), consistent with previous studies that demonstrate poor recognition of short peptidyl substrates by ADAMTS13^{17,21}.

VWF73(P3-P3') selection by ADAMTS13. To address the role of exosite interactions in ADAMTS13 substrate recognition, the P3-P3' residues within VWF73 were replaced with random amino acids. The VWF73(P3-P3') library contained $\sim 2.5 \times 10^7$ independent clones, and high throughput sequencing showed the expected nucleotide composition (Figure S7), although amino acid frequencies deviated from expected (Figure S8), likely reflecting biases in displayed peptides due to phage production.

Following treatment of the VWF73(P3-P3') library with ADAMTS13, 1670 cleaved peptides were detected ($p_{\text{FDR}} < 0.05$, Supplementary data 2). Overall, bulky aliphatic amino acids were preferred in the enriched peptide pool, whereas acidic amino acids, as well as proline and cysteine, appeared to antagonize ADAMTS13 substrate recognition (Fig. 4A,B). The native amino acid sequence for VWF within this interval (Leu-Val-Tyr-Met-Val-Thr) was among the top peptides identified ($p_{\text{FDR}} = 8 \times 10^{-5}$) (Fig. 4C), and none of the most significantly cleaved peptides exhibited faster substrate performance than wild type VWF73 (Table 2). As a result, peptides with lower P-values than wild type VWF73 are not necessarily cleaved more efficiently.

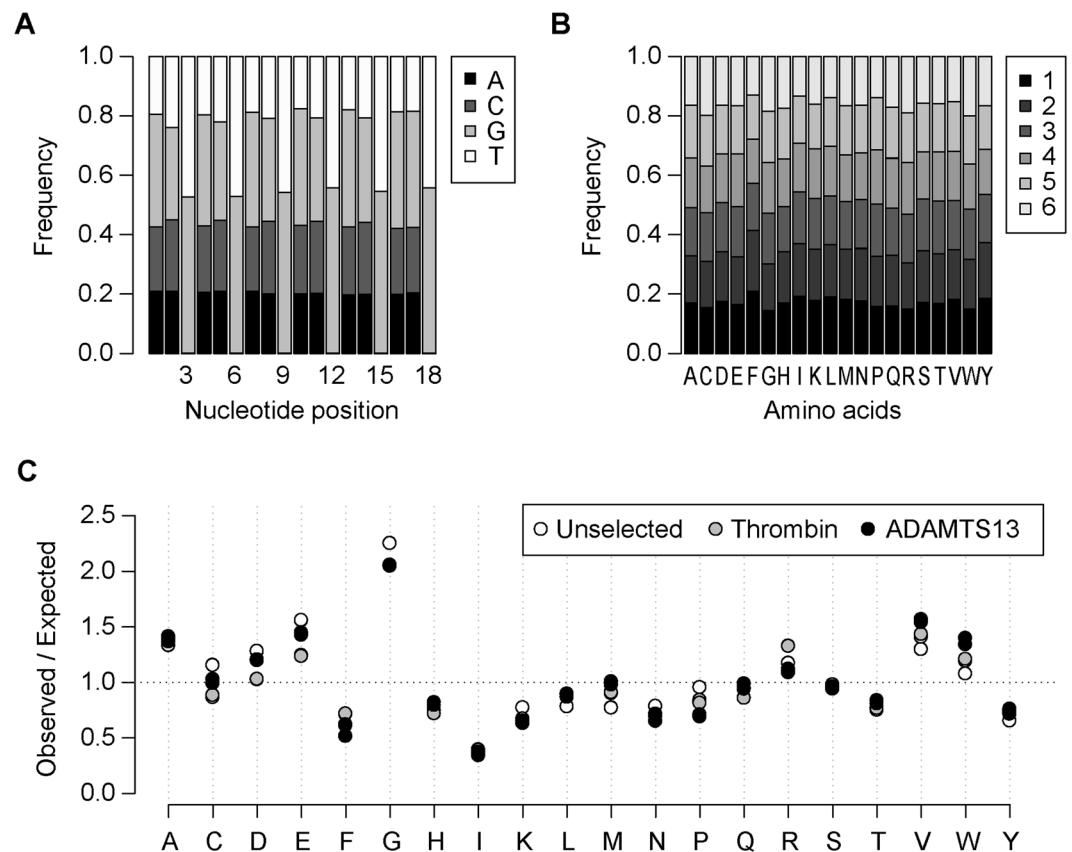


Figure 1. Nucleotide and Amino Acid distribution in NNK library. **(A)** The frequency of each nucleotide was calculated at all 18 positions within the unselected random 6 amino acid library. As expected, the first 2 positions of each codon contain all 4 nucleotides (N), in roughly equal proportions, while the third position contains only G and T (K). **(B)** The frequency of each amino acid was calculated for each position of the library. The representations for each amino acid across the 6 positions are roughly equivalent. TAG (91.8% of stops) was over-represented in the library compared to TAA (1.7% of stops) and TGA (6.5% of stops), consistent with NNK randomization. The presence of stop codons within the library suggests a mechanism to bypass stop codons at low levels during phage assembly, either through ribosomal read-through or alternative translation start sites. Because adenine should not be present in the third nucleotide position in the NNK randomization scheme, the lower representation of TAA and TGA compared to TAG likely represents background sequencing errors within the dataset. **(C)** The frequency of each amino acid was normalized to the number of codons in the NNK library for the unselected library (FLAG) and after selection of peptides cleaved by thrombin or ADAMTS13. In the unselected library and after both selections, the frequency of each amino acid varies from the frequency expected based on the number of codons for that amino acid in the NNK randomization scheme. The amino acid distribution within the unselected library likely reflects both codon usage for each amino acid as well as biased phage production for clones bearing peptides with particular amino acid content. The preference of either thrombin or ADAMTS13 for peptides with a specific amino acid content should also be observed as deviation from frequencies in the unselected library.

Comparing the cleaved and uncleaved VWF73(P3-P3') peptides reveals a coherent ADAMTS13 substrate recognition motif. At 5 out of 6 positions, the corresponding residue in VWF is among the most significantly enriched amino acids (Fig. 5). Approximately 75% of cleaved peptides contained Leu, with 34% containing Leu at amino acid position 1. Hierarchical cluster analysis of cleaved peptides trained against uncleaved peptides indicates a general preference for bulky aliphatic residues and exclusion of electrostatic amino acids and proline (Fig. 6A). Specifically, (Leu/Ile)1, Tyr3, (Leu/Tyr/Met/Phe)4 provided 75% of the predictive capacity for ADAMTS13 substrate recognition, relative to a randomly selected uncleaved sample (Fig. 6B), indicating their dominant roles in substrate recognition by ADAMTS13. In wild type VWF73, position 4 corresponds to the P1' residue. This residue has previously been shown to be critical for metalloprotease substrate recognition^{31–33}, consistent with the dominant feature for bulky aliphatic residues at position 4. Overall, these data suggest a substrate specificity profile for ADAMTS13 largely dictated by bulky aliphatic amino acids.

We recently reported a comprehensive kinetic characterization for nearly every amino acid substitution at each position of VWF73³⁰. Comparing these previous results with the current analysis revealed a strong correlation between experimental datasets (Fig. 7). Variants at positions 1, 3, and 5 showed the strongest correlation

Sample	Total reads	Passed filter (%)	Unique peptides	Mean count	Median count	Min-Max
Unselected	12,327,473	89.8	5,536,697	2.00	1	1–144
	13,296,736	89.7	5,791,291	2.06	1	1–170
Thrombin	12,334,072	89.7	5,324,398	2.08	1	1–254
	12,240,464	90.0	5,316,032	2.07	1	1–225
ADAMTS13	11,166,160	89.1	5,211,441	1.91	1	1–74
	12,076,701	89.9	5,379,005	2.02	1	1–71

Table 1. Random 6 amino acid peptide library. The results of the high throughput sequencing data analysis pipeline are shown for two samples of the unselected random 6 amino acid peptide library, and following two selections of this library by thrombin and ADAMTS13. The table shows the total number of sequencing reads per sample (total reads), the percentage of reads that passed the quality filters (passed filter) and the total number of unique peptides that were ultimately identified (unique peptides). Also shown is the average number of sequencing counts for each unique peptide (mean count), the median number of counts for each peptide (median count), and the range of counts for each unique peptide (min-max).

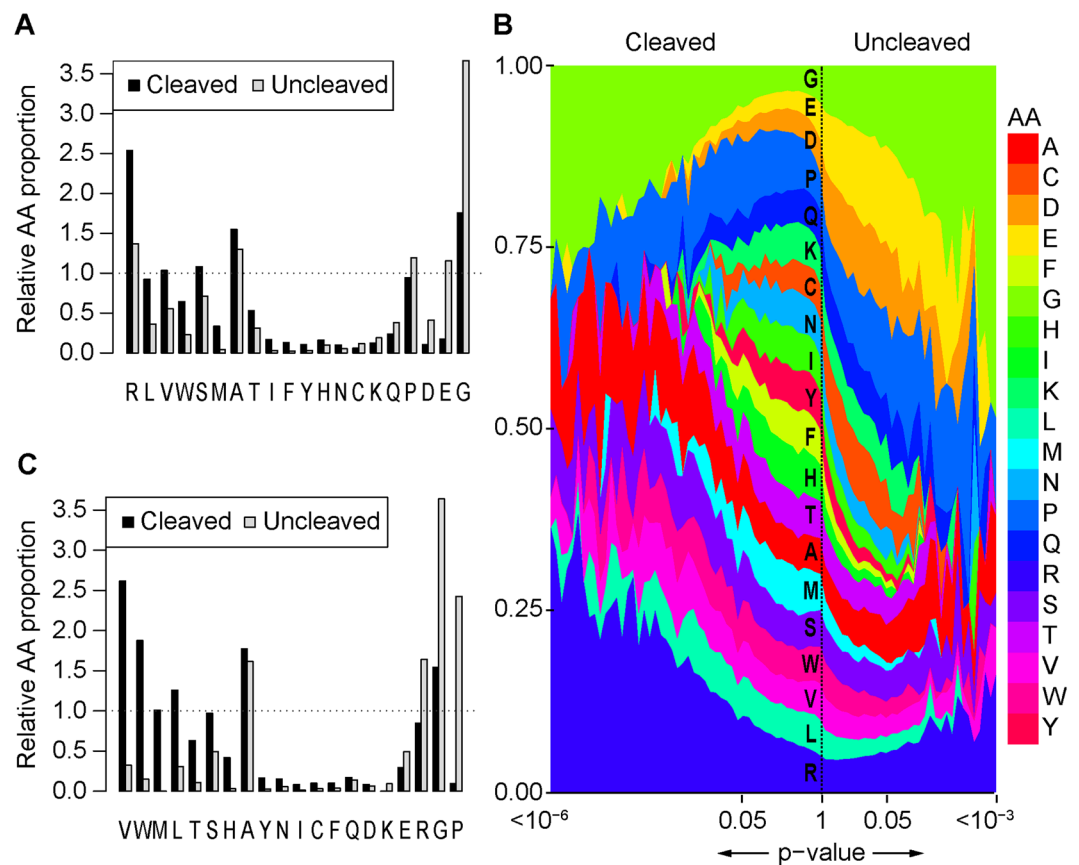


Figure 2. Amino acid enrichment and depletion for thrombin and ADAMTS13. **(A)** The relative proportion of each amino acid in the significantly cleaved or uncleaved peptides following thrombin selection compared to all peptides assessed in the experiment. Amino acids are sorted according to the difference in their relative proportion in cleaved peptides compared to uncleaved peptides. **(B)** The frequency of each amino acid was calculated for peptides grouped by p-value, with the most significantly enriched peptides on the left and most significantly depleted peptides on the right. Each data point is calculated as an average of frequencies for more than 10 peptides belonging to the same p-value subgroup on the logarithmic scale. Arg is the most abundant amino acid among the significantly enriched peptides, whereas Asp and Glu are the most abundant among the depleted peptides, and are not present at all among the enriched peptides. Gly is abundant in both enriched and depleted peptides, but is more abundant within the depleted peptides, suggesting a net-antagonistic role in thrombin substrate specificity. **(C)** Same as **(A)** but for ADAMTS13 selection.

($R > 0.7$), whereas position 6 exhibited the weakest correlation (R of ~ 0.33). These data indicate that while most amino acid substitutions in the P3-P3' interval inhibit proteolysis relative to wild type VWF³⁰, many changes are tolerated and can ultimately be cleaved by ADAMTS13.

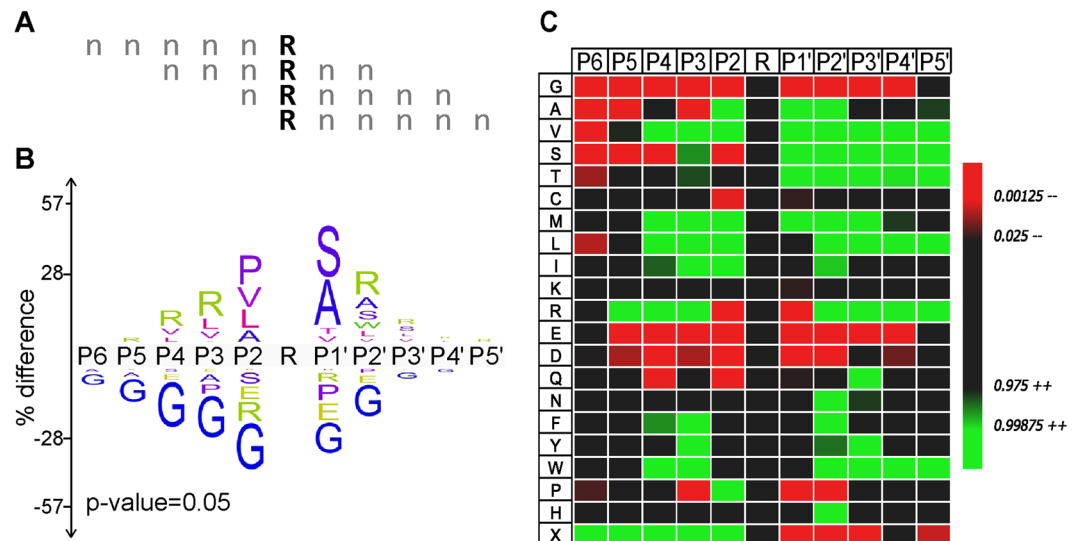


Figure 3. Thrombin Substrate motif. **(A)** Schematic overview for alignment of peptides assuming Arg as P1. For peptides containing multiple Arg residues, the center-most Arg was aligned, with preference given to position 3 Arg residues in peptides containing Arg residues at position 3 and 4. **(B)** The iceLogo plot representing the relative frequency of every amino acid in the cleaved (top) and uncleaved (bottom) peptide pools following thrombin selection. **(C)** The iceLogo heatmap shows the preference for each amino acid at every position in the cleaved (green) or uncleaved (red) peptide pools. Residues at positions that do not appreciably populate either pool are also indicated (black).

Discussion

We have generated a comprehensive catalog of the substrate specificity for thrombin and ADAMTS13 based on a phage display library of random 6 amino acid peptides. As expected, thrombin exhibited strong preference for Arg and a weaker preference for Lys, consistent with the P1 requirements of known natural substrates. In addition to defining preferred amino acids, our data reveal negative regulators of thrombin substrate recognition including acidic amino acids, Pro at any position except P2, and Ser at any position except P1'. In contrast to thrombin, ADAMTS13 exhibited poor recognition of random hexamer substrates. The number of cleaved peptides by ADAMTS13 was expanded >17-fold when residues P3-P3' of VWF73, a known ADAMTS13 substrate that also contains exosite binding residues, were replaced by random 6 amino acid peptides. These data suggest that exosite interactions are required for substrates to gain access to the ADAMTS13 active site. Overall, these data provide the most comprehensive set of substrate recognition peptides for these proteases, and illustrate distinct modes of specificity determination.

Thrombin. Thrombin is the final effector protease of blood coagulation and participates in both amplification and attenuation of the clotting system. As such, thrombin is one of the most widely characterized proteases in the human genome^{7,34–36}. Our data identify a comprehensive set of thrombin-recognized peptides, exceeding 6,700 unique peptide sequences, expanding on previous reports. Our findings demonstrate the most restricted amino acid diversity within the P2-P1' interval, consistent with the idea that 3 amino acid peptide substrates can effectively discriminate thrombin specificity³⁴. Although P2 Pro dominates thrombin natural substrates, it is not an absolute requirement for proteolysis at P1 Arg, with other low molecular weight amino acids such as Ala, Val, and Leu also found at P2 in our data. By contrast, bulky or electrostatic amino acids at this position abrogated substrate recognition. These observations are consistent with the crystal structure of thrombin, illustrating an apolar S2 pocket marked by Trp215 of thrombin^{37,38}. The comparably shallow S1' pocket³⁹ also excludes bulky amino acids, consistent with the over-representation of lower molecular weight amino acids at the P1' position. Bulky hydrophobic amino acids, such as Trp, Tyr, and Met, emerged in the extended substrate positions (P5-P3 and P2'-P4') consistent with previous library screens and natural thrombin substrate alignments⁷. These amino acids are expected to fill vacant pockets previously observed in crystal structures of thrombin complexed with hirudin, and likely stabilize substrate interactions^{39–41}.

These data are consistent with previous phage display screens of thrombin (summarized in⁷), but with a few notable exceptions. Although previous studies have demonstrated a preference for Gly at P2, our data show a much higher proportion of Gly in the uncleaved peptides, suggesting a net-antagonistic role in thrombin substrate recognition. This difference can likely be explained by the fact that Gly was the most abundant amino acid in our library. As a result Gly is expected to be found in cleavable peptides by chance and does not itself support thrombin interactions with substrates. Indeed, Gly was previously identified at all amino acid positions⁷, further supporting a nonspecific role in substrate recognition. These data also highlight the power of high throughput sequencing coupled to substrate phage display. The simultaneous quantifying of enrichment and depletion for millions of unique peptide sequences in a single protease reaction provides greater power to detect subtle effects on substrate recognition than was previously possible.

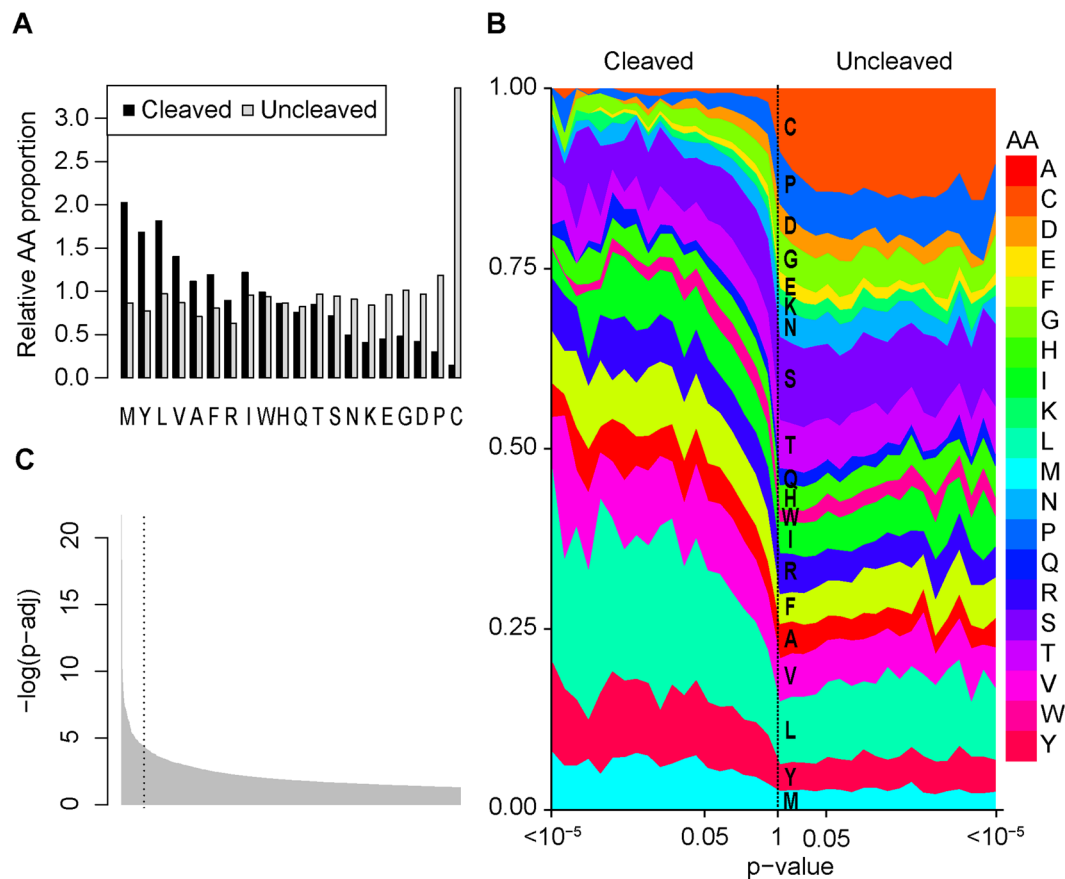


Figure 4. ADAMTS13 selection of VWF73(P3-P3') random peptide library. **(A)** The relative proportion of each amino acid from significantly cleaved or uncleaved peptides following ADAMTS13 selection of the VWF73(P3-P3') library compared to all unique peptides. **(B)** The frequency of each amino acid was calculated for peptides grouped by p-value, with the most significantly cleaved peptides on the left and most significantly uncleaved peptides on the right. Each data point is calculated as an average of frequencies for more than 10 peptides belonging to the same p-value subgroup on the logarithmic scale. **(C)** Rank-order p-value of all significantly cleaved peptides following VWF73(P3-P3') selection by ADAMTS13. The P3-P3' residues of native VWF (Leu-Val-Tyr-Met-Val-Trp), dashed line, was among the most enriched peptides.

VWF73 P3-P3' peptide	Fold-change (\log_2)	k_{cat}/K_M ($\times 10^5 \text{ M}^{-1} \text{ min}^{-1}$)
LVYMVT (WT)	0.84	300
LELYLS	0.49	0.15
IQLFLA	0.70	0.54
RLRYFL	0.79	1.74
IMMFLG	0.65	3.18
LRYSM	0.66	1.44
LGLEHS	-1.20	No cleavage
LSVYGS	-1.09	No cleavage
NLQLIF	-1.79	No cleavage
SSWVMC	-1.76	No cleavage
APPVDS	-1.68	No cleavage

Table 2. Kinetic characterization of VWF73 library peptides. The top-ranked peptides from the VWF73(P3-P3') library following selection by ADAMTS13 were identified from the most significantly enriched (Log_2 fold change >0) or most significantly depleted (Log_2 fold change <0) based on adjusted P-value. Phage were cloned with wild type VWF73 P3-P3' residues replaced with the residues listed. Each individual phage clone was reacted with ADAMTS13 and cleavage was monitored at various reaction time points using AlphaLISA as previously described^{30,53}. Individual k_{cat}/K_M values were calculated for clones exhibiting detectable proteolysis by ADAMTS13. Clones with no detectable proteolysis are indicated as 'no cleavage'.

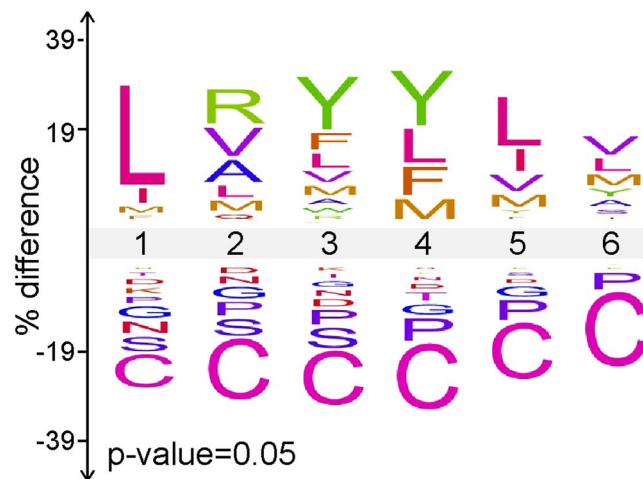


Figure 5. ADAMTS13 substrate recognition motif. The iceLogo plot representing the relative frequency of every amino acid in the cleaved (top) and uncleaved (bottom) peptide pools following ADAMTS13 selection of the VWF73(P3-P3') library.

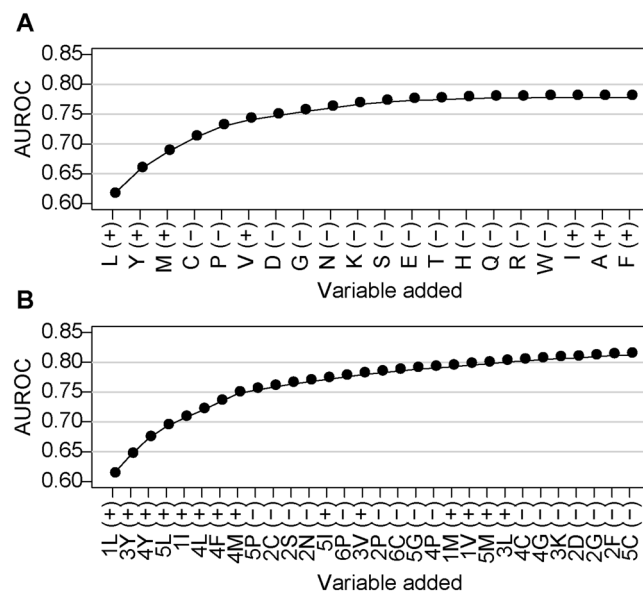


Figure 6. AUROC of the ADAMTS13 substrate motif. Logistic regression with forward stepwise feature selection was applied to ADAMTS13 selection of the VWF73(P3-P3') library, with significantly cleaved peptides trained against neither significantly cleaved nor significantly depleted. This analysis provided a measure of which amino acid requirements were most useful for predicting ADAMTS13 substrate recognition, and how well a model based on the presence of these amino acids predicted peptide cleavage. Outcomes were iterated for amino acid regardless of position (A) or amino acids constrained to position (B). Both positive (+) and negative (-) regulators of ADAMTS13 substrate recognition were identified. Amino acid content alone yielded an Area Under Receiver Operating Curve (AUROC) of ~0.78 (A), whereas amino acid content at defined positions yielded a maximum AUROC of 0.8 (B).

ADAMTS13. VWF is currently the only known substrate for ADAMTS13¹¹, which could suggest a narrow substrate profile. Consistent with this hypothesis, ADAMTS13 cleaved only 96 peptides from the random peptide library. The enriched peptides preferentially contained bulky hydrophobic amino acids but revealed no coherent motif, suggesting poor substrate recognition within this comprehensive library. These findings are consistent with a previous report demonstrating a greater than 1500-fold reduction in the k_{cat}/K_M for proteolysis of VWF by ADAMTS13 in the absence of exosite interactions¹⁷. Our library theoretically surveys all possible 6 amino acid peptide sequences, and therefore confirms the notion that ADAMTS13 does not efficiently recognize short peptidyl substrates^{17,21}.

Recently, a mechanism of ADAMTS13 auto-regulation was described in which COOH-terminal CUB domains interacting with the NH₃-terminal spacer domain^{42,43}. A mechanism was proposed whereby exosite

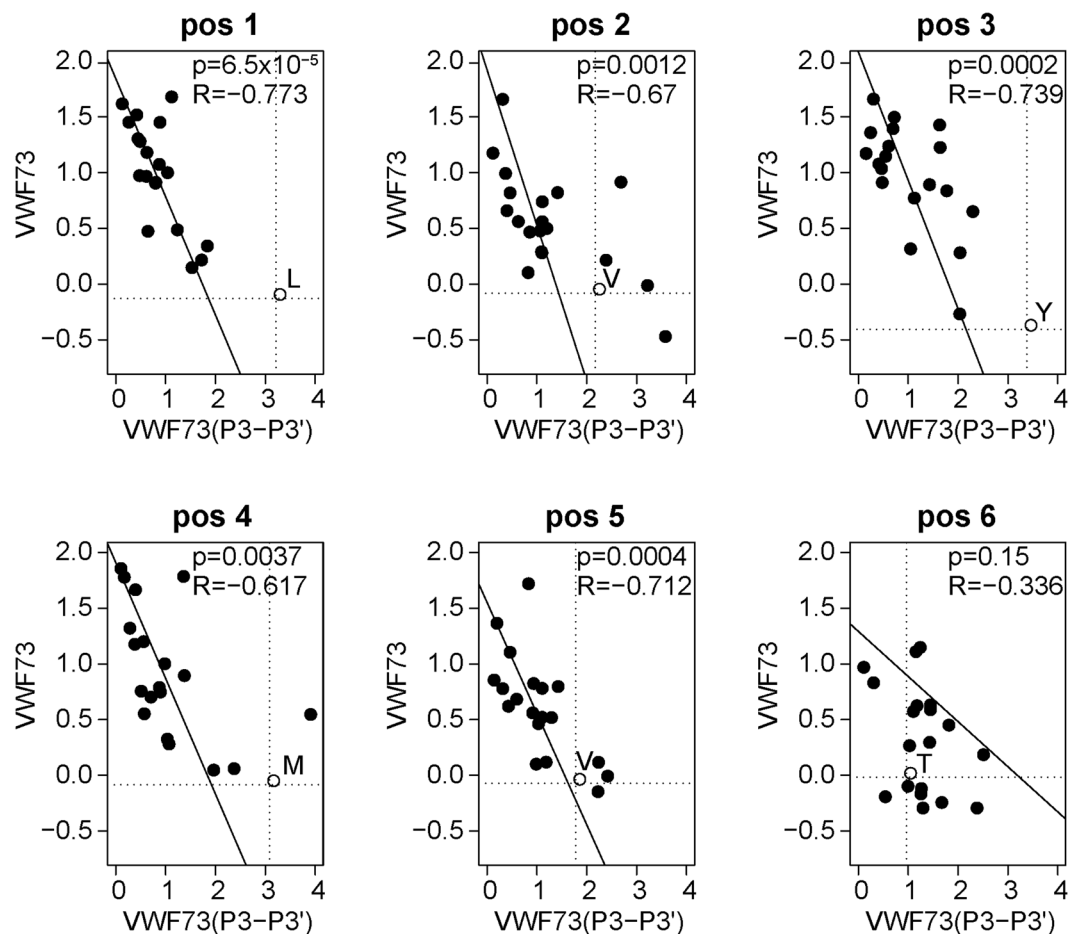


Figure 7. Correlation between ADAMTS13 selection of VWF73 mutagenesis and VWF73(P3-P3') libraries. ADAMTS13 selection of the VWF73(P3-P3') library was compared to previously reported VWF73 mutagenesis data³⁰, where substitutions tended to inhibit cleavage of wild type VWF73. Thus, the enrichment for each substitution within the P3-P3' interval³⁰ was correlated to the relative proportions of each amino acid in enriched compared to depleted peptides from the VWF73(P3-P3') library selection. Each panel represents an amino acid position, and each data point represents an amino acid substitution. The wild type amino acid at each position is indicated at the intersection of the dashed lines. Values represented in each plot correspond to p-values calculated for each correlation (top) and the R value (bottom) for the linear regression (line).

engagement activates ADAMTS13 by relieving this auto-regulation in addition to aligning the substrate scissile bonds toward the active site. Consistent with this mechanism, we observed that ADAMTS13 recognition of a random peptide library was expanded >17-fold when the random peptide library was expressed within the context of an exosite-binding substrate. These data may suggest that access to the active site is impaired when ADAMTS13 adopts its closed conformation.

Alignment of cleaved peptides revealed a distinct substrate recognition motif for ADAMTS13. Our data indicate that long-chain aliphatic amino acids at P3 (including Leu, Ile, and Met) are a dominant feature for ADAMTS13 substrate recognition, consistent with previous findings which highlight the importance of the P3 residue for ADAMTS13 substrate recognition⁴⁴. Overall, substrate recognition for ADAMTS13 exhibits a general requirement for aliphatic and aromatic residues throughout, including Tyr at P1, and Leu, Tyr, Met, and Phe at P1'. Although no crystal structure of the ADAMTS13 metalloprotease domain is currently available, the structure for the corresponding domain in ADAMTS5 (which shares 28% amino acid sequence identity and 42% similarity with ADAMTS13) has been solved⁴⁵. This structure reveals a hydrophobic active site cleft with a deep S1' pocket, characteristic of other metalloproteases of the metzincin family, that is known to accept bulky aliphatic residues at the P1' position of substrates. However, the structure of the ADAMTS5 protease domain does not identify a binding site for the P3 residue^{30,44}. Previous studies demonstrated that ADAMTS13 residues Asp187-Arg193 forms a subsite within the metalloprotease domain that flanks the active site and contributes to recognition of the VWF scissile bond⁴⁶. Interestingly, the charged residues within this loop (D187, R190, and R193) appeared to make the greatest contribution to substrate recognition. How these residues influence the selectivity of peptides containing bulky hydrophobic amino acids in the VWF73(P3-P3') library remains to be determined. Overall, these data suggest that ADAMTS13 is capable of recognizing and cleaving proteins other than VWF only if exosites are

simultaneously engaged. The consensus motif and list of cleavable peptides may facilitate the discovery of novel physiological substrates of ADAMTS13.

We previously interrogated the interaction between ADAMTS13 and VWF73 using a comprehensive mutagenesis substrate phage display library and showed that the P3-P2' interval is among the most critical regions driving ADAMTS13 substrate recognition³⁰. The data reported here are highly concordant with this previous report, providing a more detailed investigation of the P3-P3' interval. Together, these studies provide a broad framework for comprehensive protease profiling that complement or expand upon existing technologies^{3,10,47,48}.

However, we acknowledge a number of potential limitations to our approach. First, this technique does not define the P1-P1' site of cleavage for each peptide identified. In the case of thrombin, the strategy of aligning peptides by fixing an Arg residue is supported by extensive investigation over many decades, as well as the identification of very similar motifs for peptides containing a single Arg compared to peptides containing multiple Arg residues. For ADAMTS13 cleavage of VWF73(P3-P3'), exosite interactions within VWF73 may restrict ADAMTS13 cleavage to the 3rd or 4th position of the hexamer library, though cleavage elsewhere in the P3-P3' interval cannot be excluded. For example, the presence of Tyr and Phe at position 4 of cleaved peptides may be indicative of the P1 residue shifting from position 3 in certain peptides. As a result, the motif generated from the VWF73(P3-P3') library may be incomplete.

The reaction conditions employed here are expected to result in the proteolytic reaction proceeding to completion, providing great sensitivity to detect even weak substrates, but limiting quantitative comparison among cleaved peptides. For example, 5 of the most significantly cleaved peptides from the VWF73(P3-P3') library (Supplementary data 2) did not cleave as efficiently as wild type VWF73, which was 135th most heavily selected by the cleavage assay (Table 2). Thus, the possibility that select peptides within in this library may still exhibit increased efficiency as ADAMTS13 substrates compared to WT cannot be excluded.

Despite these limitations, our findings demonstrate the power of coupling substrate phage display to high throughput sequencing to provide a rapid and robust platform for comprehensive protease profiling. Current high throughput sequencing technology provides the capacity to sequence ~300 million molecules in parallel (Illumina). This capacity allows precise enrichments to be calculated for every library clone, and statistical interpretations of the data after a single round of selection. This approach avoids biases in phage infection and re-amplification that commonly confound traditional phage display biopanning experiments⁴⁹. Furthermore, recent advances in oligonucleotide array synthesis allow for rationally designed substrate libraries and more precise control over library composition^{50,51}. As these technologies continue to improve, the capacity to investigate more comprehensive libraries will expand and yield new insights into protease specificity determination. Ultimately, these studies could facilitate the identification of novel physiological protease substrates, development of more specific biochemical or clinical tools to assess protease activity, and support the development of specific protease inhibitors to treat important human diseases.

Methods

Phagemid Modification. The fUSE55 vector⁵² was modified to contain a cotranslational-translocation signaling sequence and NH₂- and COOH-terminal epitope tags (See Table S1 for complete oligonucleotide list). A FLAG tag was first inserted into the phagemid, pAY-E⁵³, at the NotI and SgrAI sites using annealed oligomers, **P1** and **P2**, generating pAY-FE. Tandem FLAG and E epitope tags followed by a glycine-serine rich linker were amplified from pAY-FE with primers, **P3** and **P4**, and inserted into fUSE55 at the BglII site, generating fUSE65. The TorT (i.e., cotranslational-translocation) signaling sequence was fused to transcriptional regulatory elements of fUSE55 by PCR using primers **P5-P7**, and subsequently inserted at the BsrGI and SfiI sites of fUSE65 to generate fUSE66. For fUSE67, oligomers **P8** and **P9** were annealed and extended using standard PCR protocols and inserted into fUSE66 at the SfiI and SgrAI sites. The resulting features of fUSE67 vector are arranged: 5'-TorT signaling sequence, FLAG tag, T7 tag, multiple cloning site, E tag, glycine-serine rich linker, and gIII-3'. All expected modifications were verified by Sanger DNA sequencing. All oligonucleotides were from Integrated DNA Technologies (Coralville, Iowa).

Construction of substrate phage display libraries. Three distinct phage display libraries were generated to evaluate the substrate recognition patterns of thrombin and ADAMTS13. The random nucleotide libraries were either inserted into fUSE67, or designed to contain a FLAG-tag 5' to the variable region before cloning into the fUSE55 phage display vector^{52,54}. Both fUSE67 and fUSE55 place the substrate on all copies of the PIII protein of M13 filamentous phage.

To construct the random 6 amino acid substrate phage display library, the NNK degenerate codon series was used, where N represents an equal 25% proportion of A, C, G, and T, and K represents equal 50% proportion of G and T. Thus, 10 ng of the NNK oligonucleotide **L1** was used as a template in a PCR reaction containing 1 μM **S1** and 1 μM **AS1** primers (Table S2) using the following thermal profile for 30 cycles: 95 °C (30 s), 60 (30 s), 72 (30 s).

The PCR product was gel purified on 1.5% agarose and extracted using the QIAquick Gel Purification Kit (Qiagen), and digested with BglI (NEB). All restriction digested products were prepared for ligation using agarose gel purification followed by electroelution using the ELUTRAP system (GE Healthcare). The digested and purified oligonucleotides were ligated into 1 μg of fUSE55 using a 6:1 molar ratio (insert:vector). The ligation mixture was incubated at 16 °C overnight, precipitated, and resuspended in TE buffer (20 mM TRIS-HCl, pH 8.0, 1 mM EDTA). The ligation product was electroporated into MegaX DH10B *E. coli* (Invitrogen), and the library was titrated, revealing a total library depth of 2.5×10^8 independent clones.

Random 6 amino acid peptide libraries were also constructed in the context of VWF73 (Asp1596-Arg1668 of VWF), replacing the codons for Leu8-Thr13 with the degenerate codon series, NNK. Two approaches for the library construction were undertaken. In the first approach (VWF73(P3-P3')-1), the NNK randomization was

tailed onto the forward primer with 1 ng of VWF cDNA in pBlueScript SK+ used as template in a PCR reaction containing 1 μ M S2 and 1 μ M AS2 (Table S4), using Herculase II (Agilent). The PCR product was gel purified as above and used as template in a PCR reaction containing 1 μ M S3 and 1 μ M AS2. The PCR product was gel purified as above and used as template in a final PCR reaction containing 1 μ M S4 and 1 μ M AS2. The PCR product was gel purified as above. In all cases, the PCR thermal profile was: 95 °C (30 s), 62 (30 s), 72 (30 s), repeated for 20 cycles. A second library was constructed (VWF73(P3-P3')-2, where the randomized oligonucleotide was used as a template to account for possible nucleotide bias in VWF73(P3-P3')A. A single PCR reaction was assembled containing 1 nM L2, 1 nM AS3, 1 nM AS4, 1 μ M AS5, and 1 μ M S5 (Table S5) using Herculase II. The PCR thermal profile was: 95 °C (30 s), 60 (30 s), 72 (30 s), repeated for 30 cycles.

In all approaches, the PCR products were digested with either BglI or AscI and NotI, gel purified using ELUTRAP, then ligated into 1 μ g FUSE55 or FUSE67 at a 6:1 molar ratio (insert:vector) overnight at 16 °C. The ligation product was precipitated, resuspended in TE buffer, and electroporated into MegaX DH10B *E. coli*. The libraries were titrated onto 30 μ g/mL tetracycline Luria Broth (LB) agar plates revealing 3×10^7 independent clones for VWF73(P3-P3')A and 1×10^7 independent clones for VWF73(P3-P3')B. For the two VWF73(P3-P3') libraries, no major differences in library composition were detected by high throughput sequencing, and datasets were combined for final analysis.

Panning. The phage libraries were prepared as previously described³⁰. Approximately 1×10^{10} phage were added to 1 mL TBS-B (20 mM Tris-HCl pH 7.4, 150 mM NaCl, 1% BSA) containing 50 μ L anti-FLAG agarose beads (Sigma), and mixed at room temperature for 2 hr. The beads were recovered by gentle centrifugation ($3000 \times g$ for 1 min) and washed 5 times with TBS-B. The phage-coated beads were then resuspended with 500 μ L reaction buffer (20 mM Tris-HCl, pH 7.4, 150 mM NaCl, 5 mM CaCl₂, 10 μ M ZnCl₂, and 1% BSA) containing 5 nM thrombin (Hematologic Technologies) or 5 nM ADAMTS13 (R&D Systems). These reaction conditions have previously been shown to result in efficient hydrolysis of peptidyl substrates for both thrombin⁵⁵ and ADAMTS13⁵⁶. The reaction was incubated overnight with end-over-end mixing at room temperature. The beads were recovered by centrifugation, and the supernatant containing phage displaying cleaved peptides was recovered. For the control samples containing no protease, unreacted phage bound to anti-FLAG beads were eluted using 500 μ L 0.15 mg/ml 3X FLAG peptide. Single stranded DNA (ssDNA) was prepared as previously described³⁰.

Deep sequencing. Unselected and selected phage ssDNA were used as templates in PCR reactions to prepare samples for high throughput sequencing to evaluate enrichment following panning, as previously described³⁰. For all samples, an initial barcoding PCR was performed using primers listed in Table S3 for the random peptide substrate phage display library and Table S6 for VWF73(P3-P3'). The thermal profile was: 98 °C (30 s), 62 °C (30 s), 72 °C (30 s). The number of cycles was determined empirically to prevent product laddering, assessed by agarose gel electrophoresis. To complete the assembly of Illumina library adapters, a second PCR was performed using 10 ng of the barcoded PCR product as template and 0.5 μ M of PE1seq and PE2seq primers (Table S3). The thermal profile was: 98 °C (30 s), 60 °C (30 s), and 72 °C (30 s). PCR products were gel purified on 1% agarose.

Illumina library quality was assessed by qPCR using the Library Quantification Kit (KK4835, Kapa Biosystems) and the Agilent DNA 1000 Bioanalyzer kit (5067-1504, Agilent), according to manufacturer's instructions. Libraries were sequenced on a HiSeq2500 (Illumina) using paired-end 50 base pair reads in Rapid Mode.

Recombinant phage and peptide validation. The results of the VWF73(P3-P3') screen were validated in part using purified recombinant peptide clones. Recombinant phage and peptides were purified and k_{cat}/K_M values determined as previously described⁵³. All oligonucleotides used to assemble the clones are provided in Table S8.

Sequencing analysis pipeline and QC analysis. Sequence filtering and peptide analysis were performed using an in-house pipeline written in Python and are available for download (github.com/tombergk/NNK_VWF73/). A number of quality filters were applied to the paired-end reads from the.fastq files (Figure S1). First, one of the reads from each pair (forward or reverse) was compared to one of three 8 bp seed sequences within the forward primer region to orient the sequence. The multiple seeds allowed for sequencing errors to be tolerated at this initial stage without discarding the read. Second, a perfect match of nucleotides between the sense and antisense reads was required within the variable coding region. This highly stringent quality filter should reduce sequencing errors within the library to 0.01%, assuming a 1% error rate per sequence⁵⁷. Finally, a base pair quality score of at least 5 out of 40 was required from each position within variable coding region. Stop codons were evaluated (see Results) but removed from subsequent analyses. Because the FUSE55 (and FUSE67) phage display system places a displayed peptide on all PIII proteins, a stop codon within the library should abrogate PIII production and prevent phage assembly. As a result, any occurrence of stop codons in the library is likely due to sequencing errors, although occasional ribosome read-through cannot be excluded. All paired-end sequences that passed the above quality filters were translated into corresponding peptides and the occurrence of each unique peptide was recorded. Biases in amino acid content between the random 6 amino acid peptide library and VWF73(P3-P3') are shown in Table S7.

Generated.fastq files have been deposited to the NCBI Sequence Read Archive (project accession number #PRJNA356764) found at <https://www.ncbi.nlm.gov/sra>. The project encompasses 3 sets of paired-end high throughput sequencing.fastq files used in our pipelines: #SRR5097080, #SRR5097081, #SRR5097082.

Motif definition and determination. Peptides containing a minimum of 4 reads combined in selected and unselected controls were analyzed. Enrichment and depletion of peptides was assessed using the DESEQ2 software package⁵⁸, which estimated variance-mean dependence in peptide counts from selected and unselected phage and tested for differential expression using a negative binomial distribution. Peptides with Benjamini-Hochberg⁵⁹ adjusted p-values (p_{FDR}) < 0.05 were considered significant for both enrichment and depletion. All significantly enriched and depleted peptides from the selections are available as supplemental files. Amino acid frequency plots and heatmaps were created using the iceLogo package⁶⁰, where the ratio of amino acid frequencies in the enriched peptides was compared to depleted peptides. In the case of thrombin, all peptides containing a single Arg were aligned and centered around Arg to assess the amino acid dependency in this context.

References

1. Drag, M. & Salvesen, G. S. Emerging principles in protease-based drug discovery. *Nature reviews. Drug discovery* **9**, 690–701, <https://doi.org/10.1038/nrd3053> (2010).
2. Yi, L. *et al.* Engineering of TEV protease variants by yeast ER sequestration screening (YESS) of combinatorial libraries. *Proceedings of the National Academy of Sciences of the United States of America* **110**, 7229–7234, <https://doi.org/10.1073/pnas.1215994110> (2013).
3. Turk, B. E., Huang, L. L., Piro, E. T. & Cantley, L. C. Determination of protease cleavage site motifs using mixture-based oriented peptide libraries. *Nature biotechnology* **19**, 661–667, <https://doi.org/10.1038/90273> (2001).
4. Chen, E. I. *et al.* A unique substrate recognition profile for matrix metalloproteinase-2. *The Journal of biological chemistry* **277**, 4485–4491, <https://doi.org/10.1074/jbc.M109469200> (2002).
5. Kridel, S. J. *et al.* A unique substrate binding mode discriminates membrane type-1 matrix metalloproteinase from other matrix metalloproteinases. *The Journal of biological chemistry* **277**, 23788–23793, <https://doi.org/10.1074/jbc.M111574200> (2002).
6. Hedstrom, L. Serine protease mechanism and specificity. *Chemical reviews* **102**, 4501–4524 (2002).
7. Gallwitz, M., Enoksson, M., Thorpe, M. & Hellman, L. The extended cleavage specificity of human thrombin. *PLoS One* **7**, e31756, <https://doi.org/10.1371/journal.pone.0031756> (2012).
8. Kridel, S. J. *et al.* Substrate hydrolysis by matrix metalloproteinase-9. *The Journal of biological chemistry* **276**, 20572–20578, <https://doi.org/10.1074/jbc.M100900200> (2001).
9. Ratnikov, B. I. *et al.* Basis for substrate recognition and distinction by matrix metalloproteinases. *Proceedings of the National Academy of Sciences of the United States of America*, <https://doi.org/10.1073/pnas.1406134111> (2014).
10. Eckhard, U. *et al.* Active site specificity of the matrix metalloproteinase family: Proteomic identification of 4300 cleavage sites by nine MMPs explored with structural and synthetic peptide cleavage analyses. *Matrix biology: journal of the International Society for Matrix Biology*, <https://doi.org/10.1016/j.matbio.2015.09.003> (2015).
11. Crawley, J. T., de Groot, R., Xiang, Y., Luken, B. M. & Lane, D. A. Unraveling the scissile bond: how ADAMTS13 recognizes and cleaves von Willebrand factor. *Blood* **118**, 3212–3221, <https://doi.org/10.1182/blood-2011-02-306597> (2011).
12. Zhang, Q. *et al.* Structural specializations of A2, a force-sensing domain in the ultralarge vascular protein von Willebrand factor. *Proceedings of the National Academy of Sciences of the United States of America* **106**, 9226–9231, <https://doi.org/10.1073/pnas.0903679106> (2009).
13. Zhang, X., Halvorsen, K., Zhang, C. Z., Wong, W. P. & Springer, T. A. Mechanoenzymatic cleavage of the ultralarge vascular protein von Willebrand factor. *Science* **324**, 1330–1334, <https://doi.org/10.1126/science.1170905> (2009).
14. Xu, A. J. & Springer, T. A. Mechanisms by which von Willebrand disease mutations destabilize the A2 domain. *The Journal of biological chemistry* **288**, 6317–6324, <https://doi.org/10.1074/jbc.M112.422618> (2013).
15. Levy, G. G. *et al.* Mutations in a member of the ADAMTS gene family cause thrombotic thrombocytopenic purpura. *Nature* **413**, 488–494, <https://doi.org/10.1038/35097008> (2001).
16. Kokame, K., Matsumoto, M., Fujimura, Y. & Miyata, T. VWF73, a region from D1596 to R1668 of von Willebrand factor, provides a minimal substrate for ADAMTS-13. *Blood* **103**, 607–612, <https://doi.org/10.1182/blood-2003-08-2861> (2004).
17. Gao, W., Anderson, P. J. & Sadler, J. E. Extensive contacts between ADAMTS13 exosites and von Willebrand factor domain A2 contribute to substrate specificity. *Blood* **112**, 1713–1719, <https://doi.org/10.1182/blood-2008-04-148759> (2008).
18. de Groot, R., Lane, D. A. & Crawley, J. T. The role of the ADAMTS13 cysteine-rich domain in VWF binding and proteolysis. *Blood* **125**, 1968–1975, <https://doi.org/10.1182/blood-2014-08-594556> (2015).
19. de Groot, R., Bardhan, A., Ramroop, N., Lane, D. A. & Crawley, J. T. Essential role of the disintegrin-like domain in ADAMTS13 function. *Blood* **113**, 5609–5616, <https://doi.org/10.1182/blood-2008-11-187914> (2009).
20. Lynch, C. J., Lane, D. A. & Luken, B. M. Control of VWF A2 domain stability and ADAMTS13 access to the scissile bond of full-length VWF. *Blood* **123**, 2585–2592, <https://doi.org/10.1182/blood-2013-11-538173> (2014).
21. Gao, W., Anderson, P. J., Majerus, E. M., Tuley, E. A. & Sadler, J. E. Exosite interactions contribute to tension-induced cleavage of von Willebrand factor by the antithrombotic ADAMTS13 metalloprotease. *Proc Natl Acad Sci USA* **103**, 19099–19104, <https://doi.org/10.1073/pnas.0607264104> (2006).
22. Hills, R. *et al.* Identification of an ADAMTS-4 cleavage motif using phage display leads to the development of fluorogenic peptide substrates and reveals matrilin-3 as a novel substrate. *The Journal of biological chemistry* **282**, 11101–11109, <https://doi.org/10.1074/jbc.M611588200> (2007).
23. Carlos, F. B. Dennis, R. B. III, Jamie K. S. & Gregg J. S. *Phage Display: A laboratory manual*. Vol. 1 (Cold Spring Harbor Laboratory Press, 2001).
24. Shendure, J. & Lieberman Aiden, E. The expanding scope of DNA sequencing. *Nature biotechnology* **30**, 1084–1094, <https://doi.org/10.1038/nbt.2421> (2012).
25. Fowler, D. M. *et al.* High-resolution mapping of protein sequence-function relationships. *Nature methods* **7**, 741–746, <https://doi.org/10.1038/nmeth.1492> (2010).
26. Araya, C. L. *et al.* A fundamental protein property, thermodynamic stability, revealed solely from large-scale measurements of protein function. *Proceedings of the National Academy of Sciences of the United States of America* **109**, 16858–16863, <https://doi.org/10.1073/pnas.1209751109> (2012).
27. Larman, H. B. *et al.* Autoantigen discovery with a synthetic human peptidome. *Nature biotechnology* **29**, 535–541, <https://doi.org/10.1038/nbt.1856> (2011).
28. Zhu, J. *et al.* Protein interaction discovery using parallel analysis of translated ORFs (PLATO). *Nature biotechnology* **31**, 331–334, <https://doi.org/10.1038/nbt.2539> (2013).
29. Larman, H. B., Liang, A. C., Elledge, S. J. & Zhu, J. Discovery of protein interactions using parallel analysis of translated ORFs (PLATO). *Nature protocols* **9**, 90–103, <https://doi.org/10.1038/nprot.2013.167> (2014).
30. Kretz, C. A. *et al.* Massively parallel enzyme kinetics reveals the substrate recognition landscape of the metalloprotease ADAMTS13. *Proc Natl Acad Sci USA* **112**, 9328–9333, <https://doi.org/10.1073/pnas.1511328112> (2015).
31. Bode, W. *et al.* The X-ray crystal structure of the catalytic domain of human neutrophil collagenase inhibited by a substrate analogue reveals the essentials for catalysis and specificity. *EMBO J* **13**, 1263–1269 (1994).

32. Eckhard, U. *et al.* Active site specificity profiling of the matrix metalloproteinase family: Proteomic identification of 4300 cleavage sites by nine MMPs explored with structural and synthetic peptide cleavage analyses. *Matrix Biol* **49**, 37–60, <https://doi.org/10.1016/j.matbio.2015.09.003> (2016).
33. Minond, D. *et al.* The roles of substrate thermal stability and P2 and P1' subsite identity on matrix metalloproteinase triple-helical peptidase activity and collagen specificity. *J Biol Chem* **281**, 38302–38313, <https://doi.org/10.1074/jbc.M606004200> (2006).
34. Vindigni, A., Dang, Q. D. & Di Cera, E. Site-specific dissection of substrate recognition by thrombin. *Nat Biotechnol* **15**, 891–895, <https://doi.org/10.1038/nbt0997-891> (1997).
35. Ng, N. M. *et al.* Discovery of amino acid motifs for thrombin cleavage and validation using a model substrate. *Biochemistry* **50**, 10499–10507, <https://doi.org/10.1021/bi201333g> (2011).
36. Ng, N. M. *et al.* The effects of exosite occupancy on the substrate specificity of thrombin. *Archives of biochemistry and biophysics* **489**, 48–54, <https://doi.org/10.1016/j.abb.2009.07.012> (2009).
37. Bode, W. *et al.* 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. *The EMBO journal* **8**, 3467–3475 (1989).
38. Bode, W. & Stubbs, M. T. Spatial structure of thrombin as a guide to its multiple sites of interaction. *Seminars in thrombosis and hemostasis* **19**, 321–333, <https://doi.org/10.1055/s-2007-993283> (1993).
39. Matthews, J. H., Krishnan, R., Costanzo, M. J., Maryanoff, B. E. & Tulinsky, A. Crystal structures of thrombin with thiazole-containing inhibitors: probes of the S1' binding site. *Biophysical journal* **71**, 2830–2839, [https://doi.org/10.1016/S0006-3495\(96\)79479-1](https://doi.org/10.1016/S0006-3495(96)79479-1) (1996).
40. Qiu, X. *et al.* Structure of the hirulog 3-thrombin complex and nature of the S' subsites of substrates and inhibitors. *Biochemistry* **31**, 11689–11697 (1992).
41. Qiu, X., Yin, M., Padmanabhan, K. P., Krstenansky, J. L. & Tulinsky, A. Structures of thrombin complexes with a designed and a natural exosite peptide inhibitor. *The Journal of biological chemistry* **268**, 20318–20326 (1993).
42. South, K. *et al.* Conformational activation of ADAMTS13. *Proc Natl Acad Sci USA* **111**, 18578–18583, <https://doi.org/10.1073/pnas.1411979112> (2014).
43. Muia, J. *et al.* Allosteric activation of ADAMTS13 by von Willebrand factor. *Proc Natl Acad Sci USA* **111**, 18584–18589, <https://doi.org/10.1073/pnas.1413282112> (2014).
44. Xiang, Y., de Groot, R., Crawley, J. T. & Lane, D. A. Mechanism of von Willebrand factor scissile bond cleavage by a disintegrin and metalloproteinase with a thrombospondin type 1 motif, member 13 (ADAMTS13). *Proc Natl Acad Sci USA* **108**, 11602–11607, <https://doi.org/10.1073/pnas.1018559108> (2011).
45. Shieh, H. S. *et al.* High resolution crystal structure of the catalytic domain of ADAMTS-5 (aggrecanase-2). *The Journal of biological chemistry* **283**, 1501–1507, <https://doi.org/10.1074/jbc.M705879200> (2008).
46. de Groot, R., Lane, D. A. & Crawley, J. T. The ADAMTS13 metalloprotease domain: roles of subsites in enzyme activity and specificity. *Blood* **116**, 3064–3072, <https://doi.org/10.1182/blood-2009-12-258780> (2010).
47. O'Donoghue, A. J. *et al.* Global identification of peptidase specificity by multiplex substrate profiling. *Nature methods* **9**, 1095–1100, <https://doi.org/10.1038/nmeth.2182> (2012).
48. Ratnikov, B., Cieplak, P. & Smith, J. W. High throughput substrate phage display for protease profiling. *Methods in molecular biology* **539**, 93–114, https://doi.org/10.1007/978-1-60327-003-8_6 (2009).
49. t Hoen, P. A. *et al.* Phage display screening without repetitious selection rounds. *Anal Biochem* **421**, 622–631, <https://doi.org/10.1016/j.ab.2011.11.005> (2012).
50. Kosuri, S. & Church, G. M. Large-scale de novo DNA synthesis: technologies and applications. *Nature methods* **11**, 499–507, <https://doi.org/10.1038/nmeth.2918> (2014).
51. Bonde, M. T. *et al.* Direct mutagenesis of thousands of genomic targets using microarray-derived oligonucleotides. *ACS synthetic biology* **4**, 17–22, <https://doi.org/10.1021/sb5001565> (2015).
52. Parmley, S. F. & Smith, G. P. Antibody-selectable filamentous fd phage vectors: affinity purification of target genes. *Gene* **73**, 305–318 (1988).
53. Desch, K. C. *et al.* Probing ADAMTS13 substrate specificity using phage display. *PLoS One* **10**, e0122931, <https://doi.org/10.1371/journal.pone.0122931> (2015).
54. Scott, J. K. & Smith, G. P. Searching for peptide ligands with an epitope library. *Science* **249**, 386–390 (1990).
55. Stevens, W. K., Cote, H. F., MacGillivray, R. T. & Nesheim, M. E. Calcium ion modulation of meizothrombin autolysis at Arg55-Asp56 and catalytic activity. *J Biol Chem* **271**, 8062–8067 (1996).
56. Anderson, P. J., Kokame, K. & Sadler, J. E. Zinc and calcium ions cooperatively modulate ADAMTS13 activity. *J Biol Chem* **281**, 850–857, <https://doi.org/10.1074/jbc.M504540200> (2006).
57. Ross, M. G. *et al.* Characterizing and measuring bias in sequence data. *Genome biology* **14**, R51, <https://doi.org/10.1186/gb-2013-14-5-r51> (2013).
58. Moderated estimation of fold change and dispersion for RNA-Seq data with DESeq2 (Bioconductor, 2014).
59. Hochberg, Y. B. Y. Controlling the False Discovery Rate: A Practical and Powerful Approach to Multiple Testing. *Journal of the Royal Statistical Society. Series B* **57**, 289–300 (1995).
60. Colaert, N., Helsens, K., Martens, L., Vandekerckhove, J. & Gevaert, K. Improved visualization of protein consensus sequences by iceLogo. *Nature methods* **6**, 786–787, <https://doi.org/10.1038/nmeth1109-786> (2009).

Acknowledgements

The authors would like to thank Isabel Wang and Vivian Cheung (University of Michigan) for assistance with deep sequencing experiments, and Jennifer Fox for technical assistance. This work was supported by the National Institutes of Health (R01 HL039693 and P01-HL057346). D.G. is an investigator of the Howard Hughes Medical Institute. C.A.K. holds a McMaster University Department of Medicine Internal Career Award. K.T. is an International Fulbright Science and Technology Fellow and the recipient of an American Heart Association Predoctoral Fellowship Grant.

Author Contributions

C.A.K., K.T., and A.Y. designed all experiments. C.A.K. and K.T. performed all experiments. C.A.K., K.T., A.V.E., and D.G. analyzed and interpreted the data. C.A.K., K.T., and D.G. wrote the manuscript.

Additional Information

Supplementary information accompanies this paper at <https://doi.org/10.1038/s41598-018-21021-9>.

Competing Interests: The authors declare no competing interests.

Publisher's note: Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.



Open Access This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons license, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons license, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons license and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this license, visit <http://creativecommons.org/licenses/by/4.0/>.

© The Author(s) 2018