Science Advances
AAAS

advances.sciencemag.org/cgi/content/full/5/8/eaaw2851/DC1

Supplementary Materials for

Thiol-to-amine cyclization reaction enables screening of large libraries of macrocyclic compounds and the generation of sub-kilodalton ligands

S. S. Kale, M. Bergeron-Brlek, Y. Wu, M. G. Kumar, M. V. Pham, J. Bortoli, J. Vesin, X.-D. Kong, J. Franco Machado, K. Deyle, P. Gonschorek, G. Turcatti, L. Cendron, A. Angelini, C. Heinis*

*Corresponding author. Email: christian.heinis@epfl.ch

Published 21 August 2019, *Sci. Adv.* **5**, eaaw2851 (2019)
DOI: 10.1126/sciadv.aaw2851

The PDF file includes:

Other Supplementary Material for this manuscript includes the following:

(available at advances.sciencemag.org/cgi/content/full/5/8/eaaw2851/DC1)

Data S1 (.pdf format). Raw data of peptide macrocyclization reactions.
Data S2 (.pdf format). Poster showing all 432 different macrocycle scaffolds in the pilot-scale library.

Supplementary Results

Inhibitors of KLK5

Four hits identified in the KLK5 screen (indicated in fig. S5A) were characterized in detail based on the inhibition efficiency and SAR data. Specifically, the two peptides 20-Y-R-C and 20-W-R-C cyclized with linker **5** were characterized in detail because they belong to a group of macrocycle hits that share a similar structure, differing in only the second amino acid position (right upper corner in fig. S5A). The two peptides R-75-G-C and R-75-20-C cyclized with linker **6** were analyzed in detail because of their good activity, the latter one being the best hit of the screen, and the high structural similarity (the only amino acid position that differs is occupied by the similar residues Gly and βAla). The chemical structures, the analytical HPLC chromatograms, and the KLK5 inhibition data of the four macrocycles, with and without the Gly-Gly-Trp tag, are shown in fig. S6C. The four macrocycles synthesized without the Gly-Gly-Trp tag showed K_is in the medium micromolar range and thus weaker activities than expected based on the extent of inhibition observed in the screen (fig. S6C). We speculated that the GGW tag, that is conjugated to the macrocycles in the high-throughput screen, contributes to the binding. Indeed, when the four macrocycles were synthesized with the Gly-Gly-Trp tag, the macrocycles inhibited KLK5 with K_is in the high nanomolar to low micromolar range (fig. S6C). The best inhibitor was R-75-20-C+6, showing a K_i of 144 ±15 nM. This macrocycle was the best hit in the high-throughput screen, confirming that the screening strategy enables the identification of the most active macrocycles.

Overall structure of human α-thrombin

Human α-thrombin consists of two polypeptide chains of 36 (L-chain) and 259 amino acid residues (H-chain) covalently linked via a disulfide bridge (Cys122 of H-chain with Cys1 of L-chain). The L-chain of human α-thrombin can be traced unambiguously from Glu1C to Gly14M. The first five amino terminal residues (Thr1H to Gly1D) and the carboxyl-terminal residue Arg15 are undefined and not visible in the Fourier map. The electron density of the H-chain is clearly visible for all residues with the exception of eight amino acids part of the surface flexible autolysis loop (Thr147 to Lys149E). The carboxyl-terminal residue Glu247 lacks adequate

electron density. The H- and L-chains of human α-thrombin are not organized in separate domains and form a single contiguous spherical molecule that exhibits the characteristic topology of a trypsin-like serine protease. The L-chain is mainly organized in a multiple-turn conformation and is positioned along the H-chain molecular surface opposite to the active-site cleft. The H-chain structure consists of two opposed six-stranded β-barrels that are folded in an antiparallel manner and connected by turn structures and four helical regions (Ala56 – Leu59, Arg126 – Leu129C, Arg165 – Ser171 and Val231 – Gln244). Like other serine proteases, the human α-thrombin has three disulfide bridges (Cys42 – Cys58, Cys168 – Cys182 and Cys191 – Cys220) and an active site containing the catalytic triad His57, Asp102 and Ser195 residues that are located at the junction of both barrels. The overall structure of human α-thrombin in complex with P2 does not show any striking rearrangements of the main backbone if compared to other human α-thrombin structures that were determined alone or in complex with different inhibitors. Indeed, a structural similarity search against the PDB archive done by the server Dali showed that our structure displayed root mean square deviations (RMSD) of the Cα-atoms never exceeding 0.6 Å within the first 120 closest structural neighbors, composed of a pool of representative human α-thrombin structures obtained from crystals belonging to the same or different space groups in complex with a variety of inhibitors.

Overall structure of P2 macrocycle

The electron density of the macrocycle P2 is well-defined allowing an unambiguous assignment of group orientations. The numbering of the atoms in P2 is shown in fig. S7. No classical secondary structure elements are found in the molecule. The side chains of Arg, β-hydroxyproline, Cys and the *N*-2-(hydroxymethyl)benzyl group point in opposite directions. The β-homoproline is in the cis-configuration and its pyrrolidine ring (C8 – C11) forms a plane with the C23 – C28 phenyl ring of the linker **2**. The two rings are perpendicular to both hydroxyl methyl-benzyl and arginine guanidine groups. The macrocycle forms an extended structure with two intra-molecular interactions that appear to confer structural constrains to the molecule (Table S3A). The hydroxy group (O3) of the hydroxymethyl-benzyl substituent forms an intra-molecular hydrogen bond with the main chain nitrogen (N6). Additionally, the hydroxy group (O3) forms a hydrogen bond with the main chain carbonyl oxygen (O2). Intriguingly, the

macrocycle exhibits two slightly different conformations in the two-thrombin molecules present within the asymmetric unit. This could be attributed to a certain degree of flexibility existing at the level of the main chain amide bond connecting the β-hydroxyproline and the arginine groups of the macrocycle. In one conformation, the main chain oxygen O1 of P2 points toward the main chain nitrogen of Gly216 (Gly216 N) whereas in the flipped conformation the nitrogen N4 of P2 is oriented toward the main chain oxygen of the same Gly216 (Gly216 O). Based on the quality of the electron density maps we can conclude that, in the tested conditions, the latter conformation involving a hydrogen bond between N4 of P2 and the oxygen of Gly216 residue is less represented in the asymmetric unit and thus slightly less favored in the formation of the complex.

Interactions between human α-thrombin and P2

The P2 macrocycle fits very well into the cleft formed by the active site and the surrounding substrate pockets covering a protein surface of 201 Å2 (Table S3D). The pyrrolidine and aromatic rings are perfectly oriented to form a large number of hydrophobic interactions with the adjacent enzyme residues whereas the guanidine group and the main chain backbone are engaged in hydrogen bonds with nearby enzyme residues. Most interactions of P2 with human α-thrombin are mediated by the guanidine group of arginine that accommodates in the primary specificity S1 pocket. The guanidine group forms a salt bridge with the side chain carboxylic group of Asp189 (P2 N1 with Asp189 OD2) and two hydrogen bonds with the main chain oxygen of Gly219 (P2 N1 with Gly219 O) and a molecule of water (P2 N2 with H_2O 10). Additionally, the guanidine group forms two polar contacts with the main chain oxygen of Ala190 (P2 N1 with Ala190 O) and the side chain carboxylic group of Asp189 (P2 N2 with Asp189 OD1). The main chain oxygen O1 and O5 of P2 form hydrogen bonds with the main chain nitrogen of Gly216 (Gly216 N) and Gly219 (Gly219 N), respectively (Table S3B). The macrocycle' conformations and interactions are very similar in the two active sites of the two-thrombin molecules present in the asymmetric unit. P2 folds in very similar fashion in the two sites and the hydrogen-bonding networks are highly conserved. The only exception is the hydrogen bond existing between the main chain oxygen O1 of P2 and Gly216 N in one conformation and the hydrogen bond between the main chain nitrogen N4 of P2 and Gly216 O in the flipped conformation (Table S3B).

Importantly, the binding of P2 to human α-thrombin is mediated by multiple hydrophobic interactions as summarized in Table S3C. The P2 β-hydroxyproline pyrrolidine ring (C8 – C11) point towards the hydrophobic cage shaped by the side chains of residues His57, Tyr60A, Trp60D (proximal S2 pocket) and Leu99 (distal S3 pocket). The aromatic ring (C15 – C20) of the P2 hydroxyl methyl-benzyl group forms hydrophobic interactions with main and side chains of nearby residues Trp96, Arg97, Glu97A, Asn98 and Leu99. Additionally, the C23 – C28 phenyl ring establishes hydrophobic interactions with side chains of adjacent Gly216 and Glu217, and residues of the lyophilic distal S3 pocket (Thr172, Trp215). Furthermore, the residues of the S3 pocket mediate interaction with ten atoms of P2 (N1, N2, N3, O1, O5, C1, C4, C5, C22 and C33). Three P2 atoms (N1, C1 and N2) interact with the side chain of Asp189 (S1 pocket). Lastly, five atoms (C1, C2, N1, N2 and N3) form interactions with both main and side chains of Ala190, a residue that differentiates the specificity of thrombin from trypsin.

Supplementary Figures

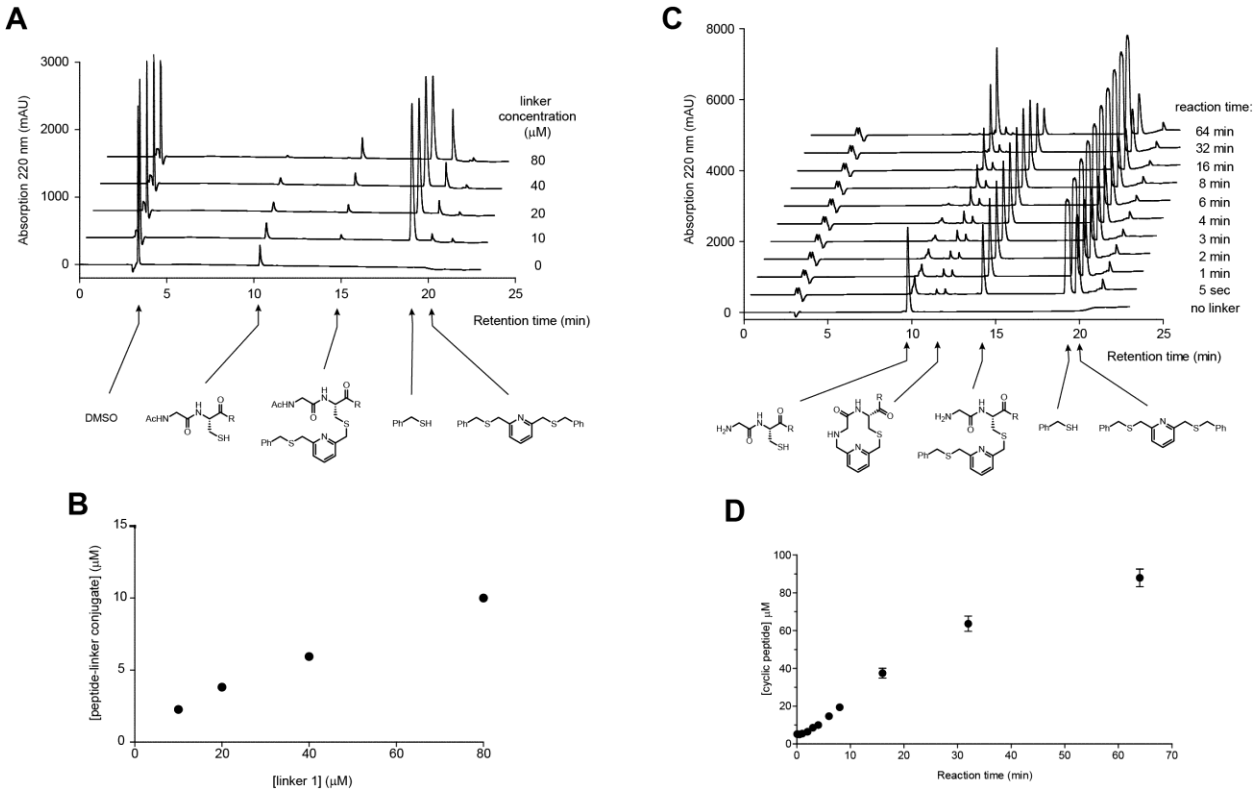


Fig. S1. Reaction kinetics of the thiol-to-amine cyclization reaction. (**A**) Inter-molecular alkylation reaction. The N-terminally capped peptide Ac-Gly-Cys containing a Gly-Trp C-terminal appendix (Trp was appended to facilitate detection at 280 nm) at a concentration of 10 μM was incubated with different concentrations of linker 1 for 5 min at 30 °C, and the reaction stopped by addition of excess benzyl mercaptan (800 μM). The reactions were analyzed by RP-HPLC. (**B**) The concentration of peptide-1-thiobenzyl adduct, quantified by integrating the area under the peaks, is shown in dependence of the linker **1** concentrations in the reaction. (**C**) Reaction kinetics of the macrocyclization reaction. The peptide Gly-Cys containing a Gly-Gly-Trp C-terminal appendix at a concentration of 100 μM was incubated with 800 μM linker **1** at 30 °C for different time periods and the reaction quenched by addition of excess benzyl mercaptan (8 mM). The reactions were analyzed by RP-HPLC. (**D**) The concentration of the expected macrocyclic product, quantified by integrating the area under the peaks, is shown in dependence of the reaction time. Mean values and variations of two measurements are shown.

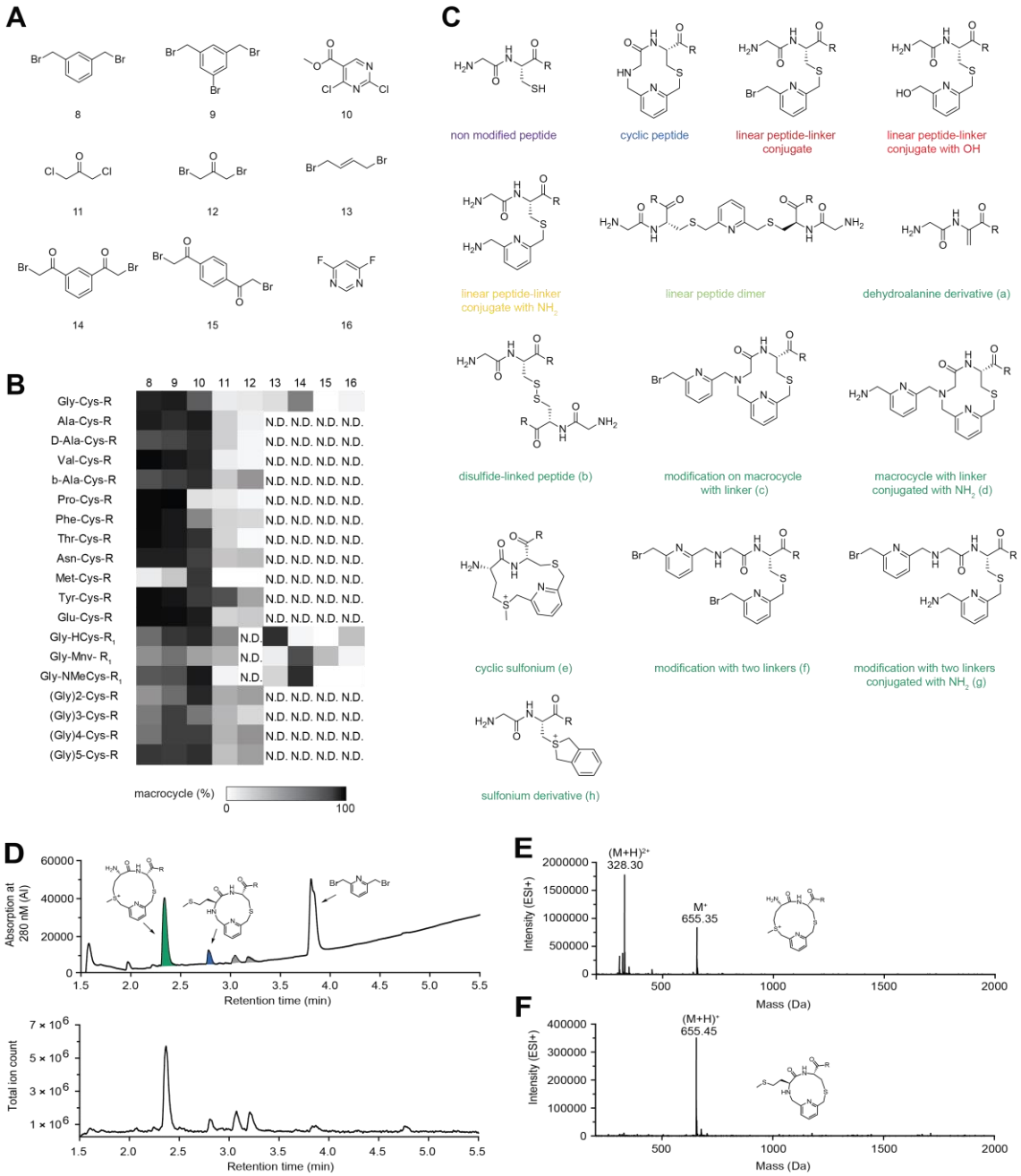


Fig. S2. Thiol-to-nitrogen cyclization reagents and side products. (**A–B**) Thiol-to-nitrogen cyclization reaction with reagents **8-16**. Chemical structures are shown in (**A**) and the yields of the desired macrocycle products are shown as percentage of peptide incubated with the linkers in (**B**). Black and white indicate 100% and 0% macrocyclization product, respectively. R = GGW,

R_1 = GW. N.D. = not determined. (**C**) Products and potential side products of the thiol-to-nitrogen macrocyclization reaction illustrated with the dipeptide Gly-Cys and linker reagent **1**. The indicated colors and letters are the same as those used to highlight the HPLC peaks in the chromatograms and to indicate the products in the tables of the Supplementary Data S1. The peptide Met-Cys is used to illustrate the cyclic sulfonium side product (e). (**D**–**E**) Incubation of Met-Cys-R (R = GGW) with di-eletrophiles yielded cyclic sulfonium products in most cases. (**D**) LC-MS analysis of Met-Cys-R (100 μM) incubated with 1 (800 μM) for 1 hr at 30 °C. Absorption (upper panel) and total ion count (lower panel) were recorded. Two peaks with products having the mass of the desired macrocyclic product were observed (highlighted in green and blue). (**E**) MS analysis of the first peak (green) showed a singly and doubly charged molecule that is expected for the cyclic sulfonium product. (**F**) MS analysis of the second peak (blue) showed a singly charged molecule expected for the thiol-to-amine cyclized peptide.

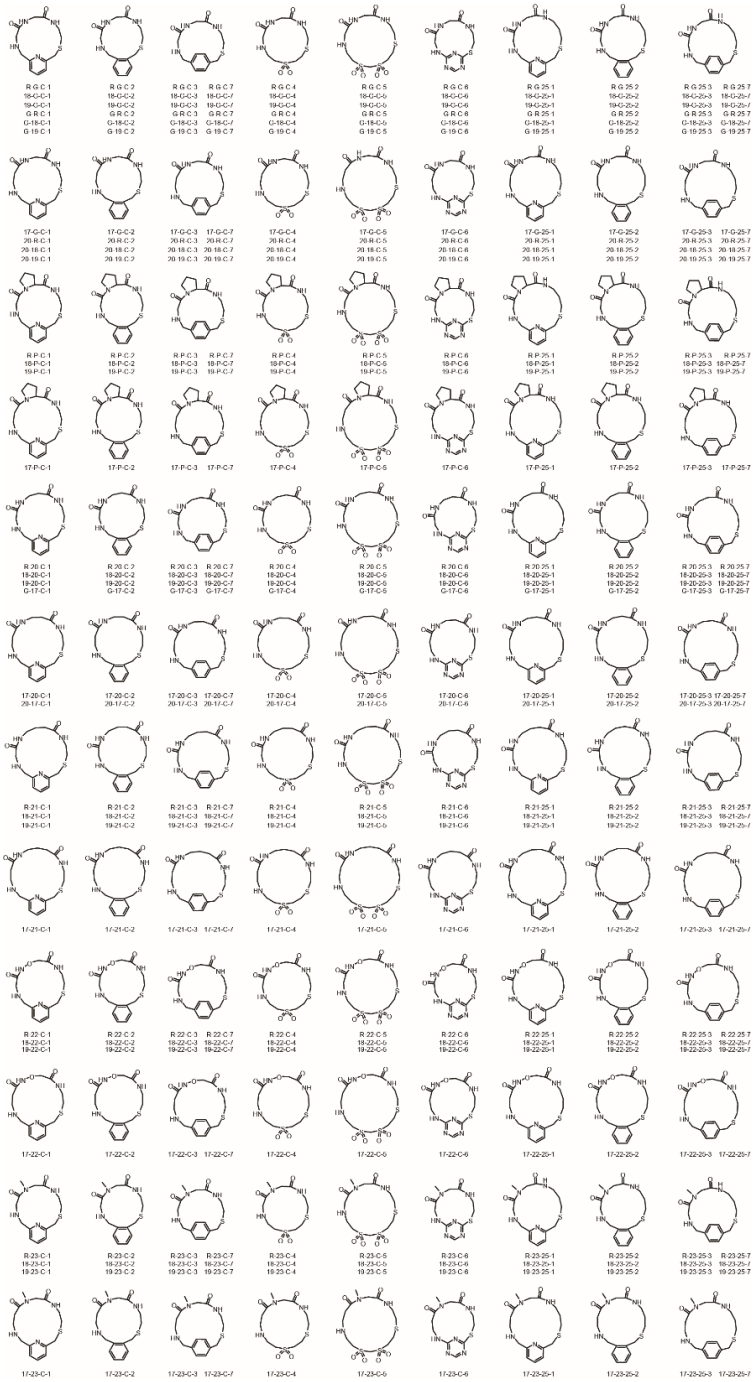


Fig. S3. Scaffold diversity in pilot-scale macrocycle library. The chemical structures of the 432 different backbones present in the library are shown. The names of the macrocycles containing the scaffolds are shown below. The backbones are arranged as the macrocycles shown in Fig. 2A.

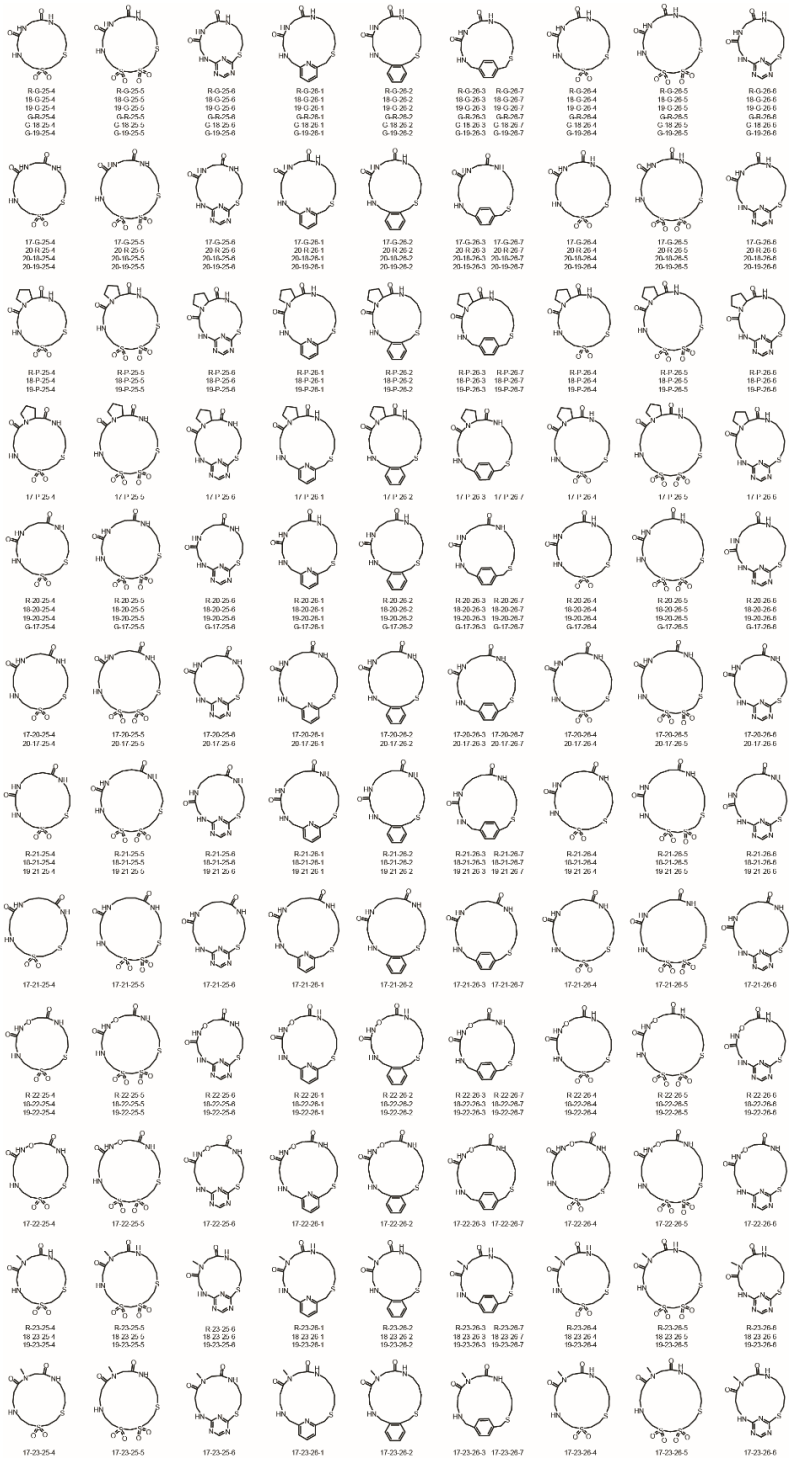


Fig. S3. Continued

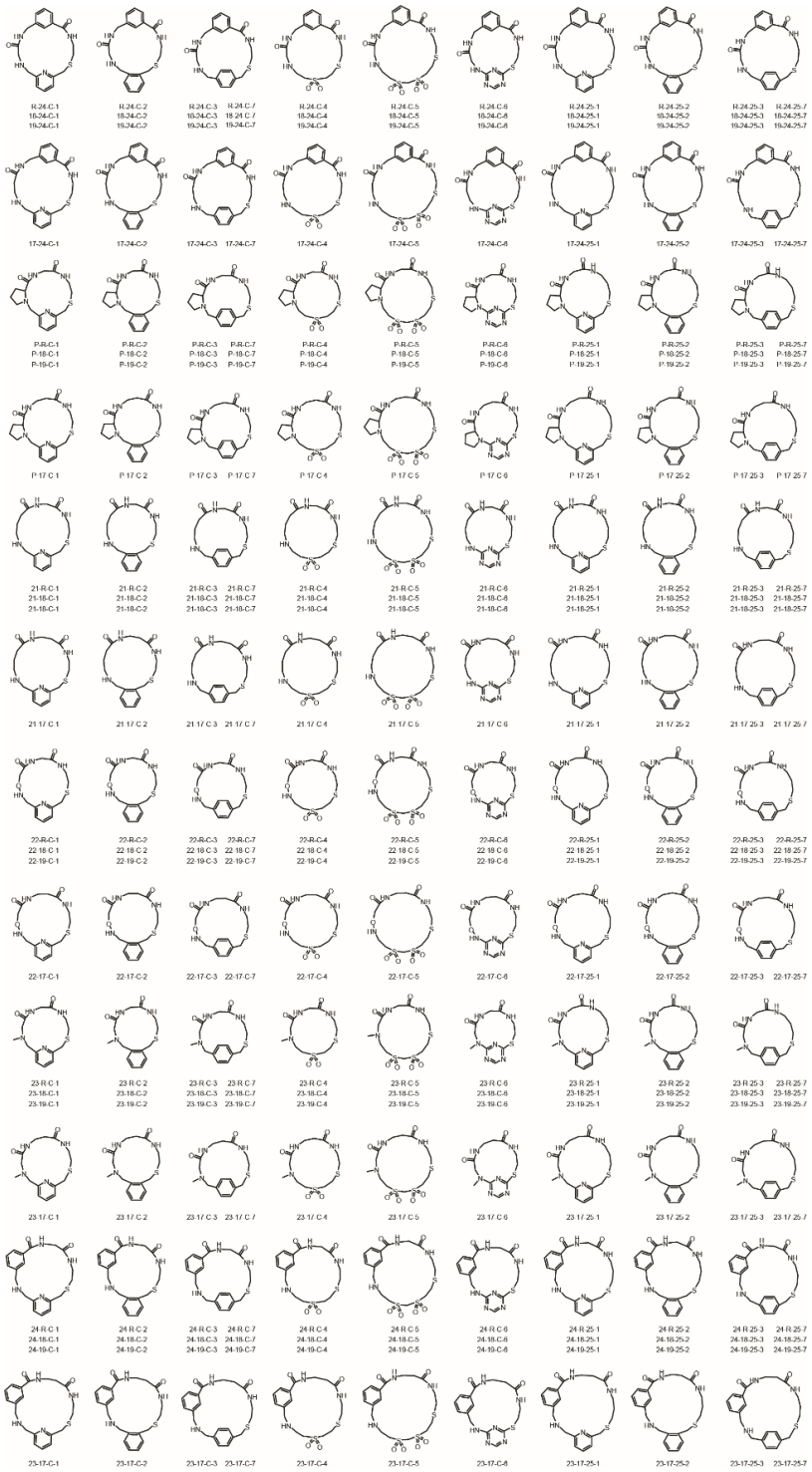


Fig. S3. Continued

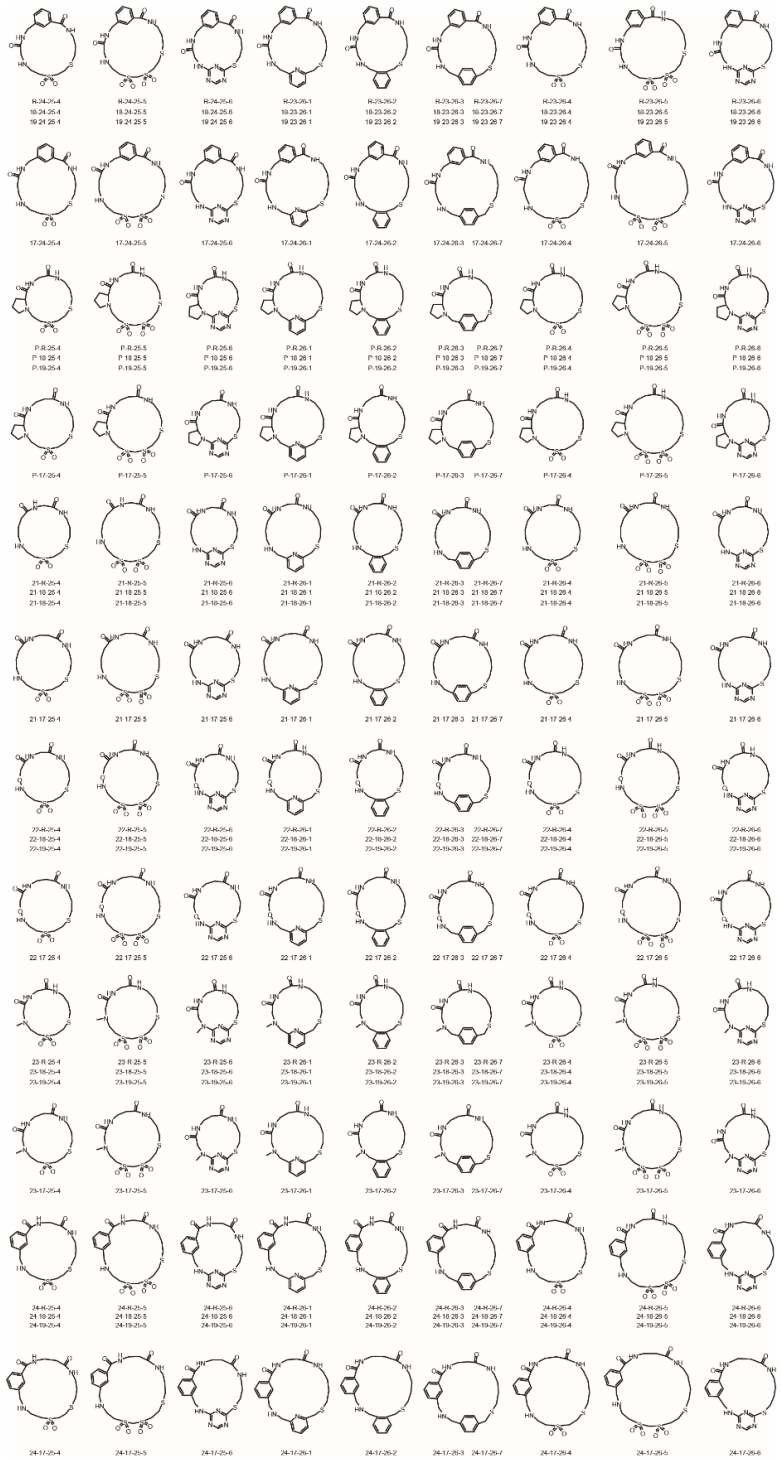


Fig. S3. Continued

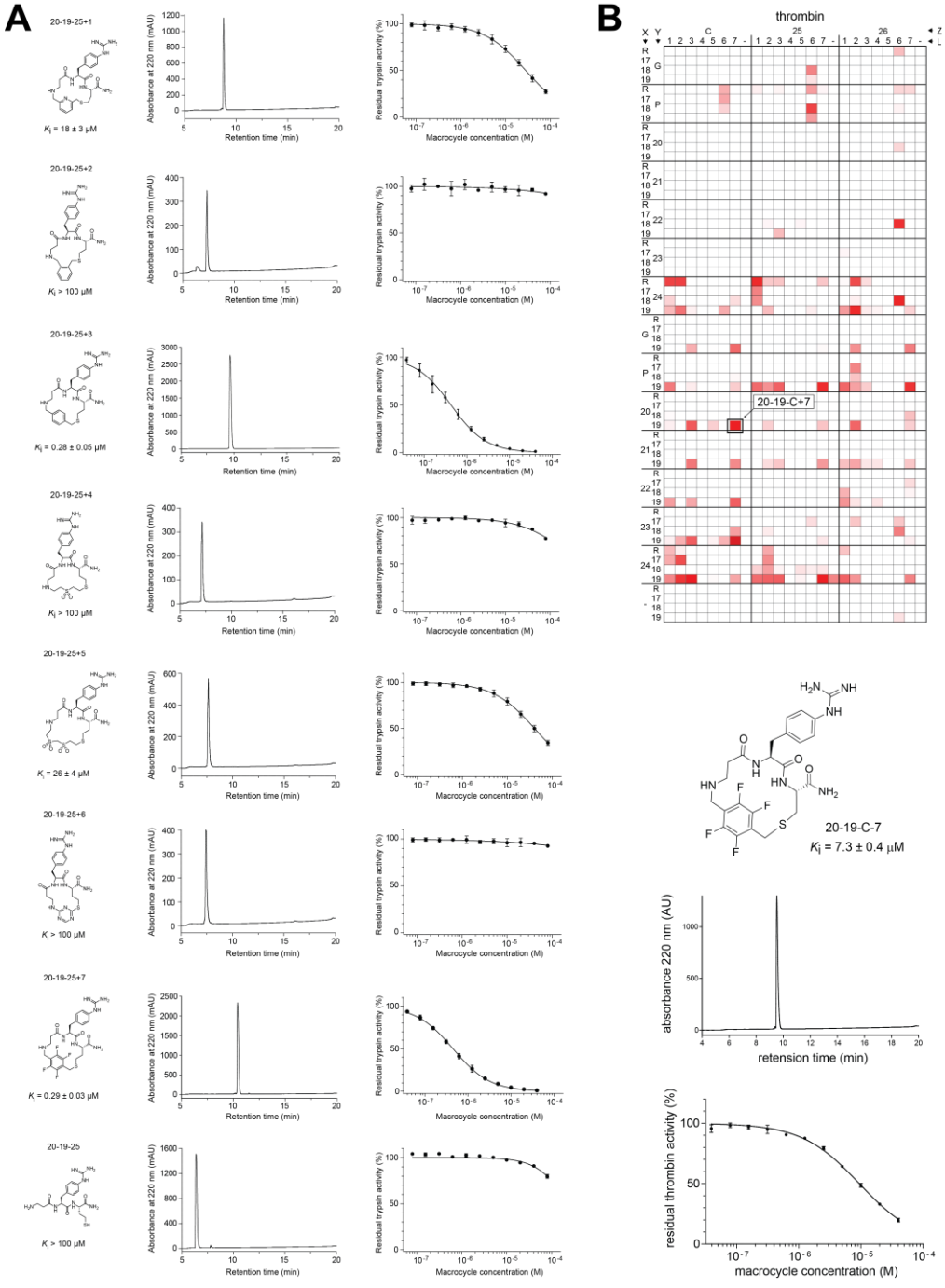


Fig. S4. Characterization of trypsin and thrombin hits. (**A**) Macrocycles based on tri-peptide 20-19-25 and inhibition of trypsin. The chemical structure and the inhibition constant for trypsin (K_i) are shown on the left. The last number in the macrocycle name indicates the chemical linker.

Mean values and SD of three measurements are indicated for the K_is. Analytic HPLC chromatograms of the macrocycles and trypsin inhibition dose-response curves are shown in the middle and on the right hand side. (**B**) Thrombin inhibitor 20-19-C+7 identified in the pilot-scale library. The chemical structure and the K_i (top right), analytical HPLC chromatogram (middle right) and the trypsin inhibition dose-response curve (bottom right) are shown. The mean value and SD of the K_i were determined based on three measurement.

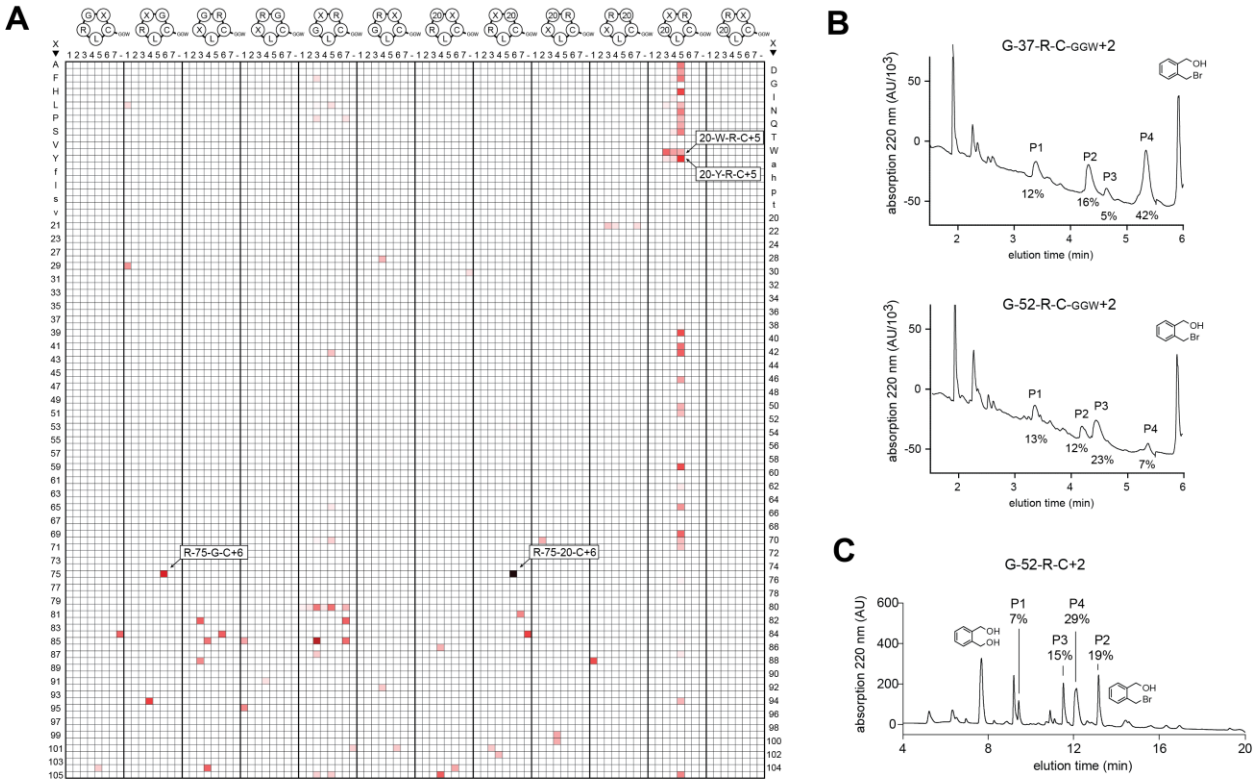
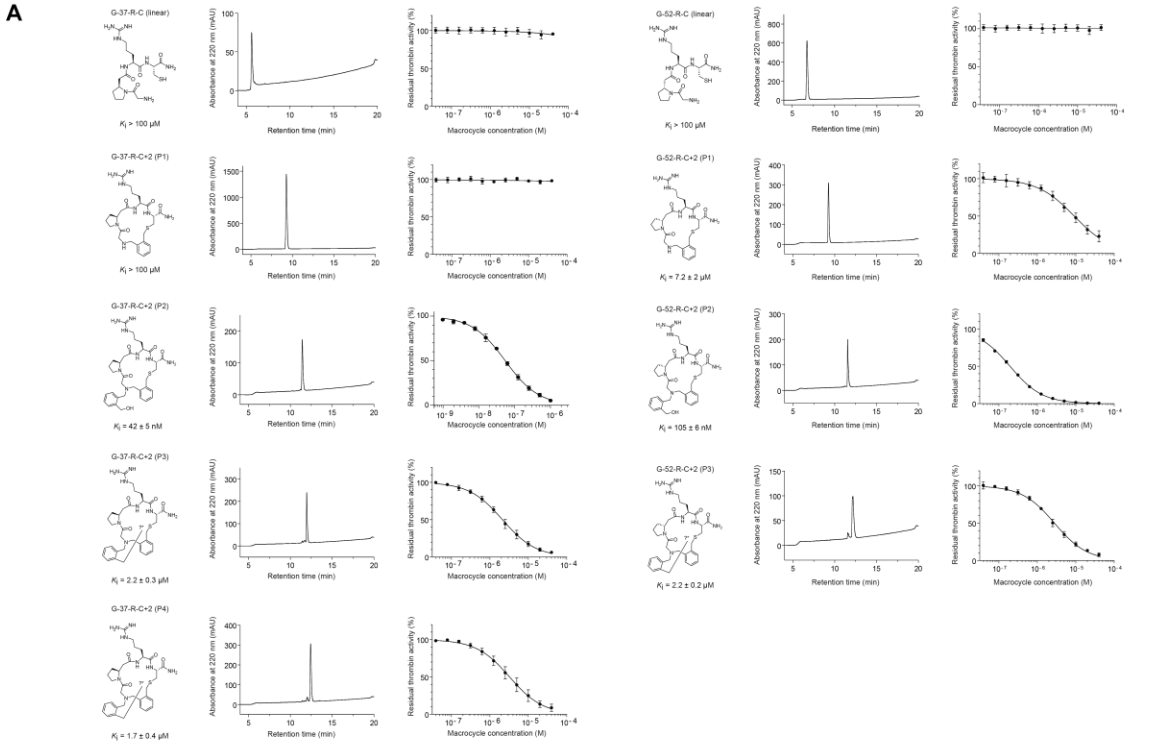


Fig. S5. KLK5 screen and HPLC analysis of thrombin hits. (**A**) KLK5 inhibitor screen. 8988 macrocycles of the 12 formats shown on the top were screened. The macrocycles are composed of an arginine residue for potential binding to S1 substrate binding pocket of KLK5, one of 107 amino acids with diverse side chains or structures (indicated with X; Table S1), a glycine or β-alanine (**20**) residue, a cysteine and one of the linkers **1–7**. The library was screened against KLK5 using a fluorogenic protease substrate and macrocycle concentrations of 13 μM (or slightly lower depending on the macrocyclization yields). The extent of KLK5 inhibition is indicated with color (linear black-red-white gradient for 100% to 0% inhibition). Reaction products that were characterized further are indicated with arrows and are labeled with names (the number and letters indicate amino acids and the last number the linker). (**B**–**C**) Analysis of thrombin hits. Reversed-phase HPLC and MS analysis of peptides G-37-R-C and G-52-R-C reacted with linker **2**. The yields of P1-P4 are indicated as % of total peptide products. The chemical structures of reagent **2** hydrolyzed once or twice are indicated. (**B**) The two peptides

carrying a GGW appendix were incubated with reagent **2** using exactly the same conditions as in the screen (concentrations, solvents, volumes, pipetting order) and the products analyzed by LC-MS. (**C**) The reaction of the peptide G-52-R-C without GGW appendix reacted with reagent **2** was analyzed by analytical RP-HPLC. The products of the equivalent experiment with G-37-R-C is shown in Fig. 3B.


A
G-37-R-C (linear)
K_i > 100 μM
G-37-R-C+2 (P1)
K_i > 100 μM
G-37-R-C+2 (P2)
K_i = 42 ± 5 nM
G-37-R-C+2 (P3)
K_i = 2.2 ± 0.3 μM
G-37-R-C+2 (P4)
K_i = 1.7 ± 0.4 μM
G-52-R-C (linear)
K_i > 100 μM
G-52-R-C+2 (P1)
K_i = 7.2 ± 2 μM
G-52-R-C+2 (P2)
K_i = 105 ± 6 nM
G-52-R-C+2 (P3)
K_i = 2.2 ± 0.2 μM
Absorbance at 220 nm (mAU)
Retention time (min)
Residual thrombin activity (%)
Macrocycle concentration (M)

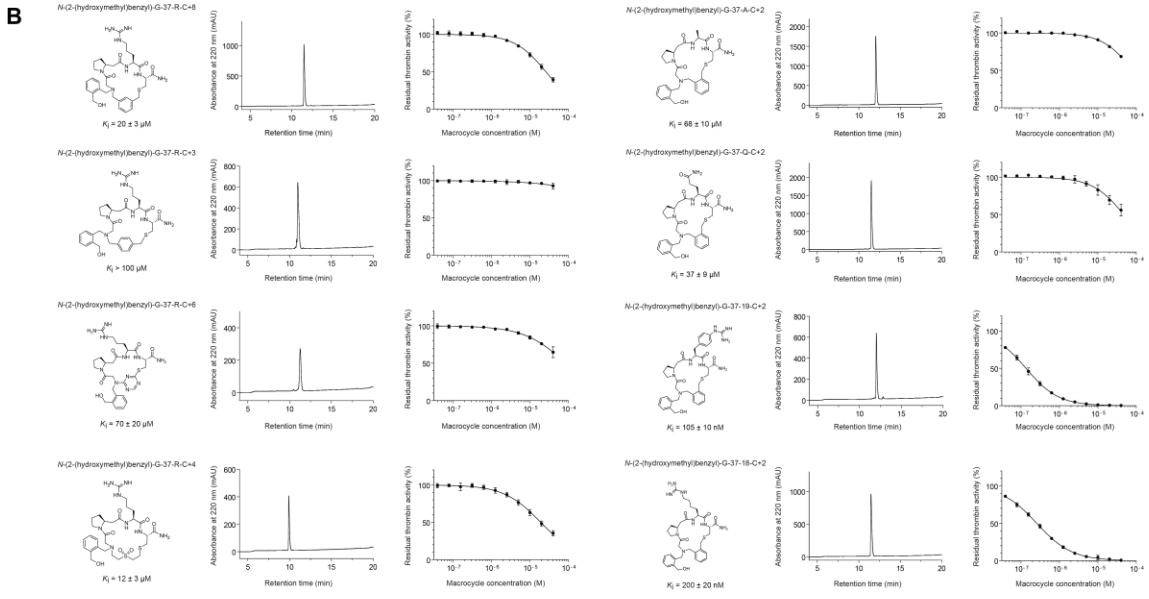

B
N-(2-(hydroxymethyl)benzyl)-G-37-R-C+8
K_i = 20 ± 3 μM
N-(2-(hydroxymethyl)benzyl)-G-37-R-C+3
K_i > 100 μM
N-(2-(hydroxymethyl)benzyl)-G-37-R-C+6
K_i = 70 ± 20 μM
N-(2-(hydroxymethyl)benzyl)-G-37-R-C+4
K_i = 12 ± 3 μM
N-(2-(hydroxymethyl)benzyl)-G-37-A-C+2
K_i = 68 ± 10 μM
N-(2-(hydroxymethyl)benzyl)-G-37-Q-C+2
K_i = 37 ± 9 μM
N-(2-(hydroxymethyl)benzyl)-G-37-19-C+2
K_i = 105 ± 10 nM
N-(2-(hydroxymethyl)benzyl)-G-37-18-C+2
K_i = 200 ± 20 nM
Absorbance at 220 nm (mAU)
Retention time (min)
Residual thrombin activity (%)
Macrocycle concentration (M)

Fig. S6. Structures, HPLC analysis, and activities of macrocycles. (**A**) Structures and activities of peaks 1-4 (P1-P4) obtained by reacting peptides G-37-R-C and G-52-R-C with reagent **2**. The masses of P3 and P4 correspond to macrocycles that contain an additional linker **2** connecting two nucleophilic groups. While the first nucleophilic group is most likely the secondary amine of glycine, the nature of the second nucleophilic group is not known. K_is are mean values of three measurements and SD are indicated. Analytic HPLC chromatograms and thrombin inhibition dose-response curves are shown. (**B**) Structure-activity relationship (SAR) analysis of P2. Structures and activities of P2 variants shown in Fig. 3C. K_is are mean values of three measurements. The SD is indicated. Analytic HPLC chromatograms and thrombin inhibition dose-response curves are shown. (**C**) Structures and activities of KLK5 hits indicated on fig. S5A. The macrocycles were synthesized with and without the GGW appendix and characterized. K_i values are means of three measurements. SDs are indicated. Analytic HPLC chromatograms and KLK5 inhibition dose-response curves are shown.

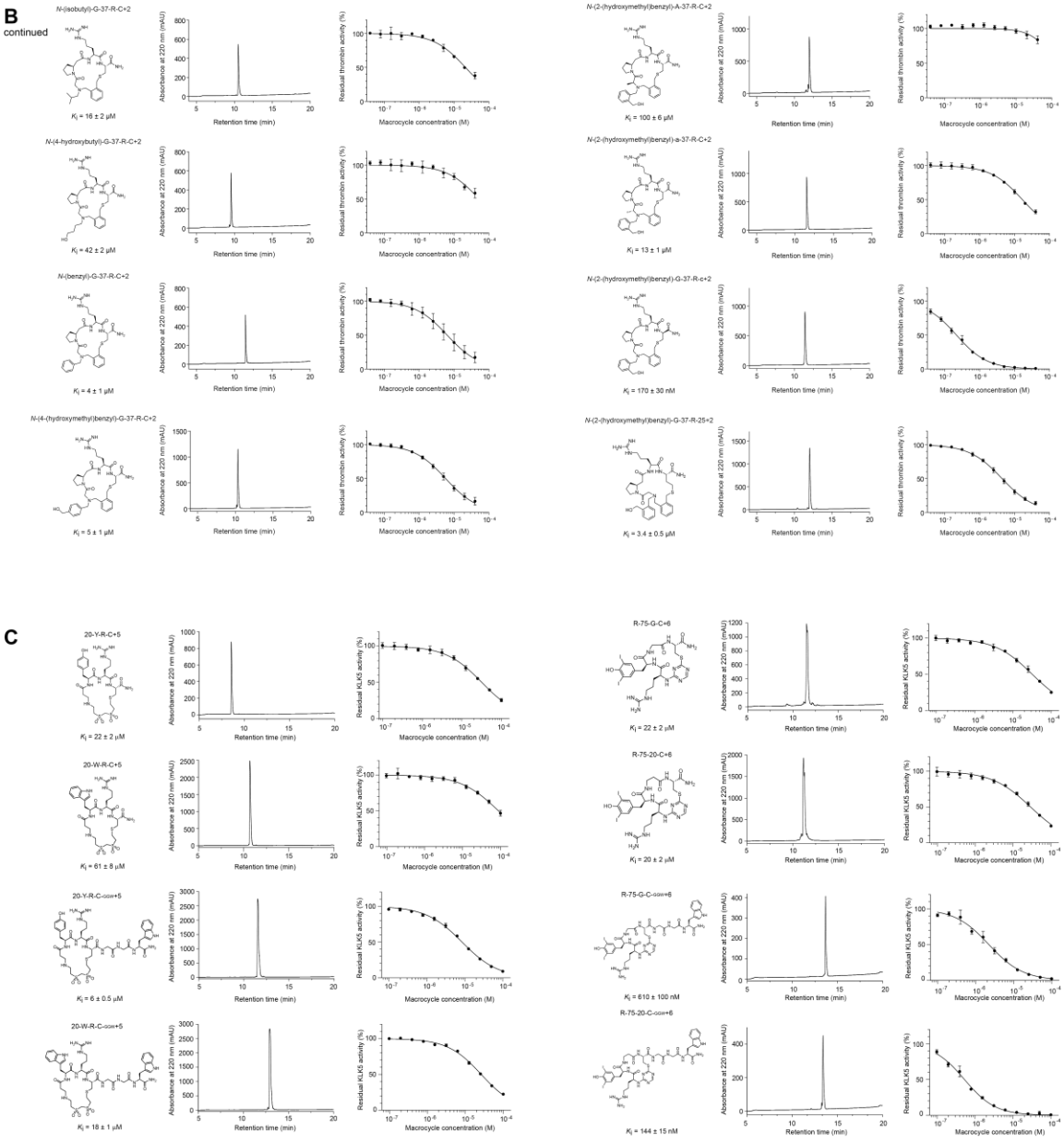


Fig. S6. Continued

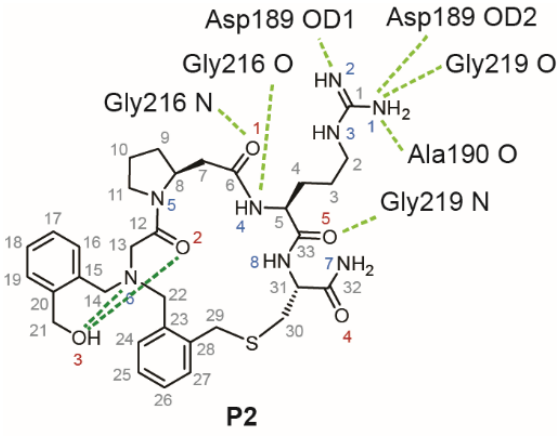
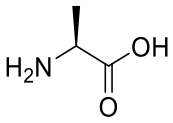
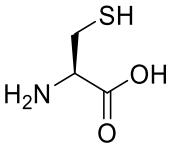
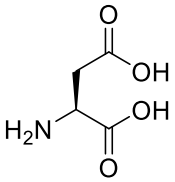
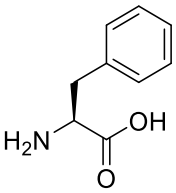
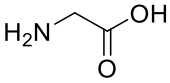
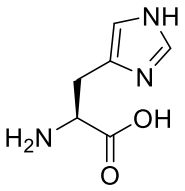
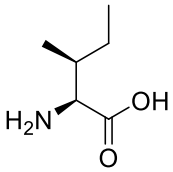
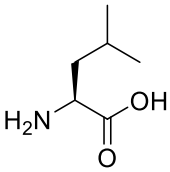
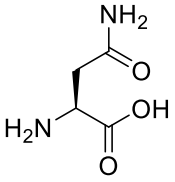
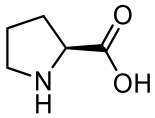
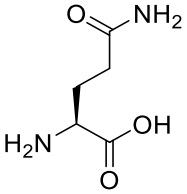
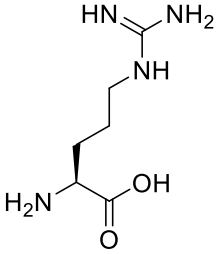
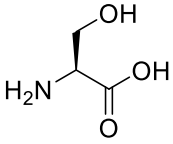
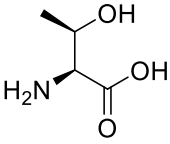
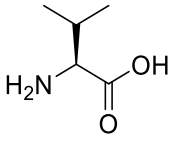
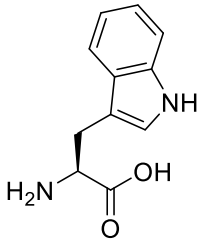
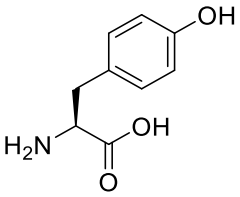
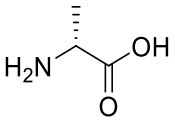
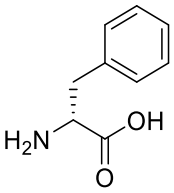
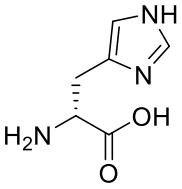
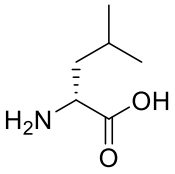
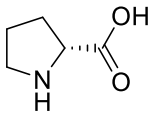
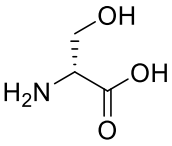
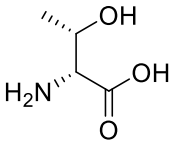
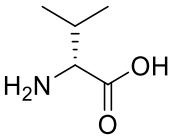
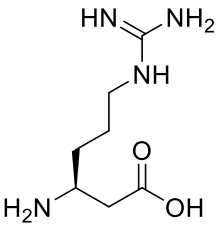
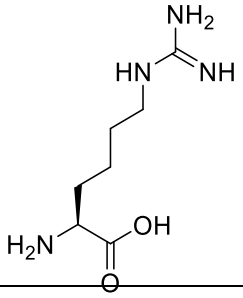
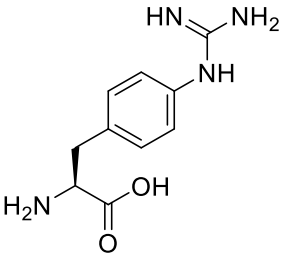
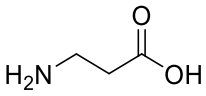
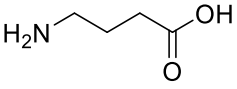
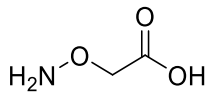
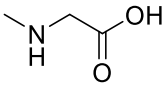
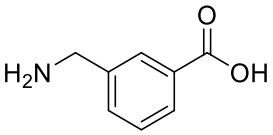
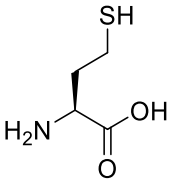
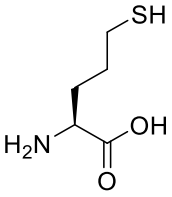
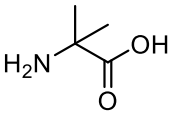
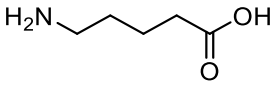
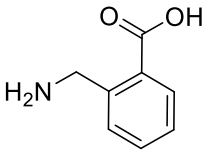
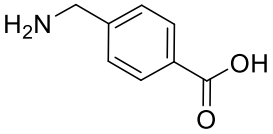
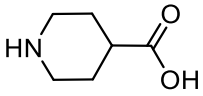
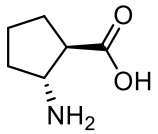
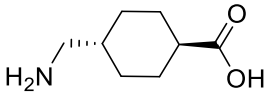
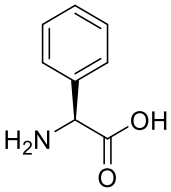
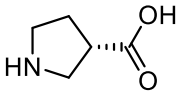
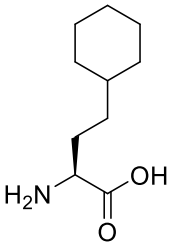


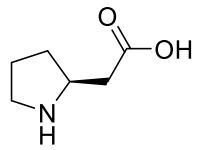
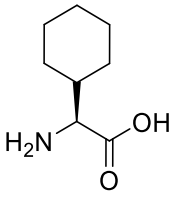
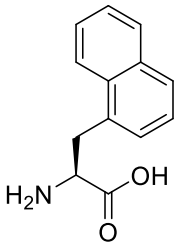
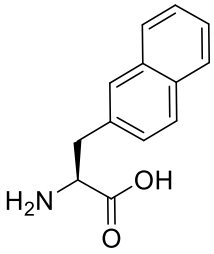
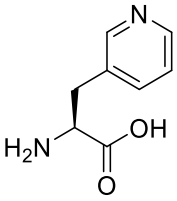
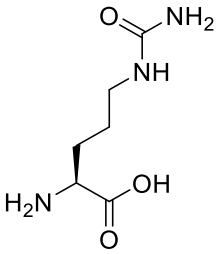
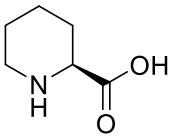
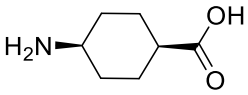
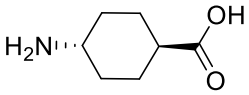
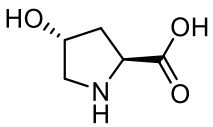
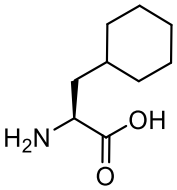
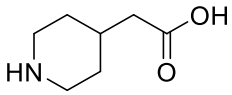
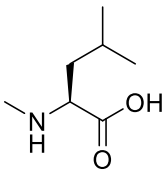
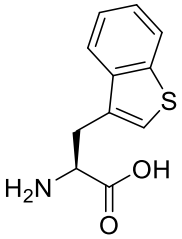
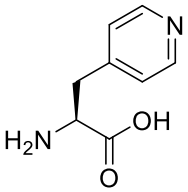
Fig. S7. Chemical structure of P2. Atom numbers used in the PDB file are indicated. Intra-molecular hydrogen bonds are shown as green dashed lines. Inter-molecular hydrogen bonds are shown as dashed light green lines. The macrocycle exhibits two slightly different conformations in the two-thrombin molecules present within the asymmetric unit. In one conformation, the main chain oxygen O1 of P2 points toward the main chain nitrogen of Gly216 (Gly216 N) whereas in the second conformation, the nitrogen N4 of P2 is oriented toward the main chain oxygen of Gly216 (Gly216 O).

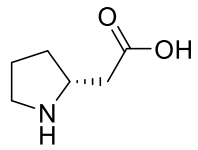
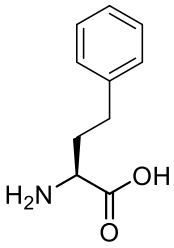
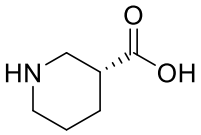
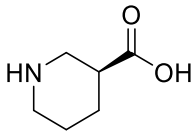
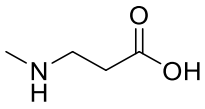
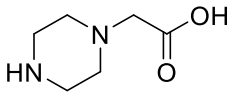
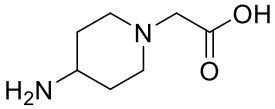
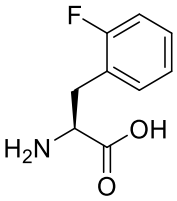
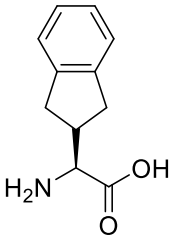
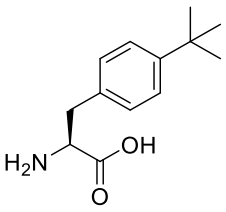
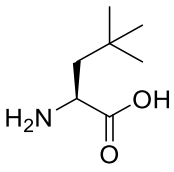
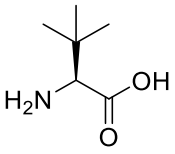
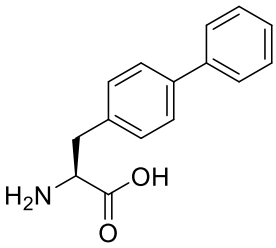
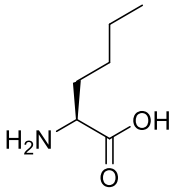
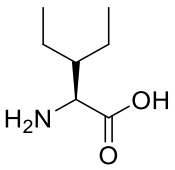
Supplementary Tables

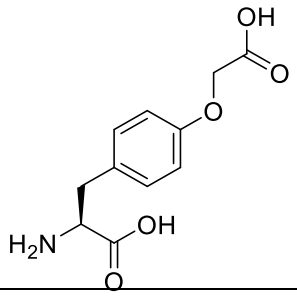
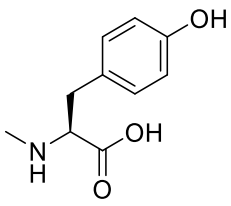
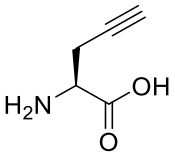
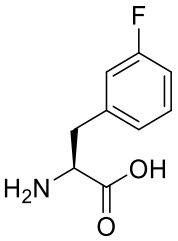
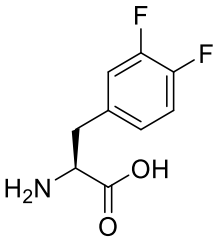
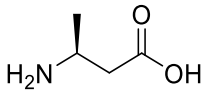
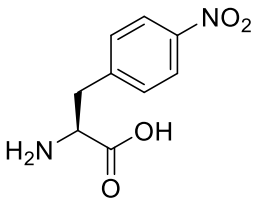
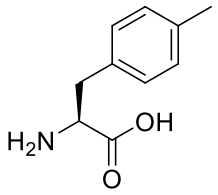
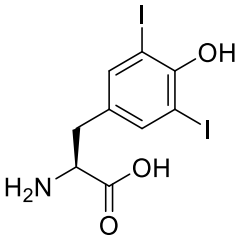
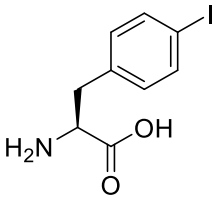
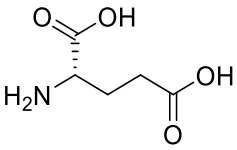
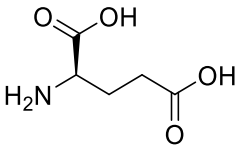
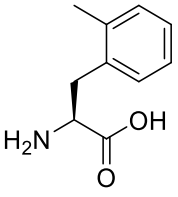
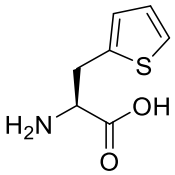
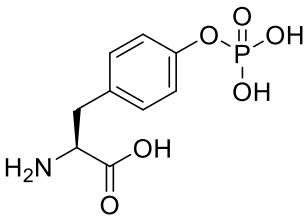
A		**C**		**D**	
F		**G**		**H**	
I		**L**		**N**	
P		**Q**		**R**	
S		**T**		**V**	

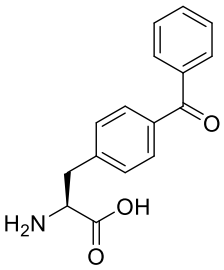
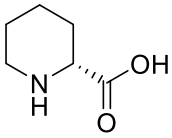
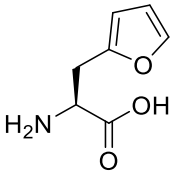
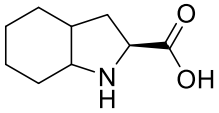
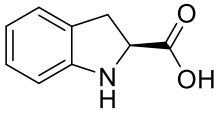
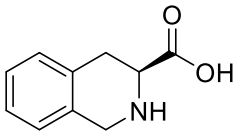
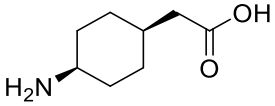
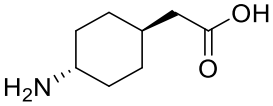
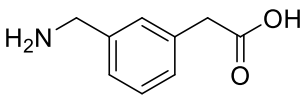
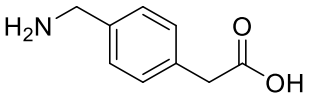
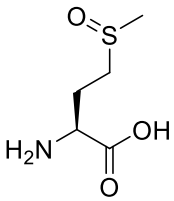
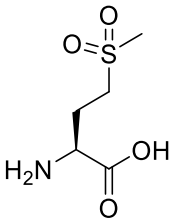
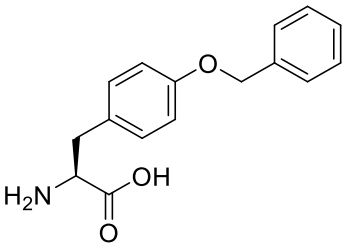
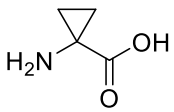
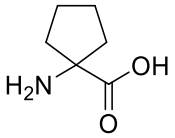
W		**Y**		**a**	
f		**h**		**l**	
p		**s**		**t**	
v		**17**		**18**	
19		**20**		**21**	

22		**23**		**24**	
25		**26**		**27**	
28		**29**		**30**	
31		**32**		**33**	
34		**35**		**36**	

37		**38**		**39**	
40		**41**		**42**	
43		**44**		**45**	
46		**47**		**48**	
49		**50**		**51**	

52		**53**		**54**	
55		**56**		**57**	
58		**59**		**60**	
61		**62**		**63**	
64		**65**		**66**	

67		**68**		**69**	
70		**71**		**72**	
73		**74**		**75**	
76		**77**		**78**	
79		**80**		**81**	

82		**83**		**84**	
85		**86**		**87**	
88		**89**		**90**	
91		**92**		**93**	
94		**95**		**96**	

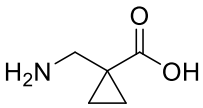
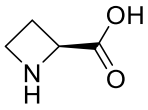
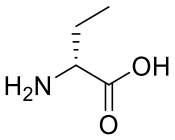
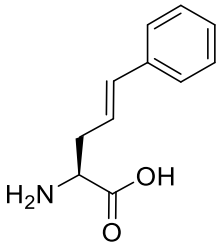
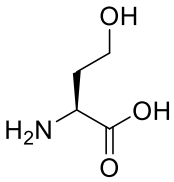
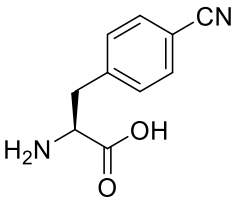
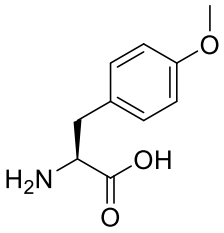
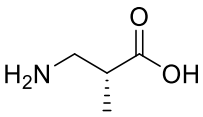
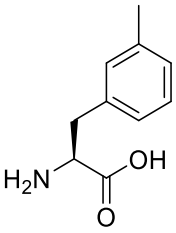
97		**98**		**99**	
100		**101**		**102**	
103		**104**		**105**	

Table S1. Chemical structures and codes of amino acids. The codes correspond to the numbers in position X of macrocycles in the large-scale library.

Data collection *	
Wavelength (Å)	1.0725
Space group	$P2_12_12_1$
Cell parameters	
a, *b*, *c* (Å); α, β, γ (°)	55.94, 80.91, 159.43; 90, 90, 90
Resolution (Å)	79.71 – 2.30 (2.38 – 2.30)
Independent reflections	32984 (3269)
Multiplicity	5.7 (5.9)
R_{merge}	0.10 (0.65)
R_{pim}	0.068 (0.44)
$<I / \sigma(I)>$	9.0 (2.0)
Completeness (%)	99.90 (99.91)

Refinement	
No. reflections (Used for R_{free} calculation)	32984 (3269)
R_{work} / R_{free}	0.20 / 0.236
Number non-hydrogen atoms	4975
protein (chains A, B, H, L)	4636
ligand (YW6)	141
solvent	150
others (EDO, NAG, Na)	48

Geometry	
RMSD values	
bond lengths (Å)	0.007
bond angles (°)	1.26
Ramachandran plot (%)	

most favored	96.89
additionally allowed	3.11
outliers	0.0
Rotamers outliers (%)	1.58
Average B-factor	47.85

Table S2. Statistics on X-ray structure data collection and refinement. 3300 frames were measured in 0.1° oscillation steps. (*). A single crystal was used to collect all diffraction data. Highest-resolution shell statistics are shown within brackets. Atoms with multiple conformations are counted as multiple and distinct scatterers.

A

P2 atom 1	P2 atom 2	Distance (Å)	Interaction
O3	N6	3.0	HB
O3	O2	3.1	PI

B

P2 atom	thrombin atom / residue	Distance (Å)	Interaction
O1	N / Gly216	3.04	HB*
O5	N / Gly219	3.30	HB
N1	O / Ala190	3.10	PI
N1	O / Gly219	3.02	HB
N1	OD2 / Asp189	2.72	SB
N2	OD1 / Asp189	3.40	PI
N2	H_2O 10	2.90	HB
N4	O / Gly216	3.05	HB#

C

P2 atom	thrombin atom / residue	Distance (Å)
C9	CG / His57	3.84
C9	CD2 / His57	3.43
C9	NE2 / His57	3.80
C10	CE2 / Tyr60A	3.50
C10	CZ / Tyr60A	3.74
C10	CZ2 / Trp60D	3.88
C11	CZ2 / Trp60D	3.64
C10	CH2 / Trp60D	3.77
C11	CH2 / Trp60D	3.70
C17	O / Trp96	3.73
C18	O / Trp96	3.51
C18	O / Arg97	3.63

C17	C / Glu97A	3.71
C18	C / Glu97A	3.72
C17	O / Glu97A	3.55
C18	O / Glu97A	3.45
C17	N / Asn98	3.56
C17	CA / Asn98	3.46
C17	C / Asn98	3.67
C17	N / Leu99	3.67
C16	CG / Leu99	3.90
C8	CD1 / Leu99	3.64
C25	OG1 / Thr172	3.75
N1	CG / Asp189	3.37
C1	OD1 / Asp189	3.72
N1	OD1 / Asp189	3.38
N2	OD1 / Asp189	3.39
C1	OD2 / Asp189	3.69
N1	OD2 / Asp189	2.72
C1	C / Ala190	3.89
C1	O / Ala190	3.33
C2	O / Ala190	3.85
N1	O / Ala190	3.15
N2	O / Ala190	3.81
N3	O / Ala190	3.68
N2	CB / Ala190	3.90
C2	CA / Cys191	3.85
C7	O / Ser214	3.66
O1	CA / Trp215	3.33
O1	C / Trp215	3.70
N2	O / Trp215	3.69
O1	CB / Trp215	3.36

C24	CE3 / Trp215	3.87
C24	CZ3 / Trp215	3.85
N3	N / Gly216	3.85
O1	N / Gly216	3.04
C1	CA / Gly216	3.85
N1	CA / Gly216	3.80
C4	C / Gly216	3.89
C22	O / Gly216	3.42
C23	O / Gly216	3.88
C33	O / Gly216	3.72
C4	O / Gly216	3.01
C5	O / Gly216	3.76
O1	O / Gly216	3.15
O5	O / Gly216	3.47
C25	CB / Glu217	3.60
C26	CB / Glu217	3.72
C25	CG / Glu217	3.79
C26	CG / Glu217	3.36
C27	CG / Glu217	3.63
C26	CD / Glu217	3.38
C25	OE1 / Glu217	3.50
C26	OE1 / Glu217	3.31
O5	N / Gly219	3.30
O5	CA / Gly219	3.63
C1	O / Gly219	3.89
C2	O / Gly219	3.27
C4	O / Gly219	3.74
N1	O / Gly219	3.02
C2	SG / Cys220	3.68
N2	CA / Gly226	3.65

D

	Thrombin	P2	Complex
Volume ($Å^3$)	41442	831	42337
Surface area ($Å^2$)	11271	598	11070

Table S3. Interactions between P2 and thrombin. (**A**) Atoms forming intra-molecular interactions in the macrocycle P2. Optimal inter-molecular hydrogen bond (HB) and polar interaction (PI) were identified and measured with the program PYMOL. (**B**) Atoms of the macrocycle P2 forming inter-molecular interactions with atoms and residues of human α-thrombin (chymotrypsin numbering). Optimal inter-molecular hydrogen bonds (HB), salt bridges (SB) and polar interactions (PI) were defined using the web server PROFUNC. Note: H-bonds present exclusively in one of the two conformations explored by P2 are indicated as * and #, respectively. **(C)** Atoms of the macrocycle P2 forming hydrophobic interactions with atoms and residues of human α-thrombin (chymotrypsin numbering). Interactions have distances shorter than 4.0 Å and were defined using the software LIGPLOT+ by the web server PROFUNC. (**D**) The solvent excluded volume and the corresponding buried surface were calculated using the 3V web server and a spherical probe of 1.5 Å radius. The buried surface area between human α-thrombin and P2 is 201 $Å^2$.

Captions for Supplementary Data

Data S1. Raw data of peptide macrocyclization reactions.

Raw data of peptide macrocyclization reactions**.** The numbers in the tables indicate the quantities of the macrocyclic product and side products that were formed with the indicated cyclization reagents. The reactions were analyzed by LC-MS monitoring the absorption at 220 nm and measuring the total ion count (TIC) in positive mode. The quantity of macrocycle and side products were determined based on the area under the peaks and is indicated as percentage of total peptide products. Peptides with a Gly-Gly-Trp (R) or Gly-Trp (R_1) appendix were used as substrates. For each cyclization reagent, the RP-HPLC and MS chromatograms are shown for the peptide Gly-Cys-R as an example below the tables. The peak of the desired cyclic peptide product is colored in blue. The peaks of frequently found side products are shown in different colors as illustrated with the example of the peptide Gly-Cys-R and the reagent **1** in fig. S2C. The peaks of less often observed side products are shown in green. They are a) dehydroalanine derivative, b) disulfide-linked peptide, c) modification with two linkers, d) modification with two linkers of which one has reacted with NH_3, e) cyclic sulfonium. The peaks of side products that were not identified but are suspected to contain a peptide moiety (based on the mass) are colored in grey. Abbreviations of amino acids: hCys = homocysteine, Mnv = 5-mercapto norvaline, MeCys = N-methyl cysteine.

Data S2. Poster showing all 432 different macrocycle scaffolds in the pilot-scale library.

Poster showing the skeletal diversity of macrocycles in the pilot-scale library. The chemical structures of all different backbones are shown and the names of macrocycles containing the backbones are indicated below. The letters and numbers in the names indicate the amino acids and cyclization linkers.