Supporting Information

Directed evolution driven increase of structural plasticity is a prerequisite for binding of complement lectin pathway blocking MASP-inhibitor peptides

Zsolt Dürvanger[a], Eszter Boros[b], Zoltán Attila Nagy[b], Rózsa Hegedüs[c], Márton Megyeri[d], József Dobó[d], Péter Gál[d], Gitta Schlosser[e] Annamária F. Ángyán[f], Zoltán Gáspári[f], András Perczel[a,g], Veronika Harmat[a,g], Gábor Mező[c,h], Dóra K. Menyhárd[a,g,*], Gábor Pál[b,*]

[a] Laboratory of Structural Chemistry and Biology, Institute of Chemistry, ELTE Eötvös Loránd University, Pázmány Péter sétány 1/A, H-1117, Budapest, Hungary

[b] Department of Biochemistry, ELTE Eötvös Loránd University, Pázmány Péter sétány 1/C, H-1117 Budapest, Hungary

[c] MTA-ELTE Research Group of Peptide Chemistry, Pázmány Péter sétány 1/A, Budapest, H-1117, Hungary

[d] Institute of Enzymology, Research Centre for Natural Sciences, Magyar tudósok krt 2, H-1117, Budapest, Hungary

[e] Department of Analytical Chemistry, MTA-ELTE Lendület Ion Mobility Mass Spectrometry Research Group, Institute of Chemistry, ELTE Eötvös Loránd University, Pázmány Péter sétány 1/A, Budapest, H-1117, Budapest, Hungary

[f] Faculty of Information Technology and Bionics, Pázmány Péter Catholic University, Práter u. 50/A, H-1083 Budapest, Hungary

[g] MTA-ELTE Protein Modelling Research Group, Eötvös Loránd Research Network, Pázmány Péter sétány 1/A, H-1117, Budapest, Hungary

[h] Department of Organic Chemistry, ELTE Eötvös Loránd University, Pázmány Péter sétány 1/A, Budapest, H-1117, Hungary.

* Corresponding authors: Dóra K. Menyhárd dora.k.menyhard@ttk.elte.hu, and Gábor Pál, gabor.pal@ttk.elte.hu

Index

Materials and Methods

Crystallographic data collection and refinement statistics
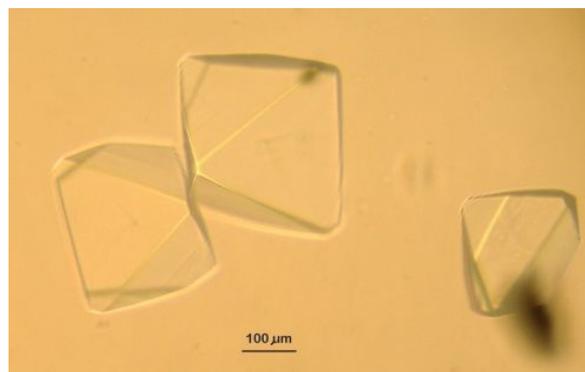
Supplementary Figure 1. Crystals of the MASP-1 / SFMI1 complex.

Supplementary Table 1. X-ray diffraction data collection and refinement data. Data for the highest resolution shell are given in parentheses

Data collection	
Unit cell dimensions a, b, c (Å), α, β, γ (°)	69.3 69.3 161.8 90.0 90.0 120.0
Space group	P$3_1$21
Resolution range (Å)	19.85-2.40 (2.50-2.40)
No. of unique refl. / observed refl.	17832 / 139615
$< I / \sigma(I) >$	14.74 (1.01)
R_{meas}	0.106 (2.080)
Resolution at $< I / \sigma(I) > = 2.0$ (Å)	2.70
Completeness (%)	97.1 (76.7)
$CC_{1/2}$	99.9 (31.4)
Refinement	
Resolution range (Å)	19.85-2.40 (2.50-2.40)
R / R_{free} (No. of obs.)	0.2560 (16446) / 0.2849 (855)
No. of atoms: protein / inhibitor / solvent	2867 / 77 / 33
B-factor of protein, inhibitor, solvent (Å^2)	76.29 / 92.91 / 63.57
Wilson B-factor (Å^2)	70.29
RMS dev. bond length (Å)	0.002
RMS dev. bond angles (°)	0.452
Ramachandran fav. / all. / disall.	386 / 14 / 0

Supplementary results

Hydrogen bonds in the crystal structure of the MASP-1 / SFMI1 complex

Supplementary Table 2. Donor-acceptor pairs for H-bonding in the crystal structure of the MASP-1 / SFMI1 complex. Default geometric criteria used by PyMOL were employed (a cutoff distance of 3.2 Å for ideal geometry and 3.6 Å for minimally accepted geometry)

Intermolecular hydrogen bonds			Intramolecular hydrogen bonds (SFMI1)		
SFMI1	MASP-1	Distance / Å	Donor	Acceptor	Distance / Å
Gly1-N	Asp671-OD1	2.6	Ser6-OG	Pro8=O	3.1
Gly1=O	Asp671-N	2.8	Ile10=O	Ser4-N	3.0
Cys3-N	Gly669=O	3.1			
Cys3=O	Gly669-N	3.5			
Arg5-N	Ser667=O	3.1			
Arg5=O	Gly644-N	2.8			
Arg5=O	Asp645-N	3.0			
Arg5=O	Ser646-N	2.8			
Arg5-NH2	Asp671=O	3.3			
Arg5-NH2	Asp640-OD2	3.1			
Arg5-NH1	Asp640-OD1	3.2			
Arg5-NH1	Ala641=O	2.6			
Leu7-N	Phe474=O	3.0			

Characterization of SFMI2 and derived peptides by HPLC-MS

For the disulfide bond containing peptides, a single product could be isolated. However, for some of the thioester bound containing derivatives, two separate isomers were found: only one product could be isolated by HPLC in the case of Lys-, Orn-, and Agl-containing peptides, while two peaks were detected in the case of Dab- and Dap-containing ones. However, the two isomers were stable only in the case of the Dap containing peptide, while

the isomers of Dab could not be separated. We assume that the explanation for a single conformer of Agl-containing cyclic peptide is the proximity of two amide planes that prevent its rotation. This was reflected by the molecular dynamics simulations and subsequent energy calculations, where in the case of Dap we found a 12.9 kJ/mol difference between conformers with the less favorable *cis*- and the preferred *trans*-amide configuration. In the case of Dab this energy difference was only 2.1 kJ/mol, while in the case of Agl, it was 9.3 kJ/mol. The energy level of the transition state for the *trans* ↔ *cis* isomerization relative to the energy level of the *trans* conformer i.e. the activation energy was estimated using a dihedral drive (along the amide bond in 10° steps using the OPLS3 forcefield) resulting in 46.9 kJ/mol for Dap, 56.1 kJ/mol for Dab and 79.2 kJ/mol for Agl. Thus if we assume that synthesis results in the formation of the more stable *trans* isomer in the case of Dap and Agl, we should expect an accumulation of the *cis* form only in the case of Dap (due to its much more favorable activation energy), while all solutions of Dab are expected to (eventually) contain both *cis* and *trans* forms (due to their very close conformational energies).

H-GY[CSRSYPPVC]IPD-OH

M(mo): 1553.6616
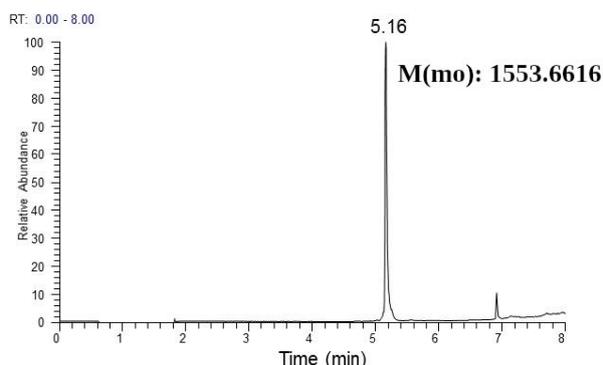
Ac-GY[CSRSYPPVC]IPD-NH$_2$

M(mo): 1594.6903
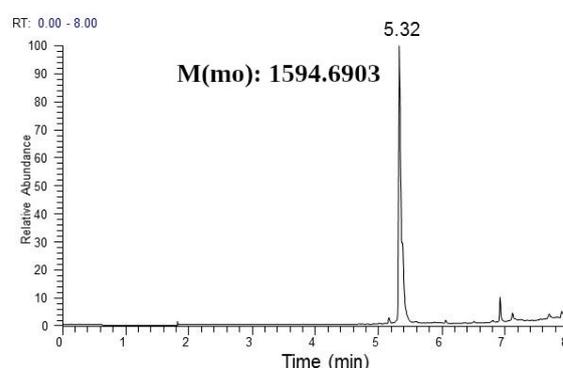
Ac-GY[Lys(-CH$_2$CO)SRSYPPVC]IPD-NH$_2$

M(mo): 1661.7862
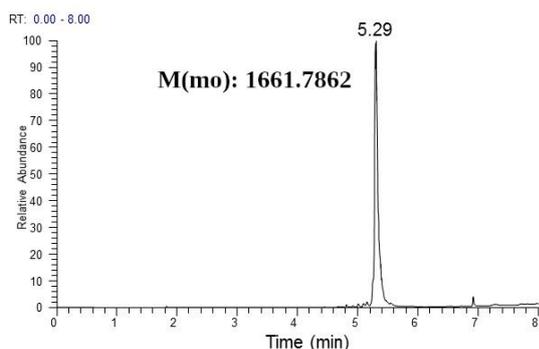
Ac-GY[Orn(-CH$_2$CO)SRSYPPVC]IPD-NH$_2$

M(mo): 1647.7707
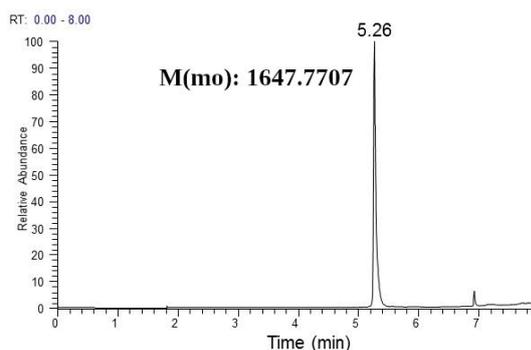
Ac-GY[Dab(-CH$_2$CO)SRSYPPVC]IPD-NH$_2$
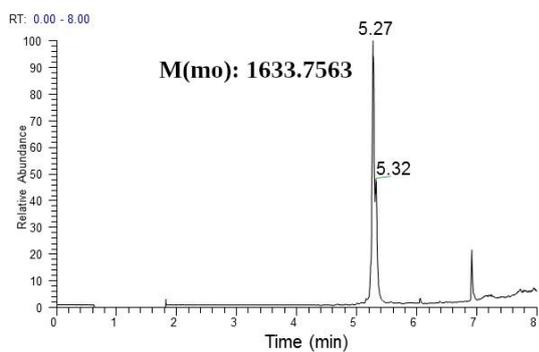
Ac-GY[Agl(-CH$_2$CO)SRSYPPVC]IPD-NH$_2$
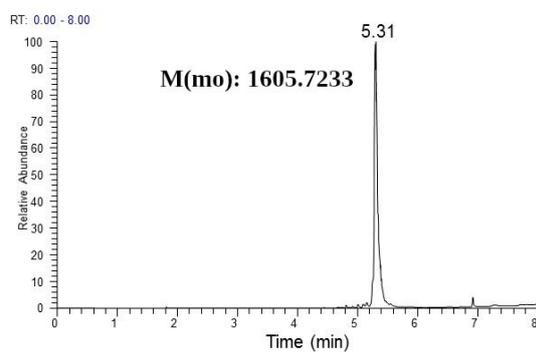
Ac-GY[Dap(-CH$_2$CO)SRSYPPVC]IPD-NH$_2$/1
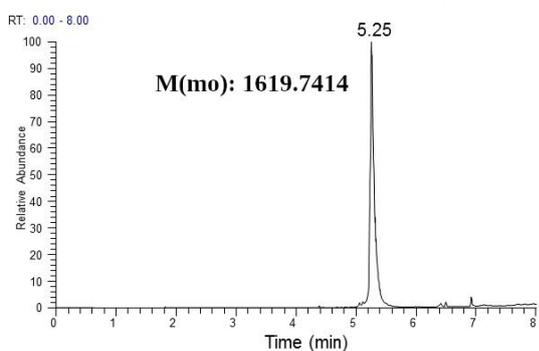
Ac-GY[Dap(-CH$_2$CO)SRSYPPVC]IPD-NH$_2$/2
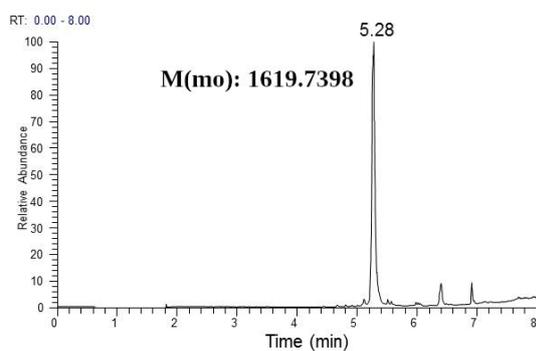
Supplementary Figure 2. HPLC-MS chromatograms of SFMI2, SFMI2cap, SFMI2cap-Lys, SFMI2cap-Orn, SFMI2cap-Dab, SFMI2cap-Agl, and two isomers of SFMI2cap-Dap, respectively.
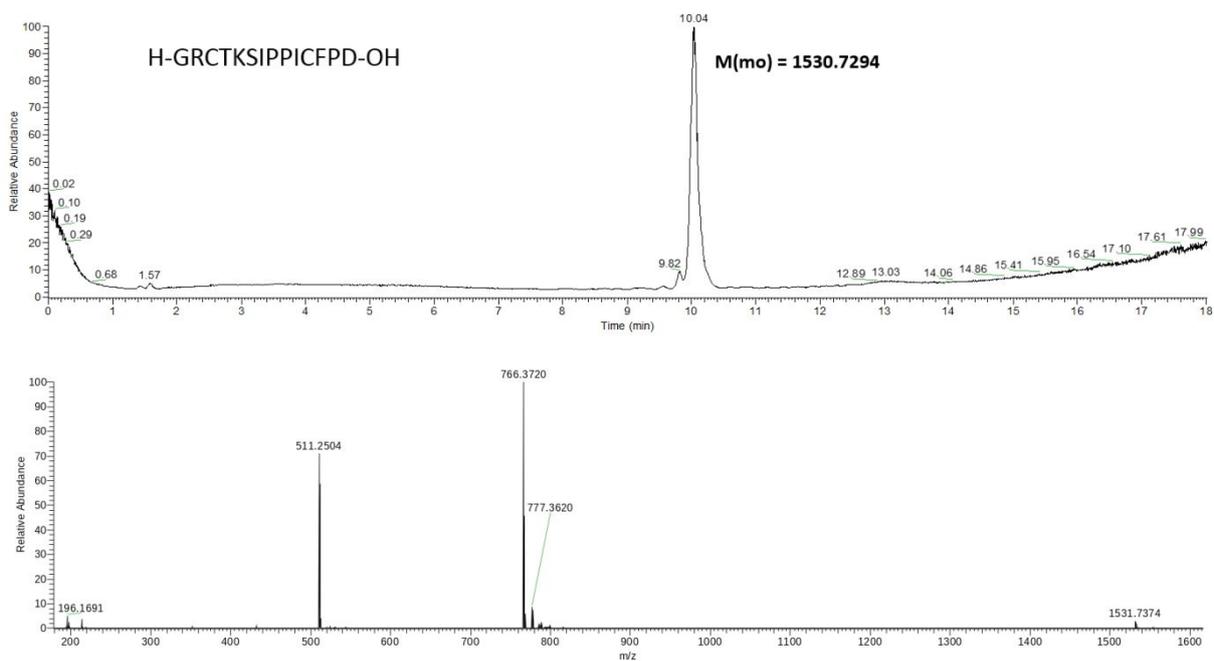
Supplementary Figure 3. HPLC-MS chromatogram and MS spectrum of the main peak of the acyclic SFTI variant.

Validation of the MD simulation protocol using the NMR structure of acyclic SFTI
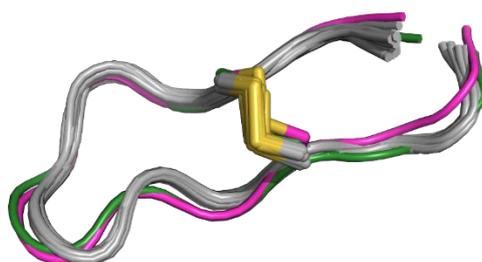
Supplementary Figure 4. Mid structures of backbone clusters of the acyclic SFTI variant obtained by MD simulation started from the NMR structure (dark green) and from the common starting structure of the other inhibitors (purple) fitted to the ensemble derived from NMR measurements (PDB ID:1JBN, gray) (1)

Testing the roles of hydrophobic core forming Phe12, Ile10 and Thr4 residues in maintaining the β-hairpin structure of SFTI-derived inhibitors
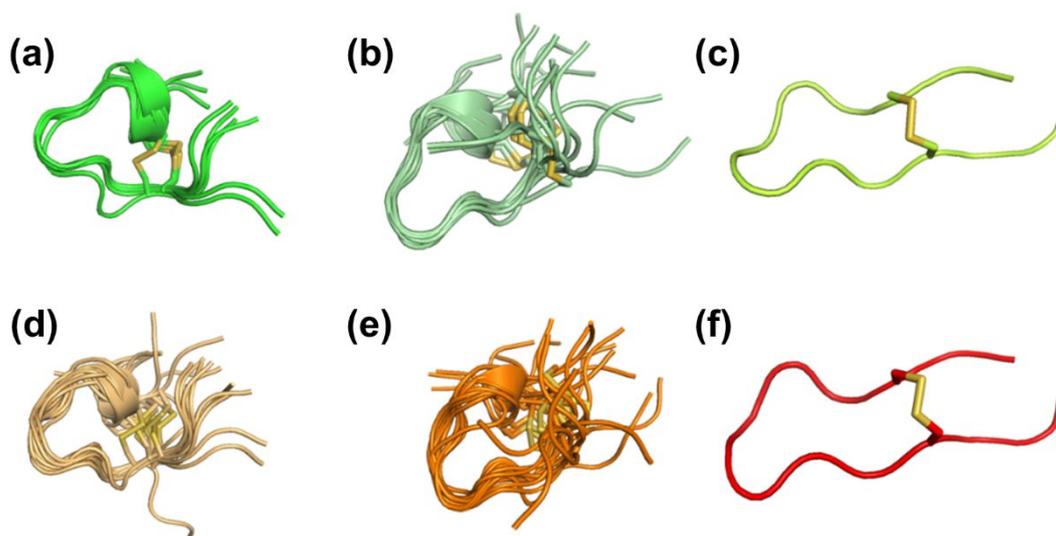
Supplementary Figure 5. Main conformers of the last 600ns of the trajectories obtained by simulating the mutated variants of the SFMI1 and SFMI2 peptides. The shown conformers are mid-structures of the backbone clusters obtained by using 1.5Å cutoff and represent at least 80% of all snapshots. **a-c,** Representative structures of SFMI1 mutants Ile12Phe, Ser4Thr and double mutant Ile12Phe/Ser4Thr, respectively. **d-f,** Representative structures of SFMI2 mutants Ile12Phe, Ser4Thr and the double mutant Ile12Phe/Ser4Thr, respectively

NMR study on SFMI2

NMR measurements confirmed that Pro8 (the P3' residue) of SFMI2 is in *cis* configuration in its solution state (as was also seen in the case of the parent SFTI molecule): the Tyr7 HA - Pro8 HA NOE wad clearly observed while the Tyr7 HA-Pro8 HD NOEs were not detected. Since Pro8 HA - Pro 9 HD NOEs were also present, the detected pattern is compatible with a cis Pro8 and trans Pro9 configuration. At the low pH (pH=3) used for the NMR measurements (necessary to suppress the amide NH exchange), both the C-terminal and the sidechain of Asp14 are (at least partially) protonated (estimated pKa of the Asp14 sidechain and the C-terminus is 4.3±0.8 and 3.4±.7, respectively), thus it was expected that the experiment might suggest a slightly different conformation from that seem in the simulations. We found that over 70% of the NOEs were reproduced by the simulations (Supplementary Table 3.). The measured NOEs are also in partial disagreement with the hairpin structure of the acyclic SFTI variant (measured also at acidic pH), supporting that SFMI2 assumes a different solution state structure from that of SFTI (Supplementary Table 3.).

Supplementary Table 3. Comparing the long distance NOEs of SFMI2 determined by NMR at pH 3 to the MD generated trajectory (2nd column) of the same system at physiological pH and the corresponding distances in the solution state structure (measured at pH 4.5) of the acyclic SFTI (PDB ID: 1JBN) (4th column) (1). In the case of the simulation, only the subset of snapshots fulfilling the given constraint (1- all, 0 – none) is shown, while in the case of the NMR structure the shortest distance that was found in the 20 structures was deposited in the database. Darkness of colors indicates the extent of resemblance to the experimentally derived parameters.

SFMI2		SFTI	
measured NOE	fraction of MD trajectory	corresponding contact in NMR structure of SFTI	distance (Å)
in region within the disulfide bridge:			
Cys3-HA Pro8-HB1,HB2	0.00	Cys3-HA Pro8- HB2	8.3
Cys3-HB1,HB2 Pro8-HB1,HB2	0.00	Cys3-HB3 Pro8- HB2	7.3
Cys3-HB1,HB2 Pro8-HG1,HG2	0.00	Cys3-HB3 Pro8- HG2	9.6
Cys3-HB1,HB2 Cys11-HA	0.99	Cys3-HB3 Cys11-HA	3.4
Cys3-HB1,HB2 Pro9-HA	0.14	Cys3-HB3 Pro9-HA	4.2
Ser4-HB1,HB2 Pro8-HA	0.30	Thr4-HB(CG2) Pro8-HA	8.8
Ser4-HB1,HB2 Pro8-HB1,HB2	0.30	Thr4-HB(CG2) Pro8- HB1,HB2	8.7
Ser4-HB1,HB2 Val10-HB	0.55	Thr4-HB(**CG2**) Ile10-HB	2.2
Ser4-HB1,HB2 Val10-HN	0.68	Thr4-HB(CG2) Ile10-HN	3.1
Ser4-HN Val10-HN	0.69	Thr4-HN Ile10-HN	2.5
Ser4-HN Cys11-HA	0.22	Thr4-HN Cys11-HA	3.5
Ser4-HN Pro8-HA	0.25	Thr4-HN Pro8-HA	7.8
Tyr7-HN Cys11-HA	0.00	Ile7-HN Cys11-HA	10.8
Pro8-HB1,HB2 Cys11-HA	0.00	Pro8-HB1,HB2 Cys11-HA	8.4
in the terminal regions:			
Tyr2-HN Cys11-HA	0.15	Arg2-HN Cys11-HA	4.6
Tyr2-HN Pro13-HD1,HD2	0.06	Arg2-HN Pro13-HD2	5.8
Tyr2-HA Ile12-HN	0.16	Arg2-HA Phe12-HN	4.5
Tyr2-HD1,HD2 Cys11-HB1,HB2	0.16	Arg2-HD2 Cys11-HB3	5.3
Cys3-HA Ile12-HN	0.35	Cys3-HA Phe12-HN	3.2
Cys3-HB1,HB2 Ile12-HN	0.69	Cys3-HB3 Phe12-HN	5.3
Cys3-HB1,HB2 Pro13-HB1,HB2	0.37	Cys3-HB3 Pro13-HB3	11.7

Ser4-HB1,HB2 Ile12-HG*	0.08	Thr4-HB(**CG2**) Phe12-(**CG**)	5.7
Arg5-HG1,HG2 Ile12-HG*	0.18	Lys5-HG2 Phe12-(**CG**)	10.3
Arg5-HD1,HD2 Ile12-HG*	0.19	Lys5-HD2 Phe12-(**CG**)	9.3

MD simulation of the free SFTI-derived peptides and MASP/peptide complexes
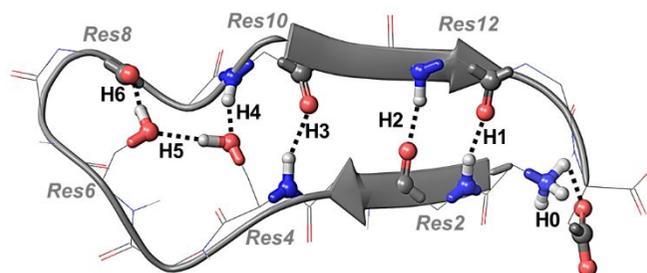

Supplementary Table 4. Formation of the canonical H-bonds during the MD simulations of the studies systems (0 indicates that the H-bond was present in none, 1 indicates that it was present in all of the snapshots of the equilibrium trajectory) - darker colors indicate increased presence. The inlet figure shows the numbering of the H-bonds, with the participating functional groups shown in ball-and-stick representation.

	H0 Nterm… Cterm		H1 res2-NH...res12-CO		H2 res2-CO...res12-NH		H3 res4-NH...res10-CO		H4 res4-OG…res6-OG		H5 res4-OG…res10-NH		H6 res6-OG…res8-CO	
	free	complex	free	complex	free	complex	free	complex	free	complex	free	complex	free	complex
MASP-1/SFMI1	0.14	0.00	0.06	0.01	0.00	0.94	0.00	1.00	0.02	0.33	0.01	0.53	0.03	0.55
MASP-1/SFMI2	0.12	0.00	0.00	0.06	0.00	0.97	0.00	1.00	0.00	0.73	0.00	0.94	0.00	0.30
MASP-2/SFMI1	0.14	0.60	0.06	0.54	0.00	0.98	0.00	1.00	0.02	0.64	0.01	0.90	0.03	0.32
MASP-2/SFMI2	0.12	0.48	0.00	0.38	0.00	0.99	0.00	0.99	0.00	0.33	0.00	0.57	0.00	0.39
MASP-2/SFMI2cap	0.04	0.05	0.00	0.41	0.02	0.99	0.01	1.00	0.00	0.78	0.26	0.96	0.02	0.11
MASP-2/SFMI2cap_Lys	0.01	0.06	0.03	0.10	0.05	0.42	0.90	0.99	0.04	0.54	0.07	0.70	0.01	0.44
MASP-2/SFMI2cap_Orn	0.01	0.00	0.00	0.00	0.00	0.41	0.00	0.95	0.00	0.06	0.10	0.11	0.01	0.75
MASP-2/SFMI2cap_Dab	0.00	0.01	0.00	0.04	0.00	0.12	0.00	0.78	0.01	0.47	0.48	0.94	0.01	0.47
MASP-2/SFMI2cap_cDap	0.00	0.03	0.57	0.00	0.70	0.93	0.80	1.00	0.06	0.66	0.37	0.96	0.01	0.21
MASP-2/SFMI2cap_tDap	0.02	0.02	0.14	0.10	0.41	0.65	0.59	0.94	0.05	0.80	0.01	0.94	0.01	0.18
MASP-2/SFMI2cap_Agl	0.01	0.01	0.03	0.02	0.00	0.03	0.00	0.46	0.00	0.04	0.00	0.18	0.01	0.41

Scaffold dependent selection of Thr or Ser in the P2 position of MASP inhibitors

Natural evolution conserved a P2 Thr both in the Pacifastin family of the SGMIs, and of the Bowman-Birk family of SFTI homologs. On both inhibitors, the hydroxyl group of P2 Thr stabilizes the canonical loop structure through intramolecular H-bonding, while its methyl group occupies a small hydrophobic S2 pocket, which, in the case of trypsin or chymotrypsin is already preformed in the free enzyme. In particular, in the trypsin-SFTI complex, P2 Thr is wedged between the active site His and the side chain of Leu99 (Supplementary Fig. 6A). In both MASPs, in place of that Leu, a Phe is present (Phe549 and Phe529 in MASP-1 and MASP-2, respectively), which would cause a steric clash with the P2 Thr methyl group.

In the uncomplexed form of MASP-1 Lys623 reaches into the substrate-binding groove. (Supplementary Fig. 6b) Docking of an inhibitor displaces this residue. In the case of the small SFTI-derived inhibitors Lys623 shifts to bind carbonyl oxygen of Asn547. This movement locks Phe549 into a where it would interfere with the methyl group of Thr at P2. Therefore, instead of the naturally conserved P2 Thr, phage display selected a Ser. Binding of the larger SGPI-2 derived inhibitor pushes loop 618-628 upward by which the Lys623 sidechain is removed from the enzyme interior and turns toward the solvent (becomes unresolved in the electron-density map), thus the Phe is provided with enough space to move and accommodate a Thr at P2.
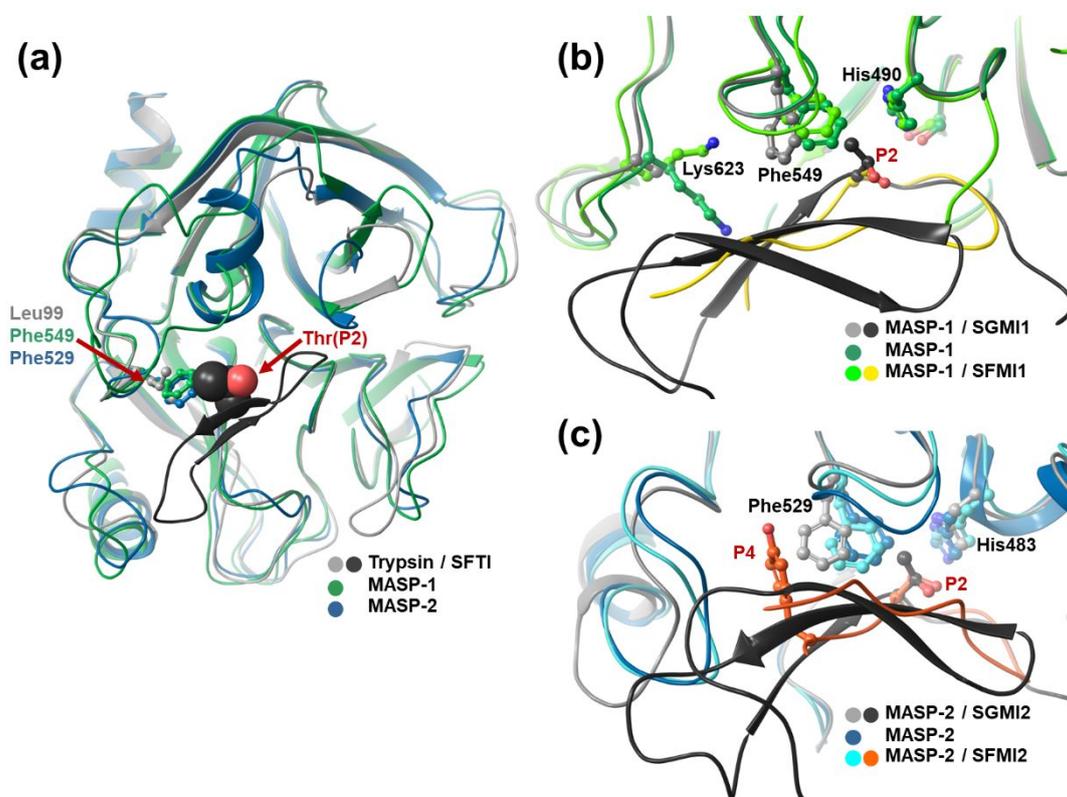
Supplementary Figure 6. a, Superimposed crystal structures of the trypsin/SFTI complex (light and darker grey, PDB ID: 1SFI) (2), MASP-1 (dark green; PDB ID: 3GOV) (3) and MASP-2 (dark blue; PDB ID: 1Q3X) (4). **b,** Superposition of the crystal structure of MASP-1/SGMI1 complex (in light and dark grey; PDB ID:

4DJZ) (5), the crystal structure of uncomplexed MASP-1 (darker green) and the MASP-1/SFMI1 complex (light green (MASP-1) and yellow). **c,** Superposition of the crystal structure of MASP-2/SGMI2 complex (in light and dark grey; PDB ID: 3TVJ) (5), the crystal structure of uncomplexed MASP-2 (green) and the modeled structure of the MASP-2/SFMI2 complex (light and dark blue).

Similarly, in case of MASP-2, binding of the larger SGMI inhibitor causes a shift of loop 602-615 allowing enough room for the aromatic ring of Phe529 to relocate and thus leave enough space for a Thr P2 residue. When the smaller inhibitor, SFMI, binds (that does not reshape loop 602-615) a gearwheel type of fit results – Phe529 becomes wedged between the P2 residue (Ser in this case) and Tyr2 (P4) of the inhibitor, which cannot move further back because of the proximity of loop 602-615. (Supplementary Fig. 6c).

Comparison of the length of loop B and loop 3 in X-ray structures of selected serine proteases.

Supplementary Table 5. Comparison of the length of loop B and loop 3 in X-ray structures of selected serine proteases.

| Enzyme	Loop 3	Length of loop 3	Loop B	Length of loop B	PDB ID
MASP-1	610-628	19	489-515	27	3GOV
MASP-2	593-615	23	482-496	15	1Q3X
Trypsin	164-179	16	56-64	9	1UTN
Matriptase	164-179	16	56-65	10	1EAX
Cathepsin G	162-179	18	56-65	10	1CGH
Chymase	164-179	16	56-65	10	3N7O
Factor XII	164-179	16	56-65	10	6GT6
Kallikrein 5	164-179	16	56-64	9	6QFE
Kallikrein 7	158-173	16	56-63	9	3BSQ
Mesotrypsin	164-179	16	56-64	9	5TP0
Neutrophil elastase	164-179	16	56-64	9	1B0F
Plasmin	706-723	18	602-615	14	6Q1U
Proteinase 3	164-179	16	56-65	10	1FUJ
Thrombin	489-504	16	362-380	19	3U69
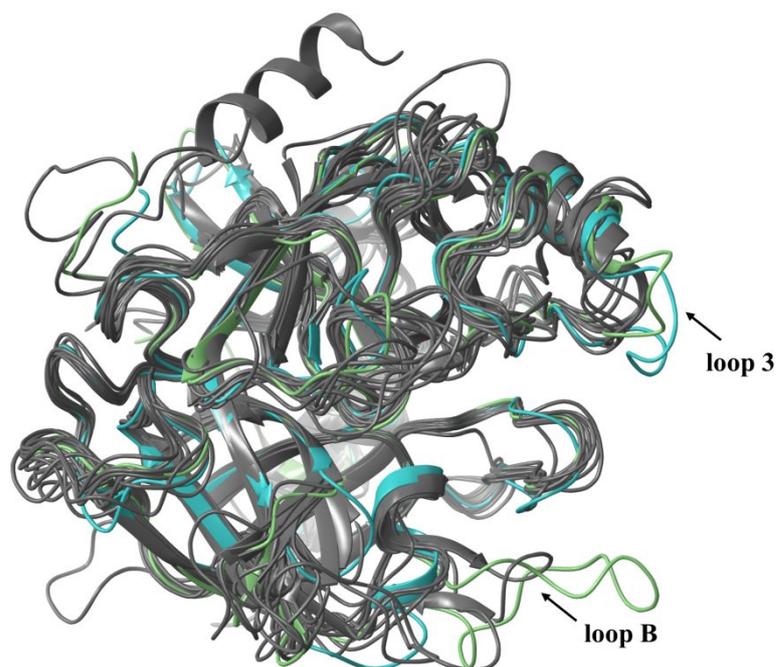
Supplementary Fig. 7. Comparison of the length of loop B and loop 3 in MASP-2 (blue), MASP-1 (green) and selected serine proteases listed in Supplementary Table 5 (gray).

ECD study of SFMI2cap and SFMI2cap-Dap under reducing conditions

(a)

(b)
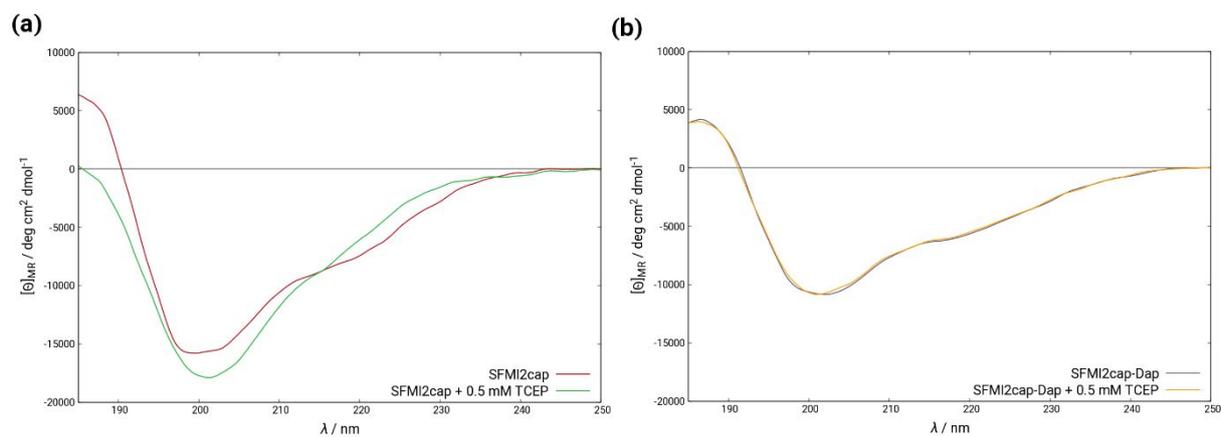
Supplementary Fig. 8. a. Comparison of the ECD spectra of SFMI2cap before and after the addition of 0.5 mM TCEP. **b.** Comparison the ECD spectra of SFMI2cap-Dap before and after the addition of 0.5 mM TCEP

MD simulations of cleaved SFMI2cap and SFMI2cap-Dap in complex with MASP-2
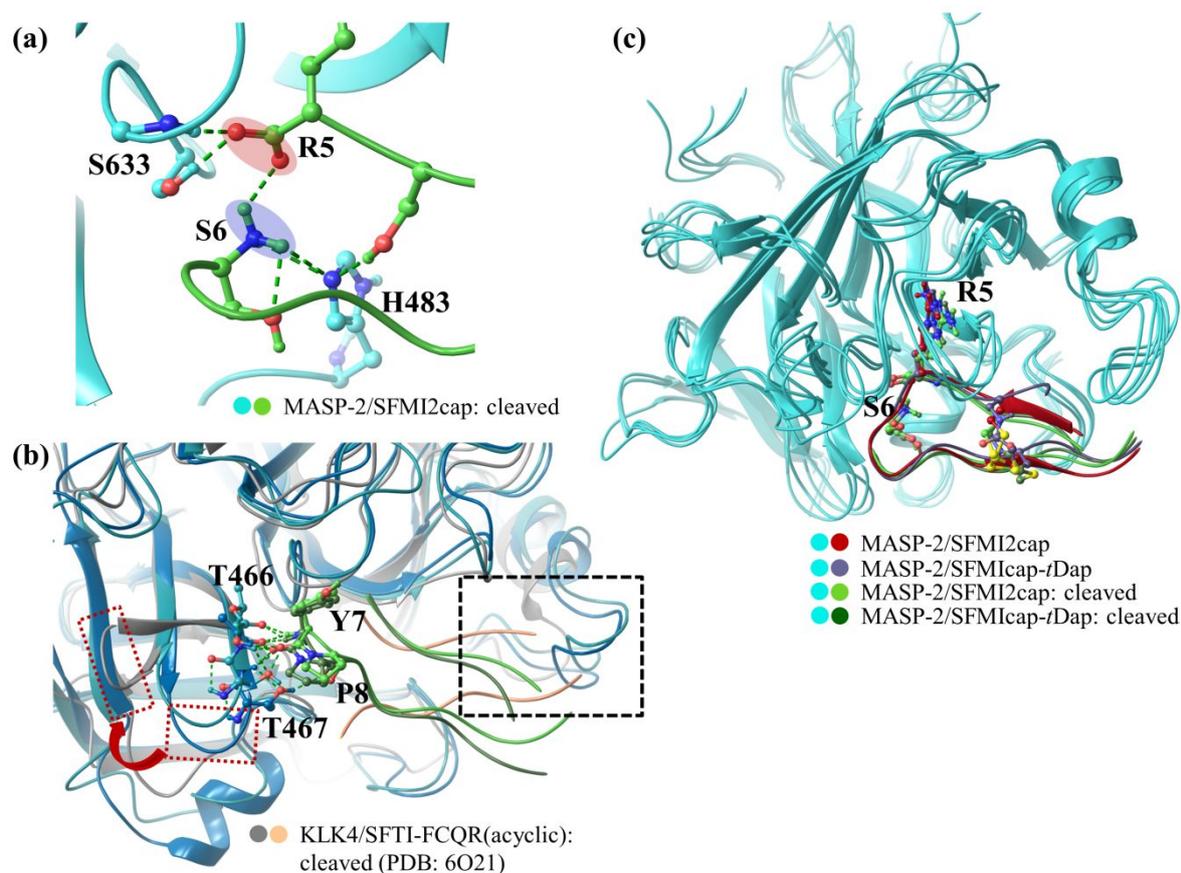
Supplementary Fig. 9. Structural consequences of the cleavage of SFMI2cap and SFMI2cap-*trans*Dap in their complexes formed by MASP-2. **a,** Interactions stabilizing the newly formed C- and N-terminus at the P1-P1' cleavage site. Structural model of the post-proteolysis state of the MASP-2/SFMI2cap complex, with the catalytic Ser633 and His483 residues coordinating the carboxyl and amino groups of the new termini. The mid structures of the most populated backbone clusters of the MD simulations are shown representing over 90% of all snapshots using 1.5Å cutoff. The newly formed carboxyl- and amino-termini between the P1 Arg5 and P1' Ser6 residues of the inhibitors are highlighted with red and blue, respectively. **b,** Comparing the calculated structures to that of KLK4 / SFTI-FCQR(Asn14)[1,14] complex (PDB ID 6O21). The red arrow and red dashed rectangles show the different conformations of the first β-hairpin turn of the N-terminal beta-barrel domains of MASP-2 and KLK4. The closeness of this segment to the inhibitor in the case of MASP-2 allows formation of H-bonds between Thr466-Thr467 of the enzyme and P2'-P3' (Tyr7-Pro8) of the SFMI2cap and SFMI2cap-*trans*Dap inhibitors. The black rectangle outlines the region where the shorter gatekeeper loop of KLK4 allows for a markedly different orientation of the N- and C-terminus of the inhibitors than in the complexes formed with long gatekeeper loop having MASP-2 enzyme. **c,** Comparing the models of the cleaved and intact forms of MASP-2/inhibitor complexes. In the case of SFMI2cap the structure of the post-proteolysis state is remarkably similar to that of the intact enzyme-inhibitor complex. In contrast, in the MASP-2/SFMI2cap-*trans*Dap complex, especially in the P5-P2 (Gly1-Ser4) region, the intact inhibitor conformation deviates, while the cleaved form relaxes to a conformation nearly identical to that seen in the case of SFMI2cap.

Supplementary References

1.   Korsinczky, M. L. J., Schirra, H. J., Rosengren, K. J., West, J., Condie, B. A., Otvos, L., Anderson, M. A., and Craik, D. J. (2001) Solution structures by 1H NMR of the novel cyclic trypsin inhibitor SFTI-1 from sunflower seeds and an acyclic permutant11Edited by M. F. Summers. *Journal of Molecular Biology*. **311**, 579–591

2.   Luckett, S., Garcia, R. S., Barker, J. J., Konarev, A., Shewry, P. R., Clarke, A. R., and Brady, R. L. (1999) High-resolution structure of a potent, cyclic proteinase inhibitor from sunflower seeds. *Journal of molecular biology*. **290**, 525–33

3.   Dobó, J., Harmat, V., Beinrohr, L., Sebestyén, E., Závodszky, P., and Gál, P. (2009) MASP-1, a promiscuous complement protease: structure of its catalytic region reveals the basis of its broad specificity. *J. Immunol.* **183**, 1207–1214

4.   Harmat, V., Gál, P., Kardos, J., Szilágyi, K., Ambrus, G., Végh, B., Náray-Szabó, G., and Závodszky, P. (2004) The Structure of MBL-associated Serine Protease-2 Reveals that Identical Substrate Specificities of C1s and MASP-2 are Realized Through Different Sets of Enzyme–Substrate Interactions. *Journal of Molecular Biology*. **342**, 1533–1546

5.   Héja, D., Harmat, V., Fodor, K., Wilmanns, M., Dobó, J., Kékesi, K. A., Závodszky, P., Gál, P., and Pál, G. (2012) Monospecific inhibitors show that both mannan-binding lectin-associated serine protease-1 (MASP-1) and -2 Are essential for lectin pathway activation and reveal structural plasticity of MASP-2. *J. Biol. Chem.* **287**, 20290–20300