Supporting Material

Fold Specificity Calculations

Based on the full atomistic force field AMBER [1], we performed simulated annealing calculations for an ensemble of several hundred conformers generated for each of the 1000 sequences using CYANA 2.1 [2, 3]. The conformers are within the upper and lower bounds on the Cα-Cα distances and dihedral angles we input into the program. This ensures our de novo design strategy observes true backbone flexibility [4]. The TINKER package [5] is subsequently used for local energy minimization of the conformers. The fold specificities of the sequences are finally calculated using the following formula:

$$f_{specificity} = \frac{\sum_{i \in \text{ new sequence conformers}} \exp[-\beta E_i]}{\sum_{i \in \text{ native sequence conformers}} \exp[-\beta E_i]}$$

where $\beta = 1/k_B T$. A workflow of stage two detailing the programs used and the number of structures that are input and output in each step is given in Fig. S2.

Details on Ensemble Generation for Approximate Binding Affinity Calculation

Structure Prediction: For each sequence from stage one, 3-dimensional structures are generated using RosettaAbinitio [6-8] part of the Rosetta++ software package. RosettaAbinitio is based upon experimental observation that the local structure of the protein is influenced, but not uniquely determined by the local sequence of the protein. The final overall protein structure is obtained when the modification of local structures come together to give a compact structure, accounting for non-local interactions such as buried hydrophobic residues, paired β strands, and specific side chain interactions. For each sequence, 1000 possible structures are generated using Monte Carlo. These structures represent local minima in free energy, not necessary the global minimum.

Clustering: The 1000 decoys from the structure prediction step are clustered based upon their φ and ψ angles using OREO [9, 10]. This groups together decoys that have similar backbone structures. The medoid, or median structure, from each of the ten largest clusters is selected to dock to the target protein. In addition, the overall lowest-energy structure is selected. This gives 11 possible backbone structures for each sequence, incorporating backbone flexibility into the ensemble generation.

When the peptides are clustered, there are typically 15 – 30 clusters formed. Within each cluster, the Cα RMSD difference among structures is on the order of 0.5 – 2 Å, so the medoid is very representative of the cluster. The Cα RMSD between medoids is on the order of 3 – 5 Å. These RMSD ranges among the members of the clusters and between clusters are similar for all sequences, whether that sequence ends up being ranked #1 or #17.

Docking Prediction: Since the binding site of the target protein is known, a refined docking prediction is done using RosettaDock [11-13]. For each sequence, each of the 11 peptide backbone structures is docked against the target protein, near the binding site. For each of the 11 docking runs per sequence, 1000 decoys are generated using Monte Carlo. The peptide is first aligned to variant E1 in the complex structure (to place it near the binding site) and is allowed to translate 3 Å normal to the binding site, 8 Å parallel to the binding site, and rotate 8°. The ten lowest energy decoys in each of the 11 runs are used as starting structures in the final rotamerically-based conformation ensemble generation (110 starting structures per sequence).

Flexibility is introduced into the ensemble by selecting a number of the lowest-energy docking conformers for each peptide structure docked to the target protein. The ten lowest-energy docking conformers that are selected for each peptide for the final ensemble generation stage differ from one another (Cα RMSD) by 0.5 – 3 Å. This spread occurs regardless of peptide structure used to dock or sequence. By selecting the lowest-energy docking conformers, this ensures that the conformers that will make the largest contribution to partition function (i.e. lowest energies) are included.

Final Ensemble Generation: The rotamerically-based conformation ensembles are generated using RosettaDesign [14]. This is done by taking a starting structure and adjusting only the rotamers on the side-chains while leaving the backbone sequence intact.

For each backbone/docked structure, a residue is randomly chosen and the rotamer changed. If this change lowers the energy, the move is accepted, if it raises the energy, the move is accepted based upon the Metropolis Criteria. These steps are repeated until thousands of rotamer substitutions are attempted. The final conformation is a low-energy rotamer conformation that will have a significant contribution to q. 200 low-energy rotamer conformations are generated per backbone/docked structure. The full-atom RMSD differences among the rotamer conformations are slighter, being on the order of 0.5 – 1 Å. These RMSD spreads (between peptide structures, docked conformers, and rotamer conformations) are consistent among all sequences, so we do not find that one ensemble repeatedly counts a similar configuration while another does not.

Predictions with Different Docking Methods

We also compared the effect of a different docking program on the K* ranking. For four sequences (3 variants plus the native sequence), the framework was run again, this time using HADDOCK (High Ambiguity Driven biomolecular DOCKing) [15, 16] as the docking step instead of RosettaDock.

HADDOCK differs from RosettaDock and most other docking programs in that it

uses experimental data, when available, to drive the docking. The experimental data can consist of chemical shift perturbation data, mutagenesis data, or bioinformatic predictions. The data is used to define the residues that are involved in the protein-protein interaction. For the cased of compstatin bound to complement component C3c, the residues involved in the peptide-protein interaction were elucidated from the paper by Janssen et al. [17].

After the docking was finished using HADDOCK, the ten lowest-scoring complexes from the lowest HADDOCK score cluster from each of the eleven docking runs were used as starting structures (110 total starting structures per sequence) for the final ensemble generation (using RosettaDesign). Table S3 shows the ranking results using both RosettaDock and HADDOCK. The

K* ranking appears to be independent of the docking program used. The sequences obtained the same rank-order whether RosettaDock or HADDOCK was used for the docking step.

Table S1: Approximate binding affinity ranking of previous compstatin variants with experimental relative activities

#	peptide	sequence	approx. binding affinity rank	relative activity	ref.
8	Ac-V4Y/H9A	Ac-ICVYQDWGAHRCT-NH_2	1	16	[18]
9	Ac-V4W/H9A (variant E1)	Ac-ICVWQDWGAHRCT-NH_2	2	45	[19]
6	Ac-I1L/H9W/T13G	Ac-LCVVQDWGWHRCG-NH_2	3	4	[20]
3	Ac-R11S	Ac-ICVVQDWGHHSCT-NH_2	4	0.5	[21]
1	Compstatin ring	CVYQDWGAHRC -NH_2	5	0.4	[22]
7	Ac-V4Y/H9F/T13V	Ac-ICVYQDWGFHRCV-NH_2	6	11	[18]
5	Compstatin (native)	ICVVQDWGHHRCT-NH_2	7	1	[23]
2	Ac-I1S/V4F/H9R/ H10L/R11A/T13P	Ac-SCVFQDWGRLACP-NH_2	8	0.5	[20]
4	R11K	ICVVQDWGHHKCT-NH_2	9	0.6	[21]

Table S2: Stage one results for the de novo design of compstatin variants

position	native residue	preferred mutations for run 1	preferred mutations for run 2
1	I	I,L,W,V,F	N,D,Q,I,F,W
2	C	-	-
3	V	-	-
4	V	W	W
5	Q	-	-
6	D	-	-
7	W	-	-
8	G	-	-
9	H	Q,K,H	N,K,T
10	H	N,F,M,Y	N,K,T
11	R	N,K	Q,K
12	C	-	-
13	T	W	W

Table S3: Comparison of ranking metric based upon docking program used: RosettaDock versus HADDOCK

sequence	approx. binding affinity rank using RosettaDock	approx. binding affinity rank using HADDOCK
SQ027	1	1
SQ086	2	2
SQ059	3	3
native	4	4

Table S4: Amino acid sequences of peptides synthesized for Biacore experiments. Brackets denote cyclization. Bold characters denote sequence differences from the native compstatin.

peptide name	amino acid sequence
linear (negative control)	I **A**VVQDWGHHR**A** T-(PEG)$_8$-K-(biotin)-NH_2
native	I[CVVQDWGHHRC]T-(PEG)$_8$-K-(biotin)-NH_2
variant E1	I[CV**W**QDWG**A**HRC]T-(PEG)$_8$-K-(biotin)-NH_2
SQ027	**W**[CV**W**QDWG**TN**RC]**W**-(PEG)$_8$-K-(biotin)-NH_2
SQ059	**D**[CV**W**QDWG**TNK**C]**W**-(PEG)$_8$-K-(biotin)-NH_2
SQ086	**Q**[CV**W**QDWG**QNQ**C]**W**-(PEG)$_8$-K-(biotin)-NH_2

Table S5: Inter-molecular contacts for the best binders of the selected 10 variants (Table 2) and comparison with E1 variant (PDB code 2QKI; [17]). Listed are C3c amino acids within 5 Å from the following atoms of compstatin variants:
Position 1, Ile($C_{\delta1}$), Asn(C_{γ}), Ser(O_{γ}), Pro(C_{γ}), Asp(C_{γ}), Gln(C_{δ}), Gly(C_{α}), Lys($N\zeta$)
Position 2, Cys(S_{γ})
Position 3, Val(C_{β})
Position 4, Trp($N_{\varepsilon1}$, $C_{\zeta2}$, $C_{\varepsilon3}$)
Position 5, Gln(C_{δ})
Position 6, Asp(C_{γ})
Position 7, Trp($N_{\varepsilon1}$, $C_{\zeta2}$, $C_{\varepsilon3}$)
Position 8, Gly(C_{α})
Position 9, Ala(C_{β}), His($N_{\varepsilon1}$), Glu(C_{δ}), Asn(C_{γ}), Thr(C_{β}), Gln(C_{δ}), Lys(N_{ζ})
Position 10, His($C_{\varepsilon1}$), Asn(C_{γ}), Lys(N_{ζ})
Position 11, Arg(C_{ζ}), Lys(N_{ζ}), Gln(C_{δ})
Position 12, Cys(S_{γ})
Position 13, Thr(C_{β}), Trp($N_{\varepsilon1}$, $C_{\zeta2}$, $C_{\varepsilon3}$)

variant	side chain contacts within 5 Å
SQ087	Asn1-Asp348/Lys385/Ser387/Ser436/Val437/Leu438/Arg439 Cys2-Ser387 Val3-Pro346/His434/Ser436/Asn451/Leu453/Leu491 Trp4-Ser436/Val437/Leu438/Leu448/Asn449/Val450/Asn451 Asp6-Asn451/Leu491 Lys9-Pro346/Asn389 Cys12-Ser387 Trp13-Val374/Ser387/Ile388/Asn389
SQ072	Asp1-Ser376/Leu377/Arg439 Cys2-Arg439 Val3-Leu438/Arg439/Thr440 Trp4-Asp348/Lys385/Ser436/Val437/Leu438/Asn451 Gln5-Asp348/Lys385/Ser387 Asp6-Val374/Gln375/Ser376/Lys385/Leu386/Ser387 Trp7-Ser387/Asn389 Gln9-Val374

Continued on Next Page …

Table S5 – Continued

variant	side chain contacts within 5 Å
SQ077	Val3-Arg458 Gln5-Arg458/Glu461/Arg485/Gly488/Asp490 Asp6-Arg458/Ala459 Trp7-Glu486/Pro487/Gly488 Gly9-Gly488 Gln11-Leu454/Arg455/Met456/Asp490 Trp13-Gly344/Met345/Pro346/Leu453/Leu454/ Arg455/Met456/Asp490
SQ040	Trp4-His391/Pro392/Ser393 Gln5-Pro392/Ser393/Gln394/Asn430 Trp7-Thr425/Asn428/Ser429/Asn430/Tyr432/Gln394 Thr9-Pro392/Asn430/Asp457 Asn10-Arg455/Met456/Asp457/Arg458 Gln11-Gly344/Arg455/Pro392 Cys12-Pro392 Trp13-Arg458
SQ098	Val3-Gly370 Trp4-Lys395 Gln5-Gly370/Glu371 Asp6-Ser398/Ile399 Trp7-Val368/Gln369/Gly370/Glu371/Ile388/Asn389/ Leu397/Ser398/Ile399/Thr400 Gly8-His391 Asn9-Glu371/Asn389/Thr390 Lys11-Glu371/Asn389 Trp13-Gly370/Glu371/Asp372/Thr373/Val374
SQ086	Gln1-Asp490/Leu491 Trp4-Met345/Pro346/Leu453/Leu491/Gly344/Asn389/Thr390 Gln5-Asp490 Asp6-Arg455 Trp7-Leu454/Arg455/Met456/Asp457/Arg458/Glu461/ Arg485/Gly488/Gln489/Asp490 Gln11-Arg458/Gly488 Cys12-Asp490

Continued on Next Page …

Table S5 – Continued

variant	side chain contacts within 5 Å
SQ024	Ser1-Pro346/Lys385/Ser387/Leu453 Cys2-Asn389 Val3-Gly344/Met345/Pro346/Leu453/Arg455 Trp4-Gly344/Asn389/Thr390/His391/Pro392 Gln5-Arg455/Asp490 Asp6-Asn430/Arg455/Met456/Asp457 Trp7-Arg458 Asn9-Arg458 Gln11-Asp490 Trp13-Asn451/Leu491/Val493
SQ059	Val3-Gly344/Thr390/His391/Pro392 Trp4-Gly344/Met345/Pro346/Leu453/Asp490 Gln5-Leu454/Asp490 Asp6-Arg455/Asp457/Arg458 Trp7-Arg458/Glu461/Gly488 Cys12-Asp490
SQ055	Val3-Gly344/Thr390/His391/Pro392 Trp4-Gly344/Met345/Pro346/Leu453/Asp490 Gln5-Leu454/Asp490 Asp6-Arg455/Asp457/Arg458 Trp7-Arg458/Glu461/Gly488 Cys12-Asp490
SQ088	Gly1-Leu454/Arg455/Met456/Asp490 Val3-Met456/Asp457/Arg458/Glu461 Trp4-Gly488 Gln5-Asp490 Asp6-Arg458 Asn10-Gly344/Asn389/Thr390/His391/Pro392 Gln11-Asn389 Trp13-Gly344/Met345/Pro346/Leu453/Leu454/Arg455/Asp490
E1	Ile1-Asn390 Cys2-Asn390 Val3-Gly345/Met346/Pro347/Leu455/Arg456 Trp4-Gly345/Thr391/His392/Pro393 Gln5-Leu455/Arg456/Met457/Asp491 Trp7- Leu455/Met457/Asp458/Arg459/Glu462/Arg486/ Gly489/Gln490/Asp491 Gly8-Gly489/Asp491 Ala9-Asp491 His10-Leu492 Cys12-Gly345

Table S6**:** Inter-molecular hydrogen bonds for the best binders of the selected 10 variants (Table 2) and comparison with E1 variant (PDB code 2QKI; [17]). The hydrogen bond criteria were: distance between donor and acceptor atom < 2.4 Å and angle hydrogen atom-acceptor heavy atom-donor heavy atom < 35°. Hydrogen bonds were calculated using MOLMOL [24].

approx. binding affinity rank	variant	donor/ acceptor	C3c	donor/ acceptor	distance (Å)	angle (°)
3	SQ087					
	Asn1	$H^{\delta 21}$	Asp348	$O^{\delta 2}$	2.4	31.7
	Trp4	$H^{\varepsilon 1}$	Asn449	O	2.0	25.8
	Lys9	$H^{\zeta 3}$	Asn389	$O^{\delta 1}$	1.9	8
4	SQ072					
	Asp1	$O^{\delta 1}$	Ser376	H^{γ}	2.0	13.9
	Asp6	$O^{\delta 1}$	Ser387	H^{N}	2.0	9.3
	Val3	O	Arg439	H^{N}	2.4	18.5
	Trp4	$H^{\varepsilon 1}$	Ser436	O^{γ}	2.1	17.4
5	SQ077					
	Gln5	$H^{\varepsilon 21}$	Gly488	O	2.2	31.0
	Asp6	$O^{\delta 1}$	Arg458	H^{ε}	1.9	12.7
	Asp6	$O^{\delta 1}$	Arg458	$H^{\eta 21}$	1.8	20.5
	Gln11	$H^{\varepsilon 22}$	Asp490	$O^{\delta 1}$	2.4	24.8
6	SQ040					
	Asn10	H^{N}	Asp457	$O^{\delta 1}$	1.8	18.9
	Gln11	$O^{\varepsilon 1}$	Arg455	$H^{\eta 1}$	1.9	32.3
7	SQ098					
	Gln5	$H^{\varepsilon 22}$	Glu371	N	2.3	10.6
	Trp13	$H^{\varepsilon 1}$	Asp372	N	2.4	21
10	SQ086					
	Gln1	$H^{\varepsilon 22}$	Asp490	$O^{\delta 1}$	2.0	13.4
	Trp4	O	Arg455	$H^{\eta 11}$	1.9	14.8
	Gln5	$H^{\varepsilon 21}$	Asp490	$O^{\delta 2}$	2.0	26.9
	Trp7	$H^{\varepsilon 1}$	Met456	O	1.8	31
	Gln11	$O^{\varepsilon 1}$	Arg458	$H^{\eta 12}$	2.1	16.5
13	SQ024					
	Trp4	O	Arg455	H^{ε}	2.0	22
	Trp4	H^{N}	Gly344	O	1.8	6.6
	Asn9	$H^{\delta 22}$	Arg458	NH_2	2.1	13.5
14	SQ059					
	Gly8	O	Arg458	H^{ε}	2.0	9.3
	Gly8	O	Arg458	$H^{\eta 21}$	2.0	19.9

Continued on Next Page …

Table S6 – Continued

approx. binding affinity rank	variant	donor/ acceptor	C3c	donor/ acceptor	distance (Å)	angle (°)
15	SQ055					
	Val3	O	Arg455	$H^{\eta 11}$	2.1	33.7
	Asp6	$O^{\delta 1}$	Arg455	$H^{\eta 12}$	2.1	3.3
	Asp6	$O^{\delta 2}$	Arg455	$H^{\eta 22}$	2.1	5.5
17	SQ088					
	Val3	O	Arg458	H^{ε}	2.0	11.6
	Val3	O	Arg458	$H^{\eta 21}$	1.9	19.1
	Gln11	$H^{\varepsilon 22}$	Asn389	$O^{\delta 1}$	1.9	4
	Trp13	$H^{\varepsilon 1}$	Leu454	O	1.8	15.5
11	E1					
	Ile1	H^{N}	Asn390	$O^{\delta 1}$	2.1	29.6
	Trp4	H^{N}	Gly345	O	2.1	4.2
	Gln5	$N^{\varepsilon 2}$	Asp491	$H^{\delta 2}$	2.0	29.9
	Trp7	$H^{\varepsilon 1}$	Met457	O	1.6	2.7
	His10	H^{N}	Asp491	$O^{\delta 1}$	1.9	20.8

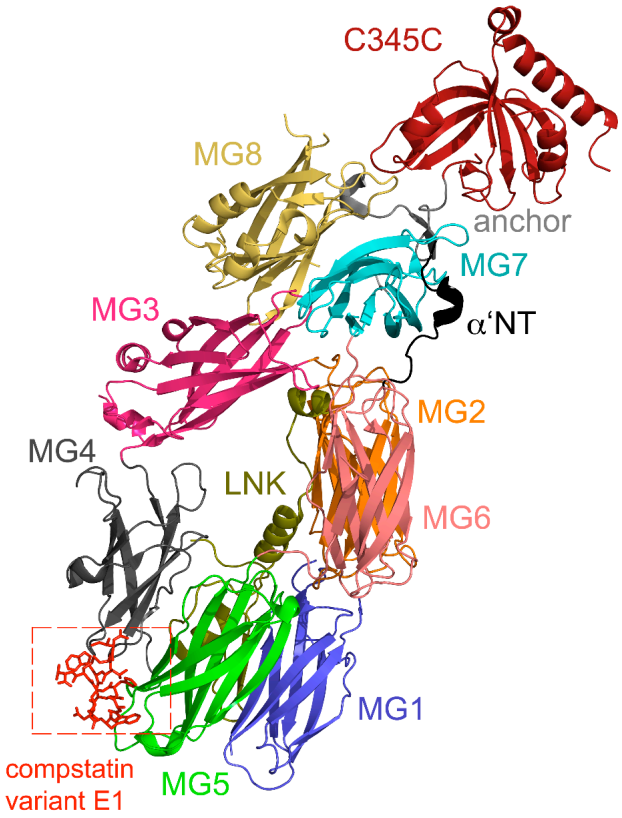

Figure S1: Ribbon representation of the crystal structure of C3c-compstatin complex as elucidated by Janssen et al. [17] at a resolution of 2.4 Å (chains A, B, C, and G in PDB file 2QKI).

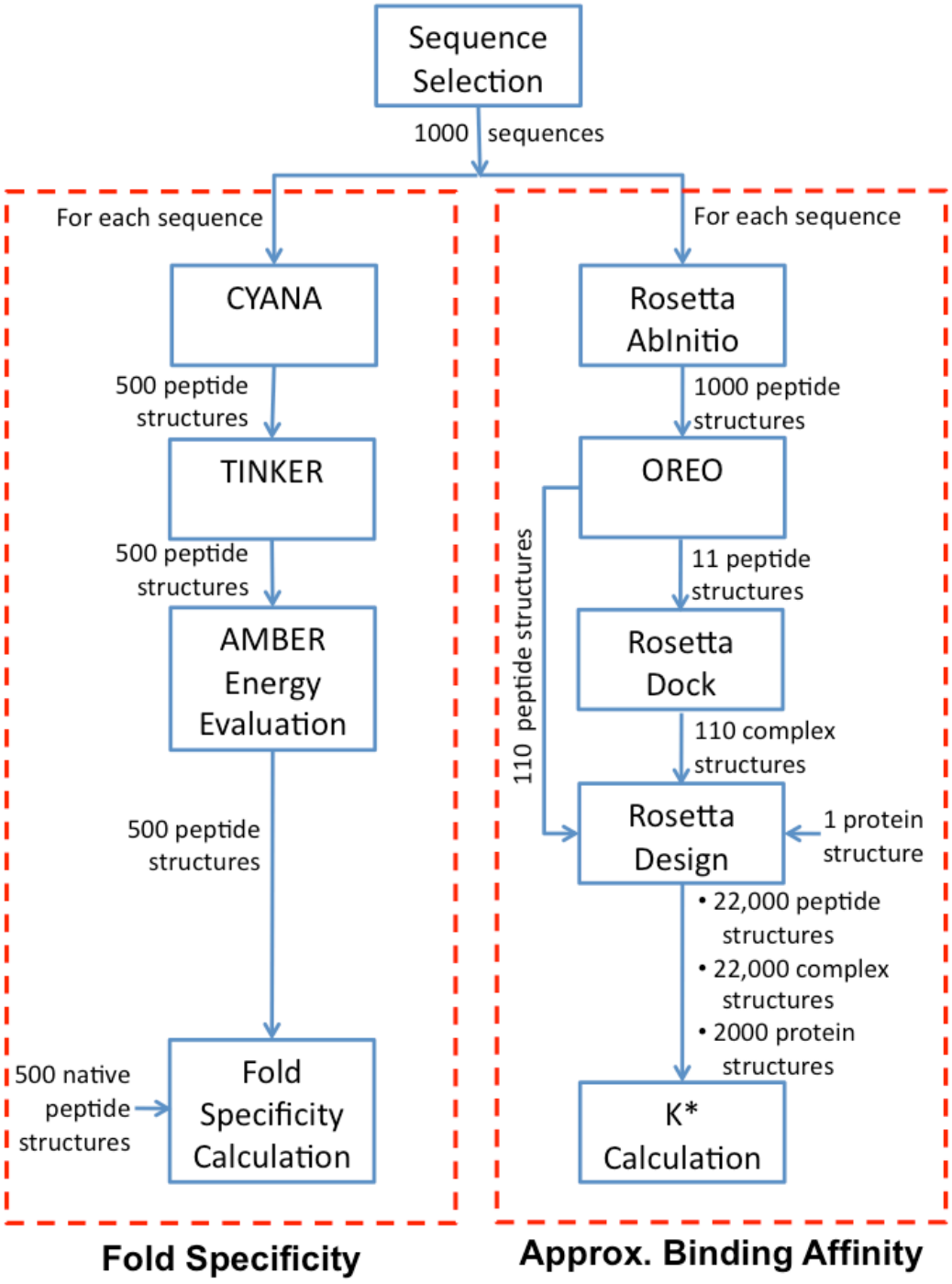

Figure S2: Workflow of two de novo design frameworks: fold specificity and approximate binding affinity

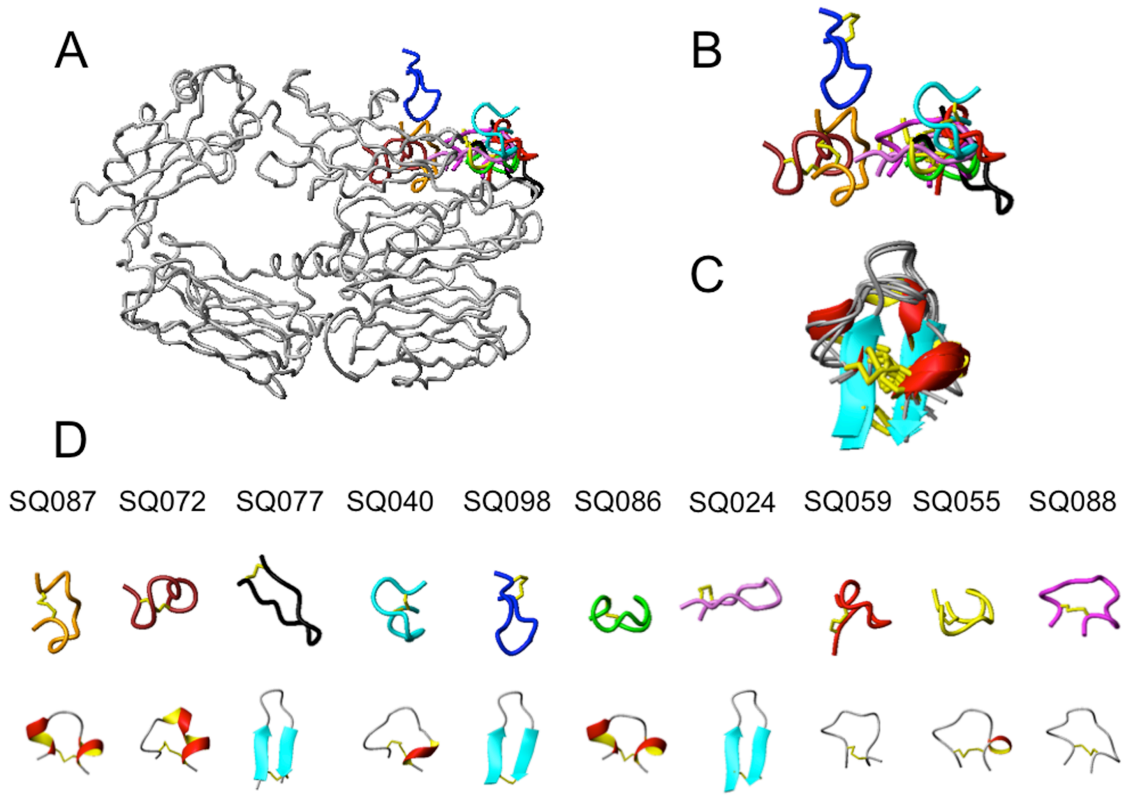

Figure S3: Structural analysis of selected 10 C3c-compstatin variant complexes of Table 2. (A) Superposition of the best binders of the selected 10 C3c-compstatin variant complexes. The backbone C_α-atoms of C3c have been used for the superposition. Only the β-chain of C3c is shown, to which compstatin binds, whereas the remaining portion of C3c has been deleted for clarity. The 10 C3c structures are identical and appear as one structure (colored in gray), whereas the compstatin variant structures are colored differently (see Panel (D) for color code). (B) Relative topology of the structures of the best binders of the selected 10 compstatin variants. The structure of C3c is deleted for clarity. The color code is explained in the middle row of Panel (D). The disulfide bridge is shown in yellow. (C) Superposition of the structures of the best binders of the selected 10 compstatin variants using the backbone C_α-atoms. The secondary structures and the disulfide bridge are shown in ribbon and stick representations, respectively. (D) The best binders of the selected 10 compstatin variant structures arranged according to their approximate binding affinities in decreasing order from left (highest) to right (lowest). The sequence id number from Table 2 is marked in the upper row. The middle row depicts the color code of Panels (A) and (B). The bottom row shows the secondary structure of each compstatin variant. Molecular graphics were prepared using the program MOLMOL [24].

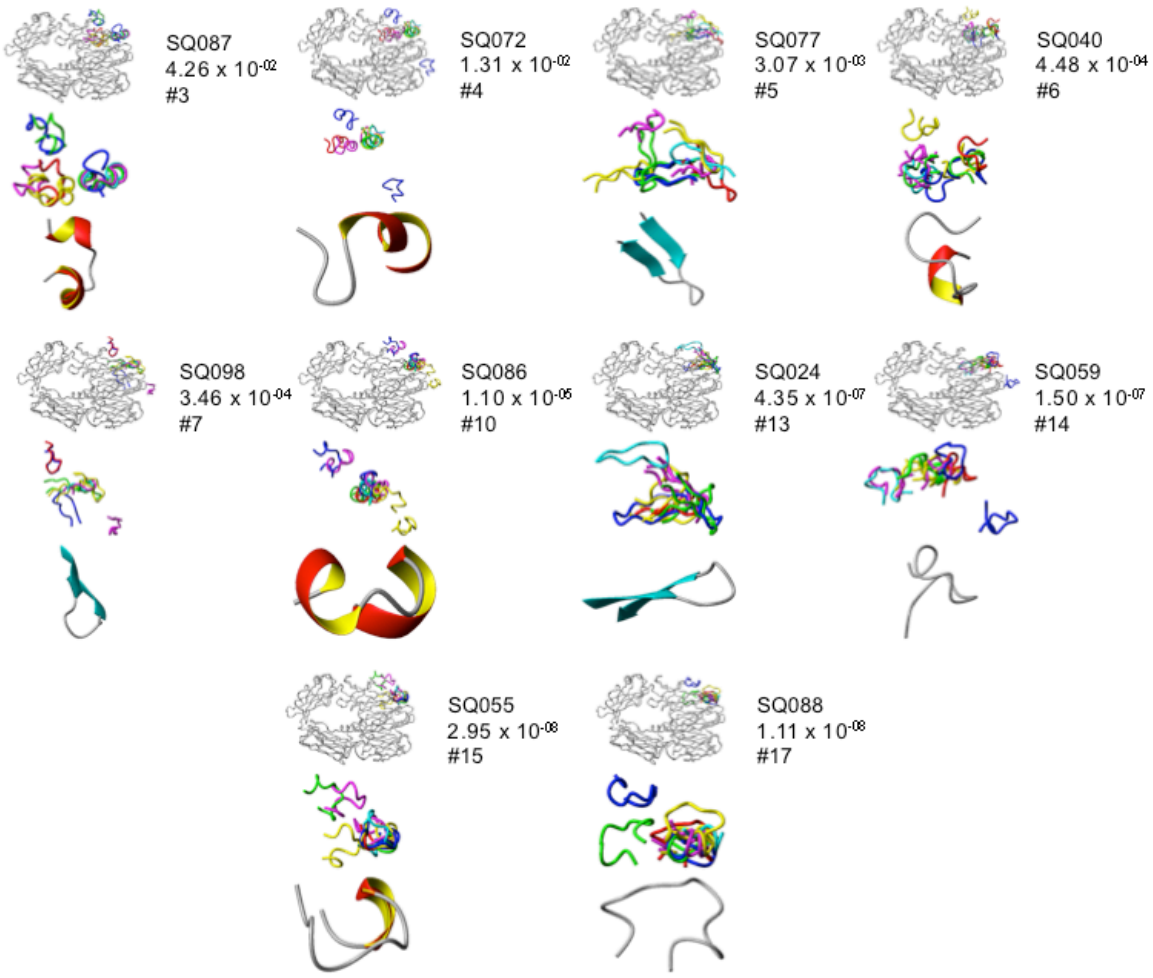

Figure S4: The molecular graphics show the 10 best binders of selected 10 compstatin variants in order of decreasing approximate binding affinity: SQ087, SQ072, SQ077, SQ040, SQ098, SQ086, SQ024, SQ059, SQ055, SQ088 (Table 2). Each figure contains fits of all 10 C3c-compstatin variant structures within each SQ# run. The top panels are C3c-compstatin variant complexes. Out of 3 C3c chains, we show only the β-chain to which the compstatin variant binds. The 10 structures are fitted using the Cα-atoms of C3c only. There are 10 identical overlapping C3c structures in each panel. The color code for compstatin variant follows the RGBYMC order, with red being the lowest energy structure and cyan being the highest energy structure from the group of 10 best binders, as follows: red = structure 1, green = structures 2-3, blue = structures 4-5, yellow = structures 6-7, magenta = structures 8-9, cyan = structure 10. The middle panels depict the compstatin variant alone in the exact same orientation as in the top panels. The color code is the same as in the top panels. The bottom panels contain the structures of the middle panels fitted to the first (lowest energy) structure using the backbone Cα-atoms. There are actually 10 identical overlapping structures. The secondary structure is shown in ribbon representation. There is some variation within a consensus binding site and significant structural and orientational variation. There are 3 β-hairpin, 2 random coil, and 5 structures with helical segments of compstatin variants.

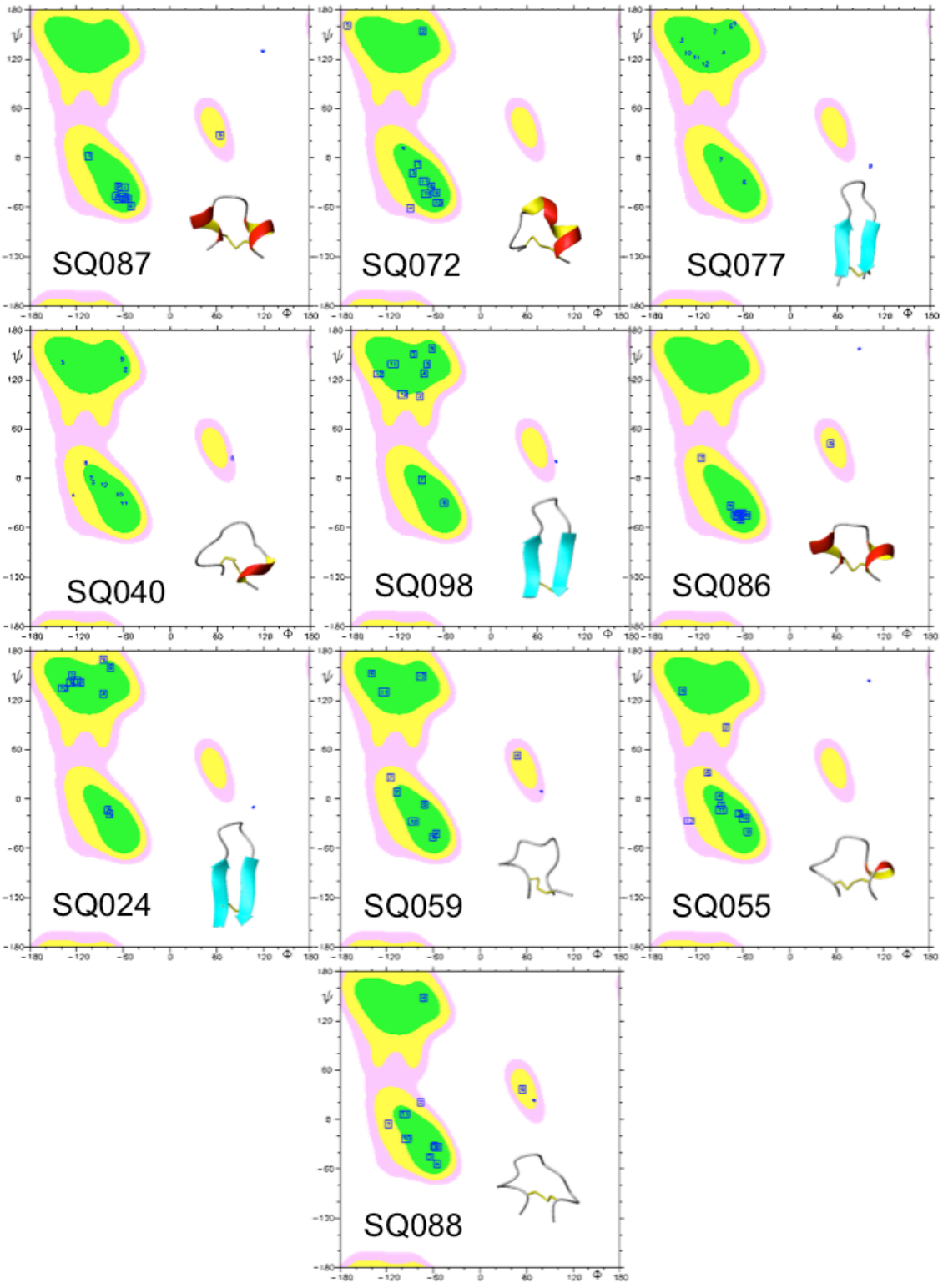

Figure S5: Ramachandran plot analysis of secondary structures for the selected 10 compstatin variants (Table 2). The backbone (ϕ, ψ)-torsion angle pairs are marked using the amino acid number. Ribbon diagrams of the structure of the peptides are also shown in the panels. The variants are arranged in decreasing order of their approximate binding affinities. The panels were prepared using the program MOLMOL [24]

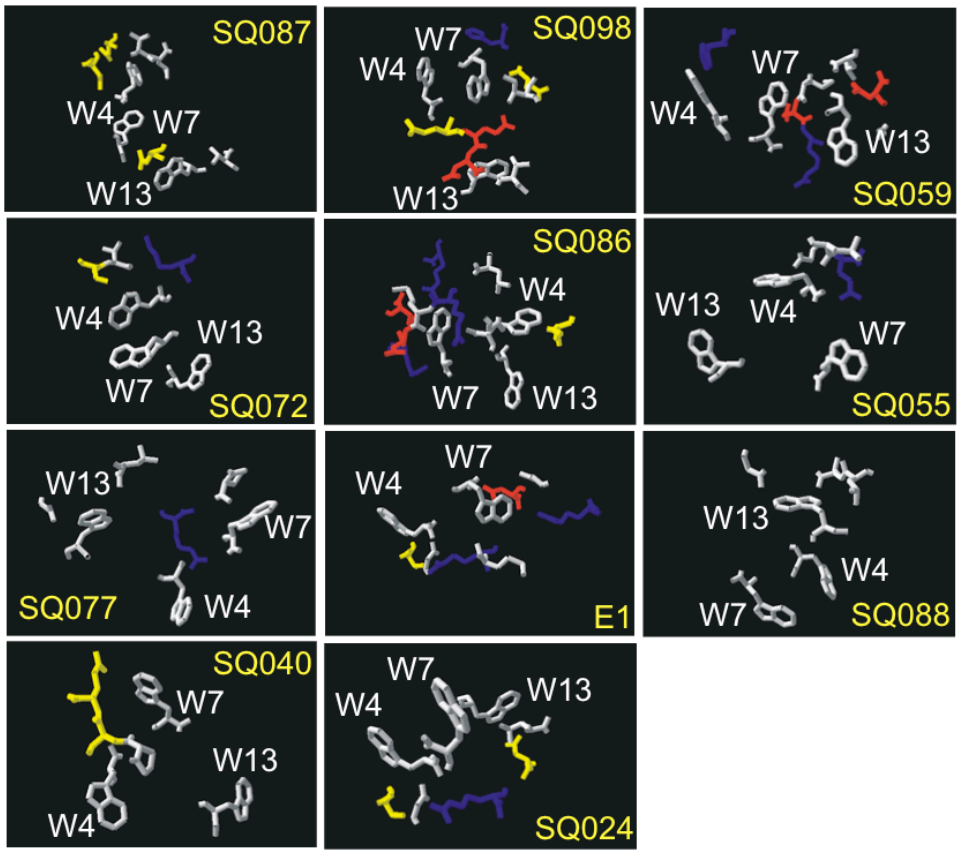

Figure S6: Side chain interactions between each of the three Tryptophans of the best binders of the selected 10 compstatin variants (Table 2) and C3c amino acids. The panels depict amino acids in the C3c binding site within 3.5 Å from Trp4, Trp7, and Trp13, using any backbone or side chain atoms. The color code for the stick representations is: gray for non-polar amino acids, red for acidic, blue for basic, yellow for neutral polar. The E1 variant (crystal structure; [17]) is also included for comparison. Molecular graphics were prepared using the program SPDBV [25].

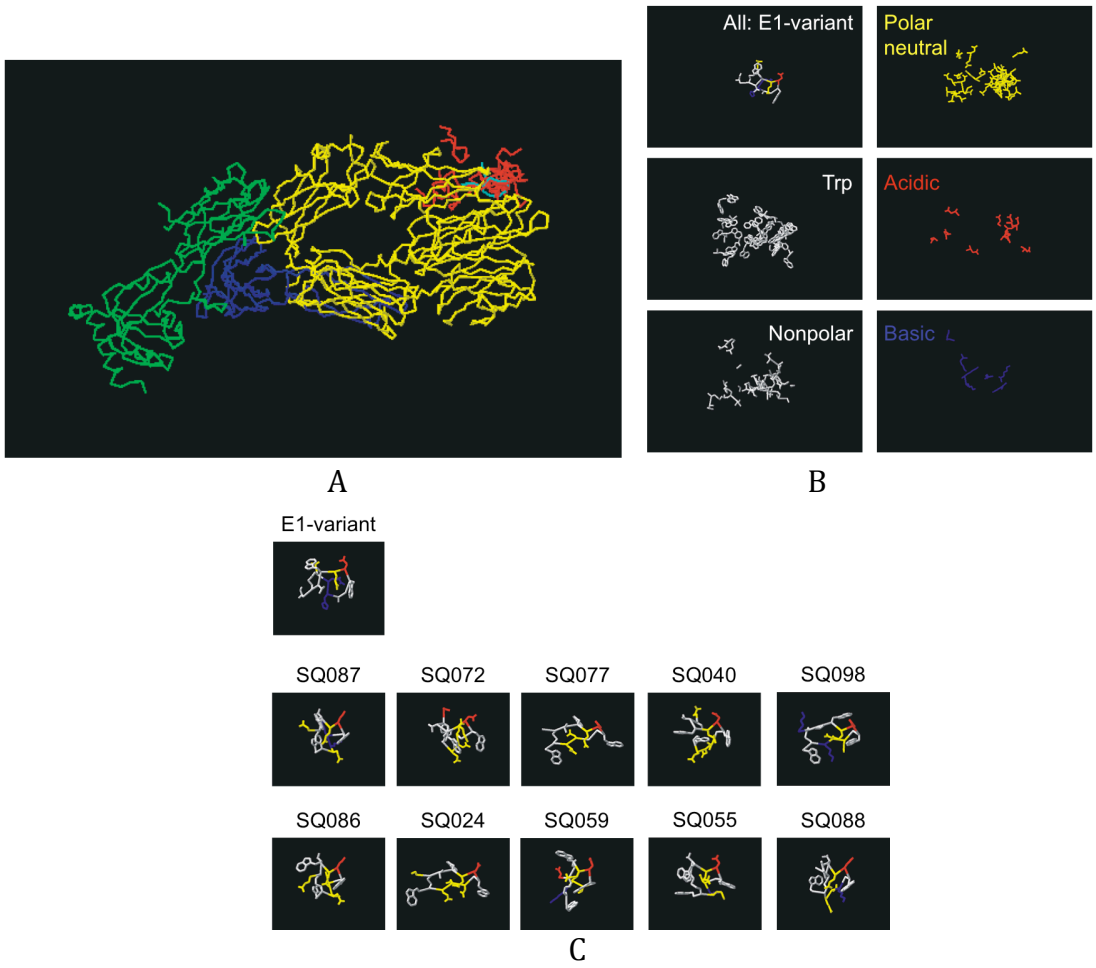

Figure S7: (A) Complexes of C3c (in yellow, green, and blue) and the best binders of the selected 10 compstatin variants (in red; from Table 2) and the E1 variant (in cyan; from crystal structure with PDB code 2QKI; [17]). The structures are superimposed using the Cα-atoms of C3c only and drawn using Cα-traces. Molecular graphics are made with SPDBV [25] (B) Side chains of E1-variant and the best binders of the selected 10 compstatin variants from Table 2. Side chain topologies and orientations as in (A). (C) Intra-molecular side chain interactions for E1 variant (PDB code 2QKI; [17]) and best binders of the selected 10 compstatin variants of Table 2 (arranged in decreasing approximate binding affinities). The structures are fitted using the backbone Cα-atoms and they are shown in the same orientation. Backbones are shown as Cα-traces. Color code as in (A).

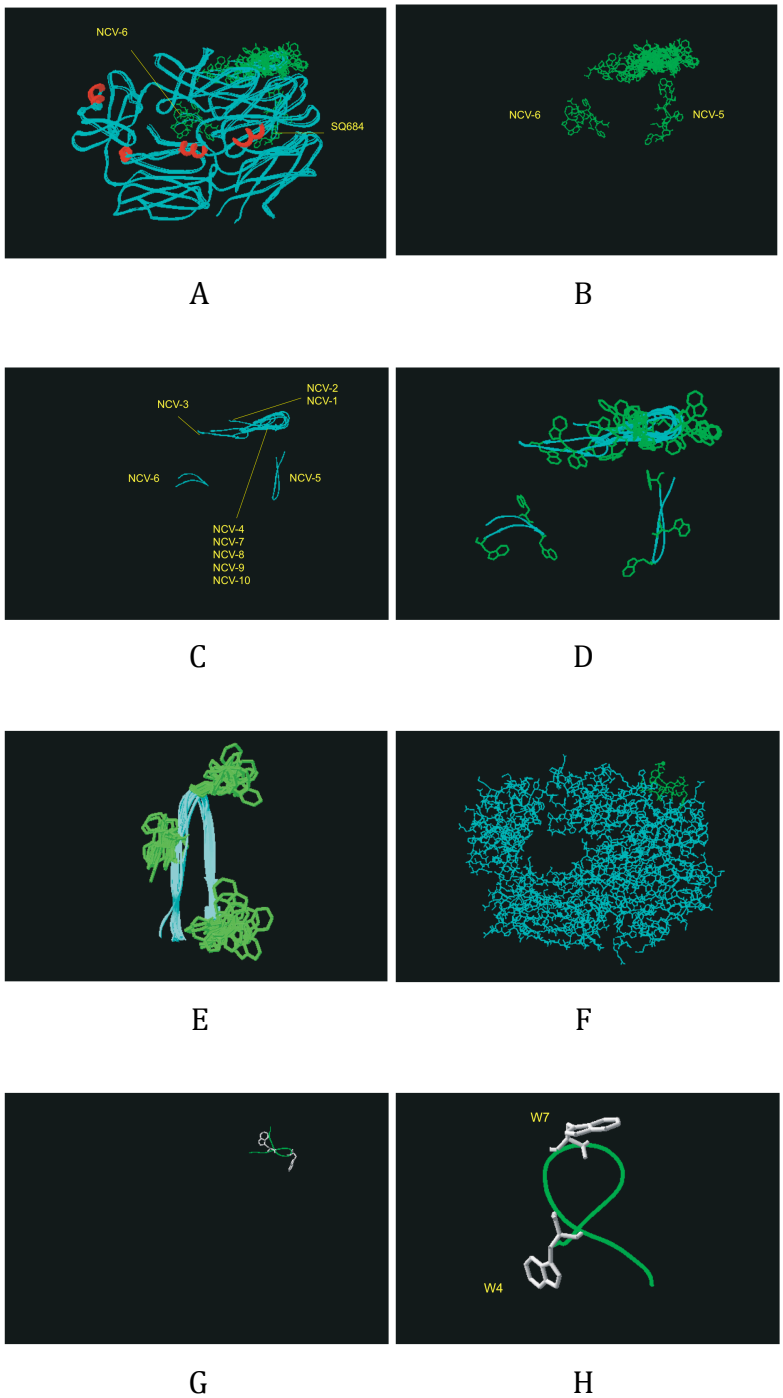

Figure S8: (A-D) Superposition of 10 structures (NCV-1 to NCV-10), using the backbone Cα- atoms of the complexes. Compstatin is colored in green. Molecular graphics were prepared using SPDBV [25] (A) - (D) show structures on the same orientations and topologies. (E) The best binders of compstatin variants NCV-1 to -10 (Table 2) superimposed using the backbone Cα-atoms. The three Tryptophans are shown (from left to right W4, W7, and W13). (F) The E1 variant (green) in complex with the β-chain of C3c (cyan). PDB code 2QKI; [17] (G) The E1 variant (green) in the same orientation and topology as in (F), with the two Tryptophans (W4 and W7) shown. (H) The E1 variant with the two Tryptophans (W4 and W7) shown.

References

[1] Cornell, W. D., P. Cieplak, C. I. Bayly, I. R. Gould, K. M. Merz, D. M. Ferguson, D. C. Spellmeyer, T. Fox, J. W. Caldwell, and P. A. Kollman, 1995. A 2nd Generation Force-Field For The Simulation Of Proteins, Nucleic-Acids, And Organic-Molecules. J. Am. Chem. Soc. 117:5179–5197.

[2] Guntert, P., C. Mumenthaler, and K. Wuthrich, 1997. Torsion Angle Dynamics for NMR Structure Calculation with the New Program DYANA. J. Mol. Biol. 273:283–298.

[3] Guntert, P., 2004. Automated NMR structure calculation with CYANA. J. Mol. Biol. 278:353–378.

[4] Floudas, C. A., 2005. Research Challenges, Opportunities and Synergism in Systems Engineering and Computational Biology. AIChE J. 51:1872–1884.

[5] Ponder, J., 1998. TINKER, Software Tools for Molecular Design. Department of Biochemistry and Molecular Biophysics, Washington University School of Medicine: St. Louis, MO.

[6] Rohl, C. A., C. E. M. Strauss, K. M. S. Misura, and D. Baker, 2004. Protein Structure Prediction Using Rosetta. Methods Enzymol. 383:66–93.

[7] Lee, M. R., D. Baker, and P. A. Kollman, 2001. 2.1 and 1.8 A° Cα RMSD Structure Predictions on Two Small Proteins, HP-36 and S15. J. Am. Chem. Soc. 123:1040–1046.

[8] Rohl, C. A., and D. Baker, 2002. De Novo Determination of Protein Backbone Structure from Residual Dipolar Couplings Using Rosetta. J. Am. Chem. Soc. 124:2723–2729.

[9] DiMaggio, P. A., S. R. McAllister, C. A. Floudas, X. J. Feng, J. D. Rabinowitz, and H. A. Rabitz, 2008. Biclustering via optimal re-ordering of data matrices in systems biology: rigorous methods and comparative studies. BMC Bioinformatics 9.

[10] DiMaggio, P. A., S. R. McAllister, C. A. Floudas, X. J. Feng, J. D. Rabinowitz, and H. A. Rabitz, 2009. Biclustering via Optimal Re-ordering of Data Matrices. Journal of Global Optimization in press.

[11] Gray, J. J., S. Moughon, C. Wang, O. Schueler-Furman, B. Kuhlman, C. A. Rohl, and D. Baker, 2003. Protein-Protein Docking with Simultaneous Optimization of Rigid-body Displacement and Side-chain Conformations. J. Mol. Biol. 331:281 – 299.

[12] Daily, M. D., D. Masica, A. Sivasubramanian, S. Somarouthu, and J. J. Gray, 2005. CAPRI Rounds 3-5 Reveal Promising Successes and Future Challenges for RosettaDock. Proteins 60:181–186.

[13] Gray, J. J., S. E. Moughon, T. Kortemme, O. Schueler-Furman, K. M. S. Misura, A. V. Morozov, and D. Baker, 2003. Protein-Protein Docking Predictions for the CAPRI Experiment. Proteins 52:118–122.

[14] Kuhlman, B., and D. Baker, 2000. Native Protein Sequences Are Close to Optimal for Their Structures. Proc. Natl. Acad. Sci. USA 97:10383–10388.

[15] Dominguez, C., R. Boelens, and A. M. J. J. Bonvin, 2003. HADDOCK: a protein-protein docking approach based on biochemical and/or biophysical information. J. Am. Chem. Soc. 125:1731–1737.

[16] de Vries, S. J., A. D. J. van Dijk, M. Krzeminski, M. van Dijk, A. Thureau, V. Hsu, T. Wassenaar, and A. M. J. J. Bonvin, 2007. HADDOCK versus HADDOCK: New features and performance of HADDOCK2.0 on the CAPRI targets. Proteins: Struc. Funct. Bioinf. 69:726–733.

[17] Janssen, B. J. C., E. F. Halff, J. D. Lambris, and P. Gros, 2007. Structure of Compstatin in Complex with Complement Component C3c Reveals a New Mechanism of Complement Inhibition. J. Biol. Chem. 282:29241–29247.

[18] Klepeis, J. L., C. A. Floudas, D. Morikis, C. G. Tsokos, E. Argyropoulos, L. Spruce, and J. D. Lambris, 2003. Integrated Structural, Computational and Experimental Approach for Lead Optimization: Deisgn of Compstatin Variants with Improved Activity. J. Am. Chem. Soc. 125:8422–8423.

[19] Mallik, B., M. Katragadda, L. Spruce, C. Carafides, C. Tsokos, D. Morikis, and J. Lambris, 2005. Design and NMR Characterization of Active Analogues of Compstatin Containing Non-natural Amino Acids. J. Med. Chem. 48:274–286.

[20] Soulika, A., D. Morikis, M. Sarias, M. Roy, L. Spruce, A. Sahu, and J. Lambris, 2003. Studies of Structure-Activity Relations of Complement Inhibitor Compstatin. J. Immunology 171:1881–1890.

[21] Morikis, D., M. Roy, A. Sahu, A. Torganis, P. Jennings, G. Tsokos, and J. Lambris, 2002. The structural basis of compstatin activity examined by structure-function-based design of peptide analogs and NMR. J. Biol. Chem. 277:14942–14953.

[22] Morikis, D., and J. Lambris, 2002. Structural aspects and design of low molecular mass complement inhibitors. Biochem. Soc. Trans. 30:1026–1036.

[23] Sahu, A., B. Kay, and J. Lambris, 1996. Inhibition of human complement by a C3-binding peptide isolated from a phage displayed random peptide library. J. Immunol. 157:884–891.

[24] Koradi, R., M. Billeter, and K. W¨uthrich, 1996. MOLMOL: A program for display and analysis of

macromolecular structures. J. Mol. Graphics 14:51– 55.
[25] Guex, N., and M. C. Peitsch, 1997. SWISS-MODEL and the Swiss- PdbViewer: an environment for comparative protein modeling. Electrophoresis 18:2714–2723.