Supplementary Information for

Conserved patterns hidden within group A *Streptococcus* M protein hypervariability are responsible for recognition of human C4b-binding protein

Cosmo Z. Buffalo[1], Adrian J. Bahn-Suh[1], Sophia P. Hirakis[1], Tapan Biswas[1], Rommie E. Amaro[1], Victor Nizet[2,3], and Partho Ghosh[1]

Department of Chemistry & Biochemistry[1], Pediatrics[2], and Skaggs School of Pharmaceutical Sciences[3], University of California, San Diego, La Jolla, CA 92093

Supplementary Table S1. Data collection, phasing and refinement statistics for native and SAD (SeMet) structures

	M2-C4BP	M2-C4BP (SeMet L46M/ L71M)	M2-C4BP (SeMet L29M/ L46M)	M28-C4BP	M22-C4BP	M49-C4BP (SeMet L29M/ L46M)	M2 (K65A/ N66A)-C4BP
PDB ID	5HYU			5HYP	5HYT	5HZP	5I0Q
Data collection							
Space group	P 4_3 3 2	P 4_3 3 2	P 4_3 3 2	P 4_3 3 2	P 2_1 2_1 2_1	P 4_3 2_1 2	P 4_3 3 2
Cell dimensions							
a, *b*, *c* (Å)	148.3 148.3 148.3	148.7 148.7 148.7	148.3 148.3 148.3	133.7 133.7 133.7	68.08 80.35 152.9	78.1 78.1 345.3	148.6 148.6 148.6
α, β, γ(°)	90, 90, 90	90, 90, 90	90, 90, 90	90, 90, 90	90, 90, 90	90, 90, 90	90, 90, 90
Wavelength	0.984 Å	0.979 Å	0.979 Å	0.979 Å	0.979 Å	0.979 Å	0.979 Å
Resolution (Å)	50.00-2.56(2.60-2.56)[a]	50.00-2.90(2.95-2.90)[a]	50.00-3.00(3.05-3.00)[a]	50.00-3.02(3.07-3.02)[a]	80.40-2.54(2.68-2.54)[a]	86.40-2.74(2.89-2.74)[a]	50.00-2.29(2.37-2.29)[a]
R_{merge}	0.18(1.00)	0.18(1.00)	0.15(1.00)	0.14(1.00)	0.14(1.00)	0.23(1.00)	0.13(1.00)
I / σ_I	67.3(4.17)	21.3(2.17)	22.4(3.44)	13.2(0.81)	10.7(1.2)	11.3(1.00)	46.4(3.22)
Completeness (%)	100(100)	99.9(100)	100(100)	99.8(100)	99.8(99.9)	99.8(99.1)	100(100)
Redundancy	42.8(43.1)	40.6(40.7)	22.9(22.6)	9.5(9.7)	6.6(6.7)	9.4(9.6)	40.5(32.6)
$cc_{1/2}$	1.00(0.86)	0.99(0.86)	0.99(0.86)	0.99(0.48)	1.00(0.60)	1.00(0.61)	0.99(0.86)
Refinement							
Resolution (Å)	35.97-2.56(2.65-2.56)			44.40-3.02(3.21-3.02)	76.50-2.54(2.60-2.54)	71.16-2.74(2.84-2.74)	49.54-2.29(2.37-2.29)
No. reflections	18514 (1808)			7741 (1084)	28328 (1975)	28700 (2449)	47375 (4655)
R_{work} / R_{free}	0.21(0.27)/ 0.22(0.28)			0.25(0.33)/ 0.29(0.39)	0.21(0.34)/ 0.27(0.42)	0.25(0.42)/ 0.31(0.43)	0.20(0.26)/ 0.22(0.29)
No. atoms							
Protein	1259			1197	4844	3146	1252
Ligand/ion	0			0	0	25	0
Water	76			0	70	20	132
B-factors							
Protein	76.7			110.9	70.6	106.7	53.4
Ligand/ion						145.2	
Water	103.8				54.6	68.2	67.0
R.m.s deviations							
Bond lengths (Å)	0.01			0.01	0.01	0.01	0.01
Bond angles (°)	1.35			1.31	1.48	1.21	1.35
MolProbity score	3.03[44th][b]			3.33[52nd] [b]	2.49[78th] [b]	3.03[53rd][b]	2.80[37th][b]
Ramachandran							
% preferred	90.3			86.7	92.6	90.4	92.9
% allowed	8.4			11.3	4.2	7.8	6.4
% disallowed	1.3			2.0	3.2	1.8	0.7
Clashscore	18.23[81st]			16.4[96th]	9.81[97th]	11.69[97th]	9.78[94th]

[a]Highest resolution bin in parentheses here and other rows.
[b]Percentile in brackets here and other rows.

Supplementary Table S2. Ionic Interaction Pair Occupancy in C4BPα2 for Quadrilateral Residues of M2 (%)

C4BP-M2	M2 Wild-type	M2 (K65A)	M2 (N66D)	M2 (N66A)	M2 (K65A/N66A)
H67-D62	64.3/94.4[a]	96.5/100	93.4/88.6	75.8/77.5	99.0/98.4
R64-E68	73.6/44.2	33.8/41.8	50.1/73.0	61.2/35.8	56.2/28.0
R66-N/D/A66	5.70/4.70	59.0/44.4	45.7/26.0	—	—

[a]The two percentages reflect the two binding sites in the 2:2 C4BP-M protein complex.

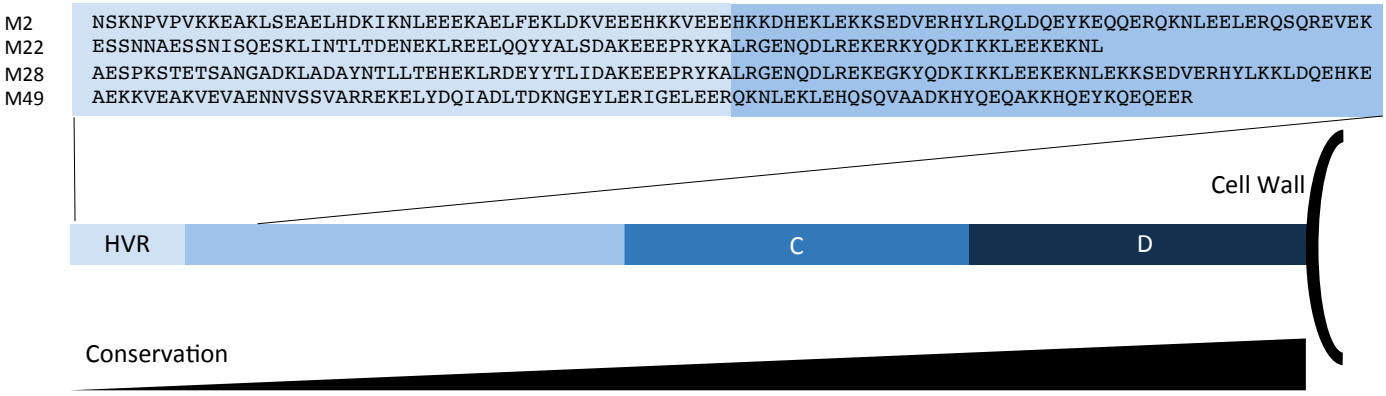
Supplementary Table S3. Ionic Interaction Pair Occupancy in Quadrilateral for residues of M49, M22, and M28 (%)

C4BP-M49		C4BP-M22		C4BP-M28	
H67-D69	75.3/90.4[a]	H67-N60	100/100	H67-N63	100/100
		R64-E65	71.4/55.7	R64-E68	52.4/30.2
R66-D73	54.1/43.3	R66-D64	95.6/96.0	R66-E70	21.1/28.8

[a]The two percentages reflect the two binding sites in the 2:2 C4BP-M protein complex.

Supplementary Figure S1

a.

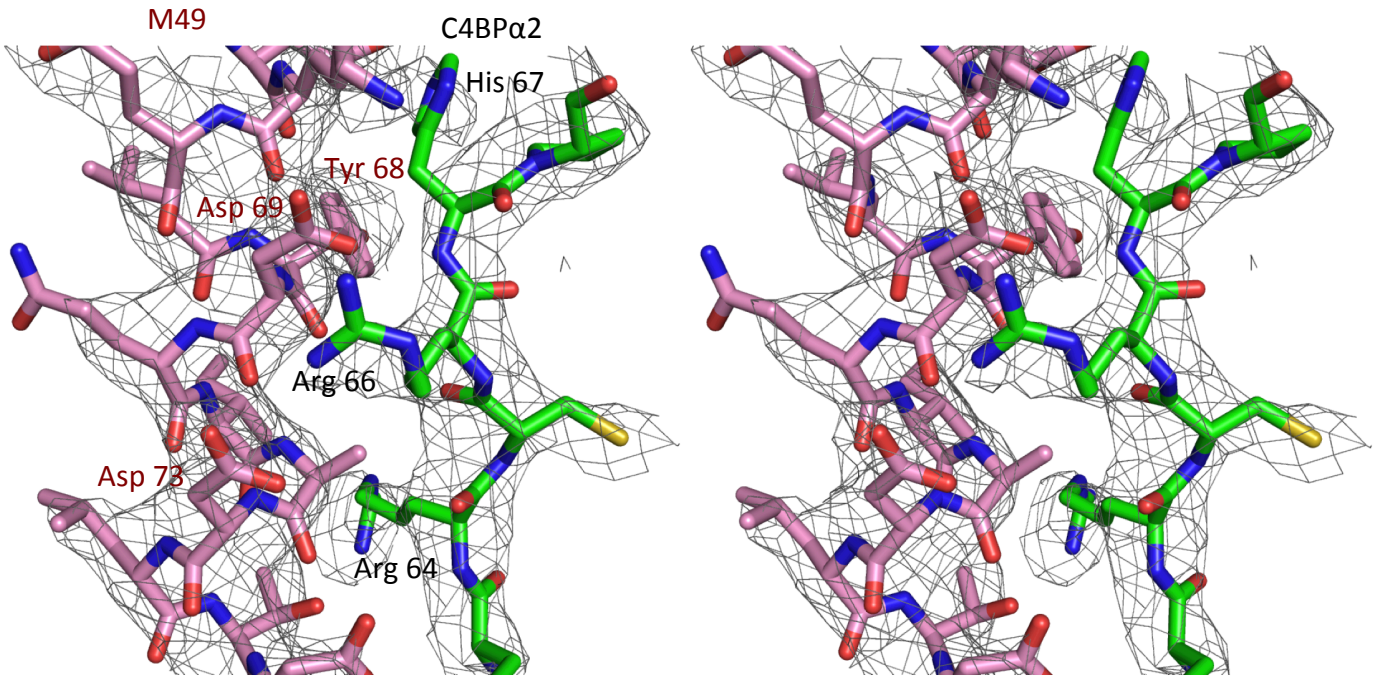


b.

```
M2          -----NSKNPVPVKKEAKLSEA--ELHDKIKNLEEEKAELFEKLDKVEEEHKKVEEE
M22         -----ESSNNAESSNISQESKLINTLTDENEKLREELQQYYALSDAKEEEPRYKA--
M28         AES--PKSTETSANGADKLADAYNTLLTEHEKLRD---EYYTLIDAKEEEPRYKA--
M49         AEKKVEAKVEVAENNVSSVARREKELYDQIADLTDKNGEYLERIGELEER-------
                  .  .  .  . :     *  :  .* :   :     .  **
```

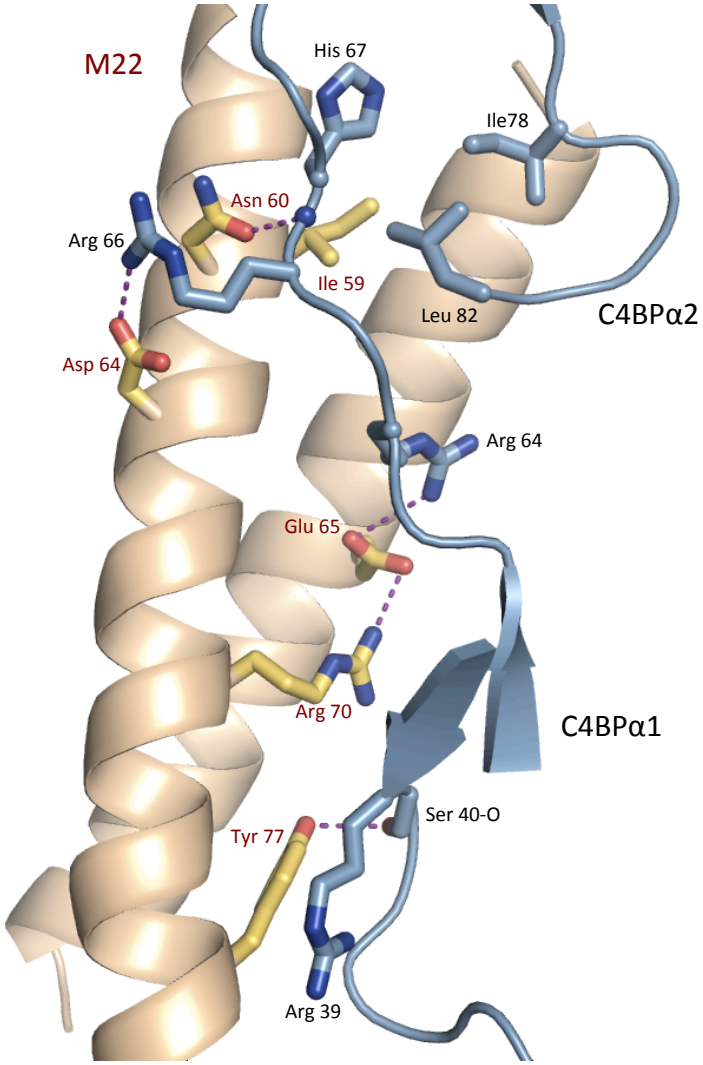
Supplementary Figure S1. Schematic of M protein domains.
a. The sequences of M2, M22, M28, and M49 co-crystallized with C4BPα1-2 are depicted, with the HVR in light blue and other regions of the protein in darker blue or black. M proteins are hypervariable at their N-termini, with conservation increasing towards their C-termini.
b. Multiple sequence alignment of the M2, M22, M28, and M49 HVRs is shown, as carried out using MUSCLE[1]. The asterisks indicate identical amino acids, the colons indicate amino acids with strongly similar properties, and the periods indicate amino acids with weakly similar properties.

Supplementary Figure S2



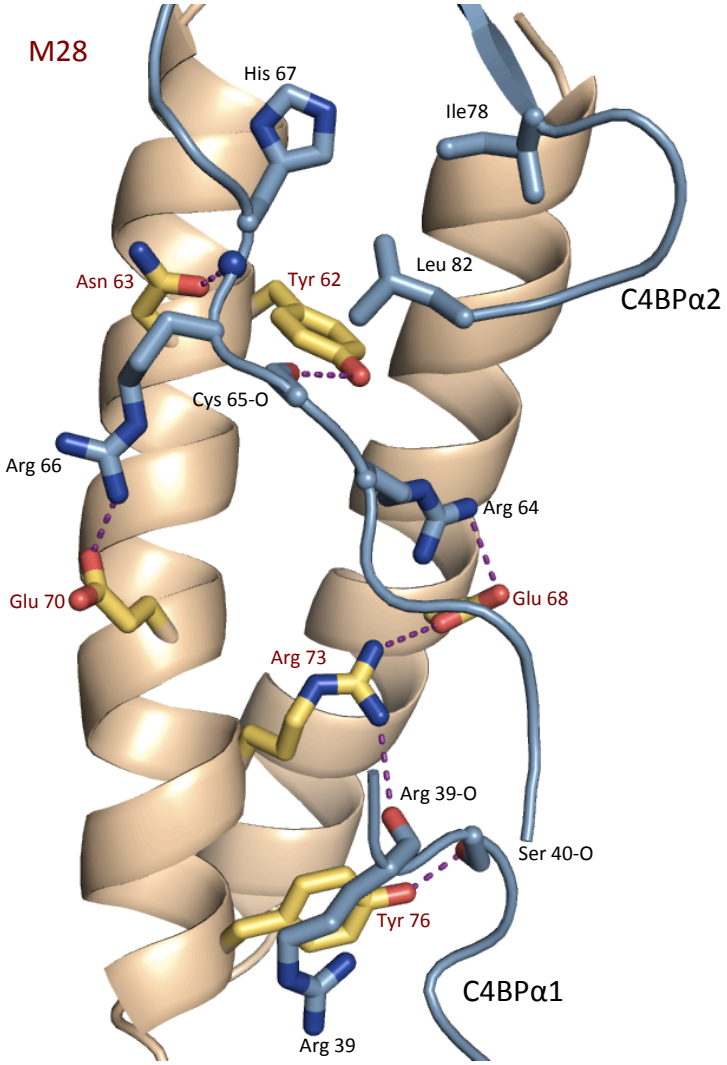
Supplementary Figure S2. Electron density for the M49 HVR-C4BPα1-2 complex.
Electron density from a simulated annealing 2Fo-Fc omit map (contoured at 1σ) for the M49 HVR-C4BPα1-2 structure. Residues M49 68-75 and C4BPα1-2 64-67 and 77-82 were excluded along with a 3.5 Å sphere around these residues, prior to performing simulated annealing (2500 K) and phase calculation. The final model is overlaid in bonds representation. M49 is depicted with pink carbons, and C4BP with green carbons; nitrogens and oxygens are blue and red, respectively, for both. The numbering of M proteins is such that the initiator Met is residue 1.

Supplementary Figure S3



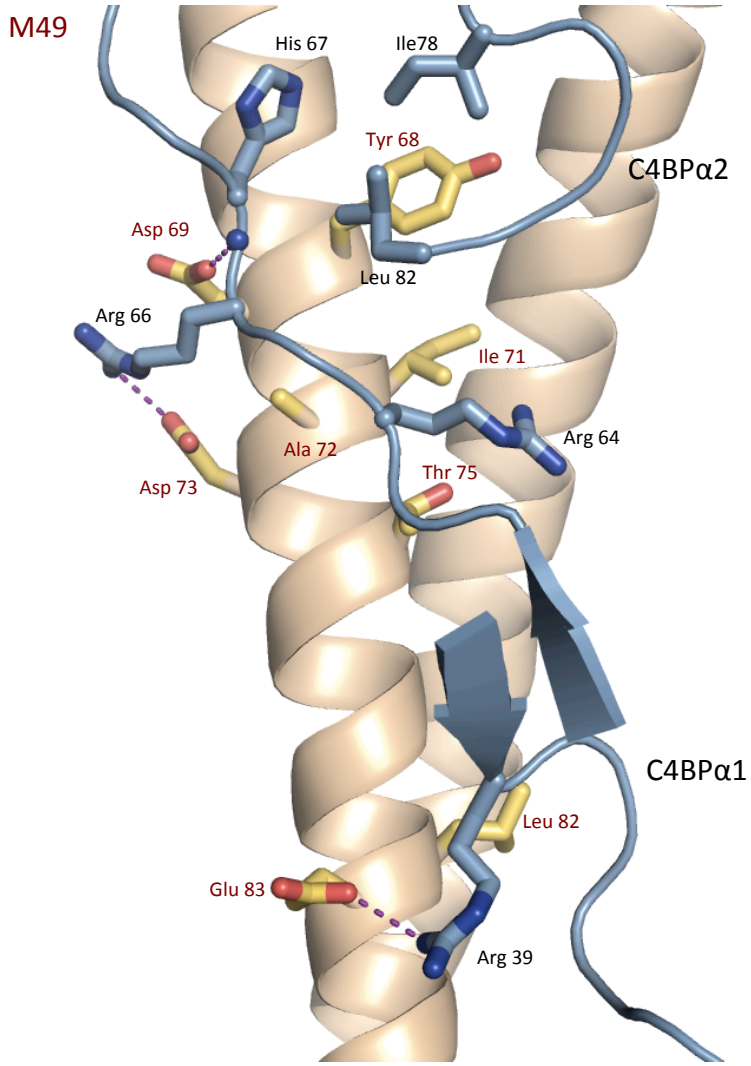
Supplementary Figure S3. Structure of M22-C4BP.
M2 is in gray ribbon representation with key side chains in bonds representation, in which carbons are yellow, oxygens red, and nitrogens blue. C4BPα1-2 is in cyan ribbon representation, with key side chains in bonds representation, in which carbons are cyan, oxygens red, and nitrogens blue. Hydrogen bonds and salt bridges depicted by dashed magenta lines.

Supplementary Figure S4



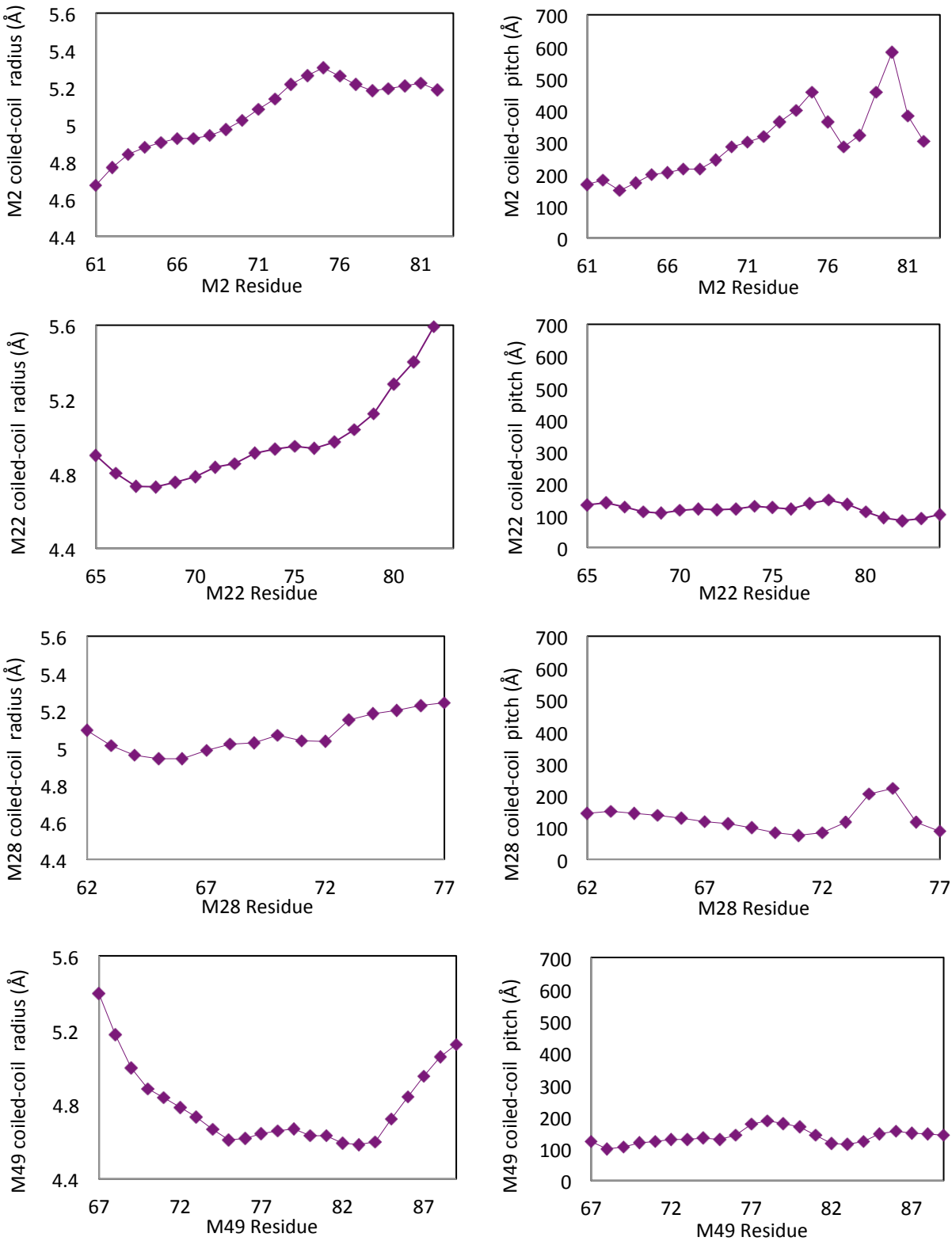
Supplementary Figure S4. Structure of M28-C4BP.
The depiction is as in Supplementary Figure S3.

Supplementary Figure S5



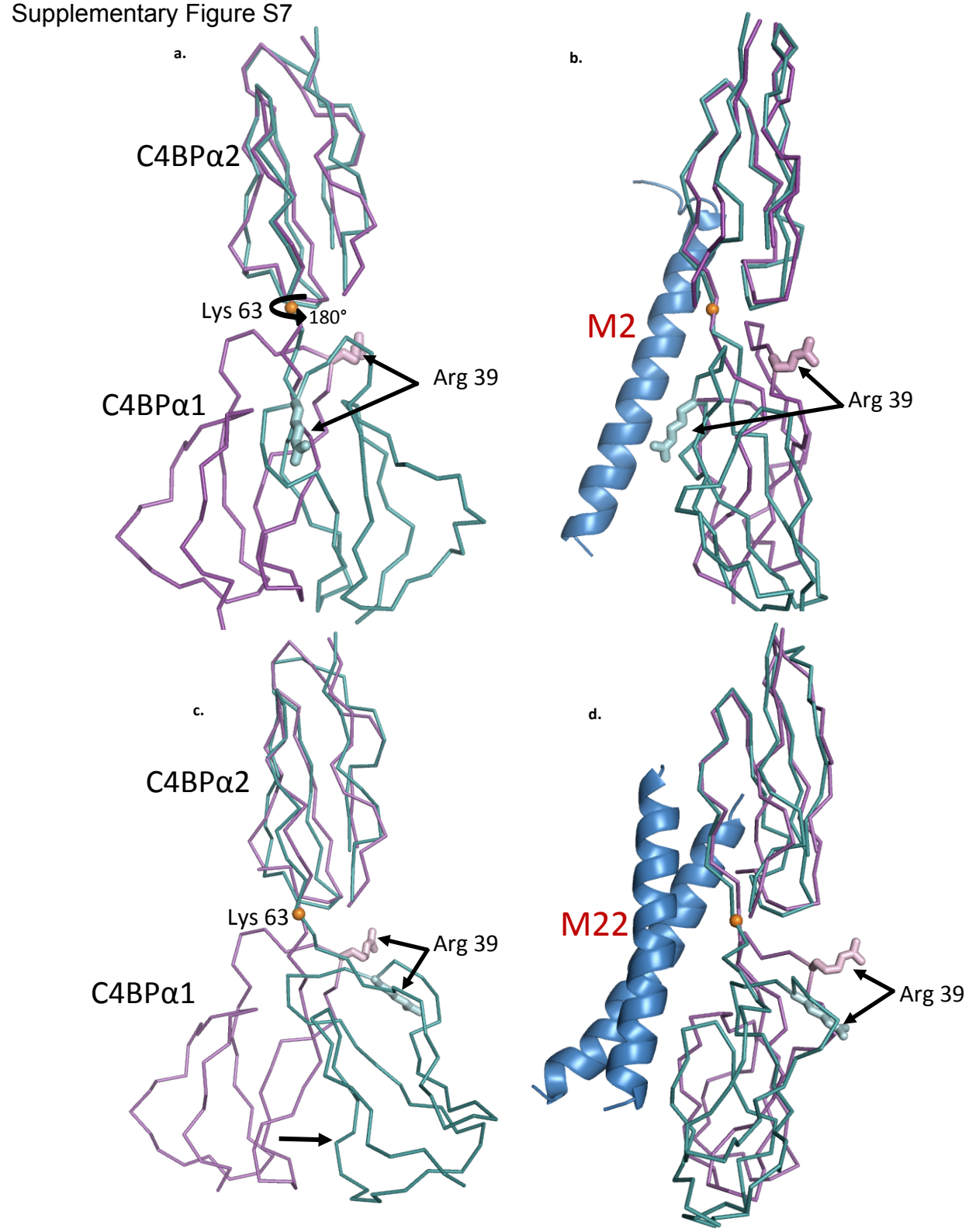
Supplementary Figure S5. Structure of M49-C4BP.
The depiction is as in Supplementary Figure S3.

Supplementary Figure S6



Supplementary Figure S6. Coiled coil parameters of M proteins.
Radius and pitch of the M protein HVR α-helical coiled coils, reporting on the residues that are at the interface with C4BPα1-2.

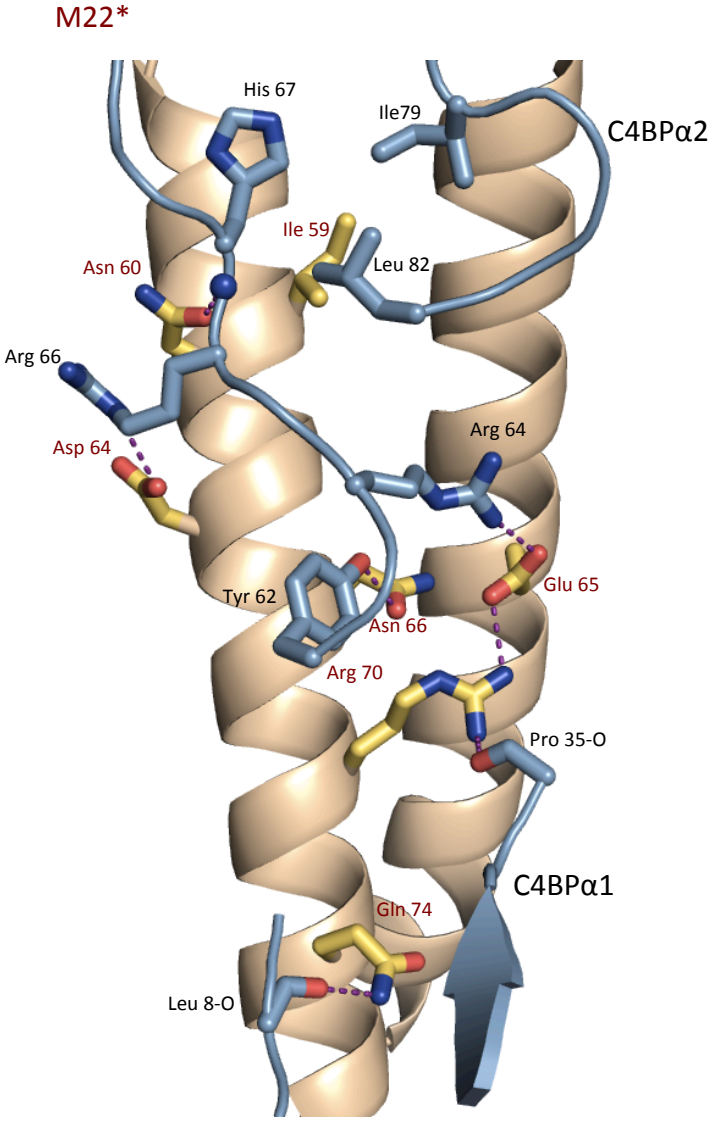
Supplementary Figure S7



Supplementary Figure S7. Rotation of C4BPα1-2.
a. Superposition of free (magenta) and M protein-bound C4BPα1-2 (cyan) based on the C4BPα2 domain, showing the 180° rotation of C4BPα1 around Lys 63 (indicated

Supplementary Figure S7 Continued.
by golden ball). The molecules are depicted as Cα chain traces. The position of C4BP Arg 39 is shown in bonds representation. The same depiction is used in the following panels.
b. A 90° rotated view of the superposition shown in panel a, with one α-helix of the M2 coiled coil shown as a blue ribbon.
c. One of the C4BPα1-2 molecules bound to M22 that is restricted from undergoing a 180° rotation by a crystal contact is depicted. Superposition of free (magenta) and the M22 protein-bound C4BPα1-2 (cyan) is based on the C4BPα2 domain. The bound C4BPα1 domain is related to free C4BPα1-2 by a tilt rather than a rotation.
d. A 90° rotated view of the superposition shown in panel c, with the M22 coiled coil shown as a blue ribbon.

Supplementary Figure S8



Supplementary Figure S8. Structure of the M22-C4BP interaction in which C4BPα1 is tilted rather than rotated.
The depiction is as in Supplementary Figure S3. M22 is denoted with an asterisk to indicate that in this complex, the C4BPα1 domain is tilted rather than rotated.

Supplementary Figure S9

```
          abcdefgabcdefgabcdefgabcdefgabcdefg
M2        KKEAKLSEAELHDKIKNLEEEKAELFEKLDKVEEE
M49       VSSVARREKELYDQIADLTDKNGEYLERIGELEER
M114      KEATKLSEAELYNKIQELEEGKAELFDKLEKVEEE
M73       KEAKKLNEAELYNKIQELEEGKAELFDKLEKVEEE
M77       EGVSVGSDASLHNRITDLEEEREKLLNKLDKVEEE
M84       ASVKKNNEEELHNKIADLLDQNEEYLNKIDELKEG
M89       SVSVKDNEKELHNKIADLEEERGEHLDKIDELKEE
M97       GPVPRSLWLREYDKNQELTKKLTEFEEKLLQN
M102.1    SSVPVKKAAELYDKIKELEEGREELLNDLDKVKED
M106      QKQNVSSNGRIYEIYDELQTKYDELQTKHEELLGE
M112      SSVSVKNEVKLHNEIAALQEEKEKLLNELDVKEEH
M118      ADSNASSVAKLYNQIADLTDKNGEYLERIEELEER
M124      ATKSKLSEAELHDKIKNLEEEKAELFEKLDKVEEE
```

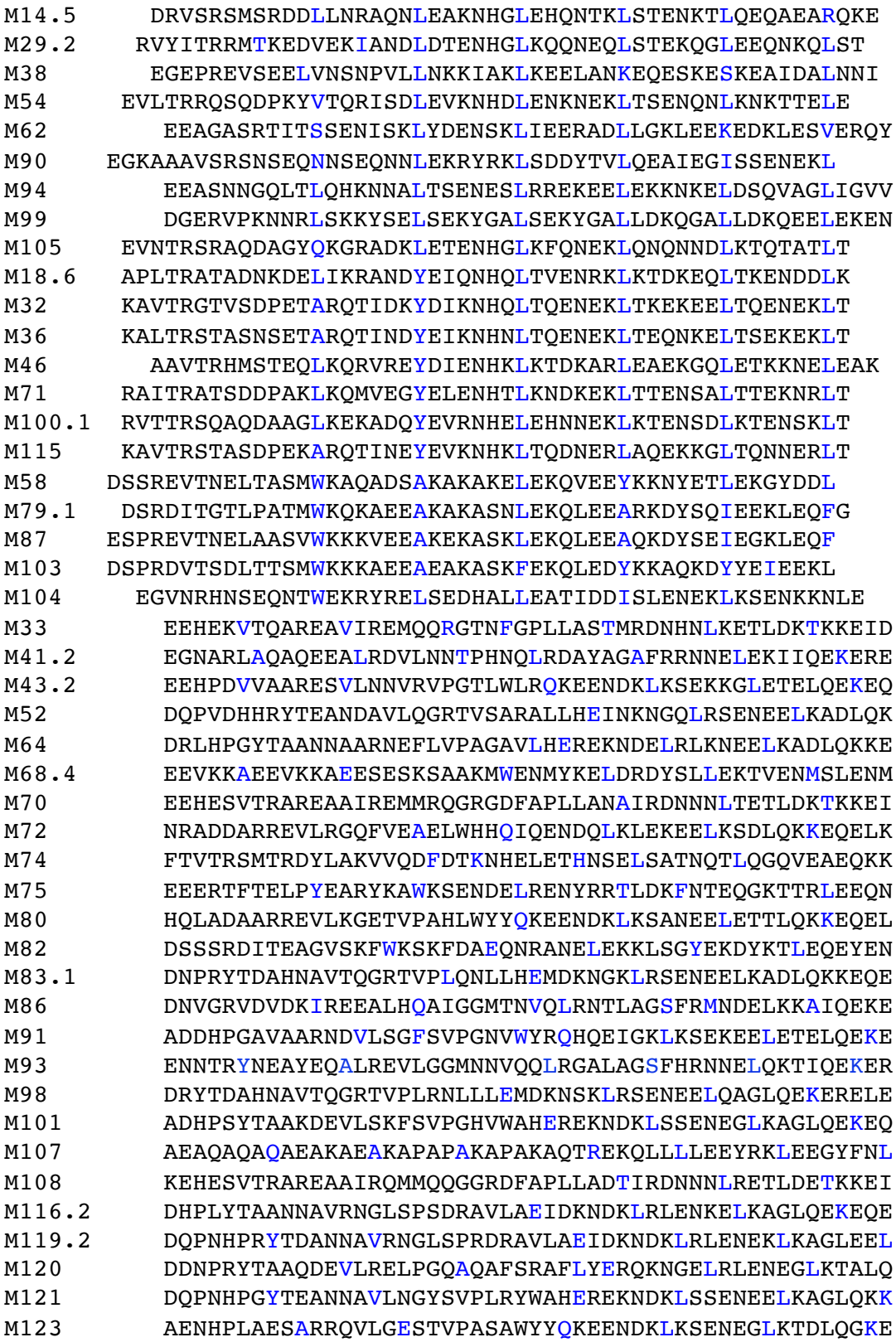
Supplementary Figure S9. Sequence alignment of C4BP-binding M protein HVRs of the M2/M49 pattern.
Residues that contact (first two lines) or are predicted to contact C4BP are in red, and the heptad register is indicated above. Residues observed (first two lines) or predicted to be at core *d* positions of the heptad register are highlighted in blue for visual reference.

Supplementary Figure S10

```
          Abcdefgabcdefgabcdefgabcdefgabcdefg
M22       ISQESKLINTLTDENEKLREELQQYYALSDAKEEE
M28       ADKLADAYNTLLTEHEKLRDEYYTLIDAKEEEPRY
M4        AWNWPKEYNALLKENEELKVEREKYLSYADDKEKD
M4.1      AWNWPKEYNALLKENEEFKVEREKYLSYADDKEKD
PrtH      NAKLVEVVETTSLENEKLKSENEENKKNLDKLSKD
M8        NEQLINELNNLIEENNDLKDKLARNLDLLDNTREK
M9.2      LSVPKTEYDKLYDDYDKLQEKSAEYLERIGELEER
M11       TNVSADLYNSLWDENKTLREKQEEYITKIQNEETK
M15.1     WKLTIEEYNKLLDENEKLKEKNEEYLEKIGEQEER
M25       AKAAEAKVDKLEKQLEGYKKLEEDYFNLEKR
M42       KVKLEVLYNSLWEENKTLREEQEEYIAKIDKLDEK
M44       GSVSLELYDKLSDENDILREKQDEYLTKIDGLDKE
M48.1     ELPPEARYKAWKSENDELRENHRKILDKFNAEQNK
M55       LYQERQRLQDLKSKFQDLKNRSEGYIQQYYDEEKN
M59       ELTLQQKYDALTNENKSLRRERDNYLNYLYEKEEL
M60       ISKERELINTLVDENNKLMEERARHLDLIDNI
M61       GSVSLELYDKLSDENDILREKQDEYLTKIDGLDKE
M63.4     KLTLEHKYNALTNENKSLIRREKDKYLYEKEELEK
M66       QNTWEKRYQKLSDDHTLLQDAIEEISSENEKLKSE
M67       GGVRLDLYDKLSKENDILREKQDEYLTKIDELGEK
M69.1     GGVSLDLYSKLLNENDILRDKQDDYLTKIDELTEK
M76       SNVSINLYNELQAEHDKLQTKHEELLAEHDALKEK
M78       SITNEQLIDKLVEENNDLKEERAKYLDLLDNREKD
M81       ENVPKQQYNALWEENEDLRGRERKYIAKLEKEEIQ
M85       TSVSADLYNSLWDENKTLREKQGEYITKIQNEETK
M88.1     ISNNERLINELTDENNELKDKLARSLDLLDNTREK
M92       SGSVSTPYNNLLNEYDDLLAKHGELLSEYDALKEK
M96       LDQFGRDYDELQKKYDKLDKENKEYASQLGK
M109      ADNLAKEYNTLLTENEKLREELQQYYALIDAKEEE
M110      HEELWKEYDILKEKLDKDQEEREKIELNYLK
M111.1    VTAPAHFWENQRREIEKLKGEIDQLKLLLGKS
M117      GSVSIDRYNELSGEYNKLLDQNGNLLDENEILREK
```

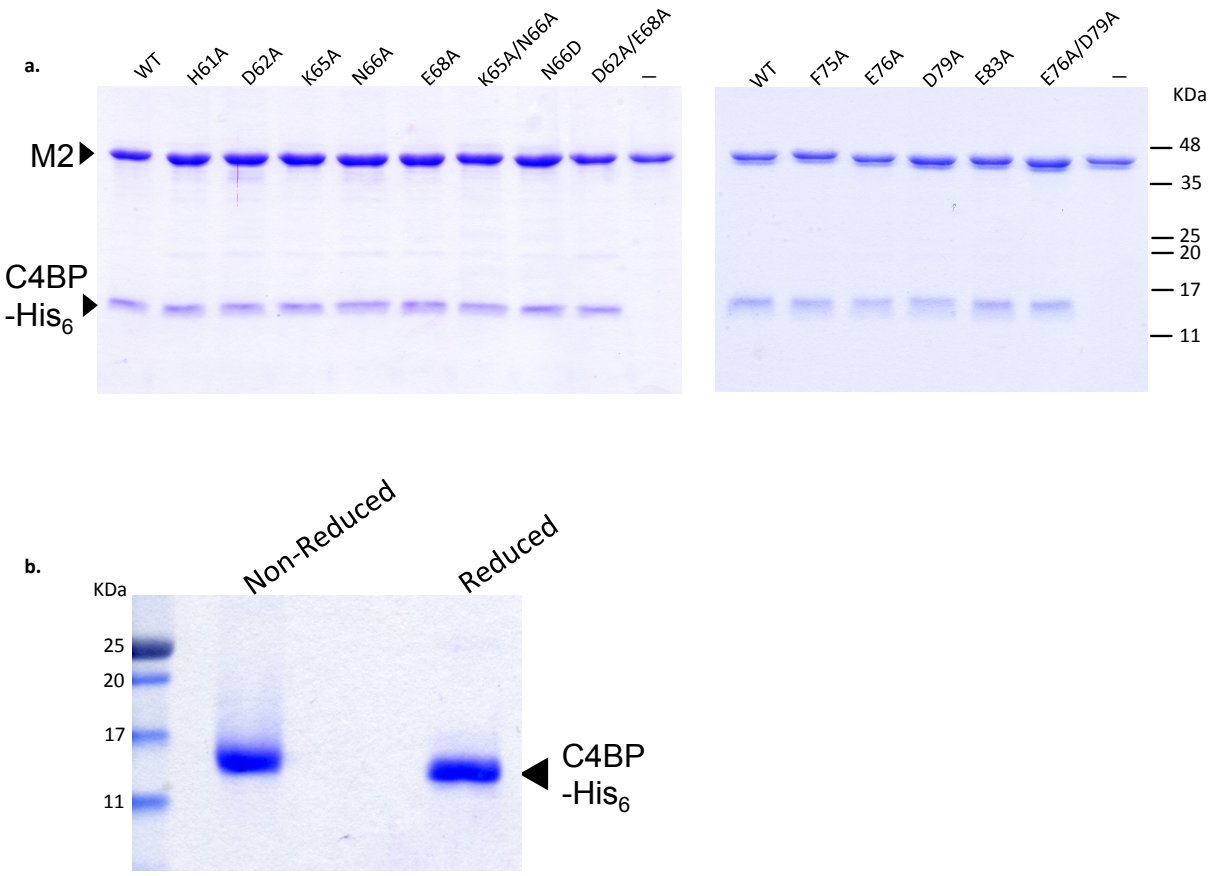
Supplementary Figure S10. Sequence alignment of C4BP-binding M protein HVRs of the M22/M28 pattern.
Residues that contact (first two lines) or are predicted to contact C4BP are in red, and the heptad register is indicated above. Residues observed (first two lines) or predicted to be at core *d* positions of the heptad register are highlighted in blue for visual reference.

Supplementary Figure S11



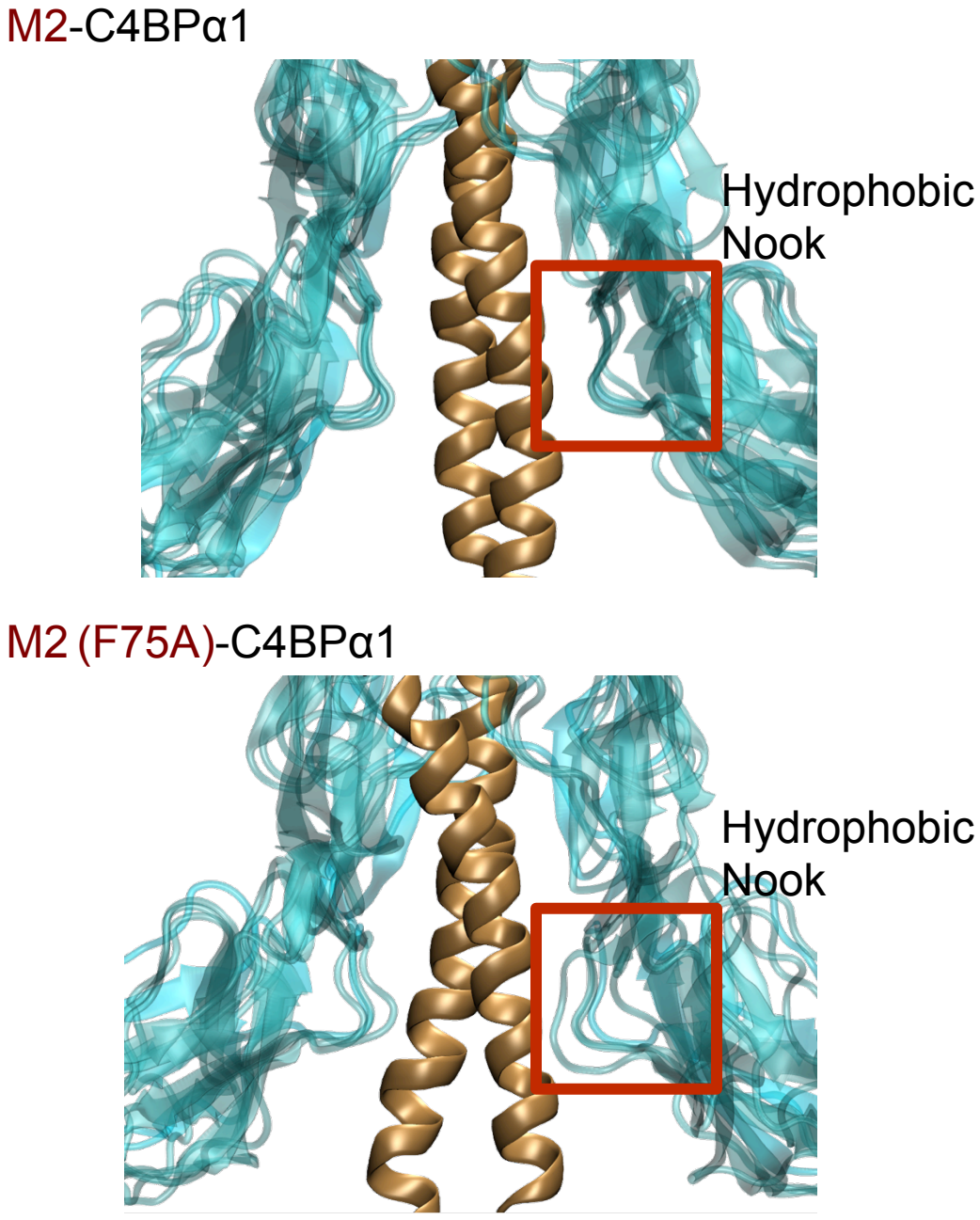
Supplementary Figure S11. C4BP-binding M protein HVRs that cannot be classified as belonging to either M2/M49 or M22/M28 patterns.
Residues predicted to be at core *d* positions of the heptad register are highlighted in blue.

Supplementary Figure S12



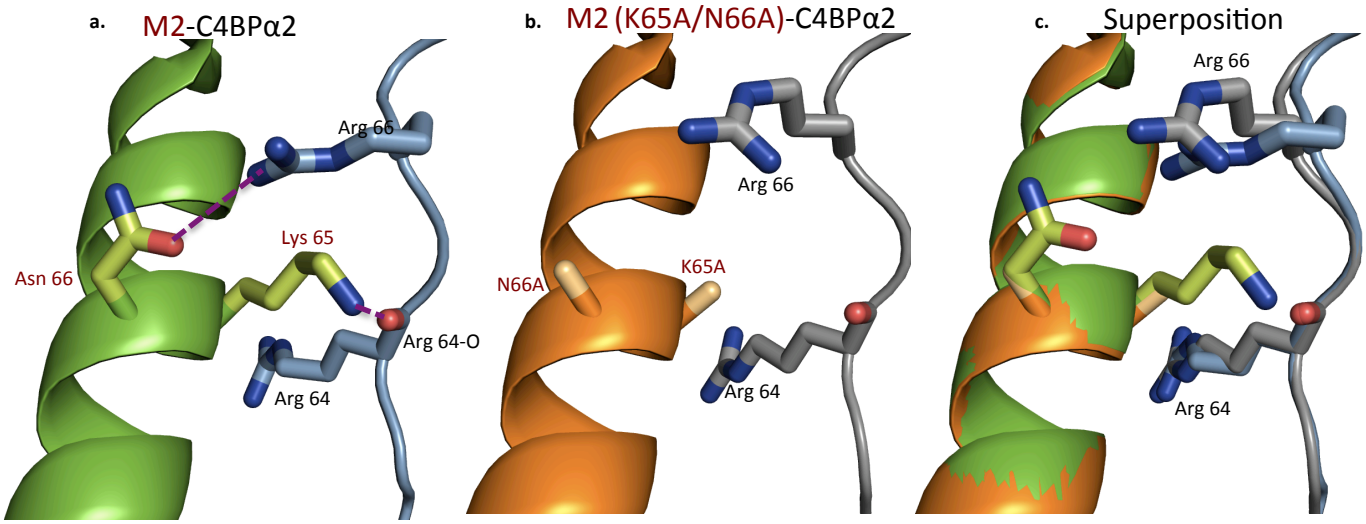
Supplementary Figure S12. C4BP-M2 interaction.
a. Input samples from the experiment shown in Figure 4a visualized by non-reducing, Coomassie-stained SDS-PAGE. Molecular mass markers were not run on these particular gels; their positions are based on measurements from equivalent gels.
b. C4BPα1-2 visualized on Coomassie-stained SDS-PAGE under non-reducing and reducing conditions, showing that C4BPα1-2 used in these experiments is intact and not nicked. Molecular mass markers are in the leftmost lane.

Supplementary Figure S13



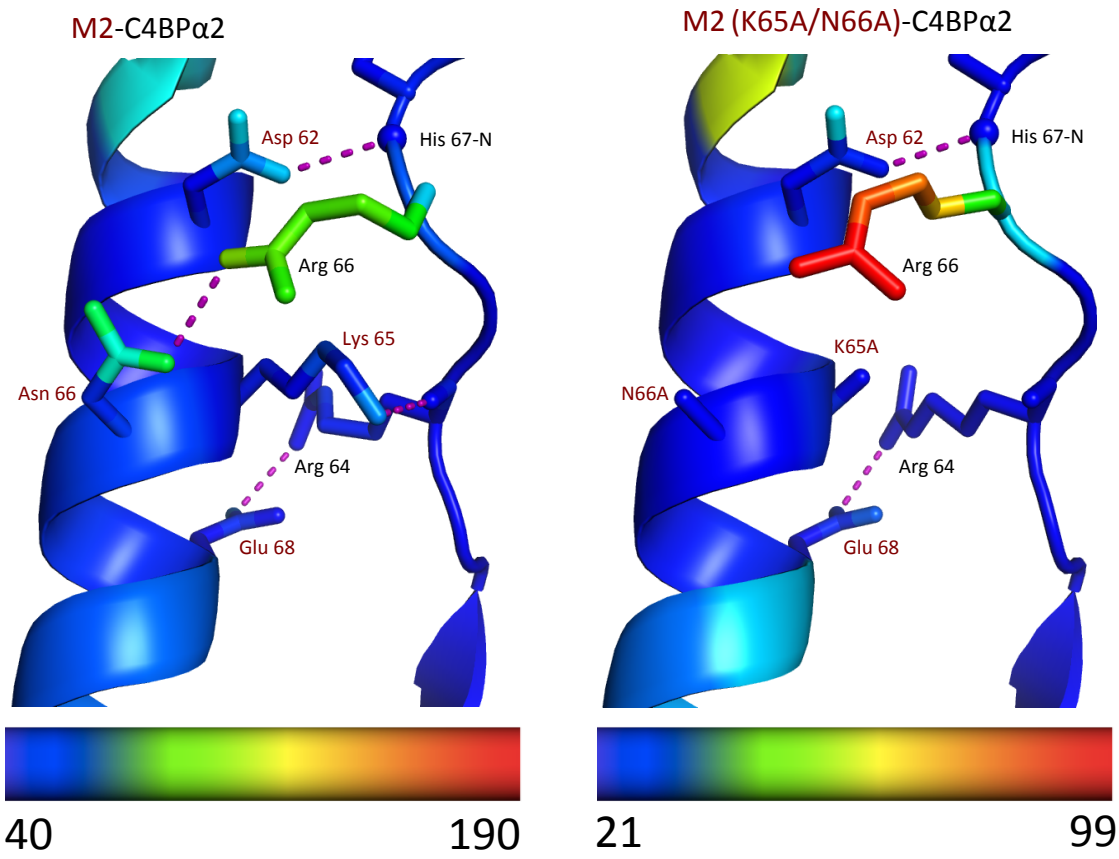
Supplementary Figure S13. Molecular dynamics simulation of the Arg39 'hydrophobic nook' interaction with wild-type M2 and M2 F75A.
Five overlaid replicates of simulated interactions between C4BPα1-2 (cyan, ribbon representation) and wild-type M2 (upper panel) or M2 F75A (lower panel); C4BPα1-2 is in cyan ribbon representation, and M2 in gold ribbon representation. The simulated coordinates of the five replicates were aligned on M2 protein. The images highlight the hydrophobic C4BP α1 Arg39 'hydrophobic nook'.

Supplementary Figure S14



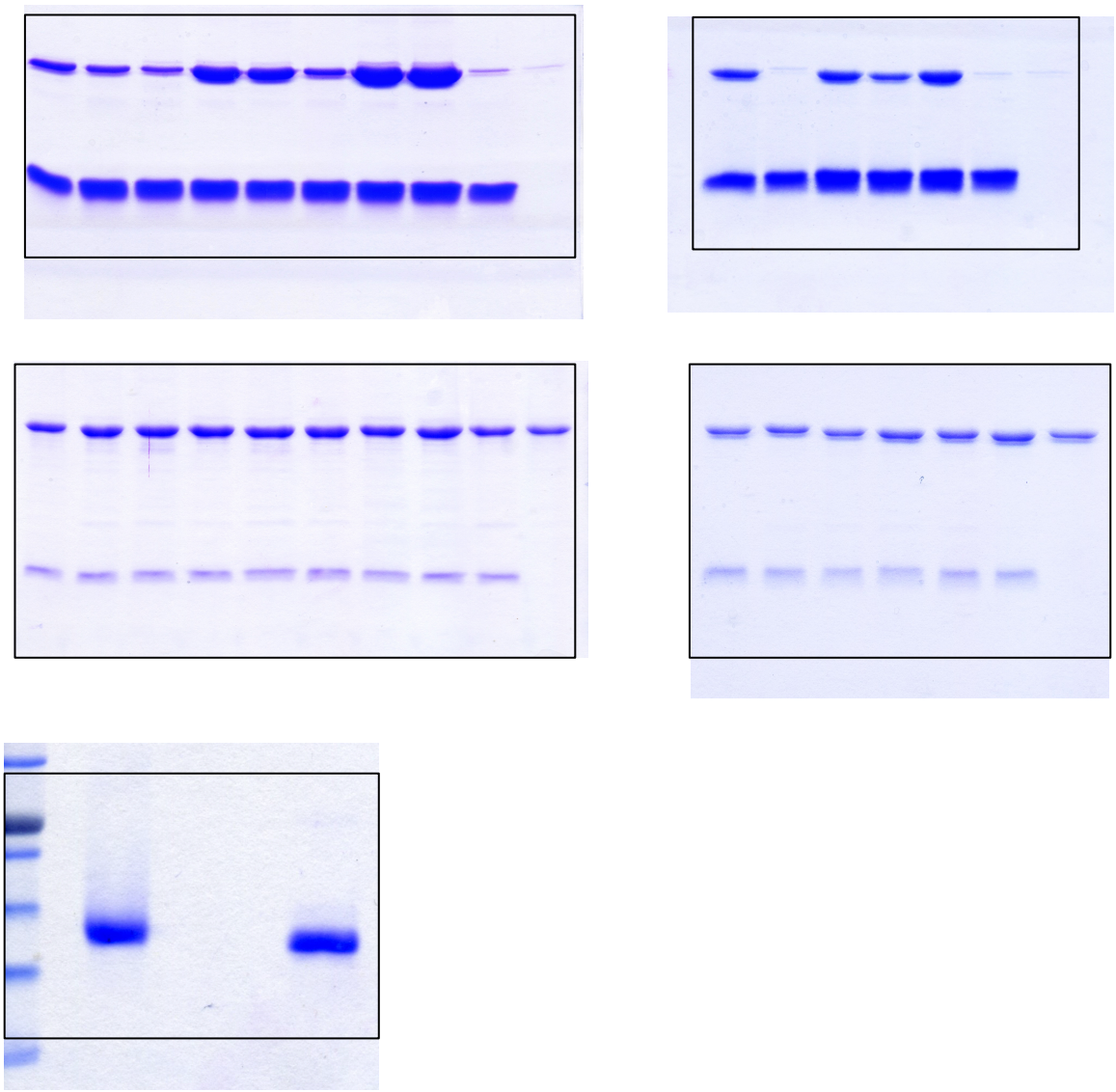
Supplementary Figure S14. Interactions of M2 and M2 (K65A/N66A) with C4BPα2.
a. M2 is depicted as a green ribbon, with Lys 65 and Asn 66 in bonds representation (carbons are green, oxygens red, and nitrogens blue). C4BPα2 is in cyan ribbon representation, with Arg 64 and its main chain carbonyl and Arg 66 in bonds representation (carbons are cyan, oxygens red, and nitrogens blue). Hydrogen bonds are depicted as red dashed lines.
b. M2 (K65A/N66A) is depicted as a gold ribbon, with Ala 65 and Ala 66 in bonds representation (carbons are light gold). C4BPα2 is in gray ribbon representation, with the same groups as in panel a shown.
c. Superposition of the structures shown in panels a and b.

Supplementary Figure S15



Supplementary Figure S15. B-factors of C4BPα2 bound to M2 or M2 (K65A/N66A).
B-factors in the M2-C4BPα1-2 (left) and M2 (K65A/N66A)-C4BPα1-2 (right) structures represented by a color spectrum. The coloring of the M2-C4BPα1-2 complex is from the lowest to highest B-factor in the structure. The scale for the M2 (K65A/N66A)-C4BPα1-2 complex was set equivalent to that of the M2-C4BPα1-2 complex, but adjusted to account for the difference in Wilson B-factors between these two structures. This was done by dividing the B-factors of the M2-C4BPα1-2 structure by the ~2-fold lower Wilson B-factor of the M2 (K65A, N66A)-C4BPα1-2 structure (30.9 vs 59.7 $Å^2$).

Supplementary Figure S16



Supplementary Figure S16. Uncropped gels from Figures 4a (top two) and Supplementary Figure S12 (bottom three).

Supplementary Video Legends

Supplementary Video S1. R39 nook in M2-C4BPα1-2.

Portion of the molecular dynamics simulation of the M2-C4BPα1-2 complex focusing on the R39 'hydrophobic nook'. M2 is in gold ribbon representation, and C4BP in cyan ribbon representation. Individual side chains are shown in bonds representation, with carbons cyans, nitrogens blue, and oxygens red. Water molecules are shown as purple spheres. Of particular note is the lack of water molecules at the M2-C4BPα1 interface, and the stability of the loop on which C4BP R39 is located.

Supplementary Video S2. R39 nook in M2 (F75A)-C4BPα1-2.

Portion of the molecular dynamics simulation of the M2 (F75A)-C4BPα1-2 complex. The depiction is the same as in Supplementary Video S1. Of particular note is the increased number of water molecules at the M2 (F75A)-C4BPα1 interface, and the instability of the loop on which C4BP R39 is located.

Supplementary Video S3. C4BPα2 contacts in M2-C4BPα1-2.

Portion of the molecular dynamics simulation of the M2-C4BPα1-2 complex. The depiction is the same as in Supplementary Video S1. Of particular note is the infrequent interaction between C4BPα2 R66 and M2 N66.

Supplementary Video S4. C4BPα2 contacts in M2 (K65A)-C4BPα1-2. Portion of the molecular dynamics simulation of the M2 (K65A)-C4BPα1-2 complex. The depiction is the same as in Supplementary Video S1. Of particular note is the increased frequency of interaction between C4BPα2 R66 and M2 N66.

Supplementary Video S5. C4BPα2 contacts in M2 (N66D)-C4BPα1-2. Portion of the molecular dynamics simulation of the M2 (N66D)-C4BPα1-2 complex. The depiction is the same as in Supplementary Video S1. Of particular note is the increased frequency of interaction between C4BPα2 R66 and M2 N66D.

Supplementary References

1. Edgar, R. C. MUSCLE: multiple sequence alignment with high accuracy and high throughput. *Nucleic Acids Res* **32**, 1792-1797, (2004).