DNA manipulation

The coding sequences of mature M2 (amino acids 42-367), M22 (42-335), M28 (42-363), and M49 (42-359) proteins were cloned from GAS strains M2 (AP2), M22 (Sir22), M28 (strain 4039-05), and M49 (NZ131), respectively, into a modified version of the pET28a vector (Novagen) modified such that it encoded an N-terminal His_6-tag followed by a PreScission^{TM} protease (GE Healthcare) cleavage site. Constructs that encoded truncated versions of these proteins, which consisted of only the N-terminal 79, 86 or 100 amino acids, were generated through the insertion of an amber stop codon at an appropriate site by site-directed mutagenesis. Site-specific mutations were also introduced into the M2 coding sequence by site-directed mutagenesis. All site-directed mutagenesis was performed according to the Agilent QuickChange^{TM} manual, except that 50 μ L reactions were set up for polymerase chain reactions (PCR) instead of 12.5 μ L reactions.

The coding sequence of the CCP1-2 domains of human C4BPα chain (C4BPα 1-2)[1] (a kind gift from G. Lindahl) was cloned into the modified pET28a vector described above, and also into a pET28b vector that encoded a non-cleavable C-terminal His_6-tag. The cleavable N-terminal His_6-tag version of C4BPα 1-2 was used for crystallographic studies, and the non-cleavable Cterminal His_6-tagged version for co-precipitation binding studies. To obtain selenomethionine (SeMet)-substituted protein to be used in phase determination, methionines were introduced in the coding sequence of C4BPα 1-2 at amino acid positions 29, 46, and/or 71 by site-directed mutagenesis.

Protein Expression and Purification

M proteins were expressed in *Escherichia coli* BL21 (DE3) and purified as described[2] with minor modifications to the procedure. Specifically, bacteria were lysed with a C-5 Emulsiflex (Avestin Inc., Ottawa, Canada) and ion exchange chromatography was omitted, and in the case of purification of M2 (wild-type and variants), imidazole was not included in the lysis and wash buffers.
C4BPα 1-2 was expressed in *E. coli* Rosetta 2 (Novagen) cells. The protein was purified and refolded as described[3], except for the use of a C5 Emulsiflex for lysis. Where needed, the N-terminal His_6-tags of M proteins and C4BPα 1-2 were removed by PreScission^{TM} protease cleavage according to manufacturer's instructions, and the cleaved protein was purified by reverse Ni^{2+}-NTA chromatography. M proteins and C4BPα 1-2 were lastly purified by size-exclusion chromatography (Superdex 200) in a buffer composed of 150 mM NaCl, 50 mM Tris, pH 8.5. Proteins were then concentrated to $\sim$ 20 mg/mL by ultrafiltration; protein concentrations were determined by absorbance at 280 nm using calculated molar extinction coefficients. Aliquots of concentrated protein were flash-frozen in liquid N_2 and stored at -80 $^\circ$ C.

SeMet was incorporated into C4BPα 1-2 (L29M/L46M), C4BPα 1-2 (L29M/L71M), and C4BPα 1-2 (L46M/L71M) using methionine pathway inhibition as described[4]. SeMet-labeled C4BPα 1-2 was purified as described above.

Crystallization and Data Collection

For preparation of complexes, M2 (amino acids 42-141), M2 (K65A/N66A) (42-141), M22 (42-120), M28 (42-141), or M49 (42-127) protein was mixed with C4BPα 1-2 (wild-type or SeMet-substituted mutant) at a 1:1 molar ratio (final concentration of complex $\sim$ 5 mg/mL), and dialyzed overnight at 4 $^\circ$ C in 10 mM Tris, pH 8. The samples were then concentrated by ultrafiltration to $\sim$ 20 mg/mL. Crystallization was performed by the hanging drop vapor-diffusion method.

The M2-C4BPα 1-2, M2 (K65A/N66A)-C4BPα 1-2, and M28-C4BPα 1-2 complexes and the SeMet-labeled M2-C4BPα 1-2 (L29M/L46M) and M2-C4BPα 1-2 (L46M/L71M) complexes were co-crystallized at 20° C by mixing 1 μ L of complex with 1 μ L of the reservoir solution, which was 1.5 M $(NH_4)_2SO_4$, 0.1 M Bis-Tris Propane, pH 7.0. These crystals were transferred to the reservoir solution supplemented with 20% ethylene glycol for cryopreservation, mounted in fiber loops, and flash-cooled in liquid N_2

. Crystals containing SeMet-labeled protein were treated similarly, except the reservoir solution was supplemented with freshly prepared 1 mM TCEP.
The M22-C4BPα 1-2 complex was co-crystallized similarly, except the reservoir solution was 2 M $(NH_4)_2SO_4$, 2% PEG 400, and HEPES pH 7.5. The SeMet-labeled M49-C4BPα 1-2 L29M/L46M complex was also co-crystallized similarly, except the reservoir solution was 1.6 M Na/K PO_4, pH 6.9. These two co-crystals were transferred to their respective reservoir solutions supplemented with 20% glycerol before being flash-cooled in liquid N_2

.

Diffraction data were collected from crystals under cryogenic conditions. Diffraction data for M2-C4BPα 1-2 were collected at the Stanford Synchrotron Radiation Lightsource (SSRL) beamline 9-2, for M22-C4BPα 1-2 at the Advanced Photon Source (APS) beamline 24-ID-C, and for M2 (K65A/N66A)-C4BPα 1-2 and M28-C4BPα 1-2 at the Advanced Lightsource (ALS) beamline 8.2.1. Single-wavelength anomalous dispersion (SAD) data were collected from SeMet-labeled M2-C4BPα 1-2 (L29M/L46M) and M2-C4BPα 1-2 (L46M/L71M) at the APS beamline 19-ID, and from SeMet-labeled M49-C4BPα 1-2 (L29M/L46M) at the APS beamline 24-ID-E.

Diffraction data from crystals of M22-C4BPα 1-2 and M49-C4BPα 1-2 (L29M/L46M) were indexed, integrated, and scaled using XDS[5], while HKL2000[6] was used for data from all other crystals.

Structure Determination and Refinement

M2-C4BPα 1-2. For structure determination of M2-C4BPα 1-2, Se sites were located from SAD data of SeMet-labeled M2-C4BPα 1-2 (L29M/L46M) and M2-C4BPα 1-2 (L46M/L71M), and phases calculated for each data set using Autosol (within Phenix[7]). The two sets of phases were combined using the Reflection File Editor program (within Phenix). From the combined phase set, four Se sites, three at substituted methionines and one at the native Met 14, were identified per asymmetric unit, which contained one M2 α -helix and one C4BPα 1-2 molecule.

Here and in all cases below, model building was carried out with Coot[8] as guided by inspection of SAD-phased maps or σ_A-weighted $2mF_o - DF_c$ and $mF_o - DF_c$ maps, and refinement was carried out with Refine (within Phenix) using default parameters. Between 15 and 75 iterative cycles of building and refinement, with each refinement step consisting of 1-rounds, were performed in each case. In later stages of refinement, TLS parameterization was used in Refine. Individual B-factors were refined isotropically. Water molecules were added in the final stages of refinement using Phenix with default parameters (3σ peak height in σ_A-weighted $mF_o - DF_c$ maps).

In order to model M2-C4BPα 1-2 (L29M/L46M/L71M), the NMR structure of C4BPα 1-2 was manually fit into SAD-phased density, with the two domains of C4BPα 1-2 being treated as individual rigid bodies. The M2 molecule was then built into density, with the register of the coiled coil being assigned from well defined density corresponding to large side chains (i.e., His 20, Phe 75, and His 85). The SeMet residues in the model were changed to leucines, and the model was then refined

against the higher resolution (2.56Å resolution limit) data collected from crystals of M2-C4BPα 1-2. TLS parameterization involved the following groups: For M2, 53-57 and 58-86; for C4BPα 1-2, 0-59 and 60-124. Continuous electron density was evident for the entire main chain of C4BPα 1-2 and for residues 53-86 of the M2 protein. Most side chains of C4BPα 1-2 were also visible, except for some in long loopswere distant from the interface with M protein.

The structure of M2 (K65A/N66A)-C4BPα 1-2 was determined by difference Fourier synthesis using the refined structure of M2-C4BPα 1-2. The set of reflections used for R_{free} calculations for the refinement of M2-C4BPα 1-2 were maintained. TLS parameterization was equivalent to that for M2-C4BPα 1-2.

M28-C4BPα 1-2. The structure of M28-C4BPα 1-2 was determined by molecular replacement using the program Phaser (within Phenix[7]). The C4BPα 1-2 molecule from the structure of the M2-C4BPα 1-2 complex served as the search model. The molecular replacement solution had a log-likelihood gain score of 379. The asymmetric unit contained one C4BPα 1-2 molecule and one M28 α -helix, whose register was determined by well defined density corresponding to large side chains (i.e. Tyr 62, Tyr 76, Tyr 77). The model was first subjected to cycles of rigid body refinement, followed by the refinement protocol described above. TLS parameterization involved the following groups: For M28, 5-; for C4BPα 1-2, 0-, , and . Continuous electron density was evident for the entire main chain of C4BPα 1-2 amino acids -of M28.

M22-C4BPα 1-2. The structure of the M22-C4BPα 1-2 complex was determined by molecular replacement using the program Phaser. The search model consisted of an M28 α -helical, dimeric coiled-coil in complex with a single C4BPα 1-2 molecule. The solution, which had a log-likelihood gain score of 166, resulted in two copies of the search model in the asymmetric unit, while the solvent content suggested that the asymmetric unit was composed of two M22 α -helical, dimeric coiled-coils and four C4BPα 1-2 molecules; this latter composition was found to be accurate. After refinement of the initial molecular replacement model, two additional C4BPα 1-2 molecules became evident in electron density maps, and were placed stepwise into density, with the two domains of C4BPα 1-2 being treated as individual rigid bodies, between rounds of iterative refinement. Both these additional copies had similar conformations to one another, and had a tilted orientation of the C4BPα 1 and C4BPα 2 domains relative to these domains in unbound C4BPα 1-2. This tilted orientation differs from the 180° rotation observed in the two other copies of C4BPα 1-2 bound to M22, as well as in copies of C4BPα 1-2 bound to M2, M28, and M49. Side chains for M22 were subsequently built into density, with the register being assigned based on well defined density corresponding to large side chains (i.e., Tyr 66 and Tyr 67). The model was then subjected to cycles of rigid body refinement followed by the refinement procedures described above. TLS parameterization involved the following groups: For M22 chain A, 52-80; for M22 chain C, 52-79; for M22 chain E, 52-; for M22 chain G, 52-; for C4BPα 1-2 chain B, 1-13, 14-27, 28-59, 60-73, 74-86, 87-102, 103-109, 110-115, and 116-124; for C4BPα 1-2 chain D, 0-59 and 60-124; for C4BPα 1-2 chain F, 1-59 and 60-124; for C4BPα 1-2 chain H, 0-13, 14-33, 34-47, 48-59, 60-

74, 75-86, 87-109, and 110-124. Continuous electron density was evident for the entire main chain of C4BPα 1-2and for residues 52-79 (or 80, depending on the chain) of M22.

M49-C4BPα 1-2. For structure determination of M49-C4BPα 1-2, Se sites were located from SAD data collected for SeMet-labeled M49-C4BPα 1-2 (L29M/L46M), and phases calculated using the program Autosol. Six Se sites were identified per asymmetric unit, which was found to contain an M49 α -helical, coiled-coil dimer and two C4BPα 1-2 molecules. This is consistent with the total of two SeMet substitutions introduced into C4BPα 1-2. The crystal structure of C4BPα 1-2 from the M2-C4BPα 1-2 co-crystal structure was manually fit into SAD-phased density, with the two domains of C4BPα 1-2 being treated as individual rigid bodies. A model of the M49 protein was then built into density, with the amino acid register for the coiled coil being assigned based on well defined density corresponding to large side chains (i.e., His 20, Phe 75, and His 85). TLS parameterization involved the following groups: For M49 chain A, 56-60 and 61-126; for M49 chain C, 56-126; for C4BPα 1-2 chain B, 0-10, 11-62 and 63-124; for C4BPα 1-2 chain D, 0-13, 14-27, 28-33, 34-44, 45-53, 54-62, 63-73, 74-86, 87-102, and 103-124. Continuous electron density was evident for most of the main chain of C4BPα 1-2, except for some of the longer loops of the C4BPα 1 domain, and for amino acids 56-124 (or 126, depending on the chain) of M49.

Validation. Structural models were validated with MolProbity[9] (Table S1). Molecular figures were made with PyMol (http://pymol.sourceforge.net). Coordinates and structure factors have been deposited in the RSCB PDB (list of codes).

Co-Precipitation Assays

Forty μ g of C4BPα 1-2-His$_6$protein was mixed with 120μ g of intact M2 protein (wild-type or mutant) in 50μ L of phosphate buffered saline (PBS) at 37° C for 30min. Fifty μ L of Ni^{2+}-NTA agarose beads were equilibrated in PBS, then added to the protein mix in a 1:1 beads:PBS (100 μ L) slurry and incubated for 30min at 37° C under agitation. The beads were washed three times with 0.5 mL of PBS supplemented with 15 mM imidazole, and eluted with 40 μ L PBS supplemented with 500 mM imidazole. Proteins in the input and eluted fractions were resolved by non-reducing SDS-PAGE and visualized by Coomassie-staining.

Molecular Dynamics

System Preparation. Heavy atom coordinates were taken from the co-crystal structures of M protein-C4BPα 1-2 complexes. Structures of complexes containing M2 substitution mutants were created by computational point mutations at the desired amino acid(s). Due to the varying resolutions of crystal structures, crystallographic waters were removed prior to solvating the system. Each structure was prepared for simulation using the Amber14SB force field[10,11,12]. The ionization states of titratable residues at pH 7 were predicted using PROPKA 3.1[13,14] and visually inspected

to ensure the accuracy of assigned states. Free cysteine residues were converted to disulfide-bonded pairs manually and built using tLeap, a system preparation program from the Amber Tools 2015 package[12]. The C-termini of proteins were capped to remove charges. The solvent was modeled explicitly using the TIP4P water model[15] and a 0.15 M NaCl concentration was applied after neutralizing the overall charge of the protein complexes. The Particle Mesh Ewald electrostatic summation method[16,17] was employed to evaluate electrostatics during simulation. In total, eight different M proteins in complex with C4BPα 1-2 were simulated: 1) M2 (amino acids 53-86), 2) M2 K65A (53-86), 3) M2 N66D (53-86), 4) M2 N66A (53-86), 5) M2 K65A/N66A (53-86), 6) M22 (52-79), 7) M28 (55-80), and 8) M49 (56-126). All systems contained residues 1 to 124 of C4BPα .

Minimization, Equilibration, and Production Molecular Dynamics. The NAMD simulation package[18,19] was used to minimize, heat, equilibrate, and simulate each system using a 2 fs time-step. Every system underwent a series of separate minimization, heating, and equilibration stages in preparation for production runs. The minimization spanned five stages in 10 ps intervals using the NVT ensemble: 1) 5,000 steps of hydrogen-only minimization, 2) 5,000 steps of solvent minimization, 3) 5,000 steps of side-chain minimization, 4) 5,000 steps of protein-backbone minimization, and 5) 5,000 steps of full-system minimization. Following minimization, the Langevin thermostat[20,21] was used to slowly heat the system to 310 K using the NVT ensemble over 250,000 steps (500 ps). The system was then subjected to three sequential equilibration stages using the NPT ensemble for 125,000 steps/stage

(250 ps/stage). The pressure was set to 1 atm and maintained using the Beredensen barrostat[22]. In the first MD production run, atoms were assigned a random starting velocity, and sequential steps carried over the velocities from the previous step. Five replicates of each system were performed to enhance sampling of the conformational landscape[23] and the total simulation time for each system was 25 ns/replicate. Therefore, the total aggregate simulation time for each system was 125 ns.

Percent Occupancy (Footprinting) Analysis

.

The five replicates comprising each system (125 ns total) were combined using cpptraj[24], a simulation processing software in the AmberTools package[12]. Trajectories were aligned against the first frame and an average structure was calculated using all atoms in the appropriate protein complex. The average conformation was used to realign the trajectories with respect to Cα atoms. The average conformation was then used to calculate the root mean squared fluctuation (Å) (RMSF) of individual residues in the protein complex. A single concatenated 125 ns trajectory consisting of the five replicates was written by cpptraj and used for the following analysis. Using VMD[25], the radial distribution function (RDF) of pairwise interactions for a number of protein-protein contacts was calculated over the duration of the concatenated trajectory[26]. Distances in the RDF analysis were explicitly calculated for the following heavy atoms of residues:

backbone nitrogen of histidine; Cβ of alanine and valine; Cγ of aspartate, leucine, and isoleucine; Cδ of glutamate; and Cζ of arginine. A 5 Å cutoff was applied to all pairwise interactions to include salt bridges and hydrogen bonds between hydrogen atoms and heavy atoms that were not explicitly analyzed. This was done to capture interactions between equivalent atoms, e.g. Oδ and Oδ ' of aspartate interacting with Hω and Hω ' of arginine.

References

1. Jenkins, H. T. *et al.* Human C4b-binding protein, structural basis for interaction with streptococcal M protein, a major bacterial virulence factor. *J Biol Chem* 281, 3690-3697, (2006).

2. McNamara, C. *et al.* Coiled-coil irregularities and instabilities in group A *Streptococcus* M1 are required for virulence. *Science* 319, 1405-1408, (2008).

3. Andre, I. *et al.* Streptococcal M protein: structural studies of the hypervariable region, free and bound to human C4BP. *Biochemistry* 45, 4559-4568, (2006).

4. Doublie, S. Production of selenomethionyl proteins in prokaryotic and eukaryotic expression systems. *Methods Mol Biol* **363**, 91-108, (2007).

5. Kabsch, W. XDS. *Acta Crystallogr D Biol Crystallogr* **66**, 125-132, (2010).

6. Otwinowski, Z. & Minor, W. Processing of X-ray diffraction data collected in oscillation mode. *Method Enzymol* **276**, 307-326, (1997).

7. Adams, P. D. *et al.* PHENIX: a comprehensive Python-based system for macromolecular structure solution. *Acta Crystallogr D Biol Crystallogr* **66**, 213-221, (2010).

8. Emsley, P. & Cowtan, K. Coot: model-building tools for molecular graphics. *Acta Crystallogr D Biol Crystallogr* **60**, 2126-2132, (2004).

9. Chen, V. B. *et al.* MolProbity: all-atom structure validation for macromolecular crystallography. *Acta Crystallogr D Biol Crystallogr* **66**, 12-21, (2010).

10. Pearlman, D. A. *et al.* Amber, a package of computer-programs for applying molecular mechanics, normal-mode analysis, molecular-dynamics and free-energy calculations to simulate the structural and energetic properties of molecules. *Comput Phys Commun* 91, 1-41, (1995).

11. Wang, J. M., Wolf, R. M., Caldwell, J. W., Kollman, P. A. & Case, D. A. Development and testing of a general amber force field. *J Comput Chem* 25, 1157-1174, (2004).

12. Case, D. A. *et al. AMBER 2015.* (University of California, San Fransisco, 2015).

13. Olsson, M. H. M., Sondergaard, C. R., Rostkowski, M. & Jensen, J. H. PROPKA3: Consistent treatment of internal and surface residues in empirical pK(a) predictions. *J Chem Theory Comput* 7, 525-537, (2011).

14. Sondergaard, C. R., Olsson, M. H. M., Rostkowski, M. & Jensen, J. H. Improved treatment of ligands and coupling effects in empirical calculation and rationalization of pK(a) Values. *J Chem Theory Comput* 7, 2284-2295, (2011).

15. Li, P. F., Roberts, B. P., Chakravorty, D. K. & Merz, K. M. Rational design of particle mesh Ewald compatible Lennard-Jones parameters for +2 metal cations in explicit solvent. *J Chem Theory Comput* 9, 2733-2748, (2013).

16. Darden, T., York, D. & Pedersen, L. Particle mesh Ewald - an N.Log(N) method for Ewald sums in large systems. *J Chem Phys* 98, 10089-10092, (1993).

17. Essmann, U. *et al.* A smooth particle mesh Ewald method. *J Chem Phys* 103, 8577-8593, (1995).

18. Nelson, M. T. *et al.* NAMD: A parallel, object oriented molecular dynamics program. *Int J Supercomput Ap* 10, 251-268, (1996).

19. Phillips, J. C. *et al.* Scalable molecular dynamics with NAMD. *J Comput Chem* 26, 1781-1802, (2005).

20. Izaguirre, J. A., Catarello, D. P., Wozniak, J. M. & Skeel, R. D. Langevin stabilization of molecular dynamics. *J Chem Phys* 114, 2090-2098, (2001).

21. Jiang, W. *et al.* High-Performance scalable molecular dynam-

ics simulations of a polarizable force field based on classical drude oscillators in NAMD. *J Phys Chem Lett* 2, 87-92, (2011).

22. Berendsen, H. J. C., Postma, J. P. M., Vangunsteren, W. F., Dinola, A. & Haak, J. R. Molecular-Dynamics with coupling to an external bath. *J Chem Phys* 81, 3684-3690, (1984).

23. Caves, L. S. D., Evanseck, J. D. & Karplus, M. Locally accessible conformations of proteins: Multiple molecular dynamics simulations of crambin. *Protein Science* 7, 649-666, (1998).

24. Roe, D. R. & Cheatham, T. E. PTRAJ and CPPTRAJ: Software for processing and analysis of molecular dynamics trajectory data. *J Chem Theory Comput* 9, 3084-3095, (2013).

25. Humphrey, W., Dalke, A. & Schulten, K. VMD: Visual molecular dynamics. *J Mol Graph Model* 14, 33-38, (1996).

26. Levine, B. G., Stone, J. E. & Kohlmeyer, A. Fast analysis of molecular dynamics trajectories with graphics processing units-Radial distribution function histogramming. *J Comput Phys* 230, 3556-3569, (2011).