nature microbiology

ARTICLES

PUBLISHED: 5 SEPTEMBER 2016 | ARTICLE NUMBER: 16155 | DOI: 10.1038/NMICROBIOL.2016.155

Conserved patterns hidden within group A *Streptococcus* M protein hypervariability recognize 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]*

No vaccine exists against group A *Streptococcus* (GAS), a leading cause of worldwide morbidity and mortality. A severe hurdle is the hypervariability of its major antigen, the M protein, with >200 different M types known. Neutralizing antibodies typically recognize M protein hypervariable regions (HVRs) and confer narrow protection. In stark contrast, human C4b-binding protein (C4BP), which is recruited to the GAS surface to block phagocytic killing, interacts with a remarkably large number of M protein HVRs (apparently ~90%). Such broad recognition is rare, and we discovered a unique mechanism for this through the structure determination of four sequence-diverse M proteins in complexes with C4BP. The structures revealed a uniform and tolerant 'reading head' in C4BP, which detected conserved sequence patterns hidden within hypervariability. Our results open up possibilities for rational therapies that target the M–C4BP interaction, and also inform a path towards vaccine design.

Group A *Streptococcus* (GAS, *S. pyogenes*) is a major cause of worldwide morbidity and mortality[1]. This bacterial pathogen is responsible for mucosal infections, acute invasive diseases and autoimmune sequelae (for example, pharyngitis, necrotizing fasciitis and rheumatic heart disease, respectively)[2]. Currently, no vaccine against GAS exists[3,4]. A major impediment to immunization is the hypervariability of the antigenic M protein, a surface-anchored virulence factor[5,6] that is also the target of neutralizing antibodies. These antibodies typically recognize the hypervariable region (HVR, N-terminal ~50 amino acids)[7–9] of M proteins, which are dimeric α-helical coiled coils, and thus confer immunity that is M-type specific. One approach to overcome hypervariability is to include multiple M protein HVRs in a vaccine, and, indeed, a vaccine candidate that includes 30 HVRs[10] has advanced into early clinical testing. However, with >200 distinct M protein HVRs[11] and the complexity of global GAS epidemiology[1], even the most-extensive multivalent vaccine is unlikely to offer universal protection. Here we offer structural details that have implications for vaccine design to overcome M protein hypervariability. This approach is based on the finding that human C4b-binding protein (C4BP) recognizes M protein HVRs with broad specificity[12], in stark contrast to the narrow type-specificity displayed by antibodies. In one study, a remarkable ~90% of GAS strains of differing M types bound C4BP (ref. 12). Although this study examined binding to whole bacteria, no protein other than M protein (or an M-like protein, such as Protein H) has been described to bind C4BP, and no region other than the M protein HVR has been described to bind C4BP. C4BP (ref. 13) is a negative regulator of the complement system that binds the complement protein C4b, and thereby disables the C3 convertase of the classical and lectin pathways. GAS recruits C4BP to its surface, as do a number of other pathogens[14,15], to evade opsonophagocytic killing[16,17].

Broad specificity in recognition is rare—it has been observed only in a few cases. A prominent example is the interaction between major histocompatibility complex (MHC) glycoproteins and peptides[18,19]. The breadth of this particular interaction is explained by the fact that MHC glycoproteins primarily make contact to the peptide main chain. To understand the basis for broad specificity in the case of M protein and C4BP, co-crystal structures of four M protein HVRs (M2, M22, M28 and M49) bound to the first two domains of the C4BP α chain were determined (Fig. 1a, Supplementary Fig. 1 and Supplementary Table 1). C4BP consists of seven α chains with disulfide bonds to a single β chain, with each of these chains being composed of multiple ~60-residue complement control protein (CCP) domains[20]. The first two CCP domains of the α chain (C4BPα1 and C4BPα2 (C4BPα1-2)) are sufficient to bind M protein HVRs[21] and C4b (refs 21,22) (Fig. 1a). Overlapping but non-identical sites on C4BP are engaged by M protein HVRs and C4b (ref. 22).

Results

Structural similarity. The structures of the four M protein HVR–C4BPα1-2 complexes (determined between 2.54 and 3.02 Å resolution limits) were astonishingly similar, given the lack of sequence relationship among the M proteins (Supplementary Fig. 1). The M protein–C4BP interface was well defined in electron density and unambiguously modelled (Supplementary Fig. 2), whereas portions of C4BPα1-2 distal to the interface were ill defined, consistent with the inherent flexibility of these domains[20]. The M protein HVRs form parallel dimeric α-helical coiled coils with two C4BPα1-2 molecules bound to each M protein dimer, as prior reports suggested[20,23] (Fig. 1b,c; a detailed view of M2 is shown later, and detailed views of M22, M28 and M49 are given in Supplementary Figs 3–5). The portions of the M proteins that contact C4BPα1-2 are in a canonical coiled-coil conformation, except for M2, which is underwound (Supplementary Fig. 6). C4BPα1 is proximal to the C-terminal portion of the M protein HVR and C4BPα2 to the N-terminal portion, in agreement with the approach of intact C4BP to the streptococcal surface (Fig. 1a).

[1]Department of Chemistry & Biochemistry, University of California, San Diego, La Jolla, California 92093, USA. [2]Department of Pediatrics, University of California, San Diego, La Jolla, California 92093, USA. [3]Skaggs School of Pharmaceutical Sciences, University of California, San Diego, La Jolla, California 92093, USA. *e-mail: pghosh@ucsd.edu

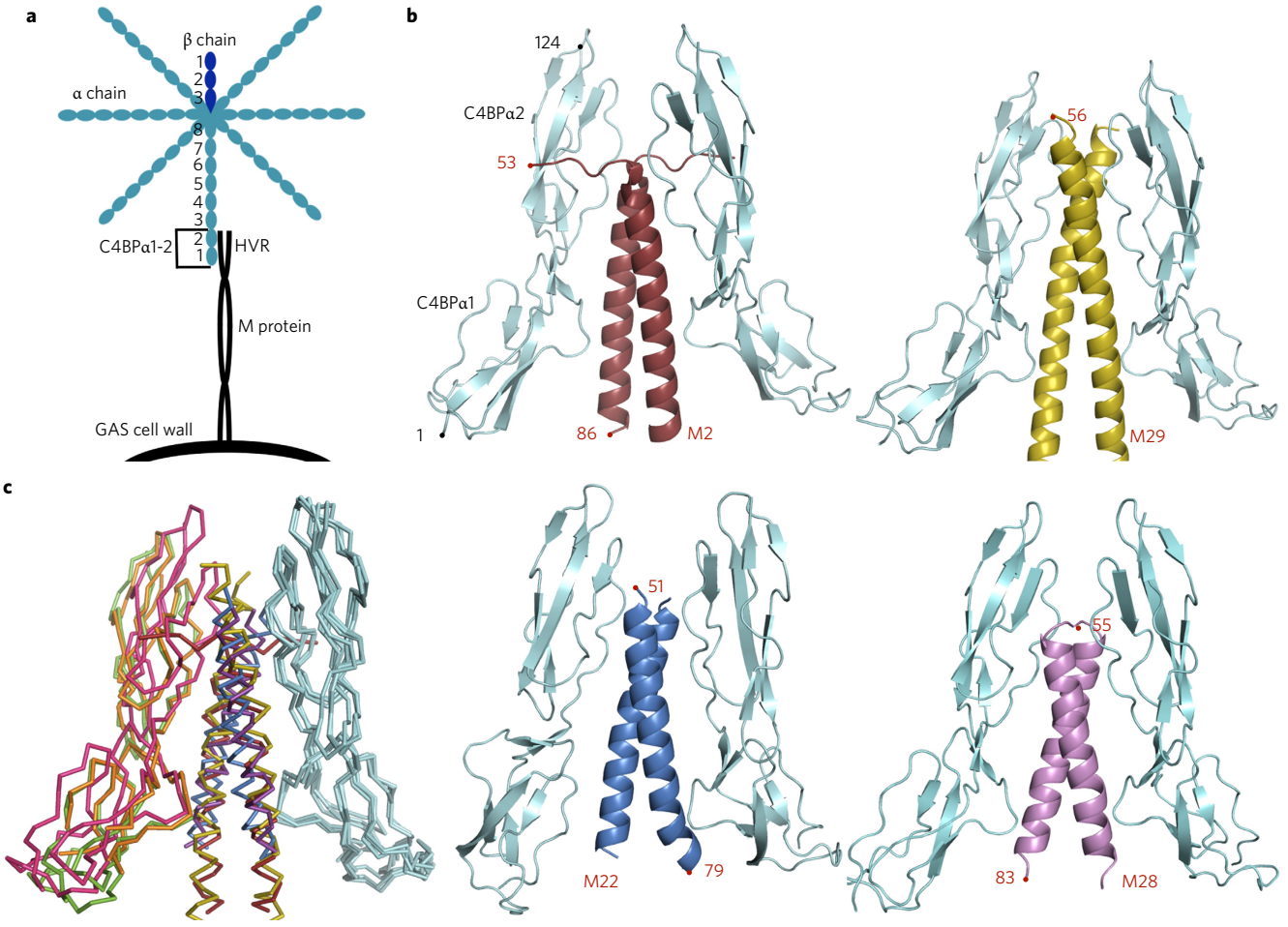

Figure 1 | Structures of M-C4BP complexes. a, The schematic of C4BP (blue) bound to surface-associated M protein (black) highlights the M HVR–C4BPα1-2 interaction. **b**, C4BPα1-2 (cyan) in complex with the HVR of M2 (red, top left), M49 (yellow, top right), M22 (blue, bottom left) and M28 (magenta, bottom right). Terminal residues are numbered. **c**, Superposition of M-C4BP complexes, based on the bound C4BPα1-2 molecule shown in cyan (right). M2 is red and its second bound C4BPα1-2 molecule is green, M49 is yellow and its second bound C4BPα1-2 is orange, M22 is blue and its second bound C4BPα1-2 is omitted (because a crystal contact restricts its orientation into an artefactual conformation) and M28 is magenta and its second bound C4BPα1-2 is pink.

The C4BPα1 and α2 domains are relatively unchanged from their unbound NMR structures[20] (average root mean square deviation (r.m.s.d.) ~1.5 and ~1.0 Å for domains 1 and 2, respectively), except that domain 1 is rotated 180° with respect to domain 2 (Supplementary Figs 7 and 8). This rotation is consistent with evidence from mutagenic[22] and structural[20] studies, and is discussed further below. The M–C4BP interface is extensive, with a total of ~1,450–1,690 Å^2 of the surface area being buried (in the 2:2 complex). Most of this surface area is polar, and the fit is far from hand-in-glove (surface complementarities 0.56–0.66)[24], except for M22 which has a better fit (0.72). These observations suggest a modest binding affinity, consistent with the 0.5 μM K_d (ref. 20) for the interaction between C4BPα1-2 and the M4 HVR. A much tighter association of picomolar K_d (ref. 25) results from the avidity between C4BP, which has multiple bundled arms[26], and surface-localized M protein.

Uniform reading head. Most significantly, the four structures revealed a uniform set of amino acids in C4BP that act as a reading head for recognizing M protein HVRs. Most of this reading head resides in C4BPα2 (Fig. 2a) and takes the form of a quadrilateral that is composed of a hydrophobic pocket that contains C4BP H67, I78 and L82 (1), a hydrogen-bonding group in the form of the main-chain nitrogen of C4BP H67 (2) and two positively charged residues, C4BP R64 (3) and C4BP R66 (4). The segment that holds this quadrilateral is structurally invariant, being stabilized by a disulfide bond at C65 and limited in conformation by P68 (not depicted). The M proteins supply amino acid side chains that interact with these C4BP residues to form complementary quadrilaterals (Fig. 2b). In all four M–C4BP structures, a hydrophobic M protein residue (usually an aromatic) fits into the hydrophobic pocket (1) and a polar M protein residue immediately following the hydrophobic residue, in sequence, hydrogen bonds to the main-chain nitrogen of H67 (2). The contacts to C4BP R64 (3) and C4BP R66 (4) are predominantly electrostatic (usually salt bridges), but in the case of M49 a polar residue is absent and C4BP R64, instead, makes hydrophobic contacts, extending its alkyl chains across several M49 residues. These data are compatible with a report that substitution of the C4BP residues R64, R66 or H67 with Gln affects binding to M4 and M22 (ref. 22). A decreased affinity results in the case of R64Q and H67Q, but increased affinity occurs for R66Q (probably through a gain-of-function).

Uniform reading head contacts from C4BPα1 were far fewer. The key C4BPα1 residue was R39, which formed electrostatic contacts through its guanidinium group as well as hydrophobic contacts through its alkyl chain to create a 'hydrophobic nook' in conjunction with main-chain atoms of C4BPα1 (Fig. 2c). Thus, out of the six C4BP residues that form uniform contacts, three are arginines. This high proportion is probably significant, as the combination

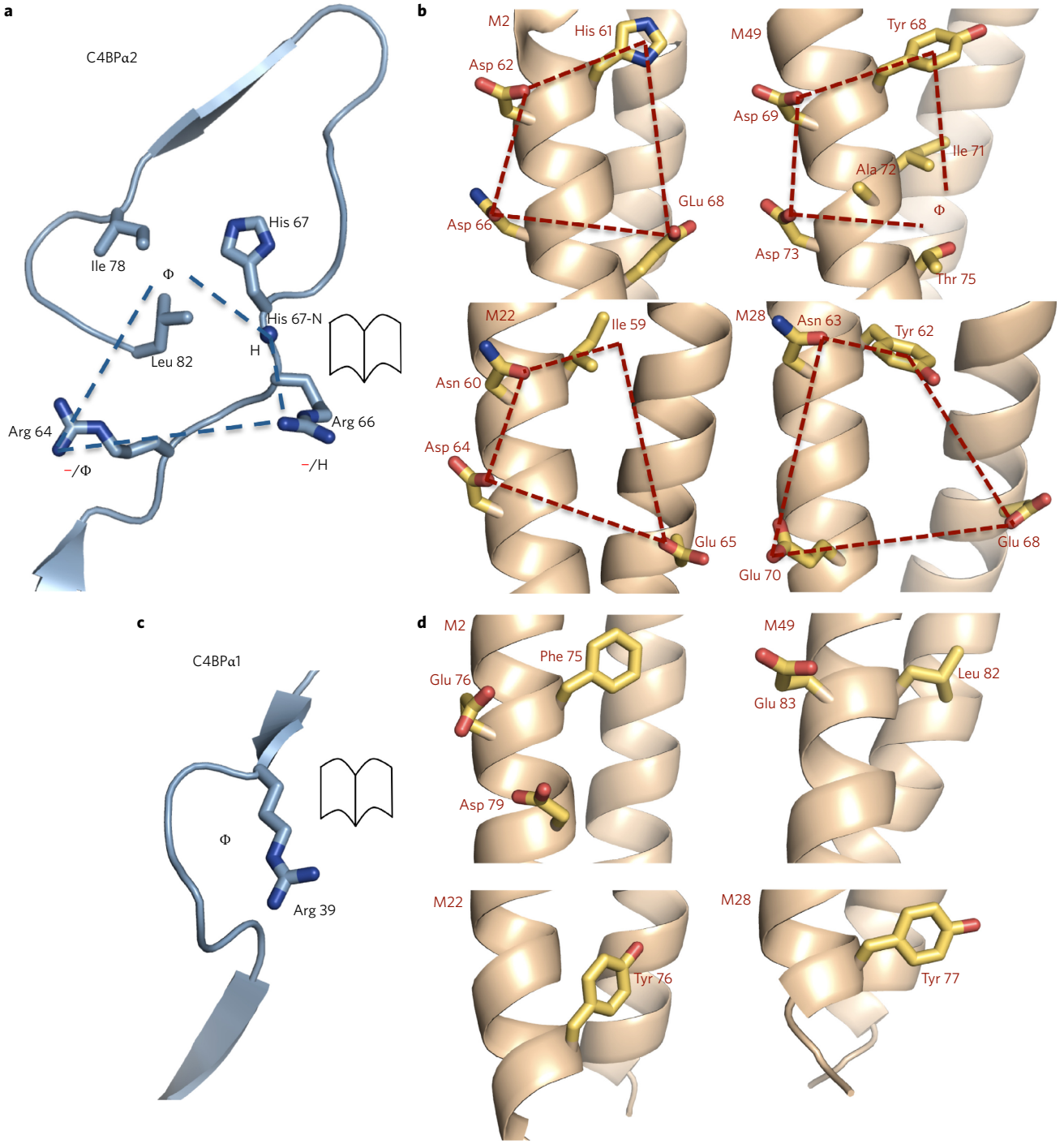

Figure 2 | C4BP-binding mode. a, The C4BPα2 quadrilateral (blue dashed lines), with the C4BPα2 backbone shown in ribbon representation and key side chains shown as bonds, in which carbons are cyan and nitrogens are blue. The chemical character of M protein residues that interact with the quadrilateral are hydrophobic (φ), negative (-) and hydrogen-bond forming (H). **b**, M2, M49, M22 and M28 residues that interact with the C4BPα2 quadrilateral and form a complementary quadrilateral (red dashed lines), shown in an open-book representation with respect to C4BPα2. The M protein backbone is in ribbon representation and key side chains are shown as bonds (C, yellow; O, red; N, blue). The numbering of the M proteins is such that the initiator Met is residue 1. **c**, The C4BPα1 Arg39 nook. The depiction and symbols are as in **a**. **d**, M2, M49, M22 and M28 residues that interact with the C4BPα1 Arg39 nook shown in an open-book representation. The depiction is as in **b**.

of polar and apolar atoms in Arg along with its chain length increase the possibilities for interactions with variable residues. Substitution of C4BP R39 with Gln results in decreased binding to M4, but increased binding to M22 (again, probably a gain-of-function)[22]. All four M proteins have hydrophobic residues that insert into the C4BPα1 'hydrophobic nook'. M2 and M49 also have negatively charged residues that interact with C4BP R39, whereas neither M22 nor M28 do. The importance of C4BP R39 provided an explanation for the aforementioned 180° rotation of C4BPα1 (around a hinge at K63 (Supplementary Fig. 7)). In free C4BP, the C4BPα1 R39 nook and the C4BPα2 quadrilateral are on opposite sides and require a 180° reorientation to interact simultaneously with M protein. This 180° rotation was seen in all four structures. However, in one of the two C4BPα1-2 molecules bound to M22,

the 180° rotation was prevented because of a crystal contact (Supplementary Figs 7c,d and 8). A similar 180° rotation appears necessary for the interaction of C4BP with C4b, as it has been demonstrated that R39 and the set of residues in C4BPα2 that interact with M protein HVRs also interact with C4b (ref. 22). The purpose of requiring a 180° rotation in C4BPα1 to transition between free and bound forms is unclear.

Sequence conservation hidden within hypervariability. The evidence gathered from these structures proved powerful in bringing to light weak sequence conservation in M protein HVRs. Comparison of the heptad position of M protein residues that interacted with C4BP made it clear that two binding patterns were observed in these structures, with chemically similar residues contacting C4BP (Fig. 3a): one for M2 and M49, and a separate one for M22 and M28. These two patterns were also evident in the way the coiled coils interacted with C4BP. In the case of M2 and M49, the coiled coils ran roughly parallel to C4BPα1-2, such that each C4BPα1-2 molecule contacted a single α-helix (Figs 1b and 2b). However, in the case of M22 and M28, the coiled coils lay crosswise across C4BPα1-2 such that each C4BPα1-2 molecule contacted both α-helices. Remarkably, these same two patterns were evident in a larger number of M proteins (Fig. 3b and Supplementary Figs 9 and 10), which had chemically similar amino acids to those in M2/M49 or M22/M28 that had been visualized to contact C4BP. We were able to assign 13 M proteins to the M2/M49 pattern and 32 (including the M-like Protein H) to the M22/M28 pattern. Thus, these two patterns may explain the interaction of nearly half of the M strains that were studied for C4BP binding[12]. A further 46 M proteins from this study[12] could not be assigned to either pattern (Supplementary Fig. 11). In these cases, it is possible that proteins besides M recruit C4BP to the GAS surface or that other regions of the M proteins do. However, as, for GAS, only M proteins and M protein HVRs have been documented to bind C4BP, we think it more likely that there are still other arrangements by which M protein HVRs interact with C4BP.

Tolerance to hypervariability. We next sought to understand the tolerance in C4BP to sequence variation in M protein, as single amino acid changes have been shown to alter recognition by antibodies but not by C4BP (ref. 12). Alanine substitutions were created in the M2 residues mentioned above that make contact with the uniform reading head. In addition, substitutions were made in two M2 residues, K65 and E83 (the numbering of M proteins is such that the initiator Met is residue 1), which make contacts observed only in M2 (Fig. 4). The M2–C4BP interaction was evaluated by a Ni^{2+}-NTA agarose (NTA, nitroloacetic acid) co-precipitation assay using His-tagged C4BPα1-2 (Fig. 4a and Supplementary Fig. 12). Of the single-site substitutions, only F75A, which binds in the C4BPα1 nook, showed substantially decreased binding. Molecular dynamics (MD) simulations of this mutation highlighted the importance of M2 F75. On alanine substitution of F75, contacts between the R39 'hydrophobic nook' and M2 diminished and waters infiltrated into this site (Supplementary Fig. 13 and Supplementary Videos 1 and 2). Strikingly, all the other residues could be mutated to Ala without substantial losses in binding (Supplementary Fig. 12) and, indeed, in two cases an increased binding was observed (see below). Providing verification for the structural observations, two double substitutions resulted in substantial loss of binding: D62A/E68A, which removed two of the polar contacts to the C4BPα2 quadrilateral, and E76A/D79A, which removed the two salt bridges to C4BPα1 R39.

Surprisingly, two of the single-site mutations, M2 K65A and N66A, increased binding, as did the K65A/N66A double mutant

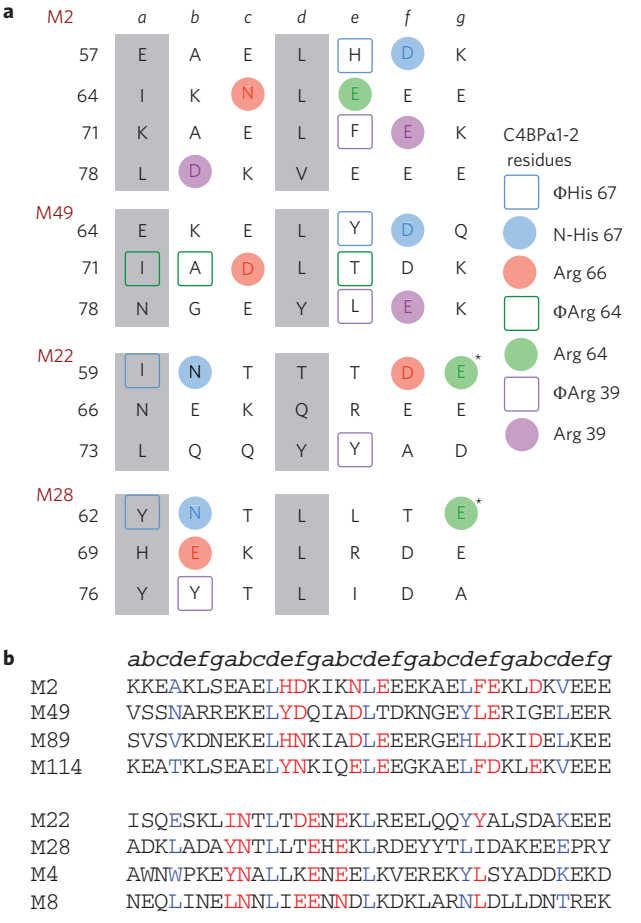

Figure 3 | C4BP-binding modes of M proteins. a, Heptad registers of M2, M49, M22 and M28 HVRs (*a* and *d* residues in grey). M protein residues that interact with C4BPα1-2 residues are highlighted according to their corresponding C4BPα1-2 interaction. **b**, Sequence alignment of M protein HVRs that belong to the M2/M49 group (top) or the M22/M28 group (bottom). Residues that contact or are predicted to contact C4BPα1-2 are in red. Residues observed or predicted to be at core *d* positions of the heptad register are highlighted in blue for visual reference.

(Fig. 4a and Supplementary Fig. 12). The structure of M2 (K65A/N66A) in complex with C4BPα1-2 was determined to a resolution limit of 2.29 Å, and no reordering of the binding site was evident (r.m.s.d. 0.15 Å) (Supplementary Fig. 14). This result suggested that M2 K65 and N66 were tolerated in the binding site, but were not optimal. Although M2 K65 formed a hydrogen bond to the main-chain oxygen of C4BP R64, it was sandwiched between two positively charged side chains (that is, C4BP R64 and R66), which provides an explanation as to why Ala substitution of K65 led to an enhanced binding. MD simulations reinforced this interpretation, as Ala substitution of K65 led to better contacts between M2 and C4BP, especially evident in the increased frequency of hydrogen bonding between C4BP R66 and M2 N66 (Supplementary Table 2). The simulations suggested that the hydrogen bond between these two residues was otherwise infrequent (Supplementary Table 2 and Supplementary Videos 3 and 4), and, indeed, these two residues had the highest B-factors in the binding site (Supplementary Fig. 15). In other M proteins that belong to the M2/M49 pattern, the equivalent of N66 is almost always Asp or Glu (Supplementary Fig. 9). Consistent with this trend, substitution of M2 N66 with Asp resulted in increased binding (Fig. 4a), and MD simulations provided evidence of the favourable interactions between C4BP R66 and negatively charged M protein amino acids

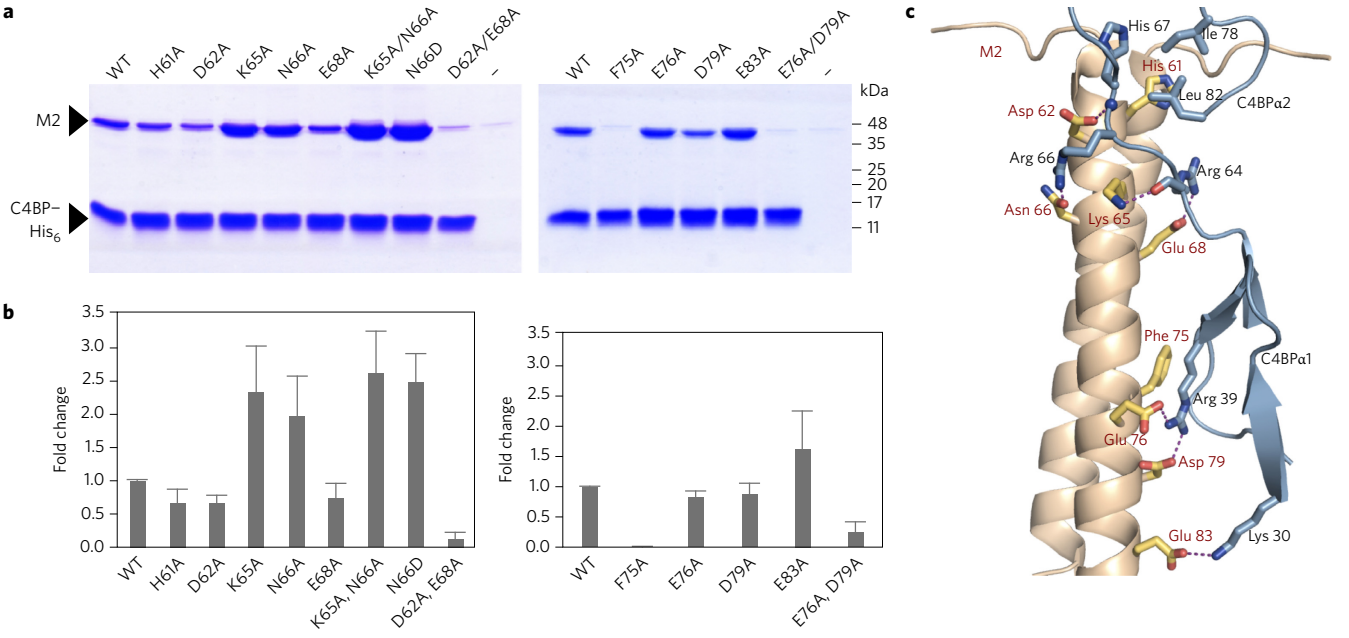

Figure 4 | M2–C4BP interaction. a, Association of His-tagged C4BPα1-2 with WT and mutant M2 HVR at 37 °C, as assessed by a Ni^{2+}-NTA agarose co-precipitation assay and visualized by non-reducing, Coomassie-stained SDS-PAGE. Only bound fractions are shown here. Input samples are shown in Supplementary Fig. 12. This gel is representative of four experimental replicates. Molecular mass markers were not run on these particular gels; their positions are based on measurements from equivalent gels. **b**, Quantification of the interaction between C4BPα1-2 and WT M2 or M2 mutant proteins. The values shown are averages of four experimental replicates, corrected for the level of background binding (that is, no C4BPα1-2) and normalized to WT M2. Standard deviations are depicted. **c**, Structure of M2 (grey ribbon representation with key side chains in a bond representation (C, yellow; O, red; N, blue)) bound to C4BPα1-2 (cyan ribbon representation, with key side chains in a bond representation (C, cyan; N, blue)). Hydrogen bonds and salt bridges are depicted by dashed magenta lines.

(Supplementary Table 3 and Supplementary Video 5). Puzzlingly, substitution of M2 N66 with Ala (and thus loss of hydrogen bonding to C4BP R66) also resulted in better binding. C4BP R66 had an even higher relative B-factor when in contact with M2 (K65A/N66A) as compared with wild-type (WT) M2 (Supplementary Fig. 15). Thus, it appears that C4BP R66 prefers a salt bridge (for example, N66D) or no interaction (for example, N66A) to a hydrogen bond, because the salt bridge provides sufficient binding energy to relieve the entropic cost of ordering the Arg, whereas the hydrogen bond does not. In short, the mutagenesis experiments reinforced the notion that the reading head in C4BP is highly tolerant to variation in the M protein.

Discussion

We have shown that the broad recognition between M proteins and C4BP is not the result of contacts to the main chain, as it is for MHC–peptide complexes. Instead, the breadth of recognition in M–C4BP complexes is explained by three unique attributes. First, the C4BP-binding site is tolerant, notably because of the prevalence of arginines. The combination of a charged head and a long alkyl body enables arginine to engage in both electrostatic and hydrophobic interactions. As a result, only loose restrictions apply to M protein side chains that interact with C4BP arginines. For example, whereas negatively charged M protein side chains were preferred for the C4BPα2 quadrilateral arginines, a set of hydrophobic residues were accommodated in M49; and for C4BPα1 R39, hydrophobic M protein side chains were the commonality. Second, there appears to be no 'hot spot' for interaction and instead the binding energy appears to be dispersed broadly over the interaction site. No single amino acid substitution in M2, except for one, reduced binding to C4BP substantially. A similar observation has been made for M22 (ref. 12). Alanine substitution of M22 E65 (E24 in Persson *et al.*[12]), which we found interacts with C4BP R64, did not change binding to C4BP, but did alter recognition by antibodies[12]. Third, the M protein coiled coil can align with C4BP in multiple ways. This enables M protein side chains that interact with C4BP to reside at different positions of the heptad repeat. Two different arrangements were seen here, but more are likely to be discovered.

C4BP is recruited by a large number of pathogens (including viruses and fungi) to prevent phagocytic uptake, the formation of the membrane-attack complex and the generation of immunostimulatory anaphylatoxins (for example, C3a and C5a)[13]. The importance of C4BP recruitment to GAS infection was demonstrated in an M22 strain. Specific loss of C4BP binding in this strain was effected through a seven-residue deletion in M22, which our results indicate eliminated interaction with C4BP R64 and R66. This C4BP-binding-deficient M22 strain was ~3- to 13-fold more susceptible to elimination by human blood as compared with the WT M22 strain[16,27]. Further evidence for the importance of C4BP recruitment was garnered recently using transgenic mice that expressed human C4BP (ref. 17) (murine C4BP does not bind M protein[21]). In particular, human C4BP transgenic mice showed a much earlier time to death as compared with non-transgenic mice when infected by a C4BP-binding GAS strain. This and other effects, including bacterial burden and levels of proinflammatory cytokines, were exacerbated when these mice also expressed human factor H, another soluble negative regulator of the complement system. Interestingly, factor H, which is composed of CCP domains like C4BP, also binds M protein HVRs[28]. Although M protein HVRs generally bind either C4BP or factor H[28], the GAS strain used in this study produced Protein H[17], which binds both[29]. Our results provide, to our knowledge, the first atomic-level understanding of the interaction between a negative regulator of the complement system and a microbial virulence factor, and open up possibilities for the rational disruption of the M-C4BP interaction for therapeutic ends.

Lastly, our work has implications for vaccine design. Broadly neutralizing antibodies (bNAbs) have been identified for several highly antigenically variable microbial pathogens, including the human immunodeficiency virus (HIV) and the influenza virus[30–33]. These antibodies target invariant structural regions that are often hidden through various factors, including glycosylation, as in the case of HIV, or steric occlusion, as in the case of both influenza and HIV. The hypervariability of M proteins has hindered the development of a GAS vaccine. Our work shows that hidden within M protein hypervariability are sequence patterns that are conserved and utilized for interaction with C4BP. This finding suggests it may be possible, using appropriate antigens, to elicit bNAbs against GAS, which would mimic the broad specificity of binding to M protein HVRs observed in C4BP. In this regard, it is notable that the 30-valent M protein vaccine displays some measure of crossreactivity to M types not included in the vaccine[10]. Indeed, 14 of the 30 M proteins in the vaccine belong to pattern 1 or 2, and crossreactivity was seen to 16 M proteins not in the vaccine belonging to pattern 1 or 2. Further work is necessary to determine if and which other conserved patterns exist in the set of M proteins that are not readily assignable to either pattern 1 or 2, but still bind C4BP. An alternative notion for conferring broad immunity is transplantation of the C4BP loops that form the uniform and tolerant reading head of C4BP to the antigen-combining site of an antibody. Such an antibody is predicted to have broad specificity against M proteins, and would provide neutralization because of targeting of the M protein HVR. This passive form of immunity could be made active by screening for antigens that bind tightly to the antibody displaying the C4BP reading head.

A potential challenge in these approaches based on the mode of C4BP binding is that the antibodies obtained through such methods may also recognize C4b. However, differences in C4BP-binding modes between M protein HVRs and C4b suggest that selectivity is possible[22]. A second challenge may be escape from such antibodies through further M protein variation. However, M protein HVRs vary from strain to strain but are stable within the type[34], which suggests that their overall sequence variation is limited by positive selection. Binding to C4BP appears to be a major evolutionary selective pressure for GAS[17]; thus, escape from such bNAbs that target M protein HVRs through further sequence variation may be limited by pressure to maintain C4BP interaction.

Methods

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 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. Each site-directed mutagenesis was performed according to the Agilent QuikChange manual, except that 50 µl reactions were set up for polymerase chain reactions instead of 12.5 µl reactions.

The coding sequence of the CCP1-2 domains of the human C4BPα chain (C4BPα1-2)[20] (a 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 C4BPα1-2 was used for crystallographic studies, and the non-cleavable C-terminal His_6-tagged version for co-precipitation binding studies. To obtain selenomethionine (SeMet)-substituted proteins to be used in the phase determination, methionines were introduced in the coding sequence of C4BPα1-2 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 previously[5] with minor modifications to the procedure. Specifically, bacteria were lysed with a C-5 Emulsiflex (Avestin) and ion-exchange chromatography was omitted, and in the case of purification of M2 (WT and variants), imidazole was not included in the lysis and wash buffers.

C4BPα1-2 was expressed in *E. coli* Rosetta 2 (Novagen) cells. The protein was purified and refolded as described previously[23], except for the use of a C-5 Emulsiflex for lysis. Where needed, the N-terminal His_6-tags of M proteins and C4BPα1-2 were removed by PreScission protease cleavage according to the manufacturer's instructions, and the cleaved protein was purified by reverse Ni^{2+}-NTA chromatography. Lastly, M proteins and C4BPα1-2 were 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 ~20 mg ml^{-1} 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 °C.

SeMet was incorporated into C4BPα1-2 (L29M/L46M), C4BPα1-2 (L29M/L71M) and C4BPα1-2 (L46M/L71M) using methionine-pathway inhibition as described previously[35]. SeMet-labelled C4BPα1-2 was purified as described above.

Crystallization and data collection. For the preparation of the 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 C4BPα1-2 (WT or SeMet-substituted mutant) at a 1:1 molar ratio (final concentration of complex ~5 mg ml^{-1}) and dialysed overnight at 4 °C in 10 mM Tris, pH 8. The samples were then concentrated by ultrafiltration to ~20 mg ml^{-1}. Crystallization was performed by the hanging-drop vapour-diffusion method.

The M2–C4BPα1-2, M2 (K65A/N66A)–C4BPα1-2 and M28–C4BPα1-2 complexes and the SeMet-labelled M2–C4BPα1-2 (L29M/L46M) and M2–C4BPα1-2 (L46M/L71M) complexes were co-crystallized at 20 °C by mixing 1 µl of complex with 1 µl 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 fibre loops and flash-cooled in liquid N_2. Crystals that contained SeMet-labelled protein were treated similarly, except the reservoir solution was supplemented with freshly prepared 1 mM tris(2-carboxyethyl)phosphine.

The M22–C4BPα1-2 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-labelled M49–C4BPα1-2 L29M/L46M complex was co-crystallized similarly, except the reservoir solution was 1.6 M Na/KPO_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–C4BPα1-2 were collected at the Stanford Synchrotron Radiation Lightsource beamline 9-2, those for M22–C4BPα1-2 at the Advanced Photon Source (APS) beamline 24-ID-C and those for M2 (K65A/N66A)–C4BPα1-2 and M28–C4BPα1-2 at the Advanced Lightsource beamline 8.2.1. Single-wavelength anomalous dispersion (SAD) data were collected from SeMet-labelled M2–C4BPα1-2 (L29M/L46M) and M2–C4BPα1-2 (L46M/L71M) at the APS beamline 19-ID, and from SeMet-labelled M49–C4BPα1-2 (L29M/L46M) at the APS beamline 24-ID-E.

Diffraction data from crystals of M22–C4BPα1-2 and M49–C4BPα1-2 (L29M/L46M) were indexed, integrated and scaled using XDS (ref. 36), whereas HKL2000 (ref. 37) was used for data from all the other crystals.

Structure determination and refinement

M2–C4BPα1-2. For the structure determination of M2–C4BPα1-2, Se sites were located from SAD data of SeMet-labelled M2–C4BPα1-2 (L29M/L46M) and M2–C4BPα1-2 (L46M/L71M), and phases calculated for each data set using Autosol (within PHENIX (ref. 38)). 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 C4BPα1-2 molecule.

Here, and in all the cases below, model building was carried out with Coot[39] as guided by the inspection of SAD-phased maps or σ_A-weighted $2mF_o - DF_c$ and $mF_o - DF_c$ maps (F_o and F_c are the experimentally measured and model-based amplitudes, respectively, m is the figure of merit and D is the σ_A weighting factor) 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 1–10 rounds, were performed in each case. In the later stages of refinement, TLS (translation, libration and screw-rotation) 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).

To model M2–C4BPα1-2 (L29M/L46M/L71M), the NMR structure of C4BPα1-2 was manually fit into SAD-phased density, with the two domains of C4BPα1-2 being treated as individual rigid bodies. The M2 molecule was then built into the density, with the register of the coiled coil being assigned from a well-defined density that corresponded to large side chains (that is, 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 (2.56 Å resolution limit) data collected from crystals of M2–C4BPα1-2. TLS parameterization involved the following groups: for M2, 53–57 and 58–86, and for C4BPα1-2, 0–59 and 60–124. Continuous electron density was evident for the entire main chain of C4BPα1-2 and for residues 53–86 of

the M2 protein. Here, and in all the cases below, electron density was visible for side chains of M protein residues, except for some solvent-exposed flexible residues (that is, Lys, Arg or Glu) distant from the interface with C4BPα1-2. Electron density was also visible for side chains of C4BPα1-2, except for some residues in long loops that were also distant from the interface with M protein. An exception to this was C4BP R66, for which electron density for the side chain was broken. Long loops of C4BPα1-2 also contained some residues whose ϕ and ψ angles were in the outlier region of the Ramachandran plot. For the M2 protein, the only residue in the Ramachandran outlier region was A58, which is the N-terminal residue of the M2 model.

The structure of M2 (K65A/N66A)–C4BPα1-2 was determined by difference Fourier synthesis using the refined structure of M2–C4BPα1-2. The set of reflections used for R_{free} calculations for the refinement of M2-C4BPα1-2 was maintained. TLS parameterization was equivalent to that for M2–C4BPα1-2.

M28–C4BPα1-2. The structure of M28–C4BPα1-2 was determined by molecular replacement using the program Phaser (within PHENIX). The C4BPα1-2 molecule from the structure of the M2–C4BPα1-2 complex served as the search model. The molecular-replacement solution had a log-likelihood gain score of 379. The asymmetric unit contained one C4BPα1-2 molecule and one M28 α-helix, whose register was determined by a well-defined density that corresponded to large side chains (that is, Tyr 62, Tyr 76 and 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, 55–83, and for C4BPα1-2, 0–59, 60–86 and 87–124. Continuous electron density was evident for the entire main chain of C4BPα1-2, except for breaks in some of its longer loops, and for amino acids 53–83 of M28.

M22–C4BPα1-2. The structure of the M22–C4BPα1-2 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 C4BPα1-2 molecule. The solution, which had a log-likelihood gain score of 166, resulted in two copies of the search model in the asymmetric unit, whereas the solvent content suggested that the asymmetric unit was composed of two M22 α-helical dimeric coiled coils and four C4BPα1-2 molecules; this latter composition was found to be accurate. After refinement of the initial molecular-replacement model, two additional C4BPα1-2 molecules became evident in the electron-density maps, and were placed stepwise into the density, with the two domains of C4BPα1-2 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 C4BPα1 and C4BPα2 domains relative to these domains in unbound C4BPα1-2. This tilted orientation differs from the 180° rotation observed in the two other copies of C4BPα1-2 bound to M22, as well as in copies of C4BPα1-2 bound to M2, M28 and M49. Side chains for M22 were subsequently built into the density, with the register being assigned based on a well-defined density that corresponded to large side chains (that is, 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 the M22 chain A, 52–80; for the M22 chain C, 52–79; for the M22 chain E, 52–79; for the M22 chain G, 52–80; for the C4BPα1-2 chain B, 1–13, 14–27, 28–59, 60–73, 74–86, 87–102, 103–109, 110–115 and 116–124; for the C4BPα1-2 chain D, 0–59 and 60–124; for the C4BPα1-2 chain F, 1–59 and 60–124; for the C4BPα1-2 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 C4BPα1-2, except for breaks in some of the longer loops, and for residues 52–79 (or 80, depending on the chain) of M22.

M49–C4BPα1-2. For the structure determination of M49-C4BPα1-2, Se sites were located from SAD data collected for SeMet-labelled M49–C4BPα1-2 (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 C4BPα1-2 molecules. This is consistent with the total of two SeMet substitutions introduced into C4BPα1-2. The crystal structure of C4BPα1-2 from the M2–C4BPα1-2 co-crystal structure was manually fit into the SAD-phased density, with the two domains of C4BPα1-2 being treated as individual rigid bodies. A model of the M49 protein was then built into the density, with the amino acid register for the coiled coil being assigned based on a well-defined density that corresponded to large side chains (that is, 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 C4BPα1-2 chain B, 0–10, 11–62 and 63–124; for C4BPα1-2 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 C4BPα1-2, except for some of the longer loops of the C4BPα1 domain, and for amino acids 56–124 (or 126, depending on the chain) of M49. The M49 residue A106 of chain A had ϕ and ψ angles that were in the outlier region of the Ramachandran plot; this residue was distant from the interface with C4BPα1-2.

Validation of structures. Structural models were validated with MolProbity[40] (Supplementary Table 1). Molecular figures were made with PyMol (http://pymol.sourceforge.net).

Co-precipitation assays. C4BPα1-2-His_6 protein (40 μg) was mixed with 120 μg of intact M2 protein (WT or mutant) in 50 μl of phosphate-buffered saline (PBS) at 37 °C for 30 minutes. Ni^{2+}-NTA agarose beads (50 μl) were equilibrated in PBS, then added to the protein mix in a 1:1 beads:PBS (100 μl) slurry and incubated for 30 minutes 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 μl of 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. Gels were scanned and ImageJ[41] was used to quantify band intensities. In total, four independent co-precipitation experiments were quantified, and band intensities were verified to be within the linear range of measurement. The intensity of the band from the lane that contains no C4BPα1-2 was subtracted as background from other measurements. Values were normalized to the value of WT M2.

Molecular dynamics

System preparation. Heavy-atom coordinates were taken from the co-crystal structures of M protein–C4BPα1-2 complexes. Structures of complexes that contain M2-substitution mutants were created by computational point mutations at the desired amino acid(s). As a result of 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[42–44]. The ionization states of titratable residues at pH 7 were predicted using PROPKA 3.1[45,46] and visually inspected to ensure the accuracy of the assigned states. Free cysteine residues were converted into disulfide-bonded pairs manually and built using tLeap, a system preparation program from the Amber Tools 2015 package[44]. The C-termini of proteins were capped to remove the charges. The solvent was modelled explicitly using the TIP4P water model[47] 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[48,49] was employed to evaluate electrostatics during the simulation. In total, nine different M proteins in complex with C4BPα1-2 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) M2 F75A (53–86), (7) M22 (52–79), (8) M28 (55–80) and (9) M49 (56–126). All the systems contained residues 1 to 124 of C4BPα.

Minimization, equilibration and production. The NAMD simulation package[50,51] 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 the production runs. The minimization spanned five stages in 10 ps intervals using the NVT (number of particles, volume and temperature) 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. After minimization, the Langevin thermostat[52,53] was used to heat the system slowly 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 (number of particles, pressure and temperature) ensemble for 125,000 steps per stage (250 ps per stage). The pressure was set to 1 atm and maintained using a Beredensen barrostat[54]. 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 the sampling of the conformational landscape[55] and the total simulation time for each system was 25 ns per replicate (40 ns per replicate for M2 F75A). Therefore, the total aggregate simulation time for each system was 125 ns (200 ns for M2 F75A).

Percent occupancy (footprinting) analysis. The five replicates that comprise each system (125 ns total for all systems, except 200 ns total for M2 F75A) were combined using CPPTRAJ[56], a simulation processing software in the AmberTools package[44]. 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) of individual residues in the protein complex. A single concatenated 125 ns (200 ns for M2 F75A) trajectory that consisted of the five replicates was written by CPPTRAJ and used for the following analysis.

Using visual molecular dynamics[57], the radial distribution function (RDF) of pairwise interactions for a number of protein–protein contacts was calculated over the duration of the concatenated trajectory[58]. Distances in the RDF analysis were calculated explicitly 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 analysed. This was done to capture the interactions between equivalent atoms, for example, Oδ and Oδ′ of aspartate interacting with Hω and Hω′ of arginine.

Data availability. Sequences for M2–C4BPα1-2, M28–C4BPα1-2, M22–C4BPα1-2, M49–C4BPα1-2 and M2 (K65A/N66A)–C4BPα1-2 have been deposited in the Protein Data Bank under accession numbers 5HYU, 5HYP, 5HYT, 5HZP and 5I0Q, respectively.

Received 29 February 2016; accepted 22 July 2016; published 5 September 2016

References

1. Carapetis, J. R., Steer, A. C., Mulholland, E. K. & Weber, M. The global burden of group A streptococcal diseases. *Lancet Infect. Dis.* **5,** 685–694 (2005).
2. Cole, J. N., Barnett, T. C., Nizet, V. & Walker, M. J. Molecular insight into invasive group A streptococcal disease. *Nat. Rev. Microbiol.* **9,** 724–736 (2011).
3. Dale, J. B. *et al.* Group A streptococcal vaccines: paving a path for accelerated development. *Vaccine* **31** (Suppl. 2), B216–B222 (2013).
4. Good, M. F., Pandey, M., Batzloff, M. R. & Tyrrell, G. J. Strategic development of the conserved region of the M protein and other candidates as vaccines to prevent infection with group A streptococci. *Expert Rev. Vaccines* **14,** 1459–1470 (2015).
5. McNamara, C. *et al.* Coiled-coil irregularities and instabilities in group A *Streptococcus* M1 are required for virulence. *Science* **319,** 1405–1408 (2008).
6. Ghosh, P. The nonideal coiled coil of M protein and its multifarious functions in pathogenesis. *Adv. Exp. Med. Biol.* **715,** 197–211 (2011).
7. Sandin, C., Carlsson, F. & Lindahl, G. Binding of human plasma proteins to *Streptococcus pyogenes* M protein determines the location of opsonic and non-opsonic epitopes. *Mol. Microbiol.* **59,** 20–30 (2006).
8. Penfound, T. A., Ofek, I., Courtney, H. S., Hasty, D. L. & Dale, J. B. The NH_2-terminal region of *Streptococcus pyogenes* M5 protein confers protection against degradation by proteases and enhances mucosal colonization of mice. *J. Infect. Dis.* **201,** 1580–1588 (2010).
9. Lannergard, J. *et al.* The hypervariable region of *Streptococcus pyogenes* M protein escapes antibody attack by antigenic variation and weak immunogenicity. *Cell Host Microbe* **10,** 147–157 (2011).
10. Dale, J. B., Penfound, T. A., Chiang, E. Y. & Walton, W. J. New 30-valent M protein-based vaccine evokes cross-opsonic antibodies against non-vaccine serotypes of group A streptococci. *Vaccine* **29,** 8175–8178 (2011).
11. McMillan, D. J. *et al.* Updated model of group A *Streptococcus* M proteins based on a comprehensive worldwide study. *Clin. Microbiol. Infect.* **19,** E222–E229 (2013).
12. Persson, J., Beall, B., Linse, S. & Lindahl, G. Extreme sequence divergence but conserved ligand-binding specificity in *Streptococcus pyogenes* M protein. *PLoS Pathog.* **2,** e47 (2006).
13. Ermert, D. & Blom, A. M. C4b-binding protein: the good, the bad and the deadly. Novel functions of an old friend. *Immunol. Lett.* **169,** 82–92 (2016).
14. Lambris, J. D., Ricklin, D. & Geisbrecht, B. V. Complement evasion by human pathogens. *Nat. Rev. Microbiol.* **6,** 132–142 (2008).
15. Blom, A. M., Hallstrom, T. & Riesbeck, K. Complement evasion strategies of pathogens—acquisition of inhibitors and beyond. *Mol. Immunol.* **46,** 2808–2817 (2009).
16. Carlsson, F., Berggard, K., Stalhammar-Carlemalm, M. & Lindahl, G. Evasion of phagocytosis through cooperation between two ligand-binding regions in *Streptococcus pyogenes* M protein. *J. Exp. Med.* **198,** 1057–1068 (2003).
17. Ermert, D. *et al.* Virulence of group A *Streptococci* is enhanced by human complement inhibitors. *PLoS Pathog.* **11,** e1005043 (2015).
18. Fremont, D. H., Matsumura, M., Stura, E. A., Peterson, P. A. & Wilson, I. A. Crystal structures of two viral peptides in complex with murine MHC class I H-2Kb. *Science* **257,** 919–927 (1992).
19. Madden, D. R., Gorga, J. C., Strominger, J. L. & Wiley, D. C. The three-dimensional structure of HLA-B27 at 2.1 Å resolution suggests a general mechanism for tight peptide binding to MHC. *Cell* **70,** 1035–1048 (1992).
20. 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).
21. Accardo, P., Sanchez-Corral, P., Criado, O., Garcia, E. & Rodriguez de Cordoba, S. Binding of human complement component C4b-binding protein (C4BP) to *Streptococcus pyogenes* involves the C4b-binding site. *J. Immunol.* **157,** 4935–4939 (1996).
22. Blom, A. M. *et al.* Human C4b-binding protein has overlapping, but not identical, binding sites for C4b and streptococcal M proteins. *J. Immunol.* **164,** 5328–5336 (2000).
23. Andre, I. *et al.* Streptococcal M protein: structural studies of the hypervariable region, free and bound to human C4BP. *Biochemistry* **45,** 4559–4568 (2006).
24. Lawrence, M. C. & Colman, P. M. Shape complementarity at protein/protein interfaces. *J. Mol. Biol.* **234,** 946–950 (1993).
25. Sanderson-Smith, M. *et al.* A systematic and functional classification of *Streptococcus pyogenes* that serves as a new tool for molecular typing and vaccine development. *J. Infect. Dis.* **210,** 1325–1338 (2014).
26. Perkins, S. J., Chung, L. P. & Reid, K. B. Unusual ultrastructure of complement-component-C4b-binding protein of human complement by synchrotron X-ray scattering and hydrodynamic analysis. *Biochem. J.* **233,** 799–807 (1986).
27. Berggard, K. *et al.* Binding of human C4BP to the hypervariable region of M protein: a molecular mechanism of phagocytosis resistance in *Streptococcus pyogenes*. *Mol. Microbiol.* **42,** 539–551 (2001).
28. Gustafsson, M. C. *et al.* Factor H binds to the hypervariable region of many *Streptococcus pyogenes* M proteins but does not promote phagocytosis resistance or acute virulence. *PLoS Pathog.* **9,** e1003323 (2013).
29. Ermert, D. *et al.* Binding of complement inhibitor C4b-binding protein to a highly virulent *Streptococcus pyogenes* M1 strain is mediated by protein H and enhances adhesion to and invasion of endothelial cells. *J. Biol. Chem.* **288,** 32172–32183 (2013).
30. Zhou, T. *et al.* Structural definition of a conserved neutralization epitope on HIV-1 gp120. *Nature* **445,** 732–737 (2007).
31. Sui, J. *et al.* Structural and functional bases for broad-spectrum neutralization of avian and human influenza A viruses. *Nat. Struct. Mol. Biol.* **16,** 265–273 (2009).
32. Ekiert, D. C. *et al.* Antibody recognition of a highly conserved influenza virus epitope. *Science* **324,** 246–251 (2009).
33. McLellan, J. S. *et al.* Structure of HIV-1 gp120 V1/V2 domain with broadly neutralizing antibody PG9. *Nature* **480,** 336–343 (2011).
34. Steer, A. C., Law, I., Matatolu, L., Beall, B. W. & Carapetis, J. R. Global *emm* type distribution of group A streptococci: systematic review and implications for vaccine development. *Lancet Infect. Dis.* **9,** 611–616 (2009).
35. Doublie, S. Production of selenomethionyl proteins in prokaryotic and eukaryotic expression systems. *Methods Mol. Biol.* **363,** 91–108 (2007).
36. Kabsch, W. XDS. *Acta Crystallogr. D* **66,** 125–132 (2010).
37. Otwinowski, Z. & Minor, W. Processing of X-ray diffraction data collected in oscillation mode. *Methods Enzymol.* **276,** 307–326 (1997).
38. Adams, P. D. *et al.* PHENIX: a comprehensive Python-based system for macromolecular structure solution. *Acta Crystallogr. D* **66,** 213–221 (2010).
39. Emsley, P. & Cowtan, K. Coot: model-building tools for molecular graphics. *Acta Crystallogr. D* **60,** 2126–2132 (2004).
40. Chen, V. B. *et al.* MolProbity: all-atom structure validation for macromolecular crystallography. *Acta Crystallogr. D* **66,** 12–21 (2010).
41. Schneider, C. A., Rasband, W. S. & Eliceiri, K. W. NIH Image to ImageJ: 25 years of image analysis. *Nat. Methods* **9,** 671–675 (2012).
42. 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).
43. 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).
44. Case, D. A. *et al. AMBER 2015* (Univ. California, San Francisco, 2015).
45. 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).
46. 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).
47. 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).
48. 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).
49. Essmann, U. *et al.* A smooth particle mesh Ewald method. *J. Chem. Phys.* **103,** 8577–8593 (1995).
50. Nelson, M. T. *et al.* NAMD: a parallel, object oriented molecular dynamics program. *Int. J. Supercomput. Appl.* **10,** 251–268 (1996).
51. Phillips, J. C. *et al.* Scalable molecular dynamics with NAMD. *J. Comput. Chem.* **26,** 1781–1802 (2005).
52. Izaguirre, J. A., Catarello, D. P., Wozniak, J. M. & Skeel, R. D. Langevin stabilization of molecular dynamics. *J. Chem. Phys.* **114,** 2090–2098 (2001).
53. Jiang, W. *et al.* High-performance scalable molecular dynamics simulations of a polarizable force field based on classical Drude oscillators in NAMD. *J. Phys. Chem. Lett.* **2,** 87–92 (2011).
54. 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).
55. Caves, L. S. D., Evanseck, J. D. & Karplus, M. Locally accessible conformations of proteins: multiple molecular dynamics simulations of crambin. *Protein Sci.* **7,** 649–666 (1998).
56. 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).
57. Humphrey, W., Dalke, A. & Schulten, K. VMD visual molecular dynamics. *J. Mol. Graph. Model.* **14,** 33–38 (1996).
58. 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).

Acknowledgements

We thank O. Ghosh for help on the project. This works was supported by National Institutes of Health (NIH) grant T32 GM007240 (C.Z.B.), American Heart Association

Predoctoral Fellowship 14PRE18320032 (C.Z.B.), NIH R01 AI096837 (P.G. and V.N.) and NIH R01 AI077780 (V.N.). The work was also funded in part by the National Biomedical Computation Resource, NIH P41 GM103426, NIH Director's New Innovator Award Program DP2-OD007237 and through the National Science Foundation XSEDE Supercomputer Resources Grant RAC CHE060073N to R.E.A. S.P.H. was supported by the Interfaces Multi-Scale Analysis of Biological Structure and Function training grant NIH T32 EB009380.

Author contributions

C.Z.B., V.N. and P.G. conceived the experiments. C.Z.B., A.J.B.-S. and T.B. carried out the structure determinations. C.Z.B. and A.J.B.-S. carried out the binding studies. S.P.H. and R.E.A. carried out and analysed the MD simulations. C.Z.B. and P.G. wrote the paper with input from all the authors.

Additional information

Supplementary information is available for this paper. Reprints and permissions information is available at www.nature.com/reprints. Correspondence and requests for materials should be addressed to P.G.

Competing interests

The authors declare no competing financial interests.