**Title**

Conservation of C4BP-binding sequence patterns in Streptococcus pyogenes M and Enn proteins

**Permalink**

https://escholarship.org/uc/item/7n11d923

**Journal**

Journal of Biological Chemistry, 300(7)

**ISSN**

0021-9258

**Authors**

Kolesiński, Piotr

McGowan, Matthew

Botteaux, Anne

et al.

**Publication Date**

2024-07-01

**DOI**

10.1016/j.jbc.2024.107478

Peer reviewed

# Conservation of C4BP-binding sequence patterns in *Streptococcus pyogenes* M and Enn proteins

**Piotr Kolesiński**[1] , **Matthew McGowan**[1] , **Anne Botteaux**[2] , **Pierre R. Smeesters**[2,3], and **Partho Ghosh**[1,*]

*From the* [1]*Department of Chemistry and Biochemistry, University of California, San Diego La Jolla, California, USA;* [2]*Molecular Bacteriology Laboratory, European Plotkin Institute for Vaccinology, ULB, Brussels, Belgium;* [3]*Department of Paediatrics, Brussels University Hospital, Academic Children Hospital Queen Fabiola, Université libre de Bruxelles, Brussels, Belgium*

Reviewed by members of the JBC Editorial Board. Edited by Ursula Jakob

Antigenically sequence variable M proteins of the major bacterial pathogen *Streptococcus pyogenes* (Strep A) are responsible for recruiting human C4b-binding protein (C4BP) to the bacterial surface, which enables Strep A to evade destruction by the immune system. The most sequence divergent portion of M proteins, the hypervariable region (HVR), is responsible for binding C4BP. Structural evidence points to the conservation of two C4BP-binding sequence patterns (M2 and M22) in the HVR of numerous M proteins, with this conservation applicable to vaccine immunogen design. These two patterns, however, only partially explain C4BP binding by Strep A. Here, we identified several M proteins that lack these patterns but still bind C4BP and determined the structures of two, M68 and M87 HVRs, in complex with a C4BP fragment. Mutagenesis of these M proteins led to the identification of amino acids that are crucial for C4BP binding, enabling formulation of new C4BP-binding patterns. Mutagenesis was also carried out on M2 and M22 proteins to refine or generate experimentally grounded C4BP-binding patterns. The M22 pattern was the most prevalent among M proteins, followed by the M87 and M2 patterns, while the M68 pattern was rare. These patterns, except for M68, were also evident in numerous M-like Enn proteins. Binding of C4BP *via* these patterns to Enn proteins was verified. We conclude that C4BP-binding patterns occur frequently in Strep A strains of differing M types, being present in their M or Enn proteins, or frequently both, providing further impetus for their use as vaccine immunogens.

*Streptococcus pyogenes* (group A *Streptococcus* or Strep A) is a widespread gram-positive bacterial pathogen that is responsible for morbidity and mortality on a global scale (1). While infection by Strep A usually results in mild conditions (*e.g.*, pharyngitis or impetigo), a significant number of cases progress to acute invasive disease, such as necrotizing fasciitis or streptococcal toxic shock syndrome, which have high mortality rates (2). In addition, prolonged Strep A infection can lead to severe autoimmune sequelae (*e.g.*, acute rheumatic

fever) (3, 4), which continues to afflict populations in developing nations (5).

Strep A is highly proficient at avoiding phagocytic killing by the immune system. A frequent means by which Strep A achieves this is by recruiting soluble human C4b-binding protein (C4BP, ~570 kDa) to the bacterial surface (6–8). Recruitment of C4BP inhibits deposition of the complement system opsonin C3b (by the classical and lectin pathways) on the bacterial surface, thereby antagonizing clearance of Strep A by phagocytic immune cells. A primary recruiter of C4BP is the essential virulence factor M protein (9, 10), which is antigenically sequence variable, with >220 M types having been identified (11). Despite this variability, C4BP binding is pervasive in Strep A, with one study identifying C4BP binding in 90 of 100 strains of differing M types (9).

M proteins form parallel, dimeric α-helical coiled coils which are attached to the bacterial cell wall and extend ~500 Å into the extracellular space (12). The N-terminal 50 amino acids of mature M proteins (*i.e.*, with the signal sequence cleaved) extend furthest into the extracellular space and are especially variable (called the hypervariable region, HVR). The HVR defines the M type. Surprisingly, the HVRs of several M types, despite lacking sequence conservation, are responsible for binding C4BP (6, 9, 13, 14). One mode of interaction between sequence-divergent M protein HVRs and C4BP was revealed by X-ray crystallography (10). Structures of four different M protein HVRs (M2, M22, M28, and M49) bound to a C4BP fragment (composed of the α1 and α2 domains, C4BPα1-2) revealed a common 3D arrangement of M protein amino acids that interact with a common set of C4BP amino acids. The C4BPα2 domain supplies four sites forming a quadrilateral composed of a (1) hydrophobic pocket (denoted ΦHis 67) that binds large hydrophobic M protein amino acids; (2) the main chain amide of H67 (denoted N-His 67) that forms a hydrogenbond with an M protein amino acid; and (3) R64 and (4) R66 (Fig. 1*A*). Arginine is versatile in providing a polar contact through its guanidinium group or hydrophobic interaction through its alkyl carbons. C4BPα2 R64 and R66 form polar contacts with M protein amino acids (D, E, or N) in M2, M22, and M28 HVRs, but in the M49 HVR, R64 forms hydrophobic contacts with hydrophobic M49 amino acids. The C4BPα1 domain provides a single site, the Arg39 nook, which
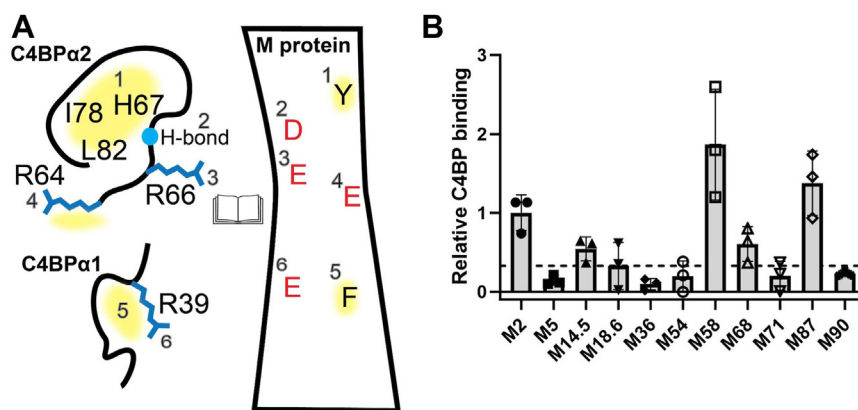
**Figure 1. Binding of C4BP to M proteins.** *A*, schematic of C4BP sites (*left*) and M protein amino acids (*right*) involved in interaction for the M2 and M22 patterns, in open book format. The M protein amino acids do not correspond to any specific M protein but show interactions common to M2, M22, M28, and M49 HVRs. *Yellow shading* corresponds to hydrophobic interactions. *B*, binding of soluble $His_6$-$M^{N100}$ protein constructs to immobilized intact C4BP, as evaluated by ELISA. Bound $His_6$-$M^{N100}$ protein constructs were detected with horseradish peroxidase-conjugated anti-$His_6$ antibodies. Values were normalized by the value of C4BP binding to $His_6$-$M2^{N100}$. Data from three biological replicates are presented with means and SDs. The *dotted line* shows the normalized value for C4BP binding to $His_6$-$M5^{N100}$ plus three SDs. C4BP, C4b-binding protein; HVR, hypervariable region.

consists of (5) a hydrophobic pocket formed by the main and side chain of R39, which interacts with a large hydrophobic M protein amino acid. In the case of the M2 and M49 HVRs, R39 also forms (6) salt bridges with negatively charged M protein amino acids.

While the C4BP-interacting amino acids in all 4 M protein HVRs are positioned similarly in 3D, they differ in their heptad positions. Coiled-coil proteins have diagnostic heptads (*abcdefg*) in their primary sequence, in which the *a* and *d* positions are typically small hydrophobic amino acids that form the hydrophobic core of the structure. The heptad pattern of C4BP-binding amino acids is similar for the M2 and M49 HVRs and defines the M2 pattern for C4BP binding, while the heptad pattern is similar for the M22 and M28 HVRs and defines the M22 pattern. The M2 and M22 patterns are evident in numerous M proteins, providing an explanation for how many Strep A strains bind C4BP (9, 10). The conserved M2 and M22 patterns occur in the midst of hypervariability and were thus unrecognized prior to structural studies.

Conservation of the C4BP-binding 3D pattern suggested its utility as a vaccine immunogen. Currently there is no vaccine against Strep A (15), with the sequence variability of M proteins being one of the major impediments (16–20). We found that an immunogen consisting of the C4BP-binding 3D pattern of M2 protein is capable of eliciting antibodies that react with M2 protein and are cross-reactive against other M protein types sharing the 3D pattern but not against those lacking it (21). These cross-reactive antibodies promote the opsonophagocytic killing of Strep A strains (21).

The M2 and M22 patterns only partially explain C4BP-binding. More than half of the 90 Strep A strains shown to bind C4BP lack an M2 or M22 pattern (9, 10). We therefore asked whether additional C4BP-binding patterns exist in M proteins. We identified C4BP-binding in several M proteins that lack an M2 or M22 pattern, and determined the structures of two, M68 and M87 protein HVRs, in complex with C4BPα1-2. Based on mutagenesis of these M proteins and effects on C4BP binding, we constructed M87 and M68

C4BP-binding patterns. Similar mutagenesis and C4BP-binding experiments were carried out for M2 and M22 proteins, enabling refinement or generation of experimentally based M2 and M22 patterns, respectively. The M2, M22, and M87 patterns were also identified and verified in numerous M-like Enn proteins. Our results indicate that C4BP-binding patterns are highly prevalent in M and Enn proteins.

## Results

### C4BP binding

We tested 9 M proteins that lack an M2 or M22 pattern for C4BP binding; these were all from strains shown to bind C4BP (9, 10). The M proteins were M14.5, M18.6, M36, M54, M58, M68, M71, M87, and M90. $His_6$-tagged constructs consisting of the N-terminal 100 amino acids of these M proteins (designated by superscripted "N100") were expressed in *Escherichia coli* and purified, and assayed for binding to intact C4BP by ELISA (Fig. 1*B*). The binding of $His_6$-$M2^{N100}$ to C4BP was used for normalization. $His_6$-$M5^{N100}$ was used as a negative control, as M5 protein does not bind C4BP (21). We distinguished binders from nonbinders using a threshold of three SDs above the C4BP-binding of $His_6$-$M5^{N100}$. $His_6$-tagged $M14.5^{N100}$, $M58^{N100}$, $M68^{N100}$, and $M87^{N100}$ crossed this threshold, while the other five of the nine did not (Fig. 1*B*). Cocrystallization with C4BPα1-2 was pursued with the four C4BP-binding $M^{N100}$ protein constructs, and $M68^{N100}$ and $M87^{N100}$ yielded cocrystals that diffracted sufficiently for atomic-level modeling of the binding interface.

### M87–C4BP interactions

The structure of the complex between $M87^{N100}$ and C4BPα1-2 was determined by molecular replacement to 2.7 Å resolution limit (Table S1). The structure revealed the same global architecture observed previously (10), with two C4BPα1-2 molecules bound by dimeric, α-helical $M87^{N100}$ (Figs. 2*A* and S2). C4BPα1-2 was unchanged in conformation (rmsd 0.74–1.18 Å) (10), although certain loops distant from
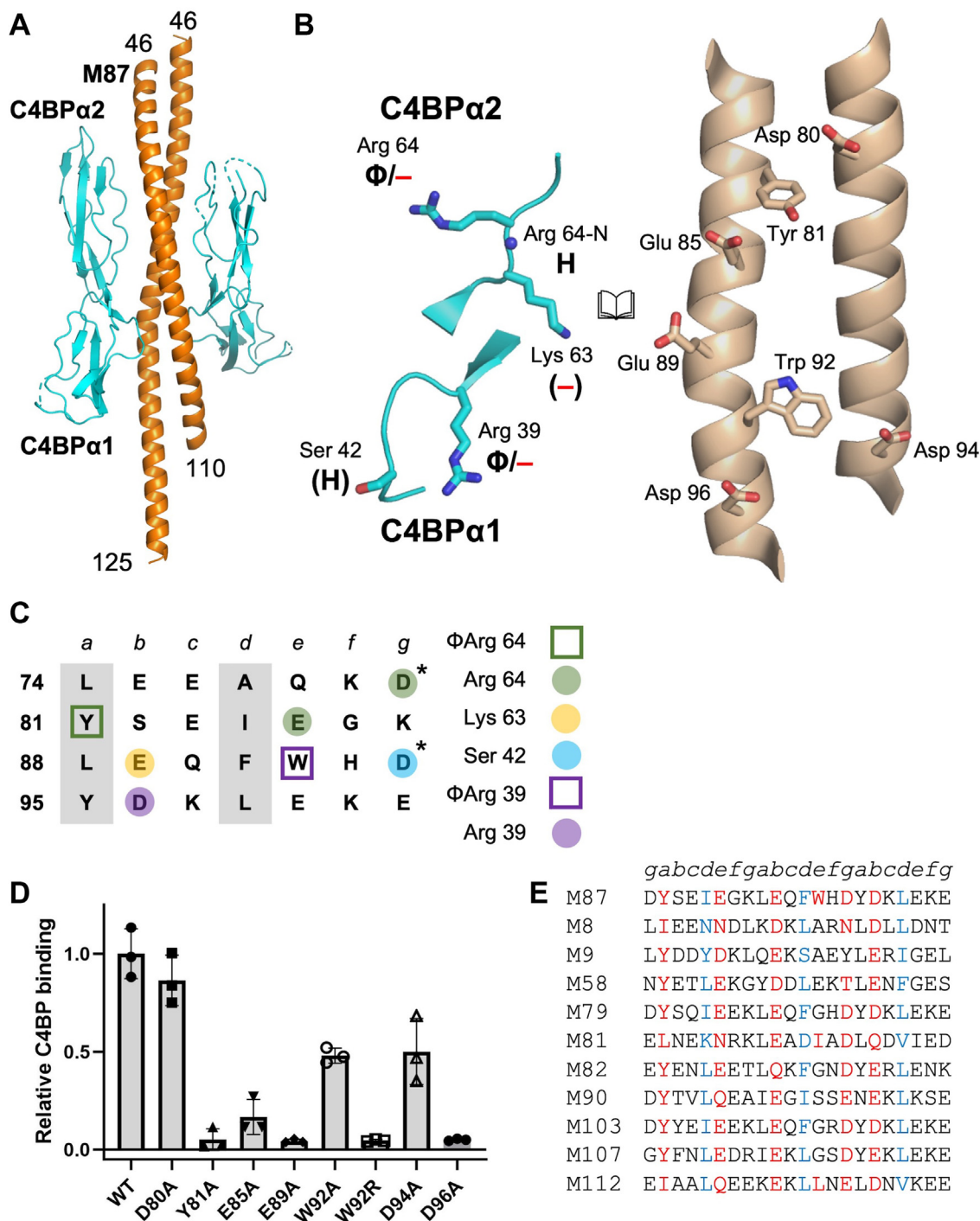
**Figure 2. M87–C4BP interactions.** *A*, structure of M87[N100] (*orange*) in complex with C4BPα1-2 (*cyan*), in *cartoon representation*. Loops of C4BPα1-2 that were not modeled are shown as *dotted lines*. *B*, interaction between C4BPα1-2 (*left*) and M87 (*right*), in *cartoon* (main chain) or *bonds* (side chains) representation, shown in open book format. Carbons are *cyan* for C4BPα1-2 and *wheat* for M87; for both, nitrogens are *blue* and oxygens are *red*. Chemical character of interactions denoted as follows: Φ, hydrophobic; —, negative; and H, hydrogen bond. Contacts that exceeded hydrogen-bonding or salt-bridging limits (≤3.5 and ≤4.0 Å, respectively) have these labels in *parentheses*. *C*, heptad register of M87 74-101 (*a* and *d* positions in *gray columns*). Amino acids marked *with an asterisk* belong to the minor helix. *Shapes* indicate type of contact (*square* for hydrophobic, Φ; *circle* for polar), and *coloring* indicates which C4BP amino acid (depicted in legend to the *right*) was contacted. *D*, binding of soluble, intact His[6]-M87 protein constructs to immobilized intact C4BP evaluated by ELISA. Bound His[6]-M87 protein was detected with horseradish peroxidase-conjugated anti-His[6] antibodies. Values were normalized by the value of C4BP-binding to WT His[6]-M87. Data from three biological replicates are presented with means and SDs. *E*, partial list of sequences of M proteins with M87 C4BP-binding patterns. Heptad positions are indicated *above* the sequences. *Red* indicates an observed (M87) or potential (other M proteins) C4BP-binding amino acid, and *blue* indicates a *d* heptad position. A full list is in Table S2. C4BP, C4b-binding protein.

the interface with M87 were flexible and not modeled. The two M87 helices (aa 46–110 and 46–125) were asymmetric, with one less coiled than the other. Nevertheless, the ~2.5 heptads

of M87 protein from D80 to D96 that bound C4BP were closely superimposable (0.68 Å rmsd) (Fig. S1). This section was stabilized by knobs-into-hole packing, except at Y81 (*a*

position), which faced outward from the coiled-coil core and contacted C4BP (Fig. 2, *B* and *C*). The two M87-C4BP interfaces were similar (0.57 Å rmsd) but not identical. The buried surface area for one interface was 1498 Å$^2$ and 956 Å$^2$ for the other. Closer inspection revealed that the higher value was due to amino acids that were not in contact (>4 Å apart) but nevertheless excluded water at their interface. The buried surface areas considering only the amino acids that were in contact distance (≤4 Å) were more similar, 583 and 600 Å$^2$. The surface complementarities of the two interfaces were 0.61 and 0.69, similar to values observed for other M protein HVR–C4BP complexes (10).

The interface was dominated by negatively charged M protein amino acids interacting with positively charged C4BP ones, supplemented with a few hydrophobic contacts (Fig. 2, *B* and *C*). Each C4BPα1-2 molecule contacted predominantly one of the two M87 helices, which we termed the "major" helix and the other the "minor" helix. The C4BPα2 quadrilateral (Fig. 1*A*) as a whole was not involved in interaction, except for R64, which formed hydrophobic contacts using its aliphatic chain with M87 Y81 of the major helix and a salt bridge with D80 of the minor helix; Y81 was the only core *a* or *d* position amino acid involved in contacts. In one of the interfaces, the main chain amide of C4BPα2 R64 formed a hydrogen bond with M87 E85 of the major helix. C4BPα2 K63, which was not observed previously to have a role in M protein interaction (10), was within near-salt bridging distance of M87 E89 of the major helix. In contrast to the C4BPα2 quadrilateral, the C4BPα1 Arg39 nook had the same role as previously observed in the M2 pattern (10), forming hydrophobic contacts and a salt bridge to M protein. Here, the former was to M87 W92 of the major helix and the latter to D96 of the minor helix. In one of the interfaces, the main chain carbonyl of C4BPα1 R39 also formed a hydrogen bond to M87 W92 of the major helix. Lastly, C4BPα1 Ser42 was proximal to M87 D94 of the minor helix and again with slight rearrangement could form a hydrogen bond to this amino acid.

The contribution of M87 amino acids to C4BP binding was evaluated by Ala- or other-substitution mutagenesis. Intact M87 protein was used and binding of His-tagged versions of WT or Ala-substituted M87 protein to intact C4BP was quantified by ELISA. Relative to WT M87 protein, Y81A, E89A, and D96A bound C4BP ~21-fold less well on average, which indicated that the hydrophobic contact to C4BP R64 and polar contacts to K63 and R39, respectively, were crucial (Fig. 2*D*). In M2 protein, the hydrophobic contact to the C4BPα1 Arg39 nook is essential (*via* F75) (10) but here C4BP binding decreased only 2-fold for M87 W92A. The M87 W92R substitution had a more dramatic effect, reducing C4BP binding 22-fold, likely due to Arg-Arg charge repulsion. Contact to C4BPα2 R64 by M87 E85 was intermediate in importance, as the M87 E85A substitution resulted in only a 6-fold decrease in C4BP binding. The two contacts from the minor M87 helix, D80 and D94, were even less consequential, with C4BP binding in D80A unchanged and in D94A decreased only 2-fold. Loss of C4BP binding for intact M87

Y81A, E85A, E89A, and D96A was not due to loss of structure or stability, as evaluated by CD spectra and melting curves of these mutant proteins compared to WT intact M87 protein (Fig. S3). M87 W92R was previously shown to have the same structure and stability as WT M87 protein (22).

We next scored a collection of 179 M protein sequences for patterns that matched the C4BP-binding motif identified structurally and functionally for M87 protein. Sequences that had identical or chemically similar amino acids corresponding to the essential M87 amino acids Y81, E85, E89, and D96 were scored highly, while sequences lacking an identical or chemically similar amino acid at any of these positions were eliminated. Ala substitution of M87 W92 and D94 had an intermediate effect on C4BP binding, and thus sequences with identical or chemically similar amino acids at these positions were scored positively but not as highly as the essential M87 amino acids. Other amino acids were tolerated at these latter positions, except for positively charged ones at the position corresponding to W92, and hydrophobic ones that cannot form hydrogen bonds at the position corresponding to D94. While M87 D80 appeared not to influence C4BP binding, this amino acid was located next to the important C4BPα2 amino acid R64, and thus sequences with a positively charged amino acid at this position were eliminated. The proper heptad position of these amino acids was also evaluated by scoring the *a* and *d* positions of three consecutive heptads. However, due to the often irregular heptads of M proteins (23, 24), only four *a* or *d* positions out of six in the three heptads needed to have coiled-coil–stabilizing amino acids (25, 26). We noticed that this scoring scheme produced a rank-ordered list that included M24 protein near the bottom (Table S2). As the M24 strain does not bind C4BP (9), we used the score for M24 protein as a cut-off and included only M proteins that scored higher. This cut-off resulted in the identification of 19 M proteins, including M87, as having an M87 pattern for C4BP binding (Fig. 2*E* and Table S2). M proteins are assigned to two clades, X and Y, with C4BP-binding assorting to clade X (27). Eleven of the 19 M proteins in the M87 pattern belong to a strain that binds C4BP, with all 11 belonging to clade X (9), whereas the other eight have not been evaluated for C4BP binding, with most of these (6 of 8) belonging to clade X.

### M68–C4BP interactions

The structure of the complex between M68$^{N100}$ and C4BPα1-2 was determined to 2.5 Å resolution limit by multiwavelength anomalous dispersion phasing using seleno-methionine (SeMet)-labeled M68$^{N100}$ (Table S1). The architecture of the complex differed substantially from that of M87 protein and those observed previously (10). M68$^{N100}$ formed an α-helical dimer but not a coiled coil (Figs. 3*A* and S4). Nearly the full extent of M68$^{N100}$ was visualized in one of the helices (aa 42–135, ~140 Å long), albeit with some breaks in main chain density, while the other helix spanned aa 52 to 127 without any breaks. The two M68 helices were asymmetrically disposed and mostly parallel to one other, except for the very
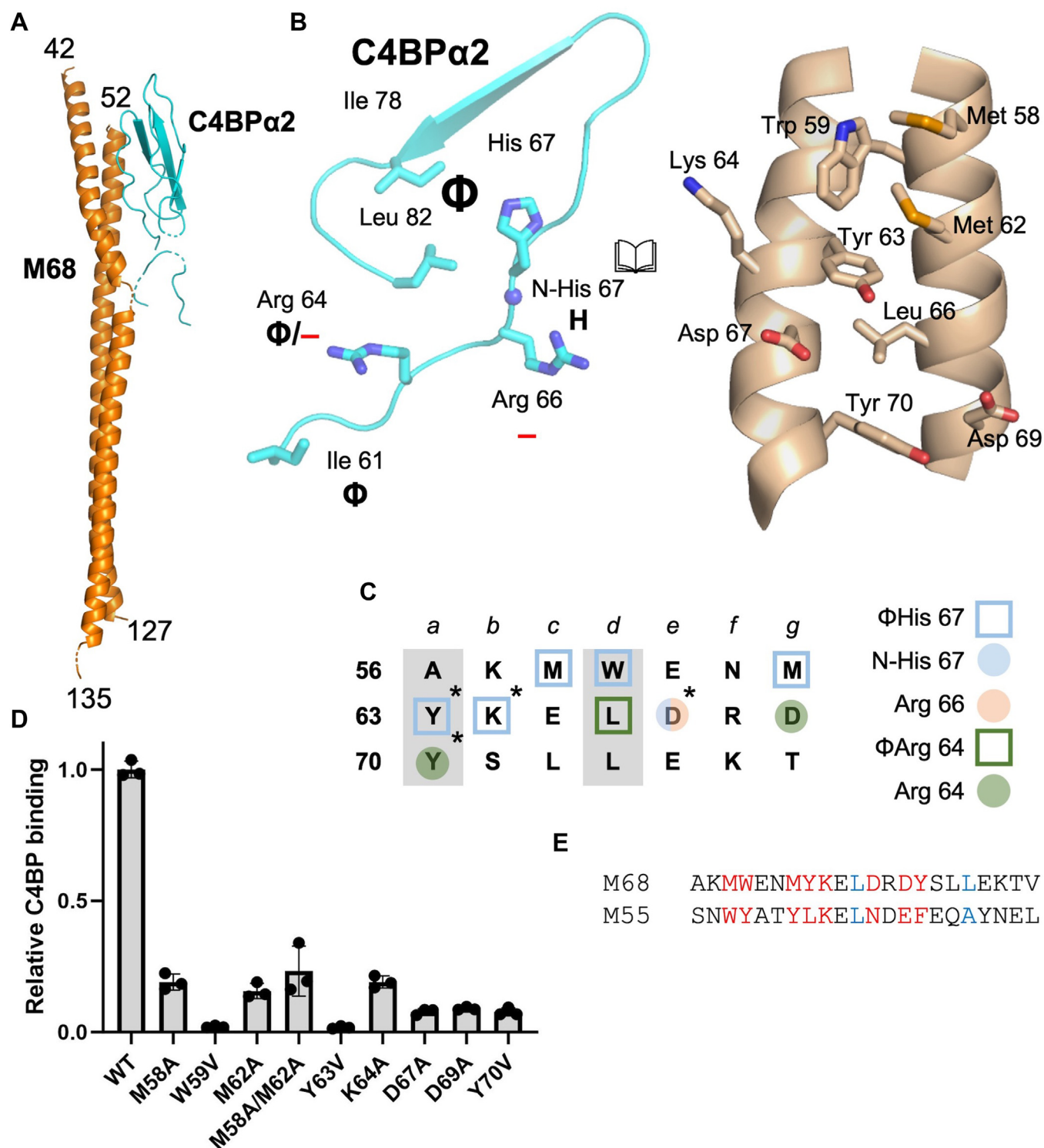
**Figure 3. M68–C4BP interactions.** *A,* structure of M68 (*orange*) in complex with C4BPα2 (*cyan*), in *cartoon representation*. *B,* interaction between C4BPα1-2 (*left*) and M68 (*right*), in *cartoon* (main chain) or *bonds* (side chains) *representation*, shown in open book format. Carbons are *cyan* for C4BPα1-2 and *wheat* for M87; for both, nitrogens are *blue* and oxygens are *red*. Chemical character of interactions denoted as follows: Φ, hydrophobic; —, negative; and H, hydrogen bond. *C,* heptad register of M68 56-76 (*a* and *d* positions in *gray columns*). Amino acids marked *without an asterisk* belong to one helix, and those *with an asterisk* to the second helix. *Shapes* indicate type of contact (*square* for hydrophobic; *circle* for polar), and *coloring* indicates which C4BP amino acid (depicted in legend to the *right*) was contacted. *D,* binding of soluble, intact WT and mutant $His_6$-M68 protein to immobilized intact C4BP, as evaluated by ELISA. Bound $His_6$-M68 protein was detected with horseradish peroxidase-conjugated anti-$His_6$ antibodies. Values were normalized by the value of C4BP binding to WT $His_6$-M68. Data from three biological replicates are presented with means and SDs. *E,* M protein sequences belonging to the M68 pattern of C4BP binding. Heptad positions are indicated above the sequences. *Red* indicates an observed or potential C4BP-binding amino acid, and *blue* indicates a *d* heptad position. C4BP, C4b-binding protein.

![ASBMB]

*J. Biol. Chem.* (2024) 300(7) 107478 **5**

N-terminal segment (aa 42–77), where one of the helices wrapped around the other. About two heptads of this N-terminal segment (aa 58–70) bound C4BP. Only a single C4BPα1-2 molecule was evident, and in this molecule the C4BPα1 domain was largely disordered. Both M68 helices contacted C4BPα2, with 1330 Å$^2$ of surface area being buried, which is typical of M protein HVR–C4BP complexes (10). The shape complementarity between the two proteins was 0.58, which is also typical (10).

The C4BPα2 quadrilateral contacted M68 protein, although with some variations (Fig. 3, *B* and *C*). The C4BPα2 quadrilateral hydrophobic pocket (ΦHis 67), instead of contacting a single hydrophobic M protein amino acid (10), contacted a cluster: M68 Met 58, Trp 59, and Met 62 from one helix and Tyr 63, along with the alkyl chain of K64, from the second. The N-His 67 site formed a hydrogen bond with M68 D67 from the second helix. This same M87 amino acid formed a salt bridge with C4BPα2 R66. The other Arg in the quadrilateral, C4BPα2 R64, formed several interactions: a salt bridge to M68 D69 from the first helix, a hydrophobic contact to L66 from the first helix, and a π–cation interaction with M68 Y70 from the second helix. M68 Y70 also made a hydrophobic contact with C4BPα2 Ile61. Notably, all of the aromatic C4BP-contacting amino acids of M87 (W59, Y63, Y70) occupied the *a* or *d* heptad position (Fig. 3*C*).

The contribution of M68 amino acids to C4BP-binding was evaluated using intact M68 protein and intact C4BP, as described above for M87 protein. The contacts to the ΦHis 67 hydrophobic site were especially important. Valine substitution of M68 W59 or Y63 reduced C4BP binding ∼50-fold relative to WT M68 protein (Fig. 3*D*). Valine was chosen since these amino acids occupied the *a* or *d* heptad position. Valine is a stabilizing amino acid at the core *a* and *d* coiled-coil positions, whereas Ala is not (25, 26). Ala substitution of Met 58, Met 62, or Lys 64, which also contacted the ΦHis 67 hydrophobic pocket, reduced C4BP binding by ∼6-fold. Likewise, dual-Ala substitution of M87 Met 58 and Met 62 reduced C4BP binding by ∼4-fold. Ala or Val substitution of M68 amino acids that contacted C4BPα2 R66 (M68 D67A) or R64 (M87 D69A and Y70V) reduced C4BP binding by ∼12-fold. As validation of the structure of this unusual complex, we examined the structure and stability of M68 M58A/M62A and D67A and found that their decreased C4BP binding was a direct functional effect rather than an indirect effect on structure or stability (Fig. S5).

As for M87 protein, we scored M protein sequences for a pattern of C4BP-contacting amino resembling that of M68 protein and for coiled-coil–stabilizing amino acids at most *a* and *d* positions. Notably, M68 has three aromatic amino acids that are significant for C4BP binding (W59, Y63, and Y70) at *a* or *d* positions. No other M protein in the set of 179 matched this pattern. If the restriction that these positions are occupied by only aromatic acids is relaxed to include hydrophobic ones larger than valine, then M55 protein is included in having an M68 pattern (Fig. 3*E* and Table S2). An M55 Strep A strain binds C4BP (9). M68 protein belongs to C4BP-binding clade X, whereas M55 protein is assigned to clade Y but is an outlier in

this scheme (27). These results indicate that the unusual mode of interaction between C4BP and M68 protein is rare among M proteins.

### M2 and M22 patterns

While contributions of M2 amino acids to C4BP binding have been evaluated by mutagenesis previously (10), this prior study used a coprecipitation assay, unlike the ELISA used here for M87 and M68 proteins. To provide a common means for comparison, we examined C4BP binding of Ala-substitution mutants of M2 protein by ELISA. We focused on three amino acids whose substitutions by Ala only slightly decreased C4BP binding in the coprecipitation assay (10). These were M2 H61, which contacts the ΦHis 67 hydrophobic pocket; D62, which forms a hydrogen bond to N-His 67; and E68, which forms a salt bridge with C4BPα2 R64 (Fig. 4*A*). Ala substitution of these amino acids in intact M2 protein led to ∼10- to 30-fold loss of C4BP binding compared to intact WT M2 protein (Fig. 4*B*). A number of M2 amino acids were previously shown by the coprecipitation assay to be significant for C4BP binding (10). These are F75, which contacts the C4BPα1 Arg39 nook, and E76 and D79, which form salt bridges with the C4BPα1 Arg39 nook, with only one of these salt bridges being required (Fig. 4*A*). We constructed a scoring scheme for the M2 pattern based on these results, in which we also included the possibility of hydrophobic contacts to C4BP2α R64, as seen in M49 protein (10). This resulted in the identification of 17 M proteins, including M2 and M49, which have the M2 motif (Fig. 3*C* and Table S2). Of these, nine are from Strep A strains shown to bind C4BP, with all nine belong to clade X (9), whereas the others have not been tested, with more than half of these (5 of 8) belonging to clade X (27).

The M22 pattern has not been probed by mutagenesis (Fig. 5*A*), and therefore we examined binding by WT or mutant intact M22 protein to intact C4BP by ELISA. Ala substitution of 5 M22 amino acids decreased C4BP binding substantially (Fig. 5*B*). The largest decrease was ∼50-fold for M22 Y77A, which makes a hydrophobic contact to the C4BPα1 Arg39 nook (10). Ala substitution of M22 I59, which makes hydrophobic contact to the C4BPα2 ΦHis 67 site, or M2 N60, which forms a hydrogen bond to C4BPα2 His 67, decreased C4BP binding by ∼26-fold. Ala substitution of M22 E65, which salt bridges with C4BPα2 R64, also decreased C4BP binding substantially (10-fold), whereas Ala substitution of M22 D64, which salt bridges with C4BPα2 R66, decreased it only slightly (∼3-fold). In M28 protein, the salt bridge to C4BPα2 R66 is made by a Glu in the following heptad (Fig. 5*A*, M22 E70). The equivalent position in M22 protein is also a Glu (E67). The M22 E67A substitution decreased C4BP binding only slightly (∼3-fold), just like M22 D64A, but the dual substitution M22 D64A/E67A decreased C4BP binding more substantially (∼13-fold), indicating that these two negatively charged M22 amino acids serve redundant roles. As above, we constructed a scoring scheme for the M22 pattern based on these results and also took into consideration that the
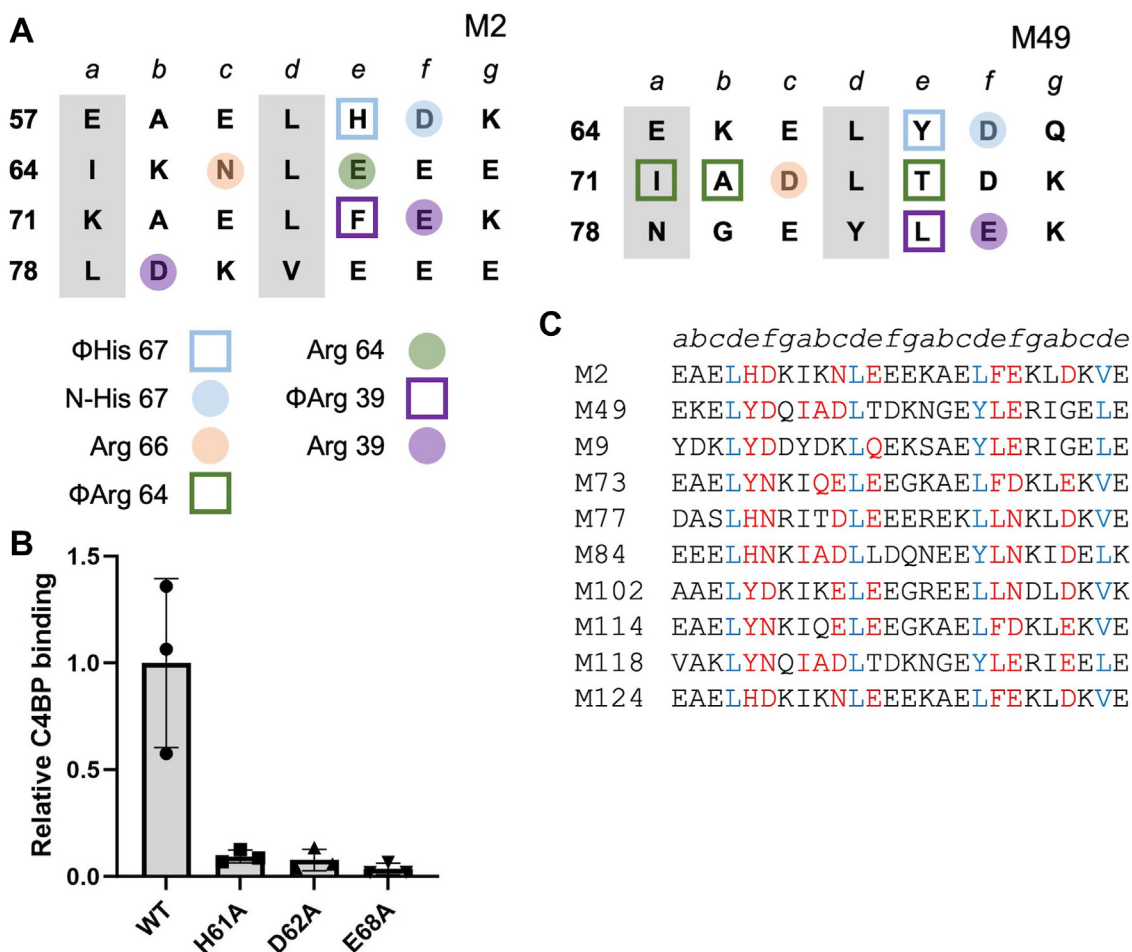
**Figure 4. M2–C4BP interactions.** *A*, heptad register of (*left*) M2 (aa 57–84) and (*right*) M49 (aa 64–84) (*a* and *d* positions in *gray columns*). *Shape* indicates type of contact (*square* for hydrophobic; *circle* for polar), and *coloring* indicates which C4BP amino acid (depicted in legend at *bottom*) was contacted. *B*, binding of soluble, intact His$_6$-M2 protein constructs to immobilized intact C4BP evaluated by ELISA. Bound His$_6$-M2 protein was detected with horseradish peroxidase-conjugated anti-His$_6$ antibodies. Values were normalized by the value of C4BP binding to WT His$_6$-M2. Data from three biological replicates are presented with means and SDs. *C*, partial list of M protein sequences belonging to the M2 pattern of C4BP binding. Heptad positions are indicated *above* the sequences. *Red* indicates an observed or potential C4BP-binding amino acid, and *blue* indicates a *d* heptad position. A full list is in Table S2. C4BP, C4b-binding protein.

hydrophobic contact to the C4BPα1 Arg39 nook could occur from a *b* heptad position as in M28 protein, rather than an *e* heptad position as in M22 protein (Fig. 5*A*). M12 appeared near the bottom of the rank-ordered list. As the M12 strain does not bind C4BP (9), we used the score for M12 protein as a threshold (Table S2). This resulted in 49 M proteins, including M22 and M28, with an M22 pattern (Fig. 5*C* and Table S2). Twenty three of these belong to M strains shown to bind C4BP, with all but one belonging to clade X (9), whereas the others have not been tested, with almost all (20 of 26) belonging to clade X (27).

### C4BP binding by Enn proteins

The M2, M22, M68, or M87 patterns defined by experimental measurements still left about half of the 90 Strep A strains shown to bind C4BP unaccounted for (9). The genomes of most these strains encode an M-like Enn protein downstream of *emm*. Enn proteins are predicted to form α-

helical coiled coils and are sequence variable but to a lesser extent than M proteins (28–30). For example, the pairwise identity of Enn proteins in a collection of 1688 globally distributed Strep A isolates is 69% compared to 45% for M proteins (29). Thus, we asked whether Enn proteins in such strains have C4BP-binding patterns. We identified an M2 pattern in 13 Enn proteins from nine different M type strains, an M22 pattern in 70 Enn proteins from 45 strains, and an M87 pattern from 30 Enn proteins in 20 strains (Fig. 6, *A–C* and Tables S3 and S4). The M2 pattern was found mainly in subgroup 5 (SG5) of Enn proteins (58%) and the M87 pattern mainly in SG1 (77%), while the M22 pattern was found across multiple subgroups (Table S3) (30). The M68 pattern continued to be rare, with no Enn proteins identified to have this pattern. To validate these observations, we expressed and purified intact Enn300, which is encoded in an M18.6 strain and has the M22 pattern, and Enn292, which is encoded by an M91 strain and can be assigned to either M2 or M22 patterns (Fig. 6*D*). Both intact

SASBMB

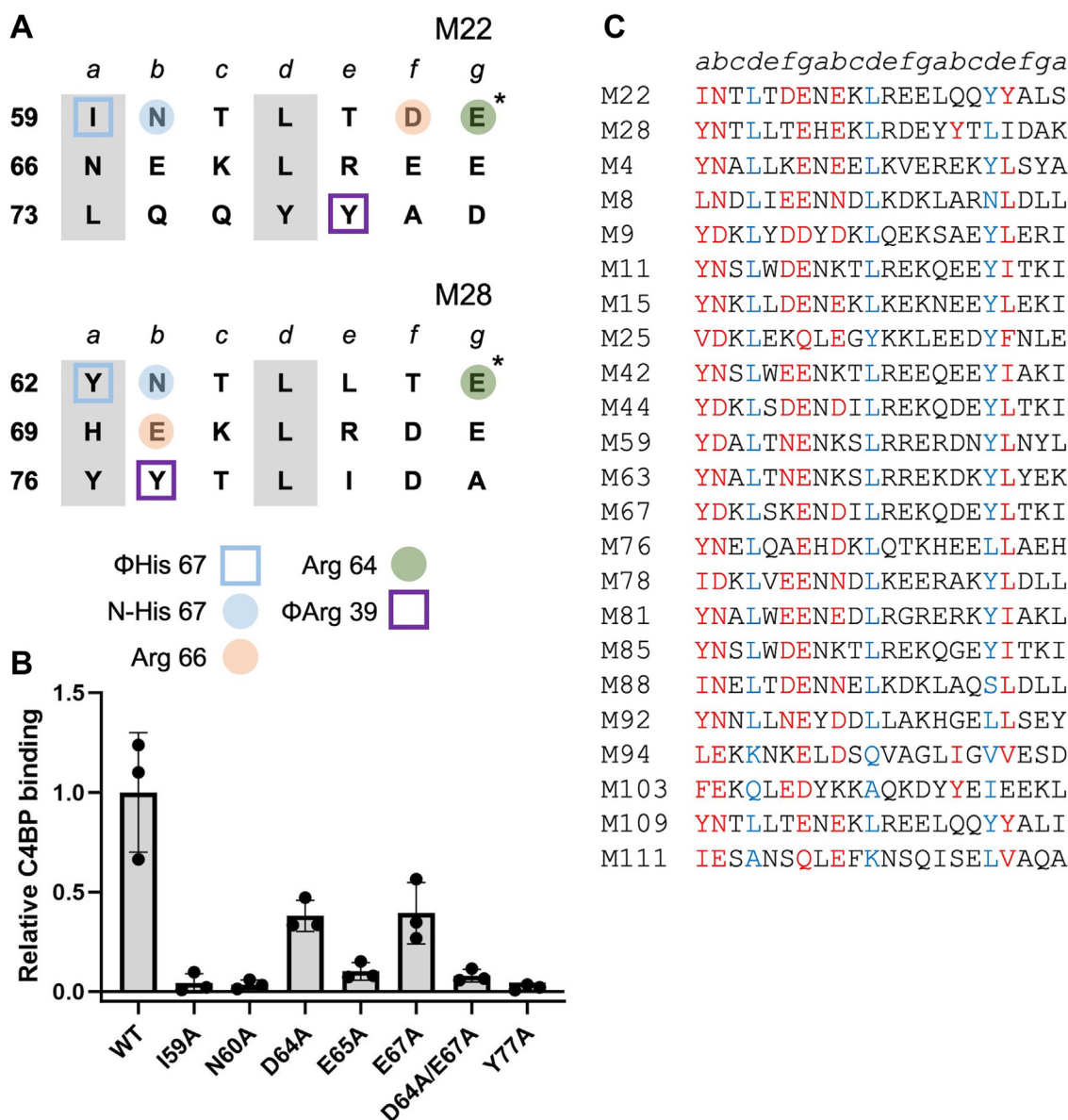*J. Biol. Chem.* (2024) 300(7) 107478 **7**

**Figure 5. M22–C4BP interactions.** *A*, heptad register of (*top*) M22 (aa 59–79) and (*bottom*) M28 (aa 62–82) (*a* and *d* positions in *gray columns*). Amino acids marked *with an asterisk* belong to the minor helix, and those *with an asterisk* to the major helix. *Coloring* indicates which C4BP amino acid (depicted in legend at *bottom*) was contacted, and *shape* indicates type of contact (*square* for hydrophobic; *circle* for polar). *B*, binding of soluble, intact $His_6$-M22 protein constructs to immobilized intact C4BP evaluated by ELISA. Bound $His_6$-M22 protein was detected with horseradish peroxidase-conjugated anti-$His_6$ antibodies. Values were normalized by the value of C4BP binding to WT $His_6$-M22. Data from three biological replicates are presented with means and SDs. *C*, partial list of M protein sequences belonging to the M22 pattern of C4BP binding. Heptad positions are indicated *above* the sequences. *Red* indicates an observed or potential C4BP-binding amino acid, and *blue* indicates a *d* heptad position. A full list is in Table S2. C4BP, C4b-binding protein.

Enn300 and En292 bound intact C4BP (Fig. 6*E*), while the corresponding M proteins did not (M18.6) or bound weakly (M91). Ala substitution of Enn300 N62 (equivalent to M22 N60), which is predicted to form a hydrogen bond to N-His 67, reduced C4BP binding by 4-fold relative to WT Enn300. Ala substitution of Enn292 D70, which is predicted to likewise form a hydrogen bond to N-His 67 in the M2 pattern or instead contact C4BPα2 Arg 66 in the M22 pattern, reduced C4BP binding by 7-fold relative to WT Enn292. These results directly demonstrated C4BP binding in Enn proteins and validated the assignment of C4BP-binding patterns to Enn proteins.

## Discussion

We sought to understand whether C4BP-binding patterns beyond the M2 and M22 patterns existed in M proteins. This was because somewhat less than half of 90 Strep A strains shown to bind C4BP have an M protein that could be assigned to either of these patterns (9, 10). To address this, several M proteins lacking an M2 or M22 pattern were identified to bind C4BP, and structures of the HVRs of two of these, M87 and M68, bound to a fragment of C4BP were determined. Neither M87 nor M68 HVRs precisely recapitulated the mode of C4BP interaction observed previously (*cf*. Fig. 1*A* with Figs. S2*C* and 4*C*) (10). For the M87 HVR, the C4BPα2 quadrilateral was not
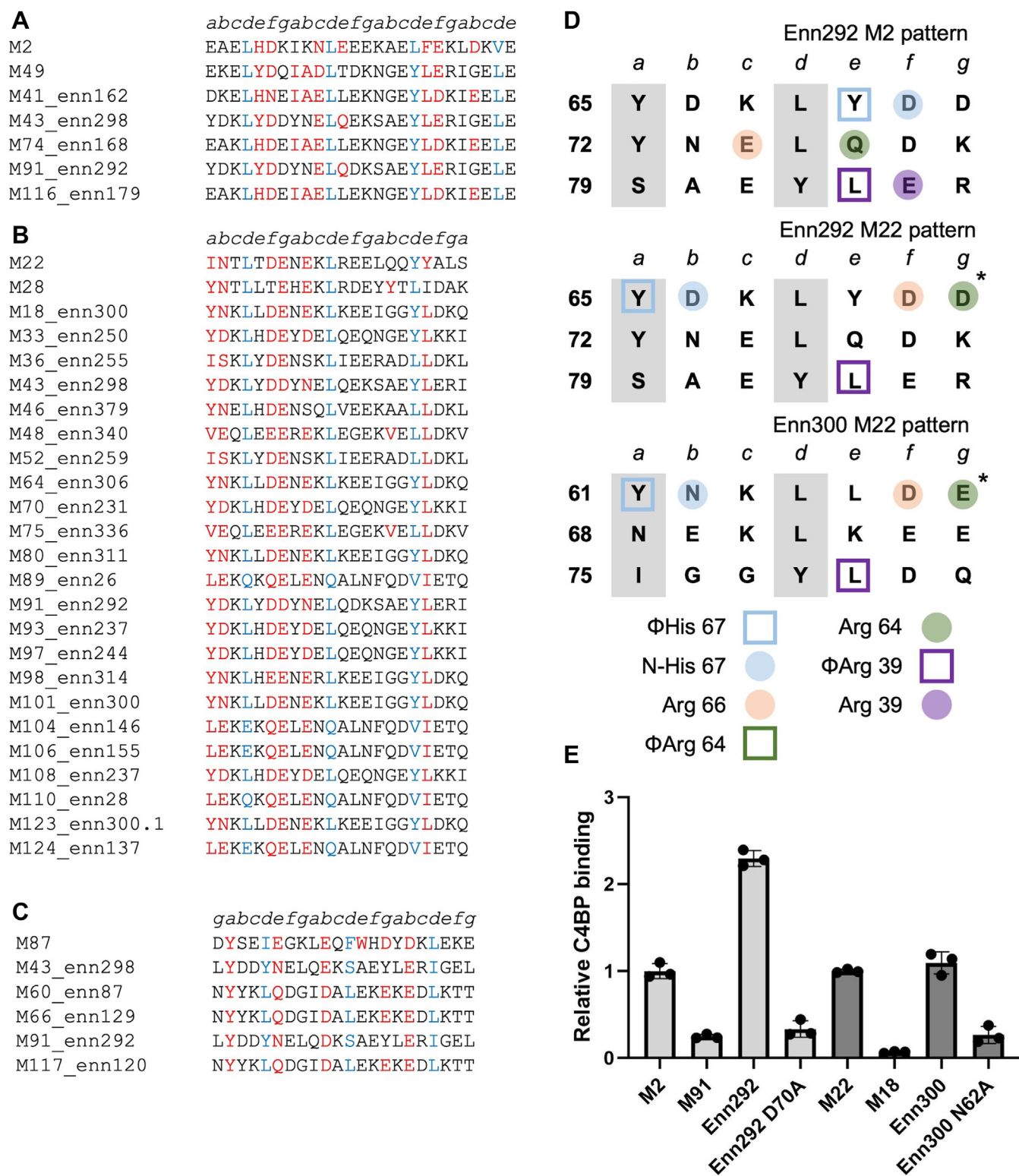
**Figure 6. Enn protein–C4BP interactions.** *A*, partial list of Enn protein sequences belonging to the M2 pattern (M2 and M49 sequences shown at *top*). Heptad positions are indicated above the sequences. *Red* indicates an observed or potential C4BP-binding amino acid, and *blue* indicates a *d* heptad position. A full list is in Table S3. *B*, same as *panel A*, except for the M22 pattern (M22 and M28 sequences shown at *top*). *C*, same as *panel A*, except for the M87 pattern (M87 sequence shown at *top*). *D*, heptad register of Enn292 (aa 64–84) in the M2 and M22 patterns and Enn300 (aa 60–80) in the M22 pattern (*a* and *d* positions in *gray columns*). Amino acids marked *with an asterisk* belong to the minor helix, and those *without an asterisk* to the major helix. *Shape* indicates type of contact (*square* for hydrophobic; *circle* for polar), and *coloring* indicates which C4BP amino acid (depicted in legend at *bottom*) is predicted to be contacted. *E*, binding of soluble, intact His$_6$-M or Enn protein constructs (WT or Ala-substituted) to immobilized intact C4BP evaluated by ELISA. Bound His$_6$-M or Enn protein was detected with horseradish peroxidase-conjugated anti-His$_6$ antibodies. Values were normalized by the value of C4BP binding to WT His$_6$-M2 protein for Enn292 and WT His$_6$-M22 protein for Enn300. Data from three biological replicates are presented with means and SDs. C4BP, C4b-binding protein.

utilized, except for one amino acid from this grouping (R64), while the C4BPα1 Arg39 nook formed hydrophobic and polar contacts similar to ones observed previously (10). In contrast, the M68 HVR utilized the C4BPα2 quadrilateral, albeit using multiple amino acids rather than a single one to contact the ΦHis 67 site, but lacked contact to the C4BPα1 Arg39 nook, with the C4BPα1 domain in this complex almost entirely disordered.

Remarkably, in M87 protein, the same site used for binding C4BP is also used for binding the human antimicrobial peptide LL-37 (22). Three M87 amino acids in particular form similar contacts to C4BP and LL-37. These are M87 Y81, E85, and W92, with the last of these forming hydrophobic contacts to an Arg in both C4BP (R39) and LL37 (R23). However, the structure of M87 protein differs considerably between these two complexes. With LL-37, the two M87 α-helices splay apart and no longer form a coiled coil in order to accommodate α-helical LL-37, which inserts between the two M87 helices, whereas with C4BP, the two M87 α-helices formed a coiled coil (Fig. S6). M87 in its unbound state also formed a coiled coil in this region (Fig. S6). Fourteen other M proteins have a sequence similar to the LL-37–binding motif of M87 (22) and all but one of these were identified to have a C4BP-binding pattern (Table S2). As expected, most have the M87 pattern, but the M22 and M68 patterns were also represented. These results point to the striking ability of a short coiled-coil sequence to have multiple functions through conformational dynamics.

The structures of M87 and M68 HVRs bound to C4BPα1-2 were used to guide mutational analysis to identify which M protein amino acids were most significant for binding. Based on these data, we searched M proteins for sequence patterns matching the C4BP-binding patterns of M87 and M68 proteins. Where appropriate, we used as a threshold the scores for M12 and M24 proteins, as these Strep A strains do not bind C4BP or have negligible binding (9). Numerous M proteins, 19 in all, have the M87 pattern, whereas the M68 pattern was rare, limited to 2 M proteins. We also carried out mutational analysis on M2 protein, taking into account its structure and that of the M49 HVR bound to C4BPα1-2, in addition to prior mutagenesis results on M2 protein (10). We also carried out mutational analysis on M22 protein, taking into account its structure and that of the M28 HVR bound to C4BPα1-2. Pattern searches based on these data revealed that the M22 pattern was the most populated, shared in 49 M proteins. The M2 pattern was also well populated, with 17 M proteins. Some M proteins had more than one sequence segment compatible with a single C4BP-binding pattern (*e.g.,* two segments of M81 have an M22 pattern), and 4 M proteins (*i.e.*, M8, M9, M81, and M103) had sequences compatible with two patterns (Table S2). Even with cut-offs applied, M3 protein from the very low C4BP-binding M3 strain was identified to have sequences compatible with M22 and M87 patterns (9). M3 protein has an unusually low coiled-coil propensity (31), even for an M protein (23, 24), and the absence of coiled-coil structure may impede M3 from binding C4BP. In all, 79 M protein types, excluding M3, out of 179

were identified as having an M2, M22, M68, or M87 C4BP-binding pattern.

M-like proteins also contribute to C4BP binding. Deletion of the M-like Enn protein in M18 and M53 strains result in loss of C4BP binding to the bacterial surface (32, 33). Complementation to restore C4BP binding and verify the role of the Enn protein was not carried out in these studies. We demonstrated direct interaction between Enn300 protein, which is expressed by the M18 strain and has an M22 pattern, with C4BP. We also validated the M22 pattern in Enn300 through mutagenesis. In the M53 strain (specifically AP53), expression of the Enn protein requires inactivation of CovRS (33), a two-component system that negatively controls a multitude of virulence factors (34). The AP53 strain expresses Enn254, which we identified has an M22 pattern (Table S3). In the study that identified C4BP binding in 90 strains, the M53 strain failed to bind C4BP (9), presumably due to lack of expression of the Enn protein. As with M proteins, some Enn proteins have at least two different types of C4BP-binding patterns (Fig. 6D, Enn292). Whether antibody reactivity against Enn proteins results in protection is not known and requires experimental investigation.

The C4BP-binding patterns were largely or exclusively near the N-terminal ends of M and Enn proteins (95% and 100%, respectively), a location that presumably enables access for multichained C4BP ($\alpha_7\beta_1$, and with protein S also bound) (35). The exceptions were a few M proteins which had the M22 pattern (*e.g.*, M221, aa 193). Further investigation is required to determine whether C4BP can access a C-terminal site closer to the bacterial cell wall. The most N terminally located pattern was that of M77 protein (at aa 49, just seven amino acids from the N terminus), and in a previous study we found that M77 protein, despite having an excellent M2 pattern, does not bind C4BP in a statistically significant manner *in vitro* (21). The very N-terminal portions of M proteins tend to lack coiled-coil character, which is true of M77 protein, and thus insufficient structuring may explain the lack of C4BP binding. Structuring of the coiled coil may be enhanced by the binding of another protein, as has been shown for the interaction of the M-like protein H (PrtH) with C4BP requiring simultaneous binding of the Fc domain of IgG to PrtH (36). PrtH lacks a known C4BP-binding pattern.

In all, we identified C4BP patterns in 111 Strep A strains of differing M types (Table S4, excluding M3). These comprise 71 of the 90 strains shown to bind C4BP (9). Among these 71 strains, 33 have both an M and an Enn protein with a C4BP-binding pattern, nine have only an M protein with a C4BP-binding pattern, and 29 have only an Enn protein with a C4BP-binding pattern. Three strains that do not bind or bind C4BP only weakly (*i.e.*, M24, M39, M53) have an Enn protein with a C4BP-binding pattern. As observed for the AP53 strain (33), interaction with C4BP in these strains may require mutation of CovRS or another regulatory system for expression of the Enn protein. In addition, 37 strains that to our knowledge have not been tested for C4BP-binding have a C4BP-binding pattern in their M proteins; Enn proteins were not investigated in these strains. In summary, our results indicate that

C4BP-binding patterns are prevalent in Strep A strains of differing M types, being present in their M or Enn proteins, or frequently both, and provide further motivation for development of conserved C4BP-binding 3D patterns as vaccine immunogens. A vaccine immunogen combining the M2, M22, and M87 patterns offers the potential of protection against about half of all known M type strains.

## Experimental procedures

### DNA manipulation

Coding sequences for M and Enn proteins were cloned into a modified pET28b vector (Novagen) that contained sequences encoding an N-terminal $His_6$-tag, followed by a PreScission protease cleavage site. Amino acid substitutions and deletions were introduced with the QuickChange II site-directed mutagenesis kit (Stratagene), according to the manufacturer's directions.

### Protein expression and purification

Expression of M proteins, Enn proteins, and C4BPα1-2 in *E. coli* Gold (DE3) BL21 and purification of these proteins were carried out as previously described (10), with minor modifications. M and Enn protein constructs composed of their N-terminal 100 amino acids were purified only on $Ni^{2+}$-NTA resin (Thermo Fisher Scientific) with the size-exclusion chromatography step omitted, followed by dialysis against 20 mM Tris–HCl, pH 8.0, 100 mM NaCl. Constructs used for crystallization, namely C4BPα1-2, $M87^{N100}$, and $M68^{N100}$, were additionally subjected to PreScission protease (Gen-Script) digestion according to manufacturer's instructions to remove the $His_6$-tag. Intact M and Enn proteins were purified by affinity chromatography on $Ni^{2+}$-NTA resin, followed by size-exclusion chromatography on a Superdex 200 16/60 column that had been pre-equilibrated with 20 mM Tris–HCl, pH 8.0, 100 mM NaCl. Fractions containing pure, nondegraded proteins of interest were pooled.

SeMet labeling of $M68^{N100}$ was performed using SelenoMet medium (Molecular Dimensions) through feedback inhibition of methionine biosynthesis, as previously described (37). SeMet-$M68^{N100}$ was purified as above, except buffers were supplemented with 5 mM DTT.

### Crystallization and data collection

SeMet-$M68^{N100}$ and $M87^{N100}$ were mixed with C4BPα1-2 in equimolar ratio. The buffer for the $M68^{N100}$ mixture was exchanged by ultrafiltration through a 3500 molecular weight cut off (MWCO) membrane (Millipore; 4500*g*, 30 min, 4 °C) to 20 mM Tris–HCl, pH 8.0, 100 mM NaCl, and samples were concentrated to 10 mg/ml by ultrafiltration through a 3500 MWCO membrane. The buffer for the $M87^{N100}$ mixture was exchanged the same way to 10 mM Tris–HCl, pH 8.0, and samples were concentrated to 15 mg/ml, as described above. All crystallization trials were carried out at 20 °C using the hanging drop vapor diffusion method.

For crystallization of SeMet-$M68^{N100}$-C4BPα1-2, 1 µl of protein sample was mixed with 1 µl of precipitant solution (100 mM MES-NaOH, pH 6.0, 18% PEG 3350, 10% ethylene glycol, 0.2 M

$CaCl_2$). Small ($50 \times 10 \times 10$ $\mu m^3$) crystals, which grew in 2 to 7 days, were crushed and diluted 100-fold with precipitant solution to constitute a seed stock. The seed stock (0.2 µl) was mixed with 0.9 µl of SeMet-$M68^{N100}$-C4BPα1-2 and 0.9 µl of 100 mM MES-NaOH, pH 6.0, 14% PEG 3350, 10% ethylene glycol, 0.2 M $CaCl_2$, and 1 mM DTT. Single crystals ($300 \times 50 \times 50$ $\mu m^3$), which grew after 4 to 7 days, were transferred to cryoprotectant composed of 100 mM MES-NaOH, pH 6.0, 14% PEG 3350, 30% ethylene glycol, 0.2 M $CaCl_2$, and 1 mM DTT and flash-cooled in liquid $N_2$ prior to data collection. Diffraction data were collected at the Stanford Synchrotron Radiation Lightsource (beamline 9–2) at 0.9789, 0.9184, and 0.9793 Å wavelength (inflection, high energy, and remote, respectively), integrated with XDS (38), and scaled with Aimless (39).

For cocrystallization of $M87^{N100}$-C4BPα1-2, 1 µl of protein sample was mixed with 1 µl of precipitant solution (0.1 M sodium citrate, pH 4.8, 0.2 M $KH_2PO_4$, 25% w/v PEG 3350). Crystals ($400 \times 80 \times 80$ $\mu m^3$), which grew after about 7 days, were cryoprotected with precipitant solution supplemented with 20% glycerol, and flash-frozen in liquid $N_2$. Diffraction data were collected at the Advanced Photon Source (beamline 24-ID-E) at 0.979 Å, and automatically processed using rapid automated processing of data, including XDS for integration and Aimless for scaling.

### Structure determination and refinement

Phase determination by multiwavelength anomalous dispersion, phase refinement, and initial model building for SeMet-$M68^{N100}$-C4BPα1-2 was carried out with AutoBuild from the Phenix suite (40). The model was modified using Coot, as guided by the inspection of σA-weighted 2Fo-Fc and Fo-Fc omit maps. Multiple cycles of building followed by refinement were performed using Refine (41) from the Phenix suite with default settings. At late stages translation-libration-screw (TLS) parameters were introduced, with each chain constituting a single TLS group. Side chains with no electron density beyond Cβ were truncated at Cβ.

Phases for the $M87^{N100}$–C4BPα1-2 complex were determined by molecular replacement using Phaser (42), with C4BPα1-2 (PDB 5I0Q) as the search model. The model was built manually into well-defined electron density using Coot and modified as guided by the inspection of σA-weighted 2Fo-Fc and Fo-Fc omit maps. Refinement was performed using Refine from the Phenix suite with default settings and TLS parameters. Each chain constituted a single TLS group. Side chains with no electron density beyond Cβ were truncated at Cβ.

Figures of molecular models were generated with PyMol (pymol.org, The PyMOL Molecular Graphics System, Version 2.5.5 Schrödinger, LLC). Interaction interfaces between M proteins and C4BP were analyzed with PISA (43).

### CD spectroscopy

CD spectra were measured on a J-1500 CD Spectrometer (Jasco). Samples were exchanged by dialysis through a 10,000 MWCO membrane into 5 mM sodium phosphate, pH 7.9, and diluted to ~0.125 to 0.250 mg/ml. Wavelength spectra were

recorded between 190 and 260 nm at 37 °C with a 0.5 nm interval. Thermal melting was carried out between 20 and 75 °C in 1 °C increments, with the CD signal being monitored at 222 nm. Wavelength scans and thermal melts were carried out in duplicate, and the data were averaged.

### Enzyme-linked immunosorbent assay

ELISAs were performed as previously described (21). In short, high-binding 96-well microtiter plates (Corning) were coated with intact C4BP (Athens Research and Technology; 5 μg/ml) in 100 μl PBS, pH 7.4 overnight at 4 °C. Wells were incubated with 100 μl of purified N terminally His$_6$-tagged M or Enn proteins (final concentration of 10 μg/ml). Horseradish peroxidase-conjugated mouse anti-His antibody (BioLegend) diluted 1:5000 in 0.1% bovine serum albumin/tris-buffered saline with Tween-20 was used for detection. Background from binding to uncoated wells was subtracted.

### Sequence scoring

Files containing 179 M protein or 143 Enn protein sequences (from NCBI GenBank) (29) were searched for C4BP-binding patterns. The sequences were in fasta format, and a python script was written to carry out the search. Sequences of 23 to 27 aa in length (*i.e.*, more than three heptads but less than four) were searched. Only sequences that had coiled-coil–stabilizing amino acids (*i.e.*, V, L, I, M, F, or Y) (25, 26) at four or more *a* or *d* positions in three consecutive heptads were considered.

For the M87 pattern, at the position equivalent to M87 D80, the occurrence of a K or R eliminated the sequence from consideration (by giving this position a score of −100; only sequences that had a total score >0 were considered). At the position equivalent to M87 Y81, the occurrence of I, L, M, F, Y, or W resulted in a score of +1. The occurrence of any other amino acid at this position resulted in elimination of the sequence. This was true for all positions at which specific amino acids were not excluded, except if stated otherwise. At positions equivalent to M87 E85, E89, and D96, the occurrence of D, E, N, or Q resulted in a score of +1 for the first two and +0.7 for the last two; this scoring scheme was used at most positions occupied by D, E, N, or Q and was called "DENQ" scoring. At the position equivalent to M87 W92, the occurrence of a K or R eliminated the sequence, but the occurrence of V, L, I, M, F, Y, or W (the term "hydrophobic amino acid" refers to this set below) resulted in a score of +0.5. At the position equivalent to M87 D94, the occurrence of L, I, M, or F eliminated the sequence, but the occurrence of D or E resulted in a score of +0.5, N or Q a score of +0.3, and S or T a score of +0.1.

For the M68 pattern, at the positions equivalent to M68 M58 and M62, the occurrence of a hydrophobic amino acid resulted in a score of +1. At the positions equivalent to M68 W59, Y63, and Y70, the occurrence of I, L, M, F, Y, or W resulted in a score of +1. At the position equivalent to M68 K64, the occurrence of K or R resulted in a score of +1. At positions equivalent to M68 D67 and D69, DENQ scoring was used.

For the M2 pattern, the scoring was as follows. At the position equivalent to M2 H61, the occurrence of a His or a hydrophobic amino acid resulted in a score of +1. At the position equivalent to M2 D62, DENQ scoring was used, in addition to scoring Ser or Thr at this position with +0.5. At the position equivalent to M2 N66, we used the observation that if the preceding position were occupied by a Lys (equivalent to M2 K65), then the occurrence of an Asp or even Ala is preferred over Asn (10). Thus, if the position equivalent to M2 K65 were occupied by a Lys or Arg, then if the position equivalent to M2 N66 were occupied by a D or E, this resulted in a score of +1, whereas if it were occupied by an N or Q, this resulted in a score of +0.5. In the absence of a positively charged amino acid at the preceding position, DENQ scoring was applied. At the position equivalent to M2 E68, DENQ scoring was used. Since the polar contact of M2 E68 with C4BPα2 R64 is replaced in M49 with hydrophobic contacts to the alkyl chain of R64, the following was taken into account. In case the position equivalent to M2 E68 was not occupied by D, E, N, or Q, then if in the following heptad the *a* position (equivalent to M49 I71) were occupied by a coiled-coil–stabilizing amino acid, the *b* position (equivalent to M49 A72) by Ala, and the *e* position (equivalent to M49 T75) by a small amino acid that can form a hydrophobic contact (T, V, or L), this resulted in a score of +1. At the position equivalent to M2 F75, the occurrence of a hydrophobic amino acid resulted in a score of +1. As M2 E76 and D79 form redundant salt bridges to C4BPα1 R39, the occurrence of D or E at either position resulted in a score of +1. If neither position were occupied by D or E, then the occurrence of an N or Q at either position resulted in a score of +0.7.

For the M22 pattern, the position equivalent to M22 I59, the occurrence of a coiled-coil–stabilizing amino acid resulted in a score of +1. At the position equivalent to M22 N60, DENQ scoring was used, in addition to scoring Ser or Thr at this position with +0.5. In M22, D64 (heptad position *f*) forms a salt bridge to C4BPα2 R66, whereas in M28, E70 (*b* heptad position, equivalent to M22 position 67) does. Thus, if a D, E, N, or Q occupied either the position equivalent of M22 positions 64 or 67, DENQ scoring was used. At the position equivalent to M22 E65, DENQ scoring was used. M22 Y77 (*e* heptad position) forms a hydrophobic contact to C4BPα1 R39, whereas M28 Y77 (*b* heptad position, equivalent to M22 position 74) does the same. Therefore, the occurrence of a hydrophobic amino acid at the equivalent of M22 position 74 or 77 resulted in a score of +1.

### Data availability

Atomic models and X-ray crystallographic data are available through the RCSB PDB (8TCB and 8TGT).

*Conflict of interest*—The authors declare that they have no conflicts of interest with the contents of this article.

*Abbreviations*—The abbreviations used are: C4BP, C4b-binding protein; HVR, hypervariable region; MWCO, molecular weight cut off; PrtH, protein H; SeMet, selenomethionine; TLS, translation-libration-screw.

## References

1. Barnett, T. C., Bowen, A. C., and Carapetis, J. R. (2018) The fall and rise of Group A Streptococcus diseases. *Epidemiol. Infect.* **147**, e4
2. Nelson, G. E., Pondo, T., Toews, K. A., Farley, M. M., Lindegren, M. L., Lynfield, R., *et al.* (2016) Epidemiology of invasive group A streptococcal infections in the United States, 2005-2012. *Clin. Infect. Dis.* **63**, 478–486
3. Carapetis, J. R., Beaton, A., Cunningham, M. W., Guilherme, L., Karthikeyan, G., Mayosi, B. M., *et al.* (2016) Acute rheumatic fever and rheumatic heart disease. *Nat. Rev. Dis. Primers* **2**, 15084
4. Nitsche-Schmitz, D. P., and Chhatwal, G. S. (2013) Host-pathogen interactions in streptococcal immune sequelae. *Curr. Top. Microbiol. Immunol.* **368**, 155–171
5. Carapetis, J. R., Steer, A. C., Mulholland, E. K., and Weber, M. (2005) The global burden of group A streptococcal diseases. *Lancet Infect. Dis.* **5**, 685–694
6. Berggard, K., Johnsson, E., Morfeldt, E., Persson, J., Stalhammar-Carlemalm, M., and Lindahl, G. (2001) 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
7. Carlsson, F., Berggard, K., Stalhammar-Carlemalm, M., and Lindahl, G. (2003) Evasion of phagocytosis through cooperation between two ligand-binding regions in *Streptococcus pyogenes* M protein. *J. Exp. Med.* **198**, 1057–1068
8. Ermert, D., Shaughnessy, J., Joeris, T., Kaplan, J., Pang, C. J., Kurt-Jones, E. A., *et al.* (2015) Virulence of group A *Streptococci* is enhanced by human complement inhibitors. *PLoS Pathog.* **11**, e1005043
9. Persson, J., Beall, B., Linse, S., and Lindahl, G. (2006) Extreme sequence divergence but conserved ligand-binding specificity in *Streptococcus pyogenes* M protein. *PLoS Pathog.* **2**, e47
10. Buffalo, C. Z., Bahn-Suh, A. J., Hirakis, S. P., Biswas, T., Amaro, R. E., Nizet, V., *et al.* (2016) Conserved patterns hidden within group A Streptococcus M protein hypervariability recognize human C4b-binding protein. *Nat. Microbiol.* **1**, 16155
11. McMillan, D. J., Dreze, P. A., Vu, T., Bessen, D. E., Guglielmini, J., Steer, A. C., *et al.* (2013) Updated model of group A *Streptococcus* M proteins based on a comprehensive worldwide study. *Clin. Microbiol. Infect.* **19**, E222–E229
12. Ghosh, P. (2018) Variation, Indispensability, and Masking in the M protein. *Trends Microbiol.* **26**, 132–144
13. Johnsson, E., Thern, A., Dahlback, B., Heden, L. O., Wikstrom, M., and Lindahl, G. (1996) A highly variable region in members of the streptococcal M protein family binds the human complement regulator C4BP. *J. Immunol.* **157**, 3021–3029
14. Morfeldt, E., Berggard, K., Persson, J., Drakenberg, T., Johnsson, E., Lindahl, E., *et al.* (2001) Isolated hypervariable regions derived from streptococcal M proteins specifically bind human C4b-binding protein: implications for antigenic variation. *J. Immunol.* **167**, 3870–3877
15. Vekemans, J., Gouvea-Reis, F., Kim, J. H., Excler, J. L., Smeesters, P. R., O'Brien, K. L., *et al.* (2019) The Path to group A Streptococcus vaccines: World Health Organization research and development Technology Roadmap and preferred Product characteristics. *Clin. Infect. Dis.* **69**, 877–883
16. Sandin, C., Carlsson, F., and Lindahl, G. (2006) Binding of human plasma proteins to *Streptococcus pyogenes* M protein determines the location of opsonic and non-opsonic epitopes. *Mol. Microbiol.* **59**, 20–30
17. Lannergard, J., Gustafsson, M. C., Waldemarsson, J., Norrby-Teglund, A., Stalhammar-Carlemalm, M., and Lindahl, G. (2011) The hypervariable region of *Streptococcus pyogenes* M protein escapes antibody attack by antigenic variation and weak immunogenicity. *Cell Host Microbe* **10**, 147–157
18. Jones, K. F., and Fischetti, V. A. (1988) The importance of the location of antibody binding on the M6 protein for opsonization and phagocytosis of group A M6 streptococci. *J. Exp. Med.* **167**, 1114–1123
19. Pandey, M., Ozberk, V., Calcutt, A., Langshaw, E., Powell, J., Rivera-Hernandez, T., *et al.* (2016) Streptococcal Immunity is Constrained by lack of Immunological memory following a single Episode of Pyoderma. *PLoS Pathog.* **12**, e1006122
20. Dale, J. B., and Beachey, E. H. (1986) Localization of protective epitopes of the amino terminus of type 5 streptococcal M protein. *J. Exp. Med.* **163**, 1191–1202
21. Wang, K. C., Kuliyev, E., Nizet, V., and Ghosh, P. (2023) A conserved 3D pattern in a Streptococcus pyogenes M protein immunogen elicits M-type crossreactivity. *J. Biol. Chem.* **299**, 104980
22. Kolesinski, P., Wang, K. C., Hirose, Y., Nizet, V., and Ghosh, P. (2022) An M protein coiled coil unfurls and exposes its hydrophobic core to capture LL-37. *Elife* **11**, e77989
23. Nilson, B. H., Frick, I. M., Akesson, P., Forsen, S., Bjorck, L., Akerstrom, B., *et al.* (1995) Structure and stability of protein H and the M1 protein from Streptococcus pyogenes. Implications for other surface proteins of gram-positive bacteria. *Biochemistry* **34**, 13688–13698
24. Stewart, C. M., Buffalo, C. Z., Valderrama, J. A., Henningham, A., Cole, J. N., Nizet, V., *et al.* (2016) Coiled-coil destabilizing residues in the group A Streptococcus M1 protein are required for functional interaction. *Proc. Natl. Acad. Sci. U. S. A.* **113**, 9515–9520
25. Tripet, B., Wagschal, K., Lavigne, P., Mant, C. T., and Hodges, R. S. (2000) Effects of side-chain characteristics on stability and oligomerization state of a de novo-designed model coiled-coil: 20 amino acid substitutions in position "d". *J. Mol. Biol.* **300**, 377–402
26. Wagschal, K., Tripet, B., Lavigne, P., Mant, C., and Hodges, R. S. (1999) The role of position a in determining the stability and oligomerization state of alpha-helical coiled coils: 20 amino acid stability coefficients in the hydrophobic core of proteins. *Protein Sci.* **8**, 2312–2329
27. Sanderson-Smith, M., De Oliveira, D. M., Guglielmini, J., McMillan, D. J., Vu, T., Holien, J. K., *et al.* (2014) 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
28. Whatmore, A. M., Kapur, V., Musser, J. M., and Kehoe, M. A. (1995) Molecular population genetic analysis of the enn subdivision of group A

streptococcal emm-like genes: horizontal gene transfer and restricted variation among enn genes. *Mol. Microbiol.* **15**, 1039–1048

29. Frost, H. R., Davies, M. R., Delforge, V., Lakhloufi, D., Sanderson-Smith, M., Srinivasan, V., *et al.* (2020) Analysis of global collection of group A Streptococcus genomes Reveals that the Majority encode a Trio of M and M-like proteins. *mSphere* **5**, e00806–e00819

30. Frost, H. R., Guglielmini, J., Duchene, S., Lacey, J. A., Sanderson-Smith, M., Steer, A. C., *et al.* (2023) Promiscuous evolution of group A streptococcal M and M-like proteins. *Microbiology (Reading)* **169**, 001280

31. Ludwiczak, J., Winski, A., Szczepaniak, K., Alva, V., and Dunin-Horka-wicz, S. (2019) DeepCoil-a fast and accurate prediction of coiled-coil domains in protein sequences. *Bioinformatics* **35**, 2790–2795

32. Perez-Caballero, D., Garcia-Laorden, I., Cortes, G., Wessels, M. R., de Cordoba, S. R., and Alberti, S. (2004) Interaction between complement regulators and Streptococcus pyogenes: binding of C4b-binding protein and factor H/factor H-like protein 1 to M18 strains involves two different cell surface molecules. *J. Immunol.* **173**, 6899–6904

33. Agrahari, G., Liang, Z., Mayfield, J. A., Balsara, R. D., Ploplis, V. A., and Castellino, F. J. (2013) Complement-mediated opsonization of invasive group A Streptococcus pyogenes strain AP53 is regulated by the bacterial two-component cluster of virulence responder/sensor (CovRS) system. *J. Biol. Chem.* **288**, 27494–27504

34. Graham, M. R., Smoot, L. M., Migliaccio, C. A., Virtaneva, K., Sturde-vant, D. E., Porcella, S. F., *et al.* (2002) Virulence control in group A Streptococcus by a two-component gene regulatory system: global expression profiling and in vivo infection modeling. *Proc. Natl. Acad. Sci. U. S. A.* **99**, 13855–13860

35. Dahlback, B. (1983) Purification of human C4b-binding protein and formation of its complex with vitamin K-dependent protein S. *Biochem. J.* **209**, 847–856

36. Ermert, D., Weckel, A., Magda, M., Morgelin, M., Shaughnessy, J., Rice, P. A., *et al.* (2018) Human IgG Increases virulence of Streptococcus pyogenes through complement Evasion. *J. Immunol.* **200**, 3495–3505

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

38. Kabsch, W. (2010) Xds. *Acta Crystallogr. D Biol. Crystallogr.* **66**, 125–132

39. Evans, P. R., and Murshudov, G. N. (2013) How good are my data and what is the resolution? *Acta Crystallogr. D Biol. Crystallogr.* **69**, 1204–1214

40. Liebschner, D., Afonine, P. V., Baker, M. L., Bunkoczi, G., Chen, V. B., Croll, T. I., *et al.* (2019) Macromolecular structure determination using X-rays, neutrons and electrons: recent developments in Phenix. *Acta Crystallogr. D Struct. Biol.* **75**, 861–877

41. Afonine, P. V., Grosse-Kunstleve, R. W., Echols, N., Headd, J. J., Mor-iarty, N. W., Mustyakimov, M., *et al.* (2012) Towards automated crys-tallographic structure refinement with phenix.refine. *Acta Crystallogr. D Biol. Crystallogr.* **68**, 352–367

42. McCoy, A. J., Grosse-Kunstleve, R. W., Adams, P. D., Winn, M. D., Storoni, L. C., and Read, R. J. (2007) Phaser crystallographic software. *J. Appl. Crystallogr.* **40**, 658–674

43. Krissinel, E., and Henrick, K. (2007) Inference of macromolecular as-semblies from crystalline state. *J. Mol. Biol.* **372**, 774–797