Isolated Hypervariable Regions Derived from Streptococcal M Proteins Specifically Bind Human C4b-Binding Protein: Implications for Antigenic Variation

Eva Morfeldt, Karin Berggård, Jenny Persson, Torbjörn Drakenberg, Eskil Johnsson, Erik Lindahl, Sara Linse and Gunnar Lindahl

*J Immunol* 2001; 167:3870-3877; doi: 10.4049/jimmunol.167.7.3870

http://www.jimmunol.org/content/167/7/3870

---

**References**

This article cites 61 articles, 30 of which you can access for free at:
http://www.jimmunol.org/content/167/7/3870.full#ref-list-1

**Subscription**

Information about subscribing to *The Journal of Immunology* is online at:
http://jimmunol.org/subscription

**Permissions**

Submit copyright permission requests at:
http://www.aai.org/About/Publications/JI/copyright.html

**Email Alerts**

Receive free email-alerts when new articles cite this article. Sign up at:
http://jimmunol.org/alerts
Isolated Hypervariable Regions Derived from Streptococcal M Proteins Specifically Bind Human C4b-Binding Protein: Implications for Antigenic Variation

Eva Morfeldt,* Karin Berggård,* Jenny Persson,* Torbjörn Drakenberg,† Eskil Johnsson,* Erik Lindahl,‡ Sara Linse,† and Gunnar Lindahl2*

Antigenic variation in microbial surface proteins represents an apparent paradox, because the variable region must retain an important function, while exhibiting extensive immunological variability. We studied this problem for a group of streptococcal M proteins in which the ~50-residue hypervariable regions (HVRs) show essentially no residue identity but nevertheless bind the same ligand, the human complement regulator C4b-binding protein (C4BP). Synthetic peptides derived from different HVRs were found to retain the ability to bind C4BP, implying that the HVR corresponds to a distinct ligand-binding domain that can be studied in isolated form. This finding allowed direct characterization of the ligand-binding properties of isolated HVRs and permitted comparisons between different HVRs in the absence of conserved parts of the M proteins. Affinity chromatography of human serum on immobilized peptides showed that they bound C4BP with high specificity and inhibition experiments indicated that different peptides bound to the same site in C4BP. Different C4BP-binding peptides did not exhibit any immunological cross-reactivity, but structural analysis suggested that they have similar folds. These data show that the HVR of streptococcal M protein can exhibit extreme variability in sequence and immunological properties while retaining a highly specific ligand-binding function. The Journal of Immunology, 2001, 167: 3870–3877.

Among the mechanisms that pathogenic microorganisms use to evade the immune system of the infected host, antigenic variation is of particular interest, because it occurs in many different pathogens and is the major obstacle to the development of new vaccines (1–3). The antigenic variability is usually due to sequence changes in a surface protein important for virulence, allowing the microorganism to evade Abs and/or cellular immunity. This situation represents an apparent paradox, because a variable region must have an important function that is retained despite the variability; otherwise, this region would be lost by mutation. Here, we study this problem for streptococcal M protein, a major bacterial virulence factor, and show that in many M proteins the extremely variable N-terminal region corresponds to a distinct ligand-binding domain that can be studied in isolated form.

Streptococcus pyogenes (group A Streptococcus) is a Gram-positive bacterium that causes several diseases, including acute tonsillitis, skin infections, rheumatic fever, and toxic shock syndrome (4). The M protein, which is the most extensively studied virulence factor of S. pyogenes, is a dimeric coiled-coil protein that inhibits phagocytosis and exhibits antigenic variation due to an N-terminal hypervariable region (HVR) (5, 6). The HVR shows little variation between strains of a given serotype, but ~100 different variants have been found in strains of different serotypes. The available evidence indicates that antigenic variation in M proteins has arisen mainly through genetic drift and selection of the most fit variants (7, 8).

The HVR of M protein is known to be a target for type-specific protective Abs (5, 6, 9), but the function in pathogenesis of this extremely variable region is unclear. A clue to this problem was obtained by the finding that the ~50-residue HVR of many M proteins promotes binding of human C4b-binding protein (C4BP), a plasma protein that inhibits activation of the complement system (10–13) (Fig. 1A). Bacteria-bound C4BP retains its complement-regulatory function and contributes to phagocytosis resistance (Ref. 10; K. Berggård and G. Lindahl, manuscript in preparation), which may explain why the HVR is a target for protective antibodies and exhibits antigenic variation. However, the specificity with which different HVRs bind C4BP is still unclear, and this problem could not be easily studied with intact M proteins, which bind several human plasma proteins (6, 9).

The studies presented here show that dimerized synthetic peptides derived from different C4BP-binding HVRs retain the ability to bind C4BP, providing the unique opportunity to analyze the specificity with which the HVRs bind C4BP, in the absence of other parts of the M proteins. Moreover, the immunological and ligand-binding properties of different HVRs could be directly compared. Our data show that different isolated HVRs bind C4BP with high specificity and that these HVRs bind to the same region in C4BP, although they exhibit almost no residue identity and lack immunological cross-reactivity. Furthermore, structural analysis indicated that the different C4BP-binding HVRs may have similar

---

Departments of *Medical Microbiology, Dermatology, and Infection and †Physical Chemistry 2, Lund University, Lund, Sweden; and ‡Department of Theoretical Physics, Royal Institute of Technology, Stockholm, Sweden

Received for publication May 10, 2001. Accepted for publication July 6, 2001.

The costs of publication of this article were defrayed in part by the payment of page charges. This article must therefore be hereby marked advertisement in accordance with 18 U.S.C. Section 1734 solely to indicate this fact.

1 This work was supported by grants from the Swedish Medical Research Council (Grant 9490), the Medical Faculty of Lund University, the Swedish Foundation for Strategic Research, the Royal Physiographic Society in Lund, the Swedish Society for Medical Research, and the Trusts of Crafoord, Golje, Hedberg, Hierta, Kock, Lundström, and Osterlund.

2 Address correspondence and reprint requests to Dr. Gunnar Lindahl, Department of Medical Microbiology, Dermatology and Infection, Lund University, Sölvegatan 23, SE-22362 Lund, Sweden. E-mail address: gunnar.lindahl@mmb.lu.se

---

3 Abbreviations used in this paper: HVR, hypervariable region; C4BP, C4b-binding protein; HA, hemagglutinin; CCP, complement control protein.
structures. These studies of streptococcal M protein demonstrate that a distinct domain in a bacterial surface protein can retain the ability to bind a ligand with high specificity, despite extreme variability in sequence and immunological properties.

Materials and Methods

M protein nomenclature

The M proteins of S. pyogenes confer resistance to phagocytosis. For historical reasons, some of these proteins have been given different names, such as ML2.1, Arp4, Sir22, etc. Following a proposal to simplify the nomenclature (8), we have previously referred to the latter proteins, which are all encoded by emm genes, as Emm proteins (14). However, the term Emm protein has not been generally used, but all proteins encoded by emm genes are usually referred to as M proteins (e.g., Ref. 15). Here, we follow this nomenclature. The proteins referred to here as M2, M4, M22, and M60 were previously designated ML2.1, Arp4, Sir22, and Arp60, respectively (16–19). However, streptococcal protein H (20), which appears to have a hybrid structure (8), will be referred to with that name, or abbreviated as PrfH.

Purified proteins and synthetic peptides

The full-length streptococcal M4 (Arp4), M5, and M22 (Sir22) proteins were purified as described, after expression of the cloned gene in Escherichia coli (11, 16, 19).

Synthetic ~50-residue peptides were purchased from the Department of Clinical Chemistry, Malmö General Hospital (Malmö, Sweden) and were ≥95% pure. These peptides were derived from the known N-terminal sequences of the processed form of different M proteins. Cysteine residues were added at the C-terminal ends, as indicated below, to allow immobilization via a disulfide bridge. Moreover, one or two tyrosine residues were added at the C-terminal ends of some of the peptides, as indicated, to facilitate radiolabeling. The 51-residue M2-N peptide corresponds to residues 1–48 of the M2 (ML2.1) protein (17), with a Tyr-Tyr-Cys sequence added. The 47-residue M4-N peptide corresponds to residues 1–45 of the M4 (Arp4) protein (16), with a Tyr-Cys sequence added. Due to technical difficulties in synthesizing peptides containing both arginine and tryptophan, the arginine (32) residue in the M4 protein was replaced with a lysine residue in the M4-N peptide. The 53-residue M5-N peptide corresponds to residues 1–50 of the M5 protein (21), with addition of the sequence Tyr-Tyr-Cys. The 53-residue M22-N peptide corresponds to residues 1–52 of the M22 (Sir22) protein (19), with addition of a cysteine residue (the peptide contains internal tyrosine residues). The M22-N peptide was also synthesized without the added cysteine (designated M22-N (no cysteine)). The 47-residue M4-N peptide corresponds to residues 1–46 of the M4 (Arp4) protein (16), with a Tyr-Cys sequence added. Due to technical difficulties in synthesizing peptides containing both arginine and tryptophan, the cysteine (32) residue in the M4 protein was replaced with a lysine residue in the M4-N peptide. The 53-residue M5-N peptide corresponds to residues 1–50 of the M5 protein (21), with addition of the sequence Tyr-Tyr-Cys. The 53-residue M22-N peptide corresponds to residues 1–52 of the M22 (Sir22) protein (19), with addition of a cysteine residue (the peptide contains internal tyrosine residues). The M22-N peptide was also synthesized without the added cysteine (designated M22-N (no cysteine)). The peptides containing a C-terminal cysteine residue were dimerized essentially as described (22). Briefly, the peptides were dissolved to 5 mg/ml in 10 mM Tris-HCl, pH 8.0, supplemented with CuCl2 (2 × 10−5 M), and NaCl (0.4 M), and the pH was adjusted to ~8.0 with 1 M NaOH. The solution was left at room temperature with gentle shaking for ~5 h and stored aliquoted at ~20°C. The 21-residue N14 peptide (sequence WNNWPKEYNNALLKEEELKVER) (11) and the 15-residue N12 peptide (sequence KDKQISDASKQGLSR) were derived from the M4 protein. The 18-residue peptides PA (sequence IYKCRHPGELRNGQVEI) and PD (sequence SRECVQDRGQGVSHPPLQ) were derived from the C4BP α-chain (23). The latter four peptides had acetylated N-termini and amidated C-termini and were purchased from Neosystem Laboratoire (Strasbourg, France).

Purified human C4BP was provided by Dr. Björn Dahlbäck or purified by affinity chromatography (J. Persson and G. Lindahl, manuscript in preparation). Human serum IgA was from Cappel-Organon-Teknika (Turnhout, Belgium), human IgG from Sigma (St. Louis, MO), and protein G from Amersham Pharmacia Biotech (Uppsala, Sweden).

Antiserum

Antiserum against the dimerized synthetic peptides were produced in rabbits. For the initial immunization, 200 μg dimerized peptide was mixed with 1 ml CFA; 2–4 booster injections of 50 μg peptide were given in IFA at 3-wk intervals. Anti-C4BP serum was a gift of Dr. Peter Bjoerkemark. Anti-protein S serum was purchased from Dako (Glostrup, Denmark).

Affinity chromatography

A solution (5 mg/ml) of dimerized peptide was dialyzed against coupling buffer (0.2 M NaHCO3, 0.5 M NaCl, pH 8.0) for 24 h at 4°C and then dialyzed against a 1-ml HiTrap column (Amersham Pharmacia Biotech). For affinity chromatography, 5 ml human serum were diluted 10-fold in PBS and passed through the column, which was then washed with 15 × 1 ml PBS.

Bound proteins were eluted with 5 × 1 ml 3 M KSCN or 6 M urea, dialyzed against PBS, and analyzed by SDS-PAGE.

Binding and inhibition tests

To analyze the ability of different peptides to bind immobilized ligands (Fig. 2B), the wells of microtiter plates (Falcon 3912; BD Biosciences, San Jose, CA) were coated overnight with 50 μl solutions of C4BP, IgA, or IgG, respectively (15 μg/ml in PBS). After blocking with PBS with 0.02% sodium azide and 0.05% Tween 20 (PBSAT), radiolabeled peptide (~15,000 cpm) was added in 50 μl PBSAT. Because binding between immobilized C4BP and synthetic peptides was improved by reducing the ionic strength of the medium, all incubations were performed in PBSAT diluted 10-fold in water (0.5× PBSAT). The wells were incubated for 2 h at 4°C and washed three times with ice-cold 0.5× PBSAT; the radioactivity associated with each well was then determined. All determinations were performed in triplicate.

To analyze the ability of different peptides/proteins to inhibit the binding between C4BP and C4BP-binding peptides (Fig. 3), C4BP was immobilized in microtiter wells by incubating the wells overnight with 50 μl of a solution (20 μg/ml) in PBS. The wells were washed three times with PBSAT, which also blocked the wells. The inhibitor was added, as indicated, together with radiolabeled peptide (~15,000 cpm/well) in a total volume of 50 μl 0.5× PBSAT, and the wells were incubated for 2 h. The wells were then washed three times with ice-cold 0.5× PBSAT, and the radioactivity associated with each well was measured. Nonspecific binding was determined in uncoated wells blocked with PBSAT. All incubations were performed at 4°C and performed in triplicate. Binding was expressed as a percentage of maximal binding of the labeled peptide to immobilized C4BP; this maximal binding varied from ~10% for M4-N to ~30% for M2-N.

To analyze the binding of antisera to C4BP-binding peptides (Fig. 5A), microtiter wells were coated overnight at 4°C with 50 μl of peptide (2 μg/ml) in PBS. After blocking with PBSAT, 50 μl antiserum (diluted in PBSAT as indicated) were added to each well. The wells were incubated at room temperature for 1 h and washed three times with PBSAT. To detect bound Abs, radiolabeled protein G (~15,000 cpm in 50 μl) was added to each well, followed by incubation for 1 h at room temperature. After washes with PBSAT, the radioactivity associated with each well was measured. Nonspecific binding was determined for uncoated wells blocked with PBSAT. For inhibition tests with antisera (Fig. 5B), microtiter wells were coated with C4BP-binding peptides (0.1 μg/ml in PBS) as described above. Antiserum to the immobilized peptide (diluted 125–250 times, to increase the sensitivity of the test) was added together with different peptides, as indicated. The ability of different peptides to inhibit binding of Abs was analyzed by the addition of radiolabeled protein G, as described above. All tests were performed in triplicate. Binding shown in Fig. 5 is expressed as a percentage of maximal binding (~60%) of radiolabeled protein G added.

CD spectroscopy

Circular dichroism spectra for estimation of secondary structure were recorded on a Jasco L-720 spectropolarimeter between 185 and 250 nm at 5°C, using a 4-s response time and a scan rate of 10 nm/min. For this analysis, stock solutions of dimerized peptides (5 mg/ml) were diluted 25–50 times.

Computational modeling

Multiple sequence alignments of the HVRs were constructed with the CLUSTALW algorithm (24) using the BLOSUM62 (25) residue substitution scoring matrices. Secondary structure was predicted by PsiPred (26), Jnet (27), and PhD (28) servers, and a consensus structure was calculated by averaging the different predictions weighted with the corresponding residue-by-residue output accuracy.

Other methods

The emm2 gene, encoding the M2 protein, was PCR amplified from strain AP2, a serotype M2 strain in our collections, cloned into plasmid pBR322 and expressed in E. coli LE392. The 3’ sequence (400 nucleotides) of this amplicon was cloned into the E. coli expression vector pCD200 and the 5’ mg was used for the expression of the published emm2 sequence (17). Radiolabeling with 125I of synthetic peptides and proteins was performed by the chloramine-T method or with Bolton and Hunter reagent (Amersham Pharmacia Biotech).
Results

The C4BP-binding regions of different M proteins show little or no residue identity

We analyzed the sequences of five C4BP-binding HVRs, including four previously studied HVRs (11) and the HVR of the M2 protein. The M2 protein is expressed by S. pyogenes strains of a common serotype and binds C4BP (Fig. 1B), and the C4BP-binding region of this M protein is located in the HVR (see below). A multiple sequence alignment (24) of these five C4BP-binding HVRs allowed identification of only three residue identities (Fig. 1C). These three identical residues represent amino acids that are very common in M proteins, and they constitute 37% of all residues in the HVRs compared here. Moreover, the shared sequence motif t-X(2)-E-X(8)-D occurs in many streptococcal proteins that do not bind C4BP, e.g., the fibronectin-binding protein F and the M5 and M6 proteins (data not shown). These data show that the HVRs of different C4BP-binding M proteins exhibit little or no residue identity.

Synthetic peptides derived from HVRs bind C4BP with high specificity

The limited length of the C4BP-binding HVRs suggested that they might correspond to a ligand-binding domain that could be characterized in isolated form. Although previous studies of a protein fragment corresponding to the HVR of a C4BP-binding M protein showed very poor binding of C4BP (11), it seemed possible that this result was due to lack of dimerization, which enhances binding of some ligands to M proteins (29). Synthetic ~50-residue peptides, derived from different C4BP-binding HVRs, were therefore dimerized via a C-terminal cysteine residue and analyzed for binding ability. Three of these peptides (designated M2-N, M4-N, and M22-N) were derived from the C4BP-binding M2, M4, and M22 proteins, and a control peptide (M5-N) was derived from the M5 protein, which does not bind C4BP (11). One of the peptides (derived from the C4BP-binding M22 protein) was also synthesized without the cysteine residue, to allow comparison of the same HVR in dimerized and nondimerized form. In SDS-PAGE, all of these peptides migrated slightly more slowly than expected (Fig. 2A). Weak bands corresponding to higher molecular mass were seen, possibly corresponding to polymeric forms.

The ligand-binding ability of the peptides was analyzed by testing radiolabeled preparations for ability to bind different human proteins immobilized in microtiter wells (Fig. 2B). The dimerized peptides derived from C4BP-binding M proteins bound to C4BP, but not to IgA or IgG. In contrast, the control peptide derived from the M5 protein did not bind to any of the immobilized ligands. This result indicated that the HVRs of C4BP-binding M proteins can indeed be characterized as isolated peptides that retain their binding ability in soluble form. The M22-N peptide lacking a C-terminal cysteine residue showed much reduced binding, compared with the cysteine-containing dimerized form of the same peptide, supporting the conclusion that dimerization is important for binding.

Additional evidence that the dimerized peptides bind C4BP with high specificity was obtained in affinity chromatography experiments, in which whole human serum was applied to columns with immobilized peptides (Fig. 2C). For all three C4BP-binding peptides, analysis of bound proteins demonstrated the presence of a single major ~70-kDa polypeptide, which was identified as the C4BP α-chain, on the basis of N-terminal sequencing and reactivity with specific antisera. A weak band corresponding to a polypeptide of ~75 kDa was similarly identified as protein S, a serum protein associated with C4BP (30). Thus, among all proteins in human serum, the immobilized peptides bound only C4BP, which represents ~0.25% of all serum proteins. This type of analysis could not have been performed with intact M proteins, which bind several human plasma proteins (6, 9).

Different isolated HVRs bind to the same region in C4BP

To determine whether the isolated HVRs bind to the same region in C4BP, inhibition experiments were performed, using radiolabeled peptides and immobilized C4BP (Fig. 3). Binding of each peptide was inhibited by the homologous peptide and by the other two C4BP-binding peptides, but not by the nonbinding M5-N peptide, indicating that the different peptides bind to the same region in C4BP (Fig. 3A).

The conclusion that all C4BP-binding peptides bind to the same site in C4BP was supported by inhibition tests with a 21-residue synthetic peptide, derived from the HVR of the M4 protein (11). This peptide, designated N14, inhibited the binding of all three
C4BP-binding peptides to C4BP (Fig. 3B), but as expected considerably higher concentrations had to be used for this relatively short peptide than for the ~50-residue C4BP-binding peptides, used as controls. The simplest explanation for this result is that the N14 peptide inhibits binding to a site in C4BP used by all three HVRs. The inhibition by the N14 peptide was not specific, because another peptide (N12) derived from the conserved repeat region of M4 caused little or no inhibition. Moreover, the N14 peptide did not inhibit the binding between IgA and a streptococcal IgA-binding peptide (31), confirming that the inhibition was specific (data not shown).

The region in C4BP used for binding was further characterized in inhibition tests using two synthetic peptides derived from the C4BP α-chain. The design of these peptides was based on previous studies suggesting that amino acid residues at the junction between the first two complement control protein (CCP) domains of the C4BP α-chain are important for binding of M protein (Fig. 1A; Refs. 12 and 32). An 18-residue peptide (PA) derived from this part of the α-chain inhibited binding of all three C4BP-binding peptides, while a peptide (PD) derived from the C-terminal part of CCP2 did not inhibit binding (Fig. 3C). The specificity of the inhibition was confirmed by control experiments similar to those described above (data not shown). These results indicate that all three HVRs bind to the same region of C4BP, at the CCP1–2 interdomain region, in agreement with studies of intact M proteins (11, 32).

**Structural analysis**

The binding properties and evolutionary relatedness of the HVRs studied here suggested that they might have similar structures, despite the extreme sequence divergence. This hypothesis was analyzed by CD spectroscopy and computational modeling.

CD spectroscopy at 5°C indicated that the three C4BP-binding peptides are 30–50% α-helical with CD spectra showing minima around 208 and 220 nm and a maximum below 200 nm (Fig. 4A). However, unlike whole M proteins (33, 34), the dimerized HVRs appeared not to have a coiled-coil structure, because the ratio of ellipticity at 220 and 208 nm was <1 (35, 36). Interestingly, the M5-N peptide, which does not bind C4BP, has according to its CD spectrum a structure different from the C4BP-binding peptides. Indeed, the CD spectrum of this peptide indicates that it is mostly random coil with only small amounts of helical structure, if any.

Computational modeling of secondary structure supported the results obtained by CD spectroscopy. The sequences of the five C4BP-binding HVRs (Fig. 1C) were submitted to three different neural network-based prediction servers (PsiPred2, Jnet, and PhD), and a consensus structure was obtained by weighted averaging of the output. With a high level of accuracy, the sequences are predicted to adopt only helix (roughly 60% in each HVR) and coil conformations. For all HVRs except the one of M2, the consensus prediction results in two helices with the connecting loops very well aligned when the sequences are positioned according to the multiple sequence alignment (Fig. 4B). Further, the helix propensity for M2 is lower in the putative loop region, and one of the three neural networks (Jnet) made a coil prediction for M2 in this region. Thus, the C4BP-binding HVRs have very similar predicted secondary structures. The predicted loop occurs at the sequence EN that is conserved in four of the five HVRs (but corresponds to the sequence KI in the M2 protein). The predicted helix propensity gradually decreases in the N- and C-terminal parts. The C-terminal helices would probably extend to the end of the HVRs when attached to the rest of the protein, whereas equilibrium between helix and coil is likely at the most N-terminal residues.
Analysis of the sequence of the M5-N peptide, which does not bind C4BP, showed that the most N-terminal helix was found in the secondary structure prediction also for this peptide, but only one of the three servers predicted a second, C-terminal, helix in this sequence. However, the limited amount of data does not make it possible to determine the significance of this difference in predicted secondary structure between M5-N and the C4BP-binding peptides.

To predict a possible tertiary structure of the C4BP-binding HVRs, the Protein Data Bank (37) was searched using iterative methods that find weak similarities by linking several intermediate sequences (38), but no homologues of known structure were found. However, analysis with several fold-recognition methods (39–42) indicated that all five C4BP-binding HVRs, including that in the M2 protein, adopt a helix-turn-helix conformation (data not shown). Thus, both CD spectroscopy and computational modeling support the hypothesis that the different C4BP-binding HVRs have similar folds.

The C4BP-binding HVRs are immunologically unrelated
Although the sequence variability in HVRs most likely is the result of an immunological selection favoring the appearance of antigenic variants, it seemed possible that different C4BP-binding HVRs might show some cross-reactivity. Indeed, the properties of the different C4BP-binding HVRs support the hypothesis that they have similar structure and binding surfaces, implying that some Abs might be expected to show cross-reactivity. This problem could not be analyzed with intact C4BP-binding M proteins, which cross-react due to extensive residue identity in the C-terminal parts (43, 44) but could be studied with the synthetic peptides described here. Antisera raised against the different peptides completely lacked cross-reactivity, as shown in a direct binding test with immobilized peptides (Fig. 5). Similar results were obtained in inhibition tests, i.e., binding of Abs to a peptide could be inhibited by the homologous peptide, but not at all by the heterologous peptides (Fig. 5B). The lack of cross-reactivity was not due to lack of Abs directed against the C4BP-binding part of the peptides.
because each antiserum inhibited the binding of C4BP to the corresponding peptide (data not shown). Together, these data show that the different C4BP-binding HVRs lack detectable immunological cross-reactivity.

Discussion

The extensive knowledge that is available about streptococcal M proteins, and the lack of variability within a given strain, make this bacterial virulence factor an attractive model system for studies of antigenic variation. An important aspect of such work is to analyze the structure and function of the N-terminal HVR, which is the major target for protective Abs (5). In this study, we have presented evidence that the HVR of many M proteins constitutes a distinct ligand-binding domain that can be studied in isolated form, as a synthetic peptide that binds human C4BP with high specificity. Importantly, ligand-binding was found to be strongly enhanced by dimerization, presumably reflecting the structure of the HVR in the intact M protein, in which the conserved part is a dimeric coiled-coil (29, 33). Because the N-terminal HVR does not have a coiled-coil structure, as shown here, it may be viewed as a frayed end, providing for high affinity binding to C4BP due to two point contact.

Although different HVRs bind the same target, human C4BP, with high specificity, they have very divergent amino acid sequences. A previous study of C4BP-binding HVRs (11) identified five shared residues, which suggested that those residues might be part of a C4BP-binding motif, but the extended analysis reported here demonstrates that different C4BP-binding HVRs almost completely lack residue identity. Thus, different HVRs do not appear to share a sequence motif that can explain their ability to bind C4BP.
although it is commonly believed that such a conserved motif must be present to allow a microbial surface protein to bind a host ligand (45, 46). However, analysis of the three-dimensional structure of different HVRs will be required to determine whether they have a conserved binding site, although a conserved sequence motif could not be identified in the alignment of different HVRs.

The simplest explanation for the ability of different HVRs to specifically bind C4BP is clearly that these HVRs have similar three-dimensional structure and similar C4BP-binding surfaces, although their primary amino acid sequences are different. This hypothesis was supported by CD spectroscopy and computational modeling. However, if different HVRs have similar C4BP-binding surfaces, they might have been expected to exhibit some immunological cross-reactivity. Indeed, when this study was initiated, we hypothesized that different C4BP-binding HVRs might show a limited cross-reactivity, which possibly could be exploited in the development of a broadly protecting vaccine. Support for this notion comes from studies of the extremely variable gp120 protein of HIV-1, indicating that even regions with highly divergent sequences may exhibit some cross-reactivity (3, 47). Although the lack of cross-reactivity between different C4BP-binding HVRs is not encouraging with regard to vaccine development, it is of considerable interest for studies of antigenic variation, because it bears witness to the extraordinary ability of microorganisms to escape immune attack.

It is instructive to consider the structure of different C4BP-binding HVRs and their lack of cross-reactivity in the light of information available about antigenic variation in the extensively studied hemagglutinin (HA) of influenza virus and gp120 of HIV-1 (3, 48–50). Both of these viruses have conserved ligand-binding sites composed of amino acid residues that are located far apart in the primary sequence but nevertheless can be identified in sequence alignments. In HA, the conserved binding site does not elicit cross-reacting Abs, possibly because this sialic-acid-binding site covers an area smaller than the footprint of an Ab, which also binds to surrounding variable regions (49, 51, 52). However, there is evidence (53) that the specificity of Ab binding may be determined mainly by the V\textsubscript{H} complementarity-determining region 3, which has a surface area similar to that of the sialic acid-binding site in HA (52, 54), whereas other parts of the Ab-binding site may be very cross-reactive. Thus, it is unclear why the ligand-binding sites in different HAs do not cross-react. The lack of cross-reactivity between different HVRs of M proteins is even more remarkable, because these HVRs participate in a protein-protein interaction and would be expected to cover an area similar to that covered by the entire Ag-binding region of an Ab (55).

In gp120 of HIV-1, an Ab that recognizes the conserved CD4-binding site can exhibit broad cross-reactivity between different gp120 molecules (3, 47), but such Abs appear to be of little importance in HIV-1 infections, possibly because the conserved ligand-binding site is poorly immunogenic, due to glycosylation, tolerance, or structural instability (3, 56). Similar mechanisms might prevent the appearance of Abs that cross-react with different HVRs of M proteins, except that the prokaryotic M proteins most likely are not glycosylated. For example, a large conformational change might occur on binding, such that the C4BP-binding site is not exposed or assembled in the free M proteins. It is also conceivable that the ligand-binding site in M proteins is not accessible to Abs because it is located in a canyon (57), but structural analysis has not provided support for this mechanism in the extensively studied rhinovirus system (58).

This study was focused on the ability of synthetic peptides derived from different HVRs to bind the same host ligand, human C4BP, rather than on the exact function of bacteria-bound C4BP in streptococcal infections. However, the demonstration that different isolated HVRs bind C4BP with high specificity fits well with the finding that this interaction contributes to phagocytosis resistance and therefore is biologically important (K. Berggård and G. Lindahl, manuscript in preparation). C4BP also binds to strains of two major Gram-negative pathogens, Bordetella pertussis and Neisseria gonorrhoeae (59–61), and was reported to contribute to serum resistance in N. gonorrhoeae (60), suggesting that several different pathogens may exploit C4BP to evade attack from the immune system of the host.

In conclusion, our studies show that the extremely variable C4BP-binding HVR in M proteins corresponds to a distinct ligand-binding domain that can be reproduced in dimerized synthetic peptides. These findings may lead to speculations that extremely variable regions in other microbial surface proteins, such as the pilin of Neisseria or the variable surface glycoprotein of Trypanosoma brucei (1, 45, 62), may also represent ligand-binding domains. Finally, our finding that C4BP-binding HVRs can be studied as isolated ligand-binding domains indicates that these regions are well suited for structural analysis. Such studies will be of interest for analysis of the evolutionary forces that give rise to antigenic variation.

Acknowledgments
We thank Margaretha Stählhammer-Carlsmalm for generous help, Björn Dahlbäck for the gift of purified C4BP, Thomas Areschoug for advice, and Ulla Regnér for technical assistance.

References