

STRUCTURAL BIOLOGY

Variability without change

Sequence variability in microbial surface proteins can be extensive in order to promote immune escape, but the changes must not interfere with the function of the protein. Structural analysis of a streptococcal protein now clarifies how an extremely variable region retains its ability to specifically bind its human target.

Gunnar Lindahl and Jenny J. Persson

All pathogenic microorganisms express one or more surface proteins that promote the establishment of an infection. In response, the host produces specific antibodies (Abs) that can block these proteins, but microorganisms have evolved several mechanisms to evade Ab attack. In particular, many microbial surface proteins vary extensively in sequence. This results in antigenic divergence and allows a pathogen expressing one protein variant to escape Abs elicited by another variant and thus keep ahead of host immune surveillance. This is a major obstacle in vaccine design, especially against several major human pathogens, including HIV and the parasite that causes malaria, where vaccines based on one or few variants will only confer limited protection.

While it may appear impossible to design a vaccine that would induce Abs that

recognize all variants of a highly variable microbial protein, the situation is not hopeless. Different variants of a microbial surface protein typically retain the same or a similar 3D structure, reflecting similar function¹⁻³. Thus, there must be severe constraints on the variability, even if this is not clear from sequence alignments. This is how it has been possible to identify monoclonal broadly neutralizing Abs (bNAbs) that can recognize structurally conserved sites in variable microbial proteins and block protein function^{2,3}. However, these Abs probably arise only rarely during an infection, and it has not yet been possible to develop any vaccine that efficiently elicits bNAbs.

In this issue of *Nature Microbiology*, Buffalo *et al.*⁴ shed fresh light on the interactions between a variable microbial protein and a host ligand, and, by inference, also on vaccine development. The authors

studied the Gram-positive bacterium group A streptococcus (GAS), also known as *Streptococcus pyogenes*, a common human pathogen that causes a variety of diseases, including acute pharyngitis (strep throat) and the autoimmune disease rheumatic fever. The most extensively studied GAS virulence factor is the surface-localized M protein, which was discovered by Lancefield almost 90 years ago and was the first bacterial surface protein implicated in virulence⁵. This fibrillar coiled-coil protein⁶ is best known for its ability to inhibit phagocytosis, a property that can be explained by the ability of M protein to prevent complement attack.

In the complement system, C3 convertases are critical proteases that cleave C3, resulting in deposition of C3b. For Gram-negative bacteria, deposition of C3b may result in the formation of a lytic complex, but Gram-positive

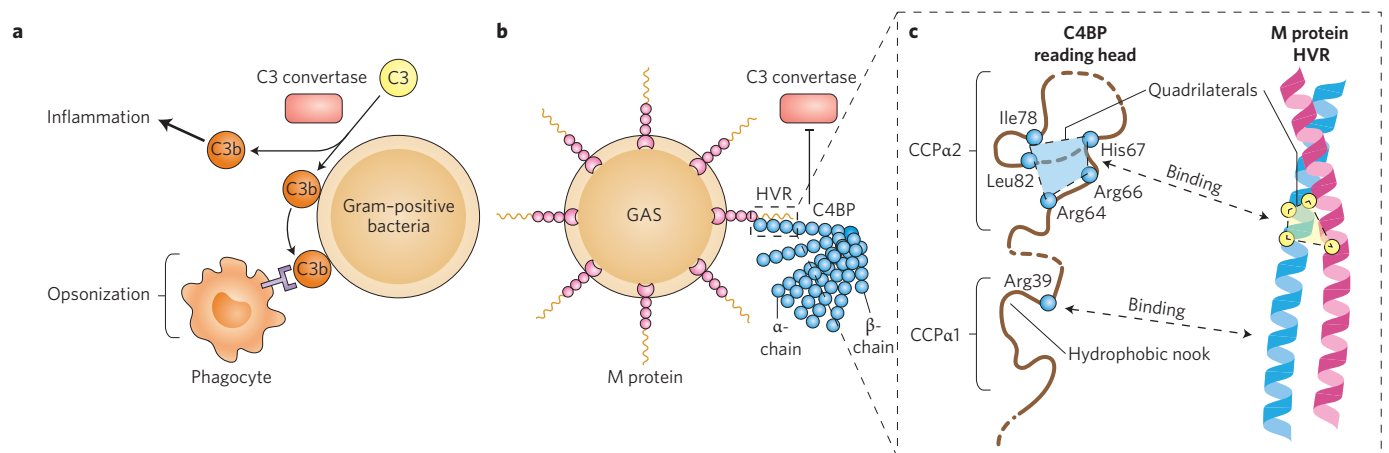


Figure 1 | C4BP interaction with GAS M protein. **a**, Simplified diagram of complement C3-dependent opsonization of a Gram-positive bacterium. A C3 convertase cleaves C3 to generate inflammatory signals and deposit complement proteins on bacterial surfaces, thereby promoting phagocytosis. Note that Gram-positive bacteria are resistant to the lytic effects of complement. **b**, The surface of GAS strains are covered by a fibrillar coiled-coil protein, M protein, that has an N-terminal HVR. The HVR varies in sequence among strains, but normally not within a strain, allowing the identification of approximately 200 distinct M protein types. In many M proteins, the HVR binds the human complement inhibitor C4BP (dashed box). The recruitment of C4BP antagonizes C3 convertase activity, leading to suppression of phagocytosis. **c**, A magnified view of the C4BP interaction shows the HVR of an M protein binding to a reading head in C4BP. Most of the reading head is located in the CCPα2 domain of C4BP, where five amino acids form a quadrilateral that interacts with a complementary quadrilateral in the HVR. The reading head also includes a functionally important 'hydrophobic nook' in CCPα1, where an arginine residue participates in both electrostatic and hydrophobic interactions with the HVR. The unique properties of the C4BP reading head allow it to be highly tolerant to sequence variation in the HVR, where it recognizes patterns of relatively conserved residues.

bacteria like GAS are resistant to this lytic function. Instead, GAS is susceptible to C3b's ability to promote opsonization and engulfment of microbes by phagocytes (Fig. 1a). Consequently, GAS harbours multiple mechanisms to avoid opsonization mediated by C3b. One major mechanism occurs through M protein recruitment of C4b-binding protein (C4BP), a high-molecular-weight host plasma protein that interferes with the function of one of the C3 convertases. In many (but not all) M proteins, C4BP binds to the N-terminal hypervariable region (HVR), and its recruitment blocks complement activation and opsonization^{7,8} (Fig. 1b). Though host-protective Abs are commonly directed against the HVR, the high variability in this region (and its weak immunogenicity⁹) enables individual strains to escape Ab attack. Remarkably, the evolution of escape mutants has resulted in the complete lack of residue identities among M protein HVR sequences⁷. But how can such extreme sequence divergence be reconciled with the fact that many of the HVRs retain their ability to bind C4BP?

Buffalo and colleagues⁴ provide molecular insight into this problem by solving the crystal structure of complexes formed between four highly divergent HVRs and the binding region of C4BP (the CCP α 1 and 2 domains). Interestingly, all four complexes had similar structure, and analysis showed that a 'reading head' in C4BP contains six key residues that allow for recognition of complementary structures in the HVRs. Most of the reading head is located in CCP α 2, where five amino-acid residues form a quadrilateral that interacts with a complementary quadrilateral in the HVR (Fig. 1c). Hidden HVR

patterns responsible for the binding were discerned — for example, a hydrophobic residue in the HVR always interacts with a hydrophobic pocket in C4BP. These HVR patterns are apparently of at least two types, each pattern being characterized by the presence of a few residues that are conserved but not identical, and they would otherwise have been undetected without structural analysis.

Several features of the C4BP reading head may contribute to its ability to recognize HVRs with highly divergent sequences. First, the reading head includes three arginine residues, which can engage in both electrostatic and hydrophobic interactions. Second, the binding energy is apparently broadly dispersed over the C4BP–HVR interaction site, as indicated by the limited impact of amino acid substitutions in the HVR. Finally, the HVR and the reading head may align in multiple ways, providing further flexibility to the system. Thus, the system is strikingly tolerant to sequence variability, a finding that can explain the emergence of HVRs that retain the ability to bind C4BP while being antigenically unrelated¹⁰.

Do these findings open the way to the design of a GAS vaccine? Although the finding that C4BP-binding HVRs have similar structure is encouraging, many problems remain, as witnessed by the extensive efforts to develop a vaccine against HIV. In that system, bNAbs that recognize conserved structures have been identified^{2,3}, but the design of a vaccine that elicits the production of such bNAbs has encountered major difficulties¹¹. In the M protein system, it may similarly be possible to identify bNAbs that recognize C4BP-binding HVRs, but the problem of vaccine

design remains. Moreover, a number of clinically important M proteins, such as M1, M3 and M5, have HVRs that do not bind C4BP, implying that these M types would not be covered by a vaccine solely focused on C4BP-binding HVRs. Finally, the use of any GAS vaccine could raise major safety concerns, as it is not clear if the immune response could trigger the autoimmune disease rheumatic fever.

In summary, the intriguing structural study by Buffalo *et al.*⁴ provides a molecular explanation for the ability of highly divergent M protein domains to bind the same human protein. It will be interesting to know whether similar mechanisms promote ligand-binding to other variable microbial domains, both in M proteins^{12,13} and in other pathogens, such as the malaria parasite¹⁴. □

Gunnar Lindahl is in the Department of Veterinary Disease Biology, University of Copenhagen, 1870 Frederiksberg C, Denmark. Jenny J. Persson is in the Department of Experimental Medical Science, Lund University, SE-221 84 Lund, Sweden.
e-mail: gunnar.lindahl@sund.ku.dk

References

- Blum, M. L. *et al.* *Nature* **362**, 603–609 (1993).
- Burton, D. R. & Hangartner, L. *Annu. Rev. Immunol.* **34**, 635–659 (2016).
- Lanzavecchia, A. *et al.* *Curr. Opin. Immunol.* **41**, 62–67 (2016).
- Buffalo, C. Z. *et al.* *Nat. Microbiol.* **1**, 16155 (2016).
- Todd, E. W. & Lancefield, R. C. *J. Exp. Med.* **48**, 751–767 (1928).
- Phillips, G. N. *et al.* *Proc. Natl Acad. Sci. USA* **78**, 4689–4693 (1981).
- Persson, J., Beall, B., Linse, S. & Lindahl, G. *PLoS Pathog.* **2**, e47 (2006).
- Sanderson-Smith, M. *et al.* *J. Infect. Dis.* **210**, 1325–1338 (2014).
- Lannergård, J. *et al.* *Cell Host Microbe* **10**, 147–157 (2011).
- Morfeldt, E. *et al.* *J. Immunol.* **167**, 3870–3877 (2001).
- Dosenovic, P. *et al.* *Cell* **161**, 1505–1515 (2015).
- Sandin, C. *et al.* *J. Immunol.* **169**, 1357–1364 (2002).
- Macheboeuf, P. *et al.* *Nature* **472**, 64–68 (2011).
- Klein, M. M. *et al.* *PLoS Pathog.* **4**, e1000147 (2008).