Published in final edited form as:

Biochem J. 2015 January 15; 465(2): 325–335. doi:10.1042/BJ20141069.

Structural Determinants of Host Specificity of Complement Factor H Recruitment by *Streptococcus pneumoniae*

David Achila^{1,†}, Aizhuo Liu^{1,†}, Rahul Banerjee¹, Yue Li¹, Erik Martinez-Hackert¹, Jing-Ren Zhang², and Honggao Yan¹

¹Department of Biochemistry and Molecular Biology, Michigan State University, 603 Wilson Road, East Lansing, MI 48824, USA

²Center for Infectious Disease Research, School of Medicine, Tsinghua University, Beijing, China

Abstract

Many human pathogens have strict host specificity, which affects not only their epidemiology but also development of animal models and vaccines. Complement factor H (FH) is recruited to pneumococcal cell surface in a human-specific manner via the N-terminal domain of the pneumococcal protein virulence factor CbpA (CbpAN). FH recruitment enables *Streptococcus pneumoniae* to evade surveillance by human complement system and contributes to pneumococcal host specificity. The molecular determinants of host specificity of complement evasion are unknown. Here we show that a single human FH domain is sufficient for tight binding of CbpAN, present the crystal structure of the complex, and identify the critical structural determinants for host-specific FH recruitment. The results offer new approaches to development of better animal models for pneumococcal infection and redesign of the virulence factor for pneumococcal vaccine development, and reveal how FH recruitment can serve as a mechanism for both pneumococcal complement evasion and adherence.

Keywords

Host microbe interaction; host specificity; complement factor H recruitment; protein-protein interaction; structural biology; *Streptococcus pneumoniae*

INTRODUCTION

The complement system is a major component of vertebrate innate immunity and one of the most important frontline defenders against microbial invaders [1, 2]. Uncontrolled or inappropriate complement activation, however, can be damaging to the host. A major player

Correspondence should be addressed to: H.Y. (yanh@msu.edu; Tel: 517-353-5282; Fax: 517-353-9334) or J.R.Z. (zhanglab@tsinghua.edu.cn; Tel: 10-6279-5892; Fax: 10-6279-5892).

[†]These authors contributed equally to this work.

AUTHOR CONTRIBUTIONS

D.A. developed protein production and purification protocols, purified all proteins, and did all ITC experiments; A.L. performed NMR experiments, NMR data analysis and structure determination, crystallization screening and optimization, and X-ray diffraction data collection; R.B. carried out molecular dynamics simulations and MM-GBSA analysis; E.M.H. phased the X-ray data using the NMR structure determined by A.L. as the search model and built and refined the crystal structure; Y.L. made the overexpression systems and site-directed mutants; J.R.Z. and H.Y. conceived the project; H.Y. supervised this work and coordinated the writing of the manuscript.

in regulating complement activation and preventing complement from damaging host tissues is complement factor H (FH) [3, 4], a 155-kDa protein with 20 homologous domains termed complement control protein (CCP) modules. FH recruitment is a mechanism widely exploited by microbial pathogens to evade complement attack [2, 3]. Recent studies revealed that human pathogens, *S. pneumoniae*, *N. meningitidis* and *Neisseria gonorrhoeae*, recruit FH in a human-specific manner, indicating FH recruitment as a contributing factor to the host specificity of these pathogens [5–7]. Conversely, the broad host range of the zoonotic bacterial pathogen *Borrelia burgdorferi* sensu lato is attributed partly to its many FH binding proteins [8, 9].

S. pneumoniae is a major cause of human acute otitis media, pneumonia, and meningitis, incurring high morbidity and mortality, particularly in children and the elderly, and enormous economic burden in both developed and developing countries [10, 11]. Under natural conditions, *S. pneumoniae* displays strict host specificity to humans [12]. Animal studies showed that the complement system is essential for immunity to *S. pneumoniae* [13–15]. Clinical surveys indicated that patients with complement deficiencies have increased susceptibility to recurrent *S. pneumoniae* infections [16–18].

Two most critical aspects of pneumococcal pathogenesis are adherence and immune evasion [19]. The former allows pneumococci to colonize in the human host and the latter enables pneumococci to evade the attack of the human immune system. CbpA, also termed PspC or SpsA, is a major pneumococcal protein virulence factor and plays an important role in both pneumococcal adherence and immune evasion [19]. CbpA is a multidomain protein attached to pneumococcal cell wall via its C-terminal phosphorylcholine-binding domain. Its middle homologous R1 and R2 domains mediate pneumococcal adherence and transcytosis across mucosal respiratory epithelial cells by binding to human polymeric immunoglobulin receptor (pIgR) [20–24]. Its N-terminal domain (CbpAN) promotes complement evasion by binding to human FH (hFH) [25, 26]. Interestingly, CbpA binds only to human pIgR and FH, but not to the counterparts of other animal species tested thus far, suggesting that these CbpA-host interactions contribute to the strict host specificity of *S. pneumoniae* [7, 21]. Animal model studies demonstrated that CbpA-deficient pneumococcal strains have attenuated capacity to colonize and cause infections [27–29]. CbpA is one of a few pneumococcal protein virulence factors that can stimulate the production of antibodies in humans [30, 31] and offers protection against challenge of virulent pneumococci in animal models [27, 28, 32–35].

While it is well established that CbpA binds to hFH via its N-terminal domain [25, 26], which part of hFH is involved in binding of CbpA has not been definitively established. Previous studies indicated that CbpA binds to hFH domains CCP6–10 [36] or CCP13–15 [37], but a more recent study indicated that CbpA interacts with two regions of hFH, CCP8–11 and CCP19–20 [26]. Furthermore, the structural basis for host specificity of FH recruitment is not known. Here we show that a single hFH domain is sufficient for binding CbpAN, although the adjacent domains also contribute to the binding to a small extent. We present the solution NMR structure of CbpAN and the crystal structure of the hFH domain in complex with CbpAN. By computational analysis and reciprocal site-directed mutagenesis, i.e., substitution of hFH residues with mouse FH (mFH) residues and vice

versa, we demonstrate that a “hydrophobic lock” on the hFH domain is a critical structural determinant for host specific FH recruitment.

EXPERIMENTAL

Construction of the CbpAN overexpression system and protein purification

The domain boundaries of CbpAN of the TIGR4 strain were defined based on the secondary structure prediction using the SOPMA server (http://npsa-pbil.ibcp.fr/cgi-bin/npsa_automat.pl?page=npsa_sopma.html) [38]. The DNA fragment encoding the CbpAN domain (Asp68 to Lys148) was amplified by PCR from the genomic DNA of the *S. pneumoniae* strain TIGR4 with the primers CbpANf and CbpANr (Supplementary Table 1). The amplified DNA fragment was cloned into the expression vector pET17bHR (lab-made, derived from pET17b) by digestion with the restriction enzymes BamHI and NdeI and ligation. The cloned DNA fragment was sequenced to ensure the correct coding sequence. The overexpression plasmid construct was transformed into the *E. coli* strain BL21(DE3)pLysS for the production of the His-tagged CbpAN. The expression system was cultured in LB media containing 20 µg/ml chloramphenicol and 100 µg/ml ampicillin with vigorous shaking at 37 °C. The culture was placed on ice when its OD₆₀₀ reached 0.8. The production of CbpAN was then induced with 0.5 mM IPTG and the culture was incubated with shaking overnight at 16 °C. The cells were harvested by centrifugation and resuspended in buffer A containing 50 mM sodium phosphate and 300 mM NaCl, pH 8.0, and lysed by sonication. The lysate was clarified by centrifugation at 16,000 g for 20 min before loading onto a Ni-NTA agarose column pre-equilibrated with buffer A. After washing with 10 mM imidazole in buffer A, the column was eluted with a linear 10–250 mM imidazole gradient. The protein-containing fractions were analyzed by SDS-PAGE. Fractions containing pure CbpAN were pooled and concentrated using an Amicon concentrator. When desired, the His-tag was cleaved with TEV protease [39]. The cleaved His-tag and the uncleaved protein were separated by a second passage through a Ni-NTA agarose column. The flow-through was collected, concentrated and loaded onto a Sephadex G-75 gel filtration column. The fractions containing pure CbpAN were pooled, concentrated, and dialyzed first against 5 mM sodium phosphate, pH 8.0, and then against water for 12 h. The dialyzed protein solution was lyophilized and stored at –80 °C. Isotopically labeled (¹⁵N- or ¹⁵N/¹³C-) CbpAN was produced and purified in the same way except that LB media were replaced with M9 media containing ¹⁵NH₄Cl as the sole nitrogen source or ¹⁵NH₄Cl and ¹³C-glucose as the sole nitrogen and the sole carbon source, respectively.

Construction of the overexpression systems for FH domains and protein purification

Synthetic hFH and mFH coding DNA with codons optimized for expression in *E. coli* were synthesized by Biomatik. DNA fragments encoding various CCPs were cloned into the expression vector pET17bHMHT (lab-made, derived from pET17b (Novagen)) using the same recombinant DNA methods as described earlier. The primers for PCR cloning were hFH9f and hFH9r for hFH CCP9, hFH8f and hFH10r for hFH CCP8-10, hFH9f and hFH10r for hFH CCP9-10, hFH13f and hFH15r for hFH CCP13-15, and hFH19f and hFH20r for hFH CCP19-20. The correct coding sequences were confirmed by DNA sequencing. The overexpression plasmid constructs were transformed into the *E. coli* strain SHuffle T7

Express *LysY* (NEB). The fusion proteins produced by the overexpression systems contained two His-tags, one before the maltose binding protein (MBP) and another between the MBP and the CCPs, and a thrombin cleavage site between the second His-tag and the CCPs. The procedure for protein production and cell harvesting was the same as described earlier. The harvested *E. coli* cells were resuspended in buffer B containing 20 mM Tris-HCl and 150 mM NaCl, pH 7.5, and lysed by sonication. The lysate was clarified by centrifugation and loaded onto a Ni-NTA agarose column pre-equilibrated with buffer B. After washing with buffer B containing 20 mM imidazole, the column was eluted with a linear 20–250 mM imidazole gradient. The protein-containing fractions were analyzed by SDS-PAGE, pooled and concentrated. An equimolar amount of lab-made His-tagged *E. coli* DsbC[40] was added to the concentrated fusion protein solution for disulfide bond reshuffling. After an overnight dialysis against buffer B at room temperature, a CaCl₂ stock solution was added to a final concentration of 0.125 mM, followed by the addition of 0.75 U thrombin/mg of the fusion protein to cleave off the MBP tag. After 3 h of incubation at room temperature, PMSF was added to a final concentration of 0.5 mM to inactivate thrombin and therefore terminate the cleavage reaction. The uncleaved fusion protein and the cleaved MBP were removed by chromatography on a second Ni-NTA agarose column eluted with a linear 0–100 mM imidazole gradient in buffer B. The fractions containing CCPs were concentrated and further purified by gel filtration chromatography on a Sephadex G-75 column. The purified protein solutions were dialyzed first against 5 mM sodium phosphate, pH 8.0, then against 2 mM sodium phosphate, pH 8.0, and lyophilized. The lyophilized proteins were stored at –80 °C.

Site-directed mutagenesis

FH mutants were generated using the PCR-based QuickChange method according to the protocol developed by Stratagene and the primers listed in Supplementary Table 1. Mutations were selected and correct coding sequences verified by DNA sequencing. Multiple mutations were generated in a sequential fashion.

Isothermal titration calorimetry (ITC)

ITC measurements were performed at 25 °C on a VP-ITC isothermal titration calorimeter (MicroCal) according to the protocol of Velazquez-Campoy and Freire [41]. The lyophilized proteins were dissolved in buffer C containing 50 mM HEPES and 50 mM KCl, pH 7.5, and dialyzed against the same buffer. The concentrations of the dialyzed protein solutions were determined by measuring OD₂₈₀ and using extinction coefficients calculated by the method of Gill and von Hippel [42] using the ProtParam tool of the ExPASy web server [43]. The sample cell was loaded with a FH protein and the syringe with CbpAN. The binding enthalpies were obtained by injecting CbpAN into the sample cell under stirring conditions. A typical ITC experiment consisted of 25 injections 10 µl each at 6 min intervals. The ITC data were analyzed using Origin 5.0.

Solution structure determination

All protein NMR samples were prepared by dissolving the lyophilized cleaved CbpAN in a pH 7.0 buffer composed of 50 mM phosphate, 50 mM NaCl, 100 µM NaN₃, and 50 µM DSS (as the internal NMR reference) in either 95% H₂O:5% D₂O or 100% D₂O. For the solution

structure determination of free CbpAN, protein concentration was typically ~1.0 mM. For the NMR assignment of CbpAN in complex with hFH CCP9, the samples were prepared by mixing 1.0 mM labeled CbpAN with 1.1 mM unlabeled hFH CCP9. For assaying the binding of CbpAN to different hFH CCPs, each NMR sample consisted of 0.2 mM ^{15}N -CbpAN and 0.5–1.0 mM unlabeled hFH CCPs.

All NMR data were acquired at 20 °C on a Bruker AVANCE 900 MHz NMR spectrometer equipped with the TCI cryo-probehead. The raw NMR data were processed using NMRPipe [44] and analyzed with NMRView [45]. Sequential resonance assignments were accomplished by the analysis of standard 3D double- and triple-resonance NMR data [46]. Backbone resonance assignment was achieved by the analysis of a set of standard triple-resonance NMR data acquired with $^{13}\text{C}/^{15}\text{N}$ -labeled CbpAN in the buffer containing 5% D_2O , including HNCA, HN(CO)CA, HNCACB, CBCA(CO)NH, HNCO, and HN(CA)CO. The assignment was confirmed by the analysis of a 3D ^{15}N -edited NOESY spectrum acquired with a ^{15}N -labeled CbpAN sample. The resonance assignment of aliphatic side chains was accomplished through the analysis of 3D HCCH-TOCSY data, and the assignment of aromatic side chains through the analysis of HBCB(CG)CDHD and HBCB(CG)(CD)CEHE NMR data. The same strategy was used for the sequential assignment of labeled CbpAN in complex with unlabeled hFH CCP9. 3D ^{13}C -dispersed NOESY spectra were recorded using a protein sample in the buffer made in 100% D_2O . The mixing times for all NOESY spectra were set to 70 ms.

The solution structure of free CbpAN was calculated using CYANA [47] (version 2.1) with inputs of inter-proton distance restraints derived from NOEs and backbone torsion-angle restraints derived from chemical shifts using TALOS [48]. The top 20 conformers generated from 50 initial random ones by the CYANA calculations (with the lowest values of the target function) were subjected to energy minimization in explicit solvent using the program AMBER 10 [49] and the ff99SB force field. Each of 20 conformers was neutralized with a Cl^- ion and solvated in a rectangular TIP3P water box with its boundaries at least 10 Å away from the protein atoms. The system was minimized in two stages. In the first stage, the water molecules and Cl^- ion were minimized by 15,000 steps of steepest descent followed 5,000 steps of conjugated gradient with the conformation of the protein molecule fixed. In the second stage, the whole system was minimized with 25000 steps of steepest descent followed by 15,000 steps of conjugated gradient with the protein molecule restrained only by the distance and torsion-angle restraints derived from the NMR data. The statistics of the AMBER-refined structure are very similar to those of the original structure obtained by the CYANA calculations except that the number of hydrogen bonds increased dramatically, from 38 before to 61 after energy minimization.

Crystal structure determination

The complex of CbpAN and hFH CCP9 were made by mixing the two proteins with a molar ratio of 1.1 to 1.0. The mixed protein solution was concentrated and further purified by FPLC using a HiLoad 26/60 gel filtration column. The complex emerged as a sharp peak well separated from the free CbpAN. The fractions containing the protein complex were pooled, concentrated to ~10 mg/ml, and dialyzed against a buffer containing 10 mM Tris-

HCl, pH 8.0. The dialyzed protein solution was concentrated up to ~20 mg/ml and centrifuged to remove any potential insoluble material. Crystallization screening was performed with a Cryphon robot (Art Robbins Instrument) using QIAGEN screening kits and the sitting drop method. The crystallization condition was optimized manually by the hanging drop vapor diffusion method at 20 °C. The optimized well solution was composed of 100 mM sodium acetate, pH 4.6, and 18–20% (w/v) PEG 3350. Crystals appeared overnight and reached within 3 days typically the size of $0.25 \times 0.25 \times 0.35$ mm. The crystals were then transferred into reservoir buffers containing sequentially higher concentrations of glycerol (5–25%, v/v) and flash frozen in liquid nitrogen. X-ray data from frozen crystals were collected at the Advanced Photon Source, LS-CAT beamline 21-ID. Crystals diffracted to ~1.10 Å Bragg spacings (for processing and refinement statistics, see Supplementary Table 3). Two datasets were used in the structure determination, including a 1.08-Å native dataset and a 2.29-Å dataset from a NaI soaked crystal. Of the 20 hFH CCPs, no structure has been reported for six CCPs, including CCP9. Phasing with the crystal structures of other CCPs was not successful. An MR solution was found for the 2.29 Å NaI dataset using Phaser [50] with the CbpAN NMR structure as the search model. Subsequent model building and refinement with ARP/wARP [51] and REFMAC5 [52] using the 1.08 Å native data set yielded essentially a complete model, including the missing hFH CCP9. Electron density was clear and continuous for both subunits and the only disordered regions were found at the N- and C-termini. The model was further refined by simulated annealing using PHENIX [53]. Small errors were corrected manually with Coot [54]. Water molecules were added during the final rounds of refinement using PHENIX [53]. Figures were prepared using PyMOL [55].

Molecular dynamics (MD) simulations and MM-GBSA analysis

All MD simulations were performed using AMBER 12 [56] and the ff12SB force field. The initial coordinate for the MD simulation of the complex of hFH CCP9 and CbpAN was taken from the crystal structure of the complex. The initial coordinate for the MD simulation of the complex of mFH CCP9 and CbpAN was obtained by homology modeling. A model of the bound mFH CCP9 was built using the SWISS-MODEL server [57–59] with the structure of the bound hFH CCP9 as the template. The sequence identity between hFH and mFH CCP9 is 65%. The model of the complex of mFH CCP9 and CbpAN was assembled with the crystal structure of the bound CbpAN and the modeled structure of mFH CCP9. The complexes of FH mutant proteins were generated by assuming that the mutations have no significant effects on the structures of the proteins. Each complex was solvated in a rectangular TIP3P water box, with the boundaries at least 10 Å away from the protein atoms, and neutralized with appropriate number of Na⁺ ions. The system was energy minimized with a combination of steepest descent and conjugated gradient methods, 5,000 steps each, first on the water molecules with all heavy atoms of the proteins restrained and then on the whole system without any positional restraint. The minimized system was heated from 0 to 300 K in 500,000 steps under the constant volume condition (NVT) and equilibrated for 4 ns under the isobaric, isothermal (NPT) conditions at 300 K. The system was simulated further under the NPT conditions at 300 K for 40 ns. Temperature was controlled by Langevin dynamics with a collision frequency of 1.0 ps⁻¹. The time step was 2 fs, with all covalent bonds to hydrogen atoms restrained using the SHAKE algorithm [60]. Long-range

electrostatic interactions were evaluated using the particle mesh Ewald method [61] with a nonbonded cutoff of 10 Å. All MD trajectories are stable based on the time evolution of RMSD (Figure S1).

MM-GBSA analysis and energy decomposition were performed with a single trajectory approach using the python script MMPBSA.py [62] provided with AmberTools 12 [56]. A generalized Born implicit solvent model with 0.1 molar salt concentration was used for the calculations. The calculations were performed on 500 snapshots for each simulation and averaged to estimate the global free energy change of the binding and the contribution from each FH9 and CbpAN residue towards the global free energy change. The entropic contribution to the free energy was not considered in the calculations as in most recent studies [56], because we were interested in only the mutational effects on binding, i.e., the relative free energies, and entropy calculation is not as accurate as the enthalpy calculation.

Accession codes

Protein Data Bank: solution NMR structure of free CbpAN, 2M6U; crystal structure of CbpAN in complex with hFH CCP9, 4K12.

RESULTS

Localization of the CbpA-binding domain in hFH

To localize which domain or domains of hFH bind to CbpA, we first produced several truncated hFH proteins, CCP8-10, CCP13-15 and CCP19-20 and measured their binding activity by the biophysical methods isothermal titration calorimetry (ITC) and NMR. Of these truncated hFH proteins, only CCP8-10 bound to CbpAN (Figure S2). The binding was tight with a K_d of ~1.5 nM and driven by a significant gain of enthalpy (Table 1). The loss of entropy is typical of the formation of a complex of two protein molecules. To check whether all three domains of CCP8-10 are needed for tight binding of CbpAN, we deleted the N-terminal CCP8 first and then the C-terminal CCP10 and measured their effects on binding CbpA (Figure S2 and 1a). The K_d values of the further truncated hFH proteins, CCP9-10 and CCP9, were similar to that of CCP8-10, all under 5 nM (Table 1). We also produced CCP10 and it showed no binding activity based on ITC measurements. These results together indicated that a single CCP9 domain is sufficient for tight binding of CbpAN. CCP8 and CCP10 enhance the affinity of CCP9 for CbpAN, but their contributions to the binding free energy are small. Of the 12 kcal/mol binding free energy for CCP8-10 binding of CbpAN, 11.4 kcal/mol is contributed by CCP9. Furthermore, we purified CCP6-8, CCP10, and CCP10-12 and measured their binding activity by ITC. None of these truncated FH proteins showed detectable binding to CbpAN.

Because CCP13–15 [37] and CCP19–20 [26] were previously reported to bind to CbpA but our ITC measurements could not detect this binding activity of either of these truncated hFH proteins, we further investigated these CCP domains by examining their effects on the chemical shifts of ^{15}N -CbpAN. The chemical shifts of CbpAN and its complex with hFH CCP9 were assigned by multidimensional NMR spectroscopy (Figure S3). Backbone amide resonances were sequentially assigned for all residues of CbpAN in complex with hFH

CCP9 and all but four residues of free CbpAN. The unassigned residues were Lys 95, Lys 96, Arg 97, and His 98, which are located between helix I and II (see the NMR structure in Figure 2a in next section) and whose amide NMR signals were very weak. Addition of hFH CCP9 caused significant changes in the chemical shifts of many residues of CbpAN (Figure 1b). Significant chemical shift changes were mostly congregated in the region spanning residues 85-111 (Figure 1c), the C-terminal segment of helix I and the N-terminal segment of helix II. In contrast, addition of 0.5 mM CCP13-15 (Figure 1d) or 1.0 mM CCP19-20 (Figure 1e) caused essentially no change in the chemical shifts of CbpAN. The NMR chemical shift perturbation experiments indicated that neither CCP13-15 nor CCP19-20 binds to CbpAN even at millimolar concentrations.

The solution NMR structure of free CbpAN and crystal structure of its complex with CCP9

To elucidate the structural basis for the binding specificity, we determined the NMR solution structure of CbpAN and the crystal structure of its complex with hFH CCP9. Over 98% of proton resonances were sequentially assigned, with only the chemical shifts of the amide protons of Lys 95, Lys 96, Arg 97 and His 98 and ϵ -protons of Lys 138 missing. The NMR structure of free CbpAN was determined with 1678 distance and 150 dihedral angle restraints. The solution structure is of high quality, with an average RMSD (to the mean structure) of 0.53 ± 0.10 Å for backbone atoms and 1.04 ± 0.09 Å for all heavy atoms and 97.7% of backbone torsion angles in the most favorable region and the rest (2.3%) in either the favorite or generally allowed region (see Table S2 for more NMR restraint and structure refinement statistics). CbpAN has a compact structure consisting of a bundle of three helices with an up-down-up antiparallel topology (Figure 2a). The first helix spans residues Asp 68–Lys 92, the second helix His 98–His 121, and the third helix residues Glu 125–Glu 147. The pIgR-binding R1/R2 domains of CbpA also consist of three helices, but the packing of the helices of CbpAN is clearly different from that of the R1/R2 domains, which form a flat, raft-like structure[23].

The crystals of the complex of CbpAN and hFH CCP9 were obtained by co-crystallization and diffracted to 1.08-Å resolution. The crystal structure of the complex was determined by molecular replacement using the NMR structure of CbpAN as the searching model, because the structure of CCP9 has not been reported and attempts to solve the complex structure by molecular replacement using the structures of other CCP domains were not successful. The final refined structure has an R_{free} value of 17.03% and R_{work} value of 15.44%.

Ramachandran analysis using MolProbity[63] revealed that 99.3 % of residues reside in geometrically favored regions (see Table S3 for more crystallographic and refinement statistics). The crystal structure of bound CbpAN is very similar to the NMR structure of the free protein with a C^α RMSD of 0.545 Å (Figure 2b), indicating that binding of CCP9 does not cause significant conformational changes in CbpAN. CCP9 has a typical CCP fold stabilized by two disulfide bonds that are formed by four invariant Cys residues (Figure 2b), one between Cys 491 and Cys 535 and the other between Cys 518 and Cys 546. Based on structural alignment, CCP9 is most similar to CCP4 (PDB ID: 2WII) with a C^α RMSD of 1.14 Å for 54 residues and CCP8 (PDB ID: 2V8E) with a C^α RMSD of 1.34 Å for 59 residues (Figure 2c). Significant deviations can be seen only in a short deletion in CCP4 and a short insertion in CCP8 (Figure 2d). The sequence identities of CCP9 to CCP4 and CCP8

both are only 30.2% with many interface residues not conserved, suggesting that the domain specificity of FH binding by CbpAN, i.e., why CbpAN binds to CCP9 but not other CCP domains, is determined by interface residues unique to CCP9.

The interface between CbpAN and hFH CCP9 contains 17 residues from CbpAN and 18 residues from CCP9. All interface residues of CbpAN are from a 22 residue region, residues 86-117, spanning the C-terminal part of the first helix and almost the entire second helix. This region matches almost exactly with the region of CbpAN showing most significant chemical shift changes upon the addition of CCP9 (Fig 1c). The interface residues of CCP9 are dispersed in amino acid sequence, from three segments, the first segment at the N-terminus, the second in the middle, and the third at the C-terminus (Figure 3). Of the 18 interface residues of CCP9, only eight residues are conserved in the sequence alignment of the CCP9 domains of 18 FH proteins (Figure 3).

CbpAN and CCP9 are held together by both polar and nonpolar interactions (Figure 2e). This many interactions are consistent with the substantial gain of enthalpy in binding (Table 1). The polar interactions include six hydrogen bonds, of which four are to the side chains of hFH residues: one to the carboxylate of Glu 523, another to the carboxylate of Asp 542, and two to the hydroxyl of Tyr 547. Of the nonpolar interactions, the most notable are those of a cluster of four hydrophobic residues of hFH, Val 495, Met 497, Leu 543 and Ile 545, with their side-chains holding the phenol group of CbpA Tyr 90 like a lock (henceforth termed “hydrophobic lock”). The phenol group of CbpA Tyr 90 is like a key inserted into the lock with its tip (the phenol hydroxyl) latched to CCP9 via a hydrogen bond to the backbone oxygen of hFH Val 495. Interestingly, of the four residues that constitute the hydrophobic lock, no residue is conserved among the 18 FH proteins (Figure 3). A complete hydrophobic lock is found only in some primate FHs, including human, chimpanzee (not shown, as the amino acid sequence of its CCP9 is identical to that hFH CCP9), *Nomascus leucogenys*, and *Pongo abelii* FHs. Only one residue, Val 495, is found in FH CCP9s of mouse and rat, two animals most frequently used as model animals for studying pneumococcal infections.

The contributions of hFH residues to binding of CbpAN and structural determinants for host specificity of FH recruitment

To investigate the relative importance of the interactions between CCP9 and CbpAN to the stability of the complex, we performed molecular dynamics (MD) simulations followed by MM-GBSA analysis. Based on the percentage of the occurrence of hydrogen bonds in the MD trajectory (Table S4), of the six hydrogen bonds between CCP9 and CbpAN, the most stable hydrogen bond is between the oxygen of hFH Val 495 and the hydroxyl of CbpA Tyr 90, indicating that the lock-key interaction is strong. The MM-GBSA computational analysis suggested that, of the 18 interface residues of hFH, the most important residues for binding of CbpA are Tyr 547 and the four hydrophobic residues that constitute the hydrophobic lock (Figure 4a). The large energetic contribution of hFH Tyr 547 is due to its two stable hydrogen bonds to CbpA (Figure 2e and Table S4) and many van der Waals interactions, with over 80% of its solvent-accessible area buried upon formation of the complex. The contribution of Val 495 to CbpA binding also includes that of the hydrogen bond from its backbone oxygen to the hydroxyl of CbpA Tyr 90. Although Asp 542 does not appear as

important for binding of CbpAN according to the computational analysis, the hydrogen bond between this residue and CbpAN is stable and consistent with the energetic contribution of Asp 542 (Table S4).

To validate the computational analysis, we replaced the side chains of the three CCP9 residues involved in intermolecular hydrogen bonding, Glu 523, Asp 542 and Tyr 547, with Ala. Although the ITC measurement of Y547A is not as accurate as those of E523A and D542A, the ITC data indicated that of these three residues, Tyr 547 is indeed the most important residue for binding CbpAN followed by Asp 542 and Glu 523 (Figure S2 and Table 1).

An issue of great interest is the structural basis of the host specificity of FH binding by CbpA, particularly why CbpA binds to hFH but not mouse FH (mFH). Biologically, this is especially relevant because mouse is the most frequently used animal model for studying pneumococcal pathogenesis and vaccine development[64]. Of the important residues for binding CbpA, based on the crystal structure and the computational and mutagenesis studies, all but three are strictly or highly conserved (Figure 3). These conserved residues include all residues involved in intermolecular hydrogen bonding. The exceptions are three of the four residues that constitute the hydrophobic lock, Met 497, Leu 543 and Ile 545, which are replaced by a negatively charged residue (Glu 497) and two polar residues (Thr 543 and Ser 545) in mFH (Figure 3). Both computational and mutagenesis analyses indicated that all these replacements have significant effects on binding of CbpAN. MM-GBSA analysis suggested that the three replacements together have large effects on binding of CbpAN, not only diminishing the favorable interactions of these replaced residues but also affecting the energetic contributions of other residues, including Val 495, Trp 540 and Tyr 547 (Figure 4b). Individually, the effect of the I545S replacement on binding energy is the largest (~4 kcal/mol), followed by that of the L543T replacement (2.68 kcal/mol) and then that of the M497E replacement (1.88 kcal/mol). In the context of mFH CCP9, the M497E replacement also introduces a large unfavorable effect on binding of CbpAN (Figure 4c). Reconstruction of the hydrophobic lock of mFH CCP9 by replacing Glu 497, Thr 543 and Ser 545 with the corresponding hFH residues gained a large binding energy (Figure 4d), approaching the magnitude of the loss of binding energy when the hydrophobic lock of hFH CCP9 was destructed by replacing the hFH residues with the corresponding mFH residues.

Experimentally, when Met 497, Leu 543 and Ile 545 of hFH CCP9 were replaced, individually, with the corresponding residues in mFH by mutagenesis, the three mutant proteins, M497E, L543T and I545S, all have a significantly reduced affinity for CbpAN (Figure S2 and Table 1). The K_d values of these mutant proteins increased by 40, 560 and 2,600-fold, respectively, relative to that of the wild-type CCP9, corresponding to a decrease in binding energy by 1.36, 4.66 and 5.33 kcal/mol, respectively. The extents of effects of the mutations on binding of CbpAN were very similar to those calculated by MM-GBSA analysis. When hFH CCP9 was altered with all the above three replacements, binding of CbpAN was no longer detectable by ITC.

To further validate that the hydrophobic lock is the critical structural determinant of host specificity of FH binding, we produced mFH CCP9. The binding of the wild-type mFH

CCP9 to CbpAN was not detectable by ITC. When Glu 497, Thr543 and Ser 545 of mFH CCP9 were replaced with the corresponding hFH residues, the reconstructed hydrophobic lock dramatically enhanced the CbpA-binding activity of mFH CCP9, resulting in a K_d of 973 nM (Figure S2), though the binding activity was still significantly lower than that of hFH CCP9, which has a K_d of 4.72 nM. MM-GBSA analysis of the “humanized” mFH CCP9 suggested that Glu 525, Lys 527, and Tyr 538 might be the culprits (Figure 4d): the contribution of Tyr 538 to binding is less than that of hFH Asn 538, the interactions of Lys 527 are unfavorable whereas those of hFH Gly 527 neutral, and the combined contributions of Glu 523, Glu 525 and Lys 527 are significantly less than those of hFH Glu 523, Asn 525 and Gly 527. To test these predictions, we made two additional mutants with the mFH triple mutant (E497M+T543L+S545I) as a template, a quadruple mutant with the addition of the mutation Y538N and a quintuple mutant with the addition of the double mutations E525N+K527G. The K_d value of the quadruple mutant protein was similar to that of the triple mutant protein (Table 1, 1,080 vs. 973 nM), indicating that the Y538N mutation does not enhance the binding activity of mFH CCP9. In contrast, the addition of the double mutations E525N+K527G lowered the K_d value (Figure S2 and Table 1) near to that of hFH CCP9 (5.78 vs. 4.72 nM), suggesting that the hydrophobic lock and these two additional residues are the key structural features recognized by CbpA.

DISCUSSION

One of major contributing factors of host specificity or tropism of human-specific pathogens is the ability of these pathogens to evade human immunity but not the immunity of other animals. The structural determinants of host specificity of immune evasion are, however, largely unknown. One of the first host defenders that pathogens encounter is the complement system, killing invading microbial pathogens via formation of the membrane attack complex and tagging them for phagocytosis [1]. In order to cause disease, *S. pneumoniae* has to escape phagocytosis [12]. To escape complement tagging for phagocytosis, like many microbial pathogens, *S. pneumoniae* arms itself by recruiting host FH to its cell surface and the FH recruitment is human-specific [7]. In this study, using a combination of biochemical, NMR, crystallographic, computational, and mutagenesis methods, we have demonstrated that a single hFH CCP domain is sufficient for tight binding of the pneumococcal protein virulence factor CbpA and identified the structural determinants for host specificity of FH recruitment by *S. pneumoniae*. It has been shown that CCP6 of hFH is required for binding of a meningococcal protein virulence factor known as fHbp, but whether a single CCP6 domain of hFH is sufficient for tight binding of the meningococcal protein is not known [65]. Intermolecular interactions in the two systems are also significantly different: hydrophobic interactions are more important than polar interactions for CCP9 binding of CbpA, whereas polar interactions such as hydrogen bonds and salt bridges are mainly responsible for CCP6-7 binding of fHbp [65]. The major structural determinant for host specificity of FH recruitment by *S. pneumoniae* is a hydrophobic lock formed by Val 495, Met 497, Leu 543 and Ile 545. Antivirulence is a novel strategy currently pursued for development of therapeutic agents against many bacterial pathogens [66, 67]. The structure-function information reported here will facilitate the development of agents against the major pneumococcal virulence factor CbpA.

Because *S. pneumonia* is a human-specific pathogen, no animal is an ideal model for studying pneumococcal pathogenesis and vaccine development. Mouse is the most frequently used animal model for these purposes, probably due to the ease of manipulation and affordability, but there are fundamental differences in interactions of *S. pneumonia* with human and mouse. One approach to overcome these differences is to “humanize” the mouse by replacement of mouse genes with human counterparts [68], and “humanized” mice have emerged as powerful tools for studying infectious diseases [69]. Our findings regarding the host specificity of FH recruitment offer a novel approach to the development of a more biologically relevant mouse model for studying pneumococcal pathogenesis and vaccine development. Our results indicate that out of more than 1,200 amino acids, only a few point mutations are needed to humanize mFH with respect to binding of CbpA. Such an approach may be preferable to a complete replacement of the mFH gene with the hFH gene, as it is not known whether hFH can regulate the complement system and serve other functions in mouse like mFH.

Limited serotype coverage is a major drawback of current capsular polysaccharide-based pneumococcal vaccines, as *S. pneumoniae* has over 90 different serotypes [70]. While the current vaccines are enormously beneficial and valuable, vaccination with these vaccines has resulted in the rise of pneumococcal strains that are not covered by the vaccines [71]. Consequently, there is an urgent need for development of broadly neutralizing, protein-based vaccines [70]. Because of its high immunogenicity, CbpAN is considered a promising candidate antigen for vaccine development [72]. Its high sequence variations, however, limit cross-strain protection. In addition, because of high serum levels of hFH [3], tight FH binding may also affect the efficacy of CbpAN as a vaccine antigen, as suggested for the meningococcal FH binding protein [65, 73]. The structural information reported here makes it possible to redesign CbpAN to enhance its efficacy as a vaccine antigen by removing its strain-specific structural features and weakening its binding of hFH.

We have further demonstrated that CCP19-20 are essentially not involved in binding of CbpA, which is consistent with studies showing that binding of glycosaminoglycans to CCP19-20 does not affect binding of FH to pneumococci [37, 74]. The function of CCP19-20 is to recognize glycosaminoglycans on the surface of host cells and prevent the complement system from attacking host tissues [3]. The ability of FH to bind CbpA on the pneumococcal cell wall and glycosaminoglycans on the surface of human epithelial cells at the same time allows pneumococci connect to human cells via FH. Thus, FH recruitment also promotes pneumococcal adherence to and uptake of pneumococci by human cells [74].

Surprisingly, Tyr 90 is conserved only in about 50% of *S. pneumoniae* strains. The structure-function relationship of CbpA in binding of FH has to be investigated in the context of the variations of the amino acid sequences of CbpAs from other *S. pneumoniae* strains. A mouse model with an engineered FH with the mutations suggested in this work may be good for studying the pathogenesis of the serotype-4 virulent *S. pneumoniae* strain TIGR4 and similar strains, but not all *S. pneumoniae* strains.

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

Acknowledgments

We thank Y. Wu and L. D. Bell for assistance in protein purification, R. J. Staples for collecting initial X-ray data, J. Liu for assistance in synchrotron X-ray data collection, S. M. Ferguson-Miller for the use of a Cryphon robot for crystallization screening, C. G. Hoogstraten for the use of a VP-ITC isothermal titration calorimeter for thermodynamic analysis, and Z. F. Burton and J. Wang for critical reading of the manuscript. We are indebted to LS-CAT at the Advanced Photon Source for access to their facilities and expertise for data collection and processing.

FUNDING

This work was supported in part by NIH grants R01DC006917 and R21DC011066, the Ministry of Science and Technology Grant 2012CB518700, and the Tsinghua University Initiative Scientific Research grant 2011Z23153. The authors have no conflict of interest.

ABBREVIATIONS

CbpA	choline binding protein A
CbpAN	the N-terminal domain of choline binding protein A
CCP	complement control protein
FH	complement factor H
hFH	human complement factor H
ITC	isothermal titration calorimetry
MBP	maltose binding protein
MD	molecular dynamics
mFH	mouse complement factor H
MM-GBSA analysis	combined molecular mechanics/generalized Born surface area analysis
pIgR	polymeric immunoglobulin receptor

References

1. Ricklin D, Hajishengallis G, Yang K, Lambris JD. Complement: a key system for immune surveillance and homeostasis. *Nat Immunol.* 2010; 11:785–797. [PubMed: 20720586]
2. Serruto D, Rappuoli R, Scarselli M, Gros P, van Strijp JAG. Molecular mechanisms of complement evasion: learning from staphylococci and meningococci. *Nature Rev Microbiol.* 2010; 8:393–399. [PubMed: 20467445]
3. Ferreira VP, Pangburn MK, Cortes C. Complement control protein factor H: The good, the bad, and the inadequate. *Mol Immunol.* 2010; 47:2187–2197. [PubMed: 20580090]
4. Zipfel PF, Skerka C. Complement regulators and inhibitory proteins. *Nat Rev Immunol.* 2009; 9:729–740. [PubMed: 19730437]

5. Granoff DM, Welsch JA, Ram S. Binding of Complement Factor H (fH) to *Neisseria meningitidis* Is Specific for Human fH and Inhibits Complement Activation by Rat and Rabbit Sera. *Infect Immun.* 2009; 77:764–769. [PubMed: 19047406]
6. Ngampasutadol J, Ram S, Gulati S, Agarwal S, Li C, Visintin A, Monks B, Madico G, Rice PA. Human factor H interacts selectively with *Neisseria gonorrhoeae* and results in species-specific complement evasion. *J Immunol.* 2008; 180:3426–3435. [PubMed: 18292569]
7. Lu L, Ma Z, Jokiranta TS, Whitney AR, DeLeo FR, Zhang JR. Species-Specific Interaction of *Streptococcus pneumoniae* with Human Complement Factor H. *J Immunol.* 2008; 181:7138–7146. [PubMed: 18981135]
8. Kurtenbach K, De Michelis S, Etti S, Schafer SM, Sewell HS, Brade V, Kraiczy P. Host association of *Borrelia burgdorferi* sensu lato - the key role of host complement. *Trends Microbiol.* 2002; 10:74–79. [PubMed: 11827808]
9. Stevenson B. Differential binding of host complement inhibitor factor H by *Borrelia burgdorferi* Erp surface proteins: a possible mechanism underlying the expansive host range of lyme disease spirochetes. *Infect Immun.* 2002; 70:491–497. [PubMed: 11796574]
10. O'Brien KL, Wolfson LJ, Watt JP, Henkle E, Deloria-Knoll M, McCall N, Lee E, Mulholland K, Levine OS, Cherian T. Burden of disease caused by *Streptococcus pneumoniae* in children younger than 5 years: global estimates. *Lancet.* 2009; 374:893–902. [PubMed: 19748398]
11. Huang SS, Johnson KM, Ray GT, Wroe P, Lieu TA, Moore MR, Zell ER, Linder JA, Grijalva CG, Metlay JP, Finkelstein JA. Healthcare utilization and cost of pneumococcal disease in the United States. *Vaccine.* 2011; 29:3398–3412. [PubMed: 21397721]
12. Musher, DM. *Streptococcus pneumoniae*. In: Mandell, GL, Bennett, JE., Dolin, RD., editors. *Principles and Practice of Infectious Diseases*. Elsevier Churchill Livingstone; Philadelphia: 2009. p. 2623–2642.
13. Gross GN, Rehm SR, Pierce AK. EFFECT OF COMPLEMENT DEPLETION ON LUNG CLEARANCE OF BACTERIA. *J Clin Invest.* 1978; 62:373–378. [PubMed: 27534]
14. Hosea SW, Brown EJ, Frank MM. The critical role of complement in experimental pneumococcal sepsis. *J Infect Dis.* 1980; 142:903–909. [PubMed: 7462698]
15. Kerr AR, Paterson GK, Riboldi-Tunncliffe A, Mitchell TJ. Innate immune defense against pneumococcal pneumonia requires pulmonary complement component C3. *Infect Immun.* 2005; 73:4245–4252. [PubMed: 15972516]
16. Sampson HA, Walchner AM, Baker PJ. Recurrent pyogenic infections in individuals with absence of the 2nd component of complement. *J Clin Immunol.* 1982; 2:39–45. [PubMed: 6980226]
17. Alper CA, Abramson N, Johnston RB, Jandl JH, Rosen FS. Increased susceptibility to infection associated with abnormalities of complement-mediated functions and of third component of complement (C3). *N Engl J Med.* 1970; 282:349. [PubMed: 4188976]
18. Carneiro-Sampaio M, Coutinho A. Immunity to microbes: Lessons from primary immunodeficiencies. *Infect Immun.* 2007; 75:1545–1555. [PubMed: 17283094]
19. Hammerschmidt S. Adherence molecules of pathogenic pneumococci. *Curr Opin Microbiol.* 2006; 9:12–20. [PubMed: 16338163]
20. Brock SC, McGraw PA, Wright PF, Crowe JE. The human polymeric immunoglobulin receptor facilitates invasion of epithelial cells by *Streptococcus pneumoniae* in a strain-specific and cell type-specific manner. *Infect Immun.* 2002; 70:5091–5095. [PubMed: 12183558]
21. Hammerschmidt S, Tillig MP, Wolff S, Vaerman JP, Chhatwal GS. Species-specific binding of human secretory component to SpsA protein of *Streptococcus pneumoniae* via a hexapeptide motif. *Mol Microbiol.* 2000; 36:726–736. [PubMed: 10844660]
22. Zhang JR, Mostov KE, Lamm ME, Nanno M, Shimida S, Ohwaki M, Tuomanen E. The polymeric immunoglobulin receptor translocates pneumococci across human nasopharyngeal epithelial cells. *Cell.* 2000; 102:827–837. [PubMed: 11030626]
23. Luo RS, Mann B, Lewis WS, Rowe A, Heath R, Stewart ML, Hamburger AE, Sivakolundu S, Lacy ER, Bjorkman PJ, Tuomanen E, Kriwacki RW. Solution structure of choline binding protein A, the major adhesin of *Streptococcus pneumoniae*. *EMBO J.* 2005; 24:34–43. [PubMed: 15616594]

24. Lu L, Lamm ME, Li HM, Corthesy B, Zhang JR. The human polymeric immunoglobulin receptor binds to *Streptococcus pneumoniae* via domains 3 and 4. *J Biol Chem*. 2003; 278:48178–48187. [PubMed: 13679368]
25. Lu L, Ma YY, Zhang JR. *Streptococcus pneumoniae* recruits complement factor H through the amino terminus of CbpA. *J Biol Chem*. 2006; 281:15464–15474. [PubMed: 16597618]
26. Hammerschmidt S, Agarwal V, Kunert A, Haelbich S, Skerka C, Zipfel PF. The host immune regulator factor H interacts via two contact sites with the PspC protein of *Streptococcus pneumoniae* and mediates adhesion to host epithelial cells. *J Immunol*. 2007; 178:5848–5858. [PubMed: 17442969]
27. Rosenow C, Ryan P, Weiser JN, Johnson S, Fontan P, Ortqvist A, Masure HR. Contribution of novel choline-binding proteins to adherence, colonization and immunogenicity of *Streptococcus pneumoniae*. *Mol Microbiol*. 1997; 25:819–829. [PubMed: 9364908]
28. Balachandran P, Brooks-Walter A, Virolainen-Julkunen A, Hollingshead SK, Briles DE. Role of pneumococcal surface protein C in nasopharyngeal carriage and pneumonia and its ability to elicit protection against carriage of *Streptococcus pneumoniae*. *Infect Immun*. 2002; 70:2526–2534. [PubMed: 11953392]
29. Iannelli F, Chiavolini D, Ricci S, Oggioni MR, Pozzi G. Pneumococcal surface protein C contributes to sepsis caused by *Streptococcus pneumoniae* in mice. *Infect Immun*. 2004; 72:3077–3080. [PubMed: 15102826]
30. McCool TL, Cate TR, Moy G, Weiser JN. The immune response to pneumococcal proteins during experimental human carriage. *J Exp Med*. 2002; 195:359–365. [PubMed: 11828011]
31. McCool TL, Cate TR, Tuomanen EI, Adrian P, Mitchell TJ, Weiser JN. Serum immunoglobulin G response to candidate vaccine antigens during experimental human pneumococcal colonization. *Infect Immun*. 2003; 71:5724–5732. [PubMed: 14500493]
32. Brooks-Walter A, Briles DE, Hollingshead SK. The *pspC* gene of *Streptococcus pneumoniae* encodes a polymorphic protein, PspC, which elicits cross-reactive antibodies to PspA and provides immunity to pneumococcal bacteremia. *Infect Immun*. 1999; 67:6533–6542. [PubMed: 10569772]
33. Ogunniyi AD, Woodrow MC, Poolman JT, Paton JC. Protection against *Streptococcus pneumoniae* elicited by immunization with pneumolysin and CbpA. *Infect Immun*. 2001; 69:5997–6003. [PubMed: 11553536]
34. Ogunniyi AD, Grabowicz M, Briles DE, Cook J, Paton JC. Development of a vaccine against invasive pneumococcal disease based on combinations of virulence proteins of *Streptococcus pneumoniae*. *Infect Immun*. 2007; 75:350–357. [PubMed: 17088353]
35. Cao J, Chen DP, Xu WC, Chen TM, Xu SX, Luo JY, Zhao Q, Liu BZ, Wang DS, Zhang XM, Shan YL, Yin YB. Enhanced protection against pneumococcal infection elicited by immunization with the combination of PspA, PspC, and ClpP. *Vaccine*. 2007; 25:4996–5005. [PubMed: 17524530]
36. Dave S, Carmicle S, Hammerschmidt S, Pangburn MK, McDaniel LS. Dual roles of PspC, a surface protein of *Streptococcus pneumoniae*, in binding human secretory IgA and factor H. *J Immunol*. 2004; 173:471–477. [PubMed: 15210807]
37. Duthy TG, Ormsby RJ, Giannakis E, Ogunniyi AD, Stroehrer UH, Paton JC, Gordon DL. The human complement regulator factor H binds pneumococcal surface protein PspC via short consensus repeats 13 to 15. *Infect Immun*. 2002; 70:5604–5611. [PubMed: 12228288]
38. Geourjon C, Deleage G. SOPMA: Significant improvements in protein secondary structure prediction by consensus prediction from multiple alignments. *Comput Appl Biosci*. 1995; 11:681–684. [PubMed: 8808585]
39. Sun P, Tropea JE, Waugh DS. Enhancing the solubility of recombinant proteins in *Escherichia coli* by using hexahistidine-tagged maltose-binding protein as a fusion partner. *Methods Mol Biol*. 2011; 705:259–274. [PubMed: 21125392]
40. Wang CC. Protein disulfide isomerase as an enzyme and a chaperone in protein folding. *Methods Enzymol*. 2002; 348:66–75. [PubMed: 11885295]
41. Velazquez-Campoy A, Freire E. Isothermal titration calorimetry to determine association constants for high-affinity ligands. *Nat Protocols*. 2006; 1:186–191. [PubMed: 17406231]
42. Gill SC, Vonhippel PH. Calculation of protein extinction coefficients from amino-acid sequence data. *Anal Biochem*. 1989; 182:319–326. [PubMed: 2610349]

43. Gasteiger, E., Hoogland, C., Gattiker, A., Duvaud, S., Wilkins, MR., Appel, RD., Bairoch, A. Protein identification and analysis tools on the ExPASy server. In: Walker, JM., editor. The Proteomics Protocols Handbook. Humana Press; Totowa, NJ: 2005. p. 571-607.
44. Delaglio F, Grzesiek S, Vuister GW, Zhu G, Pfeifer J, Bax A. NMRPipe: a multidimensional spectral processing system based on UNIX pipes. J Biomol NMR. 1995; 6:277-293. [PubMed: 8520220]
45. Johnson BA, Blevins RA. NMRView: A computer program for the visualization and analysis of NMR data. J Biomol NMR. 1994; 4:603-614. [PubMed: 22911360]
46. Cavanagh, J., Fairbrother, WJ., Palmer, AGI., Skelton, NJ. Protein NMR Spectroscopy. Academic Press; New York: 2006.
47. Güntert P. Automated NMR structure calculation with CYANA. Methods Mol Biol. 2004; 278:353-378. [PubMed: 15318003]
48. Shen Y, Delaglio F, Cornilescu G, Bax A. TALOS+ : a hybrid method for predicting protein backbone torsion angles from NMR chemical shifts. J Biomol NMR. 2009; 44:213-223. [PubMed: 19548092]
49. Case, DA., Darden, TA., Cheatham, I., Simmerling, CL., Wang, J., Duke, RE., Luo, R., Crowley, M., Walker, RC., Zhang, W., Merz, KM., Wang, B., Hayik, S., Roitberg, A., Seabra, G., Kolossváry, I., Wong, KF., Paesani, F., Vanicek, J., Wu, X., Brozell, SR., Steinbrecher, T., Gohlke, H., Yang, L., Tan, C., Mongan, J., Hornak, V., Cui, G., Mathews, DH., Seetin, MG., Sagui, C., Babin, V., Kollman, PA. AMBER 10. University of California; San Francisco: 2008.
50. McCoy AJ, Grosse-Kunstleve RW, Adams PD, Winn MD, Storoni LC, Read RJ. Phaser crystallographic software. J Appl Crystallogr. 2007; 40:658-674. [PubMed: 19461840]
51. Cohen SX, Ben Jelloul M, Long F, Vagin A, Knipscheer P, Lebbink J, Sixma TK, Lamzin VS, Murshudov GN, Perrakis A. ARP/wARP and molecular replacement: the next generation. Acta Crystallogr D Biol Crystallogr. 2008; 64:49-60. [PubMed: 18094467]
52. Murshudov GN, Skubak P, Lebedev AA, Pannu NS, Steiner RA, Nicholls RA, Winn MD, Long F, Vagin AA. REFMAC5 for the refinement of macromolecular crystal structures. Acta Crystallogr D Biol Crystallogr. 2011; 67:355-367. [PubMed: 21460454]
53. Adams PD, Afonine PV, Bunkoczi G, Chen VB, Davis IW, Echols N, Headd JJ, Hung LW, Kapral GJ, Grosse-Kunstleve RW, McCoy AJ, Moriarty NW, Oeffner R, Read RJ, Richardson DC, Richardson JS, Terwilliger TC, Zwart PH. PHENIX: a comprehensive Python-based system for macromolecular structure solution. Acta Crystallogr D Biol Crystallogr. 2010; 66:213-221. [PubMed: 20124702]
54. Emsley P, Cowtan K. Coot: model-building tools for molecular graphics. Acta Crystallogr D Biol Crystallogr. 2004; 60:2126-2132. [PubMed: 15572765]
55. DeLano, WL. The PyMOL Molecular Graphics System. DeLano Scientific; Palo Alto, CA, USA: 2002.
56. Case, DA., Darden, TA., Cheatham, I., Simmerling, CL., Wang, J., Duke, RE., Luo, R., Walker, RC., Zhang, W., Merz, KM., Roberts, B., Hayik, S., Roitberg, A., Seabra, G., Swails, J., Goetz, I., Kolossváry, I., Wong, KF., Paesani, F., Vanicek, J., Wolf, J., Wu, X., Brozell, SR., Steinbrecher, T., Gohlke, H., Cai, Q., Ye, X., Wang, J., Hsieh, M-J., Cui, G., Roe, DR., Mathews, DH., Seetin, MG., Salomon-Ferrer, R., Sagui, C., Babin, V., Luchko, T., Gusarov, S., Kovalenko, A., Kollman, PA. AMBER 12. University of California; San Francisco: 2012.
57. Arnold K, Bordoli L, Kopp J, Schwede T. The SWISS-MODEL workspace: a web-based environment for protein structure homology modelling. Bioinformatics. 2006; 22:195-201. [PubMed: 16301204]
58. Bordoli L, Kiefer F, Arnold K, Benkert P, Battey J, Schwede T. Protein structure homology modeling using SWISS-MODEL workspace. Nat Protoc. 2009; 4:1-13. [PubMed: 19131951]
59. Schwede T, Kopp J, Guex N, Peitsch MC. SWISS-MODEL: an automated protein homology-modeling server. Nucleic Acids Res. 2003; 31:3381-3385. [PubMed: 12824332]
60. Ryckaert JP, Ciccotti G, Berendsen HJC. Numerical integration of cartesian equations of motion of a system with constraints - molecular dynamics of N-alkanes. J Comput Phys. 1977; 23:327-341.
61. Essmann U, Perera L, Berkowitz ML, Darden T, Lee H, Pedersen LG. A smooth particle mesh Ewald method. J Chem Phys. 1995; 103:8577-8593.

62. Miller BR, McGee TD, Swails JM, Homeyer N, Gohlke H, Roitberg AE. MMPBSA.py: An Efficient Program for End-State Free Energy Calculations. *J Chem Theory Comput.* 2012; 8:3314–3321. [PubMed: 26605738]
63. Chen VB, Arendall WB, Headd JJ, Keedy DA, Immormino RM, Kapral GJ, Murray LW, Richardson JS, Richardson DC. MolProbity: all-atom structure validation for macromolecular crystallography. *Acta Crystallogr D Biol Crystallogr.* 2010; 66:12–21. [PubMed: 20057044]
64. Chiavolini D, Pozzi G, Ricci S. Animal Models of *Streptococcus pneumoniae* Disease. *Clin Microbiol Rev.* 2008; 21:666–685. [PubMed: 18854486]
65. Schneider MC, Prosser BE, Caesar JJ, Kugelberg E, Li S, Zhang Q, Quoraishi S, Lovett JE, Deane JE, Sim RB, Roversi P, Johnson S, Tang CM, Lea SM. *Neisseria meningitidis* recruits factor H using protein mimicry of host carbohydrates. *Nature.* 2009; 458:890–893. [PubMed: 19225461]
66. Clatworthy AE, Pierson E, Hung DT. Targeting virulence: a new paradigm for antimicrobial therapy. *Nat Chem Biol.* 2007; 3:541–548. [PubMed: 17710100]
67. Cegelski L, Marshall GR, Eldridge GR, Hultgren SJ. The biology and future prospects of antivirulence therapies. *Nature Rev Microbiol.* 2008; 6:17–27. [PubMed: 18079741]
68. Sacca R, Engle SJ, Qin WN, Stock JL, McNeish JD. Genetically engineered mouse models in drug discovery research. *Methods Mol Biol.* 2010; 602:37–54. [PubMed: 20012391]
69. Legrand N, Ploss A, Balling R, Becker PD, Borsotti C, Brezillon N, Debarry J, de Jong Y, Deng HK, Di Santo JP, Eisenbarth S, Eynon E, Flavell RA, Guzman CA, Huntington ND, Kremsdorf D, Manns MP, Manz MG, Mention JJ, Ott M, Rathinam C, Rice CM, Rongvaux A, Stevens S, Spits H, Strick-Marchand H, Takizawa H, van Lent AU, Wang CY, Weijer K, Willinger T, Ziegler P. Humanized mice for modeling human infectious disease: challenges, progress, and outlook. *Cell Host & Microbe.* 2009; 6:5–9. [PubMed: 19616761]
70. Gamez G, Hammerschmidt S. Combat Pneumococcal Infections: Adhesins as Candidates for Protein-Based Vaccine Development. *Curr Drug Targets.* 2012; 13:323–337. [PubMed: 22206255]
71. Weinberger DM, Malley R, Lipsitch M. Serotype replacement in disease after pneumococcal vaccination. *Lancet.* 2011; 378:1962–1973. [PubMed: 21492929]
72. Ricci S, Janulczyk R, Gerlini A, Braione V, Colomba L, Iannelli F, Chiavolini D, Oggioni MR, Bjorck L, Pozzi G. The factor H-binding fragment of PspC as a vaccine antigen for the induction of protective humoral immunity against experimental pneumococcal sepsis. *Vaccine.* 2011; 29:8241–8249. [PubMed: 21911026]
73. Johnson S, Tan L, van der Veen S, Caesar J, De Jorge EG, Harding RJ, Bai XL, Exley RM, Ward PN, Ruivo N, Trivedi K, Cumber E, Jones R, Newham L, Staunton D, Ufret-Vincenty R, Borrow R, Pickering MC, Lea SM, Tang CM. Design and evaluation of meningococcal vaccines through structure-based modification of host and pathogen molecules. *Plos Pathogens.* 2012; 8:e1002981. [PubMed: 23133374]
74. Agarwal V, Asmat TM, Luo S, Jensch I, Zipfel PF, Hammerschmidt S. Complement regulator Factor H mediates a two-step uptake of *Streptococcus pneumoniae* by human cells. *J Biol Chem.* 2010; 285:23486–23495. [PubMed: 20504767]

SUMMARY STATEMENT

The complex of human complement factor H and pneumococcal protein CbpA enables *Streptococcus pneumoniae* to evade phagocytosis and contributes to pneumococcal host specificity. This work elucidated the atomic structure of the complex and the structural determinants of its host specificity.

Figure 1. ITC and NMR analysis of the binding of hFH and CbpA

a, ITC measurement of the binding of hFH CCP9. For the ITC experiment, the syringe was loaded with a 75 μM solution of CbpAN and the sample cell with 5 μM hFH CCP9. The injections were made over a period of 120 min with a 6 min interval between subsequent injections. The sample cell was stirred at 310 rpm. **b**, Overlaid NMR spectra of CbpAN (blue) and CbpAN + CCP9 (red). The ^1H - ^{15}N HSQC NMR spectra were obtained with 1.0 mM ^{15}N -CbpAN in the absence and the presence of 1.1 mM hFH CCP9. Complete sequential resonance assignments are presented in Supplementary Figure 1. **c**, Perturbation of the chemical shifts of the backbone amides of CbpAN by the binding of hFH CCP9. The loop region with missing assignment for the free protein is indicated by a gray bar. Chemical shift changes were calculated using the equation: $\delta = \{0.5[\delta(^1\text{H}^{\text{N}})^2 + (0.2\delta(^{15}\text{N}))^2]\}^{1/2}$. **d**, Overlaid NMR spectra of CbpAN (blue) and CbpAN + CCP13-15 (red). The ^1H - ^{15}N

HSQC NMR spectra were obtained with 0.2 mM ^{15}N -CbpAN in the absence and the presence of 0.5 mM CCP13-15. **e**, Overlaid NMR spectra of CbpAN (blue) and CbpAN + CCP19-20 (red). The ^1H - ^{15}N HSQC NMR spectra were obtained with 0.2 mM ^{15}N -CbpAN in the absence and the presence of 1.0 mM CCP19-20.

Author Manuscript

Author Manuscript

Author Manuscript

Author Manuscript

Figure 2. Structural analysis of CbpAN and its complex with hFH CCP9

a, NMR structure of free CbpAN. The C^α traces of the top 20 conformers are drawn. CbpAN has a compact structure consisting of a bundle of three helices with an up-down-up antiparallel topology. **b**, Crystal structure of the complex of CbpAN (green) and hFH CCP9 (orange) superposed with the NMR structure of free CbpAN (cyan). The side-chains that form the two disulfide bonds (S-S) are drawn in purple lines. **c**, Superposed crystal structures of hFH CCP9 (blue), CCP4 (red), and CCP8 (green). C^α RMSDs between CCP9 and CCP4 (PDB ID: 2WII) and between CCP9 and CCP8 (PDB ID: 2V8E) are 1.139 Å for 54 residues and 1.344 Å for 59 residues, respectively. **d**, Structure-based amino acid sequence alignment of hFH CCP9, CCP4, and CCP8. The conserved residues between the three CCPs are shaded in black and CCP9 residues at the interface of binding CbpAN are colored in red. The numbering is that of CCP9. **e**, Hydrogen bonding and hydrophobic locking between hFH CCP9 and CbpAN. hFH residues are lettered in orange, CbpA residues are lettered in cyan, and hydrogen bonds are indicated by dash lines. The

hydrophobic lock (a cluster of hydrophobic residues, Val 495, Met 497, Leu 543 and Ile 545) of hFH CCP9 is drawn in a surface representation.

Author Manuscript

Author Manuscript

Author Manuscript

Author Manuscript

Figure 3. Amino acid sequence alignment of CCP9s from 18 mammals

The residues conserved among the 18 CCP9s are shaded in black. The interface residues of hFH CCP9 in the binding of CbpAN are indicated with ◆ and the residues that constitute the "hydrophobic lock" with ■. The residues of mFH CCP9 substituted with the corresponding residues of hFH CCP9 in the mutagenesis study are indicated with #. The abbreviations for various mammals are: NOMLE, *Nomascus leucogenys* (Northern white-cheeked gibbon); PONAB, *Pongo abelii* (Sumatran orangutan); MACFA, *Macaca fascicularis* (crab-eating macaque); MACMU, *Macaca mulatta* (Rhesus macaque); CALJA, *Callithrix jacchus* (white-tufted-ear marmoset); LOXAF, *Loxodonta africana* (African elephant); HETGA, *Heterocephalus glaber* (naked mole rat); OTOGA, *Otolemur garnettii* (small-eared galago); AILME, *Ailuropoda melanoleuca* (giant panda).

Figure 4. Computational analysis of the binding of CCP9 and CbpAN

Binding energy decompositions were obtained by MM-GBSA analysis for the wild-type hFH CCP9 (a), the hFH CCP9 triple mutant protein with Met 497 replaced by Glu, Leu 543 by Thr, and Ile 545 by Ser (b), the wild-type mFH CCP9 (c), and the mFH CCP9 triple mutant protein with Glu 497 replaced by Met, Thr 543 by Leu, and Ser 545 by Ile (d).

Table 1

Thermodynamic binding parameters of human and mouse FH domains *

Protein	K_d (nM)	G (kcal/mol)	H (kcal/mol)	T S (kcal/mol)
hFH CCP				
CCP6-8	ND			
CCP8-10	1.49 ± 0.5	-12.02 ± 0.17	-18.45 ± 0.46	-6.44 ± 0.43
CCP9-10	3.4 ± 1	-11.54 ± 0.15	-19.63 ± 0.67	-7.92 ± 0.66
CCP9	4.72 ± 0.4	-11.35 ± 0.05	-21.12 ± 0.67	-9.86 ± 0.62
CCP10	ND			
CCP10-12	ND			
CCP13-15	ND			
CCP19-20	ND			
hFH CCP9 Mutant Protein				
M497E	206 ± 20	-9.11 ± 0.05	-15.53 ± 0.32	-6.4 ± 0.29
E523A	47 ± 4	-9.99 ± 0.04	-15.5 ± 0.6	-5.5 ± 0.66
D542A	1640 ± 40	-7.88 ± 0.01	-18.35 ± 2.6	-10.5 ± 2.6
L543T	3580 ± 1800	-7.42 ± 0.24	-36.0 ± 0	-28.4 ± 0.48
I545S	12300 ± 590	-6.69 ± 0.03	-12.8 ± 0.8	-5.75 ± 0.44
Y547A	38000 ± 1200	-6.02 ± 0.03	-2.04 ± 0.2	-3.68 ± 0.05
M497E+L543T	17400 ± 850	-6.49 ± 0.02	-2.44 ± 0.3	-4.06 ± 0.29
M497E+L543T+I545S	ND			
mFH CCP9 and Mutant Protein				
CCP9	ND			
E497M+T543L+S545I	973 ± 21	-8.19 ± 0.01	-12.17 ± 0.13	-4.02 ± 0.18
E497M+T543L+S545I+Y538N	1080 ± 130	-8.13 ± 0.07	-13.2 ± 0	-5.69 ± 0.24
E497M+T543L+S545I+E525N+K527G	5.78 ± 0.6	-11.23 ± 0.06	-14.5 ± 0	-3.41 ± 0.17

* ND, not detectable.