Newcastle University ePrints

Blaum BS, Hannan JP, Herbert AP, Kavanagh D, Uhrin D, Stehle T. [Structural basis for sialic acid-mediated self-recognition by complement factor H.](#)
Nature Chemical Biology 2015, 11(1), 77-82.

Copyright:

This is the author's accepted manuscript of an article published in its final form by Nature Publishing Group, 2015.

For the terms for reuse for this manuscript please see:

<http://www.nature.com/authors/policies/license.html>

Link to published article: <http://dx.doi.org/10.1038/NCHEMBIO.1696>

Date deposited: 3rd February 2015 (made available 24th May 2015)

ePrints – Newcastle University ePrints

<http://eprint.ncl.ac.uk>

Structural basis for sialic acid-mediated self-recognition by complement factor H

Bärbel S. Blaum^{1*}, Jonathan P. Hannan², Andrew P. Herbert³, David Kavanagh⁴, Dušan Uhrín³, Thilo

Stehle^{1,5}

¹Interfaculty Institute of Biochemistry, University of Tübingen, Tübingen, Germany

²Department of Medicine, University of Colorado School of Medicine, Aurora, Colorado, USA

³The School of Chemistry, University of Edinburgh, West Mains Road, Edinburgh EH9 3JJ, U.K.

⁴Institute of Genetic Medicine, International Centre for Life, Central Parkway, Newcastle upon Tyne, NE1 3BZ, U.K.

⁵Department of Pediatrics, Vanderbilt University School of Medicine, Nashville, Tennessee, USA

*to whom correspondence should be addressed: baerbel.blaum@uni-tuebingen.de

Abstract

1 The serum protein complement factor H (FH) ensures down-regulation of the complement alternative
2 pathway, a branch of innate immunity, upon interaction with specific glycans on host cell-surfaces.
3 Using ligand-based NMR, we screened a comprehensive set of sialylated glycans for binding to FH and
4 solved the crystal structure of a ternary complex formed by the two C-terminal domains of FH, a
5 sialylated trisaccharide, and the complement C3b thioester-containing domain (TED). Key residues in the
6 sialic acid binding site are conserved from mice to men and residues linked to atypical hemolytic uremic
7 syndrome (aHUS) cluster within this binding site, suggesting a possible role for sialic acid as host marker
8 also in other mammals and a critical role in human renal complement homeostasis. Surprisingly, sialic
9 acid binding to FH is structurally homologous to two evolutionary unrelated proteins. The crystal
10 structure also advances our understanding of bacterial immune evasion strategies.

Introduction

1 The ability to distinguish between host cells, pathogens, and altered host cells is the most fundamental
2 property of the immune system. Adaptive immunity equips mammals for the encounter of an ever
3 increasing number of rapidly evolving pathogens, while less specific but immediate defense carried out
4 by the innate immune system. Adaptive and innate immunity are interwoven in the complement
5 system¹. Complement deficiencies can cause chronic disease or render vulnerable against infection with
6 Gram-negative bacteria and certain viruses²⁻⁴. Inadvertent action of complement on host cells is
7 prevented by dedicated regulatory proteins, some of which are expressed at the cell surface⁵. Pathogens
8 frequently mimic or recruit complement regulators to escape host immunity⁶.

9 The fluid-phase regulator factor H (FH) distinguishes cells that deserve protection from those that
10 require elimination. FH is abundant in serum and is composed of twenty small, ellipsoidal domains of
11 the complement control protein (CCP)/short consensus repeat (SCR) fold (see ref. 7 and references
12 therein). FH accelerates the decay of the complement alternative pathway (AP) C3 convertase C3bBb,
13 thus preventing local formation of more C3b, the central player of the complement amplification loop
14 and a powerful opsonin. C3b is deposited on cells in a nonselective, irreversible manner via covalent
15 attachment of its thioester-containing domain (TED) to cell-surface nucleophiles⁸. As a cofactor to the
16 serine protease factor I, FH also regulates proteolytic degradation of already deposited C3b. FH-
17 mediated complement regulation is critical on host cells that do not possess additional, membrane-
18 bound regulators⁹ as well as on cells that do¹⁰. In order to confine C3b degradation to host cells, FH is
19 thought to recognize specific host markers. Heterologous cells feature markers that may be structurally
20 similar, conferring protection from complement to sheep erythrocytes and gram-negative bacteria¹¹⁻¹³.
21 Two classes of glycans are thought to act as genuine self-markers or molecular mimics of such:
22 glycosaminoglycans (GAGs) and sialic acid glycans¹⁴. While binding of FH to GAGs is relatively well

understood¹⁵⁻¹⁷, the molecular details governing the interaction of FH with sialylated glycans have not been established.

Sialic acids are nine-carbon pyranoses that cap the non-reducing end of glycan chains on human glycoproteins and gangliosides and mediate highly specific and diverse immunological recognition processes^{18,19}. They feature various naturally occurring chemical modifications; prevalent in humans is α -N-acetylneuraminic acid (Neu5Ac). Enzymatic removal of Neu5Ac turns non-activators of complement into activators^{11,12} as a consequence of reduced FH binding²⁰. Sialylation contributes to complement evasion of pathogens such as *Neisseria* via recruitment of FH¹³. Since poly-sialic acid (formed by Neu5Ac α 2-8Neu5Ac α repeats) does not regulate the FH affinity for C3b, it is likely that a larger specific epitope exists, encompassing an essential Neu5Ac moiety plus additional components^{20,21}. It is not known whether FH binds simultaneously to cell-bound C3b and host markers, or whether sequential and perhaps even mutually exclusive binding to C3b and host glycans occurs^{22,23}.

We present a structural characterization of the low-affinity FH-Neu5Ac interaction at the molecular level. Our crystal structure of the FH C-terminal recognition domains in complex with a sialylated trisaccharide and the C3b TED suggest that C3b and sialylated glycans act in concert to target FH to host cell surfaces. Combination of our structural data with previous functional studies indicates that mutations in the *CFH* gene which predispose to atypical hemolytic uremic syndrome (aHUS) and are functionally impaired suffer from faulty sialic acid recognition. From the evolutionary point of view, the key amino acids in the FH sialic binding site are conserved across mammals. Unexpected structural homology with other sialic acid binding proteins that are not members of the CCP family hints to convergent evolution of a widespread sialic acid binding motif.

Results

Identification of sialic acid host glycans that bind FH

The most common glycan caps in humans are Neu5Ac α 2-3Gal, Neu5Ac α 2-6Gal(NAc), and Neu5Ac α 2-

8Neu5Ac²⁴. We first tested representatives of these glycan types (Fig. 1) for binding to full-length FH in solution using saturation transfer difference (STD) NMR spectroscopy (Fig. 2). This technique is based on selective irradiation of protein protons and subsequent detection of magnetization transfer to small ligands with fast off-rates (above 100 Hz), allowing to detect weak affinity binding with K_d values of 1 μ M and higher^{25,26}. All glycans for which magnetization transfer from FH was observed contained the 3'sialyl lactose (3'SL, Neu5Ac α 2-3Gal β 1-4Glc) trisaccharide (Fig. 1a-d). Binding was not prevented by several modifications of the lactose (Gal β 1-4Glc) stem. These included an alternative glycosidic linkage (Gal β 1-3Glc, in disialyl lacto-*N*-tetraose, DSLNT, and in the GM1b and GM1 glycans), branching of the Gal ring (in the GM1 glycan), acetylation of the reducing end Glc (in the GM1b glycan and DSLNT), and branching of the Glc ring (in DSLNT). No glycans terminating in Neu5Ac α 2-6 or Neu5Ac α 2-8 bound FH (Fig. 1e-f, Fig. 2). Resonances that can be unambiguously assigned in the 3'SL STD NMR spectrum are highlighted in Fig. 2. The strongest signal was observed for the Neu5Ac methyl group, suggesting that this group is a key component of the binding epitope. In contrast, both H3 protons received negligible saturation transfer. Other Neu5Ac proton resonances are located in the crowded spectral region between 3.5 ppm and 4.0 ppm where partial resonance overlap complicates the assignment, but peaks belonging to H5 and H7 are distinct and were observed in the STD difference spectrum. Magnetization transfer was also observed for the Gal H1, H3, and H4 protons, all of which lie on one side of the Gal pyranose ring. Resonances for Gal H5 and Neu5Ac H4 overlap - one of these or both are also part of the binding epitope. For the Glc ring, only the β -anomer H1 and H2 can be clearly assigned in the difference spectrum. The NMR experiments establish Neu5Ac α 2-3Gal as the minimum epitope for the sialic acid-FH interaction and indicate that the dissociation constant for the complex is in the μ M or mM range. Such low affinities are not unusual for glycan-protein complexes.

To test if FH also binds 3'SL in the presence of other serum proteins, we performed a double-difference saturation transfer (STDD) NMR experiment using FH-depleted serum instead of buffer²⁷. Serum

contains numerous proteins that bind Neu5Ac and are saturated along with FH in the selective irradiation step of the STD experiment. We therefore first collected a reference STD NMR spectrum of FH-depleted serum supplemented with 3'SL. Not surprisingly, some transfer to 3'SL from serum proteins other than FH was observed (Supplementary Results, Supplementary Fig. 1 online). We then repeated the experiment with a sample to which full length-FH was added. A comparison of both STDD spectra shows that additional saturation is transferred to 3'SL in the FH containing sample, suggesting that FH recognizes 3'SL also in the presence of other human serum proteins. The observed decrease of the STDD response may reflect competition of 3'SL with serum glycoproteins on which the 3'SLN glycan (3'sialyl-*N*-acetylglucosamine, Neu5Ac α 2-3Gal β 1-4GlcNAc) is abundant, as well as competition of FH with other serum proteins that bind 3'SL.

Crystal structure of FH19-20 bound to C3d and 3'SL

The sialic acid binding site in FH is likely located in its two most C-terminal domains^{13,28}. STD NMR experiments were therefore also acquired with a recombinant construct comprising these domains (FH19-20). The STD spectra were very similar for full-length FH and FH19-20 (Supplementary Fig. 2 online), indicating that domains 19 and 20 indeed mediate sialic acid binding. In the presence of both 3'SL and 6'SL (6' sialyl lactose, Neu5Ac α 2-6Gal β 1-4Glc) all observed resonances could be attributed to 3'SL (Supplementary Fig. 2 online), similar to what was observed for full-length FH. Next, FH19-20 and C3d (which corresponds to the cell-bound TED in C3b) were co-crystallized²² and a ternary complex was obtained by soaking with 3'SL (Supplementary Tables 1 and 2 online). Location, overall arrangement, and orientation of three crystallographic distinct 3'SL copies per asymmetric unit were essentially identical (Supplementary Fig. 3 and 4 online). One of the glycan binding sites did not participate in crystal contacts (chain F in the pdb), while in the other chains (D and E) the Glc exhibited crystal contacts. The Neu5Ac and Gal rings were not affected by crystal packing.

1 The Neu5Ac binding site is located in a recession between strand β 3 and the hypervariable loop
2 (residues 1182 to 1189) that forms one of the CCP-atypical lateral bulges in FH domain 20 (Fig. 2). A
3 network of hydrogen bonds and salt bridges anchors the Neu5Ac ring to the surrounding protein side
4 chains and backbone amides (Fig. 3a-d). The O8 and O9 hydroxyl groups of the glycerol chain are
5 hydrogen-bonded to the backbone amide and carbonyl groups of E1198, respectively, and O9 forms
6 another hydrogen bond with the W1183 backbone. The *N*-acetyl methyl group displays hydrophobic
7 interactions with the side chains of residues L1181 and V1197, and furthermore with portions of the
8 S1191 and E1195 side chains (Fig. 3a). The Neu5Ac amide group is hydrogen-bonded to the S1196
9 backbone carbonyl atom and the O4 hydroxyl group to the E1195 side chain. The Neu5Ac carboxyl group
10 forms a salt bridge with the side chain of R1215 (located in strand β 4) and is coordinated by an ordered
11 water molecule that is hydrogen-bonded to the S1196 amide. R1182, W1183, E1198, L1181 and F1199
12 confine the glycerol chain to a narrow pocket (Fig. 3a and d). Perhaps the most striking feature of the
13 binding site is the orientation of the W1183 indole ring, which packs against the 3'SL Gal and Glc rings
14 (Fig. 3b). This interaction is an important determinant for the orientation of the 3'SL lactose stem. The
15 Gal and Glc rings do not exhibit further interactions with the protein residues. This observation is in
16 agreement with the NMR experiments, which show that modifications to the lactose stem can be
17 accommodated. The strong Neu5Ac methyl peak and the observation of the Neu5Ac H4, H5, and H7
18 protons in the 3'SL difference spectrum are in agreement with the glycan conformation observed in the
19 crystal structure, and so is the Gal ring for which we observed magnetization transfer predominantly to
20 the face that is oriented towards the protein in the crystal.

21 To independently validate the observed interactions, two aHUS-associated *CFH* mutations were
22 evaluated for FH Neu5Ac binding in STD NMR experiments with 3'SL, W1183R and R1215Q. No
23 magnetization transfer was observed to either mutant, in agreement with the roles of the W1183 and
24 R1215 side chains in coordination of 3'SL in our crystal structure (Supplementary Fig. 2 online). In total,

more than a dozen *CFH* mutations and single nucleotide polymorphisms that predispose to aHUS²⁹ affect residues that are located in or close to the sialic acid binding site of FH (Fig. 3c-d).

The specificity of FH for sialylated glycans appears primarily determined by two factors: extensive and specific interactions with a single, non-reducing end Neu5Ac, and stacking interactions with the neighboring Gal. The latter interactions provide specificity for the α 2-3 linkage as compounds with α 2-6 or α 2-8 linked Neu5Ac are unlikely to form similar contacts. In a crystal structure of FH 19-20/C3d soaked with 40 mM 6'sialyl lactose (6'SL,) only weak electron density for Neu5Ac and Gal was observed, indicating lower affinity. The Gal protruded further from the binding pocket and did not stack efficiently with the W1183 side chain. The situation would be even less favorable for the Neu5Ac α 2-8Neu5Ac α disialyl cap. Notably, no magnetization transfer was observed to 6'SL or the Neu5Ac α 2-8 capped GD3 oligosaccharide in STD NMR experiments (Fig. 2).

A survey of the available crystallographic and NMR structures reveals that the FH domain 20 hypervariable loop residues 1184-1188 are inherently flexible. They adopt different conformations in the context of different crystal contacts (see also Supplementary Fig. 4 online), lack electron density in the absence of such contacts, and display a significant drop in ¹H, ¹⁵N-NOE values³⁰. Although in close proximity to 3'SL in our complex structure, these residues do not interact directly with the glycan. A comparison with the unliganded FH19-20/C3d structure²² shows that glycan binding nevertheless rigidifies this loop.

The ternary host recognition complex

The C3d binding site in C-terminal FH involves mostly domain 19 and the inter-domain linker^{22,23}, while the Neu5Ac binding site is contained entirely within domain 20. The two atoms that, in a physiological context, anchor both FH ligands to the cell surface (the Glc O1 in Neu5Ac and C δ of C3d residue Q20, Q1013 in the pre-processed C3b) are 44 Å apart (Fig. 4, the 1013 side chain is displayed as glutamate as after reaction with a cell nucleophile). The C3b Q1013 side chain dictates the orientation that the FH C-

terminal domains have to adopt in order to bind to cell-bound C3b. Both the Glc O1 and the C3b Q1013 side chain must lay on the same face of the ternary complex if 3'SL is reflective of a cell surface sialoglycoconjugate that is simultaneously bound by FH. This orientation is observed in our crystal structure (Fig. 4) which therefore supports the existence of a ternary host recognition complex formed by cell-bound C3b, FH, and sialylated glycans.

Cross-species conservation of the FH Neu5Ac binding site

Cross-species conservation of the FH sialic acid binding site was assessed by sequence comparison of human FH with five other mammalian sequences (Supplementary Fig. 5 online). Identities ranged from 98 % (chimpanzee) to 61 % (mouse) for the full-length protein. The amino acids that are most critical for sialic acid recognition, observed in our crystal structure, exhibit no or conservative mutations in all investigated sequences. W1183 and R1215 are strictly conserved. E1195 is subject to a conservative mutation in non-ape mammals. E1198, which may have a role in positioning the R1215 and W1183 side chains in the observed orientations, is also strictly conserved. R1182, stacking with the W1183 side chain, is conserved or subject to a conservative mutation. The hydrophobic base of the Neu5Ac binding site is also well-conserved: I1169, L1181, S1191, and F1199 are strictly conserved, and V1197 shows limited variation.

AP activation upon Neu5Ac-removal depends on properdin

We conducted competition experiments with Neu5Ac-containing glycans and multivalent neoglycoconjugates (3'SL and 6'SL coupled to serum ovalbumin). The hemolytic assay in which sheep erythrocytes (E_s) are exposed to normal human serum with Mg-EGTA served as a platform for AP-specific competition experiments. E_s activate the human AP only after treatment with neuraminidase^{11,12}. Direct competition with added sialic acid glycans or neoglycoconjugates at millimolar concentrations did not lead to increased cell lysis (data not shown). One possible explanation for this observation is that the high density of sialic acids on cell surfaces is critical for the low-affinity

interaction with FH and cannot be matched by soluble compounds. However, it is also possible that neuraminidase treated E_s do not only lose FH binding but also recruit an AP activator. Recent reports revived the so-called Pillemer-model in which properdin serves as a direct AP activator^{31,32}. Neuraminidase-treated E_s were therefore exposed to properdin-depleted serum (Supplementary Fig. 6 online). No lysis was observed in the absence of properdin, while supplementation with purified properdin resurrected cell lysis in a dose-dependent manner, suggesting that AP activity on erythrocytes is regulated by both FH and properdin.

Discussion

Our structural analysis of the FH-sialic acid interaction can be corroborated by past functional studies. These studies have found that the glycerol chain and carboxyl group of Neu5Ac are critical for FH recognition^{11,33,34} and that recognition is not determined by the presence of the Neu5Ac ring alone^{20,21}. Our crystal structure and NMR as well as mutagenesis experiments are in agreement with these observations as both the glycerol chain and carboxyl group undergo directed interactions with residues in the FH binding site and the recognition is not independent of the Neu5Ac glycosidic linkage. Incorporation of the GM3 ganglioside protects liposomes from complement-mediated lysis³⁵, and the 3'SL used here is in fact identical with the GM3 glycan moiety. The crystal structure allows to rationalize known mutations that affect the FH C-terminal domain, predispose to aHUS²⁹, and have been found to functionally impair FH^{36,37}. Some mutations remove the most important interactions with 3'SL (W1183C/F/L/R, R1215G/Q), clash with the ligand (L1189F/P/R, S1191L/W, S1191L+V1197A double mutant), or disrupt the hydrophobic pocket into which the glycerol chain inserts (F1199S) (Fig 3). Others may affect positioning of the W1183 side chain (E1198A, R1182S, E1198K) or alter the Neu5Ac methyl group pocket (V1197A, I1169L³⁸). T1184R and G1194D could alter the overall conformation of the binding site, although neither side chain directly coordinates Neu5Ac. In combination with the available biochemical data (erythrocyte neuraminidase treatment and

hemolytic assays), our findings suggest that 3'SL is a better proxy for erythrocytes than GAGs. Recombinant FH19-20 loses its capacity to lyse sensitized human erythrocytes via competition with full-length FH after neuraminidase treatment of the cells³⁹. FH6-8 and GAGs are the main determinants for FH host recognition in the human macula while the two C-terminal FH domains mediate recognition in the glomerulus, the kidney compartment affected by aHUS¹⁷. While GAGs play an important role for FH binding in the kidney, there must be additional host markers in this organ (but not in the eye)¹⁷. These findings, together with those presented here, provide strong evidence that sialic acid serves as a host marker in the glomerulus, imply distinct roles for sialic acid glycans and GAGs for complement regulation in the kidney, and indicate that sialic acid recognition is critical for aHUS predisposal. Sialic acid is abundant on the glomerular endothelium and necessary for glomerular integrity in mice models⁴⁰.

We would like to propose that the term 'host polyanions' to group GAGs and sialic acid glycans should be abandoned. Sulfated GAGs contain disaccharide repeats, adopt linear, helical structures⁴¹, are chemically inhomogeneous, and can have large molecular weights and numerous negative charges. In contrast, sialosides form comparably short sequences with often a single negative charge (the sialic acid carboxyl group), and can adopt branched structures. Our findings and past reports^{11,33,34} suggests that a specific recognition process rather than overall negative charge underlie FH binding to sialic acid, although the negative charge of the glycocalyx may well have a steering effect. We would like to suggest the terminology 'glycan markers' instead of 'polyanion markers' where both classes of FH glycan ligands are concerned.

Infection with *Streptococcus pneumonia* can cause sporadic HUS, possibly due to bacterial neuraminidase activity and exposure of the Thomson-Friedenreich (TF) antigen (Gal β 1-3GalNAc) to preformed IgM antibodies against this glycan⁴². In the light of our results, it is plausible that imbalanced regulation of complement after local reduction of FH binding may additionally contribute to the symptoms observed in pneumococcal HUS.

1 The FH19-20/3'SL complex further harbors implications for pathogenic immune evasion⁶. The outer
2 surface protein E (OspE) of lyme disease causing *Borrelia burgdorferi* specifically binds to the FH C-
3 terminus⁴⁴. Intriguingly, OspE binds FH via the W1183, S1196, E1198, and R1215 side chains that also
4 provide the most important interaction for 3'SL³⁸, assuring that FH bound to *B. burgdorferi* adopts an
5 orientation that resembles its host cell orientation. The host marker binding sites in FH are also likely
6 subject to evolutionary restrictions, an additional benefit for the FH-binding pathogens that bind to
7 these sites. Certain *Neisseria gonorrhoeae* strains can modify their lipooligosaccharide layer to present
8 an epitope similar to the human 3'SLN¹³. Thus, different bacteria have devised similar yet distinct
9 mechanisms to capture FH: either using dedicated proteins that bind FH via its C-terminal sialic acid
10 binding site or via direct glycan mimicry.

11 Properdin initiates AP C3-convertase formation on rabbit erythrocytes (E_R), activators of the AP
12 independent of neuraminidase-treatment³¹. Our findings indicate that on non-activating surfaces such
13 as E_S, removal of sialic acid might not only decrease FH binding but, additionally, augment properdin
14 binding. Both regulators may compete for binding to different cell surface types, independent of the
15 mechanism by which properdin activates the AP.

16 A comparison of the human FH sequence with other mammals revealed that key amino acids in the sialic
17 acid binding site are conserved among different mammals, suggesting a possible role for sialic acid self-
18 recognition by FH also in non-hominid mammals. Notably, mice which transgenically express mouse FH
19 without domains 17-20 spontaneously develop aHUS⁴⁵. While the key residues are conserved there are
20 non-conservative substitutions in proximate amino acids that do not directly interact with the Neu5Ac
21 ring, for instance in the N-terminal part of the flexible loop (amino acids 1184-1186). These could
22 modulate the relative sialoglycan binding affinities of individual FH variants somewhat, maybe reflecting
23 species-specific variations in the sialic acid family. Human FH is insensitive to some but not all occurring
24 variations of Neu5Ac^{33,34}. As an innate immune regulator FH does not possess the adaptability of

1 complex adaptive immunity recognition molecules, it therefore seems plausible that FH evolution
2 accounts for some variation in sialic acid type while keeping a firm grip on the overall glycan scaffold. It
3 is conceivable that the many human FH relatives that display varying degrees of conservation in their
4 glycan binding sites⁴⁶ also act to modulate the relatively broad FH sialic acid (and also GAG) specificity.
5 The Neu5Ac binding site in FH exhibits unexpected homology with human Siglecs (Sialic acid-binding
6 immunoglobulin-type lectins)⁴⁷ and a canine virus attachment protein^{48,49}. In all three proteins, the
7 Neu5Ac ring and glycerol chain align with an antiparallel β -sheet with β -sheet like interactions, while the
8 Neu5Ac carboxyl group is coordinated by an arginine residue protruding from a neighboring strand
9 (shown for FH in Fig. 3a). Given that the three proteins adopt different folds and have distinct biological
10 functions it seems plausible that they converged to a common sialic acid binding mode and that further
11 lectins with a homologous sialic acid binding site will be identified in the future.

ACKNOWLEDGMENTS

12 The authors thank Dr. Remco Sprangers, MPI for Development Biology (Tübingen, Germany), for
13 assistance with NMR spectra collection, the SLS beamline staff (Paul Scherrer Institute, Villigen,
14 Switzerland) for assistance with crystallographic data collection, and Dr. Holger Hengel, University
15 Hospital Tübingen, for venipuncture and assistance with serum preparation. Funding was provided by
16 the University of Tübingen Program for the Promotion of Junior Researchers and the University Hospital
17 Tübingen fortune Program (2186-0-0, to BSB). DK is a Wellcome Trust intermediate clinical fellow.

AUTHOR CONTRIBUTIONS

18 B.S.B. designed, conducted, and analyzed experiments, J.P.H. advised crystallization and D.U. NMR
19 experiments, J.P.H., A.P.H., D.K., and D.U. provided proteins, D.U. provided glycans, T.S. supervised the
20 project and provided vital reagents, all authors discussed findings, B.S.B. and T.S. wrote the manuscript.

COMPETING FINANCIAL INTERESTS

21 The authors declare no competing interests.

Figures

Figure 1: Chemical formulas of glycans tested in the STD NMR experiments with full-length FH. (a) 3'sialyl lactose (3'SL, Neu5Ac α 2-3Gal β 1-4Glc, R=OH) and 3'sialyl-*N*-acetylactosamine (3'SLN, Neu5Ac α 2-3Gal β 1-4GlcNAc, R=NHAc), (b) GM1b glycan (Neu5Ac α 2-3Gal β 1-3GalNAc β 1-4Gal β 1-4Glc), (c) GM1 glycan (Gal β 1-3GalNAc β 1-4(Neu5Ac α 2-3)Gal β 1-4Glc), (d) disialyl lacto-*N*-tetraose (DSLNT, Neu5Ac α 2-3Gal β 1-3(Neu5Ac α 2-6)GlcNAc β 1-3Gal β 1-4Glc) (e) 6'sialyl lactose (6'SL, Neu5Ac α 2-6Gal β 1-4Glc, R=OH) and 6'sialyl-*N*-acetylactosamine (6'SLN, Neu5Ac α 2-6Gal β 1-4GlcNAc, R=NHAc) (f) GD3 glycan (Neu5Ac α 2-8Neu5Ac α 2-3Gal β 1-4Glc).

Figure 2: FH binds α 2-3 linked sialic acid glycans in its C-terminal domain. NMR reference spectra of pure glycan (top) and STD difference spectra with FH (bottom) using (a) 3'sialyl lactose (b) 6'sialyl lactose and (c) the GD3 glycan. Glycans are depicted schematically and select proton resonances are highlighted. Highlighted Glc resonances belong to the β -anomer. The Neu5Ac resonances are labeled '3e' for the H3 equatorial proton and '3a' for the H3 axial proton. (d) Overall structure of FH19-20 without ligand. (e) Enlarged view of domain 20 with bound 3'SL (cyan). A schematic display of the ligand reflects its orientation in the binding site. N- and C-terminus of the respective domains are labeled, and red coloring of residues 1182-1199 in domain 20 highlights the ligand binding site in panels d and e. Diagrams were generated with Topspin and PyMOL.

Figure 3: The sialic acid FH binding site is a hotspot for aHUS mutations. (a) Enlargement of the sialic acid binding site with key amino acids highlighted (orange) and labeled. For clarity, only the sialic acid ring of 3'sialyl lactose is shown (cyan). Salt bridges and hydrogen bonds are shown as black dotted lines and a water molecule is displayed as a red sphere. (b) Rotated view of (a) with the full ligand (cyan) and only the W1183 side chain (orange) displayed. Inserted is the ligand in stick display with glycosidic linkages labeled. (c) and (d) Residues associated with aHUS shown in green. Wild type and disease-

associated variants of each amino acid are listed. Residue L1181 is not disease-associated but shown (orange) for completeness of the glycerol chain hydrophobic interactions.

Figure 4: Orientation of the ternary host recognition complex on cellular surfaces. (a) Superposition of the C3d/FH19-20/3'SL complex with C3b structure and schematic positioning on a cell surface. A C3d molecule of our ternary complex (chain B) was aligned with C3b TED (TED bluish grey, rest of C3b blue) in the published C3b structure (2I07.pdb⁸). FH19-20 and 3'SL are in red and cyan, respectively. The cell surface attachment points of the complex, the C3b residue 1013 (reacted to glutamate in 2I07.pdb and after attachment of cell surface hydroxy groups) and the 3'SL trisaccharide, are shown in stick representation.

Online Methods

Protein expression and purification.

Proteins were purchased (full-length FH) from Complement Technology, Inc. or expressed recombinantly in *E. coli* (C3d) and *Pichia pastoris* (FH19-20). Expression and purification of the recombinantly expressed proteins is described elsewhere^{22,37}.

Glycans.

Glycans were purchased from Carbosynth (3'SL(N) and 6'SL(N)) or Elicityl (GM1b glycan, GM1 glycan, GD3 glycan and DSLNT). Identity and purity of all glycans was assessed by ¹H and ¹³C NMR spectroscopy of individual glycans resolved in pure D₂O (Sigma Aldrich). In solution, the reducing end rings in oligosaccharides exchange between α- and β-anomeric forms, while their configurations are defined in the glycoconjugates which they decorate. Oligosaccharides used in this study are therefore denoted without stereo chemical specification for the reducing end ring throughout the text.

NMR spectroscopy.

NMR spectra were recorded at 288 K and 283 K using 3 mm tubes (200 μL sample volume) and a Bruker AVIII-600 spectrometer equipped with a room temperature probe head and processed with TOPSPIN 3.0

1 (Bruker). Samples for STD NMR spectra contained 2 mM of each of the different glycans (Carbosynth or
2 Elicityl) and varying concentrations ($10 \mu\text{M} \pm 2 \mu\text{M}$) of full-length FH (Complement Technology, Inc.) or 30
3 μM of recombinant FH19-20. Proteins were buffer-exchanged prior to NMR experiments in 3 kDa MWCO
4 centrifugal concentrators to 20 mM potassium phosphate pH 7.4, 150 mM NaCl in D_2O . Glycan samples
5 of the same concentration but with no protein present were used to verify that no direct excitation of
6 the glycans took place during the on-resonance irradiation step of the STD pulse program. Off- and on-
7 resonance irradiation frequencies were set to -30 ppm and 7.3 ppm, respectively. The irradiation power
8 of the selective pulses was 57 Hz, the saturation time was 2 s, and the total relaxation delay was 3 s. A 50
9 ms continuous-wave spin-lock pulse with a strength of 3.2 kHz was employed to suppress residual
10 protein signals. A total number of 512 scans were recorded. A total of 10,000 points were collected, and
11 spectra were multiplied with a Gaussian window function prior to Fourier transformation. Spectra were
12 referenced to 298 K using the α -D-Glc anomeric proton as an internal standard⁵⁰. For the double-
13 difference saturation experiment with serum proteins, 0.8 mL FH-depleted human serum (Complement
14 Technology, Inc.) was concentrated to 0.4 mL in a 10 kDa MWCO centrifugal concentrator. The
15 concentrated serum was diluted to 4 mL with 20 mM potassium phosphate pH 7.4 with 150 mM NaCl in
16 D_2O , supplemented with a 1 M EDTA stock to a final EDTA concentration of 20 mM, and concentrated
17 again to 0.4 mL, then split in two. 250 μg FH was added to one of the samples, yielding a concentration
18 of 8 μM fl-FH, and both samples were buffer-exchanged 3 times using the D_2O buffer without EDTA.
19 These buffer exchange steps served to remove small molecules from the serum and to exchange H_2O
20 against D_2O since no water suppression was used in the STD NMR spectra in order not to suppress the
21 glycan anomeric proton resonances. STD NMR spectra were collected with the aforementioned
22 parameters on both samples before and after addition of 3 mM 3'SL. Difference spectra obtained
23 without 3'SL were subtracted from difference spectra obtained with 3'SL in both cases to yield double-
24 difference saturation transfer (STDD) spectra of twofold concentrated serum proteins with 3'SL and no

FH, and with both 3'SL and FH.

Crystallography.

Co-crystals of FH19-20 and C3d were grown at 20 °C by hanging drop vapor diffusion essentially as described²². Crystals were transferred into a 40 mM 3'SL soaking solution prepared from 1 uL of a 400 mM 3'SL stock (in 20 mM Tris-HCl pH 7.0, 50 mM NaCl) and 9 uL reservoir solution (0.1 M Tris-HCl pH 9.0, 8 % (w/v) PEG 8K), stepwise cryo-protected with 20 % (v/v) glycerol in reservoir solution after 20 min of glycan soaking, and flash-frozen in liquid nitrogen. 3600 diffraction images with 0.1°/frame were collected at the X06DA beamline of the Swiss Light Source in Villingen, Switzerland, and processed with XDS⁵¹.

Model building.

The native FH19-20/C3d complex structure²² was used for initial refinement against the newly collected data in *REFMAC*⁵². Difference density maps revealed unambiguous electron density for the oligosaccharide in all three copies of FH 20. Water molecules were removed, the ligand was placed manually in the difference map, and the model went through several cycles of real-space refinement in COOT⁵³ and restrained reciprocal refinement using *REFMAC*. After the final refinement step, all 3'SL molecules were removed from the model and an unbiased omit map was produced in PHENIX⁵⁴ using the default parameters for simulated annealing. Model quality was assessed using MolProbity⁵⁵ and the Carbohydrate Structure Suite⁵⁶.

Protein sequence alignment.

FH19-20 protein sequences from other mammals were identified using a BLAST search with the human FH database entry (Uniprot P08603). The sequences of chimpanzee, orangutan, rabbit, horse, cow, and mouse were aligned with the human FH sequence using COBALT⁵⁷. The sheep sequence was not used because all sheep FH entries in the database are denoted 'fragment', with domain 20 missing

Hemolytic assay.

1 Sheep erythrocytes were obtained from defibrinated sheep blood (TCS Biosciences), properdin and
2 properdin-depleted serum were obtained from Complement Technologies Inc. 1 mL of sheep blood was
3 diluted with 10 ml cold buffer 1 (20 mM HEPES, 145 mM NaCl, 0.1 % (w/v) gelatin (from pork skin, Fluka)
4 and 10 mM EDTA, pH 7.3 at 25 °C). The cell suspension was mixed gently and spun for 10 min at 500 X g,
5 at 4 °C. The supernatant and a layer of leucocytes were aspirated off and the procedure was repeated
6 two more times with buffer 1 and subsequently three times with an identical buffer without EDTA but
7 with 5 mM MgCl₂ and with 5 mM EGTA (buffer 2). Thus obtained cells were treated with neuraminidase
8 from *Clostridium perfringens* (New England Biolabs) as follows: roughly 6 x 10⁹ cells were resuspended in
9 a total reaction volume of 100 µL using buffer 1 without EDTA and 500 units of neuraminidase, and the
10 reaction was incubated for 30 min at 37 °C. Subsequently, 150 µL of buffer 2 was added and the cells
11 were spun for 1 min at 500 g, at 4°C. The supernatant was removed and the washing step repeated two
12 more times. For the hemolytic assay, approximately 2.3 x 10⁸ neuraminidase-treated cells and cells that
13 were treated identically except for the omission of neuraminidase were used. The cells were mixed with
14 20 µL of human serum (intact, heat-inactivated at 56 °C for 30 min, or properdin-depleted) and buffer 2
15 and/or properdin to give an reaction volume of 40 µL. The reaction was incubated for 20 min at 37 °C,
16 stopped by addition of 150 µL cold buffer 1 without gelatin, and spun at 1500 g for 5 min at 4°C. Lysis
17 was determined by measuring the absorbance of the supernatant at 415 nm (A415) and compared to
18 the A415 value for the same amount of cells lysed with pure water.

19 **Accession codes.** Protein Data Bank: Coordinates and structure factors for the FH19-20/C3d/3'SL
20 complex were deposited under the accession number 4ONT.

REFERENCES

1. Ricklin, D., Hajishengallis, G., Yang, K. & Lambris, J. D. Complement: a key system for immune surveillance and homeostasis. *Nat. Immunol.* **11**, 785–797 (2010).
2. Zipfel, P. F. Complement and immune defense: From innate immunity to human diseases. *Immunol. Lett.* **126**, 1–7 (2009).
3. Ram, S., Lewis, L. A. & Rice, P. A. Infections of people with complement deficiencies and patients who have undergone splenectomy. *Clin. Microbiol. Rev.* **23**, 740–780 (2010).

- 1 4. Blue, C. E., Spiller, O. B. & Blackbourn, D. J. The relevance of complement to virus biology. *Virology*
2 **319**, 176–184 (2004).
- 3 5. Kim, D. D. & Song, W.-C. Membrane complement regulatory proteins. *Clin. Immunol.* **118**, 127–136
4 (2006).
- 5 6. Lambris, J. D., Ricklin, D. & Geisbrecht, B. V. Complement evasion by human pathogens. *Nat. Rev.*
6 *Microbiol.* **6**, 132–142 (2008).
- 7 7. Schmidt, C. Q. *et al.* The central portion of factor H (modules 10–15) is compact and contains a
8 structurally deviant CCP module. *J. Mol. Biol.* **395**, 105–122 (2010).
- 9 8. Janssen, B. J. C., Christodoulidou, A., McCarthy, A., Lambris, J. D. & Gros, P. Structure of C3b reveals
10 conformational changes that underlie complement activity. *Nature* **444**, 213–216 (2006).
- 11 9. Ferreira, V. P. & Pangburn, M. K. Factor H mediated cell surface protection from complement is
12 critical for the survival of PNH erythrocytes. *Blood* **110**, 2190–2192 (2007).
- 13 10. Renner, B. *et al.* The complement inhibitors Crry and factor H are critical for preventing autologous
14 complement activation on renal tubular epithelial cells. *J. Immunol.* **185**, 3086–3094 (2010).
- 15 11. Fearon, D. T. Regulation by membrane sialic acid of beta1H-dependent decay-dissociation of
16 amplification C3 convertase of the alternative complement pathway. *Proc. Natl. Acad. Sci. U. S. A.*
17 **75**, 1971–1975 (1978).
- 18 12. Pangburn, M. K. & Müller-Eberhard, H. J. Complement C3 convertase: cell surface restriction of
19 beta1H control and generation of restriction on neuraminidase-treated cells. *Proc. Natl. Acad. Sci. U.*
20 *S. A.* **75**, 2416–2420 (1978).
- 21 13. Ram, S. *et al.* A novel sialic acid binding site on factor H mediates serum resistance of sialylated
22 *Neisseria gonorrhoeae*. *J. Exp. Med.* **187**, 743–752 (1998).
- 23 14. Pangburn, M. K., Ferreira, V. P. & Cortes, C. Discrimination between Host and Pathogens by the
24 Complement System. *Vaccine* **26**, I15–I21 (2008).
- 25 15. Clark, S. J. *et al.* Impaired binding of the age-related macular degeneration-associated complement
26 factor H 402H allotype to Bruch's membrane in human retina. *J. Biol. Chem.* **285**, 30192–30202
27 (2010).
- 28 16. Prosser, B. E. *et al.* Structural basis for complement factor H linked age-related macular
29 degeneration. *J. Exp. Med.* **204**, 2277–2283 (2007).
- 30 17. Clark, S. J. *et al.* Tissue-specific host recognition by complement factor H is mediated by differential
31 activities of its glycosaminoglycan-binding regions. *J. Immunol.* **190**, 2049–2057 (2013).
- 32 18. Paulson, J. C., Macauley, M. S. & Kawasaki, N. Siglecs as sensors of self in innate and adaptive
33 immune responses. *Ann. N. Y. Acad. Sci.* **1253**, 37–48 (2012).
- 34 19. Crespo, H. J., Lau, J. T. Y. & Videira, P. A. Dendritic Cells: A Spot on Sialic Acid. *Front. Immunol.* **4**, 491
35 (2013).
- 36 20. Meri, S. & Pangburn, M. K. Discrimination between activators and nonactivators of the alternative
37 pathway of complement: regulation via a sialic acid/polyanion binding site on factor H. *Proc. Natl.*
38 *Acad. Sci. U. S. A.* **87**, 3982–3986 (1990).
- 39 21. Meri, S. & Pangburn, M. K. Regulation of alternative pathway complement activation by
40 glycosaminoglycans: specificity of the polyanion binding site on factor H. *Biochem. Biophys. Res.*
41 *Commun.* **198**, 52–59 (1994).
- 42 22. Morgan, H. P. *et al.* Structural basis for engagement by complement factor H of C3b on a self
43 surface. *Nat. Struct. Mol. Biol.* **18**, 463–470 (2011).
- 44 23. Kajander, T. *et al.* Dual interaction of factor H with C3d and glycosaminoglycans in host-nonhost
45 discrimination by complement. *Proc. Natl. Acad. Sci. U. S. A.* **108**, 2897–2902 (2011).
- 46 24. Varki, A. *et al.* *Essentials of Glycobiology*. (Cold Spring Harbor Laboratory Press, Cold Spring
47 Harbor/NY, 21 2009).

25. Mayer, M. & Meyer, B. Group epitope mapping by saturation transfer difference NMR to identify segments of a ligand in direct contact with a protein receptor. *J. Am. Chem. Soc.* **123**, 6108–6117 (2001).
26. Williamson, M. P. Applications of the NOE in Molecular Biology. In *Annual Reports on NMR Spectroscopy* Vol. 65 (ed. Graham A. Webb) 77–109 (Academic Press, London, 2009).
27. Claasen, B., Axmann, M., Meinecke, R. & Meyer, B. Direct observation of ligand binding to membrane proteins in living cells by a saturation transfer double difference (STDD) NMR spectroscopy method shows a significantly higher affinity of integrin $\alpha(\text{IIb})\beta 3$ in native platelets than in liposomes. *J. Am. Chem. Soc.* **127**, 916–919 (2005).
28. Ferreira, V. P., Herbert, A. P., Hocking, H. G., Barlow, P. N. & Pangburn, M. K. Critical role of the C-terminal domains of factor H in regulating complement activation at cell surfaces. *J. Immunol.* **177**, 6308–6316 (2006).
29. Kavanagh, D., Goodship, T. H. & Richards, A. Atypical hemolytic uremic syndrome. *Semin. Nephrol.* **33**, 508–530 (2013).
30. Herbert, A. P., Uhrin, D., Lyon, M., Pangburn, M. K. & Barlow, P. N. Disease-associated sequence variations congregate in a polyanion recognition patch on human factor H revealed in three-dimensional structure. *J. Biol. Chem.* **281**, 16512–16520 (2006).
31. Spitzer, D., Mitchell, L. M., Atkinson, J. P. & Hourcade, D. E. Properdin Can Initiate Complement Activation by Binding Specific Target Surfaces and Providing a Platform for De Novo Convertase Assembly. *J. Immunol.* **179**, 2600–2608 (2007).
32. Kemper, C., Atkinson, J. P. & Hourcade, D. E. Properdin: emerging roles of a pattern-recognition molecule. *Annu. Rev. Immunol.* **28**, 131–155 (2010).
33. Varki, A. & Kornfeld, S. An autosomal dominant gene regulates the extent of 9-O-acetylation of murine erythrocyte sialic acids. A probable explanation for the variation in capacity to activate the human alternate complement pathway. *J. Exp. Med.* **152**, 532–544 (1980).
34. Michalek, M. T., Mold, C. & Bremer, E. G. Inhibition of the alternative pathway of human complement by structural analogues of sialic acid. *J. Immunol.* **140**, 1588–1594 (1988).
35. Okada, N., Yasuda, T. & Okada, H. Restriction of alternative complement pathway activation by sialosylglycolipids. *Nature* **299**, 261–263 (1982).
36. Ferreira, V. P. *et al.* The binding of factor H to a complex of physiological polyanions and C3b on cells is impaired in atypical hemolytic uremic syndrome. *J. Immunol.* **182**, 7009–7018 (2009).
37. Herbert, A. P. *et al.* Structural and functional characterization of the product of disease-related factor H gene conversion. *Biochemistry* **51**, 1874–1884 (2012).
38. Maga, T. K., Nishimura, C. J., Weaver, A. E., Frees, K. L. & Smith, R. J. H. Mutations in alternative pathway complement proteins in American patients with atypical hemolytic uremic syndrome. *Hum. Mutat.* **31**, E1445–1460 (2010).
39. Schmidt, C. Q. *et al.* Rational Engineering of a Minimized Immune Inhibitor with Unique Triple-Targeting Properties. *J. Immunol.* **190**, 5712–5721 (2013).
40. Galeano, B. *et al.* Mutation in the key enzyme of sialic acid biosynthesis causes severe glomerular proteinuria and is rescued by N-acetylmannosamine. *J. Clin. Invest.* **117**, 1585–1594 (2007).
41. Mulloy, B., Forster, M. J., Jones, C. & Davies, D. B. N.m.r. and molecular-modelling studies of the solution conformation of heparin. *Biochem. J.* **293** (Pt 3), 849–858 (1993).
42. Klein, P. J. *et al.* THOMSEN FRIEDENREICH ANTIGEN IN HÆMOLYTIC-URÆMIC SYNDROME. *The Lancet* **310**, 1024–1025 (1977).
43. Copelovitch, L. & Kaplan, B. S. Streptococcus pneumoniae-associated hemolytic uremic syndrome. *Pediatr. Nephrol.* **23**, 1951–1956 (2008).
44. Bhattacharjee, A. *et al.* Structural basis for complement evasion by Lyme disease pathogen *Borrelia burgdorferi*. *J. Biol. Chem.* **288**, 18685–18695 (2013).

- 1 45. Pickering, M. C. *et al.* Spontaneous hemolytic uremic syndrome triggered by complement factor H
2 lacking surface recognition domains. *J. Exp. Med.* **204**, 1249–1256 (2007).
- 3 46. Skerka, C., Chen, Q., Fremeaux-Bacchi, V. & Roumenina, L. T. Complement factor H related proteins
4 (CFHRs). *Mol. Immunol.* **56**, 170–180 (2013).
- 5 47. Zhuravleva, M. A., Trandem, K. & Sun, P. D. Structural Implications of Siglec-5-Mediated Sialoglycan
6 Recognition. *J. Mol. Biol.* **375**, 437–447 (2008).
- 7 48. Rademacher, C. *et al.* A Siglec-like sialic-acid-binding motif revealed in an adenovirus capsid protein.
8 *Glycobiology* **22**, 1086–1091 (2012).
- 9 49. Seiradake, E. *et al.* The Cell Adhesion Molecule ‘CAR’ and Sialic Acid on Human Erythrocytes
10 Influence Adenovirus In Vivo Biodistribution. *PLoS Pathog* **5**, e1000277 (2009).
- 11 50. Nicholson, J. K., Foxall, P. J., Spraul, M., Farrant, R. D. & Lindon, J. C. 750 MHz ¹H and ¹H-¹³C NMR
12 spectroscopy of human blood plasma. *Anal. Chem.* **67**, 793–811 (1995).
- 13 51. Kabsch, W. Integration, scaling, space-group assignment and post-refinement. *Acta Crystallogr. D*
14 *Biol. Crystallogr.* **66**, 133–144 (2010).
- 15 52. Murshudov, G. N., Vagin, A. A. & Dodson, E. J. Refinement of Macromolecular Structures by the
16 Maximum-Likelihood Method. *Acta Crystallogr. Sect. D* **53**, 240–255 (1997).
- 17 53. Emsley, P. & Cowtan, K. Coot : model-building tools for molecular graphics. *Acta Crystallogr. D Biol.*
18 *Crystallogr.* **60**, 2126–2132 (2004).
- 19 54. Adams, P. D. *et al.* PHENIX: a comprehensive Python-based system for macromolecular structure
20 solution. *Acta Crystallogr. D Biol. Crystallogr.* **66**, 213–221 (2010).
- 21 55. Davis, I. W. *et al.* MolProbity: all-atom contacts and structure validation for proteins and nucleic
22 acids. *Nucleic Acids Res.* **35**, W375–383 (2007).
- 23 56. Frank, M., Lütteke, T. & von der Lieth, C.-W. GlycoMapsDB: a database of the accessible
24 conformational space of glycosidic linkages. *Nucleic Acids Res.* **35**, 287–290 (2007).
- 25 57. Papadopoulos, J. S. & Agarwala, R. COBALT: constraint-based alignment tool for multiple protein
26 sequences. *Bioinformatics* **23**, 1073–1079 (2007).