

Two distinct conformations of factor H regulate discrete complement-binding functions in the fluid phase and at cell surfaces

Received for publication, July 9, 2018, and in revised form, September 10, 2018. Published, Papers in Press, September 14, 2018, DOI 10.1074/jbc.RA118.004767

Amy J. Osborne¹, Ruodan Nan (南若丹)¹, Ami Miller, Jayesh S. Bhatt, Jayesh Gor, and  Stephen J. Perkins²

From the Department of Structural and Molecular Biology, Darwin Building, University College London, Gower Street, London WC1E 6BT, United Kingdom

Edited by Karen G. Fleming

Factor H (FH) is the major regulator of C3b in the alternative pathway of the complement system in immunity. FH comprises 20 short complement regulator (SCR) domains, including eight glycans, and its Y402H polymorphism predisposes those who carry it to age-related macular degeneration. To better understand FH complement binding and self-association, we have studied the solution structures of both the His-402 and Tyr-402 FH allotypes. Analytical ultracentrifugation revealed that up to 12% of both FH allotypes self-associate, and this was confirmed by small-angle X-ray scattering (SAXS), MS, and surface plasmon resonance analyses. SAXS showed that monomeric FH has a radius of gyration (R_g) of 7.2–7.8 nm and a length of 25 nm. Starting from known structures for the SCR domains and glycans, the SAXS data were fitted using Monte Carlo methods to determine atomistic structures of monomeric FH. The analysis of 29,715 physically realistic but randomized FH conformations resulted in 100 similar best-fit FH structures for each allotype. Two distinct molecular structures resulted that showed either an extended N-terminal domain arrangement with a folded-back C terminus or an extended C terminus and a folded-back N terminus. These two structures are the most accurate to date for glycosylated full-length FH. To clarify FH functional roles in host protection, crystal structures for the FH complexes with C3b and C3dg revealed that the extended N-terminal conformation accounted for C3b fluid-phase regulation, the extended C-terminal conformation accounted for C3d binding, and both conformations accounted for bivalent FH binding to glycosaminoglycans on the target cell surface.

The alternative pathway of complement is activated by the spontaneous hydrolysis of C3 into C3u, also known as C3(H₂O). This leads to a positive-feedback amplification of C3 cleavage to form activated C3b that opsonizes pathogenic cells.

This work was supported in part by a Ph.D. studentship grant from Alexion to Complement UK, Mizutani Foundation Award 180218, CCP-SAS Project through joint EPSRC Grant EP/K039121/1, National Science Foundation Grant CHE-1265821, and MRC Grant G0801724. The authors declare that they have no conflicts of interest with the contents of this article.

✂ Author's Choice—Final version open access under the terms of the Creative Commons CC-BY license.

This article contains [supporting material](#).

¹ Both authors contributed equally to this work.

² To whom correspondence should be addressed. Tel.: 020-7679-7048; Fax: 020-7679-7193; E-mail: s.perkins@ucl.ac.uk.

To prevent unwanted C3b-mediated host cell damage, complement factor H (FH)³ regulates complement by acting as a cofactor for factor I (FI) to cleave C3b (1–3), competing with factor B (FB) to interfere with the formation of the C3 convertase C3bBb (4), and accelerating the decay of the C3bBb complex (2, 5). These activities occur in the fluid phase and less effectively at the host cell surface (6).

Disrupted complement regulation is associated with age-related macular degeneration (AMD), the most common cause of blindness in the West (7–10), and also with atypical hemolytic uremic syndrome (aHUS), C3 glomerulopathy (C3G), and Alzheimer's disease (11–13). In early stage AMD, sub-retinal pigment epithelial deposits known as drusen develop within Bruch's membrane, which is an extracellular matrix layer interposed between the retinal pigment epithelium and the choroidal vasculature (14). Drusen contains oxidized lipids, carbohydrates, cellular materials, and over 200 proteins, including FH and other complement components (10, 15, 16).

FH is composed of 20 short complement regulator (SCR) domains of ~61 residues connected by linkers of 3–8 residues in length (Fig. 1) (17). Eight out of nine potential N-glycosylation sites in FH are occupied by biantennary disialylated glycans (18). Functionally, FH has multiple binding sites for its major ligands C3b, the C3d fragment, C-reactive protein (CRP), and two host-cell markers heparin, which is an analogue of the natural ligand heparan sulfate (HS), and sialic acid (19–21). Each ligand binds to FH weakly with micromolar K_D values (22) at specific SCR domains summarized in Fig. 1. For example, FH has two binding sites for heparin at SCR-7 and SCR-20 (22–26), and weak independent interactions with heparin at SCR-7 and SCR-20 result in bivalent binding with an overall greater affinity than either site on their own (21, 27). The dimerization of FH at high concentrations was first reported by X-ray and neutron scattering (28). FH also self-associates at each of SCR-6/8 and SCR-16/20 with micromolar K_D values; the sequential daisy-chaining of these two dimer sites leads to FH oligomers

³ The abbreviations used are: FH, factor H; AMD, age-related macular degeneration; SCR, short complement regulator; SPR, surface plasmon resonance; SAXS, small-angle X-ray scattering; FI, factor I; NT, N-terminal; CT, C-terminal; COM, center of mass; AUC, analytical ultracentrifugation; PDB, Protein Data Bank; PSF, protein structure file; RU, response unit; TAMC, torsion angle Monte Carlo; CRP, C-reactive protein; HS, heparan sulfate; PCA, principal component analysis; DOPE, discrete optimized protein energy; aHUS, atypical hemolytic uremic syndrome; C3G, C3 glomerulopathy; HSA, human serum albumin.

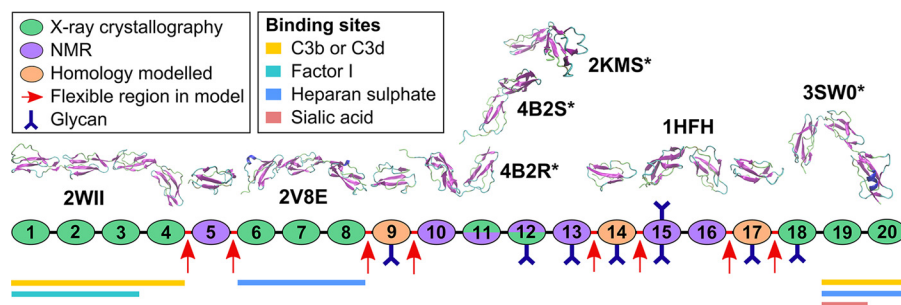


Figure 1. Cartoon of the 20 SCR domains in FH. Each SCR domain is represented by an ellipse color-coded to indicate the starting atomistic structure (green, X-ray crystallography; purple, NMR; orange, homology modelling). Ribbon views of the seven SCR structures used to model FH are shown above the cartoon, together with their PDB codes. The asterisks signify newer SCR structural models that became available after the previous full-length FH models were published. Below the cartoon, the FH functional binding sites are denoted by horizontal bars for each of C3b, C3d, factor I, heparan sulfate, and sialic acid.

(26, 29–31). The formation of dimers and tetramers of FH may also be involved in host cell recognition essential for complement regulation (32). However, it is often presumed that FH is monomeric (33).

Genetic variants in FH are associated with disease (34). In AMD, FH SCR-7 harbors the AMD-risk polymorphism Y402H (7–10). Individuals homozygous for His-402 have a 6-fold increased risk of developing AMD compared with 2.5-fold for His-402 heterozygotes (35). The presence of His-402 weakened heparin binding to both FH SCR-6/8 and full-length FH (36, 37), FH binding to CRP (38, 39), and full-length FH binding to HS in Bruch's membrane (40). For the development of AMD, age-related changes in glycosaminoglycan structures such as a decrease in the sulfation of HS, together with the presence of FH His-402, have been proposed to reduce FH binding over time. At host cell surfaces, FH His-402 showed weaker SCR-7 binding to eye tissue-specific HS only (41). In opposition to these observations of weaker HS binding for the His-402 allotype, His-402 in SCR-6/8 was seen to be bound to the highly-sulfated glycosaminoglycan analogue sucrose octasulfate using X-ray crystallography, whereas no such crystals were reported for the Tyr-402 allotype (42). Although the statistical association of Y402H with AMD is one of the strongest, many other complement variants in *CFH* and other complement genes associate also with AMD, suggesting that AMD is caused by several different molecular mechanisms. For example, at the other heparin-binding site on FH, SCR-20 harbors the AMD-risk rare variant R1210C (43). AMD is associated with another 12 rare variants in FH in its signal peptide or in SCR-1, SCR-3/5, SCR-8, SCR-16, or SCR-18 (44). AMD is also associated with the protective polymorphism V62I in SCR-1 (10, 45), common synonymous and noncoding variants in a region overlapping the *CFH* gene (36, 46, 47), and genetic variants in complement C3, *CFI*, *CFB*, and *C9* (44, 49).

Additional insights into molecular mechanisms for AMD may be gained by studying both the self-association and overall solution structure of FH in the context of the Y402H polymorphism. To date, full-length FH has proven too large, flexible, and glycosylated for high-resolution structural determination by crystallography or NMR. Preliminary molecular modeling of small angle X-ray scattering (SAXS) curves revealed that full-length FH in solution has a folded-back SCR structure that is affected by ionic strength (26, 50, 51). To examine the effect of the Y402H polymorphism on FH self-association and its solu-

tion structure, we submitted each of the homozygous FH Tyr-402 and His-402 to analytical ultracentrifugation (AUC), SAXS, MS, and surface plasmon resonance studies. Both FH allotypes self-associated to form oligomers, and SAXS curve extrapolation to zero concentration showed that both FH allotypes exhibited similar overall structures. Our preliminary FH modeling was based on just 11–14 high-resolution SCR structures and an initial molecular dynamics approach to model the inter-SCR linkers (26, 51). Here, the FH modeling is much improved by starting from 17 high-resolution SCR structures (Fig. 1; Table 1) (26, 45, 52, 53) and utilizing a much improved Monte Carlo atomistic modeling method for the full-length FH structure, including the eight FH glycan chains (54). The resulting SAXS best-fit structural ensembles corresponded to the most accurate FH models determined to date. These ensembles revealed two different extended and folded-back FH domain arrangements that favor two different modes of FH binding to either C3b or C3d. Both arrangements accounted for the bivalent binding of FH to HS-coated cell surfaces. The outcome of two different FH conformations provides new insights into the way in which FH regulates complement C3b activity.

Results

Sedimentation velocity analyses of the two FH allotypes

AUC studies use macromolecular sedimentation behavior under a high centrifugal force to determine sizes and shapes (55). The full-length FH Tyr-402 and FH His-402 allotypes were purified from the plasma of genotyped volunteers (Table 3) as single peaks from size-exclusion chromatography (Fig. 2) that gave clean single bands by SDS-PAGE (29). Here and below, FH was genotyped for both Y402H and I62V polymorphisms, with our focus on the former, whereas the presence of Val-62 or Ile-62 in the FH samples was noted for completion but was not investigated further (56). Previous fluid-phase activity assays using C3u and FI with both FH allotypes showed that both demonstrated similar C3u cleavage rates (57).

The higher oligomers previously observed for heterozygous FH (26, 29–31) were reassessed using AUC sedimentation velocity experiments on five homozygous FH Tyr-402 and five homozygous FH His-402/Val-62 samples at three concentrations between 0.5 and 2.1 mg/ml in HEPES buffers (10 mM HEPES, 137 mM NaCl, pH 7.4). Similar sedimentation boundaries were observed for the two allotypes (Fig. 3A), and these

Solution structures of FH Tyr-402 and His-402

Table 1
Sources of molecular SCR structures in FH

FH domain(s)	PDB code	Ref.	Method	Used in previous FH model
SCR-1/4	2WII	52	X-ray crystallography	51
SCR-1/3	2RLP/2RLQ	45	NMR	26
SCR-4		112	Homology	26
SCR-5		95	NMR	26, 51
SCR-6/8	2V8E	113	X-ray crystallography	26, 51
SCR-9 ^a		112	Homology	26, 51
SCR-10/11 ^b	4B2R	114	NMR	
SCR-10/13		112	Homology	26
SCR-11/12 ^b	4B2S	114	X-ray crystallography	
SCR-12/13	2KMS	53	NMR	51
SCR-14 ^a		112	Homology	26, 51
SCR-15/16	1HFH	115	NMR	26, 51
SCR-17 ^a		112	Homology	26, 51
SCR-18		112	Homology	26
SCR-18/20 ^b	3SW0	79	X-ray crystallography	
SCR-19/20	2G7I	48	X-ray crystallography	26, 51

^a Only homology models are available for this (Fig. 1).

^b Newer SCR structural models published after the previous full-length FH models were published (Fig. 1).

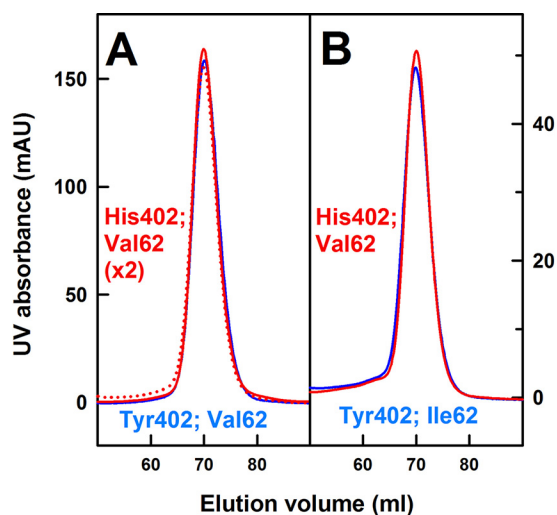


Figure 2. Size-exclusion gel filtration of homozygous FH. Pairs of homozygous purified FH Tyr-402 and His-402 at A, 0.9 mg/ml, and B, 0.3 mg/ml, were loaded onto a Superose™ 6 prep grade XK 16/60 column immediately after AUC sedimentation velocity experiments. Here and below, data for the Tyr-402 and His-402 allotypes are denoted in *blue* and *red*, respectively.

were fitted using size-distribution analyses $c(s)$. Good fits with low residuals were obtained in all 10 cases. The mean sedimentation coefficient $s_{20, w}$ of monomeric FH (Tyr-402) at 50,000 rpm was 5.69 ± 0.03 S and that for FH (His-402) was identical at 5.67 ± 0.04 S (*peak 1* in Fig. 3B). Both values (Table 3) agree well with previous $s_{20, w}$ values of 5.3 ± 0.1 S (50) and 5.65 ± 0.05 S (29) for WT monomeric FH in 137 mM NaCl. Conversion of the $c(s)$ distribution to mass distribution plots $c(M)$ showed that the major peak 1 corresponded to a molecular mass of 147 ± 10 kDa for FH Tyr-402 and 147 ± 11 kDa for FH His-402. Both values agreed with the sequence-calculated molecular mass of 154 kDa for both allotypes and previous $c(M)$ mass determinations of 142 ± 2 kDa for native FH (29). Integration showed that the dimer and trimer peaks comprised 4.0 ± 1.9 and $2.1 \pm 0.7\%$, respectively, for FH Tyr-402 and 5.7 ± 1.5 and $2.5 \pm 0.6\%$, respectively, for FH His-402. Averaging the $c(s)$ distributions established the observation of these dimer and trimer peaks (Fig. 3B). Hence, at 1.1 mg/ml, the His-402 allotype formed slightly more dimer and trimer than the Tyr-402 allotype, but the difference was small (*peaks 2 and 3* in Fig. 3B). The mean

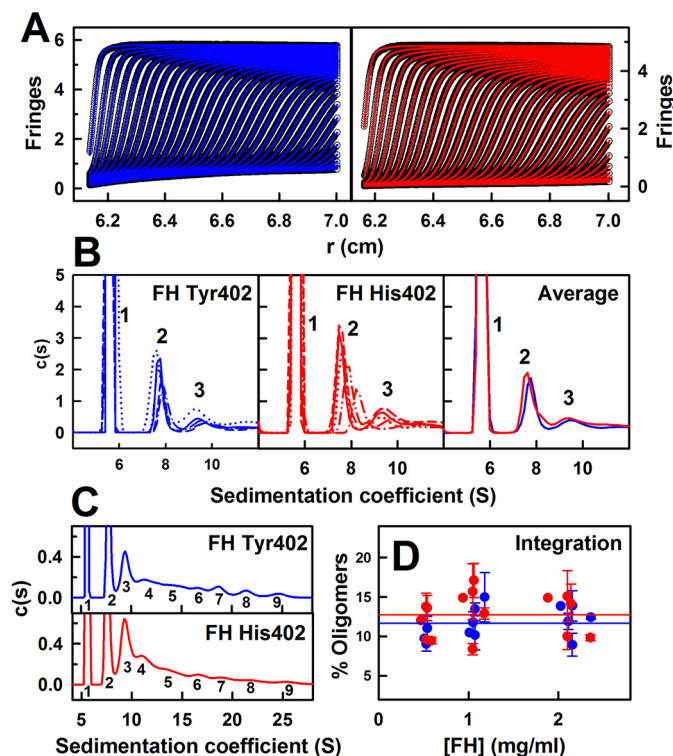


Figure 3. Sedimentation velocity $c(s)$ size-distribution analyses for FH Tyr-402 and FH His-402. All analyses correspond to 1.1 mg/ml FH. A, representative sedimentation boundary fits corresponding to FH Tyr-402 (*blue*) and FH His-402 (*red*) are shown using only every 10th scan for clarity. The experimental data are shown in *black*. B, peaks 1–3 in the $c(s)$ analyses for five FH Tyr-402 and five FH His-402 samples (*red*) are shown in different *line styles*. The averaged $c(s)$ plots for FH Tyr-402 and FH His-402 are shown at the *right*. The monomer peak 1 at 5.7 S is normalized to 100. C, full $c(s)$ analyses for FH Tyr-402 and FH His-402 reveal up to nine FH oligomer species starting with peaks 1 and 2 for monomer and dimer and extending to peak 9 for presumed nonamers. D, proportion of FH oligomers was derived by integration of the $c(s)$ analyses for peaks 2–9 for the five FH Tyr-402 and FH His-402 samples at concentrations between 0.5 and 2.4 mg/ml. Statistical *error bars* are shown where visible. The mean proportion of oligomers is shown as *horizontal lines*.

$s_{20, w}$ value of dimeric FH Tyr-402 was 7.57 ± 0.36 S, and that for dimeric FH His-402 was similar at 7.69 ± 0.40 S (*peak 2* in Fig. 3B). The mean $s_{20, w}$ value of trimeric FH Tyr-402 was 9.3 ± 0.7 S and that for trimeric FH His-402 was similar at 9.4 ± 0.8 S (*peak 3* in Fig. 3B). These values agree well with previous values of 7.3 ± 0.2 and 9.2 ± 0.2 S for heterozygous WT FH dimers and

trimers, respectively, in a range of NaCl concentrations and pH values, and their size increases with the increase in FH concentration (26, 29). In general, reaction boundaries give single $c(s)$ peaks that correspond to the co-sedimentation of two different species in rapid equilibrium and exhibit well-defined sedimentation coefficients between the values of the two different species (58). The observation of distinct dimer and trimer $c(s)$ peaks thus corresponded to slow equilibria between monomer, dimer, and trimer FH on the time scale of sedimentation.

AUC had showed previously that FH was described by a reversible monomer–dimer equilibrium up to 1.3 mg/ml and that heterozygous FH above 2 mg/ml forms irreversible high oligomers (29). To assess these with homozygous FH, size-exclusion chromatography was performed after the proteins had sedimented by AUC. After resuspending by gentle shaking, size-exclusion chromatography showed that the single homogeneous FH peak between 65 and 75 ml was supplemented by small additional peaks corresponding to irreversible FH oligomers that eluted earlier between 55 and 65 ml in all cases, as expected (Fig. 2). The main monomer peak was eluted at the same position in each case, indicating no structural differences between the allotypes.

In heterozygous WT FH, oligomers from tetramers to nonamers have previously been observed at concentrations as low as 0.4 mg/ml (26, 29). Here, their existence was also observed for FH Tyr-402 at 1.1 mg/ml, and the $s_{20,w}$ values of peaks 4–9 (Fig. 3C) were 11.9 ± 1.0 S, 14.7 ± 0.8 S, 17.1 ± 0.7 S, 19.4 ± 1.0 S, 22.1 ± 1.0 S, and 25.0 ± 0.7 S in that order. For FH His-402 at 1.1 mg/ml, the $s_{20,w}$ values of peaks 4–9 (Fig. 3C) were in agreement at 11.9 ± 0.9 S, 14.3 ± 0.5 S, 16.7 ± 0.4 S, 18.7 ± 0.6 S, 21.3 ± 0.5 S, and 24.8 ± 0.7 S in that order. Integrations of peaks 2–9 yielded the percentages of oligomers (Fig. 3D). Values of 11.7 ± 2.0 and $12.8 \pm 2.7\%$ of oligomers were seen for FH Tyr-402 and FH His-402, respectively, with this being marginally higher for FH His-402. Both percentages are consistent with an estimate of 17% of dimers predicted from a self-association K_D value of $28 \mu\text{M}$ (29).

Mass spectrometry of the two FH allotypes

Because FH self-association had only previously been reported by AUC, MS of homozygous FH was used to further examine self-association. MS observes macromolecular assemblies and determines their stoichiometry (59). One pair of homozygous FH Tyr-402/Val-62 and His-402/Val-62 allotypes was studied shortly after their concentration to about 4 mg/ml in order to promote the formation of oligomers. The use of 140 mM ammonium acetate buffer, pH 7.0, and 1 M ammonium acetate buffer, pH 7.0, evaluated two ionic strengths, given the decrease in oligomers with an increase in the NaCl concentration (26). Ammonium acetate was used because this is volatile in the gas phase used for measurements, unlike the HEPES buffer used for AUC. All four samples showed signals with five to six components at mass/charge (m/z) ratios that corresponded to the FH monomer with charges of +23 to +27 (Fig. 4). For FH Tyr-402, the monomer mass was determined to be 150 kDa in 140 mM ammonium acetate and 157 kDa in 1 M ammonium acetate (Fig. 4, A and B). Similar masses were obtained for FH His-402, giving 156 kDa in 140 mM ammonium

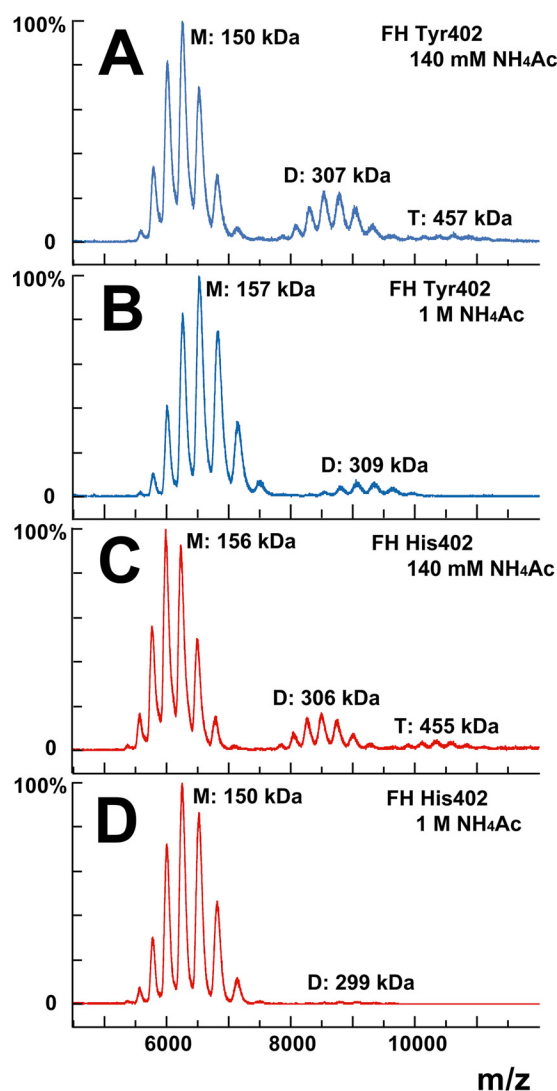


Figure 4. Mass spectrometry of FH Tyr-402 and FH His-402. The FH monomer, dimer, and trimer peaks are marked by *M*, *D*, and *T* and their measured molecular weights. *A*, homozygous FH Tyr-402/Val-62 was studied at 3.9 mg/ml ($25.3 \mu\text{M}$) in 140 mM ammonium acetate buffer, pH 7.0. *B*, same FH Tyr-402/Val-62 sample was studied at 4.4 mg/ml ($28.6 \mu\text{M}$) in 1 M ammonium acetate buffer, pH 7.0. *C*, homozygous FH His-402/Val-62 was studied at 3.9 mg/ml ($25.3 \mu\text{M}$) in 140 mM ammonium acetate buffer, pH 7.0. *D*, same FH His-402/Val-62 sample was studied at 3.6 mg/ml ($23.4 \mu\text{M}$) in 1 M ammonium acetate buffer, pH 7.0.

acetate and 150 kDa in 1 M ammonium acetate (Fig. 4, C and D). These masses agreed well with those from AUC (26, 29). They also agreed with the sequence-calculated mass of 154 kDa for FH based on eight glycan chains (18). The variability of the mass determinations was attributed to glycan heterogeneity in FH.

Dimer and trimer FH were also observed in the mass spectra (Fig. 4). For both FH Tyr-402 and His-402, the dimer and trimer were seen in 140 mM ammonium acetate, whereas only the dimer was seen in 1 M ammonium acetate. Their molecular masses were determined to be 307 and 457 kDa and 309 kDa, respectively (Fig. 4, A and B). Similar results were obtained for FH His-402, now giving masses of 306 and 455 kDa in 140 mM ammonium acetate, and 299 kDa in 1 M ammonium acetate (Fig. 4, C and D). Peak integrations gave the proportions of monomer, dimer, and trimer as 73, 24, and 4% for FH Tyr-402

Solution structures of FH Tyr-402 and His-402

and 76, 17, and 6% for FH His-402 in 140 mM ammonium acetate (Fig. 4, A and C). The appearance of dimer and trimer peaks in the mass spectra validates the observation of dimer and trimer by AUC (Fig. 3B), and the summed integration was comparable with 11.7–12.8% by AUC from Fig. 3D. At 4 mg/ml, dimer formation was estimated as 37% from a K_D value of 28 μM , with this being comparable with the observed amounts of dimer. The proportions of monomer and dimer changed to 90 and 10% for FH Tyr-402 and 98 and 2% for FH His-402 in 1 M ammonium acetate (Fig. 4, B and D). The reduction of dimer and the disappearance of trimer in 1 M ammonium acetate agrees well with the effect of increased ionic strength in reducing FH self-association (26). No difference was detectable between the two allotypes.

Surface plasmon resonance of the two FH allotypes and FH SCR-6/8

Surface plasmon resonance (SPR) of the binding of solution-phase FH to surface-immobilized FH further enabled FH self-association to be monitored (60). 800 RU of each of homozygous FH Tyr-402 and FH His-402/Val-62 were immobilized on CM3 chips. Equilibrium experiments were performed by flowing FH Tyr-402 over the chip at eight concentrations between 1.3 and 10.9 μM and FH His-402 at concentrations between 1.3 and 12.3 μM in HEPES buffer. The binding of both FH allotypes to their immobilized FH partner was observed at all concentrations (Fig. 5, A and B) indicating that FH self-association proceeded in physiologically relevant conditions. In equilibrium analyses, the individual sensorgrams could not be fitted to a 1:1 binding kinetic model (data not shown). This key outcome indicated that FH self-association could not be explained as a 1:1 model and that more than one self-association site existed in FH with different affinities, in agreement with the multiple oligomers seen by AUC (29). Alternative models to fit the SPR data were not attempted because of the complication that both the ligand and analyte showed bivalent binding for each other. The maximum binding responses were used to fit a steady-state affinity model. An estimated dissociation constant K_D value of 8.0 μM with a χ^2 value of 2.6 RU^2 was determined for FH Tyr-402, which was almost identical to an estimated K_D value of 6.7 μM with a χ^2 value of 6.5 RU^2 determined for FH His-402 (Fig. 5, A and B). Both values were smaller than the K_D value of 28 μM determined for WT FH in solution by sedimentation equilibrium AUC (29). The stronger self-association seen by SPR than by AUC may result from different avidity effects at a surface that amplifies the strength of binding, although the linearity of the binding data in a comparatively low FH concentration range precluded a quantitative comparison with the AUC value.

The SPR results were examined further using the SCR-6/8 fragment. Each SCR-6/8 allotype was immobilized to a CM4 chip up to 150 RU, and the His-402 allotype was also immobilized up to 400 RU on a third chip (Fig. 5, C and D). The binding of SCR-6/8 Tyr-402 was studied between 0.1 and 0.8 mg/ml (5–40 μM). The binding of SCR-6/8 His-402 was studied between 0.1 and 0.6 mg/ml (5–28 μM). Both allotypes self-associated with each other. The immobilized Tyr-402 allotype showed the same binding affinities to both allotypes (Fig. 5C). However, immobilized His-402 allotype showed a higher self-

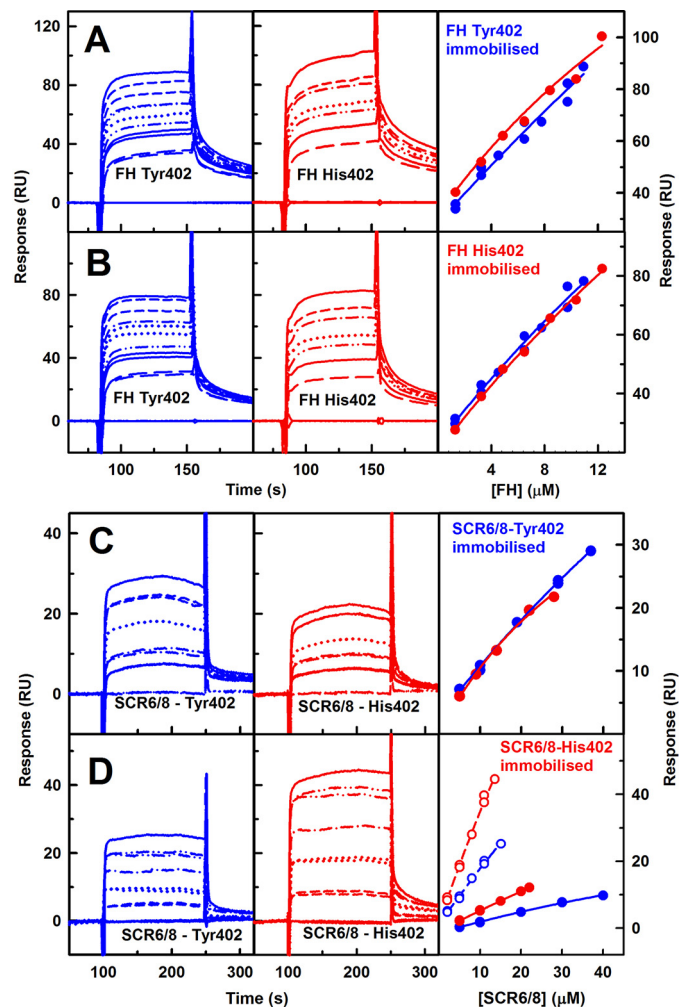


Figure 5. Surface plasmon resonance analysis of FH Tyr-402 and FH His-402 self-association. The sensorgrams are shown after subtraction of that recorded with the buffer only. In the right panels, no K_D values were determined because these were greater than 40 μM . A, 800 RU of homozygous FH Tyr-402/Val-62 was immobilized on a CM3 chip. FH Tyr-402/Val-62 was analyzed at eight concentrations of 0 μM (measured twice), 1.30 μM (twice), 3.24 μM (twice), 4.53 μM , 6.48 μM (twice), 7.77 μM , 9.72 μM (twice), and 10.93 μM from bottom to top. FH His-402/Val-62 was analyzed at eight concentrations of 0 μM (twice) and 1.30, 3.24, 4.86, 6.48, 8.42, 10.37, and 12.31 μM . Here and below, binding affinities were fitted using the maximum response values. B, 800 RU of homozygous FH His-402/Val-62 was immobilized on a CM3 chip. The same FH Tyr-402/Val-62 and FH His-402/Val-62 concentrations were used. C, 150 RU of SCR-6/8 Tyr-402 was immobilized on a CM4 chip. SCR-6/8 Tyr-402 was analyzed at concentrations of six concentrations of 0 μM (twice), 5 and 10 μM (twice), 19 μM , 29 μM (twice), and 37 μM from bottom to top. For the His-402 allotype, six concentrations of 0 μM (twice), 5 and 10 μM (twice), 19 and 29 μM (twice), and 37 μM were used, each measured in duplicate. D, 400 RU of SCR-6/8 His-402 was immobilized on a CM4 chip. The binding affinities were likewise fitted using the maximum response values (open circles). A second experiment with 150 RU of SCR-6/8 His-402 immobilized on a CM4 chip and six concentrations of 0, 5, 10, 15, 20, and 22 μM SCR-6/8 is also shown (filled circles).

association with His-402 than with Tyr-402 (Fig. 5D). This supported earlier AUC experiments suggesting that the His-402 allotype of SCR-6/8 self-associated more than the Tyr-402 allotype (30). The K_D values could not be determined from steady-state affinity fits because the K_D values were weaker than 40 μM ; however, this outcome agreed with the K_D value of 40 μM from AUC (30). Comparison of this value with the lower K_D values of 28 μM for FH self-association by AUC (29) and 7–8 μM by SPR

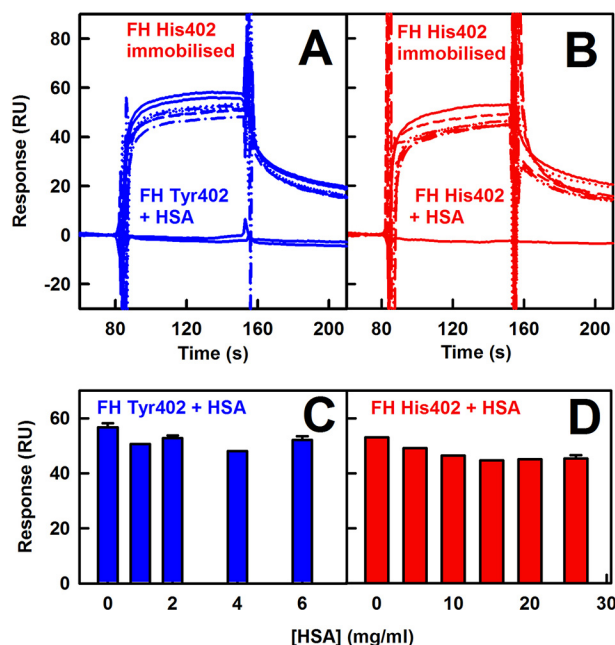


Figure 6. Surface plasmon resonance analysis of FH self-association in the presence of HSA. 800 RU of FH His-402 was immobilized on a CM3 chip. The sensorgrams are shown after subtraction of the sensorgram with those of HSA at the corresponding concentrations. The buffer sensorgram is shown at the bottom. A, FH Tyr-402/Val-62 at 4.53 μM was analyzed with HSA at concentrations between 0 and 6 mg/ml. B, FH His-402 at 4.53 μM was analyzed with HSA at concentrations of 0–26 mg/ml. C and D, bar charts show the maximum responses from A and B, respectively. Error bars are shown where large enough to be observed.

(above) showed that other self-association sites must exist in FH as well as in SCR-6/8.

Human serum albumin is the most abundant protein in serum, being present at 35–50 mg/ml. As a control of potential nonspecific protein–protein interactions that might influence FH self-association, further SPR experiments were performed with solution-phase FH Tyr-402 and His-402 flowed over immobilized FH His-402 in the presence of up to 6 and 25 mg/ml human serum albumin (Fig. 6, A and B). FH self-association continued to be observed even in the presence of high amounts of albumin (Fig. 6, C and D).

X-ray scattering of the two FH allotypes

X-ray scattering is a diffraction method to determine solution structures of macromolecules in random orientations (61). Here, two pairs of homozygous FH Tyr-402 and FH His-402/Val-62 samples (Table 3) were each freshly prepared and studied in two different sessions in five concentrations between 0.4 mg/ml (2.6 μM) and 3.3 mg/ml (21.4 μM) in HEPES buffer. Excellent signal-to-noise ratios were obtained with no detectable effect from radiation damage. Linear Guinier fits at low Q values (where $Q = 4\pi \sin \theta/\lambda$; 2θ = scattering angle; λ = wavelength) gave the radius of gyration (R_g) in satisfactory $Q \cdot R_g$ ranges below 1.1 (Fig. 7A). The R_g value monitors the degree of elongation of FH. At larger Q values, the cross-sectional Guinier R_{XS} fits monitor the structural proximity between non-neighboring SCR domains (denoted as R_{XS-1}) and between neighboring SCR domains (denoted as R_{XS-2}) (50). Both the R_{XS-1} and R_{XS-2} parameters were likewise obtained from linear

fits within satisfactory $Q \cdot R_{XS}$ ranges (Fig. 7B). This permitted comparative data analyses for the four FH samples. At 0.4 mg/ml, the R_g values were 7.54 and 7.97 nm for both FH Tyr-402/Val-62 samples and were 7.45 and 7.91 nm for both FH His-402/Val-62 samples. The R_{XS-1} values were 2.35 and 2.36 nm for FH Tyr-402 and 2.15 and 2.25 nm for FH His-402. The R_{XS-2} values were 1.77 and 1.76 nm for FH Tyr-402 and 1.77 and 1.75 nm for FH His-402. Because the R_g , R_{XS-1} , and R_{XS-2} parameters at 0.4 mg/ml were similar for both allotypes, no conformational differences between the allotypes were detectable.

Both FH Tyr-402 and FH His-402 showed concentration dependences for $I(0)/c$, R_g , and R_{XS-1} but not for R_{XS-2} (Fig. 7, D–G). These concentration dependences showed self-association for both FH allotypes, the extent of which was similar to WT FH. The $I(0)/c$ value is proportional to molecular mass. Assuming the slopes in Fig. 7D corresponded to a monomer–dimer equilibrium, the K_D values were estimated to be 16 ± 4 and $18 \pm 2 \mu\text{M}$ for the two FH Tyr-402 samples and 11 ± 3 and $22 \pm 8 \mu\text{M}$ for the two FH His-402 samples. These K_D values were comparable with that of 28 μM from AUC (29) and those of 7–8 μM from SPR (Fig. 1). The extrapolation of the R_g values to zero concentration, in order to represent the monomer, gave values of 7.22–7.77 nm (Fig. 7E; Table 3). These values were lower than that of 8.90 ± 0.19 nm for WT FH (29); this reduction was attributed to reduced oligomer formation in improved homozygous FH preparations. The lower R_g value of 7.22–7.77 nm indicated that FH was even more folded-back in its SCR domain arrangement than previously thought. The R_{XS-1} values at zero concentration were 2.02–2.27 nm, which were lower when compared with the previous R_{XS-1} value of 2.51 ± 0.06 nm (29). The mean R_{XS-2} value was 1.75–1.77 nm, being similar to the previous R_{XS-2} value of 1.79 ± 0.01 nm (29).

The distance distribution function $P(r)$ reports the distances between all pairs of atoms within FH (Fig. 7C). The $P(r)$ curves gave another calculation of the R_g and $I(0)$ values for FH Tyr-402 and FH His-402 (Table 3). At 0.4 mg/ml, the Guinier and $P(r)$ R_g values showed good agreement. The maximum dimension L of FH resulted from the intercept of the $P(r)$ curves with zero at large r . The r value of the maximum M in the $P(r)$ curve gives the most commonly occurring distance within FH. The two FH allotypes showed similar concentration dependences in their $P(r)$ curves and L values (Fig. 7C), indicating similar solution structures for the two allotypes. For FH Tyr-402, a concentration-independent peak $M1$ was seen at $r = 5.3 \pm 0.7$ nm, and a concentration-dependent second peak $M2$ was seen at $r = 11$ nm at 3.3 mg/ml. For FH His-402, $M1$ was unchanged at 4.9 ± 0.4 nm, and $M2$ was also 11 nm at 3.3 mg/ml. The concentration dependence of $M2$ reflected FH oligomer formation. At 0.4 mg/ml, the L values were 28–29 nm for FH Tyr-402, and 26–28 nm for FH His-402 (Table 3), but were increased at higher concentrations (Fig. 7C). These L values were lower than those previously reported (29), and this is attributable to a reduced FH oligomerization at lower concentrations.

In normalized Kratky analyses of $(Q \cdot R_g)^2 \cdot I(Q)/I(0)$ versus $Q \cdot R_g$ for the four FH experimental curves (62), a clear peak was obtained in all four cases that tailed off at large $Q \cdot R_g$ values (data shown below). These are representative of extended multidomain proteins tethered by linkers of varying conformations.

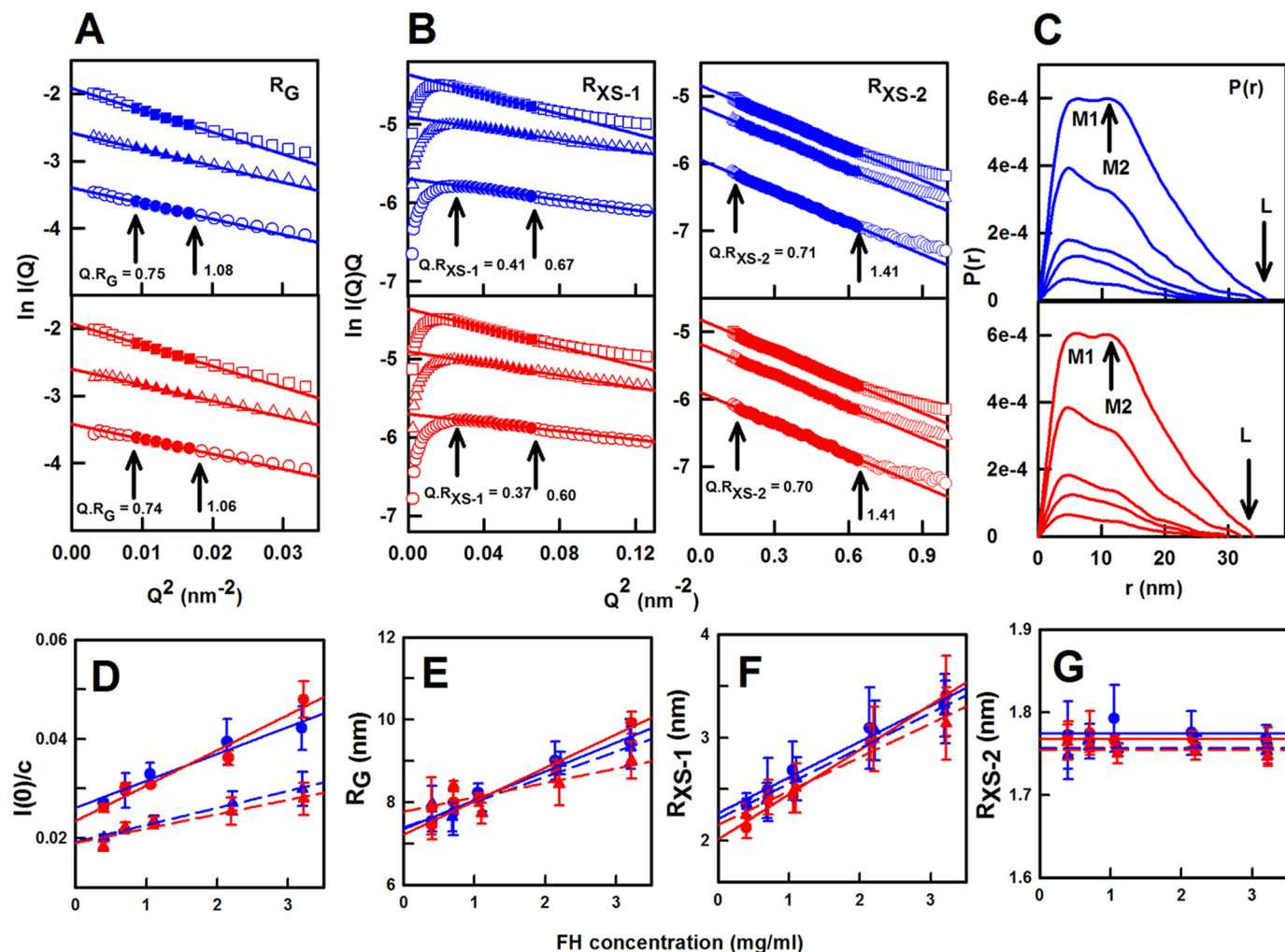


Figure 7. X-ray scattering analyses of FH Tyr-402 and FH His-402. In the R_g and R_{XS} Guinier analyses, the filled symbols correspond to the $I(Q)$ data used to determine the R_g or R_{XS} values, and the straight lines correspond to the best fit through those points. The $Q \cdot R_g$ fit ranges are arrowed. A, R_g plots of $\ln I(Q)$ versus Q^2 at low Q values for FH Tyr-402 (blue) and FH His-402/Val-62 (red) at concentrations of 1.1 mg/ml (\circ), 2.2 mg/ml (\triangle), and 3.2 mg/ml (\square). The Q fit range was $0.09 - 0.13 \text{ nm}^{-1}$. B, corresponding cross-sectional R_{XS-1} and R_{XS-2} fits of $\ln I(Q)$. Q versus Q^2 values are shown using Q fit ranges of $0.16 - 0.26$ and $0.4 - 0.8 \text{ nm}^{-1}$, respectively. C, distance distribution function $P(r)$ analyses are shown at concentrations of 0.4, 0.7, 1.1, 2.2, and 3.3 mg/ml. D–G, concentration dependences of the $I(0)/c$, R_g , R_{XS-1} , and R_{XS-2} parameters are shown for FH Tyr-402 and FH Tyr-402/Val-62 (blue; \bullet and \blacktriangle) and FH His-402/Val-62 (red; \bullet and \blacktriangle). Each value was measured in quadruplicate and then averaged and fitted by linear regression except in G when the mean was shown. Statistical error bars are shown where visible. D, $I(0)/c$ intensities were measured in two beam sessions, resulting in two different pairs of lines. E, the R_g values at zero concentration were 7.39 ± 0.25 and $7.35 \pm 0.13 \text{ nm}$ (FH Tyr-402) and 7.22 ± 0.15 and $7.77 \pm 0.27 \text{ nm}$ (FH His-402) (Table 3). F, R_{XS-1} values at zero concentration were 2.21 ± 0.06 and $2.27 \pm 0.06 \text{ nm}$ (FH Tyr-402), and 2.02 ± 0.06 and $2.15 \pm 0.06 \text{ nm}$ (FH His-402). G, averaged R_{XS-2} values were 1.77 and 1.77 nm (FH Tyr-402) and 1.76 and 1.75 nm (FH His-402).

The Tyr-402 and His-402 allotypes showed very similar peaks at $Q \cdot R_g$ values of 4.22 and 4.28 respectively, indicating that the two allotypes exhibited similar extended flexible structures.

FH models for the Tyr-402 and His-402 allotypes

The atomistic modeling of the two FH Tyr-402 and two His-402 scattering curves was initiated using 17 high-resolution SCR structures, three SCR homology models, and eight glycan chains to construct an energy-minimized initial FH model (Fig. 1 and Tables 1 and 2; see under “Materials and methods”). By varying the torsion angles at eight inter-SCR linkers, 510,000 FH models were created in four Monte Carlo simulations, from which 29,715 models were retained for analysis as being physically realistic without steric clashes. Comparison of the four extrapolated experimental scattering curves with the 29,715

theoretical curves gave an R -factor versus R_g distribution with a minimum close to the experimental extrapolated R_g value (Fig. 8). An R_g filter of $\pm 5\%$ of the experimental value and an R -factor filter of $> 5\%$ gave 1240/2678 and 1219/310 best-fit models for the two Tyr-402 and two His-402 allotypes, respectively; these are 1–9% of the 29,715 trial models (blue/red, Fig. 8). The best-fit 100 FH models were selected on the basis of the lowest R -factors (orange, Fig. 8) of 2.3–3.2 and 2.3–3.6%, respectively (Table 3).

The $s_{20,w}^0$ values of the best-fit 100 FH models were calculated using HYDROPRO to be 5.35–5.77 and 5.21–5.61 S and for the centroid models to be 5.49–5.52 and 5.44–5.51 S, both for FH Tyr-402 and FH His-402, respectively (Table 3). These values agreed well with the experimental values of 5.66–5.69 S (Table 3) and confirmed the outcome of the atomistic modeling, given that the mean difference between the modeled and

Table 2
Homology modeling of the SCR-9, SCR-14, and SCR-17 domains

Target domain	Template domain	PDB code	Residues	Maximum score	Total	Query cover	E value	Residue identities
SCR-9	SCR-4	2UWN	148–186	40.1	41.2	%	0.0004	%
SCR-14	SCR-15	3ZD1	5–59	37.5	37.5	68	0.002	44
SCR-17	SCR-18	3SW0	3–59	42.2	77.6	100	0.00007	36
						100		37

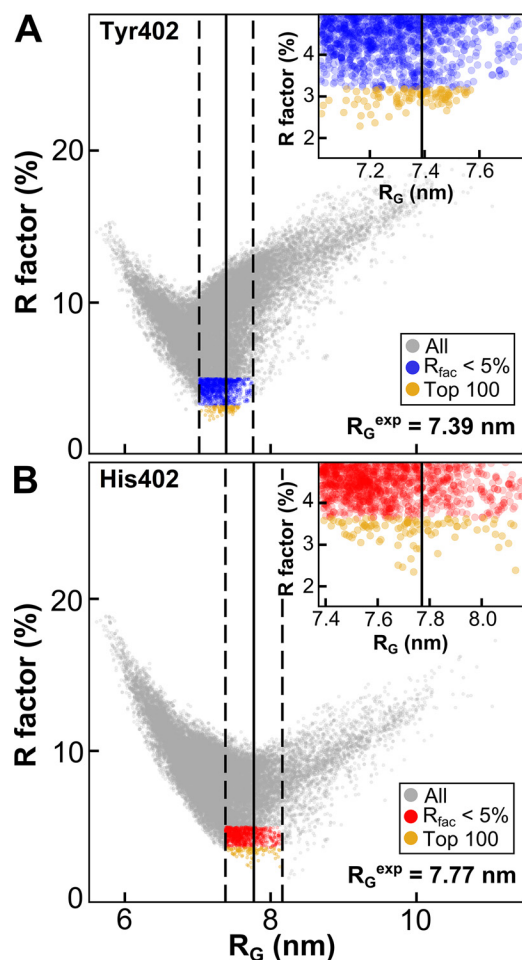


Figure 8. Atomistic modeling searches for the FH solution structure. The 29,715 FH models were fitted to one experimental scattering curve for each of homozygous FH Tyr-402/Val-62 (A) and FH His-402/Val-62 (B), each extrapolated to zero concentration. Gray corresponds to all FH models. The blue/red subsets corresponds to FH models that passed two filters, namely R_g within $\pm 5\%$ of the experimental R_g and an R -factor of $< 5\%$. Orange corresponds to the 100 best-fit FH models (see inset). The vertical black lines represent the experimental R_g for FH Tyr-402 (7.39 nm) and FH His-402 (7.77 nm). The dashed lines represent the $\pm 5\%$ upper and lower boundaries of these R_g values.

experimental values should be ± 0.21 S for related macromolecules, including antibodies (63). Like antibodies that are also considered to be flexible molecules, HYDROPRO was effective in calculating the $s_{20,w}^0$ values; this was attributed to the use of the averaged best-fit solution structures. The calculations were much improved to those of the 2009 FH modeling, which gave 4.91–5.01 S from older scattering models that in retrospect were too elongated (26).

NT-COM and CT-COM separation distributions

The best-fit SCR arrangements within the 29,715 trial FH models were shown to be bimodal using histograms of their

N-terminal α -carbon (NT) and C-terminal α -carbon (CT) to COM separations (Fig. 9). Longer separations corresponded to extended SCR arrangements in FH; shorter separations corresponded to a folded-back bent one.

(i) For all 29,715 models, the NT-COM and CT-COM separations corresponded to similar ranges of 0.5–21.8 and 0.8–22.4 nm, respectively (Fig. 9, A and D). However, the most populated NT-COM separations were short ones (3–5 nm) that tailed off after > 16 nm. In distinction, the most populated CT-COM separations were long ones (14–16 nm) followed by short ones (~ 5 nm) (Fig. 9, A and D). The NT-COM separations were right-skewed with a skewness coefficient of 0.53, whereas the CT-COM separations were left-skewed and bimodal with a skewness coefficient of -0.55 . Overall, the separation distributions were uneven and not normally distributed. This unexpected outcome was attributed to the seven glycans in SCR-12/18 that perturbed the random generation of physically-realistic FH structures (Fig. 1) and the asymmetric distribution of the longest inter-SCR linkers in FH (26).

(ii) For the 1219–1240 filtered FH models, both the NT-COM and CT-COM distance distributions corresponded to small subsets of those for all 29,715 models (Fig. 9, B and E). Density plots showed that these models occupied limited conformations when compared with all 29,715 models (insets, Fig. 9, B and E). The separations for both allotypes were bimodal, being left-skewed for the NT-COM distances and right-skewed for the CT-COM distances. Overall, no differences were seen between the two allotypes.

(iii) For the two sets of 100 best-fit FH models, the two NT-COM and CT-COM distance distributions were again largely bimodal distributions (Fig. 9, C and F). Similar bimodal results were obtained for the fits with the other pair of FH Tyr-402 and FH His-402 experimental curves (distributions not shown).

For the filtered and top 100 models, statistical analyses were performed to confirm that the separation distance distributions (Fig. 9) were unaffected by sampling bias. The best-fit 100 models were not necessarily a converged subset of the filtered models as seen in other modeling (64). A nonparametric (*i.e.* no underlying assumption) Kolmogorov-Smirnov test (data not shown) was thus used to indicate that the best-fit distances for NT-COM and CT-COM were either short or long and were independent of any sampling bias observed in all 29,715 models. This suggested that FH existed in two conformations.

The full FH SCR domain arrangement was examined using pairs of NT-COM and CT-COM distances. The 29,715 separations were nonrandomly distributed between 1 and 25 nm, showing that a broad range of FH conformations had been generated as desired (Fig. 10, A and B). The most sampled best-fit NT-COM and CT-COM separations occurred at 1–10 and 12–17 nm, respectively, at the top left-hand corner. The second

Table 3
 Experimental X-ray and analytical ultracentrifugation data for FH Tyr-402/Val-62 and His-402/Val-62 and their modeling fits

	Filter	Models	R_g	R_{XS-1}	R_{XS-2}	L	R factor	$s_{20,w}^0$
			nm	nm	nm	nm	%	S
Experimental data								
FH Tyr-402/Val-62			7.39 ± 0.25	2.21 ± 0.06	1.77 ± 0.01	28–29		5.66 ± 0.08
			7.35 ± 0.13	2.27 ± 0.06	1.77 ± 0.01			
			7.61 ± 0.01 ^a					
FH His-402/Val-62			7.77 ± 0.27	2.02 ± 0.06	1.76 ± 0.01	26–28		5.69 ± 0.02
			7.22 ± 0.15	2.15 ± 0.06	1.75 ± 0.01			
			7.54 ± 0.01 ^a					
Atomistic modeling								
Library of 29,715 FH models	None	29,715	5.61–11.08	NA ^b	NA	NA	2.8–25.3	NA
FH Tyr-402/Val-62 (sample 1)	R_g and R -factor	1240	7.02–7.75	NA	NA	NA	2.3–4.9	NA
	Best fit	100	7.02–7.57	2.34–2.98	1.85–2.10	NA	2.3–3.2	5.35–5.77
	PCA group 2 (NT out)	49	7.34–7.57	2.34–2.62	1.95–2.10	NA	2.7–3.2	5.38–5.77
	Centroid	1	7.41	2.50	2.03	NA	3.1	5.52
FH Tyr-402/Val-62 ^c (sample 2; data not shown)	R_g and R -factor	2,678	6.98–7.72	NA	NA	NA	1.9–4.9	NA
	Best fit	100	7.03–7.47	2.71–3.05	1.81–2.04	NA	1.9–2.7	5.33–5.70
	PCA group 14 (NT in)	48	7.11–7.43	2.71–2.98	1.81–1.99	NA	1.9–2.7	5.36–5.70
	Centroid	1	7.32	2.84	1.95	NA	2.2	5.49
FH His-402/Val-62 (sample 1)	R_g and R -factor	1,219	7.38–8.16	NA	NA	NA	2.3–4.9	NA
	Best fit	100	7.39–8.13	2.32–3.15	1.48–1.97	NA	2.4–3.6	5.21–5.61
	PCA group 5 (NT in)	79	7.39–7.95	2.67–3.07	1.64–1.94	NA	2.5–3.6	5.21–5.55
	Centroid	1	7.48	2.92	1.87	NA	3.2	5.44
FH His-402/Val-62 (sample 2; data not shown)	R_g and R -factor	310	6.86–7.58	NA	NA	NA	2.7–4.9	NA
	Best fit	100	6.86–7.57	2.09–2.93	1.82–2.10	NA	2.7–4.0	5.38–5.88
	PCA group 19 (NT out)	84	6.86–7.57	2.28–2.82	1.89–2.10	NA	2.7–4.0	5.38–5.78
	Centroid	1	7.40	2.48	2.03	NA	3.8	5.51

^a The R_g , R_{XS} , and L values correspond to the SAXS curve extrapolated to zero concentration. The two R_g values correspond to the Guinier fits; the third corresponds to the $P(r)$ analyses.

^b NA is not available.

^c In this instance I62V was heterozygous; the other three were homozygous for Val-62.

most sampled best-fit separations occurred as a broad cluster at 5–17 and 2–8 nm, respectively. For the filtered Tyr-402 models, two structural regions were visible as two dense clusters in *blue*, in which the top 100 Tyr-402 models were shown in *orange* (Fig. 10A). For the filtered His-402 models, the same two clusters were seen in *red* and *orange*, respectively (Fig. 10B). In contrast, FH models with long NT-COM and CT-COM separations (>15 nm) showed large radii of gyration and poor fits with R -factors >5% (*gray*, Fig. 10, A and B). This analysis showed that FH either has a bent-inwards N-terminal region and an extended C-terminal region or vice versa.

Best-fit conformations for FH Tyr-402 and His-402

After superimposition of the top 100 Tyr-402 and His-402 FH structures, visual inspection showed multiple conformers. To understand these conformers, they were clustered into conformational families using principal component analysis (PCA) (Fig. 11) (65). PCA determines the correlated motions of protein residues as linearly uncorrelated variables termed principal components (66). These “essential motions” are extracted from a covariance matrix of the atomic coordinates of the frames in the trajectory. The eigenvectors of this matrix each have an associated eigenvalue that characterizes a mode of the motion or variance.

The PCA analyses confirmed that two families of FH structures with either a bent-inwards N-terminal region and an extended C-terminal region, or vice versa, were seen. Four PCA groups accounted for 98.5 and 86.9% of the variances in the top 100 models of each Tyr-402 and His-402, respectively (Fig. 11, C and G). For each PCA group, the combined NT-COM and CT-COM separation distances were plotted (Fig. 11, D and H). The conformational density of each group was represented as

either *blue* (NT) or *red* (CT) density, with the centroid FH model depicted as a *black cartoon* (Fig. 12). The 60 Tyr-402 models in PCA groups 2–4 corresponded to an extended N terminus and a bent inwards C terminus (*red*, *green*, and *blue*, Figs. 11D and 12, B–D). The 40 Tyr-402 models in PCA group 1 corresponded to a bent inwards N terminus and an extended C terminus (*black*, Figs. 11D and 12A). The 79 His-402 models in PCA group 5 corresponded to a bent inwards N terminus and an extended C terminus (*black*, Figs. 11H and 12E). The 21 His-402 models in PCA groups 6–8 corresponded to a bent inwards C terminus and an extended N terminus (*red*, *green*, and *blue*, Figs. 11H and 12, F–H).

Each of the PCA groups 1–8 gave FH structures whose theoretical scattering curves $I(Q)$ fitted well with the experimental Tyr-402 or His-402 curves with low R -factors between 2.4 and 3.8% (Fig. 12). The visual agreement was good to at least $Q = 1.0 \text{ nm}^{-1}$. The centroid models with an extended N terminus and a bent inwards C terminus (Fig. 12, B–D, and F) gave fits similar to those with a bent inwards N terminus and an extended C terminus (Fig. 12, A, E, G, and H). The theoretical distance distribution curves $P(r)$ also showed good agreements with the experiment. Thus, the eight centroid FH models gave very similar MI peak values that agreed with the experimental $P(r)$ curves, where MI was 4.7–5.0 nm (Fig. 7C). Their intensities were comparable, although these showed oscillations around the experimental $P(r)$ curves. Of note, the experimental $P(r)$ curves generally appeared smoother than the theoretical curves. This is likely due to both the lower resolution and the fluctuation between the alternative conformations of the solution structure studied by the experimental curves. Only eight of the 200 best-fit structures had conformations in which both the N- and C-terminal regions were inwardly bent, and none

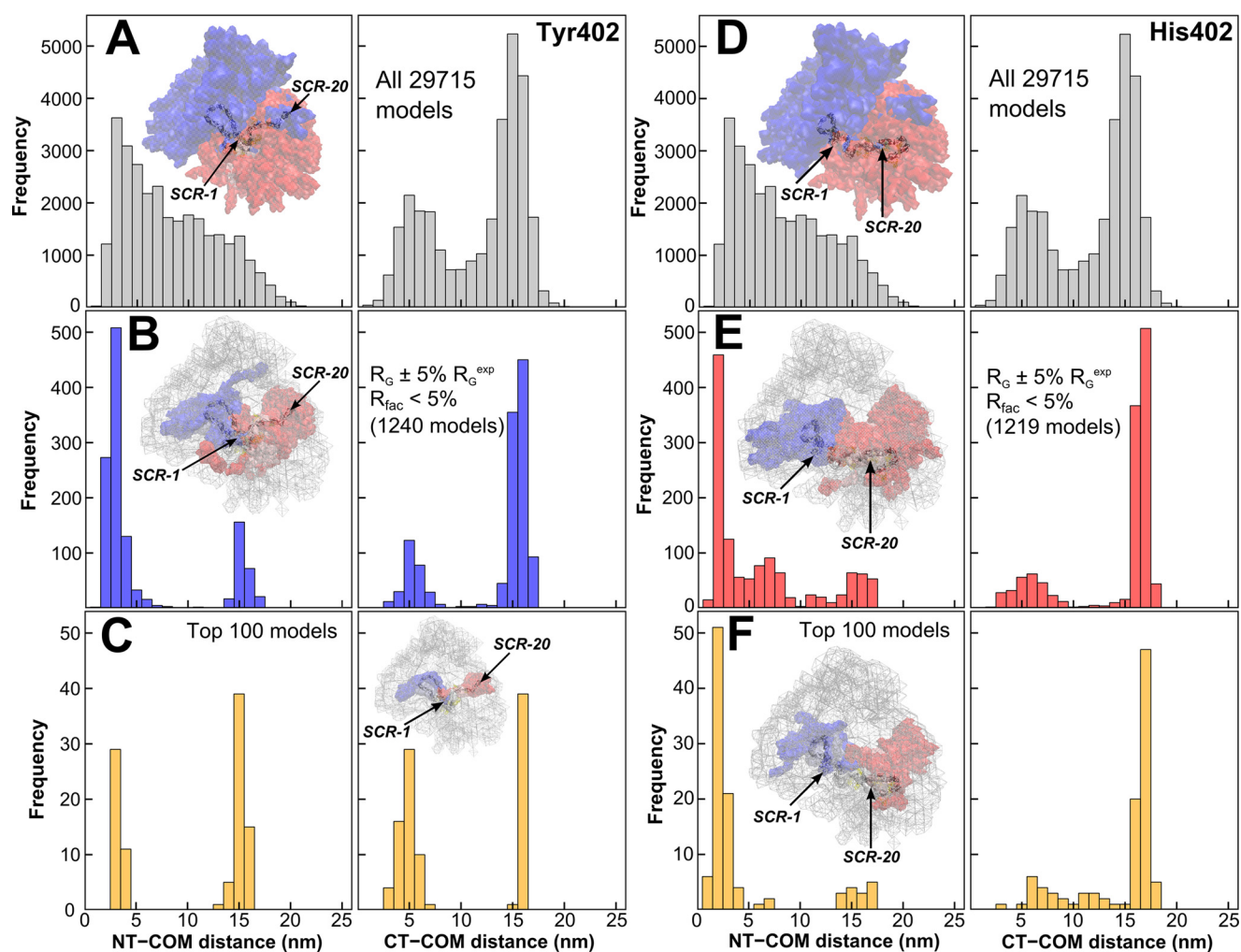


Figure 9. Center-of-mass separation frequencies in the FH Tyr-402 and FH His-402 models. *A* and *D*, top two panels (gray) show the separations (NT-COM and CT-COM) between the N-terminal α -carbon (NT) and C-terminal α -carbon (CT) to the center of mass (COM) for all 29,715 FH models. *B* and *E*, separations are shown for the filtered 1240 and 1219 FH models (Fig. 8). *C* and *F*, separations are shown for the 100 best-fit FH models. For each set of FH models, the density plot shows the N-terminal half in blue, the C-terminal half in red, and the best-fit model as a cartoon.

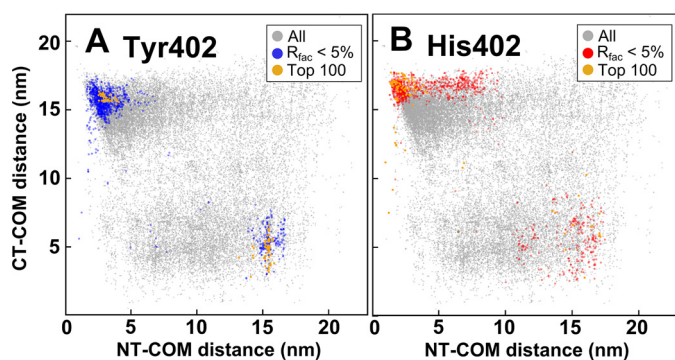


Figure 10. Separation densities in the FH Tyr-402 and FH His-402 models. For each of the Tyr-402 (*A* and *B*), joint NT-COM and CT-COM separations for the FH Tyr-402 and FH His-402 models are compared with each other. Blue/red corresponds to the filtered FH models, and orange corresponds to the 100 best-fit FH models (Fig. 8).

showed that both regions were extended in any structure. The same PCA analyses for the other pair of available Tyr-402 and His-402 curves gave similar structural outcomes (data not shown).

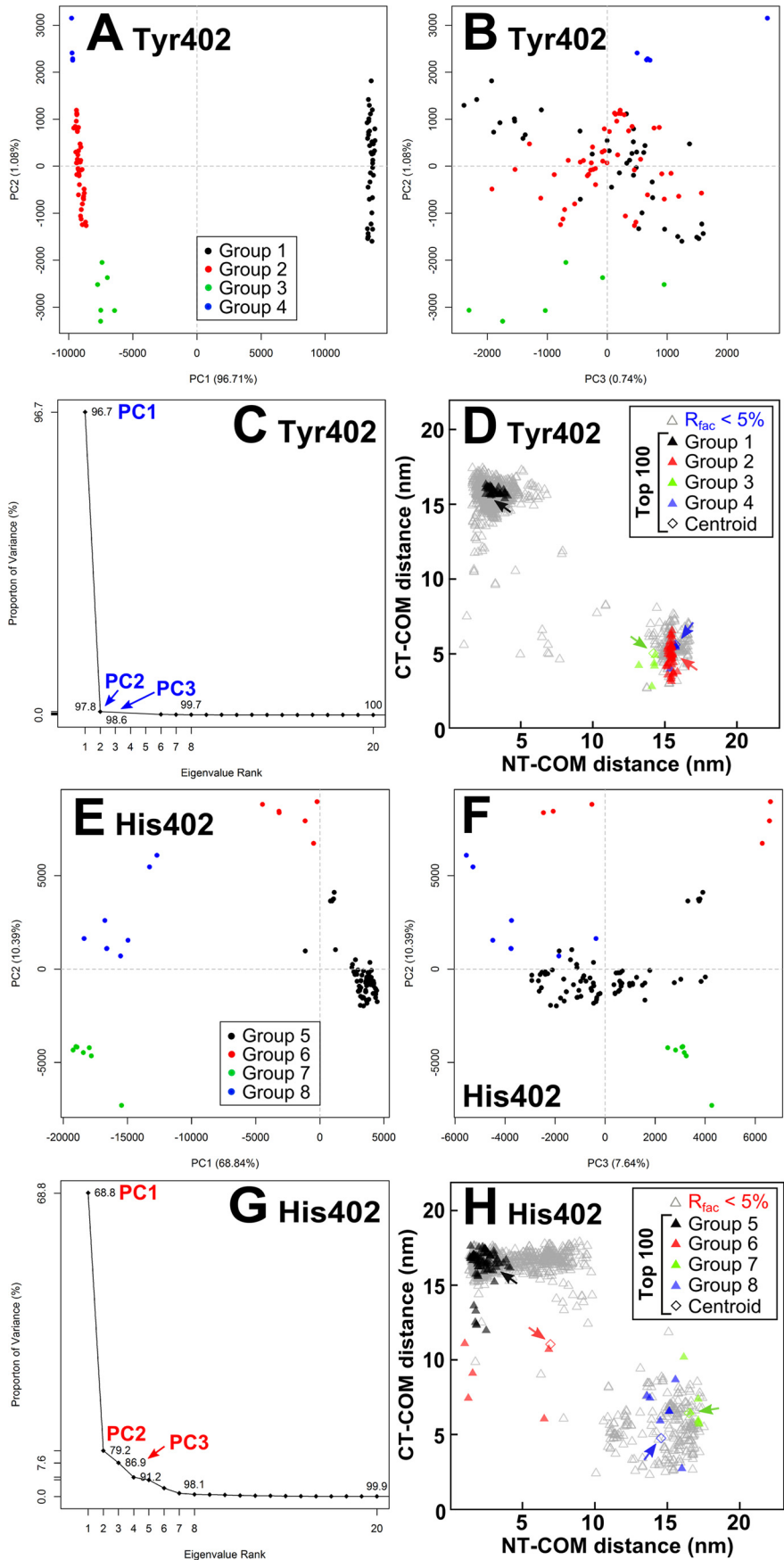
The normalized Kratky analyses for six modeled best-fit curves corresponding to the NT-in and NT-out centroid con-

formations from Fig. 12 each showed a clear peak that was similar in shape to the experimental curves (Fig. 13). The $Q \cdot R_g$ range of the modeled peaks were 3.90–4.12 and 3.33–4.30, respectively, in good accord with the observed values of 4.22–4.48. This similarity showed that the modeling had mostly replicated the extended flexible multidomain protein structures. The reduced $Q \cdot R_g$ peak positions for some of the models may indicate an insufficient representation of flexible extended linkers in FH. Overall, this comparison validated our analyses that FH is a multidomain protein whose domains were connected by semi-extended linkers (62).

Discussion

FH is an essential regulatory glycoprotein that protects host cells from complement activation both at the cell surface and in the fluid phase. In this study, the most accurate molecular models to date have been produced for full-length FH, and these suggest that FH exists in two distinct conformations that enable its complement functions. Common and rare genetic variants in FH are associated with major and rare inflammatory diseases, including AMD, Alzheimer's, aHUS, and C3G (34). For AMD, the Y402H and V62I polymorphisms are associated with

Solution structures of FH Tyr-402 and His-402



increased and decreased risk, respectively. To perform its regulatory activities, FH binds to multiple physiological ligands, including to itself, with weak K_D values in the micromolar range, and has a highly glycosylated 20-domain structure (Fig. 1). For FH, characterization of its self-association is important for determining its solution conformation and investigating the development of drusen in AMD. Up to now, the self-association of FH has only been characterized for native heterozygous FH purified from plasma, and this has not been widely accepted. Likewise, the only FH molecular structures up to now had been determined by preliminary modeling of X-ray scattering curves, resulting in two different folded-back structures whose significance was unclear (26, 51, 67). Here, working with genotyped FH, we have now shown using four independent methods that WT FH Tyr-402 and risk-associated FH His-402 both significantly self-associate. The V62I polymorphism involves a minor structural rearrangement within the core of SCR-1 and is protective for AMD due to its subtly better capacity to bind C3b, inhibit proconvertase formation, and catalyze inactivation of fluid-phase and surface-bound C3b (45, 56). For both polymorphisms, it was important to demonstrate that neither affected FH oligomerization by self-association. More importantly, by working with genotyped FH samples with reduced sample aggregation, and also by utilizing a much-improved atomistic modeling procedure based on molecular dynamics and Monte Carlo procedures and statistical assessments of the outcome, we have determined similar detailed molecular models for both allotypes of full-length FH, including its eight FH glycans. Our statistical validation of two alternative FH conformations significantly improves our understanding of ligand binding to FH and how this may be perturbed in its function.

Self-association of FH

FH self-association is important for FH function because this will increase the local concentration of its ligand-binding sites. FH oligomer formation in solution was first observed by X-ray scattering (28). Another FH study at that time used EM *in vacuo* and gel filtration to deduce that FH was monomeric; however, the low concentrations in use in that study would not have favored dimer formation, and gel filtration is insensitive to FH self-association (Fig. 2) (33). The FH fragments SCR-6/8 and SCR-16/20 each formed dimers (26, 29–31). This outcome predicted that full-length FH would form weak daisy-chained multimers through these two independent dimerization sites. These multimers were indeed observed for heterozygous FH previously by AUC as a series of small peaks at large S values, in which the sedimentation equilibrium fits suggest that FH dimers and trimers are initially formed reversibly below 1 mg/ml (26, 29). Higher FH concentrations favor the formation of larger oligomers by further daisy-chaining to result in irreversible oligomers that can be removed by size-exclusion chromatography. Here, these small peaks were now observed again

for homozygous FH at 1.1 mg/ml (Fig. 3C). Biologically, given that the plasma concentration of FH is around 0.7 mg/ml (19), reversible dimers and trimers are expected to form *in vivo*. The resulting clustering of FH on cell surfaces, when bound to surface markers and C3b/C3d, is likely to enable broader surface protection, thus being functionally significant (68). However, the accumulation of FH by clustering through self-association may also relate to the formation of soft drusen in AMD via chronic inflammation, in particular if FH oligomer formation becomes irreversible (69). Consistent with this, studies with full-length FH showed that its affinity for both heparin and HS was not altered by the Y402H polymorphism (39, 70, 71). It was also found that FH multimer formation is enhanced in the presence of zinc, which inhibits FH function (57). In this context, the self-association of FH Tyr-402 and FH His-402 becomes of interest. In this study, AUC showed that 11.7–12.8% multimers were present for FH His-402 and FH Tyr-402 (Fig. 3D).

Because the occurrence of FH self-association had been questioned (33, 72), other orthogonal techniques were utilized in our study to confirm this. MS confirmed the existence of as much as 23–28% FH dimer and trimer for both FH allotypes (Fig. 4). SPR revealed self-association for both full-length FH allotypes and for both SCR-6/8 allotypes (Fig. 5). SAXS analyses of the Guinier $I(0)/c$, R_g , and R_{XS-1} parameters showed clear concentration effects attributable to self-association (Fig. 7, D–F). One immediate consequence of this outcome was the need to allow for self-association effects in the SAXS structural studies to model the monomeric FH structure. It was necessary to extrapolate the scattering curves to zero concentration. FH self-association has had a marked effect on earlier SAXS data collection. The first FH studies gave an R_g value of 12.4 nm in 1991 (28), which was reduced to 11.4 nm in 2001 (50), then to an extrapolated value of 8.90 nm in 2008 (26, 29), and now to an extrapolated value of 7.22–7.77 nm (Fig. 7E). This reduction in R_g is attributed to reduced aggregation in FH because of improved protein handling (e.g. FH forms aggregates in storage or when in transit); its occurrence may have caused the lack of progress with the protein crystallization of FH to date.

Atomistic modeling of two FH conformations

The new molecular modeling of the FH scattering curves has benefitted from five improvements (54) as follows: (i) the application of molecular dynamics to generate an initial full FH structure; (ii) the use of 17 high-resolution SCR structures, in distinction to up to 11–14 SCR structures previously; (iii) the inclusion of eight energy-refined glycan chains in FH; (iv) the use of rapid Monte Carlo simulations with the inter-SCR linkers to generate a large number of 29,715 physically realistic trial FH structures; and (v) statistical analyses of the resulting ensemble of best-fit structures. Interestingly, our previous two 2008 modeling studies had yielded alternative folded-back FH structures with either the N- or C-terminal SCR domains in an

Figure 11. Principal component analyses of the 100 best-fit models for FH Tyr-402 and FH His-402. For FH Tyr-402 (A and B) and FH His-402 (E and F), the 100 best-fit models were grouped by PCA into four groups, 1–4 (black, red, green, and blue, respectively), as exemplified by the first three principal components (PC2 versus PC1 and PC3 versus PC2). C and G, first three eigenvalue rankings (PC1 to PC3) accounted for variances of 98.5 and 86.9% in the 100 best-fit FH models. D and H, black, red, green, and blue triangles correspond to the PCA groups 1–4 and 5–8 in each set of best-fit 100 models. The four arrowed diamonds correspond to the average (centroid) model for each PCA group.

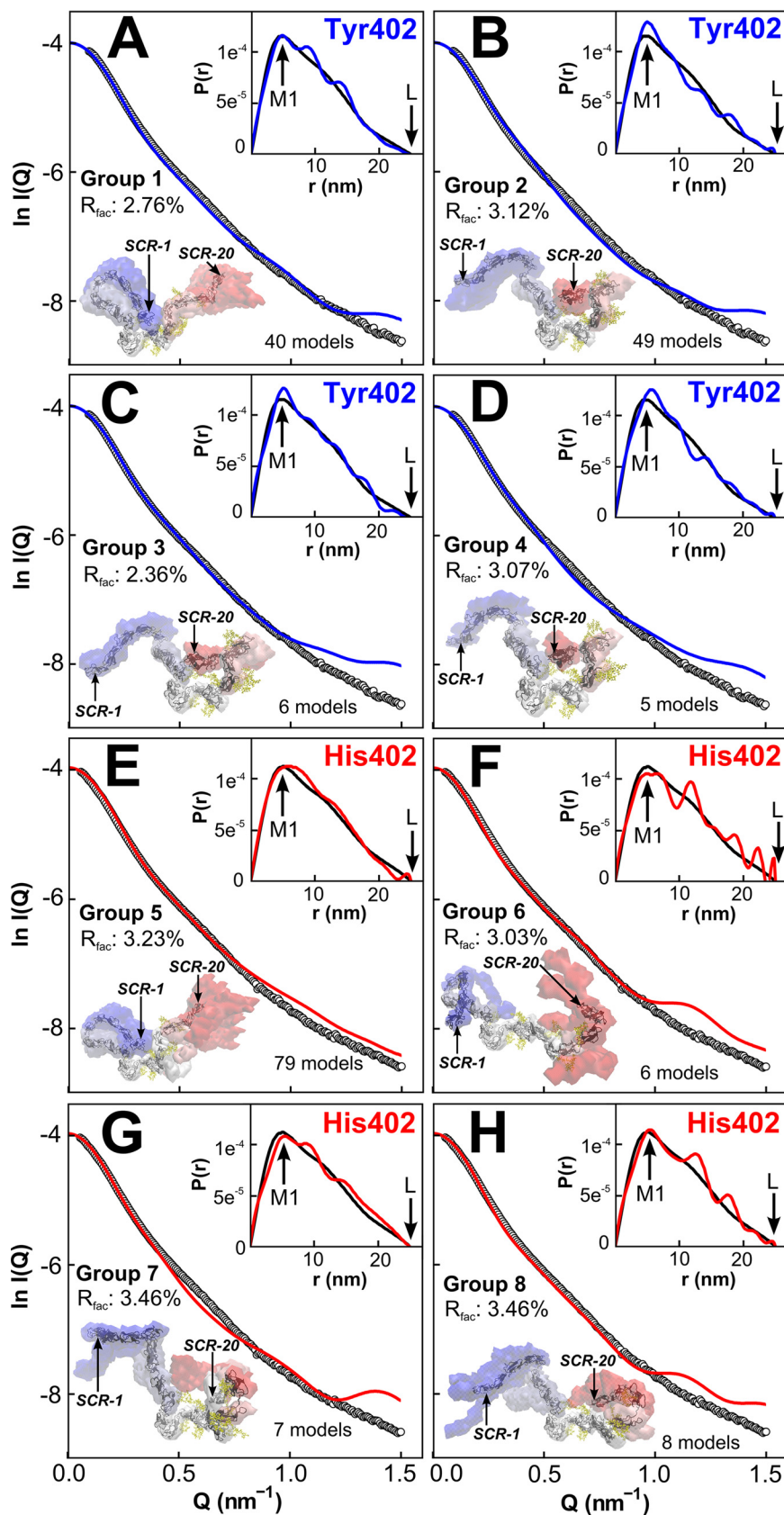


Figure 12. Scattering curve fits for the centroid PCA models for FH. For FH Tyr-402 (A–D) and FH His-402 (E–H), the four panels display the centroid FH model for the PCA groups 1–4 and 5–8, respectively, in density plots. The number of FH models in each PCA group is displayed. The experimental curve is denoted by black circles, and the theoretical curve is denoted as solid blue/red lines. The inset shows the $P(r)$ curves for the experimental data (black line) and modeled curve (blue/red line). For each PCA group, the conformational space is represented as blue (NT), white or red (CT) density. The average (centroid) model of each PCA group is depicted as a black cartoon in which SCR-1 and SCR-20 are arrowed.

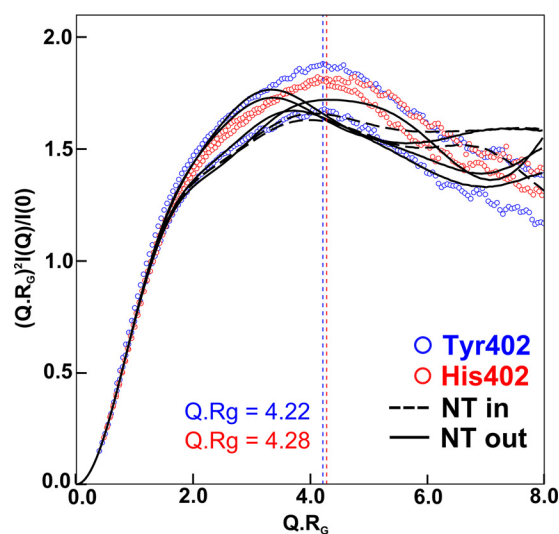


Figure 13. Normalized Kratky plots for the experimental and best-fit FH curves. The four experimental curves are shown in blue and red for the Tyr-402 and His-402 allotypes, respectively, which peak at $Q \cdot R_g$ values of 4.22 and 4.28, respectively. For both allotypes, the theoretical curves for four N-terminal extended and two C-terminal extended FH models are shown as black solid and dashed lines, respectively. These correspond to the two sets of centroid models of Fig. 12, B, C, D, and H and Fig. 12, A and E, respectively.

extended conformation (26, 51, 67), yet their significance was unclear. Here, the statistical analysis of the 29,715 curve fits showed that both alternative FH conformations were valid outcomes from the SAXS modeling (Fig. 14, A and B). In the 200 best-fit structures, some 81 and 119 structures, respectively, corresponded to (i) an extended N terminus and a bent inwards C terminus, and (ii) a bent inwards N terminus and an extended C terminus. In principal, either of these two FH structures may be the actual solution structure, or both FH structures may co-exist as an equilibrium in solution. Consideration of C3b and C3d binding (below) suggested both FH conformations exist. Further statistical analyses of the 200 best-fit structures showed that the two FH Tyr-402 and His-402 allotypes displayed the same solution structures.

Biological significance of the FH Tyr-402 and His-402 models

To summarize FH activity, SCR-1/5 is involved in both cofactor and decay-acceleration activities (73, 74), whereas SCR-19/20 is involved in the recognition of host cell surfaces (75). These two functions are mediated by N-terminal SCR-1/4 binding to both C3b and FI and C-terminal SCR-19/20 binding to C3d, heparin, and sialic acid. C3dg and C3d, being two fragments of C3 that each contain the thioester domain for host-surface binding, are formed by the breakdown of C3b by FI and/or serum proteases, and both bind to complement receptor 2 to trigger immune complex clearance and to stimulate the immune response (1). C3d is also an autologous helper T-cell target (76). During complement regulation at host cell surfaces, an increase in the number of C3d-binding sites for SCR-19/20 would enable more FH to be recruited at the host cell surface for a more rapid breakdown of C3b (77). For cell-surface attachment of FH, an optimal combination of affinities for each of C3b and polyanion ligand is essential for C-terminal FH binding (68), for which both SCR-7 and SCR-19/20 bind bivalently via heparin (21, 27).

The functions of FH were clarified by superimposing the two alternative best-fit conformations of FH with crystal structures for the SCR-1–4–C3b complex and the SCR-19–20–C3dg complex (PDB codes 2WII and 5NBQ) (52, 78).

(i) The N-terminal extended FH conformation accounted for binding to C3b at SCR-1/4, and to heparin and sialic acid, but less so for C3dg. The binding of C3b to SCR-1/4 showed minimal steric clashes with the rest of FH (black arrow, Fig. 14A). Likewise, for binding to fluid phase FI, the SCR-1/3 domains were accessible, including the FH residues 101, 106, 112, and 119 that were important for this interaction (52). For bivalent SCR-7 and SCR-19/20 binding to heparin and/or sialic acid on cell surfaces, the essential residues in each of SCR-7 (Tyr/His-402) and SCR-19/20 (FH residues 1181–1183, 1191, 1195, 1196, 1198, and 1199) were solvent-accessible. However, C3dg binding to SCR-19 resulted in significant steric clashes with SCR-9, SCR-12/15, and the glycans (black arrows, Fig. 14A), implying that some structural rearrangement will be required for C3dg binding.

(ii) In contrast, the C-terminal extended FH model accounted for binding to C3dg at SCR-19 or SCR-20 and to heparin and/or sialic acid on cell surfaces but not to C3b at SCR-1/4. The binding of C3b to SCR-1/4 resulted in large steric clashes with the SCR-6/8 domains (black arrows, Fig. 14B). The shortest distance between the SCR-1/4 and SCR-6/9 domains was ~ 6 nm between SCR-1 and SCR-9. For binding to fluid phase FI, the SCR-1/3 domains were accessible. The accessibility of the Tyr/His-402 residue in SCR-7 was not compromised. For bivalent binding to heparin and/or sialic acid on cell surfaces, each of SCR-7 and the essential residues in SCR-19/20 (above) were accessible. For C3dg binding to SCR-19 or SCR-20 (79), no steric clash was seen (Fig. 14B).

In summary, the two extended N-terminal and extended C-terminal FH models (Fig. 14, A and B) showed that SCR-7 and SCR-19/20 were spatially aligned to permit simultaneous bivalent FH binding to glycosaminoglycans at the host cell surface. However, in order for FH to bind C3b via SCR-1/4, potential steric hindrance from SCR-6/8 would need to be removed, most likely by moving the flexible linkers between SCR-4/5 and SCR-5/6 (52). Such large conformational changes in other complement proteins have already been reported for the thioester domain in C3b, and the Bb fragment from C3b within the C3 convertase (52). Interestingly, in the FH model with the compact N terminus, the smallest distance between SCR-1/4 and SCR-6/9 involved Arg-78 in SCR-1 and Asp-538 in SCR-9, which bear opposite charges (80). We speculate that these domains may electrostatically attract each other; indeed, conformational changes in FH with increase in ionic strength have been reported (26). Further studies are required to investigate such a proposed mechanism.

The existence of two FH conformations with different C3b and C3d binding strengths accounts for previously reported K_D values for each of SCR-1/4 binding to C3b (10–14 μM) and SCR-19/20 binding to C3b or C3d (1–5 and 0.2–8 μM , respectively). These K_D values are similar to the K_D value of full-length FH binding to C3b (0.6–3 μM) (21), indicating no synergy between the two binding sites at SCR-1/4 and SCR-19/20, and meaning that these two C3b and C3d sites are independent of

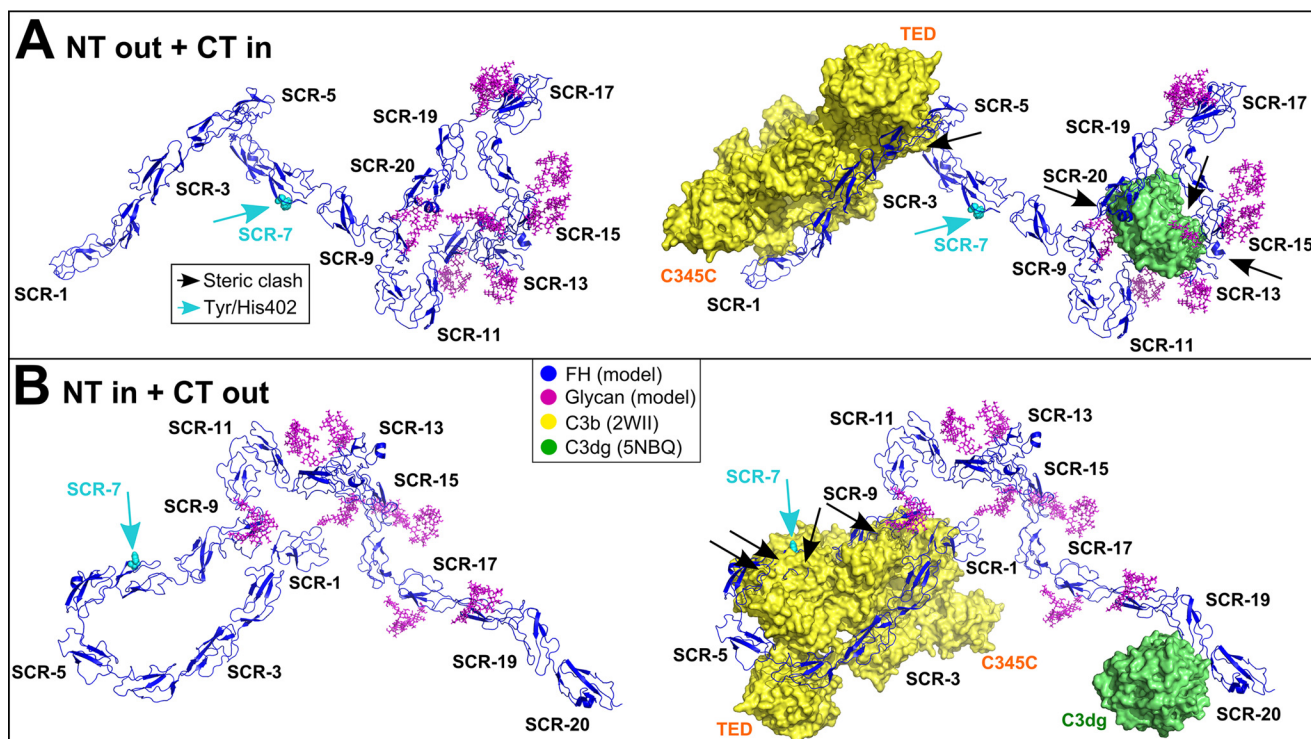


Figure 14. Best-fit FH centroid models superimposed onto C3b and C3dg. For both allotypes, the most representative N-terminally extended (A) and C-terminally extended (B) FH models are shown as *blue ribbon traces* (left). These correspond to the centroid models shown in Figs. 12, B and E, respectively. The coordinates are available for download in [supporting Materials](#). These models were superimposed onto the SCR-1/4-C3b and SCR-19/20-C3dg crystal structures, in which C3b is shown as a *yellow surface* and C3dg is shown as a *green surface* (right, PDB codes 2WII and 5NBQ). The eight FH glycans are shown in *magenta*. The *cyan arrow* indicates Tyr-402/His-402. *Black arrows* indicate steric clashes of FH with C3b or C3dg.

each other. This outcome is in concord with our N-terminally extended and C-terminally extended FH models that bind to C3b or C3d separately but not both together. In summary, FH is seen with an N-terminally extended conformation that is able to dismantle fluid phase C3 convertases, and also with a distinct C-terminally extended conformation that binds to cell surfaces via deposited C3d for enhanced local protection from complement activation. Both alternative conformations are able to bind to glycosaminoglycans at the host cell surface, thus protecting this against excess C3b activity. Overall, the SAXS data showed that FH underwent significant self-association, and the modeling of these data for the monomer has resulted in the proposal of two distinct FH conformations that would enable FH to perform separate C3b or C3d binding functions in either the fluid phase or at host cell surfaces, respectively.

Materials and methods

Protein purification of FH and FH SCR-6/8

Thirty aliquots of 50–100 ml of human blood was collected from anonymous healthy volunteers with ethical approval (University College London, ID: 6986/001). The studies abide by the Declaration of Helsinki principles. The Y402H and I62V polymorphisms were identified by direct DNA sequencing of their PCR product. PCR was performed with Extensor Hi-Fidelity PCR[®] master mix from ABgene following the manufacturer's protocol. Bidirectional sequencing of the PCR product was carried out with Big Dye Terminator version 3.1 on a 3730 DNA Analyzer (Applied Biosystems, Cheshire, UK). Homozygous FH Tyr-402, which was homozygous or heterozygous at I62V,

and double homozygous FH His-402/Val-62 were purified using monoclonal affinity chromatography with an MRC-OX23–Sepharose column (29, 81). Bound FH was eluted from the column using 3 M MgCl₂, pH 6.9, and then dialyzed into HEPES buffer (10 mM HEPES, 137 mM NaCl, pH 7.4) in the presence of 0.5 mM EDTA to remove Mg²⁺. To avoid cross-contamination between different FH samples, the MRC-OX23 column was washed with guanidine (0.2 M Tris, 4 M guanidine-HCl, pH 8.0) between purifications. Each FH sample was passed through a Hitrap[™] protein G HP column to remove residual IgG contaminant. Nonspecific aggregates and human serum albumin were removed by gel filtration on a Superose[™] 6 prep grade XK 16/60 column (Fig. 2). Protein concentrations were determined using an absorption coefficient of 16.2 for FH (Tyr-402) and 16.1 for FH (His-402) (1%, 280 nm, 1 cm path length), revised in the light of its known glycosylation (26). After sedimentation velocity experiments on the purified FH allotypes, the proteins were re-passed through the Superose[™] 6 prep grade XK 16/60 column to analyze the presence of FH oligomers (Fig. 2).

Recombinant FH SCR-6/8 was purified as described previously (30). SCR-6/8 Tyr-402 was expressed in BL21 (DE3) *Escherichia coli* cells using a pET21ab vector system, whereas SCR-6/8 His-402 was expressed in BL21 (DE3)pLysS *E. coli* cells using a pET14b vector system. The protein was solubilized from inclusion bodies, and the refolded protein was passed through a Hitrap[™] 5-ml heparin HP column equilibrated in Tris buffer (50 mM Tris, 150 mM NaCl, 1 mM EDTA, pH 7.4). The bound protein was washed with Tris buffer and eluted

with up to 1 M NaCl in Tris buffer. Proteins were dialyzed into HEPES buffer for experiments, and their integrity was routinely checked by SDS-PAGE before and after scattering and ultracentrifugation.

Sedimentation velocity data

AUC data were obtained on two Beckman XL-I instruments equipped with AnTi50 and AnTi60 rotors. Sedimentation velocity experiments were performed at 20 °C at rotor speeds of 50,000 and 60,000 rpm in two-sector cells with column heights of 12 mm for five FH Tyr-402 samples at 0.5, 1.1, and 2.1 mg/ml (3.2–13.6 μM), and repeated for five FH His-402/Val-62 samples, all in HEPES buffer. Size-distribution analyses $c(s)$ were performed using SEDFIT (82, 83). SEDFIT provided shape and size data using direct boundary Lamm fits of up to 300 scans using a fixed resolution of 200, and floated the meniscus, the bottom of the cell, the baseline, and the average frictional ratio f/f_0 (initial value 1.781), until the overall root mean square deviation and agreement between the observed and calculated sedimentation boundaries were satisfactory. The percentage fraction of oligomers in the total loading concentration was derived using the $c(s)$ integration function. Other details are described elsewhere (26, 29).

Mass spectrometry data

Mass spectra were acquired on a Q-ToF (quadrupole-TOF) 2 MS instrument for high mass operation and equipped with a Z-spray nanoflow source (Waters, Manchester, UK) (84). The following instrumental parameters were used: capillary voltage 1.7 kV; cone voltage 80–120 V; cone gas 100 liters/h; collision cell voltage up to 200 V; ion transfer stage pressure 0.004–0.02 mbar; and argon collision gas at a collision cell pressure of 2.5 bar. Mass spectra were analyzed using the software Data-Explorer. Purified FH samples were concentrated to a final concentration of about 4 mg/ml (26 μM), and then dialyzed overnight into 140 mM or 1 M ammonium acetate at pH 7.5 for measurement the next day. One pair of FH Tyr-402 and His-402 allotypes was analyzed in both buffers.

Surface plasmon resonance data

The binding of the FH Tyr-402 and His-402 allotypes to immobilized FH and that of the SCR-6/8 Tyr-402 and His-402 allotypes to immobilized FH SCR-6/8 were studied using a Biacore X100 with version 1.1 of its evaluation software (GE Healthcare, Uppsala, Sweden). For both the self-association studies and the human serum albumin control, FH was coupled to a carboxymethylated dextran (CM3) research grade sensor chip with a shorter carboxymethylated dextran matrix than a standard CM5 chip, which gave nonspecific binding (85). Each FH SCR-6/8 allotype was coupled to a CM4 research grade sensor chip with a low degree of carboxymethylation compared with a CM5 chip. Immobilizations utilized a standard amine coupling procedure according to the manufacturer's protocol. FH (20 $\mu\text{g}/\text{ml}$) or SCR-6/8 (10 $\mu\text{g}/\text{ml}$) in 10 mM acetate buffer, pH 5.5 (FH), was injected over flow cell 2 to reach a maximum response of 150 RU. A control surface cell was prepared identically on flow cell 1 but without protein immobilization. Equilibrium analyses were performed at 25 °C using Biacore X100

wizards at a flow rate of 30 $\mu\text{l}/\text{min}$. The running buffer was HEPES buffer. FH Tyr-402 was injected for 180 s at six concentrations between 0.1 and 1.5 mg/ml (0.7–9.7 μM). FH His-402 was injected at six concentrations between 0.1 and 1.8 mg/ml (0.7–11.7 μM). SCR-6/8 Tyr-402 was injected for 150 s over the CM4 chip immobilized with either the Tyr-402 or the His-402 allotype at concentrations of 0.1–0.8 mg/ml (5–40 μM). SCR-6/8 His-402 was injected over the Tyr-402 CM4 chip at concentrations of 0.1–0.6 mg/ml (5–28 μM) and over the His-402 CM4 chip at concentrations of 0.1–0.5 mg/ml (5–22 μM). For the SCR-6/8 analyses, the CM4 chips were used with 0.3 mg/ml nonspecific binding reducer (GE Healthcare, Uppsala, Sweden) added to each sample to reduce nonspecific binding to dextran surfaces. Regeneration after each run was achieved by pulsing with 10 mM sodium acetate buffer, 5 M NaCl, 0.05 mM EDTA, pH 7.0, across both flow cells twice for 45 s. The maximum binding response value in each run was fitted to a steady-state 1:1 affinity model.

X-ray scattering data

SAXS data were acquired in two sessions on Instrument ID02 (86) at the European Synchrotron Radiation Facility (Grenoble, France) with a ring energy of 6.0 GeV and operating in four-bunch mode. Storage ring currents were 30–44 and 68–89 mA, and the sample-detector distances were 2 and 3 m, respectively. Radiation damage was eliminated by the continuous movement of the sample in its capillary flow cell during beam exposure, the use of 10 time frames of duration between 0.1 and 0.5 s each during each acquisition, and on-line checks for the absence of radiation damage at low Q . Two FH Tyr-402 and two FH His-402 allotypes were each studied at five concentrations between 0.4 mg/ml (2.6 μM) and 3.3 mg/ml (21.4 μM). Measurements were done in HEPES buffer. Other details including data reduction are described elsewhere (30, 31).

In a given solute-solvent contrast, the radius of gyration (R_g) is a measure of structural elongation if the internal inhomogeneity of scattering densities within the protein has no effect. Guinier analyses at low Q gives the R_g , and the forward scattering at zero angle $I(0)$ (87) as shown in Equation 1,

$$\ln I(Q) = \ln I(0) - R_g^2 Q^2 / 3 \quad (\text{Eq. 1})$$

This expression is valid in a $Q \cdot R_g$ range up to 1.5. If the structure is elongated, the mean radius of gyration of cross-sectional structure R_{XS} and the mean cross-sectional intensity at zero angle ($I(Q)Q$) $_{Q \rightarrow 0}$ are obtained from Equation 2,

$$\ln(I(Q)Q) = (I(Q)Q)_{Q \rightarrow 0} - R_{XS}^2 Q^2 / 2 \quad (\text{Eq. 2})$$

The R_g and R_{XS} parameters were calculated using the SCT suite of programs (88). Indirect transformation of the scattering data $I(Q)$ in reciprocal space into real space to give the distance distribution function $P(r)$ was carried out using the program GNOM: (89)

$$P(r) = \frac{1}{2\pi^2} \int_0^\infty I(Q) Q r \sin(Qr) dQ \quad (\text{Eq. 3})$$

Solution structures of FH Tyr-402 and His-402

$P(r)$ corresponds to the distribution of distances r between volume elements. For this, the X-ray $I(Q)$ curve utilized up to 207 data points for Q between 0.06 and 1.50 nm⁻¹ for 0.4 mg/ml FH, and increasing up to 308 data points for Q between 0.06 and 2.10 nm⁻¹ for 3.3 mg/ml FH. Other details are described elsewhere (29–31, 50).

Atomistic FH scattering modeling

The first scattering modeling of the 20 SCR domains in FH used NMR structures for three SCR domains and homology models for the remaining 17 SCR domains (Fig. 1) (50). The second modeling used NMR and crystal structures for 11–14 SCR domains, and homology models for the remaining 9 SCR domains (26, 51). Here, a combination of MODELLER version 9.14 (90) and monomer Monte Carlo (SASSIE-web) (91) was used to build a starting FH model from NMR and crystal structures for 17 SCR domains and three improved SCR homology models for SCR-9, SCR-14, and SCR-17 (Tables 1 and 2). For the homology modeling, a template SCR structure without linker residues was identified using the basic local alignment search tool to search the Protein Data Bank server (BLAST-PDB) (Table 2) (92), and its sequence was aligned to the template sequence using Clustal Omega multiple sequence alignment (93). By this, low sequence identity was mostly confined to the loop regions, and the secondary structure was conserved. These loop regions were modeled using the loop optimization protocol in MODELLER. The best SCR model was selected from a final dataset of 100 generated models using the normalized discrete optimized protein energy (DOPE) score (94). For the SCR-9, SCR-14, and SCR-17 homology models and the NMR-based SCR-5 model (95), strained high-energy bonds were relaxed to give physically stable models by energy minimization (500,000 steps) and molecular dynamics simulations (1,000,000 steps) in the simulation engine NAMD version 2.9 (96). The PDB Reader input generator tool (97, 98) converted the PDB file into CHARMM readable format, generated a protein structure file (PSF), including the disulfide bonds, and provided the CHARMM36 force-field files (99, 100). The quality of the secondary structure was verified using Ramachandran plots on the RAMPAGE server (101), the DSSP program (102, 103), and visual comparison with other SCR domains. As required, the Ramachandran plots showed a minimal number of outliers in these four SCR models. To create the initial full-length FH structure (Fig. 1), the 11 SCR fragment models (SCR-1/4, SCR-5, SCR-6/8(His-402), SCR-9, SCR-10/11, SCR-11/12, SCR-12/13, SCR-14, SCR-15/16, SCR-17, and SCR-18/20) were aligned to the FH sequence sequentially using MODELLER. Each of SCR-1/5, SCR-6/13, SCR-14/16, and SCR-17/20 were assembled first, before their simultaneous alignment to the full-length FH sequence. The best FH model was selected based on the normalized DOPE score and then refined in NAMD to give the final FH initial structure. FH was numbered from the N terminus of the 18-residue signal peptide to follow HGVS convention. Eight biantennary disialylated glycans (18) were added to Asn-511 in SCR-9, Asn-700 in SCR-12, Asn-784 in SCR-13, Asn-804 in SCR-14, Asn-864 and Asn-892 in SCR-15, Asn-1011 in SCR-17, and Asn-1077 in SCR-19 (Fig. 1). For each glycan–Asn Protein Data Bank (PDB) file, GlycanReader was

used to generate the CHARMM force field and PSF inputs for energy minimization in NAMD. The resulting energy-minimized glycans were positioned onto FH by superimposing the Asn residues in PyMOL and then deleting the Asn residue in the glycan PDB. Once all eight glycans had been added to FH and accepted by GlycanReader, bash scripts were used to convert the nomenclature and numbering of the glycan and protein atoms to the format required for the torsion angle Monte Carlo (TAMC) module in SASSIE-web. In this, the glycans were moved with their attached protein segment and not varied independently.

The model library of physically realistic FH structural conformations was generated by subjecting the inter-SCR linkers to the TAMC module in SASSIE-web (104). Eleven of the 19 linkers were defined by high-resolution structures and were therefore held fixed. The remaining eight linkers at SCR-4/5, SCR-5/6, SCR-8/9, SCR-9/10, SCR-13/14, SCR-14/15, SCR-16/17, and SCR-17/18 were moved (arrowed, Fig. 1). For each of these linker residues, the backbone φ and ψ torsion angles were varied in steps of up to either 30 or 180°, except for the torsion angles involving the conserved Cys residues at the start and end of each linker. To maximize the sampling, four simulations were run as follows: (i) 200,000 steps with up to 30° moves using the FH initial model; (ii) 200,000 steps with up to 180° moves using the FH initial model; (iii) 100,000 steps with up to 30° moves using an FH model from simulations i and ii with its CT bent inwards, and (iv) 10,000 steps with up to 30° moves using the best-fit FH model from simulations i–iii. In a Monte Carlo simulation, many moves result in structures with steric clashes and were discarded as physically unrealistic. In the present case, the 510,000 attempted moves resulted in 15,136 (8%), 3720 (1%), 10,316 (10%), and 543 (5%) physically-realistic, acceptable models, respectively. These were combined into a library of 29,715 models for SAXS curve fitting. A theoretical scattering curve was generated from each model using SasCalc. SasCalc calculates the scattering curve using an exact all-atom expression for the scattering intensity in which the orientations of the Q vectors are taken from a quasi-uniform spherical grid generated by the golden ratio (105).

For each of the Tyr-402 and His-402 allotypes, the experimental scattering curve with 221 data points was extrapolated to zero concentration to eliminate the contribution of self-association. These were compared with the theoretical curves using the R -factor shown in Equation 4,

$$R = \sum_{Q_i} \frac{|(I_{\text{exp}}(Q_i) - I_{\text{model}}(Q_i))|}{|I_{\text{exp}}(Q_i)|} \quad (\text{Eq. 4})$$

where Q_i is the Q value of the i th data point; $I_{\text{exp}}(Q_i)$ is the experimental scattering intensity; and $I_{\text{model}}(Q_i)$ is the theoretical modeled scattering intensity (88). For SasCalc, the lowest Q values before the Guinier R_g region in the extrapolated scattering curves were interpolated to zero Q using MATLAB. After interpolation, the original 221 $I(Q)$ values between Q of 0.0 and 1.5 nm⁻¹ were retained to define the Q spacing for SasCalc. The R -factor versus R_g graphs were similar for the extrapolated and 0.4 mg/ml curves, whereas those from curve fitting at 0.7, 1.1, and 2.2 mg/ml gave worse R -factors (data not shown), as

expected. For each of the four Tyr-402 and His-402 curves, the 29,715 models were filtered on both R_g and R -factor. Models were accepted if their R_g values were within $\pm 5\%$ of the experimental R_g values (Table 3), and their R -factor was $\leq 5\%$. The best-fit 100 models were identified by ranking the filtered models by their R -factors. The Y402H polymorphism had no effect on the curve fits, leading to an R -factor difference of only 0.0003%.

The SCR domain arrangement in the 29,715 FH models was parameterized using the separations between the FH COM and the NT and CT α -carbon atoms. Separations were calculated using Tcl scripting within Visual Molecular Dynamics (VMD) software (106), and their frequencies were displayed as histograms using R (107) for statistical analyses. To assess the normality of the separation distributions, Quantile-Quantile plots and two-tailed Kolmogorov-Smirnov tests were employed using the `ks.test` statistical function in R (107, 108). Here, the null hypothesis stated that the distribution was consistent with the normal distribution, and a Bonferroni-corrected significance level of 0.05 divided by 2 (0.025) was set. The Kolmogorov-Smirnov test calculated the maximum vertical deviation between the two curves as the test statistic D and the probability of D occurring due to chance (p value). To compare the distributions for all 29,715 models, the 310–2678 filtered models, and the best-fit 100 models of the two allotypes, their distributions were converted to cumulative density frequencies and compared using the two-tailed Kolmogorov-Smirnov test in R with the `ks.test` statistical function. Here, for each comparison, the null hypothesis stated that the distributions were the same and a Bonferroni-corrected significance level of 0.05 divided by 12 (0.00417) was set. To visualize the resulting conformations after applying the R_g and R -factor filters, the filtered models were superimposed with the reference FH model using SCR-10/13, and visualized as density plots using SASSIE-web.

Principal component analysis (PCA) provided by the Bio3d package in R (108) identified four major conformational states in the best-fit 100 models. For both allotypes, PCA satisfactorily accounted for $>80\%$ of the variance between the models. The average FH structure for each PCA group was identified using a centroid model computed using R. The SCR domain arrangement of each PCA group was identified by a density plot using SASSIE-web. To assess the conformational differences between the two allotypes, the frequencies of the two major SCR domain arrangements were compared by using the two-tailed Fisher's test for a 2×2 contingency table in GraphPad QuickCalcs (<https://www.graphpad.com/quickcalcs/>).⁴ Here, the null hypothesis stated that the proportion of FH models with an inwardly-bent NT and extended CT, compared with an extended NT and inwardly-bent CT, were the same. The significance level was set as 0.05, and the probability of the difference occurring due to chance (p value) was computed. To see whether these two conformations could be distinguished by their SAXS curves, the difference in scattering intensity between their best fitting centroid model curves was calculated and visualized using R. The FH models were assessed by superimposing the

centroid model for each of the largest ensembles with crystal structures for the FH-ligand complexes of SCR-1/4 with C3b and SCR-19/20 with C3d using PyMOL (PDB codes 2WII and 5NBQ) (52, 78).

For the AUC modeling, the theoretical $s_{20,w}^0$ values for the FH models were calculated directly from the atomic coordinates with the default value of 0.31 nm for the atomic element radius for all atoms to represent the hydration shell by using the HYDROPRO shell modeling program (109). The molecular weight and partial specific volume for FH used followed that used previously (26, 110). FH glycosylation was considered by using the compiled partial specific volumes for saccharides (111).

Links to web servers and tools

The following were used: BLAST (<https://blast.ncbi.nlm.nih.gov/Blast.cgi>)⁴; Clustal Omega (<http://www.ebi.ac.uk/Tools/msa/clustalo/>)⁴; SASSIE-web (<https://sassie-web.chem.utk.edu/sassie2/>)⁴; CHARMM-GUI Glycan Reader (<http://charmm-gui.org/?doc=input/glycan>)⁴; CHARMM-GUI PDB Reader (<http://www.charmm-gui.org/?doc=input/pdbreader>)⁴; RAMPAGE (<http://mordred.bioc.cam.ac.uk/~rapper/>)⁴; DSSP (<http://swift.cmbi.ru.nl/gv/dssp/>)⁴; and GraphPad QuickCalcs (<https://www.graphpad.com/quickcalcs/contingency1/>).⁴

Author contributions—A. J. O., R. N., and S. J. P. conceptualization; S. J. P. supervision; A. J. O., R. N., A. M., J. G., and S. J. P. investigation; A. J. O., R. N., A. M., J. S. B., and J. G. methodology; A. J. O., R. N., and S. J. P. writing-original draft; A. J. O., R. N., and S. J. P. writing-review and editing; R. N. and S. J. P. formal analysis; J. S. B. software; S. J. P. funding acquisition.

Acknowledgments—We are most grateful to Dr. Anuj Shukla and Dr. Emanuela Di Cola (European Synchrotron Radiation Facility, Grenoble, France) for instrumental support; Dr. Naushin Waseem and Prof. Shomi Bhattacharya for the FH genotyping; Dr. Adam R. McKay for MS support; Dr. Imre Lengyel and Prof. Alan Bird for useful discussions; and Georgia Ward and Dr. Louise Gavigan for biochemical support.

References

1. Law, S. K., and Reid, K. B. (1995) in *Complement* (Male, D., ed) 2nd Ed., IRL Press at Oxford University Press, Oxford, UK
2. Whaley, K., and Ruddy, S. (1976) Modulation of the alternative complement pathways by $\beta 1H$ globulin. *J. Exp. Med.* **144**, 1147–1163 [CrossRef Medline](#)
3. Pangburn, M. K., Schreiber, R. D., and Müller-Eberhard, H. J. (1977) Human complement C3b inactivator: isolation, characterization, and demonstration of an absolute requirement for the serum protein $\beta 1H$ for cleavage of C3b and C4b in solution. *J. Exp. Med.* **146**, 257–270 [CrossRef Medline](#)
4. Farries, T. C., Seya, T., Harrison, R. A., and Atkinson, J. P. (1990) Competition for binding sites on C3b by CR1, CR2, MCP, factor B and factor H. *Complement Inflamm.* **7**, 30–41 [CrossRef Medline](#)
5. Weiler, J. M., Daha, M. R., Austen, K. F., and Fearon, D. T. (1976) Control of the amplification convertase of complement by the plasma protein $\beta 1H$. *Proc. Natl. Acad. Sci. U.S.A.* **73**, 3268–3272 [CrossRef Medline](#)
6. Liszewski, M. K., Farries, T. C., Lublin, D. M., Rooney, I. A., and Atkinson, J. P. (1996) Control of the complement system. *Adv. Immunol.* **61**, 201–283 [CrossRef Medline](#)
7. Klein, R. J., Zeiss, C., Chew, E. Y., Tsai, J. Y., Sackler, R. S., Haynes, C.,

⁴ Please note that the JBC is not responsible for the long-term archiving and maintenance of this site or any other third party hosted site.

Solution structures of FH Tyr-402 and His-402

- Henning, A. K., SanGiovanni, J. P., Mane, S. M., Mayne, S. T., Bracken, M. B., Ferris, F. L., Ott, J., Barnstable, C., and Hoh, J. (2005) Complement factor H polymorphism in age-related macular degeneration. *Science* **308**, 385–389 [CrossRef Medline](#)
8. Haines, J. L., Hauser, M. A., Schmidt, S., Scott, W. K., Olson, L. M., Gallins, P., Spencer, K. L., Kwan, S. Y., Noureddine, M., Gilbert, J. R., Schetz-Boutaud, N., Agarwal, A., Postel, E. A., and Pericak-Vance, M. A. (2005) Complement factor H variant increases the risk of age-related macular degeneration. *Science* **308**, 419–421 [CrossRef Medline](#)
9. Edwards, A. O., Ritter, R., 3rd., Abel, K. J., Manning, A., Panhuysen, C., and Farrer, L. A. (2005) Complement factor H polymorphism and age-related macular degeneration. *Science* **308**, 421–424 [CrossRef Medline](#)
10. Hageman, G. S., Anderson, D. H., Johnson, L. V., Hancox, L. S., Tauber, A. J., Hardisty, L. L., Hageman, J. L., Stockman, H. A., Borchardt, J. D., Gehrs, K. M., Smith, R. J., Silvestri, G., Russell, S. R., Klaver, C. C., Barbazetto, I., *et al.* (2005) A common haplotype in the complement regulatory gene factor H (HF1/CFH) predisposes individuals to age-related macular degeneration. *Proc. Natl. Acad. Sci. U.S.A.* **102**, 7227–7232 [CrossRef Medline](#)
11. Strohmeyer, R., Ramirez, M., Cole, G. J., Mueller, K., and Rogers, J. (2002) Association of factor H of the alternative pathway of complement with agrin and complement receptor 3 in the Alzheimer's disease brain. *J. Neuroimmunol.* **131**, 135–146 [CrossRef Medline](#)
12. Saunders, R. E., Abarrategui-Garrido, C., Frémeaux-Bacchi, V., Goicoechea de Jorge, E., Goodship, T. H., López Trascasa, M., Noris, M., Ponce Castro, I. M., Remuzzi, G., Rodríguez de Córdoba, S., Sánchez-Corral, P., Skerka, C., Zipfel, P. F., and Perkins, S. J. (2007) The interactive Factor H-atypical hemolytic uremic syndrome mutation database and website: update and integration of membrane cofactor protein and Factor I mutations with structural models. *Hum. Mutat.* **28**, 222–234 [CrossRef Medline](#)
13. Zetterberg, M., Landgren, S., Andersson, M. E., Palmer, M. S., Gustafson, D. R., Skoog, I., Minthon, L., Thelle, D. S., Wallin, A., Bogdanovic, N., Andreasen, N., Blennow, K., and Zetterberg, H. (2008) Association of complement factor H Y402H gene polymorphism with Alzheimer's disease. *Am. J. Med. Genet. B Neuropsychiatr. Genet.* **147**, 720–726 [CrossRef Medline](#)
14. Bird, A. C. (2010) Therapeutic targets in age-related macular disease. *J. Clin. Invest.* **120**, 3033–3041 [CrossRef Medline](#)
15. Crabb, J. W., Miyagi, M., Gu, X., Shadrach, K., West, K. A., Sakaguchi, H., Kamei, M., Hasan, A., Yan, L., Rayborn, M. E., Salomon, R. G., and Hollyfield, J. G. (2002) Drusen proteome analysis: an approach to the etiology of age-related macular degeneration. *Proc. Natl. Acad. Sci. U.S.A.* **99**, 14682–14687 [CrossRef Medline](#)
16. Bok, D. (2005) Evidence for an inflammatory process in age-related macular degeneration gains new support. *Proc. Natl. Acad. Sci. U.S.A.* **102**, 7053–7054 [CrossRef Medline](#)
17. Lambris, J. D., and Morikis, D. (2005) in *Structural Biology of the Complement System* (Lambris, J. D., and Morikis, D., eds) 1st Ed., pp. 48–51, Taylor & Francis, Boca Raton, FL
18. Fenaille, F., Le Mignon, M., Groseil, C., Ramon, C., Riandé, S., Siret, L., and Bihoreau, N. (2007) Site-specific N-glycan characterization of human complement factor H. *Glycobiology* **17**, 932–944 [CrossRef Medline](#)
19. Perkins, S. J., Nan, R., Okemefuna, A. I., Li, K., Khan, S., and Miller, A. (2010) Multiple interactions of complement Factor H with its ligands in solution: a progress report. *Adv. Exp. Med. Biol.* **703**, 25–47 [CrossRef Medline](#)
20. Perkins, S. J., Okemefuna, A. I., and Nan, R. (2010) Unravelling protein-protein interactions between complement factor H and C-reactive protein using a multidisciplinary strategy. *Biochem. Soc. Trans.* **38**, 894–900 [CrossRef Medline](#)
21. Perkins, S. J., Nan, R., Li, K., Khan, S., and Miller, A. (2012) Complement factor H-ligand interactions: self-association, multivalency and dissociation constants. *Immunobiology* **217**, 281–297 [CrossRef Medline](#)
22. Perkins, S. J., Fung, K. W., and Khan, S. (2014) Molecular interactions between complement Factor H and its heparin and heparan sulfate ligands. *Front. Immunol.* **5**, 126 [Medline](#)
23. Blackmore, T. K., Hellwege, J., Sadlon, T. A., Higgs, N., Zipfel, P. F., Ward, H. M., and Gordon, D. L. (1998) Identification of the second heparin-binding domain in human complement factor H. *J. Immunol.* **160**, 3342–3348 [Medline](#)
24. Blackmore, T. K., Sadlon, T. A., Ward, H. M., Lublin, D. M., and Gordon, D. L. (1996) Identification of a heparin binding domain in the seventh short consensus repeat of complement factor H. *J. Immunol.* **157**, 5422–5427 [Medline](#)
25. Ormsby, R. J., Jokiranta, T. S., Duthy, T. G., Griggs, K. M., Sadlon, T. A., Giannakis, E., and Gordon, D. L. (2006) Localization of the third heparin-binding site in the human complement regulator factor H1. *Mol. Immunol.* **43**, 1624–1632 [CrossRef Medline](#)
26. Okemefuna, A. I., Nan, R., Gor, J., and Perkins, S. J. (2009) Electrostatic interactions contribute to the folded-back conformation of wild type human factor H. *J. Mol. Biol.* **391**, 98–118 [CrossRef Medline](#)
27. Khan, S., Nan, R., Gor, J., Mulloy, B., and Perkins, S. J. (2012) Bivalent and co-operative binding of complement factor H to heparan sulfate and heparin. *Biochem. J.* **444**, 417–428 [CrossRef Medline](#)
28. Perkins, S. J., Nealis, A. S., and Sim, R. B. (1991) Oligomeric domain structure of human complement factor H by X-ray and neutron solution scattering. *Biochemistry* **30**, 2847–2857 [CrossRef Medline](#)
29. Nan, R., Gor, J., and Perkins, S. J. (2008a) Implications of the progressive self-association of wild-type human factor H for complement regulation and disease. *J. Mol. Biol.* **375**, 891–900 [CrossRef Medline](#)
30. Fernando, A. N., Furtado, P. B., Clark, S. J., Gilbert, H. E., Day, A. J., Sim, R. B., and Perkins, S. J. (2007) Associative and structural properties of the region of complement factor H encompassing the Y402H disease-related polymorphism and its interactions with heparin. *J. Mol. Biol.* **368**, 564–581 [CrossRef Medline](#)
31. Okemefuna, A. I., Gilbert, H. E., Griggs, K. M., Ormsby, R. J., Gordon, D. L., and Perkins, S. J. (2008) The regulatory SCR-1/5 and cell surface-binding SCR-16/20 fragments of factor H reveal partially folded-back solution structures and different self-associative properties. *J. Mol. Biol.* **375**, 80–101 [CrossRef Medline](#)
32. Pangburn, M. K., Rawal, N., Cortes, C., Alam, M. N., Ferreira, V. P., and Atkinson, M. A. (2009) Polyanion-induced self-association of complement factor H. *J. Immunol.* **182**, 1061–1068 [CrossRef Medline](#)
33. DiScipio, R. G. (1992) Ultrastructures and interactions of complement factors H and I. *J. Immunol.* **149**, 2592–2599 [Medline](#)
34. Osborne, A. J., Breno, M., Borsa, N. G., Bu, F., Frémeaux-Bacchi, V., Gale, D. P., van den Heuvel, L. P., Kavanagh, D., Noris, M., Pinto, S., Rallapalli, P. M., Remuzzi, G., Rodríguez de Córdoba, S., Ruiz, A., *et al.* (2018) Statistical validation of rare complement variants provides insights into the molecular basis of atypical hemolytic uremic syndrome and C3 glomerulopathy. *J. Immunol.* **200**, 2464–2478 [CrossRef Medline](#)
35. Thakkinstian, A., Han, P., McEvoy, M., Smith, W., Hoh, J., Magnusson, K., Zhang, K., and Attia, J. (2006) Systematic review and meta-analysis of the association between complement factor H Y402H polymorphisms and age-related macular degeneration. *Hum. Mol. Genet.* **15**, 2784–2790 [CrossRef Medline](#)
36. Herbert, A. P., Deakin, J. A., Schmidt, C. Q., Blaum, B. S., Egan, C., Ferreira, V. P., Pangburn, M. K., Lyon, M., Uhrin, D., and Barlow, P. N. (2007) Structure shows that a glycosaminoglycan and protein recognition site in factor H is perturbed by age-related macular degeneration-linked single nucleotide polymorphism. *J. Biol. Chem.* **282**, 18960–18968 [CrossRef Medline](#)
37. Skerka, C., Lauer, N., Weinberger, A. A., Keilhauer, C. N., Sühnel, J., Smith, R., Schlötzer-Schrehardt, U., Fritsche, L., Heinen, S., Hartmann, A., Weber, B. H., and Zipfel, P. F. (2007) Defective complement control of factor H (Y402H) and FHL-1 in age-related macular degeneration. *Mol. Immunol.* **44**, 3398–3406 [CrossRef Medline](#)
38. Laine, M., Jarva, H., Seitonen, S., Haapasalo, K., Lehtinen, M. J., Lindeman, N., Anderson, D. H., Johnson, P. T., Järvelä, I., Jokiranta, T. S., Hageman, G. S., Immonen, I., and Meri, S. (2007) Y402H polymorphism of complement factor H affects binding affinity to C-reactive protein. *J. Immunol.* **178**, 3831–3836 [CrossRef Medline](#)
39. Ormsby, R. J., Ranganathan, S., Tong, J. C., Griggs, K. M., Dimasi, D. P., Hewitt, A. W., Burdon, K. P., Craig, J. E., Hoh, J., and Gordon, D. L. (2008) Functional and structural implications of the complement factor H

- Y402H polymorphism associated with age-related macular degeneration. *Invest. Ophthalmol. Vis. Sci.* **49**, 1763–1770 [CrossRef Medline](#)
40. Langford-Smith, A., Keenan, T. D., Clark, S. J., Bishop, P. N., and Day, A. J. (2014) The role of complement in age-related macular degeneration: heparan sulphate, a ZIP code for complement factor H? *J. Innate Immun.* **6**, 407–416 [CrossRef Medline](#)
 41. Clark, S. J., Ridge, L. A., Herbert, A. P., Hakobyan, S., Mulloy, B., Lennon, R., Würzner, R., Morgan, B. P., Uhrin, D., Bishop, P. N., and Day, A. J. (2013) Tissue-specific host recognition by complement factor H is mediated by differential activities of its glycosaminoglycan-binding regions. *J. Immunol.* **190**, 2049–2057 [CrossRef Medline](#)
 42. Prosser, B. E., Johnson, S., Roversi, P., Herbert, A. P., Blaum, B. S., Tyrrell, J., Jowitt, T. A., Clark, S. J., Tarelli, E., Uhrin, D., Barlow, P. N., Sim, R. B., Day, A. J., and Lea, S. M. (2007) Structural basis for complement factor H linked age-related macular degeneration. *J. Exp. Med.* **204**, 2277–2283 [CrossRef Medline](#)
 43. Raychaudhuri, S., Iartchouk, O., Chin, K., Tan, P. L., Tai, A. K., Ripke, S., Gowrisankar, S., Vemuri, S., Montgomery, K., Yu, Y., Reynolds, R., Zack, D. J., Campochiaro, B., Campochiaro, P., Katsanis, N., *et al.* (2011) A rare penetrant mutation in CFH confers high risk of age-related macular degeneration. *Nat. Genet.* **43**, 1232–1236 [CrossRef Medline](#)
 44. Geerlings, M. J., de Jong, E. K., and den Hollander, A. I. (2017) The complement system in age-related macular degeneration: a review of rare genetic variants and implications for personalized treatment. *Mol. Immunol.* **84**, 65–76 [CrossRef Medline](#)
 45. Hocking, H. G., Herbert, A. P., Kavanagh, D., Soares, D. C., Ferreira, V. P., Pangburn, M. K., Uhrin, D., and Barlow, P. N. (2008) Structure of the N-terminal region of complement factor H and conformational implications of disease-linked sequence variations. *J. Biol. Chem.* **283**, 9475–9487 [CrossRef Medline](#)
 46. Maller, J., George, S., Purcell, S., Fagerness, J., Altschuler, D., Daly, M. J., and Seddon, J. M. (2006) Common variation in three genes, including a noncoding variant in CFH, strongly influences risk of age-related macular degeneration. *Nat. Genet.* **38**, 1055–1059 [CrossRef Medline](#)
 47. Li, M., Atmaca-Sonmez, P., Othman, M., Branham, K. E., Khanna, R., Wade, M. S., Li, Y., Liang, L., Zarepari, S., Swaroop, A., and Abecasis, G. R. (2006) CFH haplotypes without the Y402H coding variant show strong association with susceptibility to age-related macular degeneration. *Nat. Genet.* **38**, 1049–1054 [CrossRef Medline](#)
 48. Jokiranta, T. S., Jaakola, V. P., Lehtinen, M. J., Pärepallo, M., Meri, S., and Goldman, A. (2006) Structure of complement factor H carboxyl-terminus reveals molecular basis of atypical haemolytic uremic syndrome. *EMBO J.* **25**, 1784–1794 [CrossRef Medline](#)
 49. Hecker, L. A., Edwards, A. O., Ryu, E., Tosakulwong, N., Baratz, K. H., Brown, W. L., Charbel Issa, P., Scholl, H. P., Pollok-Kopp, B., Schmid-Kubista, K. E., Bailey, K. R., and Oppermann, M. (2010) Genetic control of the alternative pathway of complement in humans and age-related macular degeneration. *Hum. Mol. Genet.* **19**, 209–215 [CrossRef Medline](#)
 50. Aslam, M., and Perkins, S. J. (2001) Folded-back solution structure of monomeric factor H of human complement by synchrotron X-ray and neutron scattering, analytical ultracentrifugation and constrained molecular modelling. *J. Mol. Biol.* **309**, 1117–1138 [CrossRef Medline](#)
 51. Nan, R., Ward, G., Gavigan, L., Miller, A., Gor, J., Lengyel, I., and Perkins, S. J. (2010) The His-402 allotype of complement factor H show similar self-association to the Tyr-402 allotype but exhibits greater self-association in the presence of zinc. *Mol. Immunol.* **47**, 2263 [CrossRef](#)
 52. Wu, J., Wu, Y. Q., Ricklin, D., Janssen, B. J., Lambris, J. D., and Gros, P. (2009) Structure of complement fragment C3b-factor H and implications for host protection by complement regulators. *Nat. Immunol.* **10**, 728–733 [CrossRef Medline](#)
 53. Schmidt, C. Q., Herbert, A. P., Mertens, H. D., Guariento, M., Soares, D. C., Uhrin, D., Rowe, A. J., Svergun, D. I., and Barlow, P. N. (2010) The central portion of factor H (modules 10–15) is compact and contains a structurally deviant CCP module. *J. Mol. Biol.* **395**, 105–122 [CrossRef Medline](#)
 54. Perkins, S. J., Wright, D. W., Zhang, H., Brookes, E. H., Chen, J., Irving, T. C., Krueger, S., Barlow, D. J., Edler, K. J., Scott, D. J., Terrill, N. J., King, S. M., Butler, P. D., and Curtis, J. E. (2016) Atomistic modelling of scattering data in the collaborative computational project for small angle scattering (CCP-SAS). *J. Appl. Crystallogr.* **49**, 1861–1875 [CrossRef Medline](#)
 55. Cole, J. L., Lary, J. W., P Moody, T., and Laue, T. M. (2008) Analytical ultracentrifugation: sedimentation velocity and sedimentation equilibrium. *Methods Cell Biol.* **84**, 143–179 [CrossRef Medline](#)
 56. Tortajada, A., Montes, T., Martínez-Barricarte, R., Morgan, B. P., Harris, C. L., and de Córdoba, S. R. (2009) The disease-protective complement factor H allotypic variant Ile62 shows increased binding affinity for C3b and enhanced cofactor activity. *Hum. Mol. Genet.* **18**, 3452–3461 [CrossRef Medline](#)
 57. Nan, R., Gor, J., Lengyel, I., and Perkins, S. J. (2008b) Uncontrolled zinc- and copper-induced oligomerization of the human complement regulator factor H and its possible implications for function and disease. *J. Mol. Biol.* **384**, 1341–1352 [CrossRef Medline](#)
 58. Dam, J., and Schuck, P. (2005) Sedimentation velocity analysis of heterogeneous protein-protein interactions: sedimentation coefficient distributions *c(s)* and asymptotic boundary profiles from Gilbert-Jenkins theory. *Biophys. J.* **89**, 651–666 [CrossRef Medline](#)
 59. Benesch, J. L., and Robinson, C. V. (2009) Biological chemistry: dehydrated but unharmed. *Nature* **462**, 576–577 [CrossRef Medline](#)
 60. Tanius, F. A., Nguyen, B., and Wilson, W. D. (2008) Biosensor-surface plasmon resonance methods for quantitative analysis of biomolecular interactions. *Methods Cell Biol.* **84**, 53–77 [CrossRef Medline](#)
 61. Perkins, S. J., Okemefuna, A. I., Fernando, A. N., Bonner, A., Gilbert, H. E., and Furtado, P. B. (2008) X-ray and neutron scattering data and their constrained molecular modelling. *Methods Cell Biol.* **84**, 375–423 [CrossRef Medline](#)
 62. Receveur-Brechot, V., and Durand, D. (2012) How random are intrinsically disordered proteins? A small angle scattering perspective. *Curr. Protein Pept. Sci.* **13**, 55–75 [CrossRef Medline](#)
 63. Perkins, S. J., Okemefuna, A. I., Nan, R., Li, K., and Bonner, A. (2009) Constrained solution scattering modelling of human antibodies and complement proteins reveals novel biological insights. *J. R. Soc. Interface* **6**, Suppl. 5, S679–696 [CrossRef Medline](#)
 64. Fung, K. W., Wright, D. W., Gor, J., Swann, M. J., and Perkins, S. J. (2016) Domain structure of human complement C4b extends with increasing NaCl concentration: implications for its regulatory mechanism. *Biochem. J.* **473**, 4473–4491 [CrossRef Medline](#)
 65. Hui, G. K., Wright, D. W., Vennard, O. L., Rayner, L. E., Pang, M., Yeo, S. C., Gor, J., Molyneux, K., Barratt, J., and Perkins, S. J. (2015) The solution structures of native and patient monomeric human IgA1 reveal asymmetric extended structures: implications for function and IgAN disease. *Biochem. J.* **471**, 167–185 [CrossRef Medline](#)
 66. David, C. C., and Jacobs, D. J. (2014) Principal component analysis: a method for determining the essential dynamics of proteins. *Methods Mol. Biol.* **1084**, 193–226 [CrossRef Medline](#)
 67. Rodriguez, E., Rallapalli, P. M., Osborne, A. J., and Perkins, S. J. (2014) New functional and structural insights from updated mutational databases for complement factor H, factor I, membrane cofactor protein and C3. *Biosci. Rep.* **34**, 635–649 [CrossRef](#)
 68. Ferreira, V. P., Pangburn, M. K., and Cortés, C. (2010) Complement control protein factor H: the good, the bad, and the inadequate. *Mol. Immunol.* **47**, 2187–2197 [CrossRef Medline](#)
 69. Hageman, G. S., Luthert, P. J., Victor Chong, N. H., Johnson, L. V., Anderson, D. H., and Mullins, R. F. (2001) An integrated hypothesis that considers drusen as biomarkers of immune-mediated processes at the RPE-Bruch's membrane interface in aging and age-related macular degeneration. *Prog. Retin. Eye Res.* **20**, 705–732 [CrossRef Medline](#)
 70. Kelly, U., Yu, L., Kumar, P., Ding, J. D., Jiang, H., Hageman, G. S., Arshavsky, V. Y., Frank, M. M., Hauser, M. A., and Rickman, C. B. (2010) Heparan sulfate, including that in Bruch's membrane, inhibits the complement alternative pathway: implications for age-related macular degeneration. *J. Immunol.* **185**, 5486–5494 [CrossRef Medline](#)
 71. Toomey, C. B., Johnson, L. V., and Bowes Rickman, C. (2018) Complement factor H in AMD: bridging genetic associations and pathobiology. *Prog. Retin. Eye Res.* **62**, 38–57 [CrossRef Medline](#)
 72. Schmidt, C. Q., Slingsby, F. C., Richards, A., and Barlow, P. N. (2011)

Solution structures of FH Tyr-402 and His-402

- Production of biologically active complement factor H in therapeutically useful quantities. *Protein Expr. Purif.* **76**, 254–263 [CrossRef Medline](#)
73. Sharma, A. K., and Pangburn, M. K. (1996) Identification of three physically and functionally distinct binding sites for C3b in human complement factor H by deletion mutagenesis. *Proc. Natl. Acad. Sci. U.S.A.* **93**, 10996–11001 [CrossRef Medline](#)
74. Kühn, S., and Zipfel, P. F. (1996) Mapping of the domains required for decay acceleration activity of the human factor H-like protein 1 and factor H. *Eur. J. Immunol.* **26**, 2383–2387 [CrossRef Medline](#)
75. Pangburn, M. K. (2000) Host recognition and target differentiation by factor H, a regulator of the alternative pathway of complement. *Immunopharmacology* **49**, 149–157 [CrossRef Medline](#)
76. Knopf, P. M., Rivera, D. S., Hai, S. H., McMurry, J., Martin, W., and De Groot, A. S. (2008) Novel function of complement C3d as an autologous helper T-cell target. *Immunol. Cell Biol.* **86**, 221–225 [CrossRef Medline](#)
77. Kajander, T., Lehtinen, M. J., Hyvärinen, S., Bhattacharjee, A., Leung, E., Isenman, D. E., Meri, S., Goldman, A., and Jokiranta, T. S. (2011) Dual interaction of factor H with C3d and glycosaminoglycans in host-non-host discrimination by complement. *Proc. Natl. Acad. Sci. U.S.A.* **108**, 2897–2902 [CrossRef Medline](#)
78. Kolodziejczyk, R., Mikula, K. M., Kotila, T., Postis, V. L. G., Jokiranta, T. S., Goldman, A., and Meri, T. (2017) Crystal structure of a tripartite complex between C3dg, C-terminal domains of factor H and opSpE of *Borrelia burgdorferi*. *PLoS ONE* **12**, e0188127 [CrossRef Medline](#)
79. Morgan, H. P., Schmidt, C. Q., Guariento, M., Blaum, B. S., Gillespie, D., Herbert, A. P., Kavanagh, D., Mertens, H. D., Svergun, D. I., Johansson, C. M., Uhrin, D., Barlow, P. N., and Hannan, J. P. (2011) Structural basis for engagement by complement factor H of C3b on a self surface. *Nat. Struct. Mol. Biol.* **18**, 463–470 [CrossRef Medline](#)
80. Armstrong, C. T., Mason, P. E., Anderson, J. L., and Dempsey, C. E. (2016) Arginine side chain interactions and the role of arginine as a gating charge carrier in voltage sensitive ion channels. *Sci. Rep.* **6**, 21759 [CrossRef Medline](#)
81. Sim, R. B., Day, A. J., Moffatt, B. E., and Fontaine, M. (1993) Complement factor-I and cofactors in control of complement-system convertase enzymes. *Methods Enzymol.* **223**, 13–35 [CrossRef Medline](#)
82. Schuck, P. (1998) Sedimentation analysis of noninteracting and self-associating solutes using numerical solutions to the Lamm equation. *Biophys. J.* **75**, 1503–1512 [CrossRef Medline](#)
83. Schuck, P. (2000) Size-distribution analysis of macromolecules by sedimentation velocity ultracentrifugation and Lamm equation modeling. *Biophys. J.* **78**, 1606–1619 [CrossRef Medline](#)
84. Sobott, F., Hernández, H., McCammon, M. G., Tito, M. A., and Robinson, C. V. (2002) A tandem mass spectrometer for improved transmission and analysis of large macromolecular assemblies. *Anal. Chem.* **74**, 1402–1407 [CrossRef Medline](#)
85. Okemefuna, A. I., Nan, R., Miller, A., Gor, J., and Perkins, S. J. (2010) Complement factor H binds at two independent sites to C-reactive protein in acute phase concentrations. *J. Biol. Chem.* **285**, 1053–1065 [CrossRef Medline](#)
86. Narayanan, T., Diat, O., and Bosecke, P. (2001) SAXS and USAXS on the high brilliance beamline at the ESRF. *Nucl. Instrum. Methods Phys. Res. A.* **467**, 1005–1009
87. Glatter, O., and Kratky, O. (1982) in *Small Angle X-ray Scattering* (Glatter, O., and Kratky, O., eds) Academic Press, London
88. Wright, D. W., and Perkins, S. J. (2015) SCT: a suite of programs for comparing atomistic models with small-angle scattering data. *J. Appl. Crystallogr.* **48**, 953–961 [CrossRef Medline](#)
89. Semenyuk, A. V., and Svergun, D. I. (1991) Gnom-a program package for small-angle scattering data-processing. *J. Appl. Crystallogr.* **24**, 537–540 [CrossRef](#)
90. Sali, A., and Blundell, T. L. (1993) Comparative protein modelling by satisfaction of spatial restraints. *J. Mol. Biol.* **234**, 779–815 [CrossRef Medline](#)
91. Curtis, J. E., Raghunandan, S., Nanda, H., and Krueger, S. (2012) SASSIE: a program to study intrinsically disordered biological molecules and macromolecular ensembles using experimental scattering restraints. *Comput. Phys. Commun.* **183**, 382–389 [CrossRef](#)
92. Altschul, S. F., Gish, W., Miller, W., Myers, E. W., and Lipman, D. J. (1990) Basic local alignment search tool. *J. Mol. Biol.* **215**, 403–410 [CrossRef Medline](#)
93. Sievers, F., Wilm, A., Dineen, D., Gibson, T. J., Karplus, K., Li, W., Lopez, R., McWilliam, H., Remmert, M., Söding, J., Thompson, J. D., and Higgins, D. G. (2011) Fast, scalable generation of high-quality protein multiple sequence alignments using Clustal Omega. *Mol. Syst. Biol.* **7**, 539 [Medline](#)
94. Shen, M. Y., and Sali, A. (2006) Statistical potential for assessment and prediction of protein structures. *Protein Sci.* **15**, 2507–2524 [CrossRef Medline](#)
95. Barlow, P. N., Norman, D. G., Steinkasserer, A., Horne, T. J., Pearce, J., Driscoll, P. C., Sim, R. B., and Campbell, I. D. (1992) Solution structure of the fifth repeat of factor H: a second example of the complement control protein module. *Biochemistry.* **31**, 3626–3634 [CrossRef Medline](#)
96. Phillips, J. C., Braun, R., Wang, W., Gumbart, J., Tajkhorshid, E., Villa, E., Chipot, C., Skeel, R. D., Kalé, L., and Schulten, K. (2005) Scalable molecular dynamics with NAMD. *J. Comput. Chem.* **26**, 1781–1802 [CrossRef Medline](#)
97. Jo, S., Kim, T., Iyer, V. G., and Im, W. (2008) CHARMM-GUI: a web-based graphical user interface for CHARMM. *J. Comput. Chem.* **29**, 1859–1865 [CrossRef Medline](#)
98. Jo, S., Cheng, X., Islam, S. M., Huang, L., Rui, H., Zhu, A., Lee, H. S., Qi, Y., Han, W., Vanommeslaeghe, K., MacKerell, A. D., Jr, Roux, B., and Im, W. (2014) CHARMM-GUI PDB manipulator for advanced modeling and simulations of proteins containing nonstandard residues. *Adv. Protein Chem. Struct. Biol.* **96**, 235–265 [CrossRef Medline](#)
99. Brooks, B. R., Brooks, C. L., 3rd, Mackerell, A. D., Jr., Nilsson, L., Petrella, R. J., Roux, B., Won, Y., Archontis, G., Bartels, C., Boresch, S., Caflisch, A., Caves, L., Cui, Q., Dinner, A. R., Feig, M., Fischer, S., et al. (2009) CHARMM: the biomolecular simulation program. *J. Comput. Chem.* **30**, 1545–1614 [CrossRef Medline](#)
100. Lee, J., Cheng, X., Swails, J. M., Yeom, M. S., Eastman, P. K., Lemkul, J. A., Wei, S., Buckner, J., Jeong, J. C., Qi, Y., Jo, S., Pande, V. S., Case, D. A., Brooks, C. L., 3rd, MacKerell, A. D., Jr., et al. (2016) CHARMM-GUI input generator for NAMD, GROMACS, AMBER, OpenMM, and CHARMM/OpenMM simulations using the CHARMM36 additive force field. *J. Chem. Theory Comput.* **12**, 405–413 [CrossRef Medline](#)
101. Lovell, S. C., Davis, I. W., Arendall, W. B., 3rd, de Bakker, P. I., Word, J. M., Prisant, M. G., Richardson, J. S., and Richardson, D. C. (2003) Structure validation by α geometry: ϕ, ψ and $C\beta$ deviation. *Proteins* **50**, 437–450 [CrossRef Medline](#)
102. Joosten, R. P., te Beek, T. A., Krieger, E., Hekkelman, M. L., Hooft, R. W., Schneider, R., Sander, C., and Vriend, G. (2011) A series of PDB related databases for everyday needs. *Nucleic Acids Res.* **39**, D411–D419 [CrossRef Medline](#)
103. Kabsch, W., and Sander, C. (1983) Dictionary of protein secondary structure: pattern recognition of hydrogen-bonded and geometrical features. *Biopolymers* **22**, 2577–2637 [CrossRef Medline](#)
104. Zhang, W., Howell, S. C., Wright, D. W., Heindel, A., Qiu, X., Chen, J., and Curtis, J. E. (2017) Combined Monte Carlo/torsion-angle molecular dynamics for ensemble modeling of proteins, nucleic acids and carbohydrates. *J. Mol. Graph. Model.* **73**, 179–190 [CrossRef Medline](#)
105. Watson, M. C., and Curtis, J. E. (2013) Rapid and accurate calculation of small-angle scattering profiles using the golden ratio. *J. Appl. Crystallogr.* **46**, 1171–1177 [CrossRef](#)
106. Humphrey, W., Dalke, A., and Schulten, K. (1996) VMD: visual molecular dynamics. *J. Mol. Graph.* **14**, 33–38, 27–28 [CrossRef Medline](#)
107. The R Core Team. (2013) *R: A Language and Environment for Statistical Computing*, version 3.1.2, R Foundation for Statistical Computing, Vienna, Austria
108. Grant, B. J., Rodrigues, A. P., ElSawy, K. M., McCammon, J. A., and Caves, L. S. (2006) Bio3d: an R package for the comparative analysis of protein structures. *Bioinformatics* **22**, 2695–2696 [CrossRef Medline](#)
109. García-de la Torre, J., Huertas, M. L., and Carrasco, B. (2000) Calculation of hydrodynamic properties of globular proteins from their atomic-level structure. *Biophys. J.* **78**, 719–730 [CrossRef Medline](#)
110. Ortega, A., Amorós, D., and García de la Torre, J. (2011) Prediction of

- hydrodynamic and other solution properties of rigid proteins from atomic- and residue-level models. *Biophys. J.* **101**, 892–898 [CrossRef](#) [Medline](#)
111. Perkins, S. J. (1986) Protein volumes and hydration effects. The calculations of partial specific volumes, neutron scattering matchpoints and 280-nm absorption coefficients for proteins and glycoproteins from amino acid sequences. *Eur. J. Biochem.* **157**, 169–180 [CrossRef](#) [Medline](#)
112. Saunders, R. E., Goodship, T. H., Zipfel, P. F., and Perkins, S. J. (2006) An interactive web database of factor H-associated hemolytic uremic syndrome mutations: insights into the structural consequences of disease-associated mutations. *Hum. Mutat.* **27**, 21–30 [CrossRef](#) [Medline](#)
113. Prosser, B. E., Johnson, S., Roversi, P., Clark, S. J., Tarelli, E., Sim, R. B., Day, A. J., and Lea, S. M. (2007) Expression, purification, cocrystallization and preliminary crystallographic analysis of sucrose octasulfate/human complement regulator factor H SCRs 6–8. *Acta Crystallogr. Sect. F Struct. Biol. Cryst. Commun.* **63**, 480–483 [CrossRef](#) [Medline](#)
114. Makou, E., Mertens, H. D., Maciejewski, M., Soares, D. C., Matis, I., Schmidt, C. Q., Herbert, A. P., Svergun, D. I., and Barlow, P. N. (2012) Solution structure of CCP modules 10–12 illuminates functional architecture of the complement regulator, factor H. *J. Mol. Biol.* **424**, 295–312 [CrossRef](#) [Medline](#)
115. Barlow, P. N., Steinkasserer, A., Norman, D. G., Kieffer, B., Wiles, A. P., Sim, R. B., and Campbell, I. D. (1993) Solution structure of a pair of complement modules by nuclear magnetic resonance. *J. Mol. Biol.* **232**, 268–284 [CrossRef](#) [Medline](#)