BIOCHEMISTRY
including biophysical chemistry & molecular biology

Article

pubs.acs.org/biochemistry

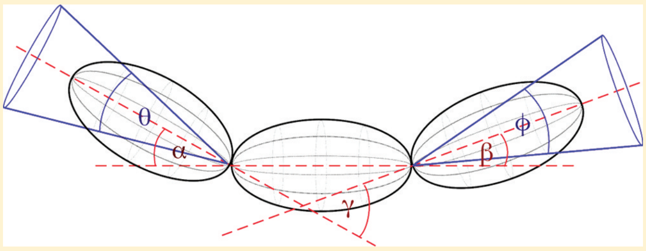
Estimation of Interdomain Flexibility of N-Terminus of Factor H Using Residual Dipolar Couplings

Mateusz Maciejewski,[†,‡] Nico Tjandra,*[,†] and Paul N. Barlow*[,‡]

[†]Laboratory of Molecular Biophysics, National Heart, Lung, and Blood Institute, National Institutes of Health, 50 Center Drive, Bethesda, Maryland 20892, United States

[‡]School of Chemistry, Joseph Black Building, University of Edinburgh, West Mains Road, Edinburgh, Scotland EH9 3JJ

Supporting Information



ABSTRACT: Characterization of segmental flexibility is needed to understand the biological mechanisms of the very large category of functionally diverse proteins, exemplified by the regulators of complement activation, that consist of numerous compact modules or domains linked by short, potentially flexible, sequences of amino acid residues. The use of NMR-derived residual dipolar couplings (RDCs), in magnetically aligned media, to evaluate interdomain motion is established but only for two-domain proteins. We focused on the three N-terminal domains (called CCPs or SCRs) of the important complement regulator, human factor H (i.e., FH1−3). These domains cooperate to facilitate cleavage of the key complement activation-specific protein fragment, C3b, forming iC3b that no longer participates in the complement cascade. We refined a three-dimensional solution structure of recombinant FH1−3 based on nuclear Overhauser effects and RDCs. We then employed a rudimentary series of RDC data sets, collected in media containing magnetically aligned bicelles (disklike particles formed from phospholipids) under three different conditions, to estimate interdomain motions. This circumvents a requirement of previous approaches for technically difficult collection of five independent RDC data sets. More than 80% of conformers of this predominantly extended three-domain molecule exhibit flexions of <40°. Such segmental flexibility (together with the local dynamics of the hypervariable loop within domain 3) could facilitate recognition of C3b via initial anchoring and eventual reorganization of modules to the conformation captured in the previously solved crystal structure of a C3b:FH1−4 complex.

Innate immunity[1] is important for defense against invading pathogens.[2] Its molecular componentry includes many examples of "multidomain proteins". The structural and dynamics properties of such proteins are, in general, poorly understood despite their high abundance and functional diversity.[3−5] In particular, better quantification of interdomain motions within multidomain proteins would shed valuable light on modes of molecular recognition.

The ∼30 plasma proteins of mammalian complement are key to innate immunity and to homeostasis.[6,7] Many of them contain multiple domains. Remarkably, the complement system is not only self-activating (via its "alternative pathway")[8] but also self-regulating. While complement acts aggressively on foreign targets,[9] a subset of complement proteins serves to ensure that host cells and surfaces are protected from complement-mediated attack. Members of this "regulators of complement activation" (RCAs)[10] family consist almost entirely of between four and 30 complement-control protein modules (CCPs), also known as SCRs or sushi domains.[11,12] Varying numbers of CCPs are also present in many other diverse extracellular proteins.[13]

Each CCP has ∼60 residues and two intramodule disulfide bridges.[14] Their C- and N-termini lie at opposite poles of prolate ellipsoid-like tertiary structures, and CCPs are connected, beads-on-a-string fashion, via mostly small intermodular interfaces.[15] Greatest variation occurs in a region termed the hypervariable loop on the side of each module.[16] The flexibility of CCP-containing proteins remains uncharacterized despite being addressed by electron microscopy,[17,18] differential scanning calorimetry,[19,20] NMR,[16,21] and other biophysical techniques.[22−26] A consensus but untested hypothesis is that flexibility varies along the length of the protein in a functionally critical manner.

The key ligand for CCP-containing complement regulators is C3b, an activated C3 cleavage product. A trickle of C3b molecules is continuously and ubiquitously generated as an outcome of the alternative pathway of complement activation.[27] C3b self-promulgates, and the job of selectively preventing C3b from accumulating on host-cell surfaces is performed chiefly by the RCAs. Factor H (FH) (155 kDa),[28] consisting of 20 CCPs, is a prominent soluble RCA. In a cocrystal structure, all four CCPs within an FH N-terminal fragment (FH1−4) contact C3b.[29] The shorter fragment FH1−3 retains at least some

Received: April 15, 2011
Revised: July 26, 2011
Published: July 27, 2011

ability to bind C3b and act as a cofactor for factor I-catalyzed cleavage of C3b to inactive iC3b.[30,31] Modest but significant changes occur in intermodule orientation between NMR-derived FH1−2 or FH2−3 structures, and FH1−4 complexed with C3b (rmsd values = 1.2 or 1.5 Å, respectively, for non-proton backbone atoms). These changes could impose an entropic penalty on binding affinity that does indeed appear weak (K_D = 10 μM)[32] considering the extensive buried surface area in the FH1−4:C3b complex. We have focused on these three N-terminal modules of FH as a test case for the application of NMR to the analysis of flexibility in a protein with multiple domains.

Our approach uses NMR to measure the residual dipolar couplings (RDCs) between nuclei that may be detectable when macromolecules are codissolved with particles that cooperatively align in magnetic fields.[33] RDCs contain information on the orientation of the interaction vectors (i.e., between the coupled nuclei) relative to a common reference alignment frame and hence, indirectly, to each other. Furthermore, implicit in the RDC expressions are dynamic parameters. These are typically absorbed into an overall ensemble scaling but can be deconvoluted. Importantly, these parameters describe motions on a picoseconds-to-milliseconds time scale, thereby pertaining to the interdomain motions (typically, in the nanoseconds time regime[34]) that likely occur within FH upon engagement with its targets.

Estimates of fast local motion of the peptide planes may be obtained from RDCs using numerous approaches, some of which are in direct analogy to those used to parametrize NMR relaxation data.[35−38] As implemented, however, these methods report on motions local to each bond vector (or rigid fragment), not the global motions of domains that are required for proteins such as FH to perform their biological roles. Moreover, these methods require at least five fully orthogonal high-quality sets of RDCs, which may be technically challenging to obtain due to the potential for incompatibility between one or more of the available alignment media and the protein of interest.

Here we illustrate a simplified approach to study global motional properties driven by availability of only a limited set of measured RDCs. We first solved the solution structure of FH1−3 using NOE-derived distance information and RDC measurements recorded in easily prepared media under three different solvent conditions. Then we addressed interdomain flexibility by considering another structural calculation in which ensembles of multiple structures that approximate extremes of motion were determined. While there are several accounts of using NMR-based ensemble approaches to infer ranges of interdomain motions in two-domain systems, most notably by Bertini et al.,[39−41] the current extension to a three-domain system has afforded insights into the rearrangements of the functionally critical triple domain at the N-terminus of FH that accompany interaction with C3b.

■ MATERIALS AND METHODS

Expression and Purification of FH1−3. *Pichia pastoris* strain KM71H transformed with pPICZαB (Invitrogen) ligated with DNA encoding FH1−3 was used for the expression as described previously.[31] ^{15}N-labeled FH1−3 was prepared in a 0.6 L volume in a 2 L fermentor. The cells were spun down, and the supernatant was mixed with ethylenediaminetetraacetic acid and phenylmethanesulfonyl fluoride (final concentrations of 5 and 1 mM, respectively) and filtered. The filtrate was diluted to 4 L with 20 mM sodium acetate buffer at pH 4.0, loaded on a prepacked Source 15S PE 4.6/100 column (GE Healthcare), and then purified using a linear salt gradient to a final concentration of 1 M NaCl. After buffer exchange to pH 6.2, size-exclusion chromatography (HiLoad 16/60 Superdex 75, GE Healthcare) was performed as a final step. The FH1−3 was visualized by Coomassie staining following sodium dodecyl sulfate (SDS)−polyacrylamide gel electrophoresis on a NuPAGE 4−12% Bis-Tris gel (Invitrogen) for validation of identity and purity.

Mesurement of RDCs. For preparation of aligned samples, two aqueous stock solutions of phospholipid bicelles (Avanti Polar Lipids, Alabaster, AL) were formed at 4 °C, both at 15% (w/v) and containing appropriate ratios of 1,2-dimyristoyl-*sn*-glycero-3-phosphocholine (DMPC), 1,2-dihexanoyl-*sn*-glycero-3-phosphocholine (DHPC), and SDS. Stock solution A was composed of a 3:1 molar ratio of DMPC:DHPC, and stock solution B contained a 30:10:1 molar ratio of DMPC:DHPC:SDS.

Three protein samples were prepared for measurement of RDCs. The first contained 0.2 mM [^{15}N]-FH1−3, 3.5% (v/v) of the phospholipid bicelle stock solution A, 20 mM phosphate buffer (pH 6.6), 2 mM arginine and 2 mM glutamine, 10% (v/v) D_2O, and 0.01% (w/v) NaN_3. The second protein sample was identical to the first but contained 3.5% (v/v) of phospholipid bicelle stock solution B instead of A. The third protein sample contained 0.2 mM [^{2}H,^{15}N]-FH1−3 instead of the nondeuterated sample, 3.4% (v/v) of phospholipid bicelle stock solution A, 50 mM phosphate buffer (pH 6.2), 2 mM arginine and 2 mM glutamine, 10% (v/v) D_2O, and 0.01% (w/v) NaN_3.

All spectra were recorded on a Bruker DRX 600-MHz spectrometer equipped with a 5 mm triple-resonance cryoprobe. The reference IPAP-HSQC[42] spectra were recorded at 298 K, the temperature at which the bicelles are in a disordered phase.[43] For the first two samples, spectra were acquired with 1024 by 128 complex points in the ^{1}H and ^{15}N dimension, respectively; for the deuterated sample, spectra contained 563 by 128 complex points. The temperature was then raised to 312 K, and the bicelles underwent a transition to the liquid crystalline phase, which was monitored by observing a stable deuterium splitting (6.70, 6.90, and 9.61 Hz for the first, second, and third sample, respectively).

The acquired spectral data were processed using NMRPipe software[44] and analyzed using Analysis software obtained from the Collaborative Computing Project for the NMR community (CcpNmr).[45] The backbone and side-chain ^{1}H and ^{15}N assignments of module pairs FH1−2 and FH2−3 are available from Biological Magnetic Resonance Data Bank (BMRB).[46] These overlapped quite well with the FH1−3 spectra,[31] and thus assignments were transferred from the pairs to the triple domain (see Supporting Information for details). The data were then converted into PIPP format[47] via CcpNmr FormatConverter software, and RDCs were extracted using an in-house script. Hereafter, the RDC data sets collected from the first, second, and third sample are referred to as set 1, set 2, and set 3, respectively.

Structural Calculations. All NMR structural calculations were performed using Xplor-NIH software[48] version 2.24. As a starting structure for the RDC and NOE-based calculations a solution structure of FH1−3 was derived from melding together previously solved structures of FH1−2 and FH2−3 as described previously.[31] The sequence of events in the subsequent structural calculations commenced with an initial Cartesian minimization, followed by 10 ps of high-temperature dynamics in torsion-angle space at 4000 K and subsequent slow cooling to the final temperature of 100 K. During cooling, NOE and RDC force constants were geometrically ramped to their final values. The last step consisted of torsion angle minimization followed by Cartesian space minimization. In those calculations CHARMM19 and CHARMM20 topologies, parameters, and force fields were used.[48]

The NOE list for the triple module (FH1−3) structure calculations was derived from the list of NOE restraints for FH1−2 and FH2−3 as retrieved from the Brookhaven Protein Data Bank (PDB) entries 2RLP and 2RLQ.[31] The overlap of NOEs in the mutual CCP 2 module was resolved by inspection. If such NOEs involved residues that were deemed close to where the missing module (i.e., CCP 1 or CCP 3) would have been, they were excluded. Hence, NOEs involving residues His87-Thr99, Lys109-Glu128, and Thr135-Val143 were removed from the FH1−2 NOE list, while NOEs from residues Lys82-Gly86, Gly100-Val108 (with the exception of NOEs between residues Thr99 and Gly100), and Cys129-Trp134 were removed from the FH2−3 list (details in Supporting Information). The remaining NOEs (4154; see Table 1 for categories) were concatenated into a single table of restraints for FH1−3 and incorporated into Xplor-NIH calculations. All three RDC sets were introduced with a harmonic target function and final force constant of 0.5 kcal mol^{-1} Hz^{-2}, with the exception of regions deemed as mobile,[31] which included the particularly dynamic Gly60 as well as residues Thr34-Ser38, Thr99-Val103, and Ser160-Glu167 constituting the three hypervariable loops. In these cases, the relevant RDCs were included with a half-open square-well target function. For each set, starting rhombicities and magnitudes of alignment tensors were first extracted from the histograms, and then grid searches were performed around the values obtained from this estimation. A total of 100 structures were calculated with the parameters outlined above.

Ensemble Calculations. For motional analysis, ensemble calculations were conducted in Xplor-NIH.[49] The two lowest-energy structures generated using NOEs (as described in the last section, see Table 1) and the three sets of RDCs were utilized as input for these calculations. Four separate calculations were performed. In each of the first two calculations, one of the two single structures was used as input, while in each of the third and fourth calculations, both structures were used simultaneously as a two-member input ensemble. In each of the four calculations 100 ensembles (each consisting of either one or two structures) were generated using the same restraints as in the original calculation, under similar force constants. For the fourth calculation only, an additional shape restraint was applied in the neighborhood of intermodular linkers (residues Pro76-Gly89 and Asp137-Thr150) as well as the whole molecule (residues Asp20-Val206), with 20 kcal mol^{-1} Å^2 force constant on both the size and the orientation components.[48] Each of the four calculations involved eight repetitions of a loop consisting of 1000 steps of dynamics at 400 K followed by 100 cooling steps from 400 to 300 K and a torsion angle space minimization followed by a Cartesian minimization. At all times, the energies from experimental restraints were computed as the uniformly weighted averages of energies from each individual ensemble member.

Estimation of Interdomain Dynamics. For FH1−3 structures, the degree of flexion in the 20% of structures with the lowest energy (from the two-member ensemble calculation) was assessed as follows. First, every second member of each ensemble was least-squares fit on CCP 2 of every first member of the same ensemble using the XtalView package program pdbfit.[50] Then the orientation of each domain was calculated for every member of every ensemble by summation of the N−C_α and C_α−C′ interatomic vectors of the regions that were

Table 1. Structure-Calculation Statistics for the 30 Lowest-Energy NOE/RDC-Refined NMR Structures

summary of NOEs used for calculation	CCP 1		CCP 2		CCP 3
sequential	439		363		355
mid-range $2 \leq \lvert i - j\rvert \leq 4$	74		90		43
long range >4	589		475		504
module to module		20		8	
module 1 to linkers		64			
module 2 to linkers		50		65	
module 3 to linkers				40	

rmsd (superposition of 30 lowest-energy structures)	CCP 1	CCP 2	CCP 3	all
all heavy atoms (Å)	0.99	0.77	0.89	1.12
backbone heavy atoms (Å)	0.47	0.41	0.40	0.79

rmsd (the lowest-energy structure, backbone heavy atoms)	CCP 1	CCP 2	CCP 3
vs starting model (Å)	0.88	0.92	0.87
vs crystal structure (Å)	0.76	1.07	1.00

RDC summary	set 1	set 2	set 3
D_a	−10.68 ± 0.29	−10.08 ± 0.42	−9.11 ± 0.28
R_h	0.19 ± 0.03	0.59 ± 0.06	0.29 ± 0.03
Q-factor	0.19	0.27	0.17

intermodular angles (between inertia axes)	CCP 1-CCP 2	CCP 2-CCP 3
x-axis (deg)	26.0 ± 9.2	39.2 ± 2.6
y-axis (deg)	31.9 ± 2.9	12.2 ± 4.6
z-axis (deg)	18.1 ± 4.7	38.5 ± 3.2

Ramachandran assessment[a]

most favored (%): 58.8

additionally allowed (%): 35.3

generously allowed (%): 5.2

disallowed (%): 0.7

buried surface area[b]

between modules 1 and 2 (Å^2): 547.39 ± 16.15

between modules 2 and 3 (Å^2): 501.48 ± 29.34

[a] The Ramachandran report was prepared using PROCHECK-NMR software.[51] Note that no dihedral restraints were used in the structure calculation. [b] Surface areas were calculated using a PyMOL[59] function getarea, using dot probes at density setting "3". They were obtained by analyzing the difference in areas of the isolated modules and the surface areas of module pairs.

previously shown to be well structured in the X-ray structure of the FH1−4:C3b complex or the NMR structures of FH1−2 and FH2−3 (residues Glu31−Thr34, Gln47-Cys52, Tyr56-Arg57, Val62-Arg67, Glu70-Ala73 in CCP 1; Gly94-Thr99, Lys109-Cys114, Tyr118-Leu121, Tyr126-Glu128, Gly133-Trp134, Ile140-Cys141 in CCP 2; and Gly155−Ile157, Ala173-Cys178, Glu189-His191 in CCP 3). For every ensemble, relative orientations of the domains were devised using these vectors. The average orientations from each ensemble were then used to derive the mean interdomain orientations along with their standard deviations. Next, two angles for each ensemble were calculated: the first angle between CCP 1 of the first ensemble member and CCP 1 of the second ensemble member (which is equivalent to the range of CCP 1−CCP 2 angles within the ensemble; see Figure 3) and the second angle between CCP 3 of the first ensemble member and CCP 3 of the second ensemble member (equivalent to the range of CCP 2−CCP 3 angles within the ensemble; see Figure 3). The ensembles were sorted and placed in bins according to their interdomain angle variability (based on the aforementioned ranges of angles), i.e., those with a value lower than 40°, lower than 30° lower than 20°, and lower than 10°.

Additionally, flexibility between terminal modules (CCP 1−CCP 3) was established by first calculating their average relative orientation (as above) for each ensemble, and then a mean orientation among the 20 lowest-energy ensembles was derived. Subsequently the mean flexibility (and the corresponding standard deviation) was established by considering all of the differences in CCP 1−CCP 3 angles between the first and second member of each of the 20 lowest-energy ensembles.

■ RESULTS

The NOE (Only)-Derived Structure of FH1−3 Is Potentially Under-Defined. In an initial exploratory calculation, we created an NMR-derived solution structure of FH1−3 by fusing together structures of FH1−2 and FH2−3 (as described previously[31]), and we subjected it to simulated annealing on the basis of 4154 previously assigned[31] NOE-derived upper distance bounds. We did not use RDC-derived restraints at this point.

The 30 lowest-energy structures that resulted from this calculation were judged as being potentially under-defined because they showed no consistent violations of NOE-derived distances, yet the root mean-square deviation (rmsd) upon overlaying their heavy backbone atoms (excluding hypervariable loop residues) was 3.2 Å. Similar overlays performed only on CCPs 1 and 2, or on CCPs 2 and 3, yielded RMSDs of 2.0 and 2.2 Å, respectively. These high values reflect low interdomain angle convergence since the same structures overlaid on heavy backbone atoms of individual modules CCP 1, CCP 2, or CCP 3 (excluding hypervariable loops and intermodular linker residues, i.e., Gln81-Pro84 and Glu142-Lys145) yielded much lower RMSDs of 0.80, 0.60, and 0.75 Å, respectively.

Incorporation of Three Sets of RDCs into Structure Calculation Yields Better Convergence. Residual dipolar couplings can be expected to decrease the likelihood of under-definition in NMR-derived structure calculations, which commonly arises from a relative paucity of intermodular NOEs as was observed in this case (see Table 1) with, for example, only 8 such NOEs detected between CCPs 2 and 3. Thus, the potentially under-restrained nature of the NOE (only)-derived structures of FH1−3 prompted the utilization within the structure calculation of three N−H^N RDC data sets recorded on FH1−3 in easily prepared bicelle-aligned media. Sets 1, 2, and 3 contained 138, 131, and 167 values, respectively (listed in Supporting Information).

Table 1 shows relevant statistics for 30 representative (lowest-energy) structures of FH1−3 derived from refinement of the "starting" (NOE-only based) model using both NOEs and the three sets of N−H^N RDCs. The alignment tensor for each data set was simultaneously refined in these calculations. The intradomain rmsd values of 0.47, 0.41, and 0.40 Å for CCP 1, CCP 2 and CCP 3, respectively, (excluding hypervariable loops and intermodular linker residues) were significantly better than in the NOE-only calculation. Additionally, the interdomain angles were better defined; a heavy backbone atom overlay (excluding the hypervariable regions) of all three modules simultaneously yielded an rmsd of 0.79 Å. Considering immediately neighboring domains (i.e., CCP 1 and CCP 2 or CCP 2 and CCP 3), the variability of the intermodular angles (Table 1) is <∼10° for the lowest-energy set of 30 FH1−3 structures, compared to ∼20° for NOE-only structures (data not shown). No consistent violations of experimental restraints were observed, implying that, within the limit of recorded data, all 30 structures are equally valid approximations of the time-averaged conformation of FH1−3.

Intermodular flexibility will in part be determined by the packing of amino acid side chains within the buried interface between modules (see Figure 4). The buried surface areas determined here (see Table 1) for the lowest-energy structures were slightly lower than those calculated in case of the starting model.[31] They were nevertheless similar to values commonly found for extended CCP pairs.[12]

A Ramachandran assessment revealed that only 0.7% of residues were in the disallowed regions[51] (Table 1), which is normally regarded as an indication of acceptable sterochemical quality. The reason for a relatively low proportion of residues being found in the most favored and additionally allowed regions is the fact that neither dihedral restraints nor knowledge-based potentials were used. The former are typically incorporated under square-well potentials such that no penalty is applied over a broad range of angles; these would be ineffectual in the subsequent stage of calculations (the ensemble calculations). Ensemble averaging would also very easily satisfy a knowledge-based potential that would therefore, likewise, have little effect.

RDC-Refined FH1−3 Structures Are Dissimilar to Structures Derived from NOEs (Only) and X-ray. The lowest-energy representative from this NOE/RDC-refined set of structures was found to be globally dissimilar both to the starting model and to a model of FH1−3 derived from the crystal structure of FH1−4 bound to C3b (PDB entry 2WII);[29,31] rmsds (backbone heavy atoms) were 2.7 and 2.8 Å, respectively. Indeed, while individual domains overlaid well (summarized in Table 1), relative domain orientations varied significantly. This is readily appreciated from Figure 1c, where the three structures are superimposed on the heavy backbone atoms of CCP 3. The rmsds of such overlays are (including all heavy backbone atoms) 9.9 Å in the case of the starting model and 18 Å for the (2WII-derived) crystal structure. These differences in interdomain orientations are

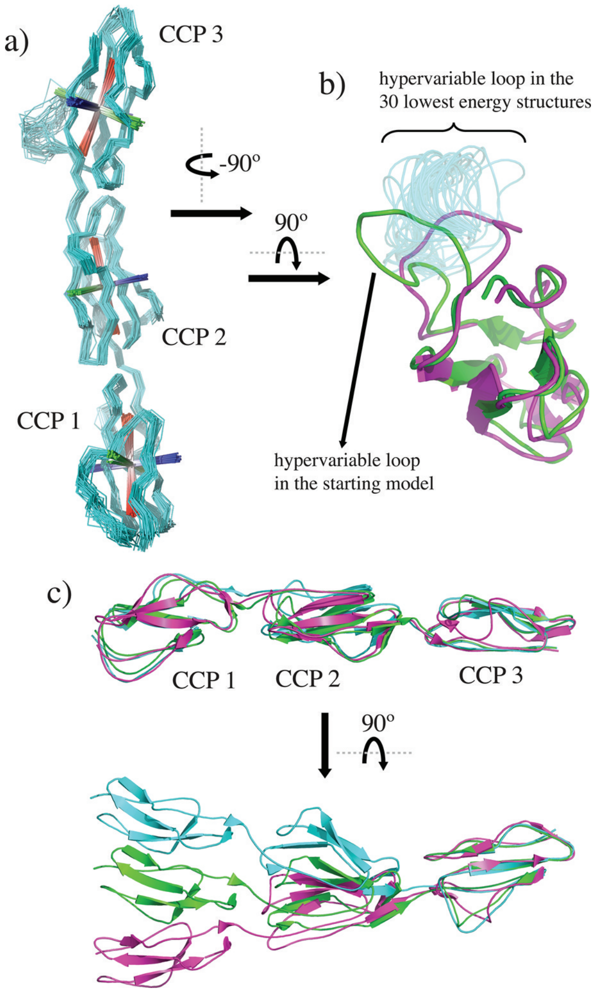


Figure 1. Structure of FH1−3 derived from NOEs and RDCs. (a) The least-squares heavy backbone atoms overlay of the NMR-derived 30 lowest-energy structures (ribbon representation). Orientations of principal axes of the inertia tensor of each domain are presented as cylinders originating from its center of mass. (b) For clarity only CCP 3 is shown, and in the case of the lowest-energy NMR structures only, the CCP 3 hypervariable loop is shown in cyan. The hypervariable loop of the third CCP is shown on the top left, the starting (NOE-only) model is represented by a green ribbon, and the crystal structure is represented by a magenta ribbon. In the starting model, the loop kinks to the left as viewed in the figure, while in the crystal structure the loop kinks to the right and approximates an extreme position of the loop in the NOE/RDC-derived structures. (c) Two views of heavy backbone atoms overlay on CCPs 3 of lowest-energy FH1−3 structure (cyan) with the starting model (green) and CCPs 1−3 of the crystal structure (magenta), rotated by 90° about the long axis. For clarity the structures are represented as smoothened ribbons. All structure visualizations were prepared in PyMOL.[59]

accommodated without major rearrangements of residues at the inter-CCP interface (Figure 4), consistent with the very few and small chemical shift differences (comparing e.g. residues in the FH1−2 interdomain interface with equivalents in FH1−3, see Supporting Information Figure S2) and with the lack of violations of relevant NOEs transplanted from FH1−2 and FH2−3 to the triple-domain construct. Note in particular that all of the 20 and 8 NOEs between CCPs 1 and 2, and CCPs 2 and 3, respectively, used in the pre-RDC structure calculations are satisfied in all of the 30 representative NOE/RDC-refined structures.

A further notable feature (see Figure 1b) is the dissimilarity of the hypervariable loop of CCP 3 (residues Ser160-Gly167) in the set of 30 lowest-energy NOE/RDC-derived structures versus both the starting model and a 2WII-inferred crystal structure of FH1−3. When, ignoring modules 1 and 2, these structures are overlaid on the backbone atoms of CCP 3 (excluding the loop), rmsds of 1.90 Å (NOE/RDC-refined versus starting model) and 2.07 Å (NOE/RDC-refined versus crystal structure) are obtained for the backbone heavy atoms of the eight residues in the hypervariable loop. Its position (see Figure 1b) within the 30 new structures lies between that of the starting model and that of the X-ray derived structure, although the majority of its conformational variations resemble that found in the crystal structure.

Alignment Tensor Analysis Suggests Interdomain Flexibility Is Absent or Low. Histograms for the three collected sets of RDCs (Figure 2) were inspected to estimate the independence of these three data sets. Set 1 and set 2 differed from set 3 in the magnitude of the alignment tensor (which measures the degree of alignment of the protein molecules), and all three sets differed in their population maxima. It was therefore anticipated that their rhombicities (which measure the asymmetry of the rhombic components of the alignment tensor) would be different. The same histograms subsequently served as a basis upon which to estimate the rhombicity and magnitude of the alignment tensors[52] for each set of RDCs; this was carried out for all three domains combined as well as for each domain separately. To obtain more accurate estimates of rhombicities and magnitudes of these tensors, grid searches were performed against the starting model[31] (see Materials and Methods) and starting values derived from respective data sets.

From the distribution of RDCs it was not clear whether one tensor per data set would suffice or whether each data set should be broken into three separate tensors (one per domain). This ambiguity is visualized in Figure 2—the domain-wise distributions of RDCs seem to be quite similar, but the extent of sampling limits definitive judgment on this issue. The most complete of the available RDC sets (set 3) was used to test this. First, one tensor for each of the three domains was used, and then one tensor was used for the whole molecule. The *Q*-factors (which measure the overall agreement of the measured RDCs with the protein structure[53]) are parametrized as

$$Q = \mathrm{rms}(\Delta\delta^{\mathrm{meas}} - \Delta\delta^{\mathrm{pred}})/\mathrm{rms}(\Delta\delta^{\mathrm{meas}})$$

where $\Delta\delta^{\mathrm{meas}}$ denotes a measured RDC and $\Delta\delta^{\mathrm{pred}}$ denotes a back-calculated RDC. The *Q*-factors were found to be 0.06, 0.10, and 0.09 for CCP 1, CCP 2, and CCP 3, respectively. For comparison, the *Q*-factor was 0.10 when a single tensor was used for all three domains. Thus, the improvement of the *Q*-factors when three tensors were used was not significant. The optimized alignment tensor components for individual domains varied by a maximum of 18%. While these relatively small variations are suggestive of possible domain motion, this need not be the case since they are also sensitive to sampling distribution within each domain and the quality of the other structural restraints.

The fact that a single alignment tensor could be used to describe RDCs in all three domains is consistent with the amount of interdomain motion being either none or very small. However, there remains the possibility that the single conformation inferred from the measured RDCs could in fact represent an average conformation of the molecule that is undergoing significant interdomain motion. To estimate how much motion was indeed present, the ensemble approach was utilized.

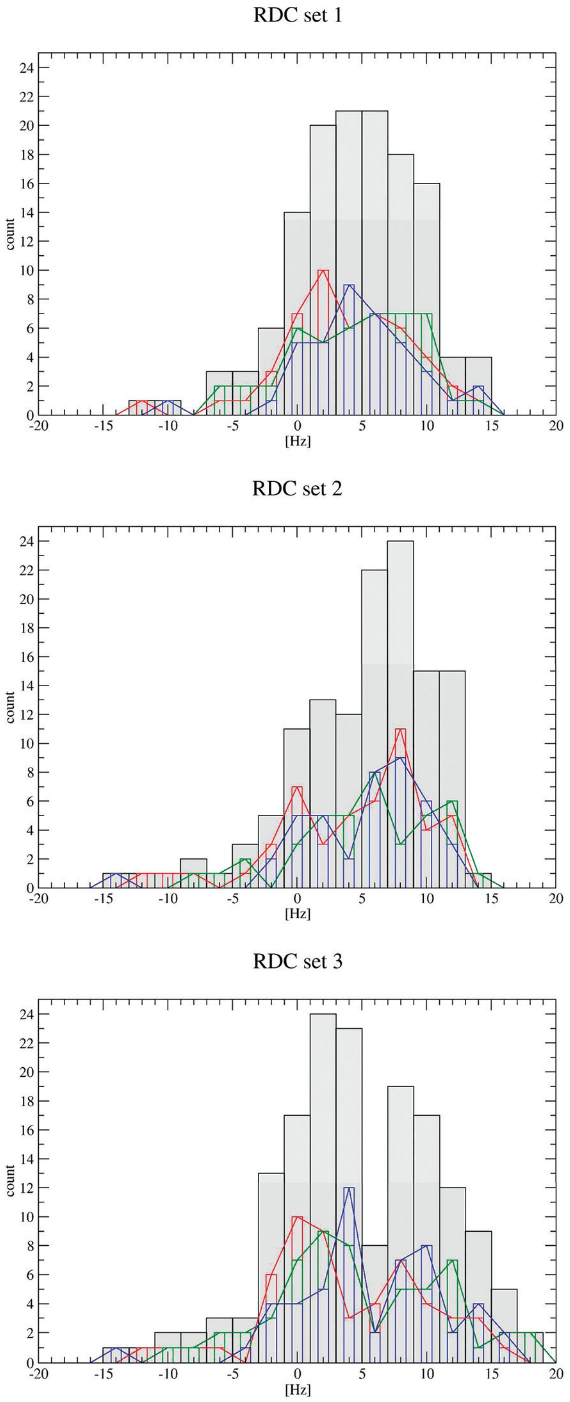


Figure 2. Histograms to show the distributions of recorded N–H^N RDC data sets. The gray histograms show the distributions of RDCs in all three domains, the red histograms show their distributions in CCP 1, the green histograms show their distributions in CCP 2, and the blue ones show the RDC distributions in CCP 3. A bin size of 2 Hz was used in all histograms.

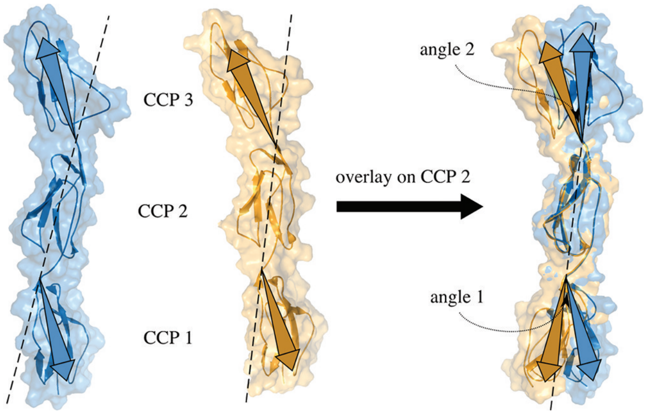


Figure 3. Two structures in blue and orange represent two members of a two-member ensemble. On the left-hand side they are viewed in the orientations emerging directly from the original structure calculation (but translated relative to one another for clarity). On the right-hand side they are shown after least-squares fitting on CCPs 2. The blue and brown arrows that represent the long axes of the N- and C-terminal domains of the ensembles are defined as described in Materials and Methods. The ranges of motion on each of the intermodular junctions are described as viewed on the right-hand side of the figure; an angle of flexure is measured between the two "arrows" on each of the terminal domains.

An Ensemble Approach to Interdomain Motional Analysis. In order to estimate the range of motion attainable in the system, the ensemble protocols were used within Xplor-NIH,[54] as described in Materials and Methods. In the ensemble approach two structures (the ensemble) that approximate to the two extremes of anisotropic motion are considered simultaneously at all times; this differs from standard calculations that consider only one structure.

Step one aimed to ascertain whether the ensemble approach describes motion better than standard structural calculations. For these purposes, the two lowest-energy structures obtained from the NOE/RDC-refined NMR structure calculations constituted the ensemble. Thus, initially these two structures were used as input for two separate standard calculations. Subsequently, a similar calculation was performed but with the two structures simultaneously used as input, i.e., serving as a two-member ensemble. The structure-calculation protocol was similar whether one or two input structures were used (see Materials and Methods). In all cases the RDC restraints were introduced under similar force constants as in the case of the NOE/RDC-based structural refinements. The average *Q*-factors of the RDCs in the 20 lowest-energy one-member and two-member "ensembles" obtained from these two approaches were then compared. For RDC set 3 the *Q*-factor was 46% lower using the two-member ensemble approach. Similarly, the *Q*-factor was 20% lower and 25% lower for RDC sets 2 and 1, respectively. Thus, it appeared that an ensemble size of two was more suitable for analysis of motion in the system under consideration than the use of single starting structures (i.e., an effective ensemble size of one). This is consistent with the existence of motion between modules.

In step two, a calculation was performed to show that the range of measured motion was not overestimated by the above analysis. To this effect a shape potential was used[54] with a relatively low force constant, to reduce the spread of energetically feasible interdomain angles in the calculations. A zero-mass inertia tensor was employed to create such potential, which ensured that during the calculation the shapes of ensemble members did not diverge to a high degree from their mean shape. These calculations were compared with an identical calculation conducted without the shape potential. For the resulting 20 lowest-energy two-member ensembles the

Q-factor of the RDC set 3 was on average 35% higher when the shape envelope was applied compared to when it was not. The corresponding values for RDC set 2 and RDC set 1 were increased by 12% and 13%, respectively. Thus, even moderate restraints on the amplitude of interdomain motions produced structures that were in poorer agreement with the measured RDC data than those produced in the calculation devoid of such restraints. This implies that the RDC terms had been appropriately adjusted in the ensemble calculations, such that the range of motion they unveiled was not overestimated. For that reason shape potentials with higher force constants were not tested.

Quantification of Interdomain Flexibility in FH1−3. The 20 lowest-energy two-member ensembles, derived without the shape potential, were analyzed in terms of interdomain angles and their variability. As summarized in Figure 5, the mean tilt angle, among all of the analyzed ensembles, between the long axes (defined as described in Materials and Methods) of CCPs 1 and 2 was 29 ± 5°; for CCPs 2 and 3, the equivalent value was 20 ± 7°. The diversity of the interdomain tilt angles within an ensemble was subsequently quantified by first performing a least-squares fit of each FH1−3 ensemble on CCPs 2; then the angle between the long axes of CCPs 1 was measured, as was the angle between the long axes of CCPs 3. The resulting two values reflect the range of interdomain tilt angles for CCP 1−CCP 2 and CCP 2−CCP 3, respectively.

The average range of tilt angles between CCP 1 and CCP 2 was 20° (± a standard deviation of 6°) while for CCP 2 and CCP 3 the equivalent value was 17° (±15°). In both cases, however, the distributions of the said parameter deviated significantly from a normal distribution (data not shown), and thus the values are better described in terms of categories (in Figure 5 embraced by cones θ and ϕ) as follows: the range of CCP 1−CCP 2 tilt angles was less than 40° within all ensembles, and was less than 30° within 95%, less than 20° within 50%, and less than 10° within 5% of ensembles; in the case of CCP 2−CCP 3 the equivalent value was less than 40° within 90% of ensembles, and was less than 30° also within 90%, less than 20° within 75%, and less than 10° within 30% of ensembles.

This suggests that, in the case of each of the two interdomain junctions, the maximum range of flexion is approximately the same at about 40° and that most motion is restricted to between 10° and 30° (depicted in Figures 5 and 6b). The presence of a similar extent of flexion at both intermodular junctions was anticipated on the grounds that the extent of interdomain contacts was similar and the spread of RDCs for each domain was comparable (Figure 2). This result was reproduced when a different pair of the lowest-energy RDC and NOE-refined NMR-derived structures of FH1−3 were chosen for input in the ensemble calculations. Moreover, similar ranges of flexional freedom were obtained when calculations were performed on larger ensembles (of four and eight structures, see Supporting Information), despite the lower *Q*-factors, again suggesting that the RDC restraints were applied under well-adjusted force constants.

The size and variability (within ensembles) of CCP 1−CCP 3 tilt angles was also determined (Figure 5). In the 20 lowest-energy ensembles produced by the calculation, the mean tilt angle is 46 ± 9°; the average angle variance over the 20 ensembles is 24 ± 13°. Individual ensemble members in only 10% of cases display a CCP 1−CCP 3 tilt angle lower than the corresponding tilt of 29° measured in the crystal structure; nonetheless, it can be readily appreciated in Figure 6a,b that the exact crystal conformation is never achieved in the ensembles. This can be explained when the interdomain angles between the neighboring modules are considered.

DISCUSSION

Segmental flexibility is likely a dominant feature of the numerous proteins that are composed from multiple domains connected end-to-end in a potentially extended arrangement. We set out to measure flexibility in the N-terminal three domains (FH1−3) of complement regulator FH as a prelude to exploring the hypothesis that interdomain flexibility is required to perform one or more of its biological roles. Many residues throughout FH1−3 contribute to an extensive interface formed when FH binds to C3b. Subsequently, these C3b-bound modules recruit factor I that then cleaves C3b to iC3b. Additionally, binding of this region of FH (along with CCP 4) to C3b in the C3 convertase (C3b.Bb) context accelerates the irreversible decay of C3b.Bb to C3b and Bb. We have explored the use of a relatively simple, NMR-based strategy for defining the limits of flexure within FH1−3.

We addressed the dynamics in FH1−3 based on ensemble interpretation of RDC data in a manner similar to Clore and Schwieters.[54] Dynamic properties are typically difficult to capture in simulated annealing protocols as they do not account for the averaging of the experimental observables; i.e., the calculations are not generally performed on an ensemble of structures, but rather on a "one structure at a time" basis. The approach suggested by Clore and Schwieters[54] involves implementation of a routine for minimization of ensembles of structures against RDCs in a simulated annealing protocol. A similar methodology was employed in the work presented here; however, our focus was not on motion at the single-residue level, but rather on putative flexibility between domains likely required for proper protein function. Indeed, neither exact correlation times of global motion nor order parameters were derived; only ranges of interdomain motion were estimated. Intermodular dynamics could be analyzed by more rigorous means if required, but this would entail appropriate formalism by analogy with progression of model-free formalism into the extended model-free formalism introduced by Clore et al.[55] and developed by Chen and Tjandra.[34] Such approaches would still suffer from difficulties typical for RDC measurements, like the requirement for five orthogonal alignments.

The approach adopted here to the study of FH1−3 is analogous to "ensemble-optimized methods" (EOM) for analysis of SAXS data,[56] which yield estimates of global flexibility. An NMR-based strategy devised by Bertini et al.[39] for analysis of interdomain motion in calmodulin is also similar to the one employed in the current work. It should be noted however that in the calmodulin study, homogeneously weighted averages of RDC values were first calculated based on multiple structures differing in domain orientations. Alignment tensors were then derived based on the comparison of such averages with experimental measurements. Nonviable (unrealistic) orientations could pass for viable ones in such an approach because of the homogeneity of weighting of RDCs calculated based on the structures selected from the pool. We avoided this difficulty by generation of realistic ensembles of structures with the help of multiple restraint types. The calmodulin study inspired Longinetti et al.[57] to adopt an inverse approach in the develop-

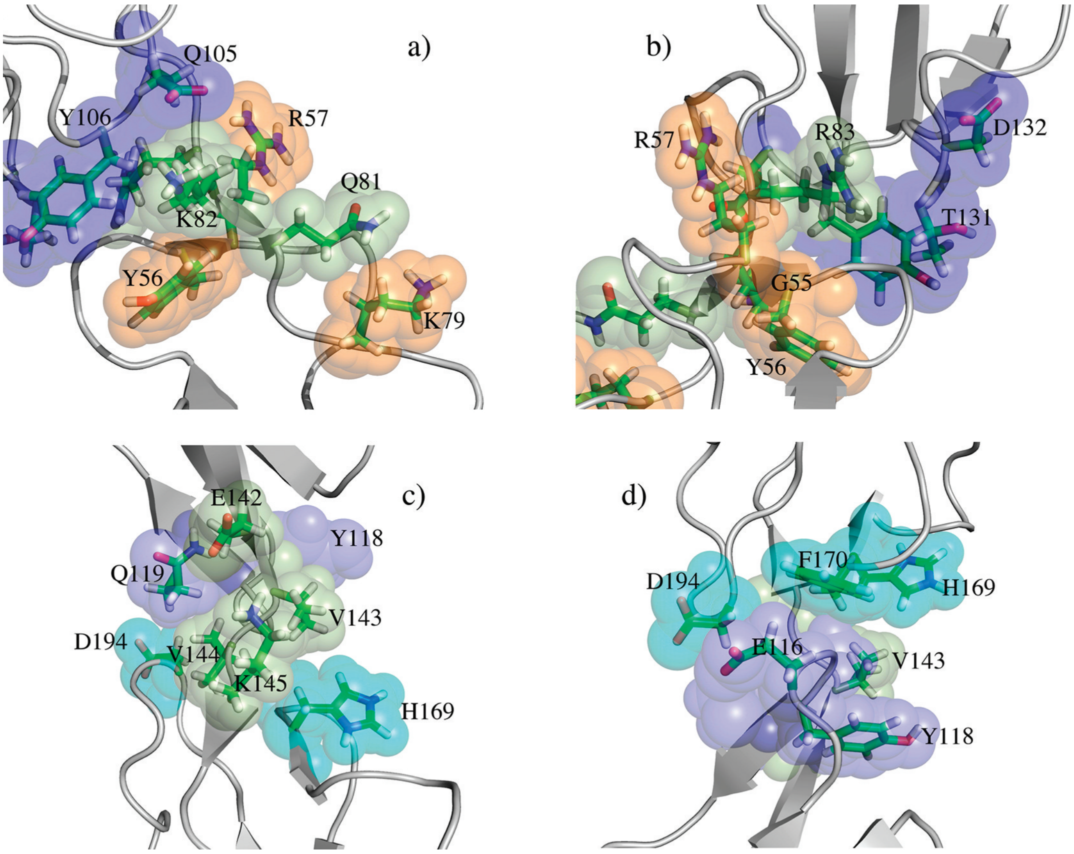


Figure 4. Representations of the linker regions of the lowest-energy NOE/RDC-derived structure. The side chains of important residues are represented by sticks and spheres. Orange spheres were used for CCP 1 (residues before Gln81), blue for CCP 2 (residues between Cys85 and Cys141, inclusive), and cyan for CCP 3 residues (those beyond Lys145). Olive spheres represent the linker residues (Gln81-Pro84 in the CCP 1−CCP 2 linker and Glu142-Lys145 in the CCP 2−CCP 3 linker). (a) Gln81 is shown to be in a rather crowded area, and close to Lys79 and Lys82 is in the proximity of Tyr56 and Glu105, and also sterically hindered. (b) Arg83 is tethered between Arg57, Glu132, and Thr131. The charged moieties of those residues are close to one another. (c) Val144 is highly buried. Val143 is constrained by nearby bulky groups as well as Lys145 and Glu142 that may interact with one another through their charges. (d) Same area as in the previous panel, but rotated by 180° about the *x*-axis. Upon interdomain motion Asp194 could clash with Gln119 and Glu116; Phe170 resides in a generally crowded position that could hamper motion, Val143 is additionally in the neighborhood of bulky Tyr118 and His169. The above panels illustrate how a network of interactions stabilizes the linkers and prevents fully unconstrained (completely isotropic) motion.

ment of “maximum allowed probability” (determined by a rotation matrix) values for various interdomain orientations, in which different conformations were assigned heterogeneous weighting factors. Such weighting was determined according to a protocol based on a simplex search of the orientational maximum allowed probability followed by simulated annealing, resulting in groups of ten conformers with weights (rotations and translations) minimized against a target function;[57] this virtually excluded unrealistic orientations from the ensemble. Such heterogeneous weighting factors were implicitly present in the approach used here. This is because, during generation of structural ensembles, violations of any introduced restraints should drive the structures away from less populated conformations. Nevertheless, since only one type of RDC restraints ($N-H^N$) was used in our calculations, the estimates of motion presented here are best regarded as upper limits on flexure of the terminal domains of the studied molecule (see Figures 5 and 6b).

The time-averaged NOE and RDC-derived NMR structures served as input for the ensemble calculations. We utilized NOEs previously collected on FH1−2 and FH2−3, along with three new sets of ($N-H^N$) RDCs (Figure 2) collected for FH1−3, to refine an existing structural model that had been built by concatenating NMR-derived structures of FH1−2 and FH2−3. The resulting structures were of good quality and back-calculated RDCs were in excellent agreement with the experimental values (Table 1). Good convergence was obtained at the level of individual modules; these also overlaid well with equivalent individual modules in the original starting model of Hocking et al.[31] and in the crystal structure of the FH1−4:C3b complex. On the other hand, comparing these RDC and NOE-derived structures with the FH1−4:C3b structure revealed that interdomain orientations were significantly different (Figure 1c). That these different orientations were accommodated without creating violations of the NOEs, transplanted from FH1−2 and FH2−3, confirmed the under-definition of NOE(only)-derived structure calculations of FH1−3.

A further (“final”) set of NOE and RDC-derived structures of FH1−3 was calculated using an ensemble-based method and was subjected to analysis of flexure (bending, in any direction, of the long axes of the prolate ellipsoid-shaped modules away from the time-averaged conformation). The volumes explored by the long axes of CCPs 1 and 3 relative to CCP 2 lie predominantly in cones (with their central axes coincident with the principal inertia tensors of the time-averaged conformations of the respective domains; see Figure 5) that have opening

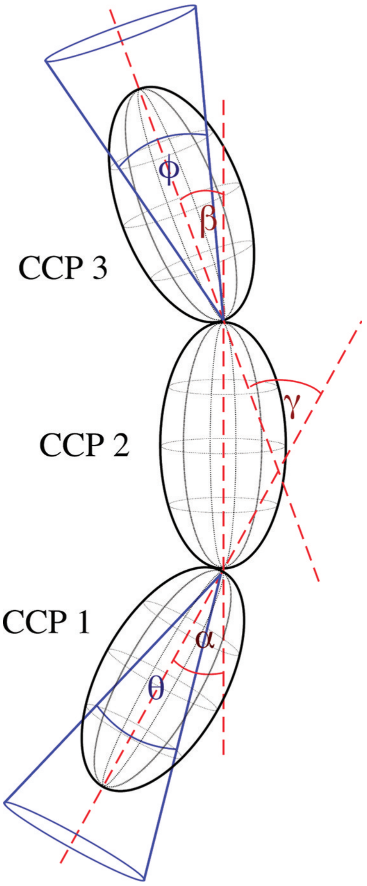


Figure 5. Summary of the flexibility analysis of lowest-energy structures resulting from the ensemble calculations. The three ellipsoids are a schematic representation of the domains of FH1−3, and the red dashed lines represent their long axes. The mean interdomain orientation $\alpha = 29 \pm 5°$, $\beta = 20 \pm 7°$, and $\gamma = 46 \pm 9°$. The interdomain tilt angle variability among structures falls within a cone of (full) angle $\theta < 40°$ in 100% structures; <30° in 95%, <20° in 50%, <10° in 5% of structures; or $\phi < 40°$ and <30° in 90% of structures; and <20° in 75%, <10° in 30% of structures.

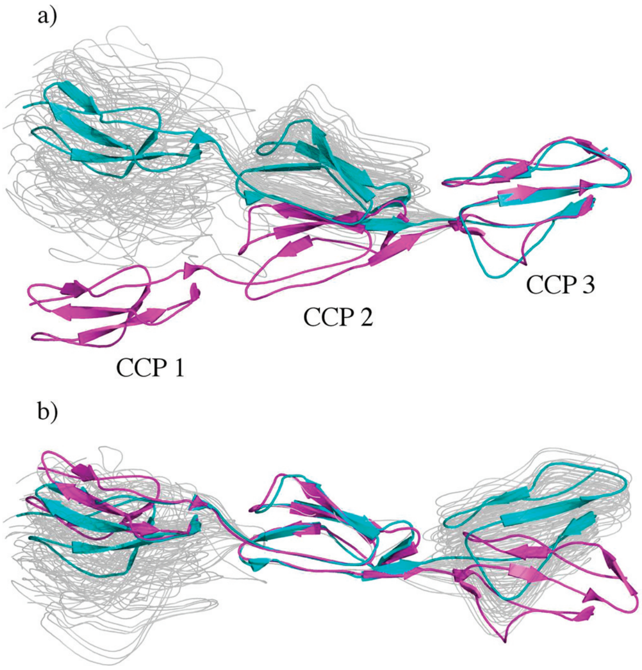


Figure 6. CCPs 1−3 of FH in crystal structure (magenta), lowest-energy NOE/RDC refined structure (cyan), and structural ensembles within the 40° cone representing flexional freedom (gray). (a) Least-squares fit on CCP 3 of the crystal structure. (b) Least-squares fit on CCP 2 of the crystal structure. None of the solution structures reproduces the crystal structure conformation.

(full) angles lower than 40°. The observed conformational variability is sufficient to embrace a conformation close to that found in the X-ray structure of FH1−4 bound to C3b (Figure 6). On the other hand, the refined NMR-derived solution structures of FH1−3 presented here are more extended than the equivalent regions in the ensemble of structural models obtained previously from analysis of SAXS data recorded for FH1−5 wherein CCP 1−CCP 2 tilt angles are 54.3 ± 19.5° and CCP 2−CCP 3 tilt angles are 63.1 ± 30.8°.[26]

In addition to establishing upper bounds on flexure our data shed light on intradomain motion. Interestingly, the inferred conformational mobility within the hypervariable loop in CCP 3 of FH1−3 is insufficient to accommodate the conformation of the equivalent loop in the FH1−4:C3b crystal structure (Figure 1b) that contacts the CUB domain of C3b.

Taking together these data pertaining to interdomain and intradomain motion has allowed us to formulate a hypothetical mode of engagement of C3b by the N-terminal region of FH. We suggest that an initial, possibly electrostatically driven, encounter between these two large proteins involves one of the domains of FH 1−3. The flexure we have observed between FH domains 1−3 aids a subsequent interdomain rearrangement into the positions observed in the cocrystal structure that buries a large surface area; note that minor rearrangements of domains in C3b also appear to accompany the binding of FH1−4.[27,29] Subsequently, the flexible hypervariable loop of CCP 3 may contribute to fidelity by binding, perhaps through an induced fit mechanism, to its cognate site on the CUB domain. The entropic cost of such rearrangements might offset enthalpic gains, thus resulting in an interaction with only medium (although adequate) affinity, but high specificity. Further work is needed to elucidate the thermodynamic contributions to complex formation in order to test this hypothesis. A further repercussion of interdomain flexure within FH, which could be explored in future work, is that the rearrangement of domains in the FH:C3b complex might create a new binding interface for recruitment and activation of factor I.

In conclusion, we have described an NMR-based strategy that yields estimates of interdomain flexibility that are difficult to access by other techniques, although it would be interesting to compare our results to the outcome of EOM[56] for fitting of SAXS curves (previously performed SAXS studies did not adopt the EOM approach[26]). We showed our method to be useful for establishing the extent of flexibility between two globular domains that are connected via short linkers to either end of a central globular domain (as exemplified by FH1−3). Such an arrangement is common within numerous multiple-domain proteins; about 8% of multicellular proteomes are reckoned to consist of proteins like FH in which repeating domains (domains belonging to the same family) occur adjacent to one another.[5] In many cases several tandem domains comprise an intact ligand-binding site. But interdomain flexibility has further, profound, repercussions for FH. FH1−3 cooperates[32,58] with at least three other sites elsewhere in the FH molecule to protect host tissue via a sophisticated multivalent host-surface recognition mechanism. These include an additional binding site for C3b at its C-terminus, 16 CCPs away from the N-terminal C3b-binding site. Other complement regulators contain as many as three discrete C3b-binding sites dispersed over 30 CCPs. Most hypotheses regarding the little understood mode of action of these proteins invoke variations in flexibility along the length of the protein and these ought to be tested, preferably through studies of longer segments of protein containing more than three domains. These could be achieved without substantial modifications to the methodology described herein. Providing assignments and RDC data are available, they could, as in the current work, be interpreted in the ensemble context to generate structures that represent the extent of motion. It would, however, be necessary in each case to determine which ensemble size is appropriate and whether additional shape restraints are required. Inclusion of more varied restraints at the phase of generation of ensembles, such as SAXS curves or different types of RDC restraints (C_α−N, C′−N, C_α−H_α) would increase the accuracy and precision of the results.

■ ASSOCIATED CONTENT

Ⓢ Supporting Information

Figure S1 portraying the violations by the FH1−3 structure of the NOEs originating from the FH1−2 and FH2−3 restraint lists that had been discarded (prior to structure calculation) due to the fact they originated from residues close to the C-terminus of FH1−2 (potentially proximal to where CCP 3 is attached) or from the N-terminus of FH 2−3 (potentially proximal to where CCP 1 is attached); Figure S2 illustrating the largest amide (^{1}H and ^{15}N) chemical shift differences between FH1−3 and FH1−2 and FH2−3; Table S1 summarizing the assignment of the FH1−3 ^{15}N-HSQC spectrum; Table S2 listing recorded RDC sets; Table S3 showing the ranges of interdomain tilt angles in ensembles of various sizes and corresponding Q-factors. This material is available free of charge via the Internet at http://pubs.acs.org.

■ AUTHOR INFORMATION

Corresponding Author

*Phone +1 (301)-402-3029, Fax +1 (301)-402-3405, e-mail tjandran@nhlbi.nih.gov (N.T.); Phone +44 (0)131 650 7546, Fax +44 (0)131 650 6453, e-mail paul.barlow@ed.ac.uk (P.N.B.).

Funding

This work was supported by the Intramural Research Program of the NIH, NHLBI, to N.T. and a Wellcome Trust/NIH PhD studentship to M.M.

■ ACKNOWLEDGMENTS

We thank Dr. Henry Hocking for making available the deuterated protein samples that were used to record RDC set 3 and Dr. Charles Schwieters for useful discussions. We acknowledge the professional skills and advice of Dr. Grzegorz Piszczek (Biophysics Core Facility, NHLBI, National Institutes of Health). We thank Dr. Christoph Schmidt and Dr. Marie-Paule Strub for advice regarding *Pichia pastoris* protein expression and protein purification. We thank Juraj Bella for NMR-related advice.

■ ABBREVIATIONS

CCP, complement control protein; FH, factor H; rmsd, root mean-square deviation; BMRB, Biological Magnetic Resonance Data Bank; PDB, Brookhaven Protein Data Bank; RCA, regulator of complement activation; NMR, nuclear magnetic resonance; NOE, nuclear Overhauser effect; RDC, residual dipolar coupling; SAXS, small-angle X-ray scattering; DMPC, 1,2-dimyristoyl-*sn*-glycero-3-phosphocholine; DHPC, 1,2-dihexanoyl-*sn*-glycero-3-phosphocholine; SDS, sodium dodecyl sulfate; IPAP-HSQC, in-phase/anti-phase heteronuclear single quantum correlation.

■ REFERENCES

(1) Litman, G. W., Cannon, J. P., and Dishaw, L. J. (2005) Reconstructing immune phylogeny: new perspectives. *Nat. Rev. Immunol. 5*, 866−879.

(2) Hoebe, K., Janssen, E., and Beutler, B. (2004) The interface between innate and adaptive immunity. *Nature Immunol. 5*, 971−974.

(3) Apic, G., Gough, J., and Teichmann, S. A. (2001) Domain combinations in archaeal, eubacterial and eukaryotic proteomes. *J. Mol. Biol. 310*, 311−325.

(4) Bork, P., Downing, A. K., Kieffer, B., and Campbell, I. D. (1996) Structure and distribution of modules in extracellular proteins. *Q. Rev. Biophys. 29*, 119−167.

(5) Ekman, D., Björklund, A. K., Frey-Skött, J., and Elofsson, A. (2005) Multi-domain proteins in the three kingdoms of life: orphan domains and other unassigned regions. *J. Mol. Biol. 348*, 231−243.

(6) Walport, M. J. (2001) Complement. *N. Engl. J. Med. 344*, 1058−1066.

(7) Ricklin, D., Hajishengallis, G., Yang, K., and Lambris, J. D. (2010) Complement: a key system for immune surveillance and homeostasis. *Nature Immunol. 11*, 785−797.

(8) Lachmann, P. J. (2009) The Amplification Loop of the Complement Pathways. *Adv. Immunol. 104*, 115−149.

(9) Kirkitadze, M. D., and Barlow, P. N. (2001) Structure and flexibility of the multiple domain proteins that regulate complement activation. *Immunol. Rev. 180*, 146−161.

(10) Hourcade, D., Holers, V. M., and Atkinson, J. P. (1989) The regulators of complement activation (RCA) gene cluster. *Adv. Immunol. 45*, 381−416.

(11) Reid, K., and Day, A. J. (1989) Structure-function relationships of the complement components. *Immunol. Today 10*, 177−180.

(12) Soares, D. C., and Barlow, P. N. (2005) Complement Control Protein Modules in the Regulators of Complement Activation, in *Structural Biology of the Complement System*, pp 19−62, CRC Press, Boca Raton, FL.

(13) Soares, D. C., Gerloff, D. L., Syme, N. R., Coulson, A. F. W., Parkinson, J., and Barlow, P. N. (2005) Large-scale modelling as a route to multiple surface comparisons of the CCP module family. *Protein Eng., Des. Sel. 18*, 379−388.

(14) Barlow, P. N., Baron, M., Norman, D. G., Day, A. J., Willis, A. C., Sim, R. B., and Campbell, I. D. (1991) Secondary structure of a complement control protein module by two-dimensional 1H NMR. *Biochemistry 30*, 997−1004.

(15) Norman, D.G., Barlow, P. N., Baron, M., Day, A. J., Sim, R. B., and Campbell, I. D. (1991) Three-dimensional structure of a complement control protein module in solution. *J. Mol. Biol. 219*, 717−725.

(16) Barlow, P. N., Stainkasserer, 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.

(17) Weisman, H.F., Bartow, T., Leppo, M. K., Marsh, H. C., Carson, G. R., Concino, M. F., Boyle, M. P., Roux, K. H., Weisfeldt, M. L., and

Fearon, D. T. (1990) Soluble Human Complement Receptor Type 1: In Vivo Inhibitor of Complement Suppressing Post- Ischemic Myocardial Inflammation and Necrosis. *Science 249*, 146−151.

(18) DiScipio, R. G. (1992) Ultrastructures and Interactions of Complement Factors H and I. *J. Immunol. 149*, 2592−2599.

(19) Kirkitadze, M. D., Krych, M., Uhrín, D., Dryden, D. T. F., Smith, B. O., Cooper, A., Wang, X., Hauhart, R., Atkinson, J. P., and Barlow, P. N. (1999) Independently Melting Modules and Highly Structured Intermodular Junctions within Complement Receptor Type 1. *Biochemistry 38*, 7019−7031.

(20) Kirkitadze, M. D., Dryden, D. T. F., Kelly, S. M., Price, N. C., Wang, X., Krych, M., Atkinson, J. P., and Barlow, P. N. (1999) Cooperativity between modules within a C3b-binding site of complement receptor type 1. *FEBS Lett. 459*, 133−138.

(21) Henderson, C. E., Bromek, K., Mullin, N. P., Smith, B. O., Uhrín, D., and Barlow, P. N. (2001) Solution structure and dynamics of the central CCP module pair of a poxvirus complement control protein. *J. Mol. Biol. 307*, 323−339.

(22) Kirkitadze, M. D., Henderson, C., Price, N. C., Kelly, S. M., Mullin, N. P., Parkinson, J., Dryden, D. T., and Barlow, P. N. (1999) Central modules of the vaccinia virus complement control protein are not in extensive contact. *Biochem. J. 344*, 167−175.

(23) 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.

(24) Furtado, P. B., Huang, C. Y., Ihyembe, D., Hammond, R. A., Marsh, H. C., and Perkins, S. J. (2008) The partly folded back solution structure arrangement of the 30 SCR domains in human complement receptor type 1 (CR1) permits access to its C3b and C4b ligands. *J. Mol. Biol. 375*, 102−118.

(25) 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 Tyr402His disease-related polymorphism and its interactions with heparin. *J. Mol. Biol. 368*, 564−581.

(26) 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.

(27) Janssen, B. J. C., Christodoulidou, A., McCarthy, A., Lambris, J. D., and Gros, P. (2006) Structure of C3b reveals conformational changes that underlie complement activity. *Nature 444*, 213−216.

(28) Ferreira, V. P., Herbert, A. P., Hocking, H. G., Barlow, P. N., and Pangburn, M. K. (2006) Critical role of the C-terminal domains of factor H in regulating complement activation at cell surfaces. *J. Immunol. 177*, 6308−6316.

(29) Wu, J., Wu, Y.-Q., Ricklin, D., Janssen, B. J. C., Lambris, J. D., and Gros, P. (2009) Structure of complement fragment C3b−factor H and implications for host protection by complement regulators. *Nature Immunol. 10*, 728−733.

(30) Gordon, D. L., Kaufman, R. M., Blackmore, T. K., Kwong, J., and Lublin, D. M. (1995) Identification of complement regulatory domains in human factor H. *J. Immunol. 155*, 348−356.

(31) Hocking, H. G., Herbert, A. P., Kavanagh, D., Soares, D. C., Ferreira, V. P., Pangburn, M. K., Uhrín, 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.

(32) Schmidt, C. Q., Herbert, A. P., Kavanagh, D., Gandy, C., Fenton, C. J., Blaum, B. S., Lyon, M., Uhrín, D., and Barlow, P. N. (2008) A New Map of Glycosaminoglycan and C3b Binding Sites on Factor H. *J. Immunol. 181*, 2610−2619.

(33) Tjandra, N., and Bax, A. (1997) Direct Measurement of Distances and Angles in Biomolecules by NMR in a Dilute Liquid Crystalline Medium. *Science 278*, 1111−1114.

(34) Chen, K., and Tjandra, N. (2008) Extended Model Free Approach To Analyze Correlation Functions of Multidomain Proteins in the Presence of Motional Coupling. *J. Am. Chem. Soc. 130*, 12745−12751.

(35) Meiler, J., Prompers, J. J., Peti, W., Griesinger, C., and Brüschweiler, R. (2001) Model-Free Approach to the Dynamic Interpretation of Residual Dipolar Couplings in Globular Proteins. *J. Am. Chem. Soc. 123*, 6098−6107.

(36) Lipari, G., and Szabo, A. (1982) Model-free approach to the interpretation of nuclear magnetic resonance relaxation in macromolecules. 1. Theory and range of validity. *J. Am. Chem. Soc. 104*, 4546−4559.

(37) Tolman, J. R. (2002) A Novel Approach to the Retrieval of Structural and Dynamic Information from Residual Dipolar Couplings Using Several Oriented Media in Biomolecular NMR Spectroscopy. *J. Am. Chem. Soc. 124*, 12020−12030.

(38) Lakomek, N.-A., Walter, K. F. A., Farès, C., Lange, O. F., Groot, B. L., Grubmüller, H., Brüschweiler, R., Munk, A., Becker, S., Meiler, J., and Griesinger, C. (2008) Self-consistent residual dipolar coupling based model-free analysis for the robust determination of nanosecond to microsecond protein dynamics. *J. Biomol. NMR 41*, 139−155.

(39) Bertini, I., Bianco, C. D., Gelis, I., Katsaros, N., Luchinat, C., Parigi, G., Peana, M., Provenzani, A., Zoroddu, M. A., and Gray, H. B. (2004) Experimentally Exploring the Conformational Space Sampled by Domain Reorientation in Calmodulin. *Proc. Natl. Acad. Sci. U.S.A. 101*, 6841−6846.

(40) Bertini, I., Gupta, Y. K., Luchinat, C., Parigi, G., Peana, M., Sgheri, L., and Yuan, J. (2007) Paramagnetism-based NMR restraints provide maximum allowed probabilities for the different conformations of partially independent protein domains. *J. Am. Chem. Soc. 129*, 12786−12794.

(41) Bertini, I., Giachetti, A., Luchinat, C., Parigi, G., Petoukhov, M. V., Pierattelli, R., Ravera, E., and Svergun, D. I. (2010) Conformational Space of Flexible Biological Macromolecules from Average Data. *J. Am. Chem. Soc. 132*, 13553−13558.

(42) Ottiger, M., Delaglio, F., and Bax, A. (1998) Measurement of J and Dipolar Couplings from Simplified Two-Dimensional NMR Spectra. *J. Magn. Reson. 131*, 373−378.

(43) Losonczi, J., and Prestegard, J. (1998) Improved dilute bicelle solutions for high-resolution NMR of biological macromolecules. *J. Biomol. NMR 12*, 447−451.

(44) Delaglio, F., Grzesiek, S., Vuister, G., Zhu, G., Pfeifer, J., and Bax, A. (1995) NMRPipe: A multidimensional spectral processing system based on UNIX pipes. *J. Biomol. NMR 6*, 277−293.

(45) Vranken, W. F., Boucher, W., Stevens, T. J., Fogh, R. H., Pajon, A., Llinas, M., Ulrich, E. L., Markley, J. L., Ionides, J., and Laue, E. D. (2005) The CCPN data model for NMR spectroscopy: Development of a software pipeline. *Proteins 59*, 687−696.

(46) Ulrich, E. L., Akutsu, H., Doreleijers, J. F., Harano, Y., Ioannidis, Y. E., Lin, J., Livny, M., Mading, S., Maziuk, D., Miller, Z., Nakatani, E., Schulte, C. F., Tolmie, D. E., KentWenger, R., Yao, H., and Markley, J. L. (2007) BioMagResBank. *Nucleic Acids Res. 36*, D402−D408.

(47) Garrett, D. S., Powers, R., Gronenborn, A. M., and Clore, G. M. (1991) A common sense approach to peak picking in two-, three-, and four-dimensional spectra using automatic computer analysis of contour diagrams. *J. Magn. Reson. 95*, 214−220.

(48) Schwieters, C. D., Kuszewski, J. J., Tjandra, N., and Marius Clore, G. (2003) The Xplor-NIH NMR molecular structure determination package. *J. Magn. Reson. 160*, 65−73.

(49) Schwieters, C. D., Kuszewski, J. J., and Marius Clore, G. (2006) Using Xplo NIH for NMR molecular structure determination. *Prog. Nucl. Magn. Reson. Spectrosc. 48*, 47−62.

(50) McRee, D. E. (1999) XtalView/Xfit - A Versatile Program for Manipulating Atomic Coordinates and Electron Denisty. *J. Struct. Biol. 125*, 156−165.

(51) Laskowski, R. A., MacArthur, M. W., Moss, D. S., and Thornton, J. M. (1993) PROCHECK: a program to check the stereochemical quality of protein structures. *J. Appl. Crystallogr. 26*, 283−291.

(52) Clore, G. M., Gronenborn, A. M., and Bax, A. (1998) A Robust Method for Determining the Magnitude of the Fully Asymmetric Alignment Tensor of Oriented Macromolecules in the Absence of Structural Information. *J. Magn. Reson. 133*, 216−221.

(53) Cornilescu, G., Marquardt, J. L., Ottiger, M., and Bax, A. (1998) Validation of protein structure from anisotropic carbonyl chemical shifts in a dilute liquid crystalline phase. *J. Am. Chem. Soc. 120*, 6836−6837.

(54) Clore, G. M., and Schwieters, C. D. (2004) How Much Backbone Motion in Ubiquitin Is Required To Account for Dipolar Coupling Data Measured in Multiple Alignment Media as Assessed by Independent Cross-Validation? *J. Am. Chem. Soc. 126*, 2923−2938.

(55) Clore, G. M., Szabo, A., Bax, A., Kay, L. E., Driscoll, P. C., and Gronenborn, A. M. (1990) Deviations from the simple two-parameter model-free approach to the interpretation of nitrogen-15 nuclear magnetic relaxation of proteins. *J. Am. Chem. Soc. 112*, 4989−4991.

(56) Bernadó, P., Mylonas, E., Petoukhov, M. V., Blackledge, M., and Svergun, D. I. (2007) Structural Characterization of Flexible Proteins Using Small-Angle X-ray Scattering. *J. Am. Chem. Soc. 129*, 5656−5664.

(57) Longinetti, M., Luchinat, C., Parigi, G., and Sgheri, L. (2006) Efficient determination of the most favoured orientations of protein domains from paramagnetic NMR data. *Inverse Probl. 22*, 1485.

(58) Morgan, H. P., Schmidt, C. Q., Guariento, M., Blaum, B. S., Gillespie, D., Herbert, A. P., Kavanagh, D., Mertens, H. D. T., Svergun, D. I., Johansson, C. M., Uhrín, 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.

(59) The PyMOL Molecular Graphics System, Version 1.4, Schrödinger, LLC.