The central portion of factor H (modules 10-15) is compact and contains a structurally deviant CCP module

Citation for published version:

Digital Object Identifier (DOI):
10.1016/j.jmb.2009.10.010

Link:
Link to publication record in Edinburgh Research Explorer

Document Version:
Publisher's PDF, also known as Version of record

Published in:
Journal of Molecular Biology

Publisher Rights Statement:
Available under Open Access

General rights
Copyright for the publications made accessible via the Edinburgh Research Explorer is retained by the author(s) and / or other copyright owners and it is a condition of accessing these publications that users recognise and abide by the legal requirements associated with these rights.

Take down policy
The University of Edinburgh has made every reasonable effort to ensure that Edinburgh Research Explorer content complies with UK legislation. If you believe that the public display of this file breaches copyright please contact openaccess@ed.ac.uk providing details, and we will remove access to the work immediately and investigate your claim.
The Central Portion of Factor H (Modules 10–15) Is Compact and Contains a Structurally Deviant CCP Module

Christoph Q. Schmidt¹, Andrew P. Herbert¹, Haydyn D. T. Mertens², Mara Guariento¹, Dinesh C. Soares¹,³, Dusan Uhrin¹, Arthur J. Rowe⁴, Dmitri I. Svergun² and Paul N. Barlow¹∗

¹Edinburgh Biomolecular NMR Unit, Centre for Chemical and Translational Biology, Schools of Biological Sciences and Chemistry, University of Edinburgh, Edinburgh EH9 3JJ, UK
²European Molecular Biology Laboratory—Hamburg Outstation, c/o DESY, Notkestrasse 85, 22603 Hamburg, Germany
³Medical Genetics Section, Molecular Medicine Centre, Institute of Genetics and Molecular Medicine, Western General Hospital, University of Edinburgh, Crewe Road, Edinburgh EH4 2XU, UK
⁴NCMH Business Centre, School of Biosciences, University of Nottingham, Sutton Bonington, Leicestershire LE12 5RD, UK

Received 1 September 2009; received in revised form 1 October 2009; accepted 8 October 2009
Available online 14 October 2009

The first eight and the last two of 20 complement control protein (CCP) modules within complement factor H (fH) encompass binding sites for C3b and polyanionic carbohydrates. These binding sites cooperate self-surface selectively to prevent C3b amplification, thus minimising complement-mediated damage to host. Intervening fH CCPs, apparently devoid of such recognition sites, are proposed to play a structural role. One suggestion is that the generally small CCPs 10–15, connected by longer-than-average linkers, act as a flexible tether between the two functional ends of fH; another is that the long linkers induce a 180° bend in the middle of fH. To test these hypotheses, we determined the NMR-derived structure of fH12–13 consisting of module 12, shown here to have an archetypal CCP structure, and module 13, which is uniquely short and features a laterally protruding helix-like insertion that contributes to a prominent electropositive patch. The unusually long fH12–13 linker is not flexible. It packs between the two CCPs that are not folded back on each other but form a shallow vee shape; analytical ultracentrifugation and X-ray scattering supported this finding. These two techniques additionally indicate that flanking modules (within fH11–14 and fH10–15) are at least as rigid and tilted relative to neighbours as are CCPs 12 and 13 with respect to one another. Tilts between successive modules are not unidirectional; their principal axes trace a zigzag path. In one of two arrangements for CCPs 10–15 that fit well with scattering data, CCP 14 is folded back onto CCP 13. In conclusion, fH10–15 forms neither a flexible tether nor a smooth bend. Rather, it is compact and has embedded within it a CCP module (CCP 13) that appears to be highly specialised given both its deviant structure and its striking surface charge distribution. A passive, purely structural role for this central portion of fH is unlikely.

© 2009 Elsevier Ltd. All rights reserved.

Keywords: complement system; short consensus repeat; NMR structure; small-angle X-ray scattering; analytical ultracentrifugation

Abbreviations used: CCP, complement control protein; fH, complement factor H; fB, complement factor B; SANS, small-angle X-ray scattering; AUC, analytical ultracentrifugation; HSQC, heteronuclear single quantum coherence; NOE, nuclear Overhauser effect; PBS, phosphate-buffered saline; PDB, Protein Data Bank; EOM, even-and-odd mode; TOCSY, total correlated spectroscopy; SA, solvent-accessible surface area.

*Corresponding author. Joseph Black Chemistry Building, University of Edinburgh, West Mains Road, Edinburgh EH9 3JJ, UK. E-mail address: Paul.Barlow@ed.ac.uk.
Introduction

The complement system is a major component of mammalian immune defences.\(^1,2\) Complement activation not only leads to assembly of lytic membrane attack complexes\(^3\) but also augments acquired immunity and generates mediators of inflammation and opsonisation. The phylogenetically ancient complement system employs comparatively rudimentary mechanisms for discrimination between self and non-self.\(^4,5\) These are based largely upon the regulators of complement activation\(^6,7\) protein family, of which complement factor H (fH) is a prominent member.\(^8\) Mutations and polymorphisms in the fH gene are linked to kidney diseases (atypical haemolytic uraemic syndrome and dense deposit disease) and age-related macular degeneration (reviewed by de Cordoba and sy" syndrome and dense deposit disease) and age-related macular degeneration (reviewed by de Cordoba and de Jorge\(^9\)) and, tentatively, to heightened risks of Alzheimer’s disease\(^10,11\) and myocardial infarction.\(^10,11\)

Factor H is an abundant 155-kDa (1213-amino-acid residue) plasma glycoprotein.\(^12\) It competes with the pro-enzyme complement factor B (fB) for binding to molecules of the pivotal complement protein C3b. C3b, a proteolytic fragment of C3, is generated on a continuous low-level basis by the alternative pathway of complement. Nascent C3b readily binds covalently to any nearby surfaces, where it triggers several inflammatory events.\(^13\) Binding of fB to C3b and subsequent cleavage of fB yield C3b.Bb, which is a C3 convertase. This bimolecular enzyme cleaves C3 to form further molecules of C3b, thereby stoking a positive feedback loop that rapidly amplifies the number of C3b molecules.\(^14\) Binding of fH to C3b, on the other hand, mediates destruction of C3b by factor I. Factor H is also able to accelerate an irreversible disassembly of C3b.Bb, providing further regulatory activity. Through poorly understood mechanisms, fH engages more effectively with C3b (or C3b.Bb) that is bound to self-surfaces bearing polyanionic markers such as glycosaminoglycans and sialic acids,\(^3,4,5\) than with these same proteins when they are non-self. These are based largely upon the regulators of complement activation\(^6,7\) protein family, of which complement factor H (fH) is a prominent member.\(^8\) Mutations and polymorphisms in the fH gene are linked to kidney diseases (atypical haemolytic uraemic syndrome and dense deposit disease) and age-related macular degeneration (reviewed by de Cordoba and de Jorge\(^9\)) and, tentatively, to heightened risks of Alzheimer’s disease\(^10,11\) and myocardial infarction.\(^10,11\)

Factor H is a prominent member.\(^7\) Mutations and polymorphisms in the fH gene are linked to kidney diseases (atypical haemolytic uraemic syndrome and dense deposit disease) and age-related macular degeneration (reviewed by de Cordoba and de Jorge\(^9\)) and, tentatively, to heightened risks of Alzheimer’s disease\(^10,11\) and myocardial infarction.\(^10,11\)

The intact fH protein molecule has not been crystallised and is too large for structure determination by nuclear magnetic resonance (NMR). It is unusually composed entirely of numerous\(^6,16,17\) examples of the complement control protein (CCP) module (\(\sim\) 60 residues, also known as a short consensus repeat or sushi domain);\(^18\) there are 20 CCPs in fH orthologues from mice to humans (Fig. 1). Three-dimensional structures of 12 of the fH CCPs (in fragments containing up to four CCPs), showing that they approximate to prolate spheroids of \(\sim 3.8\) nm by \(\sim 1.6\) nm, have been reported.\(^21\) Transmission electron microscopy,\(^22\) small-angle X-ray scattering (SAXS), and analytical ultracentrifugation (AUC)\(^23\) of fH are consistent with a molecule of about half the length that would be expected, were all CCPs stretched out in a fully extended head-to-tail arrangement. Thus, each module might be strongly tilted with respect to its neighbours, or the chain of CCPs might double back on itself at a bend imposed by the middle section of the molecule.

Factor H has two C3b-binding regions (Fig. 1), one towards either end of the molecule.\(^24–27\) The main binding sites for C3b lie within CCPs 1–4 (with CCPs 1–3 retaining residual binding affinity\(^28\)) and CCPs 19–20. These two sites must cooperate, since the surface-plasmon-resonance-derived binding affinity of intact fH for immobilised C3b is 0.6–1.6 \(\mu\)M, which is significantly tighter than either site alone (4–10 \(\mu\)M).\(^29\) Other potential C3b-binding sites—encapsulated in fH6–8\(^25\) and fH18–15\(^27\)—exhibited only weak or very weak affinities, respectively, according to a recent report.\(^27\) To help discriminate between C3b and C3b.Bb according to context (on self-surfaces versus non-self-surfaces), two additional binding sites on fH (one within CCPs 6–8\(^30,31\) and the other within CCPs 19–20\(^32,33\)) recognise polyanions such as glycosaminoglycans and sialic acids.\(^34,35\) These ligand-binding regions correspond to locations of disease-linked single-nucleotide polymorphisms and mutations\(^3,8\)—e.g., Y402H in module 7 is a major risk factor for age-related macular degeneration,\(^36\) whilst many mutations in CCPs 20 are linked to atypical haemolytic uraemic syndrome.\(^37\)

An important question as to how the four principal ligand-binding sites encompassed in CCPs 1–8 and 19–20 work in a cooperative fashion remains.\(^21\) Intervening modules are likely to be critical in this respect. An often-noted feature of the fH central region is the preponderance of relatively small modules joined by comparatively long linking sequences (Fig. 1; for simplicity in this work, we consider module boundaries as the first and the last of the four conserved cytosteins, unless otherwise stated). This, taken together with the absence of strong evidence for binding sites amongst these modules, has led to speculation that the central CCPs of fH act primarily as a flexible tether between N-terminal and C-terminal regions constraining, but not directing, the spatial distance between them; an orthogonal viewpoint is that the central CCPs cause the fH molecule to bend back on itself in a defined way, thus organising N and C termini into a multivalent composite recognition and binding site.\(^23,30,38,39\) In previous work, a version of fH from which modules 11–15 had been deleted proved not to be fully active but retained some limited functionalities.\(^25\) Its affinity for C3b was not quantified, however; thus, whether its N-terminal and C-terminal functional regions could cooperate is not known.

We set out to address the extent to which the coincidence of long linkers and small modules constitutes a functionally relevant architectural feature at the centre of the fH molecule. We therefore determined the solution structure and dynamics of the 12th and 13th CCPs of factor H (fH12–13), since this segment contains both the smallest module (CCP 13 contains just 51 residues) and the longest intermodular linker (eight residues) within the entire RCA family. This information was supplemented
with low-resolution structural data for segments of \( fH \) extended by one or two modules on either side of \( fH_{12-13} \) (i.e., \( fH_{11-14} \) and \( fH_{10-15} \)). We obtained evidence that, far from introducing flexibility, long linkers induce defined tilt angles between CCP modules.

**Results**

**Solution NMR studies reveal a rigid \( fH_{12-13} \) structure**

Good-quality NMR spectra of \( ^{15}\)N-labelled and \( ^{13}\)N,\(^{13}\)C-labelled \( fH_{12-13} \) illustrated in the \( ^{13}\)N,\(^{1}H \) heteronuclear single quantum coherence (HSQC) spectrum in Fig. 2] allowed near-complete assignment of \( ^{1}H,^{15}\)N, and \( ^{13}\)C nuclei. The oxidised state of the eight Cys residues was confirmed by mass spectrometry, whilst the disulphide pattern (Cys\(^{9}\)-Cys\(^{11}\) and Cys\(^{12}\)-Cys\(^{14}\)) was inferred from the proximities of cysteine pairs in initial structures calculated without disulphides; these were then included in subsequent calculations. The final ensemble of the 20 lowest-energy structures, determined on the basis of 2911 nuclear Overhauser effect (NOE)-derived restraints, converges well (Table 1). Predictably, overlays of individual modules (Fig. 3a and b) yield better local convergence than overlays of both CCPs at once (Fig. 3c). Nonetheless, the 1.2-Å backbone root-mean-square deviation (RMSD) for the ensemble of bimodules is not consistent with a large variation in intermodular orientations (Fig. 3d); this tallies with the relatively plentiful NOEs detected within the linker, between the linker and modules, and between modules (Table 1). Strands and connecting loops in each of modules 12 and 13 are labelled (Fig. 3e), in accordance with a convention based on the occurrence of a maximum of eight \( \beta \)-strands (\( A-H \)) in some individual CCPs.

Relaxation measurements (Fig. 4) reveal mobility in the \( A-B \) loop of CCP 12 and its short “hypervariable region” beyond strand B (e.g., Leu704 and Ser705 have small \( 15\)N \( T_1/T_2 \) ratios and low \( 1H,^{15}N \) NOEs). The very short \( A-B \) loop of CCP 13 also appears mobile (Ser756 has a small \( 1H \)–\( ^{15}N \) NOE, backbone NH signals of Asn757 could not be found, and only a weak HSQC peak was detected for Ile759 NH). The unusual helix-like insertion in the hypervariable region of CCP 13 (discussed below) is not particularly mobile on timescales probed by measurements of \( 1H,^{15}N \) NOEs and \( T_1/T_2 \) but a stretch of residues just beyond the insertion and looping around to CCP 13 strand C exhibits below-average \( 1H,^{15}N \) NOEs (Asn767, Lys768, and Glu770) and a very weak HSQC peak (Lys769). The unusual helix-like insertion in the hypervariable region of CCP 13 (discussed below) is not particularly mobile on timescales probed by measurements of \( 1H,^{15}N \) NOEs and \( T_1/T_2 \) but a stretch of residues just beyond the insertion and looping around to CCP 13 strand C exhibits below-average \( 1H,^{15}N \) NOEs (Asn767, Lys768, and Glu770) and a very weak HSQC peak (Lys769). The D–E loop of CCP 13 appears flexible; Gly783 has a weak HSQC peak, Glu785 has exceptionally low \( 1H,^{15}N \) NOE, and Gly786 NH has the lowest \( 1H,^{15}N \) NOE of any residue apart from unstructured termini.
ular interest is that residues of the linking sequence between modules appear immobile on the timescales probed; six out of eight linker residues have average or higher-than-average $^1$H,$^{15}$N NOEs, whilst $T_2$ values are just below average, and $T_1$ values are typically just above average. This is consistent with the aforementioned numerous $^1$H–$^1$H NOEs involving linker residues and the relatively low RMSD obtained upon overlaying of the bimodular ensemble (Fig. 3c); taken together, these data imply that CCPs 12 and 13 are quite rigidly associated, and that the ranges of tilt, twist, and skew intermodular angles in Fig. 3d (and Table 1) are not a consequence of overrefinement in the ensemble of final structures.

The structures of modules 12 and 13

Structure determination of fH12–13 reveals an archetypal CCP in module 12 that overlays very well ($C^\alpha$RMSD<2 Å) with 14 out of 42 experimentally determined CCP structures (Fig. 5); it superimposes particularly well ($C^\alpha$ RMSD=1.4 Å) with CCP 19 of fH (both CCPs are in sequence cluster B in accordance with Soares et al.42). Notable features include a cis-Pro at 708, the unusual occurrence of a Trp in strand B, and three successive Tyr residues prior to strand D. Its surface is predominantly electronegative (Fig. 6), with dispersed patches of lipophilicity.

Module 13 is structurally divergent (Fig. 5) and overlays poorly with CCP 12 ($C^\alpha$ RMSD over a span of 46 aligned residues, 3.5 Å); at 51 residues, CCP 13 is the smallest CCP known, but it does not overlay well with another small (53-residue) CCP (CCP 4 of fH; $C^\alpha$ RMSD=2.8 Å) or 36 other solved CCP structures (RMSD<3.0 Å). The long axis of CCP 13 is about 4 Å shorter than that of CCP 12, as is the distance between disulphide bridges, whereas its diameter is about 2.5 Å greater; hence, it is more spherical than CCP 12 (Fig. 5). Surprisingly, an insertion of three residues (766KNK768) occurs within the hypervariable loop of CCP 13 that projects in a helical-like structure laterally from the module. Together with lysines in and just beyond the linker (and basic side chains from elsewhere), this bulge on the side of module 13 contributes to a strikingly electropositive patch extending over almost an entire face of the module. A three-residue deletion occurs within the A–B loop of CCP 13, and a three-residue deletion occurs in its flexible D–E loop. Strand D is interrupted, whilst hydrophobic-residue-rich strand E of CCP 12 is replaced in CCP 13 by an exposed loop ($^{783}$GKE$^{785}$), and there is no $\beta$-strand H in CCP 13 (refer to Fig. 5).
Architecture of the Central Portion of Factor H

Table 1. Experimental input and statistics for the lowest-energy fH12–13 structures

<table>
<thead>
<tr>
<th>Number of structures in ensemble</th>
<th>20</th>
</tr>
</thead>
<tbody>
<tr>
<td>Numbers of upper-limit distance constraints</td>
<td></td>
</tr>
<tr>
<td>Intraresidue, 1–i–j ≤ 1</td>
<td>1371</td>
</tr>
<tr>
<td>Medium range, 1&lt;i–j&lt;5</td>
<td>359</td>
</tr>
<tr>
<td>Long range, 1&lt;i–j&lt;5</td>
<td>1181</td>
</tr>
<tr>
<td>Total</td>
<td>2911</td>
</tr>
<tr>
<td>Intermodular</td>
<td>10</td>
</tr>
<tr>
<td>Module 12 to linker</td>
<td>75</td>
</tr>
<tr>
<td>Module 13 to linker</td>
<td>75</td>
</tr>
<tr>
<td>Number of hydrogen bonds</td>
<td>22</td>
</tr>
<tr>
<td>RMSD</td>
<td></td>
</tr>
<tr>
<td>Bimodule (12+13) (Å)</td>
<td>All heavy atoms</td>
</tr>
<tr>
<td>Backbone heavy atoms</td>
<td>1.23</td>
</tr>
<tr>
<td>Module 12 (Å)</td>
<td>All heavy atoms</td>
</tr>
<tr>
<td>Backbone heavy atoms</td>
<td>0.33</td>
</tr>
<tr>
<td>Module 13 (Å)</td>
<td>All heavy atoms</td>
</tr>
<tr>
<td>Backbone heavy atoms</td>
<td>0.39</td>
</tr>
<tr>
<td>Intermodular angles [minimum–maximum [mean (S.D.)]]</td>
<td></td>
</tr>
<tr>
<td>Skew (°)</td>
<td>78–91 [85 (4)]</td>
</tr>
<tr>
<td>Twist (°)</td>
<td>64–122 [91 (18)]</td>
</tr>
<tr>
<td>Tilt (°)</td>
<td>61–91 [78 (9)]</td>
</tr>
<tr>
<td>Ramachandran assessment</td>
<td></td>
</tr>
<tr>
<td>Most favoured (%)</td>
<td>72.4</td>
</tr>
<tr>
<td>Additionally allowed (%)</td>
<td>22.2</td>
</tr>
<tr>
<td>Generously allowed (%)</td>
<td>2.9</td>
</tr>
<tr>
<td>Disallowed (%)</td>
<td>2.4</td>
</tr>
<tr>
<td>Coarse area buried WHAT IF score</td>
<td>−1.671</td>
</tr>
<tr>
<td>Surface area buried between modules (Å²)</td>
<td>561*</td>
</tr>
</tbody>
</table>

* PDB ID 2KMS and BioMagResBank entry number 16439.
* Not used as input for structure calculation.
* Residues from Cys¹ to Cys¹⁰ of each module.
* Closest-to-mean structure.
* Calculated by breaking fH12–13 at the midpoint of the linker.

Together with the C-terminal (strand H) region of CCPs, is normally essential for forming and stabilising the interface with the following module. It is therefore clear that either the 13–14 interface must be flexible or modules 13 and 14 adopt an arrangement that is distinct from those of other module pairs. In summary, CCP 13 appears highly specialised with regard to its overall shape, electrostatic properties, and interface with the following module.

Module 13 is tilted by ~80° with respect to module 12

Four (Val745, Ala746, Ile747, and Leu750) of the eight linker residues bury all or a substantial portion of their hydrophobic side chains in the interface between modules along with two residues (Thr724 and Ile726) from CCP 12 strand E and residues from the C–D loop (His773) and the F–G loop (Asn794 and, to a lesser extent, Ile793) of CCP 13 (Figs. 1 and 6). The β-methylene of the other linker residues (and in particular Lys752) also contribute to this miniature core, which forms a wedge between the modules. Altogether, ~560 Å² of surface area may be considered buried between modules. The hydroxyl group of Thr724 and the HN of His773 are within H-bonding distance, as are (in the lowest-energy structure) the side-chain amide of Asn794 and the backbone nitrogen of the linker residue Lys751.

The software CRYSOI was employed to fit each conformer from the NMR-derived ensemble of fH12–13 to SAXS data collected on fH12–13 (in 50 mM potassium phosphate, pH 7.4). Good fits (see Fig. 7a) of the scattering data to each ensemble member were observed (1.24<χ<1.43), supporting the structure determined by NMR (in 20 mM potassium phosphate, pH 6.6).

Size-exclusion chromatography and AUC consistent with tilted modules in fH10–15

When passed down a calibrated size-exclusion chromatography column, fH12–13 eluted a little later than globular proteins of a comparable molecular mass, consistent with the slightly elongated protein molecule observed by NMR. On the other hand, the six-module construct fH10–15 also eluted at a position similar to that of a comparable globular protein; thus, the overall shape of fH10–15 is not highly elongated (data not shown).

In a more rigorous follow-up study, fH12–13, fH11–14, and fH10–15 were submitted to analysis via sedimentation–velocity AUC, and the results are summarised in Table 2. Analysis of AUC profiles demonstrated that all three of the samples are predominantly monomeric at 0.2–0.4 mg ml⁻¹ in phosphate-buffered saline (PBS). As may be seen from Table 2, the fitted frictional ratios indicate that fH11–14 is considerably more elongated in shape than fH12–13. On the other hand, a very interesting feature of the AUC results is that fH10–15, whilst more extended than fH12–13, is less elongated in overall shape than fH11–14. A speculative suggestion is that whilst all the modules are tilted with respect to their neighbours, they are not all tilted in the same direction. This suggestion could have been interrogated further by utilising the AUC data to construct bead-based models, but this would have entailed a more extensive study, requiring chemical constraints as opposed to purely geometrical constraints. It was therefore decided to subject the samples to SAXS on the basis that the latter technique provides a larger number of geometrical parameters upon which to base molecular modelling.

SAXS confirms the compact structure of fH10–15

The experimental scattering patterns of fH12–13, fH11–14, and fH10–15 are presented in Fig. 7a, and parameters determined from these data sets are listed in Table 2. The values of molecular mass and hydrated particle volume estimated from the scattering data agree with those predicted from the sequences for all constructs. Thus, in agreement with the AUC data (collected on samples in PBS, pH 7.3), SAXS (collected on samples in 50 mM potassium phosphate, pH 7.4) showed that all constructs appear to be monomeric in solution. To further cross-validate the consistency of the structural and hydrodynamic data, we computed the
sedimentation coefficients of the bead models generated by SAXS for the three fH constructs using the program HYDROPRO. These computed values (Table 2) are in very good agreement with the experimental values, indicating that the different experimental conditions and ionic strengths employed in the structural and hydrodynamic experiments do not lead to significant changes in the overall structure.

Analysis of the radii of gyration \( (R_g) \) and maximum dimensions \( D_{\text{max}} \) of the fH constructs reveals that fH11–14 is significantly more extended than fH12–13, but that both fH11–14 and fH10–15 share a \( D_{\text{max}} \) of \( \sim 10.5 \) nm. The data thus indicate that (i) modules 11–14 do not lie in a smooth curve reminiscent of a horseshoe (more likely, they adopt a zigzag arrangement); and (ii) the terminal CCP modules of the fH10–15 construct (i.e., CCPs 10 and 15) are tilted towards the more central modules, forming an overall compact structure.

The distance distributions \( p(r) \) for all fH constructs are positively skewed, with tails at large distances (Fig. 7b), indicating that all the particles have nonspherical elongated shapes. The fH12–13 data show a bimodal distribution characteristic of dumbbell-shaped particles, with the first maximum at \( \sim 1.6 \) nm and a second maximum at \( \sim 3.4 \) nm. Along with a maximum size \( D_{\text{max}} \) of 7.1 nm, these distances are in good agreement with the ensemble of NMR-derived structures; 1.6 nm corresponds approximately to the radii of the individual CCP 12 and CCP 13 modules, and 3.4 nm corresponds very approximately to the separation of the centres of mass of the CCPs.

Both the two larger constructs fH11–14 and fH10–15 are more extended in shape than fH12–13. The position of the second maximum in the distance distributions of all three constructs is conserved, indicating that the average intersubunit distance is maintained at \( \sim 3.4 \) nm. However, the position of the first maximum is shifted to larger distances in fH11–14 and fH10–15, corresponding to an increase in the effective cross section of these constructs and the formation of compact structures.

SAXS suggests that fH12–13, fH11–14, and fH10–15 lack intermodular flexibility

An analysis of intermodular dynamics in fH12–13, fH11–14, and fH10–15, based on the SAXS data, was...
conducted using the recently developed ensemble optimisation method. For each construct, the ensemble optimisation method analysis produced a skewed narrow $R_g$ distribution for the selected ensemble of structures that best fit the SAXS data (Fig. 7c). These distributions may be compared with the broad range of $R_g$ exhibited by the initial pool of fH12–13, fH11–14, and fH10–15 conformers (Fig. 7c) with randomly generated intermodular angles; such a comparison suggests that in none of these cases are modules free to articulate at random relative to their neighbours. The selected ensemble for fH12–13 is skewed toward higher values of $R_g [R_g(\text{av})=2.1\pm 0.1 \text{ nm}]$, confirming that this structure is significantly more extended than an average random configuration. Conversely, the selected ensembles for both fH11–14 and fH10–15 are skewed towards lower values of $R_g [R_g(\text{av})=2.8\pm 0.3 \text{ nm} \text{ and } R_g(\text{av})=3.5\pm 0.8 \text{ nm}, \text{ respectively}]$, indicating that these structures are much more compact than those in the random pool.

**Modelling structures based on SAXS and NMR-derived structures**

Both ab initio shape reconstructions and SAXS-based rigid-body modelling were employed to independently determine the overall low-resolution structures of fH12–13, fH11–14, and fH10–15. In excellent agreement with NMR, the structures of fH12–13 determined by two ab initio modelling programs employing beads and dummy residues are essentially vee-shaped (Fig. 8a).

A more detailed modelling of the larger fH constructs was conducted using BUNCH to optimize a spatial distribution of structured domains and linkers (the latter represented as dummy residues) to best fit the SAXS data. The BUNCH-derived models generated for fH11–14 show a reproducible spatial distribution of the CCPs, with minor variation in their relative orientations (Fig. 8b). The long axes of the modules trace a zigzag path within a plane, compatible with the ab initio shape envelope.
Modules 12 and 13 form closer contacts with flanking modules (CCPs 11 and 14) than with each other. The most typical model provides a very good fit to the experimental data with $\chi = 1.18$ and, importantly, the fit does not improve when the CCP 12 and CCP 13 domains are treated as two separate rigid bodies. This observation further validates the use of the NMR-derived model of fH12–13 as a single rigid body in the analysis and suggests that the $\sim 80^\circ$ tilt angle between CCPs 12 and 13 is conserved in the context of the larger construct.

The most typical fH11–14 BUNCH model was initially used as a single rigid body in the BUNCH analysis of fH10–15. Poor fits ($2.13 < \chi < 3.16$) to the SAXS data suggest that the conformation of fH11–14 in solution differs from that of its constituent CCPs in the context of the fH10–15 construct. Thus, further rigid-body modelling was conducted using the NMR-derived structure of fH12–13. Whilst this strategy leads to generation of a large number of possible arrangements of CCPs, the resulting models of fH10–15 fall into two distinct clusters (Fig. 7c). These clusters share overall similarities, including the planar zigzag nature of the path described by the long axes of the modules, but they differ in the relative orientation of the central CCP 12–13 bimodule with respect to the flanking domains. In the first case ($\chi_{RB} = 1.35$), CCPs 12 and 13 are oriented approximately perpendicular to the long axis of the fH10–15 molecule, producing a kink in the structure (Fig. 8d). In the second case ($\chi_{RB} = 1.30$), the CCP 12–13 bimodule is oriented along the long axis, whilst a nearly $-180^\circ$ tilt occurs between modules 13 and 14 (Fig. 8d). Note that BUNCH was also run by simultaneously fitting both fH10–15 and fH12–13 SAXS data without constraining CCP 12 and CCP 13 subunits according to the NMR-derived structure. A larger degree of variability between models was observed, however, and the fits were not improved ($\chi_{RB} = 1.29$ and 1.99 for the fit to the most typical model using the fH10–15 and fH12–13 data, respectively).

Discussion

The structural biology of the N-terminal and C-terminal regions of fH that encompasses the known binding sites for its principal protein and carbohydrate ligands, has been studied extensively; the medium-resolution or high-resolution structures of all eight N-terminal CCPs have been experimentally determined [in fragments: fH1–2 (Protein Data Bank (PDB) ID 2RLP), fH2–3 (PDB ID 2RLQ), fH1–4 (PDB ID 2WII), and fH5 and fH6–8 (PDB ID 2UWN)] by NMR or crystallography, as have structures of four of the C-terminal six modules [fH15–16 (PDB ID 1HFH) and fH19–20 (PDB IDs 2BZM and 2G7I)]. The central region of fH, on the other hand, represents an unexplored structural territory.
We have now provided initial structural insights into the central CCPs of fH. The structure of module 12 is similar to those of other CCP structures. Indeed, experimentally derived and modelled (by homology)42 structures of fH CCP 12 are very similar overall (Cα RMSD=1.4 Å). Module 13, the smallest CCP within the RCA family, has a highly divergent structure (Fig. 5); it thus joins a small group of structurally divergent modules (all of which diverge in different ways) within fH that includes CCPs 7 and 20.42 However, unlike these modules, CCP 13 (in the context of fH12–13, fH13–15, fH8–15, and fH10–15) has been shown not to bind to the model glycosaminoglycan heparin or to enzymatically generated fragments of heparan sulphate under physiological conditions.27 This is despite CCP 13 having one face that carries an extensive positive charge (whilst the other is predominantly neutral). It seems likely that such a divergent structure and striking electrostatic properties have evolved to serve a specialised but as yet unknown purpose. In total, we now know the structures of 14 CCPs in fH; however, our knowledge of how CCPs are spatially organised to form a functional complement regulator48,49 lags behind.

Before the current work, detailed structural knowledge was available for seven intermodular interfaces (fH1–2, fH2–3, fH3–4, fH6–7, fH7–8, fH15–16, and fH19–20) that feature intermodular linkers of between three and five residues in length. Leaving aside the possibility of overestimation of flexibility (e.g., in the structure of fH15–16 based on earlier NMR studies with no isotopic enrichment of the sample50) or overestimation of rigidity (e.g., in fH3–4, fH6–7, and fH7–8 derived from crystal structures31,51 without corroboration by NMR), the arrangement of neighbouring modules in all of these fragments may be classified as “end-to-end.”

Thus, none of the structures of fH fragments examined previously suggests large tilt angles (on average) between neighbouring CCPs; moreover, very similar arrangements of the three N-terminal modules of fH were observed both in solution28 and in complex with C3b,51 implying that modular organisation within the non-complexed fragments likely reflects the architecture of the functioning molecule, at least in some cases. Importantly, the intermodular linkers amongst this set of known CCP structures are between three and five residues in length; this may be compared to linker lengths of six, six, eight, seven, and five for the five linkers (in order) in human fH10–15 (i.e., of the 358 residues within fH10–15, nearly one-tenth lie in linkers). In this respect, it is noteworthy that the eight residues between the two CCPs at the N-terminus of complement receptor type 2 promoted a side-by-
side arrangement of these modules (tilt angle = 142°; distance between the centres of mass of the two modules = 2.2 nm), as do the six linking residues between modules 2 and 3 of fB53 and complement C2. These observations had suggested the hypothesis that a similar side-by-side arrangement might be found between fH CCPs 12 and 13, which are joined by an eight-residue linker; however, long linkers also provide the theoretical potential for greater flexibility between modules, leading to an opposing hypothesis.

The current study falsifies both hypotheses, since fH modules 12 and 13 are neither side-by-side nor flexibly associated. Numerous intermodular and linker-to-module NOEs and the good convergence of the ensemble of solved structures, backbone relaxation data, and low-resolution data from size-exclusion chromatography, AUC, and SAXS all support the contention that the two modules exhibit a defined mutual tilt of ∼80°, burying a surface area of ∼560 Å². Key interface residues (T724, I726, Val745, Ala746, Ile747, Leu750, His773, and Asn794) are conserved (or conservatively replaced) amongst orthologues (Fig. 1), so this arrangement of CCPs 12 and 13 of fH is likely conserved across species; on this basis, it is also highly likely to occur between CCPs 5 and 6 of fH-related protein 5 (Fig. 1). The even-and-odd mode (EOM) analysis of the SAXS data independently verifies the rigid nature of the long linker between CCP modules 12 and 13. It also indicates that the fH11–14 and fH10–15 substructures of fH are compact and rigid. Thus, the central CCPs of fH do not provide a flexible tether between the N-terminal region and the C-terminal region.

It was initially speculated that the CCP 12–CCP 13 tilt angle might be one of a succession of same-direction tilts between adjacent modules in the middle of fH. If this were the case, addition to fH12–13 of CCPs 11 and 14 (creating fH11–14) would produce a horseshoe-shaped entity or a spiral and would not result in a more extended structure. We put these predictions to test by comparing fH12–13 with fH11–14 using AUC and SAXS. Both techniques show that fH11–14 is more extended ($D_{max} = 10.5 ± 0.5$ nm) than fH12–13 ($D_{max} = 7.0 ± 0.5$ nm). Furthermore, the use of the NMR-derived fH12–13 structure in the rigid-body modelling of the SAXS data for fH11–14 produced a relatively well-converged set of conformers, allowing an approximation of intermodular tilts to be made. These are consistent with a compact conformation in which tilt angles between modules 11 and 12, and between modules 13 and 14, are larger than the 80° tilt between CCPs 12 and 13. On the other hand, the tilt between modules 11 and 13 is small, consistent with a zigzag arrangement rather than a horseshoe shape for fH11–14. A similar comparison was made between fH11–14 and fH10–15. In this case, the additional modules did not result in a more extended structure (compared to fH11–14), implying that modules 10 and 15 are folded in towards modules 11–14.

Previous SAXS studies applied to full-length fH23,39,55 implied a $D_{max}$ of ∼34–35 nm; fH consists of 20 CCPs, each of a typical length (3.5–3.8 nm). Whilst these results clearly rule out a highly extended end-to-end arrangement of CCPs in fH, they provide no information on the tilt angles between specific modules. Low-resolution data for fragments fH1–5, fH6–8, and fH16–2056,57 were not
Table 2. Summary of SAXS and AUC data

<table>
<thead>
<tr>
<th>Sample (molecular weight kDa)</th>
<th>SAXS</th>
<th>AUC</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>$R_g$ (nm)</td>
<td>$D_{max}$ (nm)</td>
</tr>
<tr>
<td>-----------------------------</td>
<td>----------</td>
<td>--------------</td>
</tr>
<tr>
<td>fH12–13 (13.6)</td>
<td>2.2±0.1</td>
<td>7.1±0.5</td>
</tr>
<tr>
<td>fH11–14 (27.3)</td>
<td>3.1±0.1</td>
<td>10.5±0.5</td>
</tr>
<tr>
<td>fH10–15 (41.1)</td>
<td>3.1±0.1</td>
<td>10.4±0.5</td>
</tr>
</tbody>
</table>

Molecular weights calculated from sequence.

*R* in *D* corresponds to the experimentally (SAXS) determined radius of gyration, maximum particle dimension, and hydrated particle volume, respectively. $s_{20w}$ is the sedimentation coefficient computed from the SAXS bead models. $X_s$ and $X_{RB}$ are the discrepancies between the experimental data and the computed scattering curves from the most typical ab initio and rigid-body models, respectively. The $X_{RB}$ shown for fH12–13 is the value of the best-fit NMR model to the SAXS data calculated using CRYSOL.

$F_{C,M}$ is the frictional ratio calculated from the sedimentation coefficient and the known (from sequence) molecular mass. $f/f_{min}$ is the component of the frictional ratio that can be attributed to shape, on the basis that for a non-extended 'globular' protein, a value $F=1.20$ (default value in SEDFIT) is generally applicable (for the computation of $f/f_{min}$ the second and more reliable of the two estimates for $F/s$ has been divided by 1.20).
recombinant strains of *Pichia pastoris* was described previously. For the current work, isotopically enriched samples of fH12–13 were generated as follows. The fH12–13-overexpressing strain of *P. pastoris* was cultured in a 1.25-L vessel on a Bioflo 3000 fermentor (New Brunswick Scientific, Edison, NJ). For $^{15}$N labelling, cell mass was

Fig. 8 (legend on next page)
grown in 0.6 L of minimal medium containing 7.0 g of [15N] ammonium sulphate, 20 g of glucose, basal salts, trace elements, and vitamins. Prior to induction with methanol, 1.0 g of glycerol was added to facilitate derepression of the alcohol oxidase promoter. A [13C,15N]-labelled sample was prepared following the same protocol as for [15N] labelling, but [13C]glucose, glycerol, and methanol (Sigma-Aldrich, St. Louis, MO) either before purification or between the first purification step and the second purification step. Proteins were validated before purification or between the first purification step and the second purification step.

Proteins were purified as described previously. In brief, cation-exchange or anion-exchange chromatography was followed by size-exclusion or heparin-affinity chromatography. N-linked glycans were removed by incubation with endoglycosidase H maltose-binding protein fusion protein (New England Biolabs, Ipswich, MA) either before purification or between the first purification step and the second purification step. Proteins were validated by peptide mass fingerprinting using matrix-assisted laser desorption ionisation time-of-flight mass spectrometry and (in the case of fH12–13), additionally, by mass determination on a Fourier transform ion cyclotron resonance mass spectrometer.

NMR spectroscopy

A 620 μM sample of [15N,13C]fH12–13 in 20 mM potassium phosphate buffer (pH 6.6) plus 7% (vol/vol) D2O at 298 K was used for structure determination. A standard suite of NMR experiments enabled near-complete resonance assignments. Spectra were acquired on Bruker AVANCE 600-MHz and 800-MHz spectrometers equipped with 5-mm triple-resonance cryoprobes, using pulsed field gradients. Spectra were processed using TopSpin (Bruker version 1.3) or the Azara processing package [provided as part of the common computing protocol for NMR (CCPN) data model] and analysed using CCPN Analysis. Assignment of backbone [HN(CACB), CBCA(CO) NH, HNCO, HN(CA)CO, HHBHA(CBCACO)NH, and HBHANH] and side-chain [H(CC)(CO)NH total correlated spectroscopy (TOCSY), H(CC)(CO)NH TOCSY, H(CC)(CO)NH correlated spectroscopy (COSY), and H(CC)(CO)NH triple-resonance experiments, along with two-dimensional (H)CB (CGCD)HD and (H)CB(CGCCDCE)HE aromatic side-chain experiments.

Assignment, structure calculations, and analysis

The excellent quality of fH12–13 NMR spectra is illustrated in Fig. 2. Spectral folding in the 13N dimension was employed to increase resolution. With the N-terminal cloning artefact (Ala-Gly) and one expression artefact unit (Glu-Ala), 89% of triple-resonance assignments, including 98% of backbone atoms, were completed. Of the theoretically assignable backbone atoms, the following are missing: Pro707 (CO), Val715 (CO), and Asn275 (CO, N, and NH).

Two out of five X-Pro linkage residues (Pro708 and Pro799) were defined as cis, whilst the remainder were defined as trans, as judged by chemical shift differences between Pro 15N and 13C atoms; strong NOEs were detected between HN (Xaa−1 −1) and H (Pro′), and between H (Xaa−1 −1) and HN (Xaa−1 −1), and no NOE was detected between H (Xaa−1 −1) and H (Pro′). Based on accurate mass spectrometry, all eight cysteine residues were inferred to be in the oxidised state. Specific disulphide bridges were incorporated into the final rounds of structure calculations on the basis of NOEs (reinforced by a precedent established by previous studies of CCPs). The 15N-edited NOE spectroscopy spectrum was assigned manually to near-completion, and ~10% of peaks in a 13C-edited NOE spectroscopy spectrum were also assigned manually. The remaining NOEs were assigned within CYANA 2.1, which combines automated assignment and structure calculation. Upper limits of distance constraints generated from a seven-cycle routine within CYANA were transferred into the program Crystallography and NMR System70 using the program Format Converter within the CCPN suite, in order to ultimately perform structure refinement using explicit water. Whilst no hydrogen-bond restraints were used in the structure calculation, a hydrogen/deuterium exchange experiment was performed to cross-validate 22 hydrogen bonds inferred from the final structure. Lyophilised protein was transferred to 99.9% (vol/vol) deuterated buffer, and slowly exchanging amide protons were identified in a 15N/H HSQC spectrum collected 30 min after exposure to D2O. All of the slowly exchanging NH signals were engaged in hydrogen bonds, except for the Trp738 side-chain NH, which is buried in CCP 12. The quality of the data and structures was further analysed using WHAT IF71 and PROCHECK72 (see Table 1).

Secondary structure elements were identified using STRIDE. Surface potentials were determined using GRASP73 and the MOLCAD module7 of SYBYL v6.9 (Tripos Associates, St. Louis, MO). The solvent-accessible surface area (SA) and the buried surface area at the intermodular junctions were calculated using GETAREA. The buried surface area was computed as (SA module12 + SA module13) – SA bimodule12–13, where CCP 12 was considered to encompass one residue before its Cys1 and four residues after its Cys1 (i.e., Thr690-Asp748); CCP 13 boundaries were considered the fourth residue before its Cys2 and one residue after its Cys2 (i.e., Lys749-Ser804), and the bimodule consists of Thr690-Ser804. Intermodular angles for the ensemble of NMR-derived structures were determined as described previously, using (for a reference z-axis) a vector between the principal axis of the inertia tensor (the z-axis) and the Cα of the conserved Trp738 (CCP 12) or Trp797 (CCP 13), with module boundaries defined as Cys1 and Cys2. Combinatorial extension72 was employed to compare each experimentally determined CCP structure within the complement system against the closest-to-mean individual structures of CCPs 12 and 13. Structural superimpositions were performed using MAMMOTH-mult19 then depicted using RasMol v2.7.3.75

Fig. 8. Modelling from SAXS data. (a) Overlay of DAMMIF-derived ab initio shape envelope (left) and GASBOR-derived dummy residue model (right) with the NMR-derived ensemble of fH12–13. (b) Overlay of the DAMMIF-derived ab initio shape envelope with the most typical BUNCH-derived model of fH11–14 (left); ensemble of 10 BUNCH models for fH11–14 (right). (c) Overlay of the DAMMIF-derived ab initio shape envelopes with the two most typical BUNCH models of fH10–15 (I and II). (d) The two most typical BUNCH models of fH10–15 (I and II). In (a–c), lower views are rotated 90° clockwise about the horizontal axis. The CCPs are shown in green (CCP 10), cyan (CCP 11), red (CCP 12), blue (CCP 13), yellow (CCP 14), and pink (CCP 15), with orange spheres representing linker residues modelled by BUNCH as a chain of dummy atoms.
Relaxation measurements

Measurements of backbone $^{15}$N $T_1$ and $T_2$ and $^1$H,$^{15}$N NOEs were conducted at 600 MHz using 0.6 mM samples in 20 mM potassium phosphate buffer (pH 6.6). The delays used for measurements of $T_1$ and $T_2$ were 51.1, 51.1, 301.1, 501.1, 701.1, 801.1, 901.1, and 1001.1 ms, and 16.0, 16.0, 48.0, 80.1, 112.1, 128.1, 144.1, and 160.1 ms, respectively. For heteronuclear NOE measurement, a reference experiment was recorded with a 5-s relaxation delay, whilst a second spectrum was recorded with 1H 120° pulses applied for the last 3 s of the 5-s delay. NMR data were processed using the AZARA suite of programs and assigned using Analysis. Where possible, a single-exponential decay was fitted to the extracted peak heights for each residue to obtain the relaxation rates using non-linear fitting. Peaks from the following residues were excluded due to overlap: Ser706, Tyr710, Tyr711, Cys744, Leu761, His773, Asn802, and Cys803; signals from the backbone amides of the following residues were too weak to allow measurement of relaxation parameters: His735, Gly736, Ile759, His764, Lys769, and Gly783.

Size-exclusion chromatography

Samples of fH10–15 (82 μM), fH11–14 (118 μM), and fH12–13 (15 μM) [in 0.5 ml of potassium phosphate buffer (pH 6.6) supplemented with 500 mM NaCl] were loaded separately onto a Superdex 75 preparative-grade size-exclusion column (GE Healthcare, Piscataway, NJ), and elution profiles (monitored at 280 nm; theoretical extinction coefficients of 60,900, 47,400, and 28,500, respectively) were compared to the elution profiles of standards (bovine serum albumin, ovalbumin, chymotrypsinogen, ribonuclease A, and cyanocobalamin) applied under the same conditions.

Analytical ultracentrifugation

Sedimentation velocity analyses were performed in a Beckman XL-A analytical ultracentrifuge at a rotor speed of 45,000 rpm, with a sample volume of 0.4 ml and a protein concentration of 0.2–0.4 mg ml$^{-1}$ in PBS (pH 7.5) at 293±0.5 K. A series of radial scans with absorbance at 279 nm across the centrifuge cell was recorded and saved to disk via the Beckman-Coulter Proteome Lab software. The first scan was performed immediately upon attainment of the set rotor speed, and 80 subsequent scans were recorded at 2-min intervals thereafter. The total data set thus generated was analysed via the package SEDFIT to yield values of the $s$-distribution parameter $c(s)$ as a function of sedimentation coefficient ($s$). For non-linear fitting of the optical density values at 273 nm as a function of radial distance in SEDFIT, the resolution was set to 150, with a confidence factor set to 0.68. Over a series of fits, an average value for the frictional ratio ($F$, which relates to the “axial ratio” and pertains to the shape of the molecule) was determined; this was used as a default for all fittings to minimise artefactual variations.

Small-angle X-ray scattering

Synchrotron radiation X-ray scattering data were collected at the X33 beamline of the European Molecular Biology Laboratory (DESY; Hamburg), using a MAR345 image plate detector (MarResearch, Norderstedt, Germany) and 120 s of exposure time. Solutions of fH constructs were measured at 10 °C in 50 mM potassium phosphate buffer (pH 7.4) at protein concentrations of 0.7, 1.8, and 3.5 mg ml$^{-1}$ (fH12–13), 0.7 and 1.6 mg ml$^{-1}$ (fH11–14), and 2.4, 5.1, and 10.5 mg ml$^{-1}$ (fH10–15). The sample-to-detector distance was 2.7 m, covering a range of momentum transfer 0.08 $< s < 5.0$ nm$^{-1}$ (s = 4π sinθ/λ, where θ is the scattering angle, and λ = 0.15 nm is the X-ray wavelength). Based on a repeat 120-s exposure, no detectable radiation damage occurred. Data from the image plate were normalised to the incident beam intensity and averaged, and the scattering of buffer solutions was subtracted. Difference curves were scaled for solute concentration. All data manipulations were performed using the PRIMUS software package.

The forward scattering $I(0)$ and the radius of gyration $R_g$ were determined from Guinier analysis, assuming that, at very small angles ($s < 1.3/R_g$), the intensity is represented as $I(0) = I(0)exp(-sR_g^2/3)$. These parameters were also estimated from the full scattering curves using the indirect Fourier transform method implemented in the program GNOM, along with the distance–distribution function $p(r)$ and the maximum particle dimensions $D_{max}$. Molecular masses of solutes were estimated from SAXS data by comparing extrapolated forward scattering with that of a reference solution of bovine serum albumin.

Due to the uncertainty in molecular mass estimation from SAXS data that results from uncertainty in the measured protein concentrations, an excluded volume of the solutes was determined from the ab initio modelling program DAMMIF. This estimation is independent of protein concentration and can be obtained in an automated fashion with minimal user bias. For globular proteins, this hydrated particle volume (in nm$^3$) is approximately 1.5–2 times the molecular mass (in kDa).

Analysis of the NMR ensemble of fH12–13

The fit of the NMR ensemble of fH12–13 to the SAXS data was conducted using the program CRYSOL. CRYSOL calculates the partial scattering amplitudes of

---

‡ [http://www.rasmb.bbrl.org](http://www.rasmb.bbrl.org)
proteins from their atomic coordinates, taking into account the hydration layer and the excluded solvent volume.

### Analysis of interdomain flexibility by the ensemble optimisation method

Analysis of interdomain flexibility and the size distribution of possible conformers, consistent with the measured scattering data for fH12–13, fH11–14, and fH10–15, was conducted using the ensemble optimisation method. This method selects an ensemble of possible conformations from a pool of randomly generated models, using CRYSTOL to calculate the theoretical scattering profiles and a genetic algorithm GAJOE (to select the representative set). The input structures for an analysis of the scattering data employing ensemble optimisation used the following as rigid bodies: the NMR-derived structure of CCP 15 (PDB ID 1HFI) and CCP homology models of CCPs 10, 11, and 14 using the program Modeler 9v1. Linkers between the modules were represented as a flexible chain of dummy residues.

### Ab initio shape determination and molecular modelling

Low-resolution shape envelopes for the fH constructs were determined using the ab initio bead modelling program DAMMIF. DAMMIF represents the particle as a collection of M (≧1) densely packed beads inside an adaptable and loosely constrained search volume compatible with the experimentally determined Ke. In contrast to the bounded search volume used in the bead modelling program DAMMIN, an unrestricted search volume that can grow or reduce in size during the simulated annealing procedure avoids artefactual boundary effects that may occur in the case of anisotropic particles. Each bead is randomly assigned to solvent (index=0) or solute (index=1), and the particle structure in solution is described by a binary string of length M. Disconnected strings of beads are rejected, and scattering amplitudes are calculated. Simulated annealing is then used to search for a compact model that minimises the discrepancy:

\[
x^2 = \sum \frac{1}{N-1} \sum \left( \frac{I_{exp}(S_i) - cI_{calc}(S_i)}{\sigma(S_i)} \right)^2
\]

where N is the number of experimental points; I_{exp}(S_i) and I_{calc}(S_i) are the experimental and calculated intensities, respectively; c is a scaling factor; and \(\sigma(S_i)\) is the experimental error at the momentum transfer \(S_i\). Low-resolution models were also determined from the ab initio dummy residue modelling program GASBOR, which represents the particle in solution as a protein-like chain of dummy residues, thus representing more accurately the internal structure of the particle than do the shapes determined from DAMMIF.

The results of multiple DAMMIF and GASBOR reconstructions were compared using the alignment program SUPCOMB to determine the most representative (typical) model from each of the ab initio methods. Averaged DAMMIF and GASBOR models were also determined using the program DAMAVER and these models were adjusted such that they agree with the experimentally determined excluded volume using the program DAMFILT.

Molecular modelling of fH constructs was conducted using, as rigid bodies and where appropriate, the same structures as those used in the EOM. Rigid-body models were generated for the fH12–13, fH11–14, and fH10–15 constructs using the program BUNCH, where linkers/loops between the individual CCP subunits are represented as flexible chains composed of dummy residues. Domain boundaries were defined as the first and the last cysteine residues in each CCP module. For BUNCH rigid-body modelling, either single scattering curves or multiple curves were used for refinement of the fH11–14 and fH10–15 constructs (i.e., the scattering curves of smaller fH constructs were used for simultaneous refinement). During the modelling procedure, no NOE between CCP modules was enforced. The results of multiple BUNCH runs were analysed using the programs SUPCOMB and DAMAVER to identify the most representative/typical models.

The sedimentation coefficient \(s_{20w}\) of the SAXS-generated bead models of all factor H constructs was computed by HYDROPRO. The values obtained for individual models and for the averaged models provided by DAMAVER were analysed to assess uncertainty in the SAXS-derived values of \(s_{20w}\).

### Accession number

Structural data have been deposited in the PDB with accession code 2KMS (BioMagResBank entry number 16439).

### Supplementary Data

Supplementary data associated with this article can be found, in the online version, at doi:10.1016/j.jmb.2009.10.010

### References


