The solution structure of the complement deregulator FHR5 reveals a compact dimer and provides new insights into CFHR5 nephropathy

Nilufar Kadkhodayi-Kholghi\(^1\), Jayesh S. Bhatt\(^1\), Jayesh Gor\(^1\), Lindsay C. McDermott\(^2\), Daniel P. Gale\(^3\) and Stephen J. Perkins\(^1\)

\(^1\) Department of Structural and Molecular Biology, Division of Biosciences, Darwin Building, University College London, Gower Street, London WC1E 6BT, United Kingdom
\(^2\) School of Life Sciences, University of Bedfordshire, Park Square, Luton LU1 3JU, United Kingdom
\(^3\) UCL Department of Renal Medicine, Royal Free Hospital, University College London, London NW3 2PF United Kingdom

Running Title: Solution structure of FHR5

‡ Author to whom correspondence and requests for reprints should be addressed (SJP: Tel: 020-7679-7048; Fax: 020-7679-7193; email s.perkins@ucl.ac.uk).

Keywords: analytical ultracentrifugation; atomistic modelling; FHR5; Complement; molecular dynamics; molecular modelling; Monte Carlo simulations; small angle X-ray scattering

ABSTRACT

The human complement Factor H-related 5 protein (FHR5) antagonizes the main circulating complement regulator Factor H, resulting in the deregulation of complement activation. FHR5 normally contains nine short complement regulator (SCR) domains, but a FHR5 mutant has been identified with a duplicated N-terminal SCR-1/2 domain pair that causes CFHR5 nephropathy. To understand how this duplication causes disease, we characterized the solution structure of native FHR5 by analytical ultracentrifugation and small-angle X-ray scattering. Sedimentation velocity and X-ray scattering indicated that FHR5 was dimeric, with a radius of gyration \( R_G \) of 5.5 ± 0.2 nm and a maximum protein length of 20 nm for its 18 domains. This result indicated that FHR5 was even more compact than the main regulator Factor H which showed an overall length of 26-29 nm for its 20 SCR domains. Atomistic modelling for FHR5 generated a library of 250,000 physically-realistic trial arrangements of SCR domains for scattering curve fits. Only compact domain structures in this library fit well to the scattering data, and these structures readily accommodated the extra SCR-1/2 domain pair present in CFHR5 nephropathy. This model indicated that mutant FHR5 can form oligomers that possess additional binding sites for C3b in FHR5. We conclude that the deregulation of complement regulation by the FHR5 mutant can be rationalized by the enhanced binding of FHR5 oligomers to C3b deposited on host cell surfaces. Our FHR5 structures thus explained key features of the mechanism and pathology of CFHR5 nephropathy.

Complement activation and regulation is of major importance in enabling clearance of pathogens, whilst preventing complement-mediated host cell damage. Complement factor H related 5 protein (FHR5) was first identified co-localised with C3 in glomerular immune...
deposits from patients with glomerulonephritis, and is a member of a family of structurally related proteins comprising the major serum complement regulator Factor H and five complement factor H related proteins. Factor H, comprising 20 short complement regulator (SCR) domains, has been well characterised, both in terms of its structure and function, binding to activated C3b and its fragment C3d, and regulating excess C3 activation (1). However, the principal physiological function of FHR5 is poorly understood. FHR5 circulates in plasma in extremely low concentrations of 3-6 µg/ml (2), which is approximately 100-fold lower than Factor H. It is also the least abundant of the FHR proteins, yet its structure is the longest of these proteins, with a linear sequence of nine SCR domains (Figure 1). The SCR domain (3) is the major domain type found in the complement regulators. An SCR domain is characterised by a consensus sequence of approximately 61 amino acids, with four invariant cysteine residues that form two disulphide bridges (I-III and II-IV) and a conserved tryptophan residue. It folds compactly, with a hydrophobic core, in a β-sandwich arrangement of six hydrogen bonded β-strands. The key C-terminal C3b/C3d recognition sites are conserved between SCR-19/20 of Factor H and SCR-8/9 of FHR5 (Figure 1). FHR5 also interacts with heparin (2), however FHR5 has no complement regulatory domains equivalent to SCR-1/4 of Factor H. FHR5 forms native homodimers via its two N-terminal domains SCR-1/2 that exhibit increased avidity for C3b/C3d compared with the monovalent Factor H, and, although early studies using supra-physiological concentrations of FHR5 showed evidence of weak (compared with Factor H) complement regulating activity (2), more recent work has shown that, at physiological concentrations, FHR5 competitively antagonises Factor H, thus deregulating complement (4,5). Conflicting data exist on whether FHR5 forms heterodimers with other FHRs in vivo (6,7).

CFHR5 nephropathy, a monogenic cause of kidney failure endemic in Cypriots (individuals residing in or with ancestry from the island of Cyprus), is characterised in almost all affected individuals by persistent microscopic haematuria and, in a proportion of patients, episodes of kidney damage and visible blood in the urine that occur at times of otherwise trivial mucosal infections, with repeated episodes typically resulting in progressive kidney damage and eventually end stage kidney failure occurring in >80% affected males and <20% of affected females by the age of 55 years. Kidney biopsy shows predominantly mesangial-based glomerular inflammation with deposition of C3 but not immunoglobulins in the mesangium and, under electron microscopic examination, the subendothelial part of the glomerular basement membrane – appearances termed C3 glomerulopathy that suggest defective regulation of the complement system. The disease is a highly penetrant autosomal dominant disorder that is caused by heterozygosity for an in-frame duplication of exons 2 and 3 of the CFHR5 gene that results in production of an elongated FHR5 protein with an extra two N-terminal SCR-1/2 domains in tandem. No extra-renal features of the disease have been reported, despite the review of clinical data from over 100 affected individuals of all ages (8,9). The molecular mechanisms that make the kidney susceptible to complement-mediated damage in CFHR5 nephropathy and other common causes of glomerulonephritis (e.g. lupus nephritis and immunoglobulin A (IgA) nephropathy, in which flares of disease triggered by mucosal infections also occur) are not well understood.

Protein structural studies of full-length FHR5 are complicated by its large size and its eight potentially-flexible inter-domain linkers of lengths between three
and eight residues (Figure 1), both of which make it difficult to crystallise in order to determine its three-dimensional appearance. To date, atomic-level structures have not been determined for any small FHR5 fragments. However, alternative methods can be used for structural studies. Previously for full-length factor H, electron microscopy, small-angle X-ray scattering (SAXS), analytical ultracentrifugation (AUC), and molecular modelling showed that full-length factor H has a partially folded-back structure that is relevant to its regulatory function (10-12). This combination of analytical ultracentrifugation, X-ray solution scattering and atomistic modelling has been effective in determining many macromolecular structures in solution (13-15). Many of the first structural explanations for factor H-associated diseases such as atypical haemolytic uraemic syndrome were based on homology models for the SCR domains (16-18). Here, these solution structural and modelling approaches were applied to determine the solution conformation of full-length FHR5 in order to explain its role in healthy individuals and how CFHR5 nephropathy may arise through the SCR-1/2 duplication. Following SAXS and AUC data collection, full-length FHR5 was modelled using molecular dynamics, followed by Monte Carlo simulations to generate a large library of physically-realistic trial atomistic structures for the FHR5 dimer (14,19). The theoretical scattering profiles of this library were compared to the experimental SAXS curves to determine best-fit FHR5 structures. We thus defined a small subset of compact folded-back solution structures. The extra SCR-1/2 domain pair in mutant FHR5 was readily added to these structures, their presence leading to the formation of multivalent oligomers of FHR5. Our work explains how FHR5 regulates complement activation in the kidney and how CFHR5 nephropathy arises.

Results

Purification of full-length FHR5

Human FHR5 SCR-1/9 purchased from Creative Biolabs was subjected to gel filtration chromatography to ensure monodispersity and removal of aggregates prior to SAXS experiments. The protein eluted as a single symmetrical peak at approximately 15 ml elution volume (Figure 2A). This was preceded by a broader peak that was eluted between 10 and 14 ml, which was attributed to protein aggregates. Only the protein fractions between 14.3-16.3 ml (red in Figure 2A) were retained. By SDS-PAGE (Figure 2B), a single band was seen at 60-66 kDa (non-reduced) that corresponds well to the expected monomer molecular mass of 62.4 kDa. Reducing conditions resulted in another single band but at a slightly lower mass, this difference being attributed to the presence of glycan chains on FHR5.

SEC-MALLS was used to determine the mass and self-association of FHR5 in our Tris-150 purification buffer, as in previous work (4). FHR5 from a size-exclusion column was detected by UV (blue, Figure 3) and refractive index (green) measurements, in parallel with multi-angle light scattering (red) to analyse size distributions. Three peaks were observed in the elution profile. Peak 1 at 2.7-4.2 min was assigned as aggregated material, because this had a lower UV and refractive index, but high light scattering intensities that indicated very large sizes. Its molecular mass was calculated to be above 5,400 kDa. Peak 2 at 4.9-5.2 min was the FHR5 dimer that eluted with higher UV and refractive index values but with lower light scattering. Its molecular mass was estimated as 162 kDa, this being consistent with FHR5 dimer formation, given that the mass of the monomer was 62.4 kDa from its composition (20). Despite a large inherent error associated with light scattering, no evidence of a FHR5 monomer peak was detectable. A small peak 3 at 7.6-7.9 min was assigned to fragments below 30 kDa.
AUC analyses of FHR5

AUC sedimentation velocity experiments on FHR5 studied its oligomerisation and shape using size distribution $c(s)$ analyses to determine its molecular mass and sedimentation coefficient $s_{20,w}$. Absorbance data for FHR5 at 0.16 mg/ml in PBS were collected for five different salt concentrations between 20-250 mM NaCl. SEDFIT analyses involved as many as 500 absorbance scans. The experimental sedimentation boundaries (left, Figure 4) gave good fits to the Lamm equation to give the size-distribution $c(s)$ profiles (right, Figure 4), despite the low concentrations in use. These fits were obtained by floating the meniscus, bottom of the cell, the baseline and the frictional ratio $f/f_0$ of around 1.5.

Protein aggregation was visible in the earliest boundaries that sedimented rapidly at the start of the runs, to leave behind the FHR5 dimer that sedimented more slowly (Figure 4). This agreed with SEC-MALLS. A major $c(s)$ peak at 6.0 S was observed for FHR5 in PBS-137 that corresponded to an average molecular mass of 134 kDa. This mass confirmed the presence of dimer in solution. The aggregates made little contribution to the $c(s)$ analyses between 3-12 S, even though they contributed as much as half the protein present. The molecular masses for the five buffers were between 133 kDa to 139 kDa (Table 1), showing that the FHR5 dimer was stable between 20 mM to 350 mM NaCl. The $c(s)$ analyses did not reveal any FHR5 monomer at lower $s$ values. The reproducibility of these data was tested at two different rotor speeds of 40,000 rpm and 50,000 rpm, to show no difference.

The solution structure of FHR5 between 20 mM to 350 mM NaCl was monitored using the mean $s_{20,w}$ values (Table 1). A significant decrease of 0.9 S from 6.48 S to 5.35 S was seen on going from 20 mM NaCl to 350 mM NaCl. This shift in the FHR5 dimer peak was visible in the $c(s)$ distribution plots (vertical dashed lines, Figure 4). This result indicated a conformational change in FHR5, where the smaller $s_{20,w}$ values at high NaCl concentration indicated a more elongated FHR5 domain structure that formed as the ionic strength was increased (Figure 5).

SAXS analyses of FHR5

SAXS was used to study the solution structure of the FHR5 dimer in concentration series in three different buffers, two being physiological (PBS-137 and Tris-150) and one being low salt (PBS-50). The FHR5 samples were purified by gel filtration (Figure 2). In Tris-150, data were collected using 0.04-0.5 mg/ml FHR5. In PBS-137 and PBS-50, data were collected using 0.04-0.17 mg/ml FHR5. Guinier analyses of the solution structure gave high quality linear plots in two distinct regions of the $I(Q)$ curves that corresponded to the radius of gyration $R_G$ and the cross-sectional radius of gyration $R_{XS}$ from two distinct $Q$-ranges (Figure 6). These values are measures of the overall and the shorter dimensions of macromolecular elongation respectively. Their values were deduced according to Equations (1) and (2) respectively, within satisfactory $Q.R_G$ and $Q.R_{XS}$ limits close to 1.0:

(i) In the overall structural Guinier $R_G$ analyses in a low $Q$-range of 0.1 - 0.27 nm$^{-1}$ (Figure 7A), in Tris-150 and PBS-137 buffers with similar NaCl concentrations, the mean $R_G$ values were $5.36 \pm 0.14$ nm and $5.48 \pm 0.17$ nm respectively. However, in the PBS-50 buffer with lower NaCl, the mean $R_G$ value increased slightly to $5.91 \pm 0.13$ nm. This increase was attributed to trace aggregation in FHR5 that affected the lowest $Q$ values (Figure 7A). No concentration dependence was observed for the $R_G$ values between 0.04 and 0.17 mg/ml, however a slightly increased $R_G$ value of up to 0.2 nm was seen at 0.2-0.5 mg/ml FHR5.

(ii) In the cross-sectional Guinier $R_{XS}$ analyses, using a $Q$-range of 0.32 – 0.55...
nm$^{-1}$ (Figure 7B), the mean $R_{XS}$ values in each buffer were $2.41 \pm 0.06$ nm, $2.29 \pm 0.09$ nm, and $2.46 \pm 0.14$ nm for Tris-150, PBS-137 and PBS-50 respectively (Table 2). No significant changes in the $R_{XS}$ values were seen between the data sets for these NaCl and protein concentrations, indicating that the cross-sectional structure of FHR5 was unchanged in conformation.

The distance distribution function $P(r)$ in real space represents all the distances between pairs of atoms in FHR5. This was calculated from Fourier transformation of the full $I(Q)$ scattering curve following the specification of the maximum dimension $D_{max}$ (Equation (3); Figure 8). The $P(r)$ curve provided an independent $R_G$ value for FHR5 for comparison with the Guinier value (Table 2). The $R_G$ values from the $P(r)$ analyses were in good agreement with those from the Guinier analyses (Table 2). The $P(r)$ curve also gave the maximum length $L$ of FHR5 from the value of $r$ when $P(r)=0$. The mean $L$ values were $19.5 \pm 0.4$ nm in Tris-150 (Figure 8C), $19.6 \pm 0.5$ nm in PBS-137 (Figure 8B) and $21.0$ nm in PBS-50 (Figure 8A). The $L$ value for PBS-50 was slightly higher than those in Tris-150 and PBS-137, most likely due to trace aggregation that resulted from the lower ionic strength used (see above). A single maximum $M$ was observed in all the $P(r)$ curves. This corresponded to the most frequent interatomic distance within the FHR5 structure (Table 2). The mean $M$ values were $4.9 \pm 0.3$ nm, $4.9 \pm 0.1$ nm, and $5.4 \pm 0.3$ nm for Tris-150, PBS-137 and PBS-50 respectively. The $M$ values were relatively stable, although slightly higher for PBS-50 as the result of trace aggregates.

**Initial model for the FHR5 dimer**

Currently, there is no atomic level structural information on FHR5. To determine an atomistic-level solution structure for the FHR5 dimer, a starting model for the monomer was required. This was created by comparative modelling based on four known SCR crystal structures as structural templates (Figures 1B-C). Two used related crystal structures of the N-terminal FHR1 SCR-1/2 domains and the C-terminal FHR2 SCR-3/4 domains with high sequence identities of 85.2% and 61.7% respectively with SCR-1/2 and SCR-8/9 of FHR5. The SCR-3/7 domains of FHR5 shares significant sequence similarities with the SCR-10/14 domains of Factor H. Although templates for individual SCR3/7 domains in FHR5 were searched for in PDB-Blast, the best choices were these domain structures from Factor H due to their direct sequence similarities (Figure 1C). FHR5 SCR-3/4 was represented by Factor H SCR-10/11 with a high sequence identity of 57.4%. FHR5 SCR-5/6 was represented by Factor H SCR-12/13, also with a high sequence identity of 53.9%. While FHR5 SCR-7 is similar to Factor H SCR-14, no structure existed for Factor H SCR-14. Searches showed that the best template structure for FHR5 SCR-7 was that of SCR-11 of Factor H with a sequence identity of 34.5%. The individual template-target sequence alignments (Figure 1C) showed no significant indels in the structure, because the number of residues in these were well aligned. Thus the FHR5 SCR-7 and SCR-8/9 sequences had only one gap inserted in each. The individual modelled domains satisfied validation checks using PROCHECK, where the Ramachandran plots showed that 70% of the residues were in the most favoured steric regions. The FHR5 dimer was generated from its monomer structure by aligning its SCR-1/2 domains with the crystal structure of the FHR1 SCR-1/2 dimer (Experimental Procedures), followed by energy minimisation to relax this starting structure.

**Modelling the solution structure of the FHR5 dimer**

Atomistic modelling of the FHR5 scattering data established the best-fit FHR5 dimer structures, hence providing a molecular explanation for its solution
structure. The scattering curves for 0.17 mg/ml and 0.5 mg/ml FHR5 in Tris-150 were used in order to assess good quality curves with no traces of aggregation and better signal-noise ratios at 0.5 mg/ml (Figure 9). Data for 0.5 mg/ml were not available in PBS-137 or PBS-50, and traces of aggregates were present in PBS-50 buffer, thus these data sets were not used.

The starting structure for the FHR5 dimer represented an extended conformation of the 18 SCR domains (Figure 9). Each SCR domain was held fixed in conformation. Because as many as 14 linkers between the 18 domains were potentially variable, three different Monte Carlo conformational searches were set up. As detailed in Table 3, these varied all 14 linkers (Search 1), or eight linkers in which the crystal structure-observed linkers were kept fixed (Search 2), or four linkers after every third SCR domain (Search 3) (Figure 1B) (Experimental Procedures). Initial Monte Carlo conformational simulations in Searches 1-3 gave many models that were too elongated with too large $R_G$ values and few models with low $R_G$ values close to the experimental $R_G$ value of ~5.5 nm. Thus, in further simulations, models were selected with $R_G$ values closer to the experimental $R_G$ value to generate further conformers, but now using an $R_G$ cut-off of 6.0 nm as constraint to generate more compact FHR5 dimers. This resulted in more structures with lower $R_G$ values; however, many of these models were rejected by the workflow because the more compact shapes gave rise to physically-disallowed steric clashes between the SCR domains.

All six analyses from the three Searches at two FHR5 concentrations gave a clear single minimum in the distribution of $R$-factor goodness of fit values (Figure 10). A lower $R$-factor indicated a better fit to experiment. Thus all three Searches successfully generated good-fit solution structures for the FHR5 dimer. Starting from 200,000-250,000 trial structures in Searches 1-3, 86,732 structures with no steric clashes were accepted for Search 1, and likewise 72,755 structures for Search 2, and 123,776 structures for Search 3 (yellow in Figure 10). To verify the Monte Carlo-generated conformations, a grid density plot was generated for the Search 2 library of models (Figure 9). The volumetric data showed that a full conformational range of structures had been sampled, in comparison with the starting FHR5 dimer model at the centre of the grid. Significantly, the experimental $R_G$ value of 5.36 nm occurred at the left of the distribution plots in Figure 10, clearly indicating that FHR5 has a compact domain structure. In distinction, linear FHR5 models showed higher $R_G$ values of over 8 nm.

The three sets of 72,755-123,776 models were each filtered to identify the best-fit structures (Table 3). The appearance of the $R_G$ vs. $R$-factor graphs was similar in all six fits (Figure 10). This showed that the outcome of the modelling was independent of the assumption used to generate the linkers. As required, the dimer models with the lowest $R$-factors of 4-5% agreed well with the experimental $R_G$ value of 5.36 nm. The most extended FHR5 structures with the largest $R_G$ values of 8 nm and above showed the highest $R$-factors of ~30%. No models had an $R_G$ of 4.5 nm or less because such a dimer would be too compact to be sterically allowed. Filters were now used to reject poor-fit structures. First, a ± 5% experimental $R_G$ filter was used to reject models that had $R_G$ values outside this range, followed by a ± 5% $R_{XS}$ filter. Models with an $R$-factor below 6% were then selected. For the two fits of Search 1 (Table 3), totals of 28 and 131 models were identified (green in Figure 10A). For Search 2, totals of 55 and 52 models were identified (Figure 10B). For Search 3, totals of 694 and 749 models were identified (Figure 10C). These best-fit models formed a single cluster of fits at the $R$-factor minima. The best-fit models with the lowest $R$-factors (red in Figure 10) had...
R-factors of 4.5% and 4.2% for Search 1, 4.7% and 3.9% for Search 2, and 4.3% and 3.8% for Search 3. For comparison, the parameters for the best-fit 100 models were also shown in Table 3.

Visual inspection of the fits between the theoretical and experimental SAXS \( I(Q) \) and \( P(r) \) curves showed good agreement (Figure 11). The \( M \) and \( L \) values of the \( P(r) \) curves were well reproduced. Kratky plots of the SAXS curves monitor whether the protein was compact and globular or was extended and disordered in its structure. The normalised Kratky analyses of \( (Q.R_G)^2.\frac{I(Q)}{I(0)} \) vs \( Q.R_G \) for the three best-fit models from Searches 1-3 and the experimental curve at 0.5 mg/ml showed that a clear peak was seen at a \( Q.R_G \) of 2.26 (Figure 12). Good fits to the experimental curve were also obtained at larger \( Q.R_G \) values for all three best-fit models. The Kratky plot thus showed that FHR5 possessed a globular structure with little inter-domain flexibility. In comparison, our recent Factor H models showed poorer fits at larger \( Q.R_G \) values, indicating that the 20 SCR domains in Factor H had more flexibility (12). This comparison indicated that the structure of FHR5 was well-formed, and this was less flexible in structure than full-length Factor H.

Because all three searches gave similar good fits, Search 2 was selected for the final output because this most closely resembled the crystal structures for the SCR domain pairs used to construct it. To understand better the 55 best-fit structures from Search 2 (available in Supplementary Materials), they were clustered into conformational families using principal component analysis (Figure 13) (21,22). Principal component analysis determines the correlated motions of protein residues as linearly uncorrelated variables termed principal components. These “essential motions” are extracted from a covariance matrix of the atomic coordinates of the frames in the trajectory. The eigenvectors of this matrix each have an associated eigenvalue that characterises the clustering of the models based on structural coordinates (or variance). By this, the first three eigenvalue rankings (PC1 to PC3) accounted for a variance of 68.9% in the 55 best-fit FHR5 models. The median FHR5 structure from each principal component analysis group consistently revealed folded-back N-terminal domains and extended C-terminal domains (Figure 14).

**Sedimentation coefficient modelling of the FHR5 dimer**

As an independent test of the SAXS modelling, the theoretical \( s_{20,w} \) values were calculated using HYDROPRO for the best-fit FHR5 dimer models obtained from the three Searches 1-3 (Table 3). The six best-fit models gave a mean \( s_{20,w} \) value of 5.3 ± 0.2 S. This compared well with the experimental \( s_{20,w} \) value in PBS-137 of 5.97 ± 0.2 S (Table 1). The typical accuracy of the \( s_{20,w} \) calculation is ± 0.21 S (23). The difference of 0.67 S may result from potential trace aggregates remaining in the X-ray sample which would increase the experimental and modelled \( R_G \) values of FHR5 and in turn decrease the modelled \( s_{20,w} \) value.

**Discussion**

Up to now, the domain organization of FHR5 was unknown. Here we present the first protein structures for the FHR5 dimer by a combination of SAXS and AUC in conjunction with molecular simulations. Previously, it was often thought that FHR5 possessed nine SCR domains in a flexible linear conformation (4,8,24-26). Instead, our analyses now show that FHR5 is dimeric and adopts a compact domain conformation. Such a structure readily leads to FHR5 oligomer formation in the presence of mutant FHR5 protein (see below). This structure revises our understanding of how FHR5 interacts with its target ligand C3b and its C3d fragment, as well as others such as heparin-like...
analogue. It also explains the molecular defect underlying CFHR5 nephropathy.

New understandings of the FHR5 solution structure were determined:

(i) Our SEC-MALLS and AUC data showed that full-length FHR5 SCR-1/9 is a dimer (Figures 3 and 4), in agreement with previous results for the FHR proteins (4,6,7). In addition, AUC monitors macromolecular shapes through the $s_{20,w}$ values which measures macromolecular elongation. Of interest here was that, not only did the $s_{20,w}$ values correspond to a much more compact protein than expected from the 18 domains in the dimer (Table 3), but also these $s_{20,w}$ values decreased with an increase in the NaCl concentration of the buffer. This decrease implied that the compact structure became more elongated through the weakening of charge-charge interactions between the SCR domains. The predicted pI values of the N-terminal five domains SCR-1/5 were mostly acidic at 4.6, 5.4, 8.5, 4.7 and 4.3 in that order, while the predicted pI values of the four C-terminal domains SCR-6/9 were mostly basic at 9.6, 6.3, 8.9 and 8.4 in that order (http://web.expasy.org/protparam/). Differences in these individual pI values may facilitate the formation of a more compact FHR5 domain structure through charge attractions in physiological 137 mM NaCl salt.

(ii) The SAXS data provided more detailed views of the FHR5 structure. Interestingly, given that the SAXS technique is sensitive to aggregate formation, both FHR5 and Factor H turned out to be aggregation-prone. The $R_G$ value of Factor H was originally reported to be 12.4 nm in the first SAXS studies in 1991 for reason of being aggregated; with improved Factor H purifications, this value has now diminished to 7.22-7.77 nm (12). Factor H aggregates in storage conditions. FHR5 as supplied for our study showed aggregation by SEC-MALLS and AUC, and these aggregates were removed by size-exclusion chromatography. SAXS showed that the $R_G$ and $R_{XS}$ values for FHR5 were relatively constant in 50-150 mM NaCl and between 0.1-0.5 mg/ml, although residual trace aggregates were detectable in 50 mM NaCl buffer. The maximum length $L$ of FHR5 was 20-21 nm in all buffers (Table 2). A single SCR domain is about 4 nm in length. A hypothetical fully-extended FHR5 domain arrangement (Figure 1A) would be predicted to be of length 64 nm, or over three-fold longer than seen experimentally (Figure 8). Likewise Factor H is predicted to be 80 nm in length if fully-extended, but was observed to be only 26-29 nm in length, so again such an extended structure is also predicted to be three-fold longer than seen experimentally (12). Both FHR5 and Factor H thus have similar folded-back domain structures.

(iii) Because no high resolution FHR5 domain structures were available, the starting model for FHR5 was generated by standard homology modelling methods based on sequence similarities. The FHR5 SCR-1/2 and SCR-8/9 structures were readily modelled on other FHR proteins. These modelled domain pairs were notable for their short linker lengths of three residues each, suggesting that these linkers were relatively inflexible (Figure 1C). The longest inter-SCR linkers occurred between SCR-3/7, which were six, six, eight and seven residues in length respectively. Interestingly, the same linker lengths occurred in SCR-10/14 of Factor H. In fact, sequence similarities showed that these five SCR domains resembled SCR-3/7 of FHR5. These Factor H domains contributed significantly to its folded-back solution structure (11,12,27,28). These long linkers in Factor H and FHR5 contained a high proportion of charged residues, particularly lysine and glutamate, and are conserved in mouse and bovine factor H (29). Indeed, SCR-10/14 of factor H, not only has longer inter-domain linkers, but also shorter SCR sequences and higher glycosylation levels (30). These similarities imply that these middle domains act as conformational spacers that result in more compact domain
structures that enable the multiple factor H and FHR5 binding sites to act synergistically.

(iv) The Monte Carlo simulations generated a large conformational library of possible SCR arrangements in FHR5, from which best-fit structures were identified. These best-fit structures accounted for the experimental SAXS and AUC data for FHR5. Interestingly the Kratky plots (Figure 12) did not show evidence of disorder or flexibility in the FHR5 solution structure, meaning that its structure was well-defined. The molecular structures for FHR5 and Factor H show similar folded back and compact SCR structures (Figure 14 A-D). From the principal component analyses, the three best-fit FHR5 structures (Supplementary Materials) showed that, while the SCR-1/2 dimer pair was consistently buried in the dimer core in all three structures, the two SCR-3/4 domain pairs looped back across the SCR-1/2 core in a compact arrangement with SCR-1/2. The two C-terminal ends with SCR-5/9 were solvent-exposed and either extended away from the SCR-1/4 core or looped back towards this core. The functional SCR-8/9 domains thus showed a range of folded-back or extended conformations relative to a more compact SCR-1/4 core. The three best-fit conformations were able to interact with one or two C3d ligands (Figure 14A-C).

In terms of new functional insight obtained from this study, FHR5 is a complement deregulator that competitively inhibits factor H, an important regulator of C3b activation at host cell surfaces (4,5). From the nine best fit structures (Table 3), the two C3d or C3b binding sites found in dimeric FHR5 will have a C-terminal separation of around 10-20 nm. FHR5 would increase its avidity for C3d- or C3b-coated host cell surfaces only if bound C3d or C3b were present at a great enough spatial density on this surface, thus displacing the binding of the much more abundant Factor H. If the spatial density of C3b/C3d on surfaces is low, the much more abundant Factor H will preferentially bind to inhibit and degrade C3b there via Factor I-mediated cleavage. When the density of C3b/C3d is great enough to allow dimeric FHR5 binding to be functionally bivalent (Figure 14B) the FHR5-C3d interaction becomes stronger than the monovalent Factor H-C3d interaction. This reasoning indicates a mechanism for FHR5 to modulate Factor H activity.

In CFHR5 nephropathy, the heterozygous duplication of SCR-1/2 results in a more elongated FHR5 molecule that is detectable in the blood of patients (Figure 14E,F) (8). Other heterozygous genomic rearrangements that result in the production of more elongated FHR proteins with additional N-terminal SCR-1/2 domains have been described in association with autosomal dominant C3 glomerulopathy (31,32). Conversely, clear loss-of-function variants in CFHR5 occur at high frequency in the population and are not known to be pathogenic. Among the ~245,000 alleles tested in GnomAD (http://gnomad.broadinstitute.org/gene/ENSG00000134389), ~3000 variants predicted to stop FHR5 protein translation before the final exon are documented. In addition, 3% of the UK population is homozygous for a CFHR3/CFHR1 deletion polymorphism that results in the complete deficiency of FHR1 and FHR3 (33). Together, these observations suggest that a gain-of-function mechanism underlies CFHR5 nephropathy and that tandem duplication of the two N-terminal SCR-1/2 domains is necessary and sufficient to cause this. Structural simulations using our FHR5 models show that the extra SCR-1/2 domains of the mutant FHR5 protein are readily added, and these will be accessible to other FHR5 molecules (Figure 14E,F). At least two distinct mechanisms can be proposed by which the mutation in CFHR5 nephropathy causes augmented function (i.e. increased avidity for C3-coated surfaces). In one, as proposed previously, the presence of two
accessible SCR-1/2 dimerization motifs on the single mutant protein would allow trimers or higher order oligomers to form that would be tri- or multi-valent with respect to C3d (Figure 13E,F). In addition, the greater length of macromolecules containing mutant FHR5 would reduce the density of C3d on a host cell surface required for multivalent binding to occur, since the longer protein would have a greater steric range.

Overall, it is expected that different tissues will function differentially in respect of FHR5 or Factor H binding activity. It is possible that a high blood flow rate, such as that in the renal glomeruli, enables the density of C3b or C3d deposition to become high enough to allow FHR5 dimers to bind bivalently. This explains why FHR5 is enriched in C3-coated glomeruli and why CFHR5 gain-of-function mutations result in the purely renal disease of CFHR5 nephropathy, which manifests clinically at times of infection when the complement system is systemically activated. The striking clinical and histological similarity of IgA nephropathy to CFHR5 nephropathy, combined with the co-localization of FHR5 and C3 in the glomeruli in IgA nephropathy (1), raises the possibility that the FHR proteins, including FHR5, play an important role in both diseases. This possibility is supported by the observation that, in IgA nephropathy, a common polymorphic deletion of CFHR1 (which encodes the smaller dimeric complement deregulator FHR1) is protective (34). Our demonstration of a compact FHR5 dimer structure at a molecular level therefore reveals new aspects of how FHR5 antagonises Factor H function, amplifying complement activation at host cell surfaces when C3 deposition reaches a critical density, and leading to renal damage.

**Experimental Procedures**

**Purification and composition of full-length FHR5**

Mammalian-expressed (HEK293 cells) human FHR5 SCR-1/9 was purchased from Creative Biolabs (Shirley, NY, USA). This was prepared with a Histag which was cleaved off by the manufacturer. This protein was prone to aggregation. Aggregate-free FHR5 for SAXS was successfully purified from approximately 1 mg of protein that was pooled and concentrated using a Vivaspin 20 spin concentrator (Sartorius) with a 10 kDa molecular weight cut-off, then purified using a Superdex 200 10/300 GL gel-filtration column (Cytiva) equilibrated in 50 mM Tris, 150 mM NaCl, 1 mM EDTA, pH 7.4, using a Gilson HPLC system kindly made available by Dr A.J. Beavil (Kings College London). The FHR5 concentration was checked by the absorbance reading at 280 nm. Its purity and integrity was checked by SDS-PAGE before and after each SAXS and AUC experiment under reducing and non-reducing conditions using a Novex® 8-12% Bis-Tris Gel 1.0 mm (Invitrogen, Paisley, UK).

The amino acid composition of human FHR5 SCR-1/9 was determined from its sequence (SWISSPROT accession code: Q9BXR6). Two potential N-linked glycan sites were present at Asn126 and Asn400 (Figure 1A), and may be occupied by biantennary glycans as reported for Factor H (30). However, there was no evidence that these sites are occupied, in particular at Asn126, where glycan was not present in the crystal structure of HEK293-expressed FHR1 SCR-1/2 (PDB code: 3ZD2) (4). Since FHR1 SCR-1/2 has the same glycosylation sequence as that in FHR5 (Figure 1C), glycosylation was disregarded here. The mass of glycan-free wild-type FHR5 was predicted to be 62,377 Da from its sequence. Using the program SLUV (20), it has an unhydrated volume of 79.76 nm$^3$, a hydrated volume of 105.23 nm$^3$, a partial specific volume of 0.7278 nm$^3$, and an absorption coefficient of 15.59 (1%, 280 nm, 1 cm path length).
FHR5 samples were run through SEC-MALLS. This determines protein molecular masses using a standard HPLC system equipped with a Superdex 200 Increase 5/150 GL gel filtration column (Cytiva). The instrument was equipped with three detectors, namely a miniDawn detector (Wyatt Technology) which is a triple-angle light scattering detector, an Optilab DSP Interferometric Refractometer (Wyatt Technology) which measures refractive index changes, and an SPD-20A UV absorbance detector (Shimadzu Scientific). In multiple runs, 60 µl aliquots of FHR5 were loaded on the column via an injection loop. Following separation by size-exclusion, the three different detectors were combined in parallel to provide a molecular mass for the eluted sample. The chromatograms were analysed using ASTRA software (Wyatt Technology).

For AUC and SAXS experiments, FHR5 was dialysed into the appropriate buffer at 4°C prior to data collection. For AUC, these buffers were; PBS-20 (20 mM NaCl, plus 8.2 mM Na₂HPO₄, 2.6 mM KCl, 1.5 mM KH₂PO₄, pH 7.4), PBS-50 with 50 mM NaCl, PBS-90 with 90 mM NaCl, PBS-137 with 137 mM NaCl (standard physiological salt), PBS-250 with 250 mM NaCl, and PBS-350 with 350 mM NaCl. For SAXS, three buffers were used, namely Tris-150 (50 mM Tris, 150 mM NaCl, 1 mM EDTA, pH 7.4), and PBS-50 and PBS 137 as above. The experimental buffer densities were measured at 20°C using an Anton Paar DMA 5000 density meter, and their theoretical values were calculated from SEDNTERP (35). The resulting densities were 1.000538 g/ml for PBS-20 (theoretical, 1.00052 g/ml), 1.001714 g/ml for PBS-50 (theoretical, 1.00176 g/ml), 1.003382 g/ml for PBS-90 (theoretical, 1.00342 g/ml), 1.005054 g/ml for PBS-137 (theoretical, 1.00524 g/ml), 1.009960 g/ml for PBS-250 (theoretical, 1.00999 g/ml), 1.013920 g/ml for PBS-350 (theoretical, 1.01406 g/ml), and 1.00650 g/ml for Tris-150 (theoretical, 1.00603 g/ml). A solvent viscosity of 0.01002 cp was used throughout in the AUC analyses.

**Sedimentation velocity data collection and analyses for FHR5**

AUC data were obtained on a Beckman XL-I instrument, equipped with an eight-hole AnTi50 rotor (Beckman-Coulter Inc., Palo Alto, CA). Ultracentrifugation caused any aggregates present to sediment rapidly, leaving the soluble FHR5 protein visible for analysis. Approximately 400 µl of FHR5 sample was loaded into standard AUC double-sector cells for sedimentation velocity experiments at 20°C, equipped with sapphire windows and with 12 mm column heights. Sample concentrations were 0.16 mg/ml, therefore absorbance optics was used to collect data. Up to 500 consecutive scans were recorded until the protein had fully sedimented. The AUC runs were performed using two rotor speeds of 40,000 rpm and 50,000 rpm to check for reproducibility.

Data analysis was performed using SEDFIT software (version 14.6) (36, 37), using direct boundary Lamm fits of up to 50 selected scans at appropriately spaced time intervals. A c(s) size-distribution analysis was carried out, which assumes that all species have the same frictional ratio f/f0. The c(s) distribution was optimized by floating the value of the meniscus and bottom of the cell positions, the baseline and the frictional ratio f/f0 (set at 1.2 to begin with). Fits were carried out until satisfactory visual fits and overall root mean square deviations were obtained. The final SEDFIT analysis used a resolution of 200, and the sedimentation coefficient s₂₀,w for FHR5 was determined from the peak maximum in the c(s) size-distribution plot. The c(s) integration function was also used to derive the percentage of oligomers in the total loading concentration if required.

**SAXS data collection and data analyses for FHR5**
SAXS experiments were carried out in one beam session on the BM29 BioSAXS beamline at the European Synchrotron Radiation Facility, Grenoble, France, operating with a ring energy of 6.0 GeV. Data was acquired using a Pilatus 1M two-dimensional detector with a pixel size of 172 µm. The sample-to-detector distance was 3.0 m. The beamline was equipped with an automatic sample changer, and the samples were loaded using the thermo-regulated PCR tube configuration in the BsxCuBE control interface. The FHR5 samples were measured in three buffers at concentrations of 0.04 mg/ml, 0.09 mg/ml, 0.13 mg/ml, and 0.17 mg/ml. Additional data sets were collected at 0.2 mg/ml, 0.3 mg/ml, 0.4 mg/ml, and 0.5 mg/ml concentrations in Tris-150 buffer. Data was collected in triplicate from a total sample volume of 50 µl per run. An exposure time of 1 sec was used, and the absence of radiation damage was monitored from continuous automatic online checks. A total of 10 frames were collected as the sample was passed continuously through a quartz capillary tube (1.8 mm in diameter) to minimise radiation damage due to exposure. The final time-frames were merged, excluding any damaged data, to improve the signal-to-noise ratio. Between each sample measurement, the sample capillary was cleaned using Hellmanex® and water to ensure the removal of any residual protein or aggregates on the capillary walls.

The raw scattering data files were corrected by subtraction of the buffer data from the sample data. The resulting one-dimensional scattering curve \( I(Q) \) in a \( Q \)-range between 0.05 nm\(^{-1}\) to 2 nm\(^{-1}\) (where \( Q = 4\pi \sin \theta/\lambda \); \( 2\theta \) is the scattering angle and \( \lambda \) is the wavelength) represented the macromolecular structure. Guinier analysis of \( I(Q) \) against \( Q^2 \) at low \( Q \) values gave the radius of gyration \( R_G \), which is a measure of structural elongation if the internal inhomogeneity of the scattering densities has no effect, and the forward scattering at zero angle \( I(0) \).

\[
\ln I(Q) = \ln I(0) - \frac{R_G^2 Q^2}{3}
\]

(1)

\[
\ln[I(Q) \cdot Q] = [I(Q) \cdot Q]_{Q \to 0} - \frac{R_{XS}^2 Q^2}{2}
\]

(2)

The Guinier plots are usually valid in a \( Q \) range up to \( Q R_G \) values of 1.5 (39). If the macromolecular structure is elongated, the mean cross-sectional radius of gyration \( R_{XS} \) is obtained from plots of \( I(Q)Q \) against \( Q^2 \) in a larger \( Q \) range than those used for the \( R_G \) values. Using the SCT software package (40), the \( Q \) ranges for the \( R_G \) and \( R_{XS} \) values were 0.1 - 0.27 nm\(^{-1}\) and 0.32 – 0.55 nm\(^{-1}\) respectively.

Indirect Fourier transformation of the scattering curve \( I(Q) \) in reciprocal space (units in nm\(^{-1}\)) into real space (units in nm) gives the distance distribution function \( P(r) \). This transformation was carried out using the program GNOM (41).

\[
P(r) = \frac{1}{2\pi^2} \int_0^\infty I(Q)Qr \sin(Qr) dQ
\]

(3)

\( P(r) \) corresponds to the distribution of interatomic distances \( r \) in the macromolecule. In order to obtain the distance distribution \( P(r) \) curve, the full measured scattering curve was utilised. By specifying an assumed maximum dimension \( D_{max} \), the \( P(r) \) curve provides the macromolecular length \( L \) and the most common distance \( M \). The \( P(r) \) curves also provide an alternative calculation of \( R_G \) for comparison with the Guinier analysis.

**Generating of the starting model for FHR5**

Protein structural analyses of FHR5 were initiated from homology models for the nine SCR domains. Firstly, suitable
templates were selected based on high sequence and structural similarities. This was achieved by a combination of PDB-BLAST searches and sequence alignments between the five FHR-related proteins and Factor H. The final template was selected from the quality of the sequence alignment and its structural relevance. The template structures were taken from closely related structures in the Protein Data Bank (PDB). In the process, the amino acid sequence of FHR5 SCR-1/9 was used to replace the sequence of the template structure. These were constructed using closely-related structural templates using MODELLER (version 9.15) (42).

The closest template for each of the nine SCR domains, defined in terms of sequence identity and minimum insertions and deletions, was identified using CLUSTALO alignments (43). Four template structures for eight domains were used as follows (Figure 1B,C); FHR1 SCR-1/2 (PDB code: 3ZD2), Factor H SCR-10/11 (PDB code: 4B2R), Factor H SCR-12/13 (PDB code: 2KMS) and FHR2 SCR-3/4 (PDB code: 3ZD1). The ninth domain was SCR-7, for which a multiple sequence alignment (44) was performed using the NMR structures of Factor H SCR-10/11 (PDB code: 4B2R) (28) and Factor H SCR-11/12 (PDB code: 4B2S) (28), which provided an experimental structure for Factor H SCR-11. The full-length FHR5 model was evaluated using the SAVES server (https://services.mbi.ucla.edu/SAVES/), which incorporated validation criteria including PROCHECK and Ramachandran plots. The secondary structure and surface accessibilities of the FHR5 model were analysed using the Definition of Secondary Structure of Protein (DSSP) program (45). Structures were also modelled using SWISS-MODEL (46) to cross-check the models from MODELLER using another tool.

The PDB file for the dimer of FHR5 was generated by superimposing SCR-1/2 of each FHR5 monomer model onto the FHR1 SCR-1/2 dimer crystal structure (PDB code: 3ZD2), using PyMOL (DeLano Scientific). This structure was inputted directly into the atomistic modelling workflow of the SASSIE scattering curve fit package (19). First, the PDB file was manually corrected for gaps or errors in the amino acid sequence. A protein structure file (PSF), which contained molecule-specific information for the application of a force field, was generated via PSFGEN using Visual Molecular Dynamics (VMD) (version 1.9.2) (47). To create a physically-realistic atomistic model, the structure was subjected to 10 ps of energy minimisation using the molecular dynamics simulation package NAMD (version 2.9) (47, 48). The force field for this was CHARMM-36 (49, 50) and energy minimisation was performed using the conjugate gradient method.

**Molecular simulations and SAXS fitting of FHR5**

By excluding the dimerization interface at SCR-1/2 and linker L1 which do not vary in conformation (Figure 1A), FHR5 contains seven potentially flexible inter-SCR linkers L2 – L8 (Figure 1B). The linkers were subjected to peptide dihedral angle variations in the Monte Carlo simulations through the Markov sampling of backbone torsion angles (19). This allowed the rapid generation of a large conformational library of physically realistic atomistic models of the FHR5 SCR-1/9 dimer through the Complex Monte Carlo module of SASSIE. The same linkers on either monomer of the dimer were varied independently of each other, thus the resulting dimer structures were asymmetric in shape. In Search 1, all seven linkers (L2 – L8) were varied. These were defined as follows: L2 141-SFTKGE146, L3 202-KGQVRS207, L4 263-VEQVK268, L5 323-VATHQLKR330, L6 382-TEKREQF388, L7 443-VESTAY448, and L8 504-LDP506. In
Search 2, only L2, L4, L6 and L7 were varied, because these were not part of the crystal structures that were used (Figure 1B). In Search 3, only Linkers L3 and L6 were varied as a control of Searches 1 and 2. This strategy of independent simulations (Table 3) checked whether extra or fewer constraints in the linkers affected the resulting best-fit structures. During the Monte Carlo simulations, models with steric overlaps that were generated by SASSIE were excluded by specifying an atomic overlap distance cut-off of 0.3 nm. Simulations were continued to produce models with \( R_G \) values close to that of 6.0 nm obtained experimentally by filtering for a fixed range of \( R_G \) values in the FHR5 dimer models. The outputted structures were generated as binary format DCD files and visualised on VMD. In the three searches, a total of up to 250,000 models were generated in order to sample a sufficient number of conformations for the two monomers in the dimer.

Using the SCT module (40) in SASSIE, a theoretical scattering curve was calculated for each of the FHR5 dimer models. The atomic coordinates were converted into small spheres to generate a coarse-grained sphere model. A cube side length of 0.53 nm in conjunction with a cut-off of four atoms was used to generate unhydrated sphere models. Because the hydration shell was visible by X-rays, a hydration shell containing 0.3 g of H₂O/g of protein was added to each of the models by HYPRO (51). The theoretical scattering curve \( I(Q) \) for each model was calculated using the Debye equation adapted to spheres (40, 52).

The theoretical scattering curves for the dimer models were compared to the experimental SAXS curves. In the SCT Analyse module of SASSIE, the \( R_G \) and \( R_{XS} \) values were calculated from the modelled curves using the same \( Q \) ranges that were used for the experimental Guinier analyses. The curve fits were compared and filtered based on their \( R_G \) and \( R_{XS} \) values as well as their goodness-of-fit \( R \)-factor values defined as:

\[
R \text{ factor} = \frac{\sum ||I_{Exp}(Q)| - \eta I_{Theor}(Q)||}{\sum ||I_{Exp}(Q)||} \times 100
\]

(4)

where \( I_{Exp}(Q) \) and \( I_{Theor}(Q) \) were the experimental and theoretically calculated scattered intensities, and \( \eta \) was a scaling factor used to match the theoretical and experimental \( I(0) \) values. Typical best-fit \( R \)-factors for SAXS modelling are between 2% and 8% (14). To visualise the initial and best fit models for the FHR5 dimer, density plots were generated using the Density Plot module in SASSIE. The envelope was generated for the sterically-accepted trial models, sampled to produce the volumetric data, using the Gaussian cube file format. This was superimposed onto the initial FHR5 dimer model. The output files were rendered, analysed and annotated in VMD. Once the best-fit dimer models were chosen, their sedimentation coefficients were calculated for comparison with the AUC data, based on the atomic coordinates using the HYDROPRO shell modelling program (53).

Data Availability Statement
All data are contained within this manuscript. The 55 best-fit models from Search 2 and the 6 best-fit structures from Figure 11 are available in Supplementary Materials.

Acknowledgements
We thank Dr P. Pernot (ESRF, Grenoble) for excellent X-ray instrumental support, Dr David W. Wright for expert computational support, and Dr Andrew J. Beavil (King’s College London) for expert assistance with SEC-MALLS. Initial work in this project won an award for Best Poster Presentation at the XXVIth International
Complement Workshop in Kanazawa, Japan 4-8 September 2016.

Conflict of interest: The authors declare that they have no conflicts of interest with the contents of this article.

References


**FOOTNOTES**

N.K.K. was supported by the Rosetrees Trust for Medical Research (reference M187) and a UCL Impact Studentship. D.P.G. was funded by an MRC Clinician Scientist Fellowship. J.S.B. and S.J.P. were supported by a joint EPSRC (EP/K039121/1) and NSF (CHE-1265821) grant for the CCP-SAS project.
The abbreviations used are: AUC, analytical ultracentrifugation; FHR5, Factor H Related 5; MD, molecular dynamics; SAXS, small angle X-ray scattering.

**ORCID IDENTIFIERS**

<table>
<thead>
<tr>
<th>Name</th>
<th>ORCID ID</th>
</tr>
</thead>
<tbody>
<tr>
<td>Jayesh S. Bhatt</td>
<td>0000-0001-9265-755X</td>
</tr>
<tr>
<td>Jayesh Gor</td>
<td>0000-0002-1654-1392</td>
</tr>
<tr>
<td>Lindsay C. McDermott</td>
<td>0000-0003-4723-6056</td>
</tr>
<tr>
<td>Daniel P. Gale</td>
<td>0000-0002-9170-1579</td>
</tr>
<tr>
<td>Stephen J. Perkins</td>
<td>0000-0001-9218-9805</td>
</tr>
</tbody>
</table>
TABLE 1 Sedimentation velocity data for FHR5 SCR-1/9

<table>
<thead>
<tr>
<th>Sample</th>
<th>$s_{20,w}$ (S)</th>
<th>Molecular mass (kDa)</th>
<th>Frictional ratio ($f/f_0$)</th>
</tr>
</thead>
<tbody>
<tr>
<td>FHR5 in PBS-20</td>
<td>6.48 ± 0.05</td>
<td>139 ± 2</td>
<td>1.49 ± 0.0003</td>
</tr>
<tr>
<td>FHR5 in PBS-50</td>
<td>6.29 ± 0.1</td>
<td>133 ± 4</td>
<td>1.48 ± 0.02</td>
</tr>
<tr>
<td>FHR5 in PBS-90</td>
<td>6.19 ± 0.1</td>
<td>134 ± 6</td>
<td>1.53 ± 0.05</td>
</tr>
<tr>
<td>FHR5 in PBS-137</td>
<td>5.97 ± 0.2</td>
<td>134 ± 4</td>
<td>1.59 ± 0.03</td>
</tr>
<tr>
<td>FHR5 in PBS-250</td>
<td>5.91 ± 0.02</td>
<td>136 ± 8</td>
<td>1.58 ± 0.02</td>
</tr>
<tr>
<td>FHR5 in PBS-350</td>
<td>5.35b</td>
<td>138</td>
<td>1.80</td>
</tr>
</tbody>
</table>

*a All samples were measured at 0.16 mg/ml. The values represent the mean ± s.d. The data presented includes those collected both at 40,000 rpm and 50,000 rpm.

*b Only a single run was carried out for FHR5 SCR-1/9 in PBS-350
<table>
<thead>
<tr>
<th>Sample</th>
<th>$R_G$ (nm)$^a$</th>
<th>$R_G$ (nm)$^b$</th>
<th>$R_{XS}$ (nm)</th>
<th>$L$ (nm)</th>
<th>$M$ (nm)</th>
</tr>
</thead>
<tbody>
<tr>
<td>150 mM NaCl (Tris)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>0.5 mg/ml</td>
<td>5.52</td>
<td>5.85</td>
<td>2.45</td>
<td>21.0</td>
<td>4.68</td>
</tr>
<tr>
<td>0.4 mg/ml</td>
<td>5.56</td>
<td>5.92</td>
<td>2.51</td>
<td>21.0</td>
<td>4.68</td>
</tr>
<tr>
<td>0.3 mg/ml</td>
<td>5.56 ± 0.07</td>
<td>5.90</td>
<td>2.49 ± 0.05</td>
<td>21.0</td>
<td>4.52</td>
</tr>
<tr>
<td>0.2 mg/ml</td>
<td>5.63</td>
<td>5.94</td>
<td>2.38</td>
<td>20.0</td>
<td>4.92</td>
</tr>
<tr>
<td>0.17 mg/ml</td>
<td>5.28 ± 0.13</td>
<td>5.72</td>
<td>2.38 ± 0.02</td>
<td>20.0</td>
<td>4.75</td>
</tr>
<tr>
<td>0.13 mg/ml</td>
<td>5.40 ± 0.14</td>
<td>5.72</td>
<td>2.27 ± 0.08</td>
<td>19.0</td>
<td>4.98</td>
</tr>
<tr>
<td>0.09 mg/ml</td>
<td>5.37 ± 0.17</td>
<td>5.62</td>
<td>2.43 ± 0.11</td>
<td>19.5</td>
<td>4.99</td>
</tr>
<tr>
<td>0.04 mg/ml</td>
<td>5.30 ± 0.21</td>
<td>5.46</td>
<td>2.35 ± 0.20</td>
<td>19.5</td>
<td>5.60</td>
</tr>
<tr>
<td>137 mM NaCl (PBS)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>0.17 mg/ml</td>
<td>5.55 ± 0.05</td>
<td>5.67</td>
<td>2.28 ± 0.06</td>
<td>21.0</td>
<td>4.98</td>
</tr>
<tr>
<td>0.13 mg/ml</td>
<td>5.35</td>
<td>5.80</td>
<td>2.43 ± 0.05</td>
<td>21.0</td>
<td>5.00</td>
</tr>
<tr>
<td>0.09 mg/ml</td>
<td>5.61 ± 0.11</td>
<td>5.85</td>
<td>2.21 ± 0.11</td>
<td>21.0</td>
<td>4.87</td>
</tr>
<tr>
<td>0.04 mg/ml</td>
<td>5.63</td>
<td>5.88</td>
<td>2.74</td>
<td>21.0</td>
<td>4.75</td>
</tr>
<tr>
<td>50 mM NaCl (PBS)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>0.17 mg/ml</td>
<td>5.96 ± 0.09</td>
<td>6.21</td>
<td>2.41 ± 0.04</td>
<td>20.0</td>
<td>5.25</td>
</tr>
<tr>
<td>0.13 mg/ml</td>
<td>5.82 ± 0.14</td>
<td>6.11</td>
<td>2.42 ± 0.11</td>
<td>20.0</td>
<td>5.12</td>
</tr>
<tr>
<td>0.09 mg/ml</td>
<td>5.84 ± 0.05</td>
<td>6.26</td>
<td>2.37 ± 0.08</td>
<td>19.0</td>
<td>5.38</td>
</tr>
<tr>
<td>0.04 mg/ml</td>
<td>6.01 ± 0.27</td>
<td>6.33</td>
<td>2.63 ± 0.33</td>
<td>19.5</td>
<td>5.77</td>
</tr>
</tbody>
</table>

$^a$ The first $R_G$ value is from Guinier analyses and the second one is from the $P(r)$ analyses.
The data represents the mean values from all data sets.
<table>
<thead>
<tr>
<th>Search</th>
<th>Filter</th>
<th>Models&lt;sup&gt;a&lt;/sup&gt;</th>
<th>$R_G$ (nm)&lt;sup&gt;b&lt;/sup&gt;</th>
<th>$R_G$ (nm)&lt;sup&gt;b&lt;/sup&gt;</th>
<th>$R_{XS}$ (nm)</th>
<th>$D_{max}$ (nm)</th>
<th>$R$-factor (%)</th>
<th>$s_{20,w}$ (S)</th>
</tr>
</thead>
<tbody>
<tr>
<td>A, Search 1 (All linkers 2-8)</td>
<td>None</td>
<td>86,732</td>
<td>4.42 – 8.29</td>
<td>n.a.</td>
<td>0.01 – 4.23</td>
<td>n.a.</td>
<td>4.46 – 27.14</td>
<td>n.a.</td>
</tr>
<tr>
<td>Tris-150 0.17 mg/ml</td>
<td>$R$-factor</td>
<td>100</td>
<td>5.24 – 6.02</td>
<td>n.a.</td>
<td>2.00 – 2.76</td>
<td>n.a.</td>
<td>4.46 – 5.80</td>
<td>n.a.</td>
</tr>
<tr>
<td></td>
<td>$R$-factor, $R_G$, $R_{XS}$</td>
<td>28</td>
<td>5.24 – 5.54</td>
<td>n.a.</td>
<td>2.25 – 2.48</td>
<td>n.a.</td>
<td>4.46 – 5.39</td>
<td>n.a.</td>
</tr>
<tr>
<td></td>
<td>Best fit</td>
<td>1</td>
<td>5.45</td>
<td>5.67</td>
<td>2.25</td>
<td>20.0</td>
<td>4.46</td>
<td>5.2</td>
</tr>
<tr>
<td></td>
<td>Experimental</td>
<td>n.a.</td>
<td>5.44</td>
<td>5.72</td>
<td>2.39</td>
<td>20.0</td>
<td>n.a.</td>
<td>n.a.</td>
</tr>
<tr>
<td>Tris-150 0.5 mg/ml</td>
<td>$R$-factor</td>
<td>100</td>
<td>5.24 – 5.84</td>
<td>n.a.</td>
<td>2.33 – 2.65</td>
<td>n.a.</td>
<td>4.21 – 5.30</td>
<td>n.a.</td>
</tr>
<tr>
<td></td>
<td>$R$-factor, $R_G$, $R_{XS}$</td>
<td>131</td>
<td>5.28 – 5.79</td>
<td>n.a.</td>
<td>2.33 – 2.57</td>
<td>n.a.</td>
<td>4.21 – 5.30</td>
<td>n.a.</td>
</tr>
<tr>
<td></td>
<td>Best fit</td>
<td>1</td>
<td>5.31</td>
<td>5.63</td>
<td>2.44</td>
<td>20.0</td>
<td>4.21 – 5.98</td>
<td>5.2</td>
</tr>
<tr>
<td></td>
<td>Experimental</td>
<td>n.a.</td>
<td>5.52</td>
<td>5.85</td>
<td>2.45</td>
<td>21.0</td>
<td>n.a.</td>
<td>n.a.</td>
</tr>
<tr>
<td>B, Search 2 (Linkers 2,4,6,7)</td>
<td>None</td>
<td>72,755</td>
<td>4.52-8.66</td>
<td>n.a.</td>
<td>0.01-4.66</td>
<td>n.a.</td>
<td>4.68-30.89</td>
<td>n.a.</td>
</tr>
<tr>
<td>Tris-150 0.17 mg/ml</td>
<td>$R$-factor</td>
<td>100</td>
<td>5.01 – 5.88</td>
<td>n.a.</td>
<td>2.19 – 2.76</td>
<td>n.a.</td>
<td>4.68 – 6.01</td>
<td>n.a.</td>
</tr>
<tr>
<td></td>
<td>$R$-factor, $R_G$, $R_{XS}$</td>
<td>55</td>
<td>5.18 – 5.70</td>
<td>n.a.</td>
<td>2.27 – 2.57</td>
<td>n.a.</td>
<td>4.68 – 5.84</td>
<td>n.a.</td>
</tr>
<tr>
<td></td>
<td>Best fit</td>
<td>1</td>
<td>5.45</td>
<td>5.89</td>
<td>2.37</td>
<td>20.0</td>
<td>4.68</td>
<td>5.2</td>
</tr>
<tr>
<td></td>
<td>Experimental</td>
<td>n.a.</td>
<td>5.44</td>
<td>5.72</td>
<td>2.39</td>
<td>20.0</td>
<td>n.a.</td>
<td>n.a.</td>
</tr>
<tr>
<td>Tris-150 0.5 mg/ml</td>
<td>$R$-factor</td>
<td>100</td>
<td>5.01 – 6.02</td>
<td>n.a.</td>
<td>2.19 – 2.82</td>
<td>n.a.</td>
<td>3.86 – 6.15</td>
<td>n.a.</td>
</tr>
<tr>
<td></td>
<td>$R$-factor, $R_G$, $R_{XS}$</td>
<td>52</td>
<td>5.25 – 5.74</td>
<td>n.a.</td>
<td>2.57 – 2.34</td>
<td>n.a.</td>
<td>3.86 – 5.99</td>
<td>n.a.</td>
</tr>
<tr>
<td></td>
<td>Best fit</td>
<td>1</td>
<td>5.25</td>
<td>5.38</td>
<td>2.48</td>
<td>20.0</td>
<td>3.86</td>
<td>5.5</td>
</tr>
<tr>
<td></td>
<td>Experimental</td>
<td>n.a.</td>
<td>5.52</td>
<td>5.85</td>
<td>2.45</td>
<td>21.0</td>
<td>n.a.</td>
<td>n.a.</td>
</tr>
<tr>
<td>C, Search 3 (Linkers 3 and 6)</td>
<td>None</td>
<td>123,776</td>
<td>4.54 – 7.89</td>
<td>n.a.</td>
<td>0.004 – 4.57</td>
<td>n.a.</td>
<td>4.28 – 28.14</td>
<td>n.a.</td>
</tr>
<tr>
<td>Tris-150 0.17 mg/ml</td>
<td>$R$-factor</td>
<td>100</td>
<td>5.16 – 5.69</td>
<td>n.a.</td>
<td>2.14 – 2.53</td>
<td>n.a.</td>
<td>4.28 – 4.61</td>
<td>n.a.</td>
</tr>
<tr>
<td></td>
<td>$R$-factor, $R_G$, $R_{XS}$</td>
<td>694</td>
<td>5.17 – 5.69</td>
<td>n.a.</td>
<td>2.27 – 2.51</td>
<td>n.a.</td>
<td>4.28 – 6.00</td>
<td>n.a.</td>
</tr>
<tr>
<td></td>
<td>Best fit</td>
<td>1</td>
<td>5.36</td>
<td>5.58</td>
<td>2.37</td>
<td>20.0</td>
<td>4.28</td>
<td>5.4</td>
</tr>
<tr>
<td></td>
<td>Experimental</td>
<td>n.a.</td>
<td>5.44</td>
<td>5.72</td>
<td>2.39</td>
<td>20.0</td>
<td>n.a.</td>
<td>n.a.</td>
</tr>
<tr>
<td>Tris-150 0.5 mg/ml</td>
<td>$R$-factor</td>
<td>100</td>
<td>5.13 – 5.66</td>
<td>n.a.</td>
<td>2.22 – 2.69</td>
<td>n.a.</td>
<td>3.84 – 4.65</td>
<td>n.a.</td>
</tr>
<tr>
<td></td>
<td>$R$-factor, $R_G$, $R_{XS}$</td>
<td>749</td>
<td>5.26 – 5.79</td>
<td>n.a.</td>
<td>2.33 – 2.57</td>
<td>n.a.</td>
<td>3.84 – 5.99</td>
<td>n.a.</td>
</tr>
<tr>
<td></td>
<td>Best fit</td>
<td>1</td>
<td>5.34</td>
<td>5.64</td>
<td>2.52</td>
<td>20.0</td>
<td>3.84</td>
<td>5.1</td>
</tr>
<tr>
<td></td>
<td>Experimental</td>
<td>n.a.</td>
<td>5.52</td>
<td>5.85</td>
<td>2.45</td>
<td>21.0</td>
<td>n.a.</td>
<td>n.a.</td>
</tr>
</tbody>
</table>

<sup>a</sup> Total number of models accepted after Monte Carlo simulations and after model filtering. The best fit model corresponds to that with the lowest $R$-factor in the filtered models.

<sup>b</sup> The first $R_G$ value of the pair is from Guinier analyses and the second is from the $P(r)$ analyses. n.a. indicates not available.
FIGURE 1. The human FHR5 SCR-1/9 domain structure.

A, A schematic representation of the domain structure of the human FHR5 dimer and its ligand binding sites. The two putative N-linked glycosylation sites (not modelled) on SCR-2 and SCR-7 are depicted as Y-shaped symbols. Domains SCR-1/2 (green) form a head-to-tail dimer. Domains SCR-3/7 (yellow) have high sequence identities to human Factor H SCR-10/14. The C3b/C3d and heparin binding sites are indicated as arrows on SCR-8 and SCR-9 (purple).

B, Summary of the templates used for the homology modelling of FHR5. Each FHR5 domain is aligned with the corresponding crystal/NMR structure denoted by its PDB code.

C, The sequences of the nine FHR5 SCR domains are aligned with the template SCR sequences identified by their PDB codes. The conserved cysteine and tryptophan residues are highlighted in pink and yellow respectively, and the linkers in cyan. The residue numbering corresponds to the full FHR5 protein sequence (SWISSPROT accession code: Q9BXR6) including the signal peptide, with +1 corresponding to the Methionine start codon.
FIGURE 2. **Purification of FHR5.** *A*, Gel filtration profile showing the removal of aggregates from the FHR5 sample. The sample was loaded onto a Superdex™ 200 gel filtration column. Fractions containing homogenous FHR5 from the large peak as indicated were pooled. Mass determinations are shown in Figure 3. *B*, Non-reducing (NR) and reducing (R) SDS-PAGE analyses of FHR5 (~65 kDa monomeric). The molecular masses of the Mark 12™ protein standard (Invitrogen) in kDa are labelled.
FIGURE 3. **SEC-MALLS analysis of FHR5.** The elution profile (chromatogram) for FHR5 was analysed using UV detection (blue), MALLS (light scattering) detector (red), and refractive index detector (green). Three successive prominent peaks (1-3 as indicated by the pairs of numbers beneath were examined for their molecular mass. The calculated molecular masses were >5,400 kDa, 162 kDa, and 27 kDa for peaks 1, 2 and 3 respectively. The refractive index peak above 8.0 min is attributed to the end of the gel-filtration step.
FIGURE 4. Size-distribution $c(s)$ sedimentation velocity analysis of FHR5.

Left: the experimentally observed sedimentation boundaries (black) by interference optics are shown for 0.16 mg/ml FHR5 SCR-1/9 based on the following PBS buffers: A, 20 mM NaCl; B, 50 mM NaCl; C, 90 mM NaCl; D, 137 mM NaCl; and E, 250 mM NaCl. In each case, 40-50 boundaries (coloured lines) were fitted as shown.

Right: the data fits corresponded to rotor speeds of 40,000 rpm (solid lines) and 50,000 rpm (dashed lines). The resulting size-distribution analyses $c(s)$ revealed a major peak that shifted from 6.5 S in 20 mM NaCl to 5.9 S in 250 mM NaCl. The vertical dashed lines illustrate the shift in $s_{20}$ values as the salt concentrations increased.
FIGURE 5. Relationship between sedimentation coefficient and the buffer salt concentration. The mean ± s.d of the sedimentation coefficient ($s_{20,w}$) data are shown as a function of the NaCl concentration in each buffer. For PBS-350, only a single data point was available.
FIGURE 6. X-ray Guinier $R_G$ and $R_{XS}$ analyses for FHR5. In the Guinier analyses, plots of $\ln I(Q)$ against $Q^2$ for the $R_G$ analyses (left), and $\ln I(Q).Q$ against $Q^2$ for the $R_{XS}$ analyses (right) are shown. The straight lines indicate the slopes of each fit. The filled circles (●) correspond to the $Q.R_G$ and $Q.R_{XS}$ ranges used for each fit, with the Q-range used for the $R_G$ values being 0.10 - 0.27 nm$^{-1}$, and that for the $R_{XS}$ values being 0.32 – 0.55 nm$^{-1}$ (arrowed). In rows A, B and C, the fits correspond to FHR5 concentrations of 0.17 mg/ml, 0.13 mg/ml, 0.09 mg/ml, and 0.04 mg/ml from top to bottom in each panel
A, The FHR5 fits correspond to PBS-50.
B, The FHR5 fits correspond to PBS-137.
C, The FHR5 fits correspond to Tris-150.
D, The FHR5 fits correspond to 0.5 mg/ml, 0.4 mg/ml, 0.3 mg/ml, and 0.2 mg/ml from top to bottom in Tris-150.
FIGURE 7. Concentration dependence of the X-ray Guinier $R_G$ and $R_{XS}$ values. Each value was measured in triplicate and shown as the mean ± s.d at low concentrations, except for the data sets above 0.2 mg/ml which were measured once. A, The $R_G$ and $B$, $R_{XS}$ values are shown for PBS-50 (blue, ○), PBS-137 (purple, □) and Tris-150 (black, △) buffers. The dotted line in both panels represent the mean across all values.
FIGURE 8. X-ray distance distribution $P(r)$ analyses for FHR5. The $P(r)$ curves for FHR5 in panels (A-D) correspond to those shown in Figure 6. In each panel, the $P(r)$ curves were normalised for concentration and coloured according to the FHR5 concentration from light blue at the lowest concentration to dark blue at the highest concentration. The maximum $M$ depicts the most commonly occurring distances within the FHR5 structure. The length of FHR5 is signified by $L$ at the $r$ value where $P(r)$ reaches zero.
FIGURE 9. **Density plot of the conformationally-varied FHR5 structures.** The linear dimeric FHR5 starting structure is shown at the centre in blue and red for the two monomers, with the SCR-1/2 dimer at the centre. The grid shows the complete spatial extent covered by the 72,755 modelled conformations of Search 2 for each FHR5 monomer (Table 3), shown in blue and red.
FIGURE 10. *R*-factor analyses for the atomistic modelling of FHR5 dimers. The goodness-of-fit *R*-factors are compared with the calculated \( R_G \) values for the 86,732, 72,755 and 123,776 conformationally-randomised FHR5 dimer models from three different Monte Carlo searches. The yellow circles denote the models in each search. The experimental \( R_G \) values are shown by the solid vertical lines with \( R_G \) error ranges of \( \pm 5\% \) represented by the dashed lines. The 28-749 best fit models for each search after filtering for the *R*-factor, \( R_G \) and \( R_{XS} \) values are shown in green. The single best-fit model with the lowest *R*-factor at the minima (red arrow) is in red.

A, The models were compared with data for 0.17 mg/ml and 0.5 mg/ml FHR5 in Tris-150, varying linkers 2-8 (Search 1).

B, The models were compared with the same two data sets, but varying linkers 2, 4, 6 and 7 (Search 2).

C, The models were compared with the same two data sets, but varying linkers 3 and 6 (Search 3).
FIGURE 11. **X-ray scattering curve fits for best-fit FHR5 models.** The best fits correspond to the models with the lowest $R$-factor (red in Figure 9). The experimental data at 0.17 mg/ml (upper panels) and 0.50 mg/ml (lower panels) are indicated by circles (O), overlaid with the modelled scattering curve (blue line). The insets show the experimental (black) and modelled (blue) $P(r)$ curves. Panels A, B and C show the best fit from Search 1, 2 and 3 respectively (Table 3).
FIGURE 12. Normalised dimensionless Kratky plots of the best fit curves of FHR5 models. The experimental data for 0.5 mg/ml FHR5 in Tris-150 are denoted by ◆. The modelled curves for 0.5 mg/ml in Tris-150 according to Search 1, Search 2 and Search 3 are shown in red, cyan and purple respectively. The vertical solid line denotes the maximum of the Kratky plot.
FIGURE 13. **Principal component analyses of the FHR5 best-fit models from Search 2.**
The 55 best-fit models (Table 3) were grouped by principal component analysis into five groups, out of which three groups were predominant in terms of the number of 48 models they contained. These are shown as PC1, PC2 and PC3 (black, 19 models; red, 16 models; and green, 13 models) and exemplified by the first three principal components (PC2 vs. PC1 and PC3 vs. PC2).
FIGURE 14. The three best-fit FHR5 models, their interaction with C3d, and mutant FHR5. A–C, Ribbon views are shown of single FHR5 models, these being representative of the three principal component analysis (PCA) groups determined from Search 2 (Table 3). The models correspond to the X-ray curve fits in Figure 11B. The two monomers are shown in red and blue, with the N-terminal SCR-1/2 dimer pair denoted by N and shown in cyan and orange. The C-terminal SCR-9 domains are denoted by C. To their right are shown surface views of two C3d molecules bound to the two SCR-9 domains in the FHR5 dimer shown in the same orientation. If binding is sterically allowed, the C3d surface is shown in green; if this is sterically blocked, this is greyed out. D, The two best fit structures (to the same scale) of the X-ray scattering models of glycosylated Factor H are shown. Purple denotes the eight N-glycan chains in Factor H. E, The mutant FHR5 dimer structure was generated from the best-fit structure from Search 2 (Table 3) by the addition of two extra SCR domain pairs to represent the mutant. The native SCR-1/2 pair is shown in black and blue at the centre, with the extra SCR-1/2 pair shown in red at the two N-termini of the native SCR-1/2 pair. F, The putative daisy-chaining of mutant FHR5 dimers to form a tetramer is shown. Pairs of mutant SCR-1/2 domains are formed based on the crystal structure of this dimer. Such a tetramer can be extended to form hexamers and larger oligomers, with chain extension limited by binding of a wild type FHR5 molecule.
The solution structure of the complement deregulator FHR5 reveals a compact dimer and provides new insights into CFHR5 nephropathy
Nilufar Kadkhodayi-Kholghi, Jayesh S. Bhatt, Jayesh Gor, Lindsay C. McDermott, Daniel P Gale and Stephen J. Perkins

*J. Biol. Chem.* published online September 14, 2020

Access the most updated version of this article at doi: [10.1074/jbc.RA120.015132](https://doi.org/10.1074/jbc.RA120.015132)

**Alerts:**
- When this article is cited
- When a correction for this article is posted

[Click here](http://www.jbc.org/) to choose from all of JBC's e-mail alerts