

## DATABASES

# An Interactive Web Database of Factor H-Associated Hemolytic Uremic Syndrome Mutations: Insights Into the Structural Consequences of Disease-Associated Mutations

Rebecca E. Saunders,<sup>1</sup> Timothy H.J. Goodship,<sup>2</sup> Peter F. Zipfel,<sup>3</sup> and Stephen J. Perkins<sup>1\*</sup>

<sup>1</sup>Department of Biochemistry and Molecular Biology, Royal Free and University College Medical School, Darwin Building, University College London, London, United Kingdom; <sup>2</sup>The Institute of Human Genetics, University of Newcastle upon Tyne, Newcastle upon Tyne, United Kingdom; <sup>3</sup>Department of Infection Biology, Hans Knöll Institute for Natural Products Research, Jena, Germany

Communicated by Mirelle Claustres

Factor H (FH) is a central complement regulator comprised of 20 short complement repeat (SCR) domains. Nucleotide changes within this gene (*CFH*) have been observed in patients with hemolytic uremic syndrome (HUS), and also membranoproliferative glomerulonephritis and age-related macular degeneration. All parts of FH are affected, but many mutations are clustered in the C-terminal part of FH. Up to now, structural analyses of HUS have been based on SCR-20, a domain that is involved in FH interactions with C3b, heparin, and endothelial cells. In order to identify the structural and functional consequence of HUS mutations, further disease-associated mutations were analyzed in terms of homology and nuclear magnetic resonance (NMR) models for factor H SCR domains. An interactive web database of 54 human HUS-associated mutations and others was created from the literature ([www.FH-HUS.org](http://www.FH-HUS.org)). This has comprehensive search and analysis tools, integrating phenotypic and genetic data with structural analysis. Each mutation can be highlighted on the SCR structure together with the patient FH and C3 levels where available. Two new insights were obtained from our collection of data. First, phenotypic data on FH clarify our previously-proposed classification of Type I and Type II disorders that both lead to HUS, where Type I affects FH secretion and folding, and Type II leads to expressed protein in plasma that is functionally defective. Second, the new mutations show more clearly that SCR domains from SCR-16 to SCR-19 are important for the ligand binding activities of FH as well as SCR-20. This FH web database will facilitate the interpretation of new mutations and polymorphisms when these are identified in patients, and it will clarify the functional role of FH. *Hum Mutat* 27(1), 21–30, 2006. © 2005 Wiley-Liss, Inc.

KEY WORDS: factor H; complement; HUS; *CFH*; mutation database; immunodeficiency diseases; membranoproliferative glomerulonephritis; age-related macular degeneration

## INTRODUCTION

In the immune defense system, factor H (FH) is a central complement regulator in the alternative pathway of complement activation by acting as a cofactor for factor I in the breakdown of C3b to form iC3b [Law and Reid, 1995]. It also accelerates the decay of the C3 convertase C3bBb, and competes with factor B for binding to C3b. FH consists entirely of 20 short complement repeat (SCR) domains, also known as complement control protein domains, each of length about 61 residues. The cofactor and decay accelerating activity are located within the four N-terminal domains, SCR-1 to SCR-4, which bind to intact C3b, a second C3 site is located within SCR-6 to SCR-10, which binds to the C3c fragment within C3b, and a third site is located within SCR-16 to SCR-20, which binds to the C3d fragment within C3b [Sharma and Pangburn, 1996; Zipfel et al., 1999; Jokiranta et al., 2000; Pangburn, 2000] (Fig. 1). Heparin modulates the complement regulatory functions of FH, where two heparin-binding sites have been located in SCR-7 and SCR-20 [Blackmore et al., 1996,

1998] and a third heparin-binding site originally thought to be at SCR-13 has now been localised to SCR-9 [Pangburn et al., 1991; Ormsby et al., 2004]. The physiologically more relevant sialic acid binding regions of FH have also been localized to SCR-6/10,

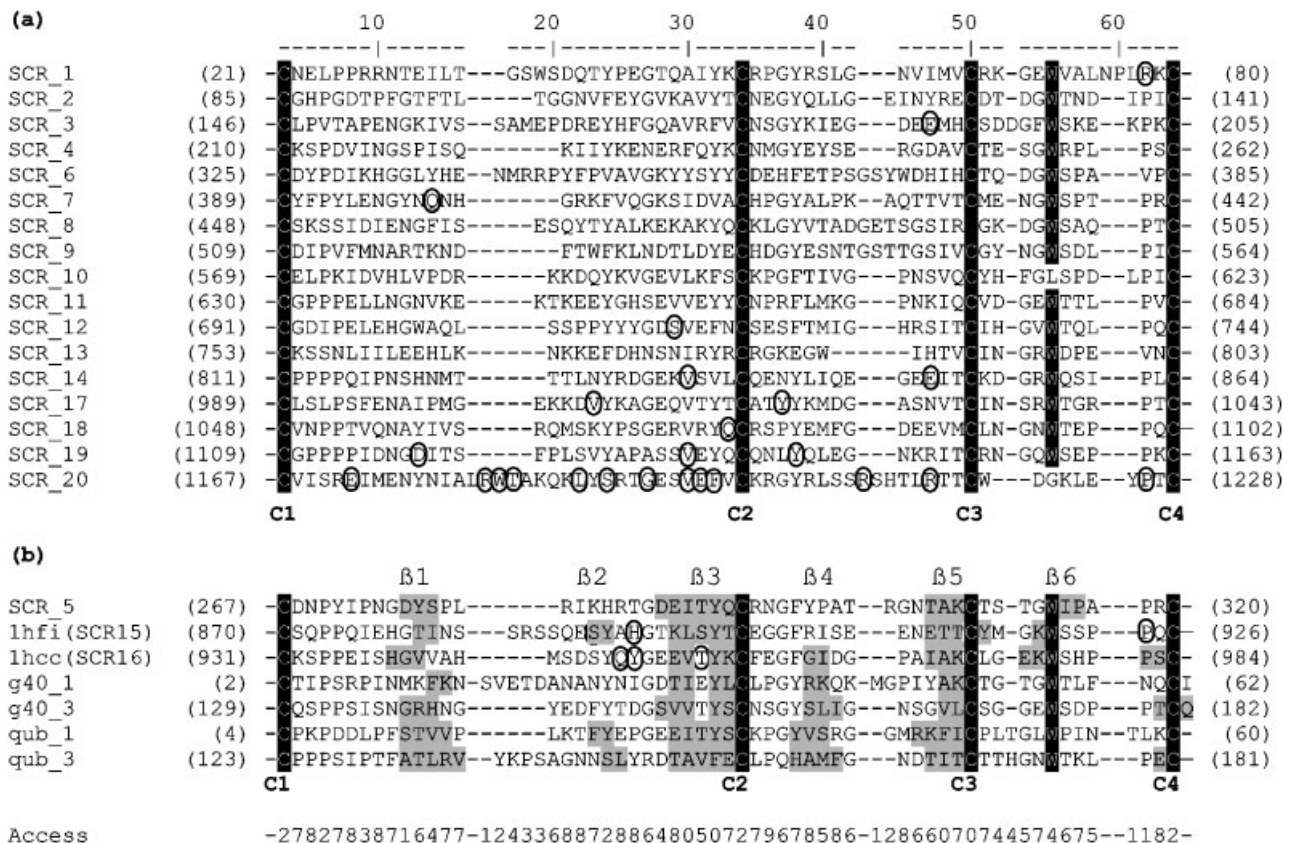
The Supplementary Material referred to in this article can be accessed at [www.interscience.wiley.com/jpages/1059-7794/suppmat](http://www.interscience.wiley.com/jpages/1059-7794/suppmat). Received 12 May 2005; accepted revised manuscript 24 August 2005.

\*Correspondence to: S.J. Perkins, Department of Biochemistry and Molecular Biology, Darwin Building, University College London, Gower Street, London WC1E 6BT, United Kingdom. E-mail: [s.perkins@medsch.ucl.ac.uk](mailto:s.perkins@medsch.ucl.ac.uk)

Grant sponsor: Medical Research Council; Wellcome Trust; National Kidney Research Fund; Robin Davies Trust.

DOI 10.1002/humu.20268

Published online 9 November 2005 in Wiley InterScience ([www.interscience.wiley.com](http://www.interscience.wiley.com)).



structural analyses if required. We believe that the use of such an interactive and contextual database alongside structural analysis will enable the significance of new mutations to be more easily understood, leading to a molecular interpretation of FH function and its mutants.

### FH-HUS MUTATION DATABASE

The FH-HUS mutation database is available at [www.FH-HUS.org](http://www.FH-HUS.org). It is maintained and updated on the server of the Department of Biochemistry and Molecular Biology at University College London. The database was constructed using MySQL ([www.mysql.com](http://www.mysql.com)) and is displayed and navigated on the web via PHP ([www.php.com](http://www.php.com)), Javascript, and HTML programming.

#### Database Functions

The website is fully interactive with a user-friendly graphical interface. A choice of three search tools is available (quick, basic, and advanced) that can be adapted to perform specific requirements such as retrieving mutations over a range of SCR domains or for a specific phenotype. The results of all searches can be downloaded in Microsoft Excel ([www.microsoft.com](http://www.microsoft.com)) format for further analysis or retrieved in a printer friendly format. The following fields of information are available for each mutation record: nucleotide alteration, codon number, initial residue, mutated residue, phenotype (classified as Type I or Type II), reference, mutation type (e.g., insertion or deletion), SCR domain, secondary structural position (taken from consensus Database of Secondary Structure in Proteins [DSSP] assignments), sequence (to highlight the mutation within a short sequence of FH), comments, and reference state (whether it is an abstract or a full citation). Following the Human Genome Variation Society (HGVS) guidelines for mutation nomenclature ([www.hgvs.org/mutnomen/](http://www.hgvs.org/mutnomen/)), the details of the specific nucleotide alteration are available for each mutation. The reference nucleotide sequence is taken from RefSeq #: NM\_000186.1 and nucleotides are numbered starting from the A of the ATG translation initiation codon. The amino acid substitution and codon number, for example, Trp1183Arg, where residue numbering includes the 18-residue signal peptide and codon 1 refers to the initiating methionine residue, is also used to describe mutations. The residue number with reference to the mature FH protein (following PDB numbering: PDB code 1haq) is also given. The following fields of information are available for each patient record: FH and C3 levels (given as low, normal, or high depending on control values used in the original reference), genotype, clinical history, inheritance (sporadic/familial), and a comments field for any extra information. Data are available describing the FH and C3 assays used in each study and can be viewed in the context of the results to allow comparison between different data sets. Statistics for the mutations within the database are displayed graphically, such as the distribution of the mutations over different SCR domains or different phenotypes. The website also provides information pages, associated links, and a reference library for FH and HUS with each reference linked to the abstract via the PubMed database ([www.ncbi.nlm.nih.gov](http://www.ncbi.nlm.nih.gov)).

New mutations are entered into the database via an electronic submission form. In order to maintain the database entries in a reliable and structured format, the database curator checks all new mutations. In particular, the database will not only accept mutations that have been published in peer-reviewed articles, but will also accept and flag mutations and patients that have not been published in peer-reviewed articles, or those that have only

been published in abstract form. The reference field of these unpublished mutations and patients will contain the name of the depositor responsible for the submission. This will permit the development of a larger, more up-to-date data set, while also enabling the source of each mutation to be identified. Users can choose to omit unpublished mutations from their searches via the advanced search form. A mailing list function is available to alert users when a new mutation is submitted. The data was first published on our website in November 2004. The structure and content of the FH-HUS database corresponds to the recommendations of the HGVS for locus-specific databases [Claustres et al., 2002].

#### Advanced Database Functions

As well as providing published information on *CFH* mutations and patients, the database also provides analysis of mutations in the context of structural models for each of the 20 FH SCR domains. Experimentally solved nuclear magnetic resonance (NMR) structures were available for SCR-5, SCR-15, and SCR-16 [Barlow et al., 1992, 1993] of human FH. The remaining 17 FH SCR domains were constructed using MODELLER v7 (<http://salilab.org/modeller/FAQ.html>) alongside the structure verification program PROCHECK [Laskowski et al., 1993]. PSI-BLAST [Altschul and Koonin, 1998] searches of the Protein Data Bank (PDB) [Berman et al., 2000] identified 27 SCR structural templates in 14 structures. Their PDB codes are listed as follows: 1hcc, 1hfi, 1g40, 1ojv, 1h03, 1gpz, 1elv, 1gkg, 1gkn, 1ly2, 1ckl, 1q3x, 1qub, and 1srz. From these 27 possible templates, the closest template for each SCR sequence in terms of sequence identity and minimum insertions and deletions was identified using CLUSTAL alignments [Higgins et al., 1994]. This is summarized in Table 1 and Figure 1. Although the sequence identity between the SCR domain structures is quite low (10–51% between themselves), they generally possess a six-stranded antiparallel  $\beta$ -sheet SCR structure ( $\beta 1$  to  $\beta 6$ ) that is subdivided into three pairs of  $\beta$ -strands, with  $\beta 1$ - $\beta 3$ - $\beta 5$  forming a central  $\beta$ -sheet and stabilized by two disulfide bridges, C1-C3 and C2-C4, at the two ends of the domain [Perkins et al., 2002].

The SCR domain structures are used to highlight the positions of point mutations with interactive views provided by the newly developed Java Applet JMol version 10 (<http://jmol.sourceforge.net/>). An in-depth analysis of each mutation provides the user with structural information such as the DSSP secondary structure assignment and side chain surface accessibility [Kabsch and Sander, 1983] and highlights known binding regions of FH that the mutation may perturb. A comparison of amino acid properties such as hydrophobicity and charge is also shown. This approach can use information contained within the database to predict a similar analysis of any new user-defined substitution at any position in FH. This flexible, yet powerful tool can be used to predict whether a novel point mutation may perturb structural stability (Type I) or biological function (Type II), or alternatively may be a non-disease-causing polymorphism. The homology models of the FH SCR domains constructed during this study are also available for download in PDB format ([www.FH-HUS.org/models.php](http://www.FH-HUS.org/models.php)), and links are provided for experimentally solved FH SCR structures that are deposited in the PDB.

#### CFH Mutations

A total of 48 mutations causing HUS were identified in *CFH* by searching the Human Gene Mutation database (<http://archive.uwcm.ac.uk/uwcm/mg/hgmd0.html>) and the literature database

TABLE 1. Summary of the Homology Modeling of 17 SCR Domains of FH\*

| SCR | Template PDB <sup>a</sup> | Protein                        | Insertions | Deletions | Sequence Identity |
|-----|---------------------------|--------------------------------|------------|-----------|-------------------|
| 1   | 1hfi                      | Human FH                       | 3          | None      | 25.0%             |
| 2   | 1qub_1                    | Human $\beta$ 2-glycoprotein I | 1          | None      | 27.6%             |
| 3   | 1qub_3                    | Human $\beta$ 2-glycoprotein I | 1          | None      | 18.3%             |
| 4   | 1hcc                      | Human FH                       | None       | 1         | 31.5%             |
| 6   | 1g40-1                    | VCCP                           | None       | 1         | 16.1%             |
| 7   | SCR5                      | Human FH                       | None       | None      | 33.3%             |
| 8   | 1hfi                      | Human FH                       | 3          | 2         | 26.7%             |
| 9   | SCR5                      | Human FH                       | 2          | None      | 23.2%             |
| 10  | 1hcc                      | Human FH                       | 1          | None      | 30.9%             |
| 11  | 1hcc                      | Human FH                       | 1          | None      | 32.7%             |
| 12  | 1hcc                      | Human FH                       | None       | None      | 33.3%             |
| 13  | 1hcc                      | Human FH                       | None       | 3         | 18.5%             |
| 14  | 1g40-3                    | VCCP                           | None       | None      | 29.6%             |
| 17  | 1g40-3                    | VCCP                           | 1          | None      | 40.0%             |
| 18  | 1g40-3                    | VCCP                           | 1          | None      | 34.5%             |
| 19  | 1g40_3                    | VCCP                           | 1          | None      | 40.0%             |
| 20  | 1g40-1                    | VCCP                           | 1, 1       | 1         | 17.5%             |

\*The template PDB files used for homology modeling of the individual FH SCR domains are described in this Table. The numbers in the insertions and deletions columns indicate how many residues had to be inserted or deleted from the template structure in order to model the FH SCR domain. Two values separated by a comma indicate there are insertions or deletions at more than one place in the model. The sequence identity is a pairwise comparison of the sequence of the FH SCR domain and the sequence of the SCR domain from the template protein.

<sup>a</sup>The PDB codes are as follows: 1hfi, 1hcc [Barlow et al., 1993]; 1qub [Bouma et al., 1999]; SCR5 [Barlow et al., 1992]; and 1g40 [Murthy et al., 2001]. VCCP, vaccinia virus complement control protein.

PubMed using the following search expressions: FH mutation, factor H mutation, *CFH* mutation, HUS mutation. The World Wide Web was also searched using the search engine Google (www.google.com) (Supplementary Table S1 [available online at www.interscience.wiley.com/jpages/1059-7794/suppmat] and Fig. 2). Three more novel unpublished mutations (Supplementary Table S1) were identified by *CFH* mutation screening of UK patients at the Institute of Human Genetics, Newcastle upon Tyne, by DHPLC (wave) and fluorescent sequencing. The phenotype was based on the clinical information provided by the referring physician. Three more novel and two previously known mutations were identified in Jena from German patients using standard assays as previously described [Neumann et al., 2003].

Each mutation and patient record has been manually entered with the aim of registering every individual described with a *CFH* mutation and not just one member of each family. At present the database records correspond to 58 patients, each possessing a unique ID number. Varying amounts of data are available for each mutation and patient, with only five mutations having no available patient information. Figure 3a highlights the variation on the amount of reported data for each of the 58 patients. This variation shows the inconsistency of the published mutation reports to date, and highlights areas where information is lacking, such as the clinical history of the patient.

### Mutation Classification

Our previous study [Perkins and Goodship, 2002] proposed the classification of the mutations into a Type I or Type II phenotype, and the presently increased total of known mutations now clarify this. Problems arise for FH-HUS mutations as there is a large variation of FH protein levels in plasma. Even with unaffected individuals, FH levels can range between 235 and 810 mg/l [Neumann et al., 2003] or 350 and 650 mg/l [Caprioli et al., 2001], and FH levels have been shown to be significantly lower in neonates and infants than in adults [Ault, 2000]. This makes it increasingly difficult to assign whether the FH level of a patient that is within the normal range accounts for one or two alleles. For example, the majority of HUS patients are heterozygous, having

one intact and one defective allele. If the mutated allele has a defective protein secretion, FH secretion is blocked and consequently the FH plasma level would be about 50% of the normal value. However, as the “normal” values cover such a large range, it is difficult to say whether this is Type I. It also should be noted that different groups use different methods for the FH assay and the serum FH values depend highly on the type of standard used. An improved classification scheme would denote Type I as either complete with no observed FH secretion (homozygous) or partial (heterozygous). Published phenotypic data for the FH and C3 levels were indicated in our study as high, normal, low, or very low according to the assignments in the literature. Details such as the method of assay and the range of values in the normal controls are stored in the database and can be displayed for comparison between patients from different studies. For future database submissions, details of the assays and ranges will be included to facilitate interpretation of FH levels in patients. At the present time, those mutations with FH levels that fall within the normal ranges are classified with an unknown phenotype unless the structural models provide strong enough evidence of a Type I or Type II phenotype. It is useful to display mutations within the SCR structural models as this can indicate areas of the SCR domain that are more susceptible to disruption than others. For example, a number of Type II mutations are found on loop regions, whereas Type I mutations are often found on buried  $\beta$ -strands.

The nine mutations in SCR-7, SCR-9, SCR-11 (two), SCR-15, SCR-16 (two), SCR-17, and SCR-19 within *CFH* that involve substitution of cysteine residues (Supplementary Table S1) provide a good example of this problem of classification. For small  $\beta$ -sheet Cys-rich domains such as the SCR of this study, and others such as the epidermal growth factor domain found in first component of complement and the thrombospondin repeat Type I domain found in properdin and the late complement components, a Cys substitution may result in the loss of the disulfide bridge. This would be predicted to cause incorrect folding, given that these small domains are largely stabilized by these Cys-Cys bridges [Creighton, 1993]. However, only six of these mutations show low FH levels in patient plasma, and, surprisingly, two mutations (Cys630Trp and Cys1043Arg) show levels falling within the

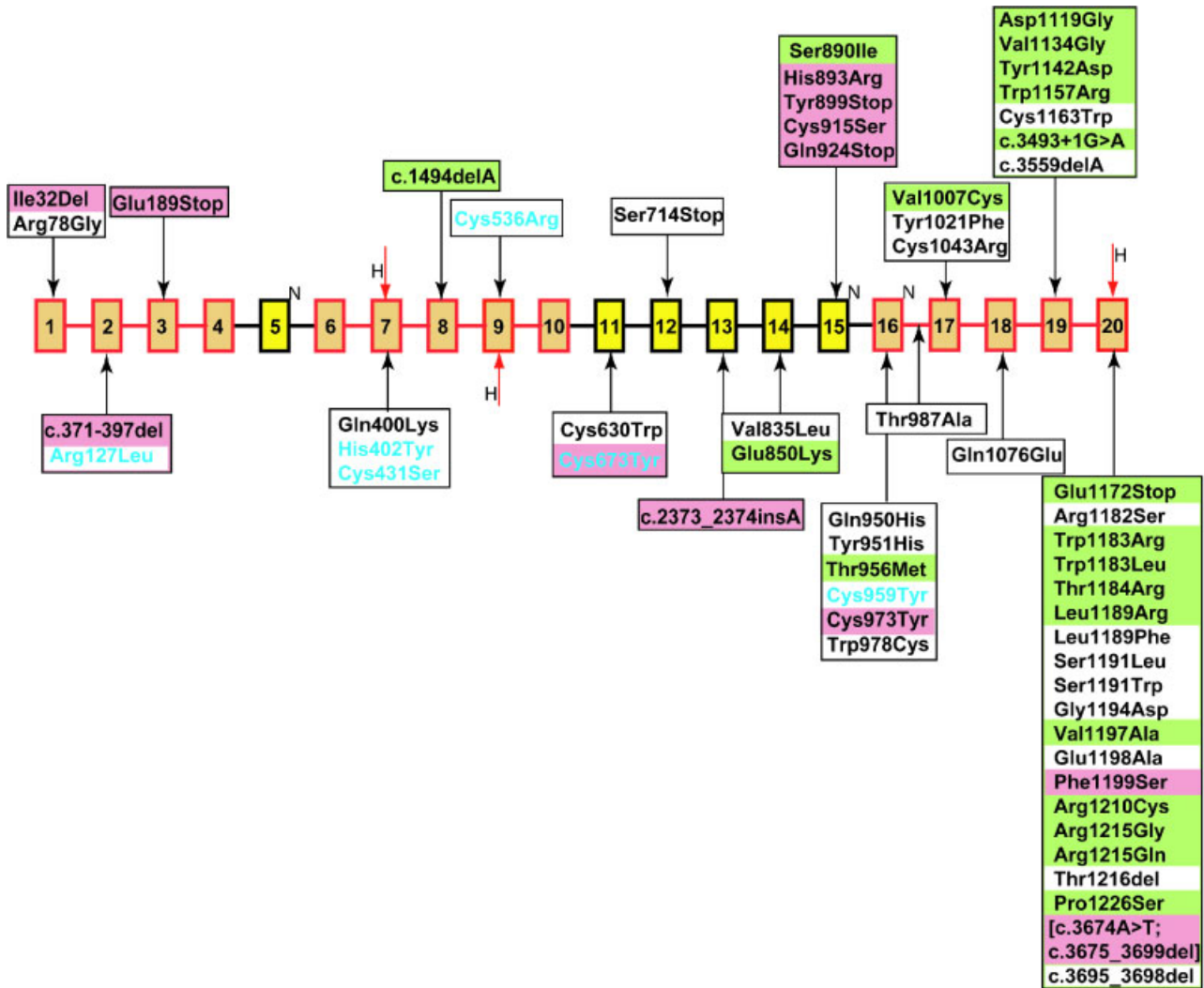


FIGURE 2. Positions of 59 mutations in *CFH*. Mutations associated with HUS within *CFH* are identified, above and below the corresponding SCR. Type I mutations are highlighted in red and Type II mutations are highlighted in green. Unclassified mutations are unshaded. The MPGN mutations and the AMD polymorphism are noted in blue text. The SCR domains involved in C3b binding are highlighted in orange with red borders, and SCR-7, SCR-9, and SCR-20 involved with heparin-binding are indicated with red arrows labeled with H. N denotes an NMR structure.

normal control range. At first glance this suggests that breakage of either the C1-C3 bond of SCR-11 (Cys630Trp) or the C2-C4 bond of SCR-17 (Cys1043Arg) does not necessarily lead to the misfolding and degradation of the FH protein. Unfortunately, it is difficult to establish whether the FH level in plasma is the consequence of only the wild-type allele or from both the wild-type and mutant alleles of a heterozygous patient. Another example can be seen with the analysis of five nonsense mutations within *CFH* (Supplementary Table S1). Three of these mutations (Glu189Stop, Tyr899Stop, and Gln924Stop) can be defined as Type I, where the mutant allele produces an unstable FH molecule and there is a lower amount of FH in patient plasma. However, one nonsense mutation (Ser714Stop) is seen in a patient with normal FH levels, and again the problem arises whether this is due to a truncated protein in the plasma, or the result of one wild-type allele. Manuelian et al. [2003] carried out functional studies on three mutations within FH, of which one was a nonsense mutation (Glu117Stop). Here the truncated protein was purified from patient plasma and shown to have reduced binding to heparin, C3b/C3d, and endothelial cells, leading to its assignment as Type II

[Manuelian et al., 2003]. In short, it is not straightforward to predict the effect of a given mutation on the FH phenotype using FH serum levels alone, and each mutation will require experimental characterization in order to fully appreciate its effect on FH structure and function.

### Distribution of Mutations in SCR-19 and SCR-20

The majority (64%) of the *CFH* nucleotide changes are point substitutions in the sequence that lead to missense and nonsense mutations, although there are a small number of insertions and deletions (11%), splice site mutations (1%), and non-disease causing polymorphisms (22%) (Fig. 3b). Of the 54 mutations, only 32 could be classified as having a Type I or Type II phenotype. The remaining 22 did not have sufficient data, or FH levels were in the normal range and it could not be identified whether the mutant protein was in patient plasma. The majority of classified mutations have a Type II phenotype (37%) with the remaining mutations classified as Type I (22%) (Fig. 3c). Although the mutations are distributed throughout the SCR domains, they cluster towards the

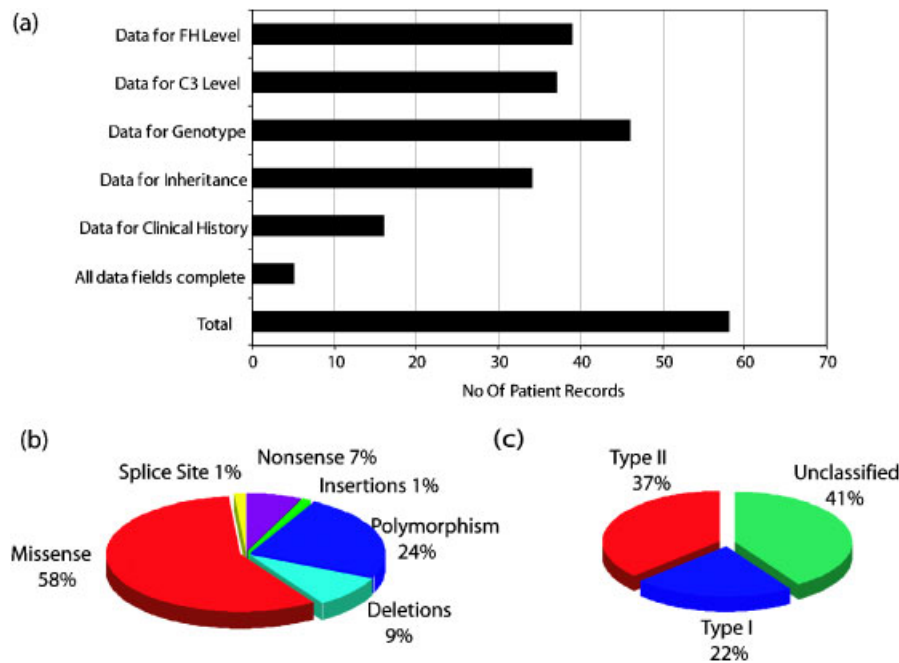


FIGURE 3. Statistical surveys of HUS patients. **a:** The coverage of 58 patient records with respect to different data fields. **b:** Pie chart of all 76 reported nucleotide changes showing their distribution between five types of change. **c:** Pie chart of all 54 reported HUS mutations to show the distribution of HUS phenotypes.

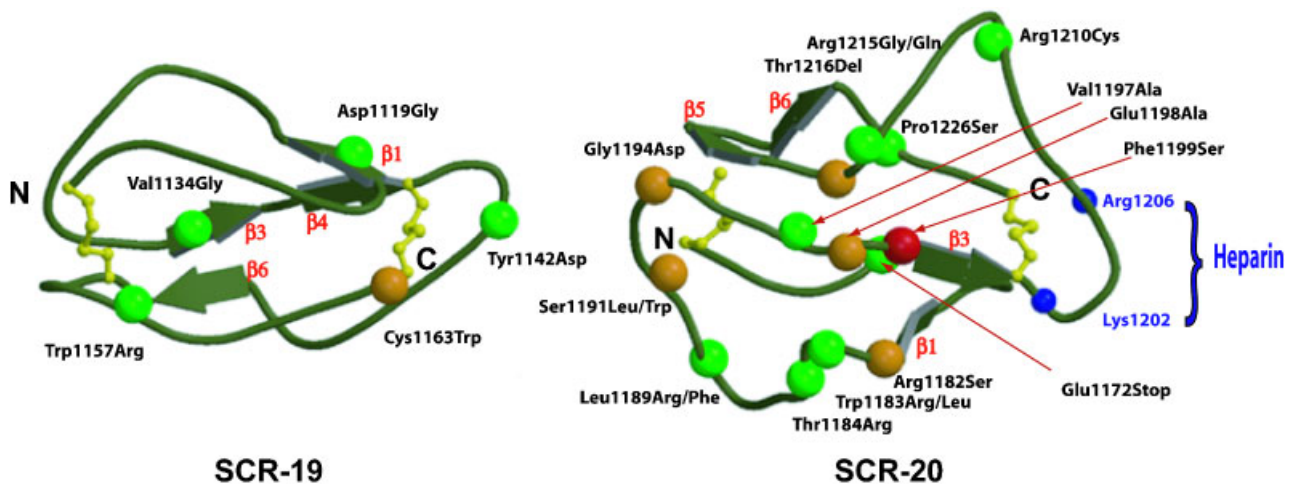


FIGURE 4. Molecular views of point mutation sites within FH SCR-19 and SCR-20 domains. Secondary structure cartoon of SCR-19 and SCR-20 in FH. The relative orientation of the two SCR domains about their longest axes is arbitrary. The four disulfide bridges are shown in yellow.  $\beta$ -strands where visible are labeled  $\beta 1$  to  $\beta 6$ . Mutations with phenotypic classification are shown in red for Type I and green for Type II. Those without classification are shown in orange. The predicted heparin-binding site at the C-terminus of SCR-20 [Ganesh et al., 2004] is highlighted by two blue spheres to denote Lys1202 and Arg1206. N and C denote the N-terminus and C-terminus, respectively, of each SCR domain. Figures were prepared with Molscript [Kraulis, 1991] and Raster 3d [Merritt and Murphy, 1994].

C-terminal domains, with nearly half (48%) of the total mutations occurring in SCR-19 and SCR-20, and a quarter occurring in SCR-15 to SCR-18. Where classified the C-terminal mutations predominantly show Type II phenotypes (72%), reflecting the functional significance of these domains. The remaining 12 mutations in SCR-1 to SCR-15 predominantly show a Type I phenotypes (50%), reflecting the structural significance of these domains.

In SCR-19/20, two of 27 mutations are Type I, 15 are Type II, and 10 have no known phenotypes (Fig. 2). Figure 4 shows that the Type I (red) and Type II (green) mutations are not localized to

any one specific region of SCR-19 and SCR-20. One of the two Type I mutations (Phe1199Ser) is on the buried  $\beta 3$  strand. A change in this area of the SCR domain may affect the secondary structure of SCR-20 and cause the incorrect folding of the SCR-20 domain effecting the secretion or degradation of the entire cellular FH protein. This is evident in the low serum levels of FH in the patient with this mutation. The 15 Type II mutations have a functional effect on FH activity. The relatively large number of Type II mutations in this region is consistent with the importance of SCR-19 and SCR-20 in heparin and C3 binding, and many of these side-chains are exposed.

TABLE 2. FH Mutations Reported in MPGN Patients

| SCR | Nucleotide change <sup>a</sup> | Molecular consequence <sup>b</sup> | DSSP and position of substitution in model | Phenotype | Reference                  |
|-----|--------------------------------|------------------------------------|--|-----------|----------------------------|
| 2   | c.377G>T                       | Arg127Leu (109)                    | E (β5)                                     | I         | Dragon-Durey et al. [2004] |
| 7   | c.1292G>C                      | Cys431Ser (413)                    | E (β5, C3)                                 | I         | Dragon-Durey et al. [2004] |
| 9   | c.1606T>C                      | Cys536Arg (518) <sup>c</sup>       | E (β3, C2)                                 | I         | Ault et al. [1997]         |
| 11  | c.2017G>C                      | Cys673Ser (655)                    | E (β5, C3)                                 | I         | Dragon-Durey et al. [2004] |
| 16  | c.2876G>A                      | Cys959Tyr (941) <sup>c</sup>       | E (β3, C2)                                 | I         | Ault et al. [1997]         |

<sup>a</sup>cDNA numbering is taken from RefSeq # NM\_000186.1 where +1 corresponds to the A of the ATG initiation codon.

<sup>b</sup>Residues are numbered starting from the initiating methionine residue at +1. The number of the residue with respect to the mature FH protein (PDB code: 1haq) is indicated in brackets.

<sup>c</sup>Patient was a compound heterozygote for C518R and C941Y.

The increased number of 27 known FH mutations in SCR-19 and SCR-20 compared to the 10 that were considered previously [Perkins and Goodship, 2002] has altered our structural understanding of how CFH mutations cause HUS. The distribution of these 27 mutations across SCR-19 and SCR-20 now argues against the previously predicted binding site for heparin at the junction of SCR-19 and SCR-20. This early model was based on the analysis of what is now seen to be too few mutations and the superimposition of a FH homology model with the crystal structure of the acidic fibroblast growth factor–heparin complex was not unambiguous [DiGabriele et al., 1998]. Another earlier prediction suggested that heparin interacted with both the tip and side of SCR-20 on the basis of manual docking and energy minimization simulations; however, this modeling was also not unambiguous [Hellwage et al., 2002]. The recent crystal structure of the vaccinia complement control protein with heparin is a closer analogue of the interaction between SCR-20 of FH and heparin [Ganesh et al., 2004]. This showed that heparin interacts with only the C-terminal tip of SCR-20. The crystal structure suggested that the FH residues Lys1202, Arg1206, and Lys1220 may make direct contact with heparin (Fig. 4a).

Interestingly, none of these three residues are implicated with HUS (Supplementary Table S1; Fig. 4a). Further support for the localization of the heparin-binding site at the tip of SCR-20 comes from sequence alignments of two factor H-related proteins FHR-3 and FHR-4. Neither of these bind to heparin at their C-terminal domains, and these residues are missing [Hellwage et al., 2002]. In addition, analyses of the six-domain SCR-15/SCR-20 fragment of FH showed that heparin-binding at SCR-20 is weakened by the following five mutants: Arg1203Glu, Arg1206Glu, Arg1210Ser, Lys1220Ser, and Arg1221Ala [Hellwage et al., 2002]. While only Arg1210 is associated with HUS (Supplementary Table S1), all five mutants are near the C-terminal tip of SCR-20 (Fig. 4a). Two of them correspond to the recent crystallography-predicted heparin-binding site on FH [Ganesh et al., 2004].

The increasing evidence for a heparin-binding site at the tip of SCR-20 also makes it less easy to account directly for the role of the 31 missense mutations across SCR-16 to SCR-20 in causing HUS. The majority of these demonstrate Type II phenotypes (Fig. 2). A single SCR domain is  $3.6 \pm 0.2$  nm in length [Aslam et al., 2003]. If FH were to be a straight molecule, its overall length is expected to be in the region of 73 nm. Instead, FH is observed to possess a folded-back SCR structure of maximal length 43 nm by solution scattering [Aslam and Perkins, 2001]. The tip of SCR-20 is required to be solvent-exposed and unhindered by the rest of the FH structure in order that it can bind to an anionic polysaccharide on a cell surface. It is possible that a residue mutation within SCR-16 to SCR-20 may affect the folded-back

inter-SCR arrangement in FH to the extent that SCR-20 is less able to interact with heparin. This interpretation may be supported by the observed conformation rearrangement of the C-terminal SCR domain relative to the other three SCR domains in the crystal structures of vaccinia coat protein with and without bound heparin [Ganesh et al., 2004]. As an alternative explanation, if the correct association of C3b with FH is important for the avoidance of HUS, it may be the case that these Type II residue substitutions perturb residues that either directly interact with C3b or with the inter-SCR arrangement that is necessary for the C-terminal domains to interact with C3b. The approximate dimensions of C3b are 18 nm × 2 nm × 10 nm [Perkins and Sim, 1986]; hence, C3b is large enough to present difficulty in forming contacts with FH if its folded back structure has been altered to offer steric hindrance to C3b binding.

## MPGN

FH deficiency is also associated with another rare kidney disease, MPGN. Whereas HUS patients show a wide range of FH plasma levels, with the majority (37%) of mutations showing Type II phenotypes, all of the patients reported with MPGN showed decreased FH plasma levels and Type I phenotypes. This lack of plasma FH leads to hypocomplementemia, progressive glomerulonephritis, mesangial cell proliferation, and an increase in mesangial matrix size. In contrast, the FH missense mutations seen in the majority of HUS patients rarely resulted in hypocomplementemia, and lead to dysfunctional FH protein present in normal amounts in plasma.

To utilize our FH structural models further, five mutations associated with MPGN (Table 2) were analyzed. MPGN is observed in patients with a complete FH deficiency (compound heterozygous or homozygous) where there is little or no FH in plasma. Interestingly, a mutation at the same position (Cys673) is seen in a heterozygous HUS patient (Cys673Tyr) and a homozygous MPGN patient (Cys673Ser) [Dragon-Durey et al., 2004]. Four of the five mutations (Cys431Ser, Cys536Arg, Cys673Ser, and Cys959Tyr) involve the substitution of a buried conserved Cys residue on either β-strand β3 or β5, whereas the remaining mutation, Arg127Leu, involves a buried Arg residue on β-strand β5 close to the disulfide bridge. Two mutations have also been reported in a Norwegian Yorkshire breed of pig with MPGN symptoms [Hegasy et al., 2002]. Leu493Val and Ile1166Arg (numbered with respect to porcine FH; GenBank accession no: AJ278470) occur in SCR-9 and SCR-20, respectively. The Leu493Val substitution most likely represents a polymorphism since a valine is seen at this position (Val513) in human FH. The Ile1166Arg substitution corresponds to Ile1179Arg in human FH. The model of SCR-20 places Ile1179 on β-strand β1.

The substitution of a large polar arginine residue for the smaller hydrophobic isoleucine residue may affect side chain packing around the  $\beta$ -strand and destabilize the SCR-20 domain to prevent the secretion of FH. This structural analysis provides explanations for defective secretion that results in the lack of FH in the kidneys of MPGN patients. It is possible that the difference between MPGN and HUS relates to the point at which a deficient FH serum level becomes effective in a patient. A patient that has little or no FH in plasma early in life may result in the onset of MPGN, while HUS may result from an immunological insult due to infection or stress that destabilizes a prevailing balance of mutant and normal serum FH in the heterozygous patient.

## AMD

FH has been implicated with AMD of the eye retina, the leading cause of blindness in the developed world, through the observation of a Tyr402His polymorphism in SCR-7 in affected patients [Klein et al., 2005; Haines et al., 2005; Hageman et al., 2005; Edwards et al., 2005]. The allele frequencies were found to be 54% for tyrosine and 46% for histidine from a control population; however, 94% of AMD cases were found to have the histidine allele [Haines et al., 2005]. This polymorphism has also been previously identified in HUS patients, but was not associated with the disease as it was seen in both patients and controls [Neumann et al., 2003]. Our homology model of SCR-7 shows His402 to be on a surface exposed loop in close proximity to positively charged exposed residues Arg404 and Lys405. FH constructs in which both Arg404 and Lys405 are replaced with uncharged alanine residues showed reduced binding to heparin, C-reactive protein (CRP), and M protein, indicating these residues are important for binding [Giannakis et al., 2003]. The change between tyrosine and histidine in close proximity to these residues could inhibit heparin-binding and lead to a dysfunctional FH molecule. It is interesting that the presence of histidine at this position might lead to a dysfunctional FH and cause AMD, but not cause HUS.

## CONCLUSIONS

To our knowledge, this database is the only up-to-date and maintained locus specific database for HUS mutations within *CFH*. It allows rapid access to the entire listing of published mutations in *CFH* including splice site mutations, point mutations, insertions and deletions, and non-disease causing polymorphisms. It is now possible to view each mutation in the full context of what has been previously elucidated for other *CFH* mutations. The structural analysis tools will allow fast and easy classification of new mutations. We would like to encourage the submission of new mutations and new patients in order to keep the database up to date and improve its utility for the clinical community. As these submissions will be via an electronic form, new data in the database will be automatically organized in the required structured and complete manner necessary to improve the utility of the database. Previously, the analysis of mutations across the *CFH* gene has been difficult due to the lack of standardized reports of mutations and patient data. One anticipated benefit of the web database is that the information contained within this may lead to the standardization of the serum levels of FH required to diagnose whether a given mutation is Type I or Type II. It may help define the level of FH necessary in serum in order to avoid the onset of HUS in patients. The restoration of FH to the serum of patients with atypical HUS has been proposed as a therapeutic strategy [Remuzzi et al., 2002].

## ACKNOWLEDGMENTS

Ms. R.E. Saunders and Professor S.J. Perkins are supported by grants from the Medical Research Council and the Wellcome Trust. Prof. T.H.J. Goodship is supported by the National Kidney Research Fund and the Robin Davies Trust. We thank Mr. Steffan Luczyn for generous computational support.

## REFERENCES

- Altschul SF, Koonin EV. 1998. Iterated profile searches with PSI-BLAST—a tool for discovery in protein databases. *Trends Biochem Sci* 23:444–447.
- Aslam M, Perkins SJ. 2001. Folded-back solution structure of monomeric Factor H of human complement by synchrotron X-ray and neutron scattering, analytical ultracentrifugation and constrained molecular modelling. *J Mol Biol* 309:1117–1138.
- Aslam M, Guthridge JM, Hack BK, Quigg RJ, Holers VM, Perkins SJ. 2003. The extended multidomain solution structures of the complement protein Crry and its chimeric conjugate Crry-Ig by scattering, analytical ultracentrifugation and constrained modelling: implications for function and therapy. *J Mol Biol* 329: 525–550.
- Ault BH, Schmidt BZ, Fowler NL, Kashtan CE, Ahmed AE, Vogt BA, Colten HR. 1997. Human factor H deficiency; mutations in framework cysteine residues and block in H protein secretion and intracellular catabolism. *J Biol Chem* 272:25168–25175.
- Ault BH. 2000. Factor H and the pathogenesis of renal diseases. *Pediatr Nephrol* 14:1045–1053.
- Ballermann BJ. 1998. Endothelial cell activation. *Kidney Int* 53:1810–1826.
- Barlow PN, Norman DG, Steinkasserer A, Horne TJ, Pearce J, Driscoll PC, Sim RB, Campbell ID. 1992. Solution structure of the fifth repeat of factor H: a second example of the complement control protein module. *Biochemistry* 31: 3626–3634.
- Barlow PN, Steinkasserer A, Norman DG, Kieffer B, Wiles AP, Sim RB, Campbell ID. 1993. Solution structure of a pair of complement modules by nuclear magnetic resonance. *J Mol Biol* 232:268–284.
- Berman HM, Westbrook J, Feng Z, Gilliland G, Bhat TN, Weissig H, Shindyalov IN, Bourne PE. 2000. The Protein Data Bank. *Nucleic Acids Res* 28:235–242.
- Blackmore TK, Sadlon TA, Ward HM, Lublin DM, Gordon DL. 1996. Identification of a heparin-binding domain in the seventh short consensus repeat of complement factor H. *J Immunol* 157:5422–5427.
- Blackmore TK, Hellwege J, Sadlon TA, Higgs N, Zipfel PF, Ward HM, Gordon DL. 1998. Identification of the second heparin-binding domain in human complement factor H. *J Immunol* 160:3342–3348.
- Bouma B, de Groot PG, van den Elsen JMH, Ravelli RBG, Schouten A, Simmelink MJA, Derksen RHW, Kroon J, Gros P. 1999. Adhesion mechanism of human  $\beta$ 2-glycoprotein I to phospholipids based on its crystal structure. *EMBO J* 18: 5166–5174.
- Buddles MR, Donne RL, Richards A, Goodship JA, Goodship THJ. 2000. Complement factor H gene mutation associated with autosomal recessive atypical hemolytic uremic syndrome. *Am J Hum Genet* 66:1721–1722.
- Caprioli J, Bettinaglio P, Zipfel PF, Amadei B, Daina E, Gamba S, Skerka C, Marziliano N, Remuzzi G, Noris M. 2001. The molecular basis of familial hemolytic uremic syndrome: mutation

- analysis of factor H gene reveals a hot spot in short consensus repeat 20. *J Am Soc Nephrol* 12:297–307.
- Caprioli J, Castelletti F, Bucchioni S, Bettinaglio P, Bresin E, Pianetti G, Gamba S, Brioschi S, Daina E, Remuzzi G, Noris M. International Registry of Recurrent and Familial HUS/TTP. 2003. Complement factor H mutations and gene polymorphisms in haemolytic uraemic syndrome: the C-257T, the A2089G and the G2881T polymorphisms are strongly associated with the disease. *Hum Mol Genet* 12:3385–3395.
- Claustres M, Horaitis O, Vanevski M, Cotton RG. 2002. Time for a unified system of mutation description and reporting: a review of locus-specific mutation databases. *Genome Res* 12: 680–688.
- Creighton TE. 1993. *Proteins, structures and molecular properties*, 2nd edition. New York: WH Freeman & Co. 507 p.
- DiGabriele AD, Lax I, Chen DI, Svahn CM, Jaye M, Schlessinger J, Hendrickson WA. 1998. Structure of a heparin-linked biologically active dimer of fibroblast growth factor. *Nature* 292:812–817.
- Dragon-Durey MA, Fremeaux-Bacchi V, Loirat C, Blouin J, Niaudet P, Deschenes G, Coppo P, Herman Fridman W, Weiss L. 2004. Heterozygous and homozygous factor H deficiencies associated with hemolytic uremic syndrome or membranoproliferative glomerulonephritis: report and genetic analysis of 16 cases. *J Am Soc Nephrol* 15:787–795.
- Edwards AO, Ritter R III, Abel KJ, Manning A, Panhuysen C, Farrer LA. 2005. Complement factor H polymorphism and age-related macular degeneration. *Science* 308:421–424.
- Ganesh VK, Smith SA, Kotwal GJ, Murthy KH. 2004. Structure of vaccinia complement protein in complex with heparin and potential implications for complement regulation. *Proc Natl Acad Sci USA* 101:8924–8929.
- Giannakis E, Jokiranta TS, Male DA, Ranganathan S, Ormsby RJ, Fischetti VA, Mold C, Gordon DL. 2003. A common site within factor H SCR 7 responsible for binding heparin, C-reactive protein and streptococcal M protein. *Eur J Immunol* 33: 962–969.
- Haines JL, Hauser MA, Schmidt S, Scott WK, Olson LM, Gallins P, Spencer KL, Kwan SY, Noureddine M, Gilbert JR, Schnetz-Boutaud N, Agarwal A, Postel EA, Pericak-Vance MA. 2005. Complement factor H variant increases the risk of age-related macular degeneration. *Science* 308:419–421.
- Hageman GS, Anderson DH, Johnson LV, Hancox LS, Taiber AJ, Hardisty LI, Hageman JL, Stockman HA, Borchardt JD, Gehrs KM, Smith RJ, Silvestri G, Russell SR, Klaver CC, Barbazetto I, Chang S, Yannuzzi LA, Barile GR, Merriam JC, Smith RT, Olsh AK, Bergeron J, Zernant J, Merriam JE, Gold B, Dean M, Allikmets R. 2005. A common haplotype in the complement regulatory gene factor H (HF1/CFH) predisposes individuals to age-related macular degeneration. *Proc Natl Acad Sci USA* 102:7227–7232.
- Hegasy GA, Manuelian T, Hogasen K, Jansen JH, Zipfel PF. 2002. The molecular basis for hereditary porcine membranoproliferative glomerulonephritis Type II. Point mutations in the factor H coding sequence block protein secretion. *Am J Pathol* 161: 2027–2034.
- Hellwage J, Jokiranta TS, Friese MA, Wolk TU, Kampen E, Zipfel PF, Meri S. 2002. Complement C3b/C3d and cell surface polyanions are recognized by overlapping binding sites on the most carboxyl-terminal domain of complement factor H. *J Immunol* 169:6935–6944.
- Higgins D, Thompson J, Gibson T. 1994. CLUSTAL W: improving the sensitivity of progressive multiple sequence alignment through sequence weighting, position-specific gap penalties and weight matrix choice. *Nucleic Acids Res* 22:4673–4680.
- Horstmann RD, Sievertsen HJ, Knobloch J, Fischetti VA. 1988. Antiphagocytic activity of streptococcal M protein: selective binding of complement control protein factor H. *Proc Natl Acad Sci USA* 85:1657–1661.
- Jokiranta TS, Hellwage J, Koistinen V, Zipfel PF, Meri S. 2000. Each of the three binding sites of factor H interacts with a distinct site on C3b. *J Biol Chem* 275:27657–27662.
- Kabsch W, Sander C. 1983. Dictionary of protein secondary structure: pattern recognition of hydrogen-bonded and geometrical features. *Biopolymers* 22:2577–2637.
- Klein RJ, Zeiss C, Chew EY, Tsai JY, Sackler RS, Haynes C, Henning AK, Sangiovanni JP, Mane SM, Mayne ST, Bracken MB, Ferris FL, Ott J, Barnstable C, Hoh J. 2005. Complement factor H polymorphism in age-related macular degeneration. *Science* 308:385–389.
- Koistinen V. 1992. Limited tryptic cleavage of complement factor H abrogates recognition of sialic acid-containing surfaces by the alternative pathway of complement. *Biochem J* 283:317–319.
- Koistinen V. 1993. Effects of sulphated polyanions on functions of complement factor H. *Mol Immunol* 30:113–118.
- Kraiczky P, Hellwage J, Skerka C, Becker H, Kirschfink M, Simon MM, Brade V, Zipfel PF, Wallich R. 2004. Complement resistance of *Borrelia burgdorferi* correlates with the expression of BbCRASP-1, a novel linear plasmid-encoded surface protein that interacts with human factor H and FHL-1 and is unrelated to Erp proteins. *J Biol Chem* 279:2421–2429.
- Kraulis PJ. 1991. MOLSCRIPT: a program to produce both detailed and schematic plots of protein structures. *J Appl Cryst* 24:946–950.
- Laskowski RA, MacArthur MW, Moss DS, Thornton JM. 1993. PROCHECK: a program to check the stereochemical quality of protein structures. *J Appl Cryst* 26:283–291.
- Law SKA, Reid KBM. 1995. *Complement*, 2nd edition. Oxford: IRL Press. 88 p.
- Manuelian T, Hellwage J, Meri S, Caprioli J, Noris M, Heinen S, Jozsi M, Neumann HP, Remuzzi G, Zipfel PF. 2003. Mutations in factor H reduce binding affinity to C3b and heparin and surface attachment to endothelial cells in hemolytic uremic syndrome. *J Clin Invest* 111:1181–1190.
- Merritt EA, Murphy MEP. 1994. Raster3D version 2.0. A program for photorealistic molecular graphics. *Acta Cryst D* 50:869–873.
- Murthy KHM, Smith SA, Ganesh VK, Judge KW, Mullin N, Barlow PN, Ogata CM, Kotwal GJ. 2001. Crystal structure of a complement control protein that regulates both pathways of complement activation and binds heparan sulfate proteoglycans. *Cell* 104:301–311.
- Neumann HP, Salzmann M, Bohnert-Iwan B, Manuelian T, Skerka C, Lenk D, Bender BU, Cybulla M, Riegler P, Konigsrainer A, Neyer U, Bock A, Widmer U, Male DA, Franke G, Zipfel PF. 2003. Haemolytic uraemic syndrome and mutations of the factor H gene: a registry-based study of German speaking countries. *J Med Genet* 40:676–681.
- Ormsby RJ, Jokiranta TS, Griggs K, Giannakis E, Sadlon T, Gordon DL. 2004. Localisation of the third heparin-binding domain in human complement factor H. *Molec Immunol* 41:289–290.
- Pangburn MK, Atkinson MAL, Meri S. 1991. Localization of the heparin-binding site on complement factor H. *J Biol Chem* 266:16847–16853.
- Pangburn MK. 2000. Host recognition and target differentiation by factor H, a regulator of the alternative pathway of complement. *Immunopharmacology* 49:149–157.

- Pérez-Caballero D, González-Rubio C, Gallardo ME, Vera M, López-Trascasa M, Rodríguez de Córdoba S. 2001. Clustering of missense mutations in the C-terminal region of factor H in atypical hemolytic uremic syndrome. *Am J Hum Genet* 68: 478–484.
- Perkins SJ, Sim RB. 1986. Molecular modelling of human complement component C3 and its fragments by solution scattering. *Eur J Biochem* 157:155–168.
- Perkins SJ, Gilbert HE, Aslam M, Hannan J, Holers VM, Goodship THJ. 2002. Solution structures of complement components by X-ray and neutron scattering and analytical ultracentrifugation. *Biochem Soc Trans* 30:996–1001.
- Perkins SJ, Goodship THJ. 2002. Molecular modelling of mutations in the C-terminal domains of factor H of human complement: a new insight into haemolytic uraemic syndrome. *J Mol Biol* 316:217–224.
- Ram S, Sharma AK, Simpson SD, Gulati S, McQuillen DP, Pangburn MK, Rice PA. 1998. A novel sialic acid binding site on factor H mediates serum resistance of sialylated *Neisseria gonorrhoeae*. *J Exp Med* 187:743–752.
- Remuzzi G, Ruggenenti P, Codazzi D, Noris M, Caprioli J, Locatelli G, Gridelli B. 2002. Combined kidney and liver transplantation for familial haemolytic uraemic syndrome. *Lancet* 359: 1671–1672.
- Richards A, Buddles MR, Donne RL, Kaplan BS, Kirk E, Venning MC, Tielemans CL, Goodship JA, Goodship THJ. 2001. Factor H mutations in hemolytic uremic syndrome cluster in exons 18–20, a domain important for host cell recognition. *Am J Hum Genet* 68:485–490.
- Rodríguez de Córdoba S, Esparza-Gordillo J, Goicoechea de Jorge E, Lopez-Trascasa M, Sanchez-Corral P. 2004. The human complement factor H: functional roles, genetic variations and disease associations. *Mol Immunol* 41:355–367.
- Sanchez-Corral P, Bellavia D, Amico L, Brai M, Rodríguez de Córdoba S. 2000. Molecular basis for factor H and FHL-1 deficiency in an Italian family. *Immunogenetics* 51: 366–369.
- Sanchez-Corral P, Perez-Caballero D, Huarte O, Simckes AM, Goicoechea E, Lopez-Trascasa M, de Córdoba SR. 2002. Structural and functional characterization of factor H mutations associated with atypical hemolytic uremic syndrome. *Am J Hum Genet* 71:1285–1295.
- Sharma AK, Pangburn MK. 1996. Identification of three physically and functionally distinct binding sites for C3b in human complement factor H by deletion mutagenesis. *Proc Natl Acad Sci USA* 93:10996–11001.
- Warwicker P, Goodship THJ, Donne RL, Pirson Y, Nicholls A, Ward RM, Goodship JA. 1998. Genetic studies into inherited and sporadic haemolytic uraemic syndrome. *Kidney Int* 53: 836–844.
- Zipfel PF, Jokiranta TS, Hellwege J, Koistinen V, Meri S. 1999. The factor H protein family. *Immunopharmacology* 42:53–60.