

Evolutionary Relationships Among Proteins Encoded by the Regulator of Complement Activation Gene Cluster

Julia Krushkal,* Olga Bat,† and Irma Gigli†

*Department of Biology and Biotechnology, Worcester Polytechnic Institute; and †Institute of Molecular Medicine for the Prevention of Human Disease, The University of Texas–Houston Health Science Center

Evolutionary relationships among members of the regulator of complement activation (RCA) gene cluster were analyzed using neighbor-joining and parsimony methods of phylogenetic tree inference. We investigated the structural and functional similarities among short consensus repeats (SCRs) of the following human proteins: the α chain of the C4b-binding protein (C4bp α), factor H (FH), factor H-related proteins (FHR-1 through FHR-4), complement receptors type 1 (CR1) and type 2 (CR2), the CR1-like protein (CR1L), membrane cofactor protein (MCP), decay accelerating factor (DAF), and the sand bass proteins, the cofactor protein (SBP1) and its homolog, the cofactor-related protein (SBCRP-1). Also included are the β chain of the human C4b-binding protein (C4bp β) and the b subunit of human blood-clotting factor XIII (FXIIIb). Our results indicate that the human plasma complement regulators, FH and C4bp α , fall into two distinct groups on the basis of their sequence divergence. Homology among RCA proteins is in agreement with their chromosomal location, with the exception of C4bp β . The evolutionary relationships among individual short consensus repeats are confirmed by the exon/intron structure of the RCA members. Structural similarities among repeats of the RCA proteins correlate with their functional activities and demonstrate the importance of the N-terminal SCRs.

Introduction

The complement system is a complex group of plasma and cell membrane proteins that participate in biologically important processes, including inflammation, host defense, and regulation of the immune response. There are more than 20 plasma proteins, which are involved in the activation and regulation of the complement system. Over half of these proteins function as enzymes, enzyme inhibitors, or enzyme cofactors. Most of the enzymes in the system circulate as inactive precursors expressing proteolytic activity only when complement is activated. Together with the plasma proteins, there are more than 10 cell membrane proteins (complement receptors) which exhibit defined ligand specificity for fragments generated during complement activation (reviewed by Müller-Eberhard 1985; Whaley and Schwaeble 1997). Some of these receptors may also serve as regulators of complement activation or may protect autologous cells from the lytic effect of complement. Receptor ligand interactions, particularly those derived from C3, mediate many of the host defense functions of complement (Weigle et al. 1985).

Activation of the complement system may occur via three distinct pathways, termed classical, alternative, and mannose-binding lectin (fig. 1) (Whaley and Schwaeble 1997; Smith, Azumni, and Nonaka 1999). Each of the three activation pathways leads to the fusion of multiple proteins into distinct enzymatic complexes termed the C3 convertases, which function to proteolytically cleave C3 into two fragments, C3b and C3a. A large portion of the C3b generated rapidly deposits on

the target, and some covalently binds to the C3 convertase itself, resulting in the formation of the C5 convertase. After C3 cleavage, the three pathways converge into the lytic event involving the terminal sequences C5, C6, C7, C8, and C9 and the oligomerization of C9 (membrane attack complex; MAC), which causes the formation of transmembrane channels and cell lysis (reviewed by Müller-Eberhard 1985).

Activation of the complement system is tightly regulated at different steps to prevent uncontrolled activation and to assure that damage to autologous cells is avoided. Regulation is achieved through the intrinsic decay of activating complexes (enzymes) and the action of plasma proteins that act as inhibitors, as well as cell membrane-anchored regulators. Soluble proteins in plasma control C3 activation either in the fluid phase or on the surface of activating particles. C4-binding protein, factor H (FH), and low-molecular-weight FH-like proteins have been shown to mediate these functions (Whaley and Ruddy 1976; Pangburn, Schreiber, and Müller-Eberhard 1977; Gigli, Fujita, and Nussenzweig 1979; Misasi et al. 1989). Similarly, a number of cell membrane-anchored regulators act to prevent uncontrolled C3 activation. They are complement receptor type 1 (CR1), decay accelerating factor (DAF/CD55), and membrane cofactor protein (MCP/CD46) (Fearon 1979; Nicholson-Weller et al. 1982; Medof, Kinoshita, and Nussenzweig 1984; Liszewski, Post, and Atkinson 1991).

In humans, a number of the regulatory proteins (table 1) are encoded by a cluster of genes located on the long arm of chromosome 1 (1q32). This region is called the regulator of complement activation (RCA) gene cluster (Weis et al. 1987; Heine-Suñer et al. 1997). Although the proteins within the RCA family vary in size, they share significant primary amino acid structure similarities. They are organized in tandem structural units termed short consensus repeats (SCRs), which are present in multiple copies in the protein. Each SCR consists

Key words: regulator of complement activation, short consensus repeats, classical pathway, alternative pathway, human, barred sand bass.

Address for correspondence and reprints: Julia Krushkal, Department of Biology and Biotechnology, Worcester Polytechnic Institute, 100 Institute Road, Worcester, Massachusetts 01609-2280. E-mail: krushkal@wpi.edu.

Mol. Biol. Evol. 17(11):1718–1730. 2000

© 2000 by the Society for Molecular Biology and Evolution. ISSN: 0737-4038

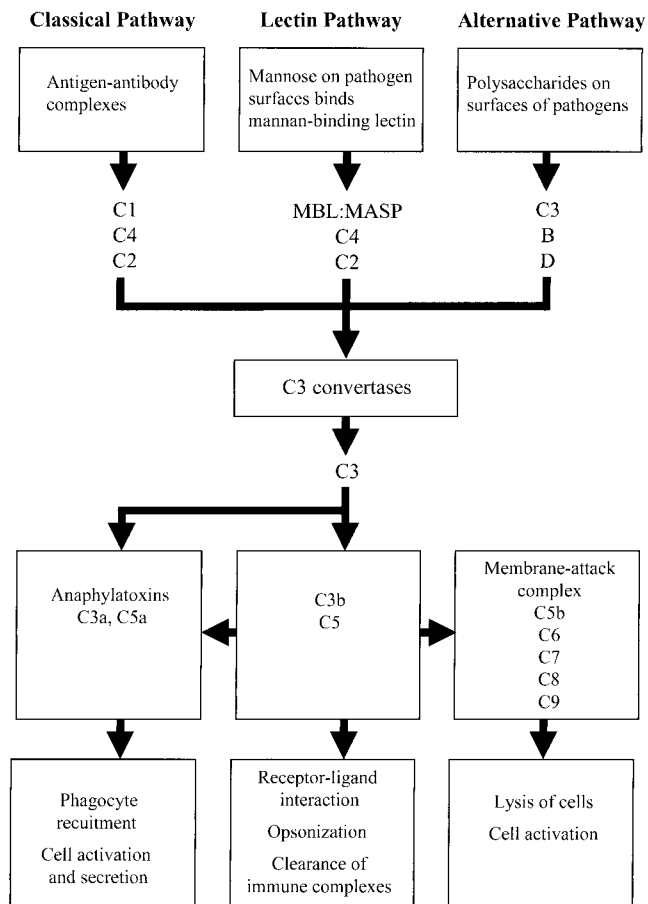


FIG. 1.—Overview of complement activation. This diagram presents an overview of the three complement pathways: classical, alternative, and mannan-binding lectin (MBL). MASP = mannan-binding lectin-associated serine protease. C1, C2, C3, C4, C5, C6, C7, C8, C9, B, and D are other complement proteins. C3a, C3b, C5a, and C5b are cleavage products of C3 and C5, respectively.

of 60–70 highly conserved amino acids, including 4 cysteines. The cysteines are disulphide-bonded, holding the SCRs in a rigid triple-loop structure (reviewed by Hourcade, Holers, and Atkinson 1989).

The two important plasma regulatory proteins encoded within the RCA cluster are FH and the α chain of the C4b-binding protein (C4bp α). They modulate the activation of the alternative and the classical pathways, respectively, by accelerating the decay of the C3 convertases. They also act as cofactors for plasma enzyme factor I (fig. 2) in the enzymatic degradation of C3b and C4b (Whaley and Ruddy 1976; Pangburn, Schreiber, and Müller-Eberhard 1977; Gigli, Fujita, and Nussenzweig 1979).

Several other genes within the RCA cluster code for integral cell membrane proteins with regulatory functions similar to those of the plasma proteins: decay accelerating factor (DAF), membrane cofactor protein (MCP), and complement receptor type 1 (CR1). DAF prevents the formation and accelerates the dissociation of both the classical and the alternative pathway C3 convertases, while MCP is a cofactor for factor I in the degradation of C3b and C4b (Nicholson-Weller et al. 1982; Pangburn, Schreiber, and Müller-Eberhard 1983; Medof, Kinoshita, and Nussenzweig 1984; Fujita et al. 1987; Seya and Atkinson 1989). CR1, the receptor for C3b/C4b, regulates complement in a manner similar to that of the plasma proteins FH and C4bp α (Fearon 1979; Iida and Nussenzweig 1981; Turner 1984). The RCA cluster also contains a gene that encodes for another protein receptor, complement receptor type 2 (CR2; the C3d/Epstein-Barr virus receptor). CR2 interacts with the cleavage fragments iC3b and C3dg that result from the factor I-mediated proteolysis of C3b (Ross et al. 1973).

In addition to these well-characterized regulatory proteins, the RCA cluster includes other homologous

Table 1
Short Consensus Report (SCR)-Containing Sequences of Human Regulation of Complement Activation Proteins and Their Close Sand Bass Homologs Analyzed

Protein	Symbol	Function	No. of SCRs	GenBank Accession No.	GenBank GI
C4-binding protein α chain	C4bp α	Complement regulation	8	AAA36507	190502
Factor H	FH	Complement regulation	20	CAA68704	31965
Factor H-related protein 1	FHR-1	Unknown	5	Q03591	543981
Factor H-related protein 2	FHR-2	Unknown	4	P36980	543983
Factor H-related protein 3	FHR-3	Unknown	5	Q02985	543982
Factor H-related protein 4 ^a	FHR-4	Unknown	5	CAA66980	1524058
Complement receptor type 1	CR1	Complement regulation	30 ^b	P17927	117310
CR1-like protein ^c	CR1L	Unknown	7	563326	563326
Complement receptor type 2	CR2	Complement regulation	15 ^b	P20023	117315
Decay accelerating factor	DAF	Complement regulation	4	P08174	1345942
Membrane cofactor protein	MCP	Complement regulation	4	P15529	1708964
Factor XIII, b subunit	FXIIIb	Blood clotting	10	P05160	119721
C4-binding protein, β chain	C4bp β	Blood clotting	3	P20851	115213
Cofactor protein ^d	SBP1	Complement regulation	17	AAA92556	639895
Cofactor related protein 1 ^d	SBCRP-1	Unknown	3	CAA67355	1438073

^a A number of other human factor H-related proteins exist, but they have not yet been characterized and their sequences are not yet known (Zipfel and Skerka 1994).

^b Several polymorphic variants of human CR1 and CR2 exist that differ from one another in number of SCRs. CR1 alleles contain 23–44 SCRs, while CR2 alleles contain 15 and 16 SCRs (Fujisaku et al. 1989; Farries and Atkinson 1991).

^c CR1L and SBCRP-1 are putative proteins predicted from nucleotide sequences.

^d SBP1 and SBCRP-1 are sand bass proteins, while all other proteins listed in the table are human RCA proteins.

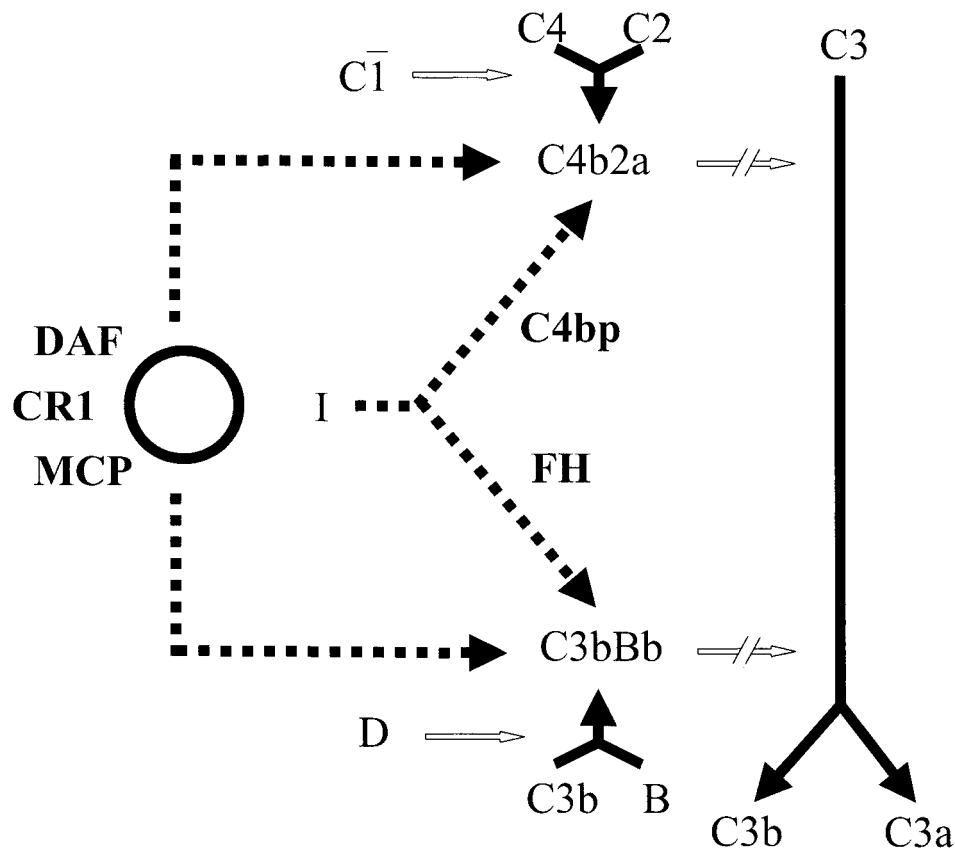


FIG. 2.—Regulation of the classical and alternative pathways by plasma and membrane regulator of complement activation (RCA) proteins. Names of complement regulators encoded by genes in the RCA cluster are shown in bold. C4bp is a C4-binding protein which consists of seven identical α chains (C4bp α) and one β chain (C4bp β). Symbols for other RCA proteins (FH, DAF, MCP, and CR1) are the same as in table 1. I is a plasma regulatory protein factor I with enzymatic activity. C1, C2, C3, C4, and factor B are complement components. C1 is an activated form of C1 that proteolytically cleaves C2 and C4. D is factor D, an enzyme that cleaves factor B in the C3bB complex. C3a and C3b are products of the proteolytic cleavage of C3. C4b2a is the classical pathway C3 convertase, and C3bBb is the alternative pathway C3 convertase. Solid black arrows indicate protein changes. White arrows show proteolytic cleavage effect, while interrupted white arrows indicate inhibition. Dashed arrows indicate regulatory activity of RCA proteins. A black circle indicates similar regulatory effects of DAF, CR1, and MCP as cofactors of factor I.

genes that contain tandem SCRs. Such genes code for proteins related to FH (FHR-1, FHR-2, FHR-3, and FHR-4) (Zipfel and Skerka 1994; Heine-Suñer et al. 1998) and another FH-like gene (Weiss and Cannich 1998) which may encode an FH-related protein not yet characterized. Two other RCA genes, CR1-like (CR1L) and MCP-like (MCPL), share a high degree of similarity to CR1 and MCP genes, respectively (Hourcade et al. 1990, 1992). Although no CR1L or MCPL protein products have yet been confirmed in humans, the CR1L protein has been identified in baboons (Birmingham et al. 1996). There are two known pseudogenes, C4BPAL1 and C4BPAL2, homologous to the gene encoding for C4bp α (Pardo-Manuel de Villena and Rodríguez de Córdoba 1995).

RCA genes that encode SCR-containing proteins with functions other than complement regulation have been identified. These proteins are the β chain of the C4-binding protein (C4bp β) and the b subunit of blood-clotting factor XIII (FXIIIb), both of which participate in coagulation (Chung, Lewis, and Folk 1974; Pangburn and Müller-Eberhard 1985; Hillarp and Dahlbäck 1988). Some human complement proteins containing SCRs do

not belong to the RCA group because their genes are not located within the RCA cluster (e.g., Reid and Day 1989).

The SCRs found in RCA proteins carry a variety of functions, including protein binding and cofactor activity in complement regulation (reviewed by Hourcade, Holers, and Atkinson 1989; Reid and Day 1989). These repeats appear to have a very ancient origin. SCR-containing genes sharing sequence similarity with SCR regions of human complement genes have been found in insects (*Drosophila*), nematodes (*Caenorhabditis elegans*), and sponges (*Geodia cydonium*) (Hoshino et al. 1993; Ainscough et al. 1998; Blumbach et al. 1998; Pahler et al. 1998). However, the function of the products of such genes in complement has not been investigated. The SCRs from different proteins, even occurring in highly divergent animal lineages, contain a number of conserved amino acid residues, which indicates their common evolutionary origin and functional importance. The degree of divergence between individual SCRs is variable, making it possible to use phylogenetic methods to study their evolution (Krushkal, Kemper, and Gigli 1998).

To date, the species most divergent from humans found to contain a functional protein regulating complement activation is the bony fish barred sand bass, *Parablax nebulifer*. The sand bass cofactor protein isolated from plasma and by recombinant cDNA expression, SBP1, has functional and structural similarity to both human C4bp α and FH. It consists of tandem SCRs, binds to both C4b and C3b, and serves as a cofactor in their enzymatic cleavage by factor I (Kaidoh and Gigli 1987, 1989; Dahmen et al. 1994; Kemper, Zipfel, and Gigli 1998). Sand bass has at least one more SCR-containing gene, and its predicted protein, sand bass cofactor related protein 1 (SBCRP-1), shares structural similarity with SBP1 (Zipfel et al. 1996).

Because RCA proteins share extensive structural similarity and functional activities, it is important to know how they originated and evolved. Farries and Atkinson (1991) proposed that the FH, C4bp α , MCP, DAF, and CR1/CR2 lineages have a starlike origin. However, because the RCA family includes a variety of plasma factors and membrane proteins participating in regulation of both the classical and the alternative pathways, these factors may have appeared at different times in evolution. Genes encoding for C3 and factor B proteins, which are involved in the alternative pathway C3 convertase, probably appeared earlier than C4 and C2, which are involved in the classical pathway convertase (Farries and Atkinson 1991; Seeger, Mayer, and Klein 1996; Sunyer, Zarkadis, and Lambris 1998). The complement regulators that interact with them may also have appeared at different times in evolutionary history.

In the present study, we investigated the homology relationships among RCA proteins and their individual SCRs. We also compared how the homology among individual domains of RCA proteins corresponded to their exon/intron structure and functional activity. Based on these comparisons, we discuss the possible evolution of the RCA members.

Materials and Methods

Sequence Data

To investigate the relationships among RCA proteins by phylogenetic tree inference, we compared their individual SCRs. Due to the high degree of sequence divergence among SCRs (Krushkal, Kemper, and Gigli 1998), our analysis was restricted to protein sequences. We therefore did not include the two pseudogenes, C4BPAL1 and C4BPAL2, and the MCPL gene, for which no protein products are known. The list of protein sequences used in this analysis and their GenBank accession numbers and GI numbers are provided in table 1.

Alignment of Sequences

Since each protein contains multiple repeats homologous to one another, it was not possible to determine a priori how to align the complete protein sequence. We therefore investigated the homology relationships among individual SCRs. The inferred relationships among repeats were used both to analyze the evolutionary relationships among proteins and to com-

pare the sequence similarities with gene structure and functional activities. One hundred thirty-two SCRs from human RCA proteins and their sand bass homologs (table 1) were aligned by an automated program, CLUSTAL W, version 1.7 (Thompson, Higgins, and Gibson 1994), and the resulting alignment was further improved manually.

Inference of Phylogenetic Trees from Individual SCRs

Phylogenetic trees were inferred by the neighbor-joining and parsimony methods. When the trees were inferred by the neighbor-joining method (Saitou and Nei 1987), we used two approaches to measure amino acid sequence dissimilarity between repeats. One approach took into account multiple amino acid substitutions and variation of substitution rate among sites (Ota and Nei 1994). We used the value 0.93 of the shape parameter α for the gamma distribution of substitution rate among sites, estimated from SCRs of human C4bp α and FH (Krushkal, Kemper, and Gigli 1998). The neighbor-joining tree based on corrected distances was inferred with 1,000 bootstrap replications using the computer software NJBOOTAG (modified from program NJBOOTLI; A. Zharkikh, personal communication).

The second approach was to use Dayhoff (1979, p.348) PAM distances. These distances were employed to infer neighbor-joining trees by programs from the PHYLIP phylogenetic package, version 3.57c (Felsenstein 1989). The trees were inferred with 500 bootstrap replications with the programs SEQBOOT, PROTDIST, and NEIGHBOR. The consensus of these 500 trees was inferred with the program CONSENSE.

We also inferred phylogenetic trees from 132 SCRs by the equally weighted parsimony method using PAUP*, version 4.0d63 (Swofford 1998). Due to the large number of sequences, all possible tree rearrangements could not be evaluated under a feasible computer time and memory. We therefore limited the search to five replications, during which over two million possible tree rearrangements were tried. After these rearrangements, 91 trees with the minimal length of 3,871 substitutions were saved, and the majority-rule (50%) consensus of these trees was inferred. Although this consensus tree may not represent the shortest possible tree, it is in good agreement with the trees inferred by the neighbor-joining method, and therefore it may be informative in the analysis of the RCA proteins.

All of the phylogenetic trees were midpoint-rooted with the help of the RETREE program from the PHYLIP package. The trees were presented using the phylogeny drawing program TreeView, versions 1.6.0 and 1.6.1 (Page 1996).

Inclusion and Exclusion of Indels

The results presented here for the tree inferred by NJBOOTAG using the neighbor-joining method with correction for the variation of substitution rate among sites are based on inclusion of insertions and deletions (indels). In this case, gaps and amino acid substitutions were treated equally. Similar results were also obtained

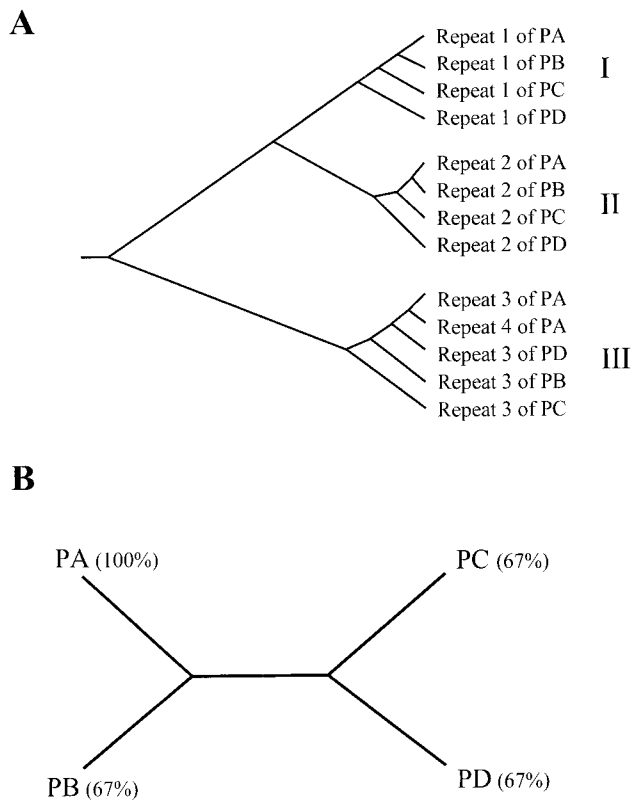


FIG. 3.—An example explaining the algorithm for inferring a summary tree for the entire protein sequences from the phylogenetic tree for individual repeats. PA, PB, PC, and PD are four hypothetical protein sequences, where PA consists of four repeats, while PB, PC and PD contain three repeats each. *A*, A hypothetical phylogenetic tree for the 13 individual repeats from sequences PA, PB, PC, and PD. *B*, A summary phylogeny for sequences PA, PB, PC, and PD inferred as a majority-rule consensus tree from the three subtrees (I–III) containing individual repeats. For each sequence, the proportion (%) of repeats supporting the summary topology is shown.

when indel positions were excluded from analysis (data not shown). In the trees inferred by PHYLIP using Dayhoff PAM distances and by PAUP* using parsimony analysis, gaps were considered an unknown state.

Inference of the Summary Tree of Relationships Among the Entire RCA Protein Sequences

The summary tree of relationships among the entire RCA protein sequences was unambiguously deduced from the phylogenetic trees of individual repeats. The summary tree was derived as a majority-rule consensus tree from the subtrees containing individual short consensus repeats. A hypothetical example of such inference is shown in figure 3. It includes four protein sequences, PA, PB, PC, and PD. PA contains four homologous repeats, while PB, PC, and PD contain three repeats each. A summary tree of relationships among PA, PB, PC, and PD is inferred as a majority-rule consensus tree from the three subtrees (I–III) of the phylogenetic tree that consists of individual repeats (fig. 3A). For each sequence, a proportion of repeats supporting the summary tree was computed (fig. 3B).

Results

The very large number (132) of SCRs used in the phylogenetic tree inference mandates that instead of presenting the entire trees, we list the nodes supported by more than 50% of the replications (tables 2 and 3). The threshold of 50% for bootstrap support was necessary because of the very short lengths of SCRs (60–70 amino acids) and their high degree of divergence, which tend to result in lower bootstrap values. We also present the summary tree that shows the relationships among the RCA proteins (fig. 4) unambiguously deduced from the trees containing individual SCRs. Table 4 shows the proportion of SCRs in each RCA sequence that support the summary tree topology shown in figure 4. In addition, figures 5 and 6 show the homology relationships among individual SCRs determined from the phylogenetic trees. The phylogenetic trees inferred from individual repeats of RCA proteins, the distances among individual SCRs, and the sequence alignment are provided as a supplement which can be viewed via a World Wide Web page (see *Supplementary Material*).

Relationships Among RCA Members and Their Chromosomal Location

Tables 2 and 3 list the clusters that were identical in the three phylogenetic trees inferred from the 132 repeats of the RCA proteins showing more than 50% support among bootstrap replicates and in a consensus tree. With the exception of two nodes that had consensus tree support of 79% and 84%, all nodes listed in tables 2 and 3 had 100% support in the consensus tree, i.e., they were present in a strict consensus tree. A large number of other clusters also had high support but a slightly different joining of lineages among the three trees. For example, clustering of SCRs 1, 8, 15, and 22 of CR1 and SCR 1 of CR1L had 99%–100% support in all three trees. In the consensus of parsimony trees, SCR 1 of CR1 and SCR 1 of CR1L were the closest to SCRs 8 and 15 of CR1, while in the neighbor-joining trees they were the closest to SCR 22 of CR1. Such clusters were not reported in tables 2 and 3, but their support values are shown in phylogenetic trees presented in the supplement (see *Supplementary Material*).

Figure 4 shows the homology relationships among RCA proteins summarized from all trees inferred, and table 4 indicates that this tree topology is strongly supported by the majority of repeats in each protein. In contrast to the previously proposed scheme of the starlike evolution of RCA members (Farries and Atkinson 1991), we found that their evolution was not starlike. The observed protein sequence similarities of RCA members mostly correlate with the physical location of the human RCA genes on chromosome 1. Our phylogenetic comparisons indicate the presence of two distinct groups which are known to be separated by a genetic distance of 7–22.5 cM (Rey-Campos, Rubinstein, and Rodríguez de Córdoba 1988; Heine-Suñer et al. 1997). One group (group 1) includes human FH, FH-related proteins FHR-1, FHR-2, FHR-3, and FHR-4, and coagulation factor FXIIIb. Indeed, the genes encoding

Table 2
Clusters of Repeats from FH, SBP1, and Their Homologs (group 1 proteins) Supported by Values Above 50%

SCRs IN LINEAGE 1	SCRs IN LINEAGE 2	SUPPORT (%)		
		NJBOOTAG	PHYLIP	PAUP*
SCR2 of FH	SCR 2 of SBP1	94	88	100
SCR 6 of FH	SCR 1 of FHR-3	100	96	100
SCRs 6 of FH, 1 of FHR-3, 1 of FHR-4	SCRs 1 of FHR-1, 1 of FHR-2	99	92	100
SCR 7 of FH	SCR 2 of FHR3	100	99	100
SCR 8 of FH	SCRs 3 of FHR-3, 2 of FHR-4	100	91	100
SCR 9 of FH	SCR 3 of FHR-4	100	98	100
SCRs 19 of FH, 4 of FHR-1*	SCR 3 of FHR-2	100	88	100
SCRs 19 of FH, 4 of FHR-1, 3 of FHR-3	SCRs 4 of FHR-3, 4 of FHR-4	96	91	100
SCR 20 of FH	SCR 5 of FHR-1	99	97	100
SCRs 20 of FH, 5 of FHR-1, 4 of FHR-2	SCRs 5 of FHR-3, 5 of FHR-4	61	52	100
SCR 1 of FHR-1	SCR 1 of FHR-2	100	100	100
SCR 4 of FHR-2	SCRs 20 of FH, 5 of FHR-1	74	89	100
SCR 3 of FHR-3	SCR 2 of FHR-4	100	92	100
SCR 4 of FHR-3	SCR 4 of FHR-4	100	100	100
SCR 5 of FHR-3	SCR 5 of FHR-4	100	98	100
SCR 1 of FHR-4	SCRs 6 of FH, 1 of FHR-3	100	96	100
SCR 6 of SBP1	SCR 1 of SBCRP-1	100	100	100
SCR 16 of SBP1	SCR 2 of SBCRP-1	86	82	100
SCR 17 of SBP1	SCR 3 of SBCRP-1	90	89	100
SCRs 17 of SBP-1, 3 of SBCRP-1	SCRs 20 of FH, 5 of FHR-1, 4 of FHR-2, 5 of FHR-3, 5 of FHR-4, 8 and 10 of FXIIIb	96	86	100
SCR 2 of FXIIIb	SCR 3 of FXIIIb	74	51	100

NOTE.—Shown are the support values for group 1 proteins in the three trees inferred from 132 individual short consensus repeats (SCRs). For NJBOOTAG, the numbers show the bootstrap support (%) from 1,000 bootstrap replications for a tree inferred by the neighbor-joining method using correction for multiple hits and substitution rate variation among sites, with indels included in the analysis. For PHYLIP, the numbers indicate the occurrence (%) of the node in a consensus tree of 500 replicates, with trees inferred by the neighbor-joining method using Dayhoff distances. For PAUP*, the numbers show the occurrence of a node in a majority-rule consensus of 91 minimal trees found after five replicates of branch-and-bound search by unweighted parsimony. Lineage 1 and lineage 2 are the two lineages that form a cluster for which the bootstrap value is reported. For each node listed, lineages 1 and 2 were identical in all three trees and had support values over 50% in all of the trees. An asterisk indicates groups of repeats that have identical amino acid sequences. Other identical repeats of group 1 proteins not listed in the table are (1) SCR 2 of FHR-1 and 2 of FHR-2 and (2) SCR 18 of FH and 3 of FHR-1.

Table 3
Clusters of Repeats from C4bp α , DAF, MCP, CR1, CR1L, and CR2 (group 2 proteins) Supported by Values Above 50%

SCRs IN LINEAGE 1	SCRs IN LINEAGE 2	SUPPORT (%)		
		NJBOOTAG	PHYLIP	PAUP*
SCR 1 of CR1	SCR 1 of CR1L	100	96	100
SCR 2 of CR1	SCR 2 of CR1L	100	100	100
SCRs 2 of CR1, 2 of CR1L	SCRs 9, 16 of CR1, 7 of CR1L	100	98	84
SCRs 4, 11, and 18 of CR1*	SCR 4 of CR1L	100	100	100
SCRs 5 and 12 of CR1*	SCR 5 of CR1L	92	93	100
SCRs 5 and 12 of CR1, 5 of CR1L	SCRs 19 and 26 of CR1	100	100	100
SCRs 6 and 13 of CR1*	SCR 6 of CR1L	81	67	100
SCRs 6 and 13 of CR1, 6 of CR1L	SCRs 20 and 27 of CR1	100	100	100
SCRs 7 and 14 of CR1*	SCRs 21 and 28 of CR1	96	88	100
SCRs 9 and 16 of CR1	SCR 7 of CR1L	100	100	100
SCR 19 of CR1	SCR 26 of CR1	100	99	100
SCR 20 of CR1	SCR 27 of CR1	77	85	79
SCR 21 of CR1	SCR 28 of CR1	100	98	100
SCR 23 of CR1	SCRs 2, 9, and 16 of CR1, 2, and 7 of CR1L	83	87	100
SCR 24 of CR1	SCRs 3, 10, and 17 of CR1, 3 of CR1L	100	91	100
SCR 25 of CR1	SCRs 4, 11, and 18 of CR1, 4 of CR1L	98	96	100
SCR 29 of CR1	SCR 14 of CR2	54	90	100
SCR 30 of CR1	SCR 15 of CR2	100	99	100
SCR 4 of CR2	SCRs 6, 13, 20, and 27 of CR1, 6 of CR1L	90	87	100
SCR 6 of CR2	SCR 10 of CR2	88	77	100
SCR 8 of CR2	SCR 11 of CR2	81	93	100
SCR 2 of DAF	SCRs 1, 8, 15, and 22 of CR1, 1 of CR1L	62	83	100
SCR 4 of MCP	SCR 4 of C4bp α	64	52	100

NOTE.—Shown are the support values for group 2 proteins in the three trees inferred from 132 individual short consensus repeats (SCRs). Notations are the same as in table 2. Other identical repeats of group 2 proteins not listed in the table are SCR 8 and 15 of CR1.

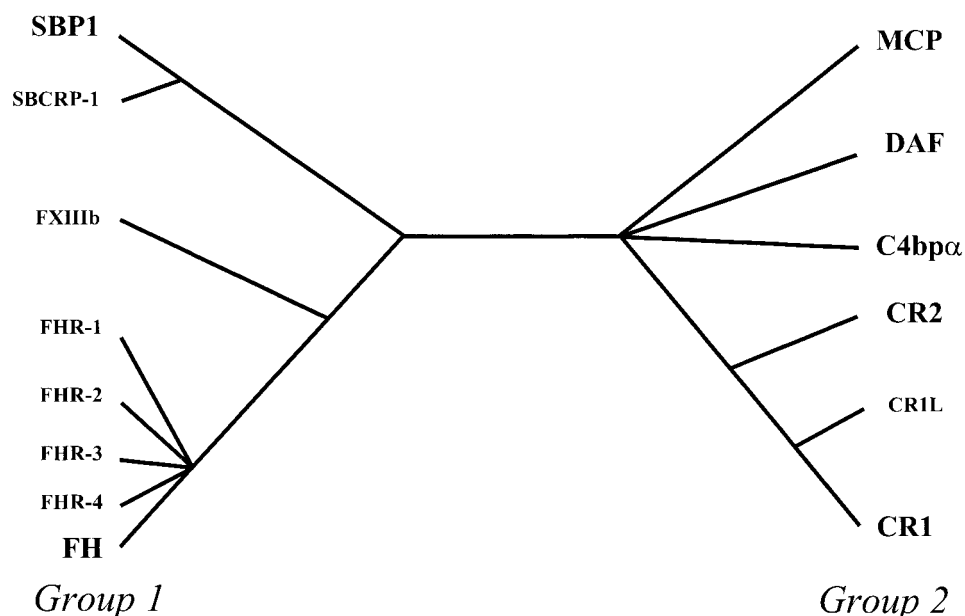


FIG. 4.—Clustering of regulator of complement activation (RCA) proteins based on the homology of their protein sequences. This summary unrooted tree reflects the homology relationships among RCA proteins based on similarity among their individual short consensus repeats (SCRs) inferred from the neighbor-joining and parsimony trees. Proteins more similar to one another are clustered together. Branch length is not proportional to degree of divergence among proteins. Names of proteins known to participate in complement regulation are shown in a larger font. C4bpβ is not included, for two of its three repeats show higher similarity to non-RCA proteins than to RCA proteins.

for these proteins are closely linked and located within 650 kb–2.2 Mb (Rey-Campos, Baeza-Sanz, and Rodríguez de Córdoba 1990; Skerka et al. 1995; Heine-Suñer et al. 1998; Weiss and Cannich 1998). These proteins express the greatest structural similarity to sand bass proteins SBP1 and SBCRP-1.

Group 2 includes the genes located within the 900-kb region which code for proteins C4bpα, MCP, DAF, complement receptors CR1 and CR2, and CR1L (Rey-Campos, Baeza-Sanz, and Rodríguez de Córdoba 1990; Pardo-Manuel de Villena and Rodríguez de Córdoba

1995). In contrast to the high support shown in tables 2 and 3 for clustering of SCRs within group 1 and group 2 proteins, clustering between repeats from groups 1 and 2 had very low support values (data not shown).

The C4bpβ gene does not fall within either group. Although the C4bpβ gene is located near the C4bpα gene (Pardo-Manuel et al. 1990), only its SCR 2 shows a close similarity to SCRs from group 2 proteins. It is similar to repeat 30 of CR1, repeat 15 of CR2, and other related repeats from group 2 proteins, all of which are similar to SCR 4 of C4bpα (see below). SCRs 1 and 3

Table 4
Proportion (%) of Short Consensus Repeats (SCRs) in Each Sequence Supporting the Summary Tree in Figure 4

PROTEIN	SYMBOL	PROPORTION OF SCRs SUPPORTING THE SUMMARY TREE ^a		
		NJBOOTAG	PHYLIP	PAUP
C4-binding protein, α chain ^b	C4bpα	50% (4)	75% (6)	63% (5)
Factor H	FH	90% (18)	95% (19)	75% (15)
Factor H-related protein 1	FHR-1	80% (4)	100% (5)	80% (4)
Factor H-related protein 2	FHR-2	75% (3)	100% (4)	75% (3)
Factor H-related protein 3	FHR-3	100% (5)	100% (5)	100% (5)
Factor H-related protein 4	FHR-4	100% (5)	100% (5)	100% (5)
Complement receptor type 1	CR1	100% (30)	100% (30)	100% (30)
CR1-like protein	CR1L	100% (7)	100% (7)	100% (7)
Complement receptor type 2	CR2	93% (14)	80% (12)	93% (14)
Decay accelerating factor	DAF	100% (4)	50% (2)	100% (4)
Membrane cofactor protein	MCP	75% (3)	75% (3)	75% (3)
Factor XIII, b subunit	FXIIIb	90% (9)	100% (10)	100% (10)
Sand bass cofactor protein	SBP1	82% (14)	76% (4)	82% (14)
Sand bass cofactor related protein 1	SBCRP-1	100% (3)	100% (3)	100% (3)

^a The proportion of supporting repeats is shown as a percentage, followed by the number of supporting repeats in parentheses. Symbols for the tree inference methods are the same as in table 2.

^b The first four N-terminal repeats (SCRs 1–4) of C4bpα support the summary tree for each method.

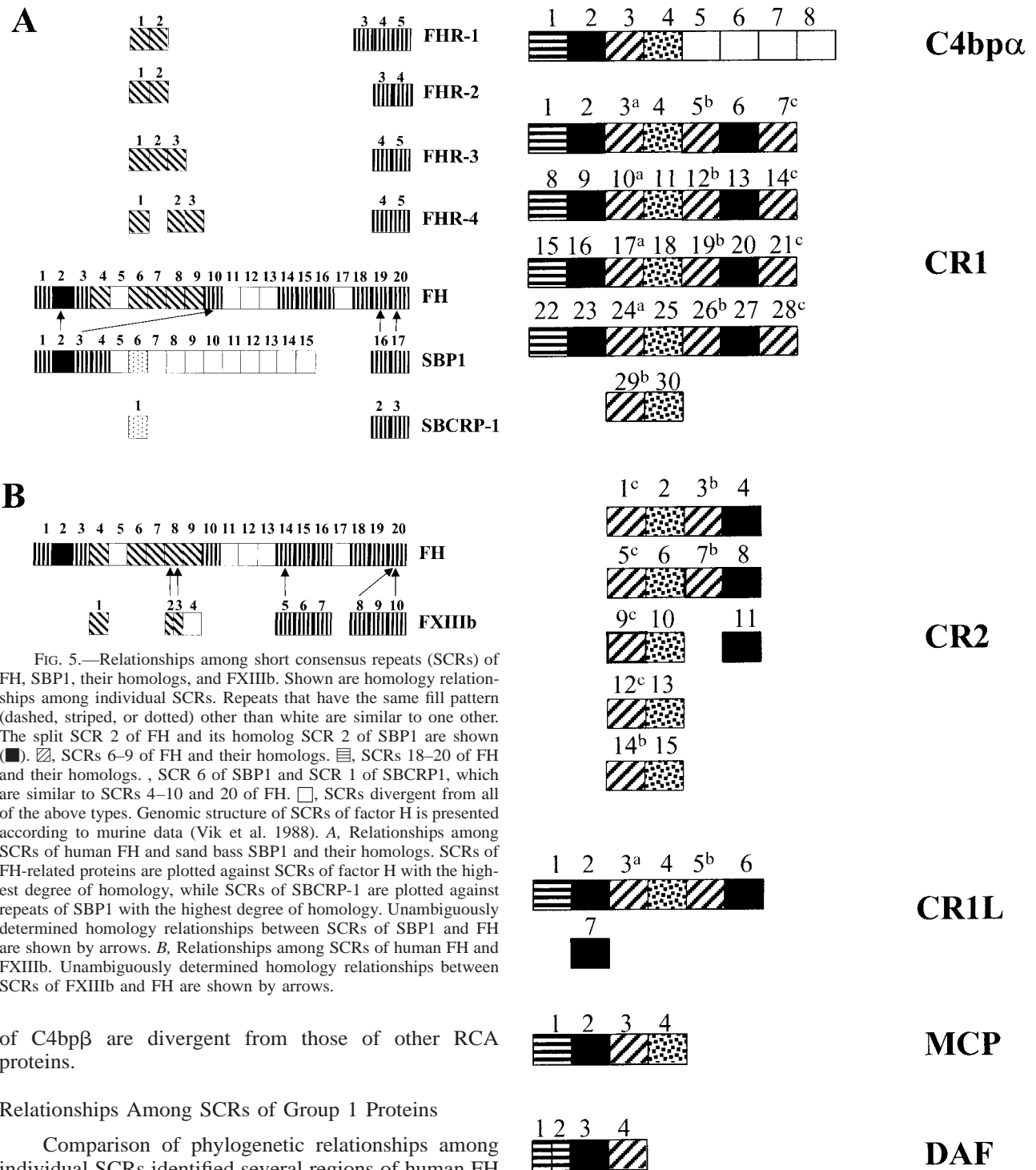


FIG. 5.—Relationships among short consensus repeats (SCRs) of FH, SBP1, their homologs, and FXIIIb. Shown are homology relationships among individual SCRs. Repeats that have the same fill pattern (dashed, striped, or dotted) other than white are similar to one other. The split SCR 2 of FH and its homolog SCR 2 of SBP1 are shown (■). ▨, SCRs 6–9 of FH and their homologs. ▩, SCRs 18–20 of FH and their homologs. □, SCR 6 of SBP1 and SCR 1 of SBCRP1, which are similar to SCRs 4–10 and 20 of FH. ◻, SCRs divergent from all of the above types. Genomic structure of SCRs of factor H is presented according to murine data (Vik et al. 1988). A, Relationships among SCRs of human FH and sand bass SBP1 and their homologs. SCRs of FH-related proteins are plotted against SCRs of factor H with the highest degree of homology, while SCRs of SBCRP-1 are plotted against repeats of SBP1 with the highest degree of homology. Unambiguously determined homology relationships between SCRs of SBP1 and FH are shown by arrows. B, Relationships among SCRs of human FH and FXIIIb. Unambiguously determined homology relationships between SCRs of FXIIIb and FH are shown by arrows.

of C4bpβ are divergent from those of other RCA proteins.

Relationships Among SCRs of Group 1 Proteins

Comparison of phylogenetic relationships among individual SCRs identified several regions of human FH that are homologous to FH-related proteins FHR-1, FHR-2, FHR-3, and FHR-4, sand bass proteins SBP1 and SBCRP-1, and human coagulation factor FXIIIb (fig. 5). Based on phylogenetic clustering, we identified four distinct subtypes of SCRs: those similar to (1) repeat 2 of FH, (2) SCRs 6–9 and related repeat 4, (3) SCRs 18–20 of FH and more distantly related repeats 1, 3, 10, and 14–16, or (4) SCRs divergent from all three of the above types.

Human FH-related proteins FHR-1, FHR-2, FHR-3, and FHR-4 seem to have diverged very recently from

FIG. 6.—Relationships among short consensus repeats (SCRs) of C4bpα, CR1, CR1L, CR2, MCP, and DAF. SCRs shown with the same fill pattern other than white are similar to one another. Split SCRs are shown (■). SCR pairs encoded by fused exons are shown in bold frames. a, b, and c indicate three types of repeats in CR1, CR1L, and CR2 that are homologous to SCR 3 of C4bpα. A 16-SCR variant of CR2 contains an additional SCR between SCRs 10 and 11 (Fujisaki et al. 1989).

the regions that correspond to SCRs 6–9 and 18–20 of FH. The support for these clusterings was very high, in most cases above 90% (table 2). These regions are also duplicated in murine FH-related proteins (Zipfel and Skerka 1994).

Among SCRs of SBP1, repeat 2 seems to be closely related to repeat 2 of FH, and their clustering is supported by 88%–100% (table 2). SCR 16 is similar to SCR 19 of FH (44%–100% support), while repeat 17 is most similar to repeat 20 of FH (86%–100% support). Repeat 3 is similar to repeat 10 of FH and distantly related to repeats 18–20 of FH. Several other more divergent SCRs of SBP1 were likely to have a common origin with repeats 6–9 or 18–20 of FH (data not shown). SCR 1 is similar to both repeat 1 and repeat 3 of FH. Repeat 4 is similar to both repeat 10 and repeat 20 of factor H, and repeat 6 is similar to repeats 4–10 and 20 (fig. 5).

As noted previously (Zipfel et al. 1996; Krushkal, Kemper, and Gigli 1998), sand bass proteins SBCRP-1 and SBP1 diverged very recently (fig. 4). The three SCRs of SBCRP-1 are closely related to repeats 6, 16, and 17 of SBP1, respectively (82%–100% support; table 2).

Comparison of the SCRs of coagulation factor FXIIIb and FH again shows groups of repeats that are likely to be related to repeats 6–9 and 18–20 of FH (fig. 5). SCRs 2 and 3 are likely to have a common origin with repeat 8 of FH, while repeats 8 and 10 clustered with 20 of FH. Repeat 1 is similar to repeats 4 and 6–9 of FH, and repeats 5–7 and 9 are related to repeats 14–16 of FH (data not shown). Finally, SCR 4 of FXIIIb was clustered with repeats 3, 4, and 6–9 of FH.

The structural homology among proteins shown in figure 5 is in agreement with the exon/intron structure of their genes. All known functional regulators of complement activation have one or more SCRs encoded by two exons (“split” SCRs), with a splice site after the second nucleotide within the Gly34 codon of the SCR consensus sequence (Hillarp et al. 1993). In murine FH, SCR 2 is the only repeat encoded by two exons (Vik et al. 1988). In contrast, sequence similarity in FH-related proteins FHR-1, FHR-2, FHR-3, and FHR-4 and coagulation factor FXIIIb shows the absence of such split SCRs. This observation has been experimentally confirmed for the FHR-2 (Skerka et al. 1995) and FXIIIb genes (Bottenus, Ichinose, and Davie 1990).

Relationships Among SCRs of Group 2 Proteins

MCP, DAF, CR1, and CR1L share structural similarity to repeats 1–4 of C4bp α (fig. 6). Table 3 lists tree nodes with over 50% support for such clustering of SCRs. DAF has a duplication of repeat 1 and lacks repeat 4 of C4bp α . The 30-SCR variant of CR1 has four tandem long homologous repeats (LHRs) similar to one another. Each LHR consists of seven SCRs closely related to SCRs 1, 2, 3, 4, 3, 2, and 3 of C4bp α , respectively. In each LHR, there are three SCRs related to SCR 3 of C4bp α . These repeats are further classified as type a, b, or c according to their degree of similarity (fig. 6).

SCRs 29 and 30 of CR1 are similar to SCRs 3 and 4 of C4bp α . CR1L consists of 7 SCRs, each similar to SCRs within an LHR of CR1, except for SCR 7, which is similar to SCR 2 of both CR1 and C4bp α (table 3).

The sequence homology of these proteins is in agreement with their gene structure. SCR 2 of C4bp α , SCR 2 of MCP, SCR 3 of DAF, SCRs 2, 6, 9, 13, 16, 20, 23, and 27 of CR1, and SCRs 2, 6, and 7 of CR1L are each encoded by two exons (Hourcade et al. 1990; Post et al. 1990; Aso et al. 1991; Post et al. 1991; Rodriguez de Cordoba, Sanchez-Corral, and Rey-Campos 1991; Vik and Wong 1993). In addition, CR1 and CR1L genes have several “fused exons” that encode two SCRs each: SCRs 3–4, 10–11, 17–18, and 24–25 of CR1, and SCRs 3–4 of CR1L (Hourcade et al. 1990; Vik and Wong 1993) (fig. 6).

CR2 has a more complex structure. The first 11 SCRs of CR2 are organized in three tandem 4-SCR arrays similar to SCRs 3, 4, 3, and 2 of C4bp α , respectively (fig. 6). In addition, the two pairs of SCRs 12–13 and 14–15 are similar to SCRs 3 and 4 of C4bp α . SCRs 4, 8, and 11 are encoded by two exons each, while SCRs 1–2, 5–6, 9–10, 12–13 and 14–15 are the result of fused exons (Fujisaku et al. 1989). Similarities between individual repeats of complement receptors CR1 and CR2 (fig. 6) suggest that the evolution of these proteins was complex and resulted from a number of crossover events. In particular, while SCRs 1–28 of CR1 are arranged so that SCRs of types a, b, and c (similar to SCR 3 of C4bp α) maintain their position in each LHR, SCRs 29 and 30 of CR1 and SCRs 1–15 of CR2 show a complex order of occurrence of b and c type repeats.

Discussion

The RCA genes are suggested to have a distinct evolutionary origin separate from other SCR-containing proteins that do not participate in the regulation of complement activation. The RCA gene cluster is likely to have evolved as a result of a series of gene duplication events (Reid et al. 1986; Farries and Atkinson 1991). This separate phylogenetic position of the RCA members is confirmed by the divergence of their sequences from other SCR-containing proteins that do not participate in complement regulation (unpublished data).

The RCA proteins fall into two distinct groups on the basis of their sequence homology (fig. 4). This division is independent of the functional activities of the proteins. For example, the two human plasma complement regulators, FH and C4bp α , belong to different groups. FH is more closely related to the blood-clotting protein FXIIIb than to many complement regulators such as C4bp α , CR1, MCP, and DAF.

From the sequence data used in this study, it was not possible to reach a conclusion as to when the divergence between the group 1 and group 2 RCA members occurred. Sequence similarity between human factor H, sand bass SBP1, and the *C. elegans* FH-like protein F36H2.3 suggest that FH, an alternative pathway member, is old in evolutionary history (Krushkal, Kemper, and Gigli 1998). Although individual repeats of hu-

man C4bp α are distinct from those of FH, SBP1 and F36H2.3, as shown by the phylogenetic tree inference, the average amino acid sequence divergence between C4bp α and FH is lower than that between C4bp α and SBP1 (Krushkal, Kemper, and Gigli 1998). Assuming that SBP1 is the single precursor of both plasma complement regulators, FH and C4bp α , the divergence between FH and C4bp α may have occurred in evolutionary history after the separation of the fish lineage. However, one cannot exclude a possibility that the split between the group 1 and group 2 proteins may have occurred before the appearance of the fish lineage and that a C4bp α -like protein in fish has yet to be identified.

The last four SCRs of C4bp α do not have close sequence similarity to SCRs of the group 2 proteins CR1, CR2, CR1L, MCP, or DAF (fig. 6). They are also divergent from SCRs of the group 1 proteins. There are two possible explanations for their divergent nature. One is that the ancestral precursor of the group 2 proteins was a C4bp α -like molecule that contained eight divergent SCRs. As a result of a series of gene duplications and unequal-crossing-over events, only the first four SCRs were propagated to give rise to MCP, DAF, CR1, CR1L, and CR2. It is also possible that the precursor of the group 2 proteins was an MCP-like molecule with only four SCRs. C4bp α may have appeared after unequal crossing over, resulting in the merging of the four SCRs of such a precursor with the four repeats of some distant SCR-containing protein.

C4bp β is distinct from both groups of the RCA proteins. Its gene may have a mosaic structure, originating as a result of an unequal crossing over between different SCR-containing genes. However, the location of the C4bp β gene within the RCA cluster in tight linkage with the C4bp α gene (Pardo-Manuel et al. 1990), as well as the presence of a split SCR 3 within C4bp β (Hillarp et al. 1993), suggests a common origin of this gene and other RCA members. Either the C4bp β gene may have appeared very early in the evolution of the RCA cluster or its sequence may have diverged from other RCA members due to diversifying selection.

The evolutionary relationships among individual SCRs may explain the functional similarities and differences expressed by RCA members within groups 1 and 2. SCRs 1–4 or 1–5 of FH are necessary for its cofactor and decay accelerating activities (Gordon et al. 1995; Kuhn and Zipfel 1996; Sharma and Pangburn 1996), while C3b binding is mediated by three sites: repeats 1–4, 6–10, and 16–20 (Sharma and Pangburn 1996). Since the N-terminal SCRs are not present in the FH-related proteins (fig. 5), it should be expected that FHR-1, FHR-2, FHR-3, and FHR-4 lack the cofactor activity but can bind C3b. The results of this theoretical analysis are in agreement with published experimental data, which show a lack of cofactor activity in FHR-1 and FHR-2 (Timmann, Leippe, and Horstmann 1991) and the existence of C3b binding by FHR-4 (Hellwage, Skerka, and Zipfel 1997).

FH and SBP1 are the only two proteins of group 1 with homology at the N-terminus; however, the functional significance of this structural finding has not yet

been elucidated. In agreement with the sequence analysis results, these proteins both display cofactor activity, decay accelerating activity, and C3b binding (Pangburn, Schreiber, and Müller-Eberhard 1977; Kaidoh and Gigli 1989). The N-terminal repeats of SBP1 are important for binding and cofactor activity for both human C4b and trout C3b (Kemper, Zipfel, and Gigli 1998). The observation that SCR 2 of SBP1 is very similar to the split SCR 2 of FH suggests their functional importance.

The functional role of SCRs in the group 2 proteins also correlates with their phylogenetic clustering and shows the importance of the N-terminal repeats, including a split SCR (repeat 2 of C4bp α , fig. 6). Such repeats are essential in the binding of human C4bp α to C4b (Dahlbäck, Smith, and Müller-Eberhard 1983). In murine C4bp α , the binding activity has been localized to SCRs 1–3 (Ogata et al. 1993). SCRs 2–4 of DAF, corresponding to repeats 1–3 of C4bp α , are functionally important (Coyne et al. 1992; Brodbeck et al. 1996). Human CR1 expresses multiple functional sites located in SCRs 1–4, 8–10, and 15–18, all of which are related to the N-terminal repeats of C4bp α (Klickstein et al. 1988; Krysch et al. 1994; Reilly et al. 1994). Binding of iC3b/C3dg to CR2 is affected by SCRs 1 and 2 (Lowell et al. 1989; Molina et al. 1995). All four SCRs of human MCP play a functional role in either cofactor activity or binding of C3b (Adams et al. 1991; Iwata et al. 1995). Therefore, similar to group 1 proteins, repeats in group 2 proteins that are related to SCRs 1–4 of C4bp α often participate in complement regulation, and the split SCR 2 is functionally important in both groups. In addition, the order of repeats also seems to play a role, since other SCRs similar to SCR 2 (split SCRs 6, 13, 20, and 27 of CR1 and 4, 8, and 11 of CR2) are not involved in complement regulation.

In summary, despite the long evolutionary history of the RCA proteins, which originated at least as early as the appearance of bony fish, a remarkable conservation of the N-terminal SCRs is observed within the group 1 and group 2 proteins, which carry a functional activity in complement interactions. Proteins lacking such N-terminal repeats seem to have lost their ability to participate in complement regulation. Future identification and analysis of complement regulators in fish, amphibians, reptiles, and mammals, as well as a detailed study of their more ancient homologs, may provide further insight into the origin and evolution of the group 1 and group 2 proteins.

Supplementary Material

The supplementary material (phylogenetic trees, distance matrices and sequence alignments) can be viewed via a World Wide Web page at http://www.uth.tmc.edu/uth_orgs/imm/centers/immunology.htm.

Acknowledgments

The authors are grateful to the two anonymous reviewers for their helpful suggestions and comments. This study was supported by grant AI20067 from the

National Institutes of Health. J.K. was supported in part by the Minnie L. Maffett Fellowship from the Minnie L. Maffett Fellowship Fund.

LITERATURE CITED

- ADAMS, E. M., M. C. BROWN, M. NUNGE, M. KRYCH, and J. P. ATKINSON. 1991. Contribution of the repeating domains of membrane cofactor protein (CD46) of the complement system to ligand binding and cofactor activity. *J. Immunol.* **147**:3005–3011.
- AINSCOUGH, R., S. BARDILL, K. BARLOW et al. (404 co-authors). 1998. Genome sequence of the nematode *C. elegans*: a platform for investigating biology. *Science* **282**:2012–2018.
- ASO, T., S. OKAMURA, T. MATSUGUCHI, N. SAKAMOTO, T. SATA, and Y. NIHO. 1991. Genomic organization of the α chain of the human C4b-binding protein gene. *Biochem. Biophys. Res. Commun.* **174**:222–227.
- BIRMINGHAM, D. J., C. M. LOGAR, X.-P. SHEN, and W. CHEN. 1996. The baboon erythrocyte complement receptor is a glycosphatidylinositol-linked protein encoded by a homologue of the human CR1-like genetic element. *J. Immunol.* **157**:2586–2592.
- BLUMBACH, B., Z. PANCER, B. DIEHL-SEIFERT, R. STEFFEN, J. MUNKNER, I. MULLER, and W. E. MULLER. 1998. The putative sponge aggregation receptor. *J. Cell Sci.* **111**:2635–2644.
- BOTTENUS, R. E., A. ICHINOSE, and E. W. DAVIE. 1990. Nucleotide sequence of the gene for the b subunit of human factor XIII. *Biochemistry* **29**:11195–11209.
- BRODBECK, W. G., D. LIU, J. SPERRY, C. MOLD, and M. E. MEDOF. 1996. Localization of classical and alternative pathway regulatory activity within the decay-accelerating factor. *J. Immunol.* **156**:2528–2533.
- CHUNG, S. I., M. S. LEWIS, and J. E. FOLK. 1974. Relationships of the catalytic properties of human plasma and platelet transglutaminases (activated blood coagulation factor XIII) to their subunit structures. *J. Biol. Chem.* **249**:940–950.
- COYNE, K. E., S. E. HALL, S. THOMPSON, M. A. ARCE, T. KINOSHITA, T. FUJITA, D. J. ANSTEE, W. ROSSE, and D. M. LUBLIN. 1992. Mapping of epitopes, glycosylation sites, and complement regulatory domains in human decay accelerating factor. *J. Immunol.* **149**:2906–2913.
- DAHLBÄCK, B., C. A. SMITH, and H. J. MÜLLER-EBERHARD. 1983. Visualization of human C4b-binding protein and its complexes with vitamin K-dependent protein S and complement protein C4b. *Proc. Natl. Acad. Sci. USA* **80**:3461–3465.
- DAHMEN, A., T. KAIDOH, P. F. ZIPFEL, and I. GIGLI. 1994. Cloning and characterization of a cDNA representing a putative complement-regulatory plasma protein from barred sand bass (*Parablax nebulifer*). *Biochem. J.* **301**:391–397.
- DAYHOFF, M. O. 1979. Atlas of protein sequence and structure. Vol. 5, Suppl. 3. National Biomedical Research Foundation, Washington, D.C.
- FARRIES, T. C., and J. P. ATKINSON. 1991. Evolution of the complement system. *Immunol. Today* **12**:295–300.
- FEARON, D. T. 1979. Regulation of the amplification C3 convertase of human complement by an inhibitory protein isolated from human erythrocyte membrane. *Proc. Natl. Acad. Sci. USA* **76**:5867–5871.
- FELSENSTEIN, J. 1989. PHYLIP—phylogeny inference package (version 3.2). *Cladistics* **5**:164–166.
- FUJISAKU, A., J. B. HARLEY, M. B. FRANK, B. A. GRUNER, B. FRAZIER, and V. M. HOLERS. 1989. Genomic organization and polymorphisms of the human C3d/Epstein-Barr virus receptor. *J. Biol. Chem.* **264**:2118–2125.
- FUJITA, T., T. INOUE, K. OGAWA, K. IIDA, and N. TAMURA. 1987. The mechanism of action of decay-accelerating factor (DAF). DAF inhibits the assembly of C3 convertases by dissociating C2a and Bb. *J. Exp. Med.* **166**:1221–1228.
- GIGLI, I., T. FUJITA, and V. NUSSENZWEIG. 1979. Modulation of the classical pathway C3 convertase by plasma proteins C4-binding protein and C3b inactivator. *Proc. Natl. Acad. Sci. USA* **76**:6596–6600.
- GORDON, D. L., R. M. KAUFMAN, T. K. BLACKMORE, J. KWONG, and D. M. LUBLIN. 1995. Identification of complement regulatory domains in human factor H. *J. Immunol.* **155**:348–356.
- HEINE-SUNER, D., M. A. DÍAZ-GUILLÉN, F. P.-M. DE VILLENA, M. ROBLEDO, J. BENÍTEZ, and S. RODRÍGUEZ DE CÓRDOBA. 1997. A high-resolution map of the regulator of the complement activation gene cluster on 1q32 that integrates new genes and markers. *Immunogenetics* **45**:422–427.
- HEINE-SUNER, D., M. A. DÍAZ-GUILLÉN, P. SÁNCHEZ-CORRAL, and S. RODRÍGUEZ DE CÓRDOBA. 1998. An integrated map of the human RCA gene cluster that positions 85 genes and polymorphic markers. *Mol. Immunol.* **35**:395.
- HELLWAGE, J., C. SKERKA, and P. F. ZIPFEL. 1997. Biochemical and functional characterization of the factor-H-related protein 4 (FHR-4). *Immunopharmacology* **38**:149–157.
- HILLARP, A., and B. DAHLBÄCK. 1988. Novel subunit in C4b-binding protein required for protein S binding. *J. Biol. Chem.* **263**:12759–12764.
- HILLARP, A., F. PARDO-MANUEL, R. R. RUIZ, S. RODRIGUEZ DE CORDOBA, and B. DAHLBÄCK. 1993. The human C4b-binding protein β -chain gene. *J. Biol. Chem.* **268**:15017–15023.
- HOSHINO, M., F. MATSUZAKI, Y.-I. NABESHIMA, and C. HAMA. 1993. *hikaru genki*, a CNS-specific gene identified by abnormal locomotion in *Drosophila*, encodes a novel type of protein. *Neuron* **10**:395–407.
- HOURCADE, D., A. D. GARCIA, T. W. POST, P. TAILLON-MILLER, V. M. HOLERS, L. M. WAGNER, N. S. BORA, and J. P. ATKINSON. 1992. Analysis of the human regulators of complement activation (RCA) gene cluster with yeast artificial chromosomes (YACs). *Genomics* **12**:289–300.
- HOURCADE, D., V. M. HOLERS, and J. P. ATKINSON. 1989. The regulators of complement activation (RCA) gene cluster. *Adv. Immunol.* **45**:381–416.
- HOURCADE, D., D. R. MIESNER, C. BEE, W. ZELDES, and J. P. ATKINSON. 1990. Duplication and divergence of the amino-terminal coding region of the complement receptor 1 (CR1) gene. *J. Biol. Chem.* **265**:974–980.
- IIDA, K., and V. NUSSENZWEIG. 1981. Complement receptor is an inhibitor of the complement cascade. *J. Exp. Med.* **153**:1138–1150.
- IWATA, K., T. SEYA, Y. YANAGI, J. M. PESANDO, P. M. JOHNSON, M. OKABE, S. UEDA, H. ARIGA, and S. NAGASAWA. 1995. Diversity of sites for measles virus binding and for inactivation of complement C3b and C4b on membrane cofactor protein CD46. *J. Biol. Chem.* **270**:15148–15152.
- KAIDOH, T., and I. GIGLI. 1987. Phylogeny of C4b-C3b cleaving activity: similar fragmentation patterns of human C4b and C3b produced by lower animals. *J. Immunol.* **139**:194–201.
- . 1989. Phylogeny of regulatory proteins of the complement system: isolation and characterization of a C4b/C3b inhibitor and a cofactor from sand bass plasma. *J. Immunol.* **142**:1605–1613.
- KEMPER, C., P. F. ZIPFEL, and I. GIGLI. 1998. The complement cofactor protein (SBP1) from the barred sand bass (*Parablax nebulifer*) mediates overlapping regulatory activities

- of both human C4b binding protein and factor H. *J. Biol. Chem.* **273**:19398–19404.
- KLICKSTEIN, L. B., T. J. BARTOW, V. MILETIC, L. D. RABSON, J. A. SMITH, and D. T. FEARON. 1988. Identification of distinct C3b and C4b recognition sites in the human C3b/C4b receptor (CR1, CD35) by deletion mutagenesis. *J. Exp. Med.* **168**:1699–1717.
- KRUSHKAL, J., C. KEMPER, and I. GIGLI. 1998. Ancient origin of human complement factor H. *J. Mol. Evol.* **47**:625–630.
- KRYSCH, M., L. CLEMENZA, D. HOWDESHHELL, R. HAUHART, D. HOURCADE, and J. P. ATKINSON. 1994. Analysis of the functional domains of complement receptor type 1 (C3b/C4b receptor; CD35) by substitution mutagenesis. *J. Biol. Chem.* **269**:13273–13278.
- KUHN, S., and P. F. ZIPFEL. 1996. Mapping of the domains required for decay acceleration activity of the human factor H-like protein 1 and factor H. *Eur. J. Immunol.* **26**:2383–2387.
- LISZEWSKI, K. M., T. W. POST, and J. P. ATKINSON. 1991. Membrane cofactor protein (MCP or CD46): newest member of the regulators of complement activation gene cluster. *Annu. Rev. Immunol.* **9**:431–455.
- LOWELL, C. A., L. B. KLICKSTEIN, R. H. CARTER, J. A. MITCHELL, D. T. FEARON, and J. M. AHEARN. 1989. Mapping of the Epstein-Barr virus and C3dg binding sites to a common domain on complement receptor type 2. *J. Exp. Med.* **170**:1931–1946.
- MEDOF, M. E., T. KINOSHITA, and V. NUSSENZWEIG. 1984. Inhibition of complement activation on the surface of cells after incorporation of decay-accelerating factor (DAF) into their membranes. *J. Exp. Med.* **160**:1558–1578.
- MISASI, R., H. P. HUEMER, W. SCHWAEBLE, E. SÖLDER, C. LARCHER, and M. P. DIERICH. 1989. Human complement factor H: an additional gene product of 43 kDa isolated from human plasma shows cofactor activity for the cleavage of the third component of complement. *Eur. J. Immunol.* **19**:1765–1768.
- MOLINA, H., S. J. PERKINS, J. GUTHRIDGE, J. GORKA, T. KINOSHITA, and V. M. HOLERS. 1995. Characterization of a complement receptor 2 (CR2, CD21) ligand binding site for C3. An initial model of ligand interaction with two linked short consensus repeat modules. *J. Immunol.* **154**:5426–5435.
- MÜLLER-EBERHARD, H. J. 1985. Introduction and overview. Pp. 1–5 in H. J. MÜLLER-EBERHARD and P. A. MIESCHER, eds. *Complement*. Springer, Berlin.
- NICHOLSON-WELLER, A., J. BURGE, D. T. FEARON, P. F. WELLS, and K. F. AUSTEN. 1982. Isolation of a human erythrocyte membrane glycoprotein with decay-accelerating activity for C3 convertases of the complement system. *J. Immunol.* **129**:184–189.
- OGATA, R. T., P. MATHIAS, B. M. BRADT, and N. R. COOPER. 1993. Murine C4b-binding protein. Mapping of the ligand binding site and the N-terminus of the pre-protein. *J. Immunol.* **150**:2273–2280.
- OTA, T., and M. NEI. 1994. Estimation of the number of amino acid substitutions per site when the substitution rate varies among sites. *J. Mol. Evol.* **38**:642–643.
- PAGE, R. D. M. 1996. TREEVIEW: an application to display phylogenetic trees on personal computers. *Comput. Appl. Biosci.* **12**:357–358.
- PAHLER, S., B. BLUMBACH, I. MULLER, and W. E. MULLER. 1998. Putative multiadhesive protein from marine sponge *Geodia cydonium*: cloning of the cDNA encoding a fibronectin-, and SRCR-, and a complement control protein module. *J. Exp. Zool.* **282**:332–343.
- PANGBURN, M. K., and H. J. MÜLLER-EBERHARD. 1985. The alternative pathway of the complement. Pp. 185–213 in H. J. MÜLLER-EBERHARD and P. A. MIESCHER, eds. *Complement*. Springer, Berlin.
- PANGBURN, M. K., R. D. SCHREIBER, and H. J. MÜLLER-EBERHARD. 1977. Human complement C3b inactivator: isolation, characterization, and demonstration of an absolute requirement for the serum protein β_1H for cleavage of C3b and C4b in solution. *J. Exp. Med.* **146**:257–270.
- . 1983. Deficiency of an erythrocyte membrane protein with complement regulatory activity in paroxysmal nocturnal hemoglobinuria. *Proc. Natl. Acad. Sci. USA* **80**:5430–5434.
- PARDO-MANUEL, F., J. REY-CAMPOS, A. HILLARP, B. DAHLBACK, and S. RODRÍGUEZ DE CÓRDOBA. 1990. Human genes for the α and β chains of the complement C4b-binding protein are closely linked in head-to-tail arrangement. *Proc. Natl. Acad. Sci. USA* **87**:4529–4532.
- PARDO-MANUEL DE VILLENA, F., and S. RODRÍGUEZ DE CÓRDOBA. 1995. *C4BPAL2*: a second duplication of the *C4BPA* gene in the human *RCA* gene cluster. *Immunogenetics* **41**:139–143.
- POST, T. W., M. A. ARCE, M. K. LISZEWSKI, E. S. THOMPSON, J. P. ATKINSON, and D. M. LUBLIN. 1990. Structure of the gene for human complement protein decay accelerating factor. *J. Immunol.* **144**:740–744.
- POST, T. W., M. K. LISZEWSKI, E. M. ADAMS, I. TEDJA, E. A. MILLER, and J. P. ATKINSON. 1991. Membrane cofactor protein of the complement system: alternative splicing of serine/threonine/proline-rich exons and cytoplasmic tails produces multiple isoforms that correlate with protein phenotype. *J. Exp. Med.* **174**:93–102.
- REID, K. B. M., D. R. BENTLEY, R. D. CAMPBELL, L. P. CHUNG, R. B. SIM, T. KRINSTENSEN, and B. F. TACK. 1986. Complement system proteins which interact with C3b or C4b: a superfamily of structurally related proteins. *Immunol. Today* **7**:230–234.
- REID, K. B. M., and A. J. DAY. 1989. Structure-function relationships of the complement components. *Immunol. Today* **10**:177–180.
- REILLY, B. D., S. C. MAKRIDES, P. J. FORD, H. C. MARSH JR., and C. MOLD. 1994. Quantitative analysis of C4b dimer binding to distinct sites on the C3b/C4b receptor (CR1). *J. Biol. Chem.* **269**:7696–7701.
- REY-CAMPOS, J., D. BAEZA-SANZ, and S. RODRÍGUEZ DE CÓRDOBA. 1990. Physical linkage of the human genes coding for complement factor H and coagulation factor XIII B subunit. *Genomics* **7**:644–646.
- REY-CAMPOS, J., P. RUBINSTEIN, and S. RODRÍGUEZ DE CÓRDOBA. 1988. A physical map of the human regulator of complement activation gene cluster linking the complement genes *CR1*, *CR2*, *DAF*, and *C4BP*. *J. Exp. Med.* **167**:664–669.
- RODRÍGUEZ DE CÓRDOBA, S., P. SANCHEZ-CORRAL, and J. REY-CAMPOS. 1991. Structure of the gene coding for the alpha polypeptide chain of the human complement component C4b-binding protein. *J. Exp. Med.* **173**:1073–1082.
- ROSS, G. D., M. J. POLLEY, E. M. RABELLINO, and H. M. GREY. 1973. Two different receptors on human lymphocytes: one specific for C3b and one specific for C3b inactivator-cleaved C3b. *J. Exp. Med.* **138**:798–811.
- SAITOU, N., and M. NEI. 1987. The neighbor-joining method: a new method for reconstructing phylogenetic trees. *Mol. Biol. Evol.* **4**:406–425.
- SEEGER, A., W. E. MAYER, and J. KLEIN. 1996. A complement factor B-like cDNA from zebrafish (*Brachydanio rerio*). *Mol. Immunol.* **33**:511–520.

- SEYA, T., and J. P. ATKINSON. 1989. Functional properties of membrane cofactor protein of complement. *Biochem. J.* **264**:581–588.
- SHARMA, A. K., and M. K. PANGBURN. 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.
- SKERKA, C., J. M. MOULDS, P. TAILLON-MILLER, D. HOURCADE, and P. F. ZIPFEL. 1995. The human factor H-related gene 2 (FHR2): structure and linkage to the coagulation factor XIIIb gene. *Immunogenetics* **42**:268–274.
- SMITH, L. C., K. AZUMNI, and M. NONAKA. 1999. Complement systems in invertebrates: the ancient alternative and lectin pathways. *Immunopharmacology* **42**:107–120.
- SUNYER, J. O., I. K. ZARKADIS, and J. D. LAMBRIS. 1998. Complement diversity: a mechanism for generating immune diversity? *Immunol. Today* **19**:519–523.
- SWOFFORD, D. L. 1998. Swofford: PAUP* 4.0 BETA for Windows. Sinauer, Sunderland, Mass.
- THOMPSON, J. D., D. G. HIGGINS, and T. J. GIBSON. 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.
- TIMMANN, C., M. LEIPPE, and R. D. HORSTMANN. 1991. Two major serum components antigenically related to complement factor H are different glycosylation forms of a single protein with no factor H-like complement regulatory functions. *J. Immunol.* **146**:1265–1270.
- TURNER, J. R. 1984. Structural and functional studies of the C3b and C4b binding proteins of a human monocyte-like cell line (U937). Masters thesis, Washington University, St. Louis, Mo.
- VIK, D. P., J. B. KEENEY, P. MUNOZ-CANOVEZ, D. D. CHAPLIN, and B. F. TACK. 1988. Structure of the murine complement factor H gene. *J. Biol. Chem.* **263**:16720–16724.
- VIK, D. P., and W. W. WONG. 1993. Structure of the gene for the F allele of complement receptor type 1 and sequence of the coding region unique to the S allele. *J. Immunol.* **151**:6214–6224.
- WEIGLE, W. O., M. G. GOODMAN, E. L. MORGAN, and T. H. HUGLI. 1985. Regulation of immune response by components of the complement cascade and their activated fragments. Pp. 323–344 in H. J. MÜLLER-EBERHARD and P. A. MIESCHER, eds. *Complement*. Springer, Berlin.
- WEIS, J. H., C. C. MORTON, G. A. P. BRUNS, J. J. WEIS, L. B. KLICKSTEIN, W. W. WONG, and D. T. FEARON. 1987. A complement receptor locus: genes encoding C3b/C4b receptor and C3d/Epstein Barr virus receptor map to 1q32. *J. Immunol.* **138**:312–315.
- WEISS, E. H., and A. CANNICH. 1998. The genomic organization of the human factor H gene family. *Mol. Immunol.* **35**:411.
- WHALEY, K., and S. RUDDY. 1976. Modulation of C3b hemolytic activity by a plasma protein distinct from C3b inactivator. *Science* **193**:1011–1013.
- WHALEY, K., and W. SCHWAEBLE. 1997. Complement and complement deficiencies. *Semin. Liver Dis.* **17**:297–310.
- ZIPFEL, P. F., C. KEMPER, A. DAHMEN, and I. GIGLI. 1996. Cloning and recombinant expression of a barred sand bass (*Parablax nebulifer*) cDNA: the encoded protein displays structural homology and immunological crossreactivity to human complement/cofactor related plasma proteins. *Dev. Comp. Immunol.* **20**:407–416.
- ZIPFEL, P. F., and C. SKERKA. 1994. Complement factor H and related proteins: an expanding family of complement-regulatory proteins? *Immunol. Today* **3**:121–126.

SHOZO YOKOYAMA, reviewing editor

Accepted July 21, 2000