STRUCTURE OF THE HUMAN B LYMPHOCYTE RECEPTOR FOR C3d AND THE EPSTEIN-BARR VIRUS AND RELATEDNESS TO OTHER MEMBERS OF THE FAMILY OF C3/C4 BINDING PROTEINS

By JANIS J. WEIS,*‡ LORRAINE E. TOOTHAKER,*‡ JOHN A. SMITH,§‖ JOHN H. WEIS,*‡ and DOUGLAS T. FEARON¶

*From the *Department of Rheumatology and Immunology, Brigham and Women's Hospital and the ‡Department of Medicine, Harvard Medical School, Boston, Massachusetts 02115; §the Departments of Molecular Biology and Pathology, Massachusetts General Hospital and the ‖Department of Pathology, Harvard Medical School, Boston, Massachusetts 02114; and the ¶Division of Molecular and Clinical Rheumatology, The Johns Hopkins University School of Medicine, Baltimore, Maryland 21205*

Complement receptor type 2 (CR2; cluster designation-21),[1] is a 145,000 M_r glycoprotein that is expressed by mature B lymphocytes (1–4), dendritic cells of the spleen (5), and pharyngeal epithelial cells (6). CR2 has specificity for the d region of the third component of complement, C3, which is most accessible in the breakdown fragments iC3b and C3dg (7, 8). CR2 enables B lymphocytes in areas distant from sites of inflammation to bind fragments of C3 in conjunction with immune complexes or bacterial particles.

The role of CR2 in the biology of the B lymphocyte is not well understood, although there are reports of growth stimulation of B lymphocytes through CR2. Certain monoclonal and polyclonal antibodies to CR2 stimulated T cell–dependent proliferation of nontransformed B lymphocytes (9–11), while others did not (12). Fragments of C3 stimulated growth of B lymphoblastoid cell lines in serum-free medium (13, 14), and human C3d, when crosslinked to Sepharose, maintained proliferation of dividing murine B lymphocytes, whereas soluble C3d was inhibitory (15). Crosslinking of CR2 and membrane IgM caused a synergistic increase in the free intracellular calcium concentration in human B lymphocytes (16). CR2 has been shown to cocap reciprocally with membrane IgM (17) and to be phosphorylated after crosslinking of membrane IgM (18, 19), further suggesting a close association of these two proteins in their transmembrane or cytoplasmic regions.

J. J. Weis is a Lucille P. Markey Scholar, and this work was supported in part by a grant from the Lucille P. Markey Charitable Trust. J. A. Smith is supported by a grant from Hoechst Aktiengesellschaft (Frankfurt am Main), Federal Republic of Germany. This work was also supported in part by National Institutes of Health grants AM-35907, AI-23401 (to the Dept. of Rheumatology and Immunology), AI-22833 (to D. T. Fearon), and AI-24158 (to J. H. Weis). It was also supported by American Cancer Society grant CD-45770 and by an American Heart Association grant-in-aid (both to H. H. Weis).

[1] *Abbreviations used in this paper:* C4bp, C4 binding protein; CR2, complement receptor type 2; DAF, decay-accelerating factor; SCR, short consensus repeat.

CR2 also serves as the receptor for the Epstein-Barr virus (EBV), mediating virus binding, internalization, and infection of B lymphocytes (20–24). The EBV membrane protein, gp350/220, has recently been identified as the viral protein that binds to CR2 (17, 25). CR2 directs the uptake and targeting of EBV to appropriate vacuoles for removal of the lipid envelope and release of the virus (26). CR2 also mediates internalization of inert particles coated with the EBV protein gp350/220 (17), demonstrating that the receptor has endocytic function when crosslinked with this ligand. The function of CR2 on dendritic cells and pharyngeal epithelial cells has not been studied. However, as pharyngeal epithelial cells are a site of EBV productive infection, CR2 may mediate viral internalization by this cell type (27).

Primary sequence analysis of the CR2 protein has demonstrated that it is a member of a gene family of complement regulatory proteins (28), including complement receptor type 1 (CR1), C4 binding protein (C4bp), factor H, and decay-accelerating factor (DAF) (29–35). CR2 is the only member of the family that has primary specificity for the d fragment of C3; the rest having higher affinity for C3b and/or C4b. The proteins of the gene family have the common structural characteristic of comprising 60–75 amino acid repeats with conserved residues at 11–14 positions. In addition, the genes for these proteins constitute a linkage group (36) that has been positioned at or near 1q32 (37, 38), and four of these genes, CR1, CR2, DAF, and C4bp, have been physically aligned in this order within a region of 750 kb (39, 40).

In this paper we report the analysis of the full-length cDNA for human CR2. This sequence has been used to identify internally homologous regions within the CR2 protein and regions of homology with other members of the gene family. The cDNA has also been used to analyze the organization of the CR2 gene.

Materials and Methods

Cell Culture. Cells of the B lymphoblastoid line Raji were cultured at 5×10^5/ml to 2×10^6/ml in RPMI 1640 medium containing 10% FCS, 2 mM glutamine, 50 U of penicillin/ml, and 50 μg of streptomycin/ml in the presence of 10% CO_2 at 37°C.

Isolation of Cytoplasmic RNA. Raji cells (1.5×10^9) were washed in 30 ml of ice-cold HBSS and suspended in 10 ml of ice-cold buffer containing 0.01 M Tris-HCl (pH 8.0), 0.14 M NaCl, and 1.5 mM $MgCl_2$. 10 ml of this buffer containing 0.8% NP-40 and 20 mM ribonucleoside vanadyl complex (New England Biolabs, Beverly, MA) were added and the suspension was incubated on ice for 2 min. The nuclei were removed by centrifugation at 800 *g* for 5 min at 4°C. The supernatant was made 1% with SDS and 5 mM with EDTA and extracted twice with 65°C phenol and once with chloroform. The aqueous phase was brought to 0.3 M with sodium acetate and precipitated with 2.2 volumes of EtOH. The RNA was recovered by centrifugation at 2,500 *g* for 30 min at 4°C. Total cytoplasmic RNA was poly(A)$^+$ selected on oligo-dT–cellulose (41) (Pharmacia Fine Chemicals, Piscataway, NJ).

Isolation and Sequencing of CR2 cDNAs. The isolation of λ4.11 from a human tonsillar cDNA library (29) in λgt11 was previously described (28). λ6.21 and λ6.11 were isolated from the same library by hybridization with probes derived from λ4.11 and labeled with α-[^{32}P]dCTP by nick translation.

The dideoxy chain termination technique (42) was used for sequencing cDNA that had been subcloned into M13 vectors by shotgun or directional cloning. DNA synthesis was polymerized with either Klenow (Boehringer Mannheim Biochemicals, Indianapolis, IN)

or Sequenase (United States Biochemical Corporation, Cleveland, OH) and dITP was incorporated in sequences displaying a high degree of GC compaction (43).

Isolation and Mapping of Overlapping Genomic Bacteriophage. A human genomic library prepared in EMBL3 (44) from DNA purified from normal peripheral blood leukocytes was screened with probes derived from the CR2 cDNA. Hybriding phage were isolated and overlapping clones were mapped by restriction endonuclease digestions with Eco RI, Hind III, and Bam HI (New England BioLabs) (45). The positions of sequences encoding CR2 protein were identified by hybridization with cDNA probes. Southern blot analysis of total genomic DNA was also used to confirm the sizes of large (>15 kb) restriction endonuclease fragments.

Primer Extension Analysis. A 24-nucleotide oligomer (GGAAGGTACCTGAACA-ACTGTACC) corresponding to nucleotides 143–166 of the antisense strand of the cloned cDNA was synthesized with a DNA synthesizer (model 380A; Applied Biosystems, Inc., Foster City, CA). The fragment was end labeled with T4 polynucleotide kinase and γ-[^{32}P]ATP to a specific activity of 10^8 cpm/μg. Poly(A)$^+$ selected cytoplasmic RNA (2 μg) from the B lymphoblastoid line Raji was annealed with 2 μg of labeled primer by incubation in 0.4 M NaCl, 10 mM Pipes (pH 6.4) for 4 min at 85°C and 15 h at 65°C. Reverse transcription was carried out by adding 80 μl of 10 mM DTT, 6 mM $MgCl_2$, 0.5 mM of each dNTP, 50 mM Tris (pH 8.2), and 25 U of AMV reverse transcriptase (Life Sciences, Tampa, FL) for 1 h at 41°C. The reaction volume was increased to 250 μl with dH_2O, extracted with phenol, precipitated with ethanol, and redissolved in 10 μl of sequencing dyes. Control M13 DNA that was sequenced and run on the same gel for determination of the length of extension was subjected to the same buffers and extraction as the primer extension sample.

Computerized Analysis of DNA and Protein Sequences. DNA sequence analysis and dot matrix homology analysis were performed at the computer facility of the Howard Hughes Medical Institute, Harvard Medical School Department of Genetics, using the Genetics Computer Group software of the University of Wisconsin Biotechnology Center (46).

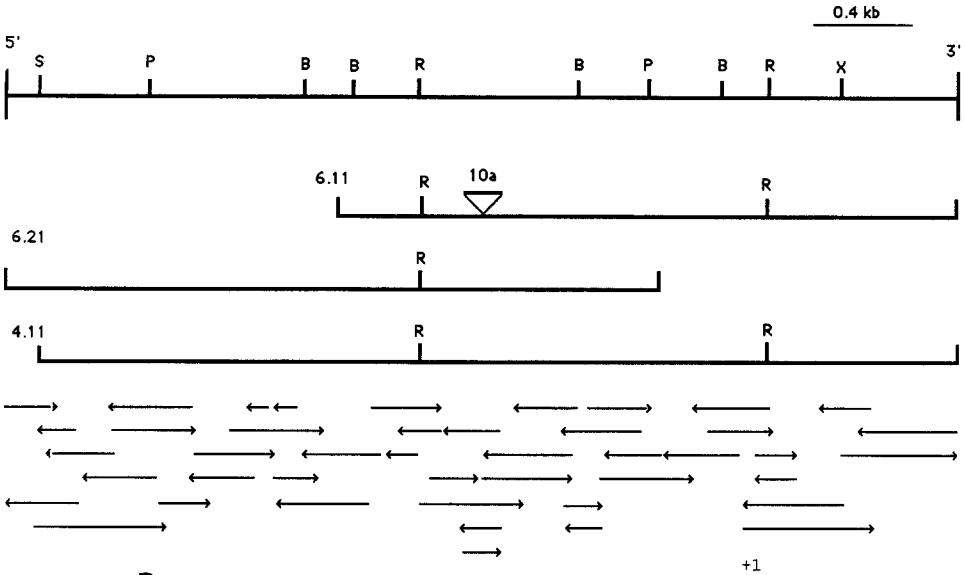
Results

Nucleotide Sequence of Human CR2 cDNA. Nucleotide sequence analysis was performed on three overlapping cDNA clones, λ6.21, λ4.11, and λ6.11, from a λgt11 tonsillar library (Fig. 1*A*). The composite cDNA from the clones λ6.21 and λ4.11 was 4 kb in length (*A*). Two internal Eco RI sites were present in the cDNA, providing three major fragments; a 1.7-kb fragment from the 5′ end of the cDNA, an internal Eco RI fragment of 1.4 kb, and a 0.7-kb fragment from the 3′ end. The full-length internal Eco RI fragment of 1.4 kb present in λ4.11 was found in two other clones in the tonsillar library (not shown). Clone λ6.21 also contained ~1 kb from the 5′ region of this internal Eco RI fragment. In contrast, the internal Eco RI fragment from clone λ6.11 contained ~200 bp of DNA inserted 283 nucleotides from the 5′ end of the fragment, which was not found in any of the other four clones extending across this region.

The complete nucleotide sequence of the CR2 cDNA is shown in Fig. 1*B*. The signal peptide DNA sequence, starting with the initiation codon ATG at position +1, is underlined, as is the sequence corresponding to the hydrophobic transmembrane regions from positions 2,923 to 2,994. The termination codon TGA at position 3,150 is double underlined. Shown in bold starting at position 3,841 is a poly-adenylation signal sequence. The arrow between nucleotides 1,974 and 1,975 indicates the position of the additional sequence identified in λ6.11, leading to the insertion of 177 nucleotides in the cDNA. The sequence of the additional DNA in λ6.11 is shown in Fig. 1*C*.

To confirm that cDNA corresponding to the entire coding region of the

A



B

+1

-55	CCCGGACGCTCGCGGGTCTCGGAACGCATCCCGCCGCGGGGGCTTCGGCCGTGGCATGGG	5
6	CGCCGCGGGCCTGCTCGGGGTTTTCTTGGCTCTCGTCGCACCGGGGGTCCTCGGGATTTC	65
66	TTGTGGCTCTCCTCCGCCTATCCTAAATGGCCGGATTAGTTATTATTCTACCCCCATTGC	125
126	TGTTGGTACCGTGATAAGGTACAGTTGTTCAGGTACCTTCCGCCTCATTGGAGAAAAAAG	185
186	TCTATTATGCATAACTAAAGACAAAGTGGATGGAACCTGGGATAAACCTGCTCCTAAATG	245
246	TGAATATTTCAATAAATATTCTTCTTGCCCTGAGCCCATAGTACCAGGAGGATACAAAAT	305
306	TAGAGGCTCTACACCCTACAGACATGGTGATTCTGTGACATTTGCCTGTAAAACCAACTT	365
366	CTCCATGAACGGAAACAAGTCTGTTTGGTGTCAAGCAAATAATATGTGGGGGCCGACACG	425
426	ACTACCAACCTGTGTAAGTGTTTTCCCTCTCGAGTGTCCAGCACTTCCTATGATCCACAA	485
486	TGGACATCACACAAGTGAGAATGTTGGCTCCATTGCTCCAGGATTGTCTGTGACTTACAG	545
546	CTGTGAATCTGGTTACTTGCTTGTTGGAGAAAAGATCATTAACTGTTTGTCTTCGGGAAA	605
606	ATGGAGTGCTGTCCCCCCCACATGTGAAGAGGCACGCTGTAAATCTCTAGGACGATTTCC	665
666	CAATGGGAAGGTAAAGGAGCCTCCAATTCTCCGGGTTGGTGTAACTGCAAACTTTTTCTG	725
726	TGATGAAGGGTATCGACTGCAAGGCCCACCTTCTAGTCGGTGTGTAATTGCTGGACAGGG	785
786	AGTTGCTTGGACCAAAATGCCAGTATGTGAAGAAATTTTTTGCCCATCACCTCCCCCTAT	845
846	TCTCAATGGAAGACATATAGGCAACTCACTAGCAAATGTCTCATATGGAAGCATAGTCAC	905
906	TTACACTTGTGACCCGGACCCAGAGGAAGGAGTGAACTTCATCCTTATTGGAGAGAGCAC	965
966	TCTCCGTTGTACAGTTGATAGTCAGAAGACTGGGACCTGGAGTGGCCCTGCCCCACGCTG	1025
1026	TGAACTTTCTACTTCTGCGGTTCAGTGTCCACATCCCCAGATCCTAAGAGGCCGAATGGT	1085
1086	ATCTGGGCAGAAAGATCGATATACCTATAACGACACTGTGATATTTGCTTGCATGTTTGG	1145
1146	CTTCACCTTGAAGGGCAGCAAGCAAATCCGATGCAATGCCCAAGGCACATGGGAGCCATC	1205
1206	TGCACCAGTCTGTGAAAAGGAATGCCAGGCCCCTCCTAACATCCTCAATGGGCAAAAGGA	1265
1266	AGATAGACACATGGTCCGCTTTGACCCTGGAACATCTATAAAATATAGCTGTAACCCTGG	1325
1326	CTATGTGCTGGTGGGAGAAGAATCCATACAGTGTACCTCTGAGGTGTGGACACCCCCTGT	1385
1386	ACCCCAATGCAAAGTGGCAGCGTGTGAAGCTACAGGAAGGCAACTCTTGACAAAACCCCA	1445
1446	GCACCAATTTGTTAGACCAGATGTCAACTCTTCTTGTGGTGAAGGGTACAAGTTAAGTGG	1505
1506	GAGTGTTTATCAGGAGTGTCAAGGCACAATTCCTTGGTTTATGGAGATTCGTCTTTGTAA	1565

```
1626  AGAAGATTTTCCATATGGAACCACGGTCACTTACACATGTAACCCTGGGCCAGAAAGAGG  1685
1686  AGTGGAATTCAGCCTCATTGGAGAGAGCACCATCCGTTGTACAAGCAATGATCAAGAAAG  1745
1746  AGGCACCTGGAGTGGCCCTGCTCCCCTGTGTAAACTTTCCCTCCTTGCTGTCCAGTGCTC  1805
1806  ACATGTCCATATTGCAAATGGATACAAGATATCTGGCAAGGAAGCCCCATATTTCTACAA  1865
1866  TGACACTGTGACATTCAAGTGTTATAGTGGATTTACTTTGAAGGGCAGTAGTCAGATTCG  1925
1926  TTGCAAAGCTGATAACACCTGGGATCCTGAAATACCAGTTTGTGAAAAAGAAACATGCCA  1985
1986  GCATGTGAGACAGAGTCTTCAAGAACTTCCAGCTGGTTCACGTGTGGAGCTAGTTAATAC  2045
2046  GTCCTGCCAAGATGGGTACCAGTTGACTGGACATGCTTATCAGATGTGTCAAGATGCTGA  2105
2106  AAATGGAATTTGGTTCAAAAAGATTCCACTTTGTAAAGTTATTCACTGTCACCCTCCACC  2165
2166  AGTGATTGTCAATGGGAAGCACACAGGCATGATGGCAGAAAACTTTCTATATGGAAATGA  2225
2226  AGTCTCTTATGAATGTGACCAAGGATTCTATCTCCTGGGAGAGAAAAAATTGCAGTGCAG  2285
2286  AAGTGATTCTAAAGGACATGGATCTTGGAGCGGGCCTTCCCCACAGTGCTTACGATCTCC  2345
2346  TCCTGTGACTCGCTGCCCTAATCCAGAAGTCAAACATGGGTACAAGCTCAATAAAACACA  2405
2406  TTCTGCATATTCCCACAATGACATAGTGTATGTTGACTGCAATCCTGGCTTCATCATGAA  2465
2466  TGGTAGTCGCGTGATTAGGTGTCATACTGATAACACATGGGTGCCAGGTGTGCCAACTTG  2525
2526  TATCAAAAAAGCCTTCATAGGGTGTCCACCTCCGCCTAAGACCCCTAACGGGAACCATAC  2585
2586  TGGTGGAAACATAGCTCGATTTTCTCCTGGAATGTCAATCCTGTACAGCTGTGACCAAGG  2645
2646  CTACCTGCTGGTGGGAGAGGCACTCCTTCTTTGCACACATGAGGGAACCTGGAGCCAACC  2705
2706  TGCCCCTCATTGTAAAGAGGTAAACTGTAGCTCACCAGCAGATATGGATGGAATCCAGAA  2765
2766  AGGGCTGGAACCAAGGAAAATGTATCAGTATGGAGCTGTTGTAACTCTGGAGTGTGAAGA  2825
2826  TGGGTATATGCTGGAAGGCAGTCCCCAGAGCCAGTGCCAATCGGATCACCAATGGAACCC  2885
2886  TCCCCTGGCGGTTTGCAGATCCCGTTCACTTGCTCCTGTCCTTTGTGGTATTGCTGCAGG  2945
2946  TTTGATACTTCTTACCTTCTTGATTGTCGTTACCTTATACGTGATATCAAAACACAGAGC  3005
3006  ACGCAATTATTATACAGATACAAGCCAGAAAGAAGCTTTTCATTTAGAAGCACGAGAAGT  3065
3066  ATATTCTGTTGATCCATACAACCCAGCCAGCTGATCAGAAGACAAACTGGTGTGTGCCTC  3125
3126  ATTGCTTGGAATTCAGCGGAATATTGATTAGAAAGAAACTGCTCTAATATCAGCAAGTCT  3185
3186  CTTTATATGGCCTCAAGATCAATGAAATGATGTCATAAGCGATCACTTCCTATATGCACT  3245
3246  TATTCTCAAGAAGAACATCTTTATGGTAAAGATGGGAGCCCAGTTTCACTGCCATATACT  3305
3306  CTTCAAGGACTTTCTGAAGCCTCACTTATGAGATGCCTGAAGCCAGGCCATGGCTATAAA  3365
3366  CAATTACATGGCTCTAAAAAGTTTTGCCCTTTTTAAGGAAGGCACTAAAAAGAGCTGTCC  3425
3426  TGGTATCTAGACCCATCTTCTTTTTGAAATCAGCATACTCAATGTTACTATCTGCTTTTG  3485
3486  GTTATAATGTGTTTTTAATTATCTAAAGTATGAAGCATTTTCTGGGGTTATGATGGCCTT  3545
3546  ACCTTTATTAGGAAGTATGGTTTTATTTTGATAGTAGCTTCCTCCTCTGGTTGGTGTTAA  3605
3606  TCATTTCATTTTTACCCTTACTTGGTTTGAGTTTCTCTACATTACTGTATATACTTTGCC  3665
3666  TTTCCATAATCACTCAGTGATTGCAATTTGCACAAGTTTTTTTAAATTATGGGAATCAAG  3725
3726  ATTTAATCCTAGAGATTTGGTGTACAATTCAGGCTTTGGATGTTTCTTTAGCAGTTTTGT  3785
3786  GATAAGTTCTAGTTGCTTGTAAAATTTCACTTAATAATGTGTACATTAGTCATTCAATAA  3845
3846  ATTGTAATTGTAAAGAAAACAAA  3868
```

FIGURE 1. *B* (continued). See legend under Fig. 1*C*.

C

1975 GGCTGCCAGCCACCTCCTGGGCTCCACCATGGTCATCATACAGGTGGAAATACGGTCTTC 2034

2035 TTTGTCTCTGGGATGACTGTAGACTACACTTGTGACCCTGGCTATTTGCTTGTGGGAAAC 2094

2095 AAATCCATTCACTGTATGCCTTCAGGAAATTGGAGTCCTTCTGCCCCACGGTGTGAA 2151

FIGURE 1. Nucleotide sequence of the human CR2 cDNA. Nucleotide sequence analysis was performed on three overlapping cDNA clones isolated from a human tonsillar cDNA library in λgt11. The composite cDNA is represented by the top line of *A*, with the positions of the following restriction endonuclease sites indicated: SSp I (S), Pst I (P), Bbv I (B), Eco RI (R), and Xba I (X). The placement of the overlapping cDNA clones is indicated below. The insert above λ6.11, labeled 10a, indicates the position of the additional sequence present in this clone but not in λ4.11 or λ6.21. The position and direction of the M13 subclones which were sequenced are indicated by the arrows in the lower portion of *A*. The nucleotide sequence of the human CR2 cDNA is shown in *B*. The sequences corresponding to the predicted signal peptide (+1–60) and the transmembrane region (2,923–2,994) are underlined. The position of the termination codon (3,150) is double underlined and the recognition sequence for polyadenylation (3,841) is in bold. The arrow between positions 1,974 and 1,975 indicates the site of insertion of the additional 177 nucleotides in λ6.11. *C* contains the nucleotide sequence of this region of λ6.11. These sequence data have been submitted to EMBL/Gen Bank Libraries under the accession number Y00649.

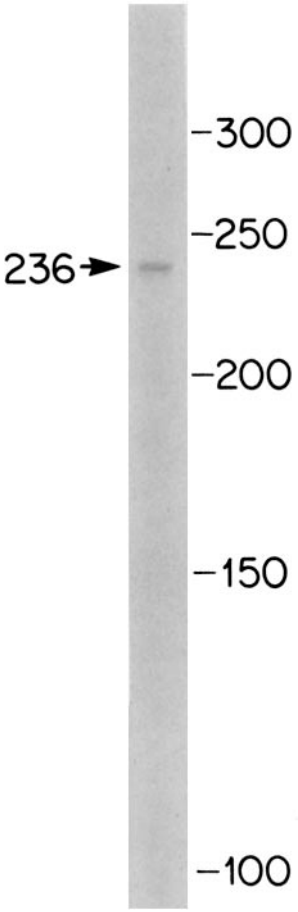


FIGURE 2. Primer extension analysis of CR2 mRNA. A 24-nucleotide oligomer derived from the positions 143–166 of the antisense strand of the cDNA was end labeled with γ-[^{32}P]ATP and used to prime reverse transcription of poly(A)$^+$ selected cytoplasmic RNA from Raji cells. Indicated on the right of the autoradiograph are the positions of migration of ssDNA as determined from a control M13 sequence. The arrow on the left indicates the position of the primer extension fragment.

FIGURE 3. Derived amino acid sequence of human CR2. The amino acid sequence derived from the nucleotide sequence of the CR2 cDNA is shown in *A*. Residues 1–20 correspond to the predicted signal peptide. Residues 21–974 contain 15 copies of the SCR sequences. These repeats are aligned by the positions of the four cysteine residues, which are boxed. Shown in italics are the positions of 11 sites for *N*-linked glycosylation. The five underlined sequences correspond to the amino acid sequences of CR2 tryptic peptides. Residues 975–998 comprise the transmembrane region and residues 999–1032 the cytoplasmic region. Indicated in bold in the transmembrane and cytoplasmic region are those residues that could serve as substrates for phosphorylation. The amino acid sequence derived from the additional 177 nucleotides of clone λ6.11 are shown in *B*, with the positions of two sites for *N*-linked glycosylation double underlined. The positions of 23 residues which are conserved in over half the CR2 repeats are shown in *C*. Residues common to other proteins having the SCR are underlined.

mRNA had been cloned, primer extension analysis was performed. A 24-nucleotide oligomer derived from positions 143–166 of the antisense strand of the cDNA was end labeled and used to prime reverse transcription of poly(A)$^+$ selected cytoplasmic RNA from the B lymphoblastoid line Raji. A single-labeled DNA fragment of 236 nucleotides in length was synthesized (Fig. 2), indicating that the site of initiation of transcription of CR2 message was 15 nucleotides upstream from the cDNA. Primer extension analysis with RNA from the B lymphoblastoid cell line SB indicated the same site was used for initiation of transcription in this cell line, while no DNA fragments were synthesized with RNA from a CR2$^-$ T cell line HSB-2 (data not shown).

Primary Structure of Human CR2. The derived amino acid sequence of human CR2 is shown in Fig. 3*A*. A hydrophobic signal peptide of ~20 amino acids was identified in the NH_2-terminal portion of the protein. Application of the von Heijne rules (47) places the cleavage site for the mature peptide at position 20,

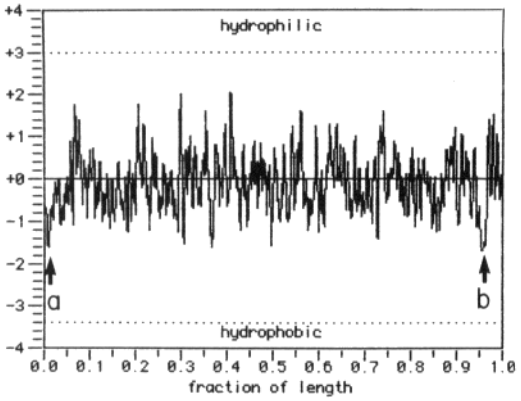


FIGURE 4. Hydrophilicity analysis of the CR2 peptide. Hopp-Woods hydrophilicity analysis was performed on the derived amino acid sequence of CR2 using the DNA Inspector II+ software (Textco, W. Lebanon, NH) on an Apple Macintosh Plus computer. The regions corresponding to the signal peptide and transmembrane region are labeled *a* and *b*, respectively.

between the glycine and isoleucine residues. However, the NH_2 terminus of CR2 is blocked (48) and the exact site of cleavage is not known. Immediately following the signal peptide are 15 copies, in tandem, of the short consensus repeat (SCR) sequence found in other members of the family of C3/C4 bp. The underlined sequences (Fig. 3*A*) correspond to sequences of tryptic peptides determined for the CR2 protein (28). The SCRs comprise 954 residues of the CR2 peptide. Immediately following the final SCR is a 24–amino acid hydrophobic sequence that may represent the transmembrane region and a 34–amino acid putative cytoplasmic domain. The assignment of the signal peptide at nucleotide +1 of the cDNA, nucleotide 70 of the mRNA, was supported by the hydrophilicity analysis of the CR2 peptide sequence (Fig. 4). Two regions of extended hydrophobicity were identified, one corresponding to the signal peptide at the NH_2-terminal portion of the protein, and the other to the transmembrane region towards the COOH terminus. Thus, the entire extracellular portion of CR2 probably comprises SCRs, making CR2 resemble CR1 but not DAF in this respect, the latter having a serine–threonine–rich region that is *O*-glycosylated (33, 34). Within the cytoplasmic region, four tyrosines, two threonines, and four serines were identified that could serve as substrates for phosphorylation (49). Two of the serines, at positions 998 and 1,010, are located close to basic residues and may be substrates for phosphorylation by protein kinase C (50).

Indicated in Fig. 3*B* is the amino acid sequence derived from the additional DNA found in the internal Eco RI fragment of clone λ6.11. This corresponds exactly to the insertion of an SCR between the 10th and 11th SCRs identified in four other cDNA clones and is labeled SCR-10a. The fact that the reading frame is maintained in λ6.11 and that the positions of residues common to other SCRs are conserved argues that this cDNA clone reflects an alternative splicing event or the transcriptional product from a different allele rather than a cloning artifact or an unspliced intervening sequence. In addition, this region is not flanked by consensus splice sequences. SCR-10a may correspond to the additional SCR reported by Moore et al. (51).

The SCRs of this gene family are characterized by the four conserved cysteines which are aligned and indicated by boxes in Fig. 3*A*. Shown in Fig. 3*C* are the positions of 23 residues conserved in over half of the CR2 repeats. The under-

lined glycine, tryptophan, and prolines are also characteristic of the SCR structure of other C3/C4 binding proteins.

The positions of 11 potential sites of *N*-linked glycosylation, Asn-X-Ser/Thr, are double underlined in Fig. 3*A*. Two additional sites are present in SCR-10a shown in *B*. These findings are consistent with the 8–11 *N*-linked oligosaccharides previously predicted for CR2 from comparison of the estimated molecular weight of the high mannose-containing precursor and the nonglycosylated forms of CR2 (52). The molecular weight of the peptide encoded by the open reading frame is 112,716, which is also similar to the molecular weight of 111,000 determined for the nonglycosylated form of CR2. The predicted signal peptide has a molecular weight of 2,162 and SCR 10-a from λ6.11 has a molecular weight of 6,234.

Analysis of Internal Homologies in CR2. Dot matrix homology analysis was performed on the entire derived amino acid sequence of CR2, with and without SCR-10a, to determine whether groups of SCRs formed repeating homologous sequences (Fig. 5). A pattern is evident, beginning at the NH_2 terminus, wherein the sequence contained in SCR-1 through -4 is tandemly repeated three times within CR2, with similar SCRs occupying the same relative positions within each group. That is, SCRs -1, -5, -9, and -12 are similar, as are SCRs -2, -6, -10, and -13, SCRs -3, -7, and -14, and SCRS -4, -8, and -11. The COOH-terminal SCR-15 does not conform to this pattern, most resembling SCRs -6, -10, and -13 rather than SCRs -4, -8, and -11. Thus, in the absence of SCR-10a, the additional sequence found in clone λ6.11, the first and second group contain four SCRs, and the third and fourth contain three SCRs (Fig. 5). Interestingly, SCR-10a is located between SCR-10 and -11 and is most homologous to SCR-3, -7, and -14, so that its presence completes the third group of linked SCRs (Fig. 5). Therefore, the 15 or 16 SCRs of CR2 are organized into four homologous groups of linked repeats. The actual breakpoints between groups cannot be determined and there is an alternative grouping of SCRs which is consistent with this organization in which group I would consist of SCRs 1–3, group II of SCRs 4–7, group III of SCRs 8–10a, and group IV of SCRs 11–14. SCRs -4 and -8 are both homologous to SCR-11 but display little homology with other SCRs, indicating they contain relatively unique sequences which could have functional importance.

Comparison of CR2 to Other Members of the Gene Family. An earlier study demonstrated that tryptic peptides of human CR2 were homologous with the amino acid sequence derived from human CR1 cDNA and that CR2 cDNA crosshybridized with CR1 cDNA under conditions of low stringency hybridization (28). To determine the extent of the relatedness between the two proteins, dot matrix homology analysis was performed between the derived amino acid sequence of CR2 and the derived amino acid sequence of CR1 (30) (Fig. 6*A*). The CR1 sequence contains three long homologous repeats that are 70–99% identical and that comprise seven SCRs. The CR1 sequence also includes the transmembrane and the cytoplasmic regions. In addition to multiple short regions of homology between CR2 and each long homologous repeat of CR1, there were four extended regions that tended to correspond to the four linked groups of SCRs in CR2. In addition, the linked group of short consensus repeats found in CR2 was partially duplicated in each long repeat of CR1 suggesting that the long

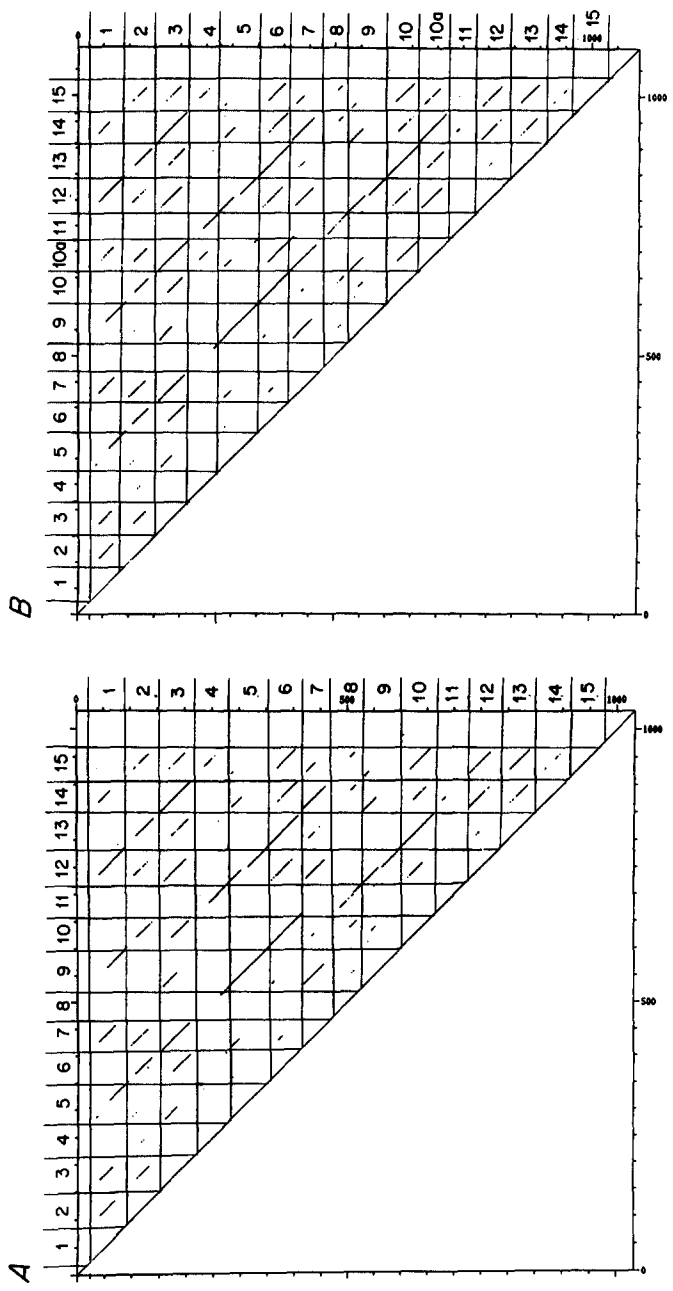


FIGURE 5. Evidence for internally repeated units within CR2. The amino acid sequence of CR2 was compared with itself by dot matrix homology (56). Shown in *A* is the analysis of the sequence derived from the composite cDNA encoding 15 SCRs while in *B* the sequence for SCR-10a, derived from clone λ6.11, has been included. The parameters were set for a window of 40 and a stringency of 20. The uninterrupted diagonal line present in *A* and *B* indicates recognition of identity of SCR with itself. Lines above this diagonal indicate regions homologous with other sequences in the CR2 peptide. The horizontal and vertical lines delineate the boundaries of each SCR, which are marked on the top and right-hand sides of the panel. The cartoon in *C* depicts the four groups of repeated sequences. Each box represents an SCR, with the interior patterns in the boxes indicating similar sequences.

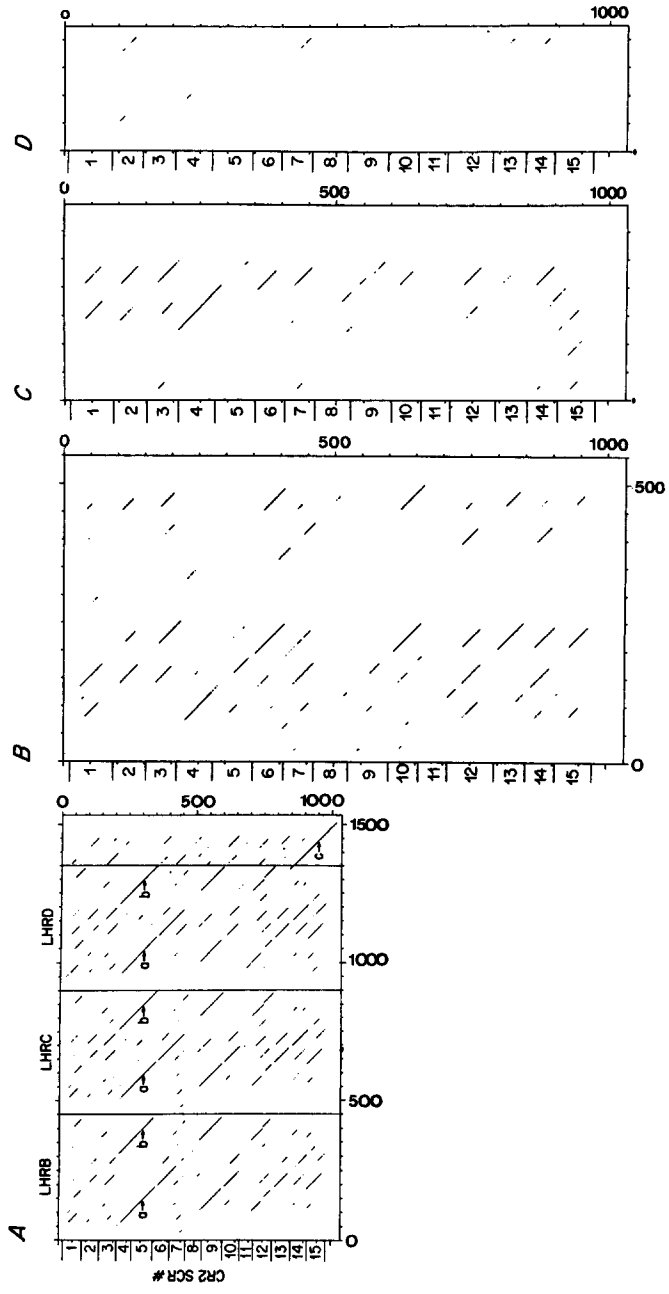


FIGURE 6. Homology analysis of CR2 with other members of the gene family of complement regulatory proteins. Dot matrix homology analysis was performed between the amino acid sequence of CR2 and the sequence of human CR1 (30) in *A*, C4bp (31) in *B*, DAF (33) in *C*, and factor H (32) in *D*. The boundaries of each SCR of CR2 are indicated by the hatch marks on the left-hand portion of each panel. The vertical lines in *A* represent boundaries of each long homologous repeat of CR1. The arrows in *A* mark the positions of extended homology between CR2 and CR1. A window of 40 and stringency of 20 were used.

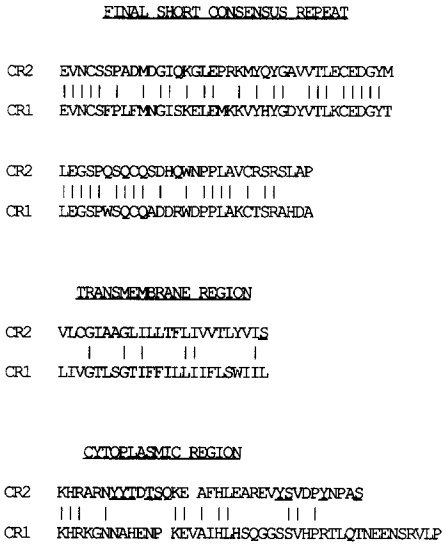


FIGURE 7. Comparison of the COOH-terminal regions of CR2 and CR1. The amino acid sequence of the final SCR, the transmembrane region, and the cytoplasmic region of CR2 is aligned with the corresponding regions from CR1 (30). Identical residues are indicated by the vertical lines between the two sequences. The residues which are underlined in the CR2 sequence could serve as substrates for phosphorylation.

homologous repeat of CR1 may have been generated by duplication of a primordial group of three to four SCRs. Less homology was detected between the NH_2-terminal SCR of each CR1 long homologous repeat and the CR2 sequence, suggesting that a CR1 function not shared by CR2 may reside in this region. Similarly, SCR-8 of CR2 may be involved in a function unique to this receptor, as it demonstrated little homology with any SCRs of CR1.

There was an extended homology between SCR-14 and -15 of CR2 and the two most COOH-terminal SCRs of CR1 (Fig. 6). The final two SCRs of CR1 are distinguished by their not being included within long homologous repeats (30) and, as noted earlier, SCR-15 of CR2 is not part of one of the groups of linked SCRs. The amino acid sequences of the final short consensus repeat, the transmembrane region, and the cytoplasmic region for both CR2 and CR1 are compared in Fig. 7. 64% of the residues in the final short consensus repeat were identical between the two proteins, with other conservative substitutions. The identity dropped to 25% in the transmembrane region. The cytoplasmic regions of both proteins began with the putative anchor sequence, Lys-His-Arg, and contained 12 conserved residues, with the CR1 sequence extending 9 residues further. These similarities between CR2 and CR1 in the sequence and spacing of the final short consensus repeat, the transmembrane region, and the cytoplasmic region provide evidence for derivation from a common membrane-bound precursor gene.

Comparison of the amino acid sequence of CR2 with the derived amino acid sequence of C4bp (31) (Fig. 6*B*) also revealed regions of homology, although much less than that observed between CR2 and CR1. The homology pattern was similar to that observed between CR2 and CR1 as the NH_2-terminal SCR of C4bp showed little homology with that of CR2, again suggesting that this SCR may be related to functions shared by CR1 and C4bp but not CR2, such as the binding of C3b and C4b. Comparison with the derived amino acid sequence of DAF (33) (Fig. 6*C*) revealed homology between CR2 and the carboxyl half of

the SCR-containing region of DAF, less in the NH_2-terminal SCRs of DAF, and none with the serine-threonine carboxyl tail of DAF. At the level of stringency used in this analysis, little homology was detected between CR2 and the partial sequence of factor H (32) (Fig. 6*D*). Thus, factor H may have diverged earlier from the presumptive ancestral gene encoding the other C3/C4 binding proteins.

Genomic Organization of Human CR2. Four EMBL3 bacteriophage containing overlapping human genomic DNA were isolated by hybridization to cDNA probes derived from CR2 cDNA and a map encompassing 43 kb of DNA was generated from restriction endonuclease sites (Fig. 8). The location of sequences hybridizing to probes derived from coding and noncoding sequences in the cDNA was determined (Fig. 8*A*). A single Eco RI to Bam HI fragment of 0.9 kb was identified that hybridized to 2–7, a 0.7-kb probe containing only 3′ untranslated sequence (Fig. 8*B*). Several fragments interspaced on 12.5 kb of DNA hybridized to 2–14, a 1.4-kb probe derived from the 3′ region of the coding sequence and containing six copies of the SCR sequence, the transmembrane region, and the cytoplasmic region. The cDNA probe 2–16, a 1.6-kb fragment from the 5′ region of the coding sequence but lacking the 165 nucleotides corresponding to the most 5′ region of the mRNA, hybridized to fragments in 6.5 kb of genomic DNA. 2–18, which contains 317 bp from the most 5′ region of the cloned cDNA and, therefore lacked only 15 nucleotides from the 5′ end of the mRNA, hybridized to a single 1.4-kb Hind III to Eco RI fragment. The 24-nucleotide oligomer derived from nucleotides 143–166 of the cDNA also hybridized only to this 1.4-kb fragment. Thus, the exons encoding the translated CR2 sequence are contained within 25 kb of genomic DNA.

Discussion

Human CR2 is a member of the family of complement regulatory proteins including CR1, C4bp, DAF, and factor H (35). In addition to their functional relatedness, these proteins comprise tandem copies of a 60–75–amino acid SCR (29–33). The positions of four cysteines, a tryptophan, several glycines, and several prolines, are conserved within each copy of this repeat. It has been determined from analysis of proteolytic fragments of β_2-glycoprotein I, a protein not involved in complement but comprising SCRs, that the first and third cysteines and the second and fourth cysteines of each repeat are disulfide linked (53). The proline residues are clustered in close proximity to the first and fourth cysteines in these proteins. This suggests that in proteins containing this repeat rigid SCRs are interspaced with angular hinges, which is consistent with the observation of flexible rodlike structures in electron micrographs of C4bp (54). Members of this family are related genetically as CR1, C4bp, and H have been chromosomally linked by family studies of protein polymorphisms (36), CR2 and CR1 have been mapped by in situ hybridization to band 1q32 (37), and all five have been placed on chromosome 1 by human–mouse hybrid mapping panels (35, 37, 38).

Although clearly a member of this family, CR2 has functionally diverged from the other members in that it has primary specificity for the d region of C3, it cannot accelerate the decay of classical or alternative pathway convertases, and its cofactor activity is restricted to particle bound iC3b (48, 55). It is unique also

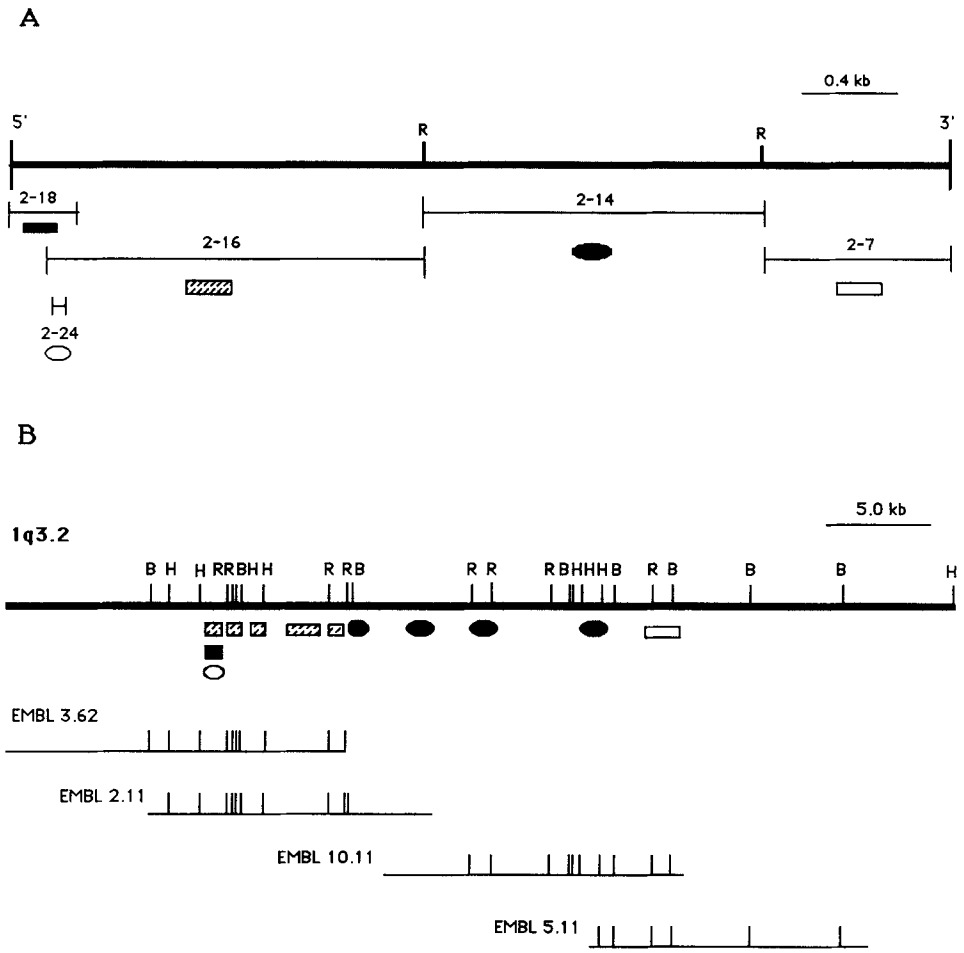


FIGURE 8. Genomic organization of human CR2. The placement of five cDNA probes used for isolation and mapping of the genomic bacteriophage is diagramed in *A*. 2–7 is a 0.7-kb fragment from the 3′ end of the cDNA that contained only 3′ untranslated sequence. 2–14 is a 1.4-kb Eco RI to Eco RI fragment from the middle of the cDNA that encodes six copies of the SCR, the transmembrane, and the cytoplasmic regions. 2–16 is a 1.6-kb fragment from the 5′ end of the cDNA, lacking 165 nucleotides from the most 5′ region of the mRNA, and encoding only SCR sequences. 2–18 is a 317-bp fragment starting 15 nucleotides downstream of the initiation codon for transcription and containing 5′ untranslated sequence, the signal peptide and 202 bp encoding short consensus repeat sequence. 2–24 if the 24-nucleotide oligomer derived from positions 143–166 of the cDNA. The map of the overlapping genomic phage is shown in *B*. The position of the following restriction endonuclease sites are shown on the composite chromosome bar in the top of *B*: Bam HI (B), Hind III (H), and Eco RI (R). The positions of the individual overlapping phage are drawn below the bar. Hybridization to the cDNA probes is indicated by placement of the circles and boxes below the chromosome 1 bar.

in its ability to bind the EBV glycoprotein gp350/220 and in being the only member of the family that is found primarily on the B lymphocyte.

Analysis of the derived amino acid sequence of CR2 has provided information pertaining to its own evolution and that of other members of the gene family. The extracellular portion of CR2 comprises four groups of linked SCRs, with

groups I and II containing four SCRs and groups III and IV containing three SCRs (Fig. 5). This structural organization suggests that SCR may have arisen by duplication of a primordial gene segment containing exons encoding four short consensus repeats. The identification of the fourth SCR of group III, SCR-10a, in λ6.11 supports the concept of linked groups of four SCRs. The functional significance of four groups of linked SCRs is not clear as it is not known whether CR2 is multivalent with respect to binding of C3d. The ligand binding site may be located within the NH_2-terminal group, with the others serving primarily to extend the receptor from the plasma membrane, as has been suggested for the long homologous repeats of CR1 (30). The C4 binding site of C4bp has also been localized to the periphery of each of the seven C4bp rods that make up the C4bp complex (54). As the entire extracellular portion of CR2 comprises SCRs, the ligand binding sites for C3d and EBV must be formed by this basic structural element.

Evidence for polymorphism of CR2 has been presented in this paper. A single cDNA clone that was isolated from the human tonsillar library contained a 177-nucleotide long insert ~2 kb from the 5′ end that encoded a single SCR (Figs. 1 and 4) that probably corresponds to that recently found by Moore et al. (51). Four other clones did not contain this DNA sequence. This cDNA may represent the product of a different allele or an alternative splicing event from the same precursor RNA. As the insert maintains reading frame, encodes an SCR, and is not flanked by consensus splice sequences, it is likely to be an in vivo transcriptional product and not an unspliced intervening sequence or an artifact of cDNA cloning. Furthermore, short exposure of autoradiographs from Northern blot analysis of mRNA from tonsils and B lymphoblastoid cells does allow discrimination of two distinct, but closely migrating mRNA species of 4.8 kb (data not shown). Although there is no evidence for polymorphism in CR2 at the protein level, mature CR2 migrates as a diffuse band because it contains complex oligosaccharides that alter the migration by 35,000 mol wt (52) and could potentially mask the difference of 6,000 mol wt between the two predicted forms.

Comparison of the sequence of CR2 with the sequence of CR1 demonstrated a high degree of relatedness between the two proteins and suggested that the long homologous repeats of CR1 could have arisen by duplication of the same primordial group of linked SCRs that have been duplicated to generate CR2. The homology between the two proteins was also evident in their COOH-terminal regions (Figs. 6 and 7). This similarity in the sequence and spacing of the final SCR, the transmembrane region, and the cytoplasmic region of CR2 and CR1 suggests that this entire portion of the CR2 and CR1 genes may have originated from an ancestral gene encoding a membrane-bound, SCR-containing peptide. However, divergence has also occurred in the cytoplasmic regions of CR2 and CR1, most notably being the presence of four tyrosine residues in CR2 that are not found in CR1. Consistent with these structural differences being the basis for differing functions of these receptors on B cells is the phosphorylation of CR2 but not CR1 during treatment of cells with phorbol esters and the stimulation of B cells that has been observed after crosslinking of CR2 but not CR1. The availability of full-length cDNA for CR2 now makes possible a

molecular analysis of its B cell–activating function induced by binding of the C3d and EBV ligands.

Summary

Human complement receptor type 2 (CR2) is the B lymphocyte receptor for C3d and the Epstein-Barr virus. This protein is also a member of a family of C3b/C4b binding proteins that regulate complement activation, comprise tandemly repeated 60–75 amino acid sequences, and whose genes map to band q32 on chromosome 1. Overlapping cDNA clones encoding the entire human CR2 protein have been isolated from a human tonsillar cDNA library. The derived amino acid sequence of 1,032 residues encodes a peptide of 112,716 mol wt. A signal peptide was identified, followed by 15 copies of the short consensus repeat (SCR) structure common to the C3/C4 binding protein family. The entire extracellular portion of the protein comprised SCRs, thus, the ligand binding sites both for C3d and the EBV protein gp350/220 are positioned within this structure. Immediately following the final SCR was a transmembrane sequence of 24 amino acids and a cytoplasmic region of 34 amino acids. One of five cDNA clones isolated contained an additional SCR, providing evidence for alternative mRNA splicing or gene products of different human alleles. The CR2 cDNAs were used to isolate CR2-specific genomic phage. The entire CR2 coding sequences were found within 20 kb of human DNA. Analysis of the CR2 cDNA sequence indicated that CR2 contained internally homologous regions and suggested that CR2 arose by duplication of a primordial gene sequence encoding four SCRs. Comparison of the CR2 peptide sequence with those of other members of the gene family has identified many regions highly homologous with human CR1, fewer with C4bp and decay accelerating factor, and very few with factor H, and suggested that CR2 and CR1 arose by duplication of the same ancestral gene sequence. The homology between CR2 and CR1 extended to the transmembrane and cytoplasmic regions, suggesting that these sequences were derived from a common membrane-bound precursor.

We acknowledge the excellent technical assistance of Sarah R. Burrow.

Received for publication 18 November 1987.

Note added in proof: The termination codon is at nucleotides 3097–3099 rather than at nucleotides 3150–3152. The sequence of CR2 deduced from a cDNA clone isolated from a Raji lymphoblastoid cell library has recently been published (57). A major difference in the sequence of SCR 12 (SCR 13 in Moore et al. [57]) is present which may reflect differences in the cellular origin of the mRNA used to prepare the cDNA libraries.

References

1. Nadler, L. M., P. Stashenko, R. Hardy, A. van Agthoven, C. Terhorst, and S. F. Schlossman. 1981. Characterization of a human B cell-specific antigen (B2) distinct from B1. *J. Immunol.* 126:1941.
2. Iida, K., L. Nadler, and V. Nussenzweig. 1983. Identification of the membrane receptor for the complement fragment C3d by means of a monoclonal antibody. *J. Exp. Med.* 158:1021.
3. Weis, J. J., T. F. Tedder, and D. T. Fearon. 1984. Identification of a 145,000 M_r

membrane protein as the C3d receptor (CR2) of human B lymphocytes. *Proc. Natl. Acad. Sci. USA.* 81:881.

4. Tedder, T. F., L. T. Clement, and M. D. Cooper. 1984. Expression of C3d receptors during human B cell differentiation: immunofluorescence analysis with the HB-5 monoclonal antibody. *J. Immunol.* 133:678.
5. Reynes, M., J. P. Aubert, J. H. M. Cohen, J. Audouin, V. Tricottet, J. Diebold, and M. D. Kazatchkine. 1985. Human follicular dendritic cells express CR1, CR2, and CR3 complement receptor antigens. *J. Immunol.* 135:2687.
6. Young, L. S., J. W. Sixbey, D. Clark, and A. B. Rickinson. 1986. Epstein-Barr virus receptors on human pharyngeal epithelia. *Lancet.* i:240.
7. Eden, A., G. W. Miller, and V. Nussenzweig. 1973. Human lymphocytes bear membrane receptors for C3b and C3d. *J. Clin. Invest.* 52:3239.
8. Ross, G. D., M. J. Polley, E. M. Rabellino, and H. M. Grey. 1973. Two different complement receptors on human lymphocytes: one specific for C3b and one specific for C3b inactivator-cleaved C3b. *J. Exp. Med.* 138:798.
9. Nemerow, G. R., M. E. McNaughton, and N. R. Cooper. 1985. Binding of monoclonal antibody to the Epstein Barr virus (EBV) CR2 receptor induces activation and differentiation of human B lymphocytes. *J. Immunol.* 135:3068.
10. Frade, R., M. C. Crevon, M. Barel, A. Vazquez, L. Krikorian, C. Charriaut, and P. Galanaud. 1985. Enhancement of human B cell proliferation by an antibody to the C3d receptor, the gp 140 molecule. *Eur. J. Immunol.* 15:73.
11. Wilson, B. S., J. L. Platt, and N. E. Kay. 1985. Monoclonal antibodies to the 140,000 mol wt glycoprotein of B lymphocyte membranes (CR2 receptor) initiates proliferation of B cells in vitro. *Blood.* 66:824.
12. Tedder, T. F., J. J. Weis, L. T. Clement, D. T. Fearon, and M. D. Cooper. 1986. The role of receptors for complement in the induction of polyclonal B-cell proliferation and differentiation. *J. Clin. Immunol.* 6:65.
13. Hatzfeld, J., E. Fischer, J. P. Levesque, A. Hatzfeld, and M. D. Kazatchkine. 1987. Binding of C3 and C3dg to the CR2 complement receptor induces growth of an Epstein-Barr virus-positive human B cell line in defined serum-free conditions. *Complement: XIIth International Complement Workshop, Chamonix, France.* 166.
14. Schulz, T. F., G. Pernegger, A. Petzer, A. Eigentler, B. L. Myones, and M. P. Dierich. 1987. Growth control of B-lymphoblastoid cells by CR2: stimulation or inhibition is dependent upon the individual or simultaneous engagement of suitable CR2 epitopes. *Complement: XIIth International Complement Workshop.* 223.
15. Melchers, F., A. Erdei, T. Schulz, and M. P. Dierich. 1985. Growth control of activated, synchronized murine B cells by the C3d fragment of human complement. *Nature (Lond.).* 317:264.
16. Carter, R. H., M. Spycher, and D. T. Fearon. 1987. Synergy between complement receptors and sIgM in activation of B lymphocytes. *Complement: XIIth International Complement Workshop, Chamonix, France.* 141.
17. Tanner, J., J. Weis, D. Fearon, Y. Whang, and E. Kieff. 1987. Epstein-Barr virus gp350/220 binding to the B lymphocyte C3d receptor mediates adsorption, capping, and endocytosis. *Cell.* 50:203.
18. Changelian, P. S., and D. T. Fearon. 1986. Tissue-specific phosphorylation of complement receptors CR1 and CR2. *J. Exp. Med.* 163:101.
19. Barel, M., A. Vazquez, C. Charriaut, M. T. Aufredou, P. Galanaud, and R. Frade. 1986. gp 140, the C3d/EBV receptor (CR2), is phosphorylated upon in vitro activation of human peripheral B lymphocytes. *FEBS (Fed. Eur. Biochem. Soc.) Lett.* 197:353.
20. Jondal, M., G. Klein, M. B. A. Oldstone, V. Bokish, and E. Yefenof. 1976. Surface

markers on human B and T lymphocytes. VIII. Association between complement and Epstein-Barr virus receptors on human lymphoid cells. *Scand. J. Immunol.* 5:401.

21. Fingeroth, J. D., J. J. Weis, T. F. Tedder, J. L. Strominger, P. A. Biro, and D. T. Fearon. 1984. Epstein-Barr virus receptor of human B lymphocytes is the C3d receptor CR2. *Proc. Natl. Acad. Sci. USA.* 81:4510.
22. Nemerow, G. R., R. Wolfert, M. E. McNaughton, and N. R. Cooper. 1985. Identification and characterization of the Epstein-Barr virus receptor on human B lymphocytes and its relationship to the C3d complement receptor (CR2). *J. Virol.* 55:347.
23. Frade, R., M. Barel, B. Ehlin-Henriksson, and G. Klein. 1985. gp140, the C3d receptor of human B lymphocytes, is also the Epstein-Barr virus receptor. *Proc. Natl. Acad. Sci. USA.* 82:1490.
24. Tedder, T. F., V. S. Goldmacher, J. M. Lambert, and S. F. Schlossman. 1986. Epstein-Barr virus binding induces internalization of the C3d receptor: a novel immunotoxin delivery system. *J. Immunol.* 137:1387.
25. Nemerow, G. R., C. Mold, V. K. Schwend, V. Tollefson, and N. R. Cooper. 1987. Identification of gp 350 as the viral glycoprotein mediating attachment of Epstein-Barr virus (EBV) to the EBV/C3d receptor of B cells: sequence homology of gp 350 and C3 complement fragment C3d. *J. Virol.* 61:1416.
26. Nemerow, G. R., and N. R. Cooper. 1984. Early events in infection of human B lymphocytes by Epstein-Barr virus: internalization process. *Virology.* 132:186.
27. Sixbey, J. W., J. G. Nedrud, N. Raab-Traub, R. A. Hanes, and J. S. Pagano. 1984. Epstein-Barr virus replication in oropharyngeal epithelial cells. *N. Engl. J. Med.* 310:1225.
28. Weis, J. J., D. T. Fearon, L. B. Klickstein, W. W. Wong, S. A. Richards, A. de Bruyn Kops, J. A. Smith, and J. H. Weis. 1986. Identification of a partial cDNA clone for the C3d/Epstein-Barr virus receptor for human B lymphocytes: homology with the receptor for fragments C3b and C4b of the third and fourth components of complement. *Proc. Natl. Acad. Sci. USA.* 83:5639.
29. Wong, W. W., L. B. Klickstein, J. A. Smith, J. H. Weis, and D. T. Fearon. 1985. Identification of a partial cDNA clone for the human receptor for complement fragments C3b/C4b. *Proc. Natl. Acad. Sci. USA.* 82:7711.
30. Klickstein, L. B., W. W. Wong, J. A. Smith, J. H. Weis, J. G. Wilson, and D. T. Fearon. 1987. Human C3b/C4b receptor (CR1): demonstration of long homologous repeating domains that are composed of the short consensus repeats characteristic of C3/C4 binding proteins. *J. Exp. Med.* 165:1095.
31. Chung, L. P., D. R. Bentley, and K. B. M. Reid. 1985. Molecular cloning and characterization of the cDNA coding for C4b-binding protein, a regulatory protein of the classical pathway of the human complement system. *Biochem. J.* 260:133.
32. Kristensen, T., R. A. Wetsel, and B. F. Tack. 1986. Structural analysis of human complement protein H: homology with C4b binding protein, β_2-glycoprotein I, and the Ba fragment of B. *J. Immunol.* 136:3407.
33. Caras, I. W., M. A. Davitz, L. Rhee, G. Weddell, D. W. Martin, and V. Nussenzweig. 1987. Cloning of decay-accelerating factor suggests novel use of splicing to generate two proteins. *Nature (Lond.).* 325:545.
34. Medof, M. E., D. M. Lublin, V. M. Holers, D. J. Ayers, R. R. Getty, J. F. Leykam, J. P. Atkinson, and M. L. Tykocinski. 1987. Cloning and characterization of cDNAs encoding the complete sequence of decay-accelerating factor of human complement. *Proc. Natl. Acad. Sci. USA.* 84:2007.
35. Reid, K. B. M., D. R. Bentley, R. D. Campbell, L. P. Chung, R. B. Sim, T. Kristensen, and B. F. Tack. 1986. Complement system proteins which interact with C3b or C4b. *Immunol. Today.* 7:230.

36. Rodriguez de Cordoba, S., D. M. Lublin, P. Rubinstein, and J. P. Atkinson. 1985. Human genes for three complement components that regulate the activation of C3 are tightly linked. *J. Exp. Med.* 161:1189.
37. Weis, J. H., C. C. Morton, G. 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.
38. Rey-Campos, J., P. Rubinstein, and S. Rodriguez de Cordoba. 1987. Decay-accelerating factor: genetic polymorphism and linkage to the RCA (regulator of complement activation) gene cluster in humans. *J. Exp. Med.* 166:246.
39. Carroll, M. C., E. A. Alicot, P. Katzman, L. B. Klickstein, and D. T. Fearon. 1987. Organization of the genes encoding CR1, CR2, DAF, and C4bp in the RCA locus on human chromosome one. *Complement: XIIth International Complement Workshop, Chamonix, France.* 141.
40. Rey-Campos, J., P. Rubinstein, and S. Rodriguez de Cordoba. 1987. Mapping of the decay-accelerating factor (DAF) to the regulator of complement activation (RCA) gene cluster in humans. *Complement: XIIth International Complement Workshop, Chamonix, France.* 217.
41. Aviv, H., and P. Leder. 1972. Purification of biologically active globin messenger RNA by chromatography on oligothymidylic acid-cellulose. *Proc. Natl. Acad. Sci. USA.* 69:1408.
42. Sanger, F., S. Nicklen, and A. R. Coulson. 1977. DNA sequencing with chain-terminating inhibitors. *Proc. Natl. Acad. Sci. USA.* 74:5463.
43. Tabor, S., and C. C. Richardson. 1987. DNA sequence analysis with a modified bacteriophage T7 DNA polymerase. *Proc. Natl. Acad. Sci. USA.* 84:4767.
44. Frischauf, A. M., H. Lehrach, A. Poustka, and N. Murray. 1983. Lambda replacement vectors carrying polylinker sequences. *J. Mol. Biol.* 170:827.
45. Maniatis, T., E. F. Fritsch, and J. Sambrook. 1982. Molecular Cloning: A Laboratory Manual. Cold Spring Harbor Laboratory, Cold Spring Harbor, New York. 545 pp.
46. Devereux, J., P. Haeberli, and O. Smithies. 1984. A comprehensive set of sequence analysis programs for the VAX. *Nucleic Acids Res.* 12:387.
47. von Heijne, G. 1986. A new method for predicting signal sequence cleavage sites. *Nucleic Acids Res.* 14:4683.
48. Weis, J. J., S. A. Richards, J. A. Smith, and D. T. Fearon. 1986. Purification of the B lymphocyte receptor for the C3d fragment of complement and the Epstein-Barr virus by monoclonal antibody affinity chromatography, and assessment of its functional capacities. *J. Immunol. Methods.* 92:79.
49. Williams, R. E. 1976. Phosphorylated sites in substrates of intracellular protein kinases: a common feature in amino acid sequences. *Science (Wash. DC).* 192:473.
50. Kishimoto, A., K. Nishiyama, H. Nakanishi, Y. Uratsuji, H. Nomura, Y. Takeyama, and Y. Nishizuka. 1985. Studies on the phosphorylation of myelin basic protein by protein kinase C and adenosine 3′:5′-monophosphate-dependent protein kinase. *J. Biol. Chem.* 260:12492.
51. Moore, M. D., N. R. Cooper, and G. R. Nemerow. 1987. Molecular cloning and sequence of a cDNA clone encoding the entire EBV/C3d receptor, CR2. *Complement: XIIth International Complement Workshop, Chamonix, France.* 197.
52. Weis, J. J., and D. T. Fearon. 1985. The identification of N-linked oligosaccharides on the human CR2/Epstein-Barr virus receptor and their function in receptor metabolism, plasma membrane expression, and ligand binding. *J. Biol. Chem.* 260:13824.
53. Lozier, J., N. Takahashi, and F. W. Putnam. 1984. Complete amino acid sequence of human plasma β_2-glycoprotein I. *Proc. Natl. Acad. Sci. USA.* 81:3640.

54. Dahlback, 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.
55. Mitomo, K., T. Fujita, and K. Iida. 1987. Functional and antigenic properties of complement receptor type 2, CR2. *J. Exp. Med.* 165:1424.
56. Maizel, J. V., and R. P. Lenk. 1981. Enhanced graphic matrix analysis of nucleic acid and protein sequences. *Proc. Natl. Acad. Sci. USA.* 78:7665.
57. Moore, M. D., N. R. Cooper, B. F. Tack, G. R. Nemerow. 1987. Molecular cloning of the cDNA encoding the Epstein-Barr virus/C3D receptor (complement receptor type 2) of human B lymphocytes. *Proc. Natl. Acad. Sci. USA.* 84:9194.