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

BY LLOYD B. KLICKSTEIN,[*§] WINNIE W. WONG,[*] JOHN A. SMITH,[‡]
JOHN H. WEIS,[*] JAMES G. WILSON,[*] AND DOUGLAS T. FEARON[*]

*From the Departments of *Medicine and ‡Pathology and the §Program in Cell and
Developmental Biology, Harvard Medical School; and the *Department of Rheumatology and
Immunology, Brigham & Women's Hospital, Boston, Massachusetts 02115; and the
‡Departments of Pathology and Molecular Biology, Massachusetts General Hospital,
Boston, Massachusetts 02114*

The human C3b/C4b receptor, termed CR1, is present on erythrocytes, monocytes/macrophages, granulocytes, B cells, some T cells, splenic follicular dendritic cells, and glomerular podocytes (1–5). A soluble form of the receptor has been found in plasma that has ligand binding activity and the same M_r as membrane-associated CR1 (6). CR1 binds C3b and C4b that have covalently attached to immune complexes and other complement activators, and the consequences of these interactions depend upon the cell type bearing the receptor (7). Erythrocyte CR1 binds immune complexes for transport to the liver (8, 9). CR1 on neutrophils and monocytes internalizes bound complexes, either by adsorptive endocytosis through coated pits (10, 11) or by phagocytosis after activation of the receptor by phorbol esters, chemotactic peptides, or proteins that are present in the extracellular matrix, such as fibronectin and laminin (12–14). Phosphorylation of CR1 may have a role in the acquisition of phagocytic activity. The function of CR1 on B lymphocytes is less defined, although treatment of these cells with antibody to CR1 enhanced their response to suboptimal doses of PWM (15). CR1 on follicular dendritic cells may subserve an antigen presentation role (16) and the function of CR1 on T cells is unknown. CR1 can also inhibit the classical and alternative pathway C3/C5 convertases and act as a cofactor for the cleavage of C3b and C4b by factor I, indicating that CR1 also has complement regulatory functions in addition to serving as a receptor (17, 18).

CR1 is a glycoprotein composed of a single polypeptide chain. An unusual structure for the receptor was first suggested by the finding of four allotypic

forms differing by increments of ~40,000–50,000 M_r. The two most common forms, the F and S allotypes, also termed the A and B allotypes, have M_r of 250,000 and 290,000 (19, 20), respectively, and two rarer forms have M_r of 210,000 and >290,000 (21, 22). These differences apparently represent variations in the polypeptide chain of CR1 because they were not abolished by treatment of purified receptor protein with endoglycosidase F (20) and they were observed when receptor allotypes were biosynthesized in the presence of tunicamycin (23). All four CR1 allotypes have C3b-binding activity, indicating that insertion or deletion of these relatively large peptide segments does not destroy the ligand binding site (19–22).

Additional evidence for an unusual structure for CR1 was the finding that two nonoverlapping restriction fragments of a CR1 cDNA crosshybridized under conditions of high stringency (24). This indication of repetitive coding sequences was supported by the observation that both cDNA probes hybridized to multiple restriction fragments of genomic DNA, most of which were common to both probes (24). In addition, the CR1 gene has been shown recently to have repetitive intervening sequences by the demonstration of crosshybridization of a genomic probe lacking coding sequences to several genomic restriction fragments (25). Further, DNA from an individual having the larger S allotype had an additional restriction fragment hybridizing to this genomic probe when compared with DNA from an individual having the F allotype, suggesting that duplication of genomic sequences occurred in association with the higher molecular weight CR1 allele (25).

In the present study, the repetitive unit of CR1 was defined by sequencing 5.5 kb of tonsillar CR1 cDNA. Three tandem, direct repeats of 1.35 kb were found that encode 50-kD peptide segments, termed long homologous repeats (LHR).[1] Each LHR is composed of seven short consensus repeats (SCR) that are characteristic of C3/C4 binding proteins.

Materials and Methods

Isolation and Sequence of CR1 Tryptic Peptides.   CR1 was purified from washed human erythrocyte membranes by sequential Matrex Red A and YZ-1 mAb affinity chromatography (26). Tryptic peptides were prepared and isolated by sequential gradient and isocratic reverse-phase HPLC as described (24). Tryptic peptide analysis was performed with a 470A Protein Sequencer (Applied Biosystems, Inc., Foster City, CA), and analysis of each degradative cycle was achieved using a 120 PTH-amino acid analyzer (Applied Biosystems, Inc.).

Isolation of cDNA Clones and Genomic Clones.   A cDNA library was constructed in λgt11 from human tonsillar poly(A)$^+$ RNA as described (24). By RNA blot hybridization, the tonsil donor was homozygous for the F allele of CR1 (24). The cDNA was selected on an agarose gel to include fractions between 2 and 7 kb before the cloning steps. The initial complexity of the library was 4.5 × 10^6 recombinants per 100 ng cDNA and the library was amplified in *Escherichia coli* strain Y1088. The library was screened (27) with CR1 probes, CR1-1 and CR1-2 (24), that had been radiolabeled to a sp act of 2–8 × 10^8 cpm/μg by nick translation. Hybridization was performed in 50% formamide, 5× SSC (1× SSC: 15 mM sodium citrate, 150 mM sodium chloride) at 43°C and filters were washed at 60°C in 0.2× SSC, conditions that do not allow the detection of CR2 cDNA

---

[1] *Abbreviations used in this paper:*   C4bp, C4-binding protein; LHR, long homologous repeat; SCR, short consensus repeat.
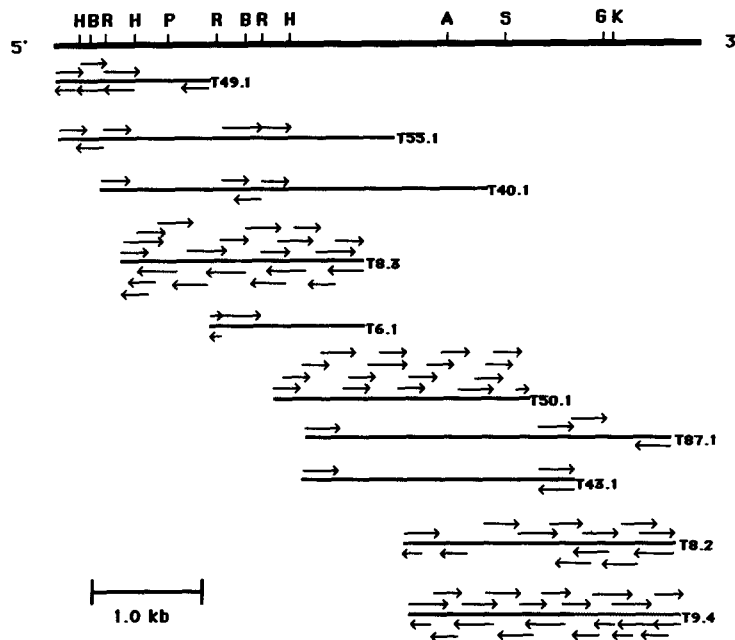
FIGURE 1. Restriction map of 5.5 kb of human CR1 cDNA. The black bar indicates the cDNA, restriction sites are H, Hind III; B, Bam HI; R, Eco RI; P, Pst I; A, Apa I; S, Sac I; G, Bgl II; K, Kpn I. The cDNA clones from which the sequence was derived are shown below the map. The arrows indicate the direction and extent of sequence analysis by the dideoxynucleotide chain termination method. cDNA clones were oriented on the basis of restriction maps and overlapping sequence identity. The grey portion of clone T9.4 indicates that this sequence was determined to be an unspliced intervening sequence as described in the text.

clones (28). Positive clones were plaque-purified twice before restriction mapping and DNA sequence analysis.

A genomic library was constructed in EMBL-3 with 15–20-kb fragments produced by partial digestion of human leukocyte DNA with Sau 3AI. The initial complexity was 1.2 × 10⁶ and the library was amplified in *E. coli* strain P2392. The library was also screened with the cDNA probes CR1-1 and CR1-2 (24).

DNA Sequence Analysis. Restriction fragments of the cDNA clones were subcloned in M13mp18 or M13mp19 and sequenced by the dideoxynucleotide chain termination method (29). Some clones were sequenced in whole or in part by first creating ordered deletion mutants using exonuclease III (30). Each region was sequenced on both strands and in most cases each region was sequenced on M13 subclones constructed from two independently isolated cDNA clones (Fig. 1). Sequence data were analyzed with the University of Wisconsin Genetics Computer Group package (Madison, WI).

Results

A size-selected tonsillar cDNA library was screened with the CR1-1 and CR1-2 probes obtained from the CR1 cDNA clone, λT8.3 (24). 15 positive phage were identified out of 1.5 × 10⁶ recombinants and 13 of these represented distinct clones. 10 were restriction mapped and sequenced in whole or in part by the dideoxynucleotide chain termination method. The cDNA clones were aligned on the basis of overlapping sequence identity (Fig. 1) and were found to span 5.5 kb (Fig. 2). A single long open reading frame was identified beginning

```
   1  CCAAGTTGAAAACCCAAACCAATGCATCTGACTTTCCCATTGGGACATCTTTAAAGTACG    60
  61  AATGCCGTCCTGAGTACTACGGGAGGCCATTCTCTATCACATGTCTAGATAACCTGGTCT   120
 121  GGTCAAGTCCCAAAGATGTCTGTAAACGTAAATCATGTAAAACTCCTCCAGATCCAGTGA   180
 181  ATGGCATGGTGCATGTGATCACAGACATCCAGGTTGGATCCAGAATCAACTATTCTTGTA   240
 241  CTACAGGGCACCGACTCATTGGTCACTCATCTGCTGAATGTATCCTCTCGGGCAATGCTG   300
 301  CCCATTGGAGCACGAAGCCGCCAATTTGTCAACGAATTCCTTGTGGGCTACCCCCCACCA   360
 361  TCGCCAATGGAGATTTCATTAGCACCAACAGAGAGAATTTTCACTATGGATCAGTGGTGA   420
 421  CCTACCGCTGCAATCCTGGAAGCGGAGGGAGAAAGGTGTTTGAGCTTGTGGGTGAGCCCT   480
 481  CCATATACTGCACCAGCAATGACGATCAAGTGGGCATCTGGAGCGGCCCGGCCCCTCAGT   540
 541  GCATTATACCTAACAAATGCACGCCTCCAAATGTGGAAAATGGAATATTGGTATCTGACA   600
 601  ACAGAAGCTTATTTTCCTTAAATGAAGTTGTGGAGTTTAGGTGTCAGCCTGGCTTTGTCA   660
 661  TGAAAGGACCCCGCCGTGTGAAGTGCCAGCCCTGAACAAATGGGAGCCGGAGCTACCAA   720
 721  GCTGCTCCAGGGTATGTCAGCCACCTCCAGATGTCCTGCATGCTGAGCGTACCCAAAGGG   780
 781  ACAAGGACAACTTTTCACCCGGGCAGGAAGTGTTCTACAGCTGTGAGCCCGGCTATGACC   840
 841  TCAGAGGGGCTGCGTCTATGCGCTGCACACCCCAGGGAGACTGGAGCCCTGCAGCCCCCA   900
 901  CATGTGAAGTGAAATCCTGTGATGACTTCATGGGCCAACTTCTTAATGGCCGTGTGCTAT   960
 961  TTCCAGTAAATCTCCAGCTTGGAGCAAAAGTGGATTTTGTTTGTGATGAAGGATTTCAAT  1020
1021  TAAAAGGCAGCTCTGCTAGTTATTGTGTCTTGGCTGGAATGGAAAGCCTTTGGAATAGCA  1080
1081  GTGTTCCAGTGTGTGAACAAATCTTTTGTCCAAGTCCTCCAGTTATTCCTAATGGGAGAC  1140
1141  ACACAGGAAAACCTCTGGAAGTCTTTCCCTTTGGAAAAGCAGTAAATTACACATGCGACC  1200
1201  CCCACCCAGACAGAGGGACGAGCTTCGACCTCATTGGAGAGAGCACCATCCGCTGCACAA  1260
1261  GTGACCCTCAAGGGAATGGGGTTTGGAGCAGCCCTGCCCCTCGCTGTGGAATTCTGGGTC  1320
1321  ACTGTCAAGCCCCAGATCATTTTCTGTTTGCCAAGTTGAAAACCCAAACCAATGCATCTG  1380
1381  ACTTTCCCATTGGGACATCTTTAAAGTACGAATGCCGTCCTGAGTACTACGGGAGGCCAT  1440
1441  TCTCTATCACATGTCTAGATAACCTGGTCTGGTCAAGTCCCAAAGATGTCTGTAAACGTA  1500
1501  AATCATGTAAAACTCCTCCAGATCCAGTGAATGGCATGGTGCATGTGATCACAGACATCC  1560
1561  AGGTTGGATCCAGAATCAACTATTCTTGTACTACAGGGCACCGACTCATTGGTCACTCAT  1620
1621  CTGCTGAATGTATCCTCTCAGGCAATACTGCCCATTGGAGCACGAAGCCGCCAATTTGTC  1680
1681  AACGAATTCCTTGTGGGCTACCCCCCAACCATCGCCAATGGAGATTTCATTAGCACCAACA  1740
1741  GAGAGAATTTTCACTATGGATCAGTGGTGACCTACCGCTGCAATCTTGGAAGCAGAGGGA  1800
1801  GAAAGGTGTTTGAGCTTGTGGGTGAGCCCTCCATATACTGCACCAGCAATGACGATCAAG  1860
1861  TGGGCATCTGGAGCGGCCCCGCCCCTCAGTGCATTATACCTAACAAATGCACGCCTCCAA  1920
1921  ATGTGGAAAATGGAATATTGGTATCTGACAACAGAAGCTTATTTTCCTTAAATGAAGTTG  1980
1981  TGGAGTTTAGGTGTCAGCCTGGCTTTGTCATGAAAGGACCCCGCCGTGTGAAGTGCCAGG  2040
2041  CCCTGAACAAATGGGAGCCAGAGTTACCAAGCTGCTCCAGGGTGTGTCAGCCGCCTCCAG  2100
2101  AAATCCTGCATGGTGAGCATACCCCAAGCCATCAGGACAACTTTTCACCTGGGCAGGAAG  2160
2161  TGTTCTACAGCTGTGAGCCTGGCTATGACCTCAGAGGGGCTGCGTCTCTGCACTGCACAC  2220
2221  CCCAGGGAGACTGGAGCCCTGAAGCCCCGAGATGTGCAGTGAAATCCTGTGATGACTTCT  2280
2281  TGGGTCAACTCCCTCATGGCCGTGTGCTATTTCCACTTAATCTCCAGCTTGGGGCAAAGG  2340
2341  TGTCCTTTGTCTGTGATGAAGGGTTTCGCTTAAAGGGCAGTTCCGTTAGTCATTGTGTCT  2400
2401  TGGTTGGAATGAGAAGCCTTTGGAATAACAGTGTTCCTGTGTGTGAACATATCTTTTGTC  2460
2461  CAAATCCTCCAGCTATCCTTAATGGGAGACACACAGGAACTCCCTCTGGAGATATTCCCT  2520
2521  ATGGAAAAGAAATATCTTACACATGTGACCCCCACCCAGACAGAGGGATGACCTTCAACC  2580
2581  TCATTGGGGAGAGCACCATCCGCTGCACAAGTGACCCTCATGGGAATGGGGTTTGGAGCA  2640
2641  GCCCTGCCCCTCGCTGTGAACTTTCTGTTCGTGCTGGTCACTGTAAAACCCCAGAGCAGT  2700
```

GTGCGTAAACTTGCGTTGGATCTTTCCCATGTCTGCAAAAGCTTCTTATGGAATTATTTC
AAATGTGGGATATGAGAAACCTTTTCTGAAAAGTGTTCGGATAGATGGAT

FIGURE 2.
1098

```
2701   TTCCATTTGCCAGTCCTACGATCCCAATTAATGACTTTGAGTTTCCAGTCGGGACATCTT   2760
2761   TGAATTATGAATGCCGTCCTGGGTATTTTGGGAAAATGTTCTCTATCTCCTGCCTAGAAA   2820
2821   ACTTGGTCTGGTCAAGTGTTGAAGACAACTGTAGACGAAAATCATGTGGACCTCCACCAG   2880
2881   AACCCTTCAATGGAATGGTGCATATAAACACAGATACACAGTTTGGATCAACAGTTAATT   2940
2941   ATTCTTGTAATGAAGGGTTTCGACTCATTGGTTCCCCATCTACTACTTGTCTCGTCTCAG   3000
3001   GCAATAATGTCACATGGGATAAGAAGGCACCTATTTGTGAGATCATATCTTGTGAGCCAC   3060
3061   CTCCAACCATATCCAATGGAGACTTCTACAGCAACAATAGAACATCTTTTCACAATGGAA   3120
3121   CGGTGGTAACTTACCAGTGCCACACTGGACCAGATGGAGAACAGCTGTTTGAGCTTGTGG   3180
3181   GAGAACGGTCAATATATTGCACCAGCAAAGATGATCAAGTTGGTGTTTGGAGCAGCCCTC   3240
3241   CCCCTCGGTGTATTTCTACTAATAAATGCACAGCTCCAGAAGTTGAAAATGCAATTAGAG   3300
3301   TACCAGGAAACAGGAGTTTCTTTTCCCTCACTGAGATCATCAGATTTAGATGTCAGCCCG   3360
3361   GGTTTGTCATGGTAGGGTCCCACACTGTGCAGTGCCAGACCAATGGCAGATGGGGGCCCA   3420
3421   AGCTGCCACACTGCTCCAGGGTGTGTCAGCCGCCTCCAGAAATCCTGCATGGTGAGCATA   3480
3481   CCCTAAGCCATCAGGACAACTTTTCACCTGGGCAGGAAGTGTTCTACAGCTGTGAGCCCA   3540
3541   GCTATGACCTCAGAGGGGCTGCGTCTCTGCACTGCACGCCCAGGGAGACTGGAGCCCTG   3600
3601   AAGCCCCTAGATGTACAGTGAAATCCTGTGATGACTTCCTGGGCCAACTCCCTCATGGCC   3660
3661   GTGTGCTACTTCCACTTAATCTCCAGCTTGGGGCAAAGGTGTCCTTTGTTTGCGATGAAG   3720
3721   GGTTCCGATTAAAAGGCAGGTCTGCTAGTCATTGTGTCTTGGCTGGAATGAAAGCCCTTT   3780
3781   GGAATAGCAGTGTTCCAGTGTGTGAACAAATCTTTTGTCCAAATCCTCCAGCTATCCTTA   3840
3841   ATGGGAGACACACAGGAACTCCCTTTGGAGATATTCCCTATGGAAAAGAAATATCTTACG   3900
3901   CATGCGACACCCACCCAGACAGAGGGATGACCTTCAACCTCATTGGGGAGAGCTCCATCC   3960
3961   GCTGCACAAGTGACCCTCAAGGGAATGGGGTTTGGAGCAGCCCTGCCCCTCGCTGTGAAC   4020
4021   TTTCTGTTCCTGCTGCCTGCCCACATCCACCCAAGATCCAAAACGGGCATTACATTGGAG   4080
4081   GACACGTATCTCTATATCTTCCTGGGATGACAATCAGCTACACTTGTGACCCCGGCTACC   4140
4141   TGTTAGTGGGAAAGGGCTTCATTTTCTGTACAGACCAGGGAATCTGGAGCCAATTGGATC   4200
4201   ATTATTGCAAAGAAGTAAATTGTAGCTTCCCACTGTTTATGAATGGAATCTCGAAGGAGT   4260
4261   TAGAAATGAAAAAAGTATATCACTATGGAGATTATGTGACTTTGAAGTGTGAAGATGGGT   4320
4321   ATACTCTGGAAGGCAGTCCCTGGAGCCAGTGCCAGGCGGATGACAGATGGGACCCTCCTC   4380
4381   TGGCCAAATGTACCTCTCGTGCACATGATGCTCTCATAGTTGGCACTTTATCTGGTACGA   4440
4441   TCTTCTTTATTTTACTCATCATTTTCCTCTCTTGGATAATTCTAAAGCACAGAAAAGGCA   4500
4501   ATAATGCACATGAAAACCCTAAAGAAGTGGCTATCCATTTACATTCTCAAGGAGGCAGCA   4560
4561   GCGTTCATCCCCGAACTCTGCAAACAAATGAAGAAAATAGCAGGGTCCTTCCTTGACAAA   4620
4621   GTACTATACAGCTGAAGAACATCTCGAATACAATTTTGGTGGGAAAGGAGCCAATTGATT   4680
4681   TCAACAGAATCAGATCTGAGCTTCATAAAGTCTTTGAAGTGACTTCACAGAGACGCAGAC   4740
4741   ATGTGCACTTGAAGATGCTGCCCCTTCCCTGGTACCTAGCAAAGCTCCTGCCTCTTTGTG   4800
4801   TGCGTCACTGTGAAACCCCCACCCTTCTGCCTCGTGCTAAACGCACACAGTATCTAGTCA   4860
4861   GGGGAAAAGACTGCATTTAGGAGATAGAAAATAGTTTGGATTACTTAAAGGAATAAGGTG   4920
4921   TTGCCTGGAATTTCTGGTTTGTAAGGTGGTCACTGTTCTTTTTTAAAATATTTGTAATAT   4980
4981   GGAATGGGCTCAGTAAGAAGAGCTTGGAAAATGCAGAAAGTTATGAAAAATAAGTCACTT   5040
5041   ATAATTATGCTACCTACTGATAACCACTCCTAATATTTTGATTCATTTTCTGCCTATCTT   5100
5101   CTTTCACATATGTGTTTTTTTACATACGTACTTTTCCCCCCTTAGTTTGTTTCCTTTTAT   5160
5161   TTTATAGAGCAGAACCCTAGTCTTTTAAACAGTTTAGAGTGAAATATATGCTATATCAGT   5220
5221   TTTTACTTTCTCTAGGGAGAAAAATTAATTTACTAGAAAGGCATGAAATGATCATGGGAA   5280
5281   GAGTGGTTAAGACTACTGAAGAGAAATATTTGGAAAATAAGATTTCGATATCTTCTTTTT   5340
5341   TTTTGAGATGGAGTCTGGCTCTGTCTCCCAGGCTGGAGTGCAGTGGCGTAATCTCGGCTC   5400
5401   ACTGCAACGTCCGCCTCCCG   5420
```

FIGURE 2. Nucleotide sequence of 5.5 kb of human CR1 cDNA. The strand corresponding to the mRNA is shown and base number 1 is the first base after the Eco R1 linker in the most 5' clone. The stop codon is underlined. The 110-bp sequence in the box was found between nucleotides 147 and 148 (*arrow*) and is believed to represent a portion of an intervening sequence as described in the text.

at the 5' end of the cDNA clones and extending 4.7 kb downstream to a stop codon. The coding sequence for CR1 in this library is expected to be 6 kb, based on an estimated 220,000 M_r for the nonglycosylated receptor (20). Thus, these clones span ~80% of the estimated coding sequence.

Clones T49.1 and T55.1 contain coding sequence at their 5' ends, indicating that additional 5' coding and noncoding sequences remain to be identified. In the 3' region, the overlapping clones, T8.2, T43.1 and T87.1, contain the transmembrane and cytoplasmic regions encoded by an identical sequence in each clone. The clone extending most 3', T8.2, contains 807 bp of untranslated sequence without a poly(A) sequence. Clone T8.3 contains a 91-bp deletion of nucleotides 1,406–1,497 and clone T40.1 contains a 9-bp deletion of nucleotides 1,498–1,507 relative to the sequences found in clones T6.1 and T55.1. These deletions occurred in regions having sequences homologous to 5' splice sites and may represent splicing errors in the mRNA. Clones T49.1 and T55.1 contain a 110-bp insertion between nucleotides 147 and 148 of the open reading frame (Fig. 2). This sequence is judged to be a portion of an intron because it did not hybridize to blots of tonsillar poly(A)$^+$ RNA, it contains a 5' splice site (31) (Fig. 2), it is flanked by cDNA sequences in CR1 genomic clones, and it shifts the reading frame. Clone T9.4 contains 0.88 kb of intervening sequence at the 3' end that does not hybridize to blots of tonsillar poly(A)$^+$ RNA (data not shown).

Dot matrix analysis of the nucleotide sequence revealed two types of internal homologies (Fig. 3). The first is represented by the bold, uninterrupted lines that indicate the presence of three tandem, direct, highly homologous repeats of 1.35 kb. These nucleotide sequences encode the LHRs of CR1. The second type of repeat is represented by the dashed parallel lines that indicate regions of lesser homology. These sequences occur every 190–210 nucleotides and encode the SCRs of CR1.

The amino acid sequence deduced from the cDNA sequence is presented in Fig. 4 and the three LHRs, designated LHR-B, LHR-C, and LHR-D, are aligned to demonstrate their homology. LHR-B extends from residue 1 through residue 438, LHR-C corresponds to residues 439–891, and LHR-D extends from residue 892 through 1,341. Residues 451–694 of LHR-C are 99% identical to residues 1–244 of LHR-B, but only 61% identical to the corresponding residues of LHR-D. In contrast, residues 695–891 of LHR-C are 91% identical to residues 1,148–1,341 of LHR-D but only 76% identical to the corresponding region of LHR-B. Thus, LHR-C appears to be a hybrid that comprises sequences most homologous to the first half of LHR-B and the second half of LHR-D. The LHRs are followed by two SCRs that are not repeated, a 25 residue hydrophobic segment and a 43 amino acid COOH-terminal region with no sequence homology to the SCRs (Fig. 4).

The uncloned 1.3 kb of the expected 6 kb of coding sequence is considered to represent a fourth LHR, LHR-A, based on analysis of tryptic peptides of erythrocyte CR1. 10 tryptic peptides have sequences identical to the amino acid sequences derived from the cDNA clones (Table I). In contrast, the sequences of three tryptic peptides are homologous to, but not identical with, the derived amino acid sequences from LHR-B, -C and -D (Table II). Tryptic peptide 56b differs from the homologous sequence in LHR-B and -C by one residue and
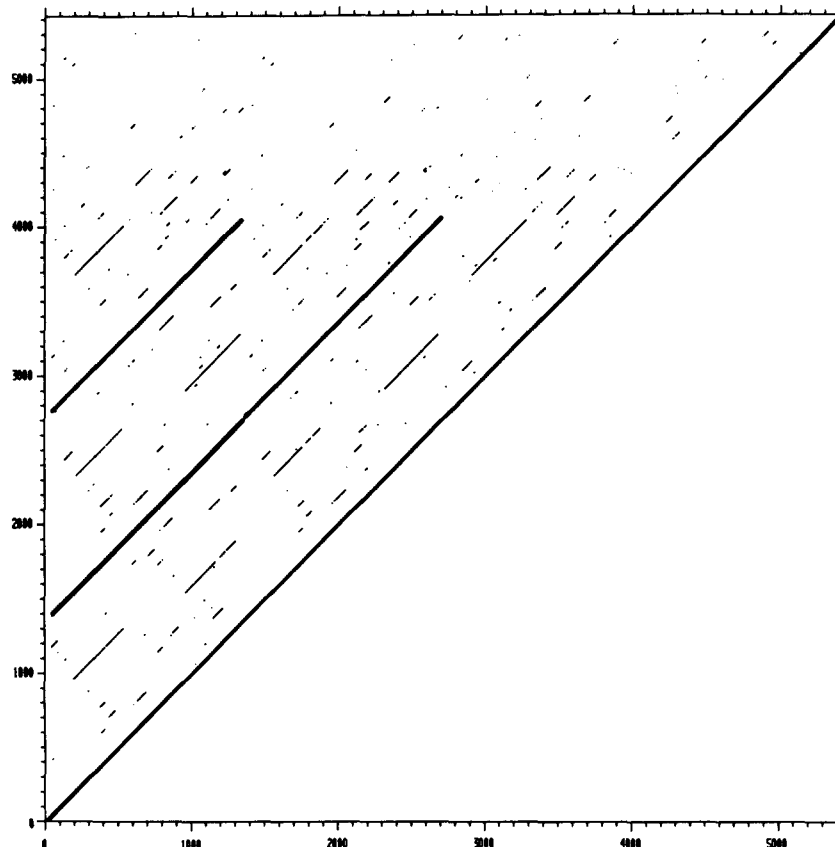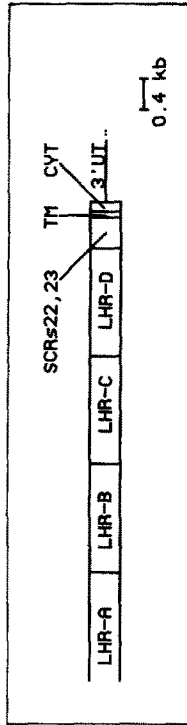
FIGURE 3.   Dot matrix analysis of the nucleotide sequence of 5.5 kb of human CR1 cDNA.
A dot was plotted if there was at least a 40 bp of 90 bp match. The dark line bisecting the
square diagonally indicates the identity of the sequence with itself. The two additional parallel
dark lines 1.35 and 2.7 kb from the line of identity represent two tandem, direct, LHRs of
1.35 kb each. The six lighter, dashed lines between two LHRs correspond to short consensus
repeats of ~0.2 kb. The SCRs extend 0.4 kb beyond the long homologous repeats.

from that in LHR-D by 8 of 18 residues. Peptide 42a differs in 3 of 19 residues
from the homologous sequence at the LHR-B/-C junction and in 13 of 19
residues from that at the LHR-C/-D junction. It is unlikely that peptide 42a is
from the NH$_2$-terminus of LHR-B in which the available derived sequence is
identical to that of the corresponding region of LHR-C (Fig. 4). The recovery
of amino acids from the first three cycles in the sequencing of these two unique
peptides was 723 ± 321 pmol (mean ± SD) and that of the peptides in Table I
corresponding to sequences in LHR-B, LHR-C, or LHR-D was 1,018 ± 545
pmol. Thus, peptides 56b and 42a are probably not derived from the S allotype
(frequency 0.2) in which a fifth LHR may be present. The third peptide in Table
II, peptide 9 (1,052 pmol), differs from the homologous sequences of LHR-B,
-C, and -D at five of nine residues, suggesting either that a more divergent
sequence exists in LHR-A, that there is an SCR in the NH$_2$-terminal region of
CR1 that is not part of a long homologous repeat, or that erythrocyte CR1, from

FIGURE 4. Deduced amino acid sequence of human CR1. Each residue is shown in the one letter code. The residues in the long homologous repeats have been aligned to illustrate their homology. All the residues in LHR-B are shown and a residue is given for LHR-C and LHR-D only where it is different from that in LHR-B. A hydropathy profile is aligned under the COOH-terminus of the protein to illustrate the presumptive transmembrane region.

A stretch of four positively charged residues immediately after the hydrophobic sequence is overlined. The six amino acid sequence with 67% homology to the site of protein kinase C phosphorylation in the epidermal growth factor receptor is underlined. A schematic diagram of the CR1 protein is shown above the sequence. (TM) transmembrane region, (Cyt) cytoplasmic region, (3'UT) 3' untranslated sequence.

TABLE I

CR1 Tryptic Peptides Found in the Derived Amino Acid Sequence

Peptide number	Amino acid sequence	Residue numbers in the derived sequence
66	V D F V CD E G F QL K G S - A	330–345
28	G A A S L - - - - QG - W S P E A P	732–749, 1,185–1,202
49	- - - - - - - - - - - - I F C - N P - A I L	805–826, 1,258–1,279
35	CQA L N K W E P E L P S C S R	228–243, 678–693
41c	D K D N F S P GQE V F Y S C E P G Y D L R	260–281
34b	A V - Y T C D P H P D R G T S F D L I G E S T I R	393–417
44d	V CQ P P P E I LHG	694–704, 1,147–1,157
54d	V F E L VG E P S I YC T S N D D Q V G I W S G P A P Q	152–179, 602–629
57b	Y E C R P E Y YG R P F S	19–31, 469–481
39b	L I GH S S A E C I L S G N A A	85–100

Tryptic peptides from human erythrocyte CR1 found in the derived amino acid sequence. The number ranges in the right-hand column indicate the location of the peptide in the derived amino acid sequence. Each dash in peptides 66, 28 and 49 indicates multiple residues were identified at that cycle. The dash in peptide 34b indicates no residue was identified at that cycle.

TABLE II

CR1 Tryptic Peptides Not Found in the Derived Amino Acid Sequence

Peptide number	Amino acid sequence	Residue numbers in the derived sequence
56b	L K T Q T - A S D F P I GT S LWY	
	L K T QTN A S D F P I GT S L K Y	2–19, 452–469
	PT I P I N D F E F P VGT S L N Y	905–922
42a	CG I L GH CQA P DH F L F AWG Y	
	CG I L GH CQA P DH F L F A K LK	435–453
	C E L S VRAGH C K T P EQF P F A S PT	885–906
9	N S VW E / T G A GH	
	N G VW S S P A P	425–433, 875–883, 1,328–1,336

Tryptic peptides not found in the derived amino acid sequence. The peptides are aligned with the closest matches from the derived amino acid sequence, the positions of which are given to the right of the sequence. Spaces were introduced to maximize the alignment. A dash indicates that no residue was identified at that position. E/T indicates that both E and T were identified at the fifth cycle in peptide 9.

which the tryptic peptides were derived, contains sequences not found in tonsillar CR1.

Each LHR comprises seven 60–70 amino acid SCRs that characterize the family of C3 and C4 binding proteins (C4bp) (Fig. 5A). Maximal homology between the 23 SCRs of CR1 was observed by introducing spaces in the alignment of the sequences (Fig. 5A). Altogether, 29 of the average 65 residues in each repeat are conserved. There are six residues that are present in all SCRs: the four half-cystines that are in similar relative positions suggesting that each may be involved in a critical disulfide linkage, and the tryptophan and the second glycine after the second half-cystine (Fig. 5A). Secondary structure analysis of the sequences between the invariant half-cystines using the algorithm of Chou and Fasman (32) predicted high probability β-turn formation and low probability α-helix formation. Sequence analysis of two CR1 genomic clones, 2.38 (Fig. 5B) and 2.46 (data not shown), indicates that SCR-14 (Fig. 5A) is encoded by a single

FIGURE 5. (A) Alignment of the SCRs of CR1. The repeats are numbered 1–23 from NH₂-terminal to COOH-terminal. Spaces have been introduced to maximize the alignment. The boxes represent invariant residues and the vertical arrows indicate positions of amino acid conservation. A residue is deemed conserved if it, or a conservative substitution, is present in at least half of the SCRs. The horizontal arrow indicates an SCR that was also sequenced from CR1 genomic clone 2.38 and is encoded by a single exon. (B) Restriction map, sequencing strategy, and partial sequence of genomic clone λ2.38. The restriction sites are: (B) Bam HI, (S) Sac I, (E) Eco RV, (K) Kpn I, (P) Pst I. The horizontal arrow indicates direction and extent of sequencing and the vertical arrows indicate the exon/intron boundaries.
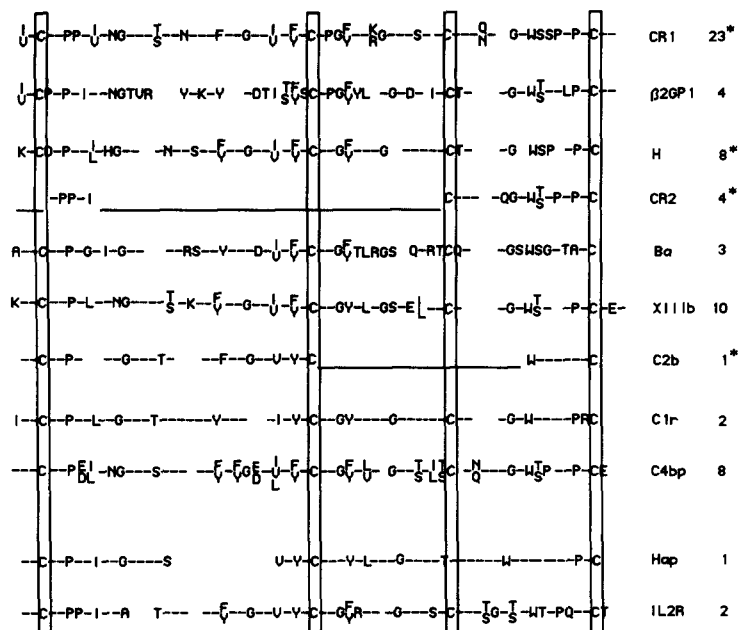
KLICKSTEIN ET AL. 1105

FIGURE 6. Alignment of the consensus sequence of the SCRs of proteins known to have this structure. Spaces were introduced to maximize the alignment. A residue is deemed conserved as in Fig. 5, except for those proteins having only one or two SCRs, in which a residue is conserved if it is present in at least half of the other proteins. The dashes correspond to nonconserved positions. The underlined portions of CR2 and C2b indicate that no sequence information has been published in this region for these proteins. The boxes indicate the invariant half-cystines. The number to the right of the sequence represents the number of SCRs used to generate the consensus sequence. The protein abbreviations and references for the sequence data used to determine the consensus sequences are: (CR1) complement receptor type 1, (H) factor H (36), (C4bp) C4 binding protein (35), (CR2) complement receptor type 2 (28), (Ba) proteolytic fragment of factor B (37), (C2b) proteolytic fragment of C2 (40), (C1r) r subunit of C1 (42), (XIIIb) b subunit of factor XIII (44), (β2GP1) β2 glycoprotein I (41), (Hap) haptoglobin (43), (IL-2-R) the IL-2-receptor (34). Asterisk indicates that incomplete sequence is available.

exon and that the COOH-terminus of SCR-23 corresponds to the end of an exon. Thus, the SCRs of CR1 may be encoded by separate exons as has been shown for the SCRs of factor B (33) and of the IL-2-R (34).

The consensus sequence of the CR1 SCRs is compared with the SCRs of the other members of the superfamily having this characteristic structure (Fig. 6). These members include not only proteins having C3/C4 binding function, CR2 (28), C4bp (35), factor H (36), factor B (37, 38), and C2 (39, 40), but also the proteins not known to have this function, the IL-2-R (34), β_2-glycoprotein I (41), C1r (42), haptoglobin α chain (43), and factor XIIIb (44). The half-cystine residues are invariant in the SCRs of all proteins, except haptoglobin that lacks the third half-cystine. The tryptophan is also invariant with the exception of the fifth SCR in β_2GP1 and two of the repeats in factor XIIIb. Other residues that are conserved but not present in each SCR tend to cluster about the half-cystines. There is only one free thiol group in factor B and C2 (45, 46), and in the SCRs of β_2-glycoprotein I the first half-cystine is disulfide-linked to the third and the second to the fourth (41).

In the derived amino acid sequence of CR1 there are 17 potential sites for N-linked oligosaccharides and all of them are in the extracellular region (Fig. 5 A). Molecular weight differences between CR1 synthesized in the presence and absence of tunicamycin (23) and analysis of glucosamine content (47) suggest the presence of only 6–8 N-linked complex oligosaccharides, indicating that all potential sites are not used. For example, the asparagine at residue 263 of the derived amino acid sequence (Fig. 4) was identified in peptide 41c (Table I), indicating absence of glycosylation at this site. In contrast, the unidentified amino acid in peptide 34b probably corresponds to a glycosylated asparagine at residue 395.

The only nonrepetitive CR1 sequences identified in the 5.5 kb of cDNA are located in the COOH-terminal region. A secondary structure analysis of this region identifies a single 25-residue putative membrane-spanning segment having strong hydrophobic character and high potential for α-helix formation (Fig. 4). This sequence is immediately followed by four positively charged residues, a characteristic of many membrane proteins. The presumed cytoplasmic region of CR1 comprises 43 residues and contains a six-amino-acid sequence, VHPRTL, which is homologous to the sequence VRKRTL, a site of protein kinase C phosphorylation in the EGF receptor and the erbB oncogene product (48, 49). There are no tyrosine residues in the cytoplasmic region of tonsillar CR1.

Discussion

~80% of the primary structure of the F allotype of CR1 has been obtained by sequencing overlapping cDNA clones. The most unusual structural feature of CR1 observed in this analysis is the presence of tandem, direct LHRs of 450 amino acids, four of which are predicted to occur in the F allotype of CR1 that has an estimated polypeptide chain length of 2,000 residues (20, 47). Three of the LHRs have been cloned and sequenced while evidence for the existence of the fourth was provided by the analysis of tryptic peptides. Each LHR is comprised of seven SCRs which are the basic structural elements of other C3/C4 binding proteins. The conservation of the four half-cystines per SCR, the probable involvement of the first and third and the second and fourth half-cystines in disulfide linkages (41) and the presence of conserved amino acids such as proline, glycine and asparagine which are frequently found in β-turns (50) lead to the proposal that an SCR forms a triple loop structure maintained by disulfide linkages (Fig. 7). This role for the half-cystine residues is supported by the finding that mildly trypsin-treated CR1 (47) and factor H (51) migrate as intact molecules when analyzed by SDS-PAGE under non-reducing conditions and as multiple tryptic fragments after reduction. This series of tandemly repeated SCRs is predicted to form an elongated structure (Fig. 7) as has been proposed for factor H and for each subunit of human C4bp (51–53). Electron microscopic studies of the subunits of C4bp have indicated dimensions of 300 × 30Å (53). As each subunit is composed of eight SCRs (35), an individual SCR is calculated to be 38 × 30Å. Assuming that the SCRs of CR1 have similar dimensions and that the F allotype has 30 SCRs, the receptor could extend as much as 1,140Å from the cell membrane. Consistent with this prediction of CR1 structure is the earlier finding that ferritin-labeled antibody bound to CR1 on
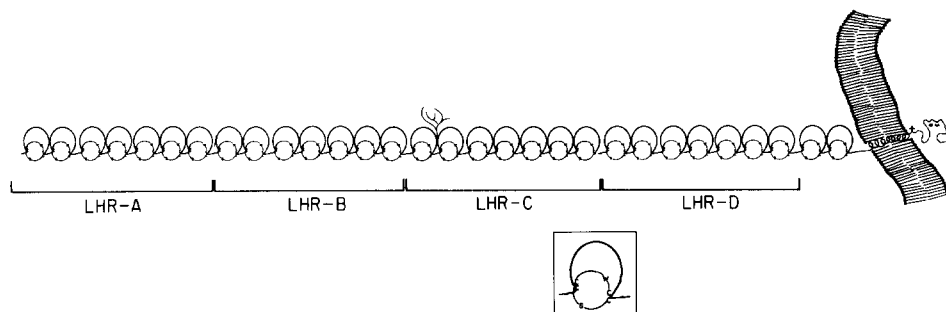
FIGURE 7.   Schematic diagram of the proposed structure of human CR1. The COOH-terminal cytoplasmic region is on the right side of the lipid bilayer. 30 SCRs are arrayed linearly on the extracellular side of the plasma membrane. The brackets indicate the LHRs. The inset is an enlargement of a single SCR to illustrate the triple loop structure.

neutrophils was frequently 500Å from the outer leaflet of the plasma membrane (11). Such an elongated structure of CR1 would facilitate the interaction of receptor-bearing cells with C3b that has covalently bound to relatively inaccessible sites within immune complexes and microbial cell surfaces.

The finding that the SCR is the major, and perhaps only, extracytoplasmic element of CR1 provides structural evidence for a close relationship between the receptor and factor H and C4bp, two plasma proteins that are exclusively or predominantly composed of SCRs (35, 36). CR1 was initially isolated as an erythrocyte membrane protein having factor H–like activity after detergent solubilization (17), and it was found subsequently to have the regulatory functions of factor H and C4bp when residing on the plasma membrane (18). By analysis of the inheritance of structural polymorphisms of CR1, factor H, and C4bp, the genes encoding these three proteins were shown to be linked (54) and the locus for this linkage group and for the structurally related receptor, CR2, have been shown recently by in situ hybridization and by the analysis of somatic cell hybrids to be on the long arm of chromosome 1, band q32 (55). Before the present study, the only evidence for a structural relationship between these proteins was a significant similarity in their amino acid compositions (26). Therefore, the present finding of at least 23 SCRs in CR1 constitutes the direct and formal demonstration of a structural relationship of the receptor with factor H and C4bp (35, 36), proteins with similar functions, and with the Ba and C2b fragments of factor B and C2 (37–40), components that form enzymatic complexes with C3b and C4b, respectively. However, the SCR is also found in several noncomplement proteins (33, 41–44) (Fig. 6), indicating that it does not necessarily represent a C3/C4 binding structure.

Among the proteins having SCRs, CR1 is unique in having organized this basic structural and genetic unit into the higher order structural unit of the LHR. The genetic basis for this organization has not been defined as yet, although analysis of a 14.5-kb Bam HI fragment of genomic DNA that is associated with expression of the S allotype has suggested that at least one repeating genomic unit in CR1 is an extended segment of DNA containing the exons encoding at least five SCRs and their flanking introns (25). These studies have also suggested that the S allele contains an additional copy of this genomic unit compared with

the number present in the F allele. This observation, combined with a tryptic peptide mapping study (56) and the present finding that an LHR represents a peptide of ~40–50 kD predicts the presence in the S allotype (290 kD) of an additional LHR relative to the estimate of four LHRs in the F allotype (M_r 250,000).

In addition to providing evidence for duplication events, the sequences of the LHRs also suggest that conversion events have occurred within the CR1 gene. LHR-B and -D are 67% identical to each other throughout their length, whereas LHR-C is 99% identical to LHR-B in the NH_2-terminal four SCRs and 91% identical to LHR-D in the COOH-terminal three SCRs. This organization could not have occurred by a single recombinational event between identical parental alleles in the origin of this hybrid LHR. Rather, the hybrid LHR may have arisen by gene conversion (57) in which sequences in an LHR-C precursor were replaced by sequences present in LHR-B or LHR-D. The near complete identity and precise alignment of homologous sequences in these LHRs (Fig. 4) also may have been maintained by a mechanism involving gene conversion. Analysis of the extent of homology between intervening sequences of those segments of the CR1 gene encoding the LHRs should determine whether gene conversion or selection based on functional constraints have strictly limited sequence divergence.

Although a previous study suggested that CR1 is monovalent (58), each LHR might represent a single C3b/C4b binding domain, which would make the receptor multivalent and adapted for the binding of complexes bearing multiple molecules of C3b and C4b. Alternatively, distinct LHRs might be responsible for binding C3b and C4b, respectively, providing a structural basis for the combination of factor H and C4bp activities in CR1. Finally, the LHRs of CR1 may represent structural domains that serve to extend CR1 from the plasma membrane, as suggested by the proposed structural model (Fig. 7), and as-yet-unsequenced SCRs at the NH_2-terminal region bind C3b and C4b, as has been found for factor H (51, 59).

Activation of protein kinase C by phorbol esters induces phosphorylation of CR1 in neutrophils, monocytes, and eosinophils (60) and the CR1 cytoplasmic domain of 43 amino acids has a sequence that is homologous to a site that is phosphorylated by protein kinase C in the EGF receptor (48, 49). However, this cytoplasmic sequence, which was found in three independent clones of the tonsillar library, is most likely that of B cell CR1, which is not phosphorylated after activation of protein kinase C (60). Determination of the cytoplasmic region of CR1 from myelomonocytic cells, which is phosphorylated when these cells are treated with phorbol esters, may provide a basis for this differential phosphorylation of the receptor.

Summary

10 overlapping CR1 cDNA clones that span 5.5 kb were isolated from a tonsillar library and sequenced in whole or in part. A single long open reading frame beginning at the 5' end of the clones and extending 4.7 kb downstream to a stop codon was identified. This sequence represents ~80% of the estimated 6 kb of coding sequence for the F allotype of CR1. Three tandem, direct, long

homologous repeats (LHRs) of 450 amino acids were identified. Analysis of the sequences of tryptic peptides provided evidence for a fourth LHR in the F allotype of CR1. Amino acid identity between the LHRs ranged from 70% between the first and third repeats to 99% between the NH_2-terminal 250 amino acids of the first and second repeats. Each LHR comprises seven short consensus repeats (SCRs) of 60–70 amino acids that resemble the SCRs of other C3/C4 binding proteins, such as complement receptor type 2, factors B and H, C4 binding protein, and C2. Two additional SCRs join the LHRs to a single membrane-spanning domain of 25 amino acids; thus, the F allotype of CR1 probably contains at least 30 SCRs, 23 of which have been sequenced. Each SCR is predicted to form a triple loop structure in which the four conserved half-cystines form disulfide linkages. The linear alignment of 30 SCRs as a semi-rigid structure would extend 1,140Å from the plasma membrane and might facilitate the interaction of CR1 with C3b and C4b located within the interstices of immune complexes and microbial cell walls. The COOH-terminal cytoplasmic domain of 43 residues contains a six-amino-acid sequence that is homologous to the sequence in the epidermal growth factor receptor that is phosphorylated by protein kinase C.

References

1. Fearon, D. T. 1980. Identification of the membrane glycoprotein that is the C3b receptor of the human erythrocyte, polymorphonuclear leukocyte, B lymphocyte, and monocyte. *J. Exp. Med.* 152:20.
2. Wilson, J. G., T. F. Tedder, and D. T. Fearon. 1983. Characterization of human T lymphocytes that express the C3b receptor. *J. Immunol.* 131:684.
3. Reynes, M., J. P. Aubert, J. H. M. Cohen, J. Audouin, V. Tricotet, J. Diebold, and M. D. Kazatchkine. 1985. Human follicular cells express CR1, CR2 and CR3 complement receptor antigens. *J. Immunol.* 135:2687.
4. Gelfand, M. C., M. L. Shin, R. B. Nagle, I. Green, and M. M. Frank. 1976. The glomerular complement receptor in immunologically mediated renal glomerular injury. *New Engl. J. Med.* 295:10.
5. Kazatchkine, M. D., D. T. Fearon, M. D. Appay, C. Mandet, and J. Bariety. 1982. Immunohistochemical study of the human glomerular C3b receptor in normal kidney and in seventy five cases of renal diseases. *Clin. Immunol. Immunopathol.* 27:170.
6. Yoon, S. H., and D. T. Fearon. 1985. Characterization of a soluble form of the C3b/C4b receptor (CR1) in human plasma. *J. Immunol.* 134:3332.
7. Fearon, D. T., and W. W. Wong. 1983. Complement ligand-receptor interactions that mediate biological responses. *Annu. Rev. Immunol.* 1:243.
8. Cornacoff, J. B., L. A. Hebert, W. L. Smead, M. E. Van Aman, D. J. Birmingham, and F. J. Waxman. 1983. Primate erythrocyte-immune complex-clearing mechanism. *J. Clin. Invest.* 71:236.
9. Medof, M. E., K. Iida, C. Mold, and V. Nussenzweig. 1982. Unique role of the complement receptor CR1 in the degradation of C3b associated with immune complexes. *J. Exp. Med.* 156:1739.
10. Fearon, D. T., I. Kaneko, and G. Thomson. 1981. Membrane distribution and adsorptive endocytosis by C3b receptors on human polymorphonuclear leukocytes. *J. Exp. Med.* 153:1615.

11. Abrahamson, D. R., and D. T. Fearon. 1983. Endocytosis of the C3b receptor of complement within coated pits in human polymorphonuclear leukocytes and monocytes. *Lab. Invest.* 48:162.

12. Newman, S. L., R. A. Musselson, and P. M. Henson. 1980. Development of functional complement receptors during *in vitro* maturation of human monocytes into macrophages. *J. Immunol.* 125:2236.

13. Wright, S. D., and S. C. Silverstein. 1982. Tumor-promoting phorbol esters stimulate C3b and C3b' receptor-mediated phagocytosis in cultured human monocytes. *J. Exp. Med.* 156:1149.

14. Wright, S. D., L. S. Craigmyle, and S. C. Silverstein. 1983. Fibronectin and serum amyloid P component stimulate C3b and C3bi-mediated phagocytosis in cultured human monocytes. *J. Exp. Med.* 158:1338.

15. Daha, M. R., A. C. Bloem, E. J. E. G. Bast, and R. E. Ballieux. 1983. Increased immunoglobulin production by human peripheral lymphocytes induced by anti-C3 receptor antibodies. *Immunobiology.* 164:227. (Abstr.)

16. Klaus, G. G. B., J. H. Humphrey, A. Kunkl, and D. W. Dongworth. 1980. The follicular dendritic cell: its role in antigen presentation and the generation of immunological memory. *Immunol. Rev.* 53:3.

17. 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.

18. Iida, K., and V. Nussenzweig. 1981. Complement receptor is an inhibitor of the complement cascade. *J. Exp. Med.* 153:1138.

19. Dykman, T. R., J. L. Cole, K. Iida, and J. P. Atkinson. 1983. Polymorphism of the human erythrocyte C3b/C4b receptor. *Proc. Natl. Acad. Sci. USA.* 80:1698.

20. Wong, W. W., J. G. Wilson, and D. T. Fearon. 1983. Genetic regulation of a structural polymorphism of human C3b receptor. *J. Clin. Invest.* 72:685.

21. Dykman, T. R., J. A. Hatch, and J. P. Atkinson. 1984. Polymorphism of the human C3b/C4b receptor. Identification of a third allele and analysis of receptor phenotypes in families and patients with systemic lupus erythematosis. *J. Exp. Med.* 159:691.

22. Dykman, T. R., J. A. Hatch, M. S. Aqua, and J. P. Atkinson. 1985. Polymorphism of the C3b/C4b receptor: Identification of a rare variant. *J. Immunol.* 134:1787.

23. Lublin, D. M., R. C. Griffith, and J. P. Atkinson. 1986. Influence of glycosylation on allelic and cell-specific M_r variation, receptor processing and ligand binding of the human complement C3b/C4b receptor. *J. Biol. Chem.* 261:5736.

24. 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.

25. Wong, W. W., C. A. Kennedy, E. T. Bonaccio, J. G. Wilson, L. B. Klickstein, J. H. Weis, and D. T. Fearon. 1986. Analysis of multiple restriction fragment length polymorphisms of the gene for the human complement receptor type 1: duplication of genomic sequences occurs in association with a high molecular weight receptor allotype. *J. Exp. Med.* 164:1531.

26. Wong, W. W., R. M. Jack, J. A. Smith, C. A. Kennedy, and D. T. Fearon. 1985. Rapid purification of the human C3b/C4b receptor (CR1) by monoclonal antibody affinity chromatography. *J. Immunol. Methods.* 82:303.

27. Maniatis, T., E. F. Fritsch, and J. Sambrook. 1982. Molecular Cloning: A Laboratory Manual. Cold Spring Harbor Laboratory, Cold Spring Harbor, NY. 545 pp.

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 of 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. Sanger, F., S. Nicklen, and A. R. Coulson. 1977. DNA sequencing with chain-terminating inhibitors. *Proc. Natl. Acad. Sci. USA.* 74:5463.

30. Henikoff, S. 1984. Unidirectional digestion with exonuclease III creates targeted breakpoints for DNA sequencing. *Gene.* 28:351.

31. Breathnach, R., C. Benoist, K. O'Hare, F. Gannon, and P. Chambon. 1978. Ovalbumin gene: evidence for a leader sequence in mRNA and DNA sequences at the exon-intron boundaries. *Proc. Natl. Acad. Sci. USA.* 75:4853.

32. Chou, P. Y., and G. D. Fasman. 1974. Prediction of protein conformation. *Biochemistry.* 13:222.

33. Campbell, R. D., and D. R. Bentley. 1985. The structure and genetics of the C2 and factor B genes. *Immunol. Rev.* 87:19.

34. Leonard, W. J., J. M. Depper, M. Kanehisa, M. Kronke, N. J. Peffer, P. B. Svetlik, M. Sullivan, and W. C. Greene. 1985. Structure of the interleukin-2 receptor gene. *Science (Wash. DC).* 230:633.

35. 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.* 230:133.

36. 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.

37. Morley, B. J., and R. D. Campbell. 1984. Internal homologies of the Ba fragment from human complement component factor B, a class III MHC antigen. *EMBO J. (Eur. Mol. Biol. Organ.)* 3:153.

38. Mole, J. E., J. K. Anderson, E. A. Davison, and D. E. Woods. 1984. Complete primary structure for the zymogen of human complement factor B. *J. Biol. Chem.* 259:3407.

39. Bentley, D. R., and R. R. Porter. 1984. Isolation of cDNA clones for human complement component C2. *Proc. Natl. Acad. Sci. USA.* 81:1212.

40. Gagnon, J. 1984. Structure and activation of complement components C2 and factor B. *Philos. Trans. R. Soc. Lond. B Biol. Sci.* 306:301.

41. 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.

42. Leytus, S. P., Kurachi, K., Sakariassen, K. S., and E. W. Davie. 1986. Nucleotide sequence of the cDNA coding for human complement C1r. *Biochemistry.* 25:4855.

43. Kurosky, A., D. R. Barnett, T.-H. Lee, B. Touchstone, R. E. Hay, M. S. Arnott, B. H. Bowman, and W. M. Fitch. 1980. Covalent structure of human haptoglobin: a serine protease homolog. *Proc. Natl. Acad. Sci. USA.* 77:3388.

44. Ichinose, A., B. A. McMullen, K. Fujikawa, and E. W. Davie. 1986. Amino acid sequence of the b subunit of human factor XIII, a protein composed of ten repetitive segments. *Biochemistry.* 25:4633.

45. Christie, D. L., and J. Gagnon. 1982. Isolation, characterization and N-terminal sequences of the CNBr-cleavage peptides from human complement factor B. *Biochem. J.* 201:555.

46. Parkes, C., J. Gagnon, and M. A. Kerr. 1983. The reaction of iodine and thiol-blocking reagents with human complement component C2 and factor B. *Biochem. J.* 213:201.

47. Sim, R. B. 1985. Large-scale isolation of complement receptor type 1 (CR1) from human erythrocytes. Proteolytic fragmentation studies. *Biochem. J.* 232:883.

48. Hunter, T., N. Ling, and J. A. Cooper. 1984. Protein kinase C phosphorylation of

the EGF receptor at a threonine residue close to the cytoplasmic face of the membrane. *Nature (Lond.).* 311:480.

49. Davis, R. J., and M. P. Czech. 1985. Tumor-promoting phorbol diesters cause the phosphorylation of epidermal growth factor receptors in normal human fibroblasts at threonine-654. *Proc. Natl. Acad. Sci. USA.* 82:1974.

50. Rose, G. D., Gierasch, L. M., and J. A. Smith. 1985. Turns in peptides and proteins. *Adv. Protein Chem.* 37:1.

51. Sim, R. B., and R. G. DiScipio. 1982. Purification and structural studies on the complement-system control protein $\beta_1 H$ (factor H). *Biochem. J.* 205:285.

52. Whaley, K., and S. Ruddy. 1976. Modulation of the alternative complement pathway by $\beta 1 H$ globulin. *J. Exp. Med.* 144:1147.

53. Dahlback, B., C. A. Smith, and H. J. Muller-Eberhard. 1983. Visualization of human C4bp-binding protein and its complexes with vitamin K-dependent protein S and complement protein C4b. *Proc. Natl. Acad. Sci. USA.* 80:3461.

54. Rodriguez de Cordoba, S., D. M. Lublin, P. Rubinstein, and J. P. Atkinson. 1985. Human genes for the three complement components that regulate the activation of C3 are tightly linked. *J. Exp. Med.* 161:1189.

55. Weis, J. H., C. C. Morton, G. A. P. Bruns, J. J. Weis, L. B. Klickstein, W. W. Wong, and D. T. Fearon. 1987. Definition of a complement receptor locus: the genes encoding the C3b receptor and the C3d/Epstein-Barr virus receptor map to human chromosome 1. *J. Immunol.* 138:312.

56. Nickells, M. W., T. Seya, V. M. Holers, and J. P. Atkinson. 1986. Analysis of C3b/C4b receptor (CR1) polymorphic variants by tryptic peptide mapping. *Mol. Immunol.* 23:661.

57. Atchison, M., and M. Adesnik. 1986. Gene conversion in a cytochrome P-450 gene family. *Proc. Natl. Acad. Sci. USA.* 83:2300.

58. Wilson, J. G., Wong, W. W., Schur, P. H., and D. T. Fearon. 1982. Mode of inheritance of decreased C3b receptors on erythrocytes of patients with systemic lupus erythematosus. *N. Engl. J. Med.* 307:981.

59. Alsenz, J., J. D. Lambris, T. F. Schulz, and M. P. Dierich. 1984. Localization of the complement-component-C3b-binding site and the cofactor activity for factor I in the 38kDa tryptic fragment of factor H. *Biochem. J.* 224:389.

60. Changelian, P. S., and D. T. Fearon. 1986. Tissue-specific phosphorylation of complement receptors CR1 and CR2. *J. Exp. Med.* 163:101.