

ANALYSIS OF MULTIPLE RESTRICTION FRAGMENT
LENGTH POLYMORPHISMS OF THE GENE FOR THE
HUMAN COMPLEMENT RECEPTOR TYPE I

Duplication of Genomic Sequences Occurs in Association with a
High Molecular Mass Receptor Allotype

BY WINNIE W. WONG, CHRISTINE A. KENNEDY,
ERMELINDA T. BONACCIO, JAMES G. WILSON, LLOYD B. KLINKSTEIN,
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 human receptor for complement fragments C3b and C4b (complement receptor type I, CR1) is found on erythrocytes, peripheral blood leukocytes, glomerular podocytes, and follicular dendritic cells and has an important role in the processing and clearance of C3b/C4b-coated microorganisms and immune complexes (1, 2). CR1 also acts as a cofactor for the factor I-mediated cleavage of C3b and C4b (3, 4), a function that might be especially relevant to the finding of a soluble form of CR1 in plasma (5). This regulatory capacity also suggested that CR1 was related to factor H and C4-binding protein (C4BP),¹ and these three proteins were subsequently shown to share a repeating 60-amino-acid motif in their primary sequence and to be genetically linked (6-9).

An unusual structural characteristic of the CR1 polypeptide chain is the existence of four allotypic forms differing in M_r by >100,000 without apparent major functional differences (10-14). The two most common allotypes, termed F and S, have M_r of ~250,000 and 290,000, respectively, when analyzed by SDS-PAGE under reducing conditions, and M_r of 235,000 and 275,000 after removal of N-linked oligosaccharides (10). Biosynthetic studies (15) have confirmed the inability of glycosylation differences to account for the allotypically different forms of CR1. A third rarer allotype designated F' has an M_r of 210,000 and a fourth variant has also been described that apparently is larger than the S allotype (12, 13, 16).

The recent identification and sequencing of cDNA clones of CR1 has revealed the presence of long homologous repeats as another unusual structural characteristic (17). Each homologous repeat is ~1.3-1.4 kb, encoding peptide sequences

This work was supported by grants AI-22833, AI-23401, and AM-36797 from the National Institutes of Health and the Arthritis Foundation Biomedical Research Center Grant. W. Wong is a research fellow of the Arthritis Foundation. J. Weis is supported by a grant-in-aid from the American Heart Association and grant IM-412 from the American Cancer Society. L. Klickstein is supported in part by NIH training grant GMO7753-06.

¹ Abbreviations used in this paper: C4BP, C4-binding protein; RFLP, restriction fragment length polymorphism.

of 46,000–52,000, and comparison of two repeats indicated 87% nucleotide and 83% amino acid identity (6). The existence of the long homologous repeats suggests that a possible mechanism for the generation of the CR1 allotypes may be the addition or deletion of these structural units. The present study addresses this possibility by defining restriction fragment length polymorphisms (RFLP) of the CR1 gene that are associated with the CR1 allotypes.

Materials and Methods

Southern Blot Analysis. Genomic DNA was isolated from peripheral blood leukocytes by proteinase K digestion followed by extraction with phenol/chloroform/isoamyl alcohol (18). 15–20 μ g of DNA from each donor were digested overnight by various restriction enzymes according to conditions recommended by the vendor (New England Biolabs, Beverly, MA). Digested samples were electrophoresed for 24–48 h in 0.8% agarose in Tris-acetate buffer (18), were transferred onto Zetabind nylon membranes (AMF-Cuno, Meriden, CT), and hybridized with 32 P-labeled probes for 36 h at 43°C in 50% formamide and 5 \times SSC. The blots were washed at 65°C in 0.2 \times SSC and exposed on Kodak XAR film at –70°C.

Analysis of RNA by Blot Hybridization. Total RNA was isolated from PBMC by extraction with guanidine-HCl and precipitation in ethanol (19). 30 μ g of RNA was denatured by heating to 55°C in 33% formamide and 6.2% formaldehyde, and was electrophoresed in 0.8% agarose in MOPS buffer containing 3.7% formaldehyde (20). RNA markers were purchased from Bethesda Research Laboratories (Bethesda, MD). The RNA was transferred onto Zetabind nylon membranes, hybridized with 32 P-labeled probe, and the blots were washed in 0.2 \times SSC at 65°C.

Genomic Libraries. DNA from selected donors was exhaustively digested with Bam HI and size-selected on a 10–40% sucrose gradient (18). The fractions containing DNA fragments of 12–18 kb were pooled, precipitated in 70% ethanol, and ligated to EMBL-3 arms (Vector Cloning Systems, San Diego, CA). The DNA was packaged and plated on *Escherichia coli* strain P2392. Positive phage clones were selected by hybridization of duplicate nitrocellulose filters with 32 P-labeled CR1 cDNA probes.

Probes. The cDNA probes, CR1-1 and CR1-2, are Eco RI fragments of CR1 cDNA clone λ T8.3 that had been subcloned in pBR327 (17). CR1-4 is the 381-bp Eco RI fragment between CR1-1 and CR1-2 that was obtained from an overlapping phage clone, λ T6.1. CR1-1 and CR1-2 crosshybridize to each other (17) but neither hybridizes to CR1-4 (Fig. 1A).

The genomic probe, GB1R1, is a 1.85-kb Eco RI fragment that contains neither human repeats nor CR1-1- and CR1-4-hybridizing sequences. This probe was obtained from λ GSB4.2, a 14.5-kb Bam HI fragment of human DNA that had been cloned into EMBL-3 and selected by hybridization to CR1-1 and CR1-2 (Fig. 1B).

The genomic probe, GB2PE, is a PvuII–Bam HI fragment of ~400 bp, which also contains no human repeats or CR1-1- and CR1-4-hybridizing sequences. This probe was derived from an EMBL-3 clone, λ GSB16.1, which contains a CR1-1⁺, CR1-4⁺, 14.5-kb Bam HI fragment of DNA from an individual having the FS phenotype (Fig. 1B). All probes were labeled with [32 P]dCTP by nick translation before use in hybridizations.

CR1 Protein Analysis. The structural phenotypes of CR1 of individual donors were assessed by immunoprecipitation of CR1 protein from erythrocyte membranes by YZ-1 monoclonal anti-CR1 coupled to Sepharose and analysis of the eluted samples by SDS-PAGE and silver staining (10, 21).

Statistical Analysis. The correlations among protein and genomic polymorphisms were assessed by χ^2 analyses.

Results

The presence of polymorphisms in the CR1 gene was sought by hybridization of CR1 cDNA probes to Southern blots of human genomic DNA fragments

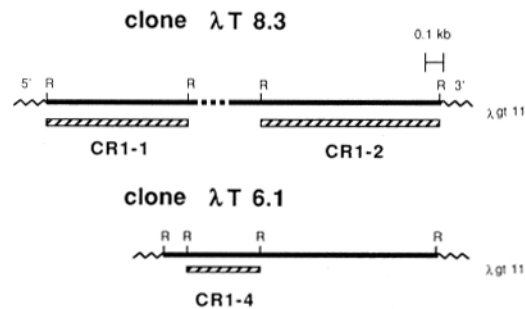
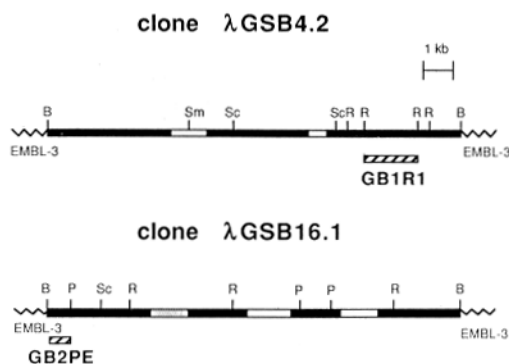
A cDNA probes:**B genomic probes:**

FIGURE 1. CR1 cDNA and genomic probes. (A) The black bars denote the overlapping cDNA clones, λ T8.3 and λ T6.1. CR1-4 is derived from the λ T6.1 clone because λ 8.3 contains a deletion of 91 bp in the CR1-4 region. The crosshatched bars represent the Eco RI fragments that have been subcloned into pBR322 or pBR327. (B) The black bars represent the 14.5-kb Bam HI genomic fragments comprising the inserts in clones λ GSB4.2 and λ GSB16.1. The white boxes designate the regions that hybridize to CR1-1 and CR1-2, and the stippled box designates the region that hybridizes to CR1-4. The crosshatched bars represent the Eco RI fragment, GB1R1, and the Pvu II-Bam HI fragment, GB2PE, which were subcloned into pBR322. The restriction sites are designated as follows: R, Eco RI; B, Bam HI; P, Pvu II; Sc, Sac I and Sm, Sma I.

generated by digestion with a panel of restriction enzymes. Although the cDNA probes, CR1-1, CR1-2, and CR1-4, constitute only 2 kb of the \sim 9 kb of CR1 mRNA (17), these probes crosshybridize to the sequences present in four to five long homologous repeats of \sim 1.4 kb each that constitute CR1 coding sequence (reference 17 and L. B. Klickstein, unpublished observations). Therefore, the use of these probes allowed for the analysis of 100–150 kb of genomic DNA. Frequently occurring polymorphisms were identified when genomic DNA was digested with Bam HI, Sac I, Eco RV, Hind III, Pvu II, and Eco RI.

Genomic Polymorphisms that Correlate with Protein Polymorphisms of CR1. Genomic DNA from six unrelated individuals having the FF, FS, or SS CR1 structural phenotypes was digested by Bam HI and analyzed by Southern blots

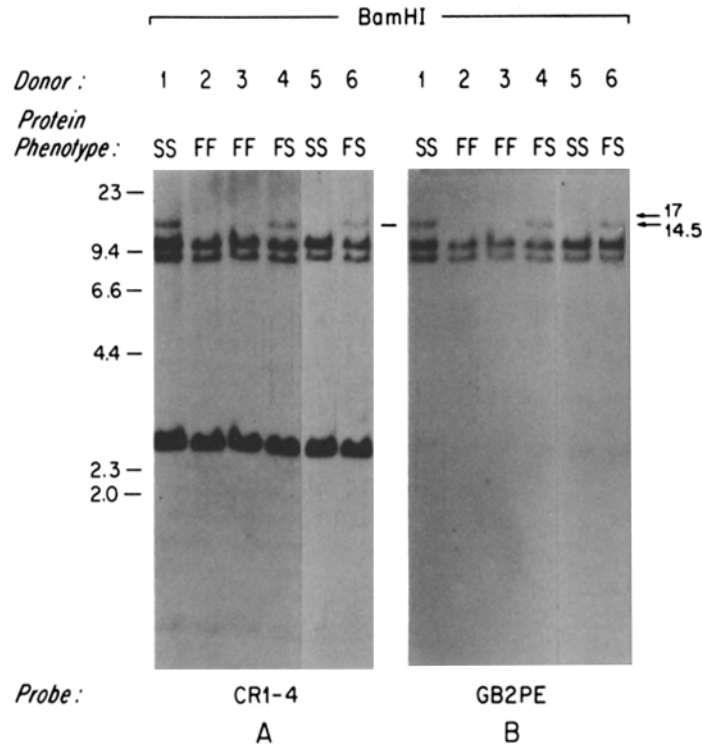


FIGURE 2. Autoradiographs of the Southern blots in which CR1-4 and GB2PE were hybridized to the Bam HI digests of the DNA of donors with different CR1 protein phenotypes. A Hind III digest of λ DNA was used as a marker and the positions of these fragments are designated on the left in kilobases. The position of the 14.5-kb restriction fragment seen in this blot is indicated at the right. The 17-kb restriction fragment was seen in the Southern blot shown in Fig. 3.

using CR1-4 as the probe. In three of the four individuals having the S allotype, a 14.5-kb fragment was identified that was absent in individuals homozygous for the F allotype (Fig. 2A). We could find no fragment present in individuals having the F allotype that was absent in the SS individuals, therefore, a restriction fragment allelic to the CR1-4⁺, 14.5-kb fragment could not be identified. In one Black individual having the SS phenotype, the CR1-4⁺, 14.5-kb Bam HI fragment was not seen (Fig. 2A, donor 5).

The extent to which the presence of the CR1-4⁺ 14.5-kb fragment correlated with the occurrence of the S allotype was determined by performing CR1 protein and DNA analyses on 58 unrelated normal individuals, excluding Black donors. No individual homozygous for the F allotype exhibited the CR1-4⁺, 14.5-kb fragment, whereas we saw this fragment in 20 of 21 individuals having the S allotype ($\chi^2 = 53.79$, $p < 0.0005$). Neither of the two persons having the F' allotype were positive for this 14.5-kb Bam HI fragment (Table I).

An intervening sequence probe for the S allotype-associated restriction fragment was obtained by digestion of the DNA from an FS donor with Bam HI and by cloning of the size-selected fragments into EMBL-3. Clones were selected by hybridization with CR1-4 and CR1-1, and one clone hybridizing with both

TABLE I
Correlation of S Phenotype with the 14.5-kb Bam HI RFLP

CR1 protein phenotype (n)	Presence of 14.5-kb fragment		
	CR1-4 ⁺	GB1R1 ⁺	CR1-4 ⁺ /GB1R1 ⁺
FF (35)	0	4	0
FS (17)	16	0	2
FF' (2)	0	0	0
SS (4)	4	0	0

probes and containing an insert of 14.5 kb, termed λ GSB16.1, was restriction mapped (Fig. 1B). The Pvu II–Bam HI fragment, GB2PE, which contained neither human repeats nor CR1-1- and CR1-4-like sequences, hybridized to the 14.5-kb Bam HI fragment characteristic of an S allele and to nonallelic fragments of 9.9 and 8.3 kb in individuals of both SS and FF phenotypes (Fig. 2B). Therefore, a noncoding genomic sequence that may be repeated once in the F allele of CR1 is repeated an additional time in the S allele.

The same blots were probed with CR1-1, which also identifies a 14.5-kb Bam HI fragment in 20 of the 21 individuals expressing the S allotype (Fig. 3A and Table I). In addition, 4 of the 35 FF individuals also had a CR1-1⁺ 14.5-kb fragment, in contrast to the absence of a fragment of this size hybridizing with CR1-4 in these persons (Figs. 2A and 3A, donor 3; Table I). These findings suggested that CR1-1, in addition to identifying the 14.5-kb restriction fragment that was associated with the S allotype, also identified a comigrating Bam HI fragment that was CR1-4⁻ and not associated with the S allotype. In the same Black individual having the SS phenotype and lacking the CR1-4⁺, 14.5-kb Bam HI fragment, CR1-1 identified a fragment of only 13 kb (Fig. 3A).

We showed the existence of this additional polymorphic Bam HI fragment by the isolation of a clone termed λ GSB4.2, which contained a CR1-1⁺, CR1-4⁻, 14.5-kb fragment, from a human genomic library constructed in EMBL-3 from size-selected, Bam HI–digested DNA. An Eco RI fragment of 1.85 kb, GB1R1, that did not hybridize with CR1-1, CR1-4, or human repeat sequences, was subcloned from λ GSB4.2 (Fig. 1B). When used as a probe in Southern blot analyses, GB1R1 hybridized to the 14.5-kb Bam HI fragment previously identified by CR1-1 in the four FF donors (Table I; Fig. 3, A and B, donor 3) and in two of the FS donors (Table I), and did not identify the 14.5-kb Bam HI fragment characteristic of the SS individuals (Fig. 3B, donor 1; Table I), confirming its capacity to distinguish between the two comigrating Bam HI fragments. However, an individual could exhibit both Bam HI polymorphisms since two of the FS donors had both CR1-4⁺ and GB1R1⁺ 14.5-kb Bam HI fragments (Table I). Also of interest was the finding that this intervening sequence probe, which did not hybridize to any mRNA species on Northern blot analysis of RNA from mononuclear cells nor any CR1 cDNA clones, identified two additional fragments of 17 and 4.6 kb in all individuals (Fig. 3B).

A Southern blot of Bam HI–digested DNA of the Black SS donor who exhibited a CR1-1⁺, 13kb Bam HI fragment was probed with GB1R1. We identified three fragments of 17, 13, and 4.6 kb (Fig. 3B, donor 5). Thus, the

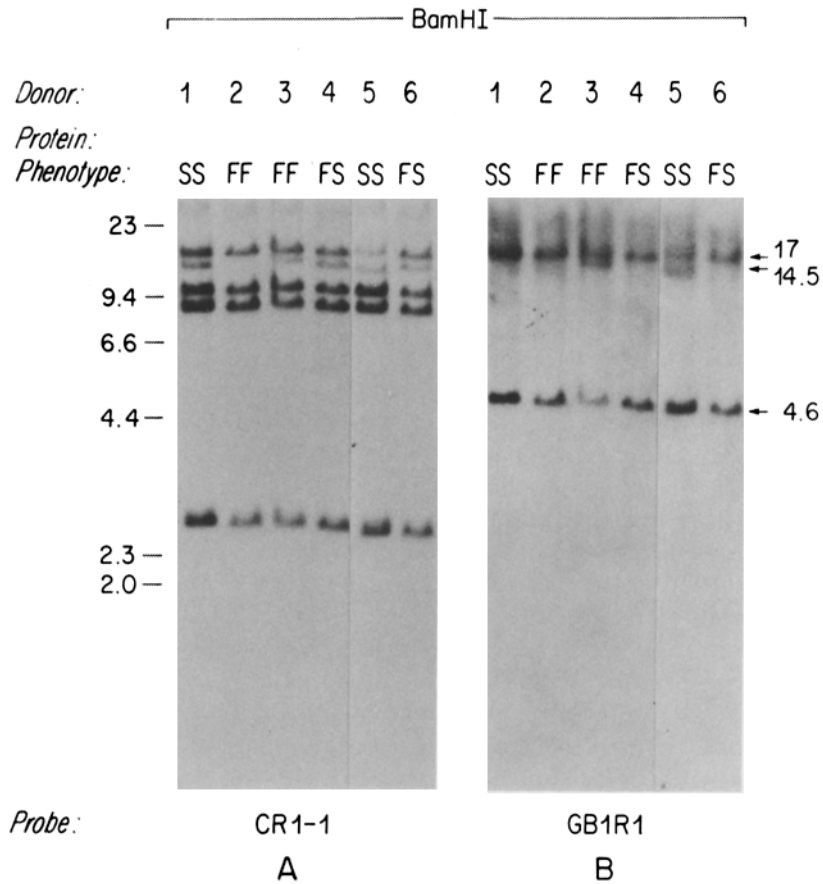


FIGURE 3. Autoradiographs of the Southern blots in which CR1-1 and GB1R1 were hybridized to the Bam HI digests of the DNA of donors with different CR1 protein phenotypes. The same blots used in Fig. 2 were probed with CR1-1 (A) and GB2PE (B) and washed as described. Donors 1-6 correspond to the same individuals shown in Fig. 2. A Hind III digest of λ DNA was used as a marker and the positions of these fragments are designated on the left in kilobases. The positions of the 17- and the 14.5-kb restriction fragments are indicated at the right.

13-kb band seen in this donor was a variant of the GB1R1⁺, 14.5-kb Bam HI restriction fragment seen in 6 of the 57 normal donors discussed above (Table I).

To obtain further evidence for the presence of additional CR1 genomic sequences in the S allele, DNA from individuals having this allele was digested with other restriction enzymes, including Apa I, Bcl I, Bgl I, Bgl II, Dra I, Eco 0109, Eco RI, Eco RV, Hind III, Pst I, Pvu II, Sac I, Taq I, and Xmn I and the Southern blots were probed with CR1-4. Only Sac I generated a polymorphic fragment correlating with the occurrence of the CR1-4⁺, 14.5-kb Bam HI fragment. RFLP analyses of DNA from 57 individuals showed that all 20 donors of the FS and SS CR1 phenotype who exhibited a CR1-4⁺, 14.5-kb Bam HI fragment also exhibited a CR1-4⁺, 19-kb Sac I fragment that was absent from the DNA of the other 35 FF donors. There was no corresponding loss of any

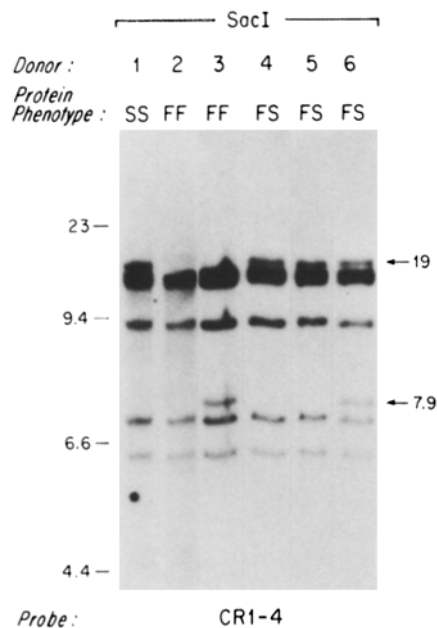


FIGURE 4. Autoradiograph of the Southern blot in which CR1-4 was hybridized to the Sac I-digested DNA of donors with the Bam HI polymorphisms. Donors 1-4 were identical to donors 1-4 in Figs. 2 and 3. Donor 6 was an FS individual having both CR1-4⁺ and GB1R1⁺ 14.5-kb Bam HI fragments. A Hind III digest of λ DNA was used as a marker and the positions of these fragments are designated on the left in kilobases. The positions of the 19- and 7.9-kb polymorphic fragments are shown at the right.

fragments in the FF individuals; therefore, an allelic Sac I fragment could not be identified (Fig. 4). Sac I also generated a CR1-1⁺, CR1-4⁺, 7.9-kb restriction fragment that was present only in the DNA of the four FF donors and two FS donors who exhibited the CR1-1⁺, GB1R1⁺, 14.5-kb Bam HI fragment (Table I; represented by donors 3 and 6 in Fig. 4). The ability to discriminate the two Bam HI polymorphisms with the use of another restriction enzyme confirms that each polymorphic region arose by mechanisms other than a single base pair change and is consistent with a polymorphism having arisen by an insertion or conversion of genomic sequences.

After these RFLPs that correlated with the 290,000 M_r S allotype were identified, a genomic polymorphism associated with the 210,000 M_r F' allotype was sought. We saw a CR1-1⁺, 18-kb Eco RV fragment in two unrelated donors who expressed the F and F' allotypes (Fig. 5). In three generations of the family of one of these FF' individuals, the 18-kb Eco RV fragment cosegregated with the F' allotype (Fig. 5). This Eco RV restriction fragment was absent in 22 other donors having the FF and FS phenotypes.

mRNA polymorphisms that correlate with protein polymorphisms. RNA was isolated from PBMC of individuals expressing the FF', FF, FS, and SS phenotypes and was analyzed by electrophoresis and blot hybridization with the CR1 cDNA probes, CR1-1 and CR1-4 (Fig. 6). We detected two mRNA species of 7.9 and 9.2 kb in RNA from the FF donor and we also saw two transcripts in an SS donor, although these were of 9.2 and 10.7 kb. Only three species of CR1

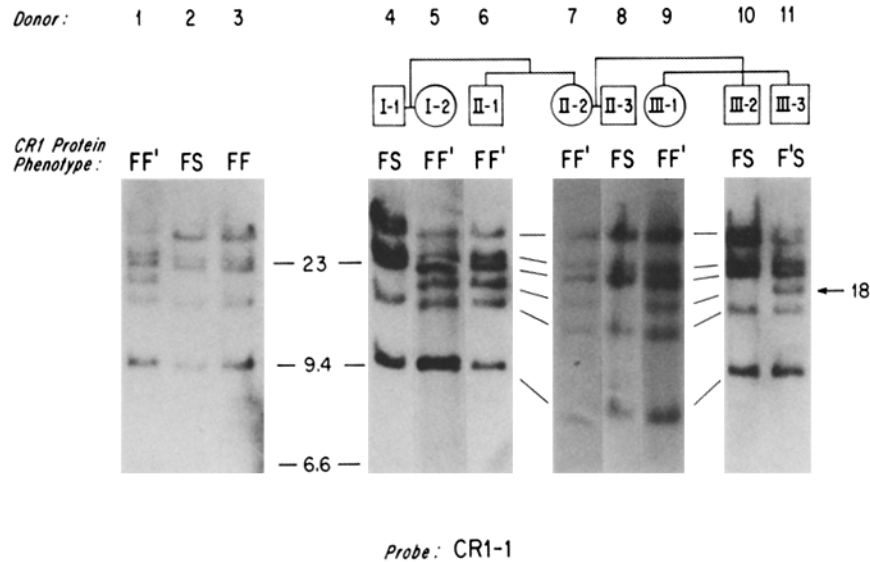


FIGURE 5. Autoradiographs of the Southern blots in which CR1-1 was hybridized to the Eco RV fragments of DNA from donors with different CR1 structural allotypes. The *left* panel represents samples from three unrelated donors having FF', FF, and FS phenotypes. The three *right* panels represent three generations in a family of another unrelated FF' individual. The three generations are designated by Roman numerals and the members of each generation are designated in Arabic numerals. A Hind III digest of λ DNA was used as a marker and the positions of these fragments (23, 9.4, and 6.6 kb) are designated. The position of the 18-kb restriction fragment characteristic of the F' allotype is shown at the right.

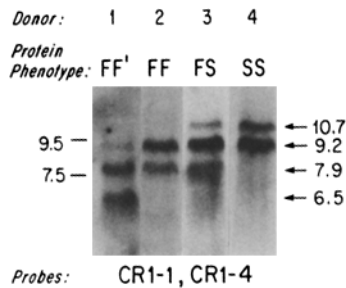


FIGURE 6. Autoradiograph of the Northern blot in which CR1 cDNA probes were hybridized to mRNA from individuals having different CR1 structural phenotypes. Total RNA isolated from PBMC of four unrelated individuals were electrophoresed in a 0.8% agarose gel for 24 h. The positions of the RNA standards are designated in kilobases on the left and the relative sizes of the CR1 mRNA bands are shown on the right.

mRNA could be distinguished in the heterozygous FS donor, and the 9.2-kb species associated with each allele in this individual could not be separated even when electrophoresis was performed for 48 h rather than 24 h hours. Similarly, the FF' heterozygote exhibited only three distinct transcripts, the 6.5-kb mRNA being unique to the F' allele. A donor homozygous for the F' allele was not available for analysis to determine if this allele also contributed an additional transcript, possibly of 7.9 kb, that comigrated with the smaller species of mRNA associated with the F allele. In summary, CR1 transcripts increased in length in

TABLE II
Distribution of the Hind III RFLP among Individuals with Different CR1 Phenotypes

CR1 protein phenotype (n)	Hind III RFLP		
	7.4 kb ⁺	7.4 kb ⁺ /6.9 kb ⁺	6.9 kb ⁺
FF (29)	13	8	8
FS (15)	11	4	0
SS (4)	4	0	0

association with the increased size of the CR1 allotypes. This unit of increment appeared to be 1.3–1.4 kb, although this conclusion is obscured by the occurrence of two CR1 transcripts for each allele that also differed by ~1.4 kb.

Other Genomic Polymorphisms in Linkage Disequilibrium with CR1 Protein Polymorphisms. Recently, allelic CR1-1⁺ Hind III fragments of 7.4 and 6.9 kb have been shown (16) to be associated with high and low numbers, respectively, of erythrocyte CR1. In the present study, among the eight individuals homozygous for the 6.9-kb Hind III fragment, all were found to express only the FF phenotype (Table II). The association between this Hind III restriction fragment and the F allotype was significant when subjected to χ^2 analysis ($\chi^2 = 8.79$, $p < 0.025$). Conversely, all four individuals expressing the SS phenotype were homozygous for the 7.4-kb Hind III fragment and no individual having the S allotype was homozygous for the 6.9-kb Hind III fragment (Table II), suggesting that the S allotype is in linkage disequilibrium with the 7.4-kb Hind III fragment.

An additional DNA polymorphism, represented by the presence or absence of a CR1-4⁺, 5.8-kb Pvu II fragment, also exhibited linkage disequilibrium with the S allotype (Fig. 7). All 12 individuals expressing the S allotype had this Pvu II fragment, whereas eight of 16 individuals having the FF phenotype lacked this fragment (Table III). Thus, the S allotype is significantly associated with the presence of a CR1-4⁺, 5.8-kb Pvu II fragment ($\chi^2 = 8.4$, $p < 0.005$), although the F allotype may be associated with either the presence or absence of this restriction fragment. We saw no linkage disequilibrium between the polymorphisms defined by Hind III and Pvu II (data not shown; $\chi^2 = 0.483$, $p > 0.3$). Among the 28 individuals analyzed, an additional CR1-4⁺ fragment of 4 kb was also present in one FF individual, indicating the existence of another rarer polymorphism revealed by the use of this restriction enzyme (Fig. 7, donor 5).

A Genomic Polymorphism Identified by CR1 cDNA not in Linkage Disequilibrium with CR1 Allotypes. Another commonly occurring RFLP involving allelic CR1-1⁺ fragments of 20 and 16 kb was seen with the use of Eco RI. In a family in which each parent exhibited only the 20 or the 16 kb fragment, both children exhibited both restriction fragments (Fig. 8). All members of this family lacked the CR1-4⁺ 14.5-kb Bam HI fragments (Fig. 8) and expressed the FF phenotype (data not shown), indicating that an F allele could be associated with either the 20-kb or the 16-kb Eco RI fragment. A lack of correlation between the Eco RI polymorphism and CR1 protein allotypes was confirmed by analysis of 57 individuals, in whom all three possible genotypes for the Eco RI RFLP were

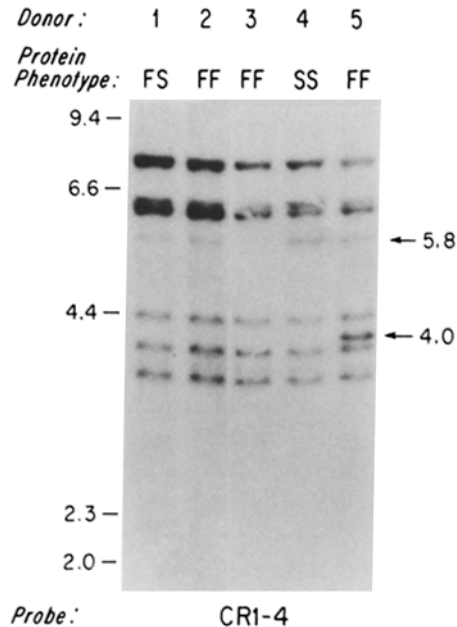


FIGURE 7. Autoradiograph of the Southern blot in which CR1-4 was hybridized to the Pvu II fragments of DNA from donors with different CR1 structural allotypes. A Hind III digest of λ DNA was used as a marker and the positions of these fragments are designated on the left in kilobases. The positions of the 5.8-kb and the 4.0-kb restriction fragments are shown on the right.

TABLE III
 Distribution of the Pvu II RFLP among Individuals with Different CR1 Phenotypes

CR1 protein phenotype (n)	Pvu II RFLP	
	5.8 kb ⁺	5.8 kb ⁻
FF (16)	8	8
FS (11)	11	0
SS (1)	1	0

found among individuals having the FF, FS, or SS phenotypes (Table IV; $\chi^2 = 1.52$, $p > 0.2$).

Although this Eco RI RFLP was not in linkage disequilibrium with the Bam HI RFLP that is associated with the S allotype, an analysis of an informative family showed the cosegregation of the 16-kb Eco RI fragment and the CR1-4⁺, 14.5-kb Bam HI fragment (Fig. 9). Thus, the Eco RI polymorphism is linked to the CR1 structural gene but may be relatively distant from the locus determining the structural phenotype. Frequencies of the alleles bearing the 20- or 16-kb Eco RI fragments were 54.5 and 45.5%, respectively, in the normal population.

Discussion

A genetic basis for the size polymorphism of CR1 was sought by analyzing CR1 genomic polymorphisms that were detected by a panel of restriction enzymes

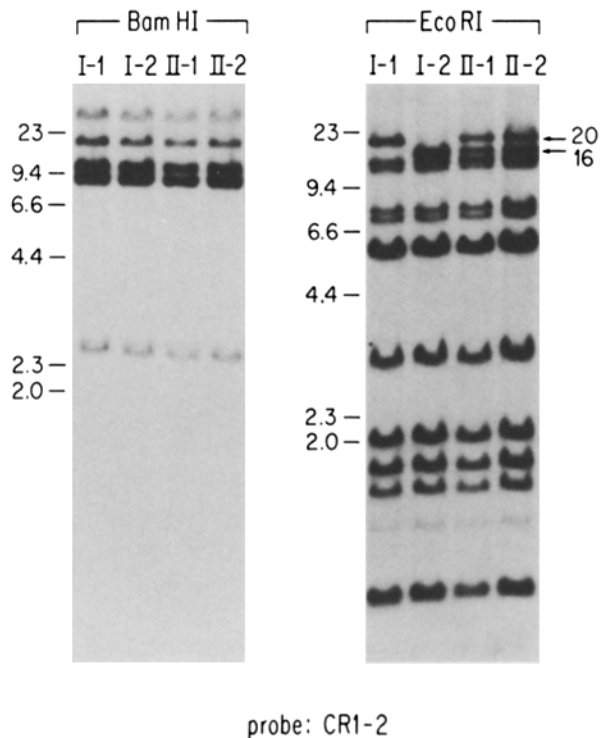


FIGURE 8. Autoradiographs of the Southern blots in which CR1-2 was hybridized to the Bam HI and Eco RI fragments of DNA from the members of a family having the FF phenotype. A Hind III digest of λ DNA was used as a marker and the positions of these fragments are designated on the left in kilobases. The positions of the 20- and 16-kb Eco RI restriction fragments are shown on the right.

TABLE IV
Lack of Correlation of the CR1 Protein Phenotypes with Eco RI RFLP

CR1 protein phenotype (n)	Eco RI RFLP		
	20 kb ⁺	20 kb ⁺ /16 kb ⁺	16 kb ⁺
FF (35)	10	20	5
FS (17)	5	7	5
FF' (1)	1	0	0
SS (4)	1	2	1

and several cDNA and genomic probes. The cDNA probes hybridized to highly homologous repetitive sequences, permitting the identification of multiple restriction fragments (17). The relationships between the DNA and protein polymorphisms were defined by pedigree analyses and by assessment of coincidence in the normal population. RFLPs were revealed by six enzymes and these RFLPs could be divided into three categories according to their relationships to the CR1 allotypes. The RFLPs generated by Bam HI, Sac I (Figs. 2 and 4; Table I), and Eco RV (Fig. 5) were found exclusively in association with a particular CR1 allotype and these may provide insight into the genetic mechanism for the

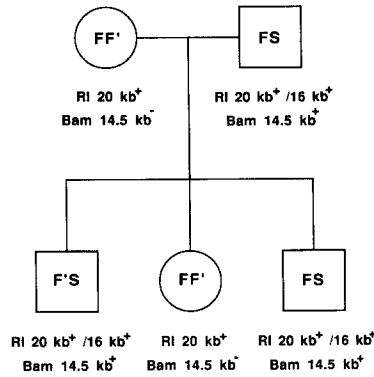


FIGURE 9. Cosegregation of the Eco RI polymorphism with the CR1-4⁺, 14.5-kb Bam HI fragment and the S allotype.

occurrence of the S, F, and F' forms of the receptor. The RFLPs generated by Hind III (Table II) and Pvu II (Fig. 7; Table III) were in linkage disequilibrium with, but not exclusively associated with, one or another of the CR1 allotypes and these may reveal evolutionary relationships between CR1 genomic, structural, and quantitative polymorphisms. The RFLP revealed by Eco RI cosegregated but was not in linkage disequilibrium with CR1 allotypes or the RFLPs generated by other restriction enzymes (Figs. 8 and 9; Table IV), perhaps reflecting a distal location within the large CR1 gene or the presence of a recombinational "hot spot" between the polymorphic Eco RI restriction site and the rest of the CR1 gene.

Among mechanisms that may account for the CR1 allotypes differing by size, three findings provided evidence for the insertion or deletion from the CR1 gene of segments containing introns and exons. First, the observation that two restriction enzymes, Bam HI and Sac I, generated new restriction fragments in 20 of 21 individuals expressing the larger S allotype and in none of 35 individuals having only the smaller F allotype indicated that the S allele differed from the F allele by more than a single base pair (Figs. 2 and 4; Table I). Two individuals, one of whom was Black (Fig. 2, donor 5), who expressed the S allotype but lacked both characteristic Bam HI and Sac I polymorphisms, were not examined further in this study, but this finding indicated that more than one mechanism could lead to the production of the higher molecular weight S form of CR1. Second, the S allele not only had new restriction fragments but also apparently contained all Bam HI and Sac I fragments that were present in the F allele (Figs. 2 and 4). This observation suggested that the new restriction sites responsible for the S-specific fragments were not located within a segment of the CR1 gene that was common to both the S and F alleles. Third, the GB2PE genomic probe derived from the S allotype-associated Bam HI fragment identified this 14.5-kb fragment as being unique to the S allele (Fig. 2B). The additional finding that this genomic probe and the GB1R1 genomic probe, both of which lacked human repeat sequences and repetitive CR1-coding sequences, hybridized to restriction fragments other than those from which the probes were derived was further evidence for the duplication of sequences within the CR1 gene (Figs. 2B and

3B). The presence of highly homologous repeat sequences in both coding and noncoding regions would suggest that homologous recombination with unequal crossovers leading to the insertion and deletion of genomic sequences may be the basis for the CR1 allotypes differing in the lengths of their polypeptide chains. If the S allele had been generated by unequal homologous recombination, then negative selection may account for the lack of occurrence at comparable frequencies of a corresponding shorter allele that constitutes the reciprocal product of such an event. Although the F' allele may be the result of a loss of genomic sequences by similar mechanisms, further analysis of this allele has been made difficult by our not having identified an F'F' homozygote among 358 persons.

Analysis of the CR1 transcripts associated with the three CR1 allotypes indicated that these differed in size, having the same rank order as the corresponding CR1 allotypes (Fig. 6). The F and S allotypes were each associated with two transcripts differing by 1.4 kb, presumably reflecting alternative polyadenylation sites. This circumstance precluded firm conclusions regarding the precise allelotypic differences in the size of the CR1 transcripts. However, the apparent minimal difference of 1.3–1.5 kb between the F', F, and S transcripts, respectively, is sufficient to encode the 40–50 kd difference in the CR1 allotypes. Also, the 1.4-kb allelic difference in transcript size is similar to the length of each of four to five highly homologous repeat sequences of CR1 cDNA that have recently been characterized (6), raising the possibility that the CR1 allotypes may differ from each other by their number of homologous repeats. Additional evidence for the possibility that repetitive sequences constitute the inserted or deleted CR1 peptides is the absence of tryptic peptides that are unique for the CR1 allotypes (22). Further analysis of the genomic sequences unique to the S allele will establish whether a genomic unit containing exons encoding a 50-kD peptide has this apparent duplicative capacity.

The number of common CR1 alleles that we observed in this study was expanded to 10 by use of the restriction enzymes, Hind III, Pvu II, and Eco RI. A 7.4-kb Hind III fragment occurred in association with either F or S, whereas the 6.9-kb Hind III fragment was associated only with F (Table II). The two types of F alleles defined by the Hind III polymorphism could be further subdivided by the presence or absence of a 5.8-kb Pvu II fragment, whereas the S allele was associated only with the presence of this 5.8-kb Pvu II fragment (Table III). Thus, we saw four different alleles associated with the F allotype whereas only a single S allele, in tight association with the 7.4-kb Hind III and the 5.8-kb Pvu II fragments, was shown (Tables II and III; Fig. 10). These findings might indicate that S evolved relatively recently or that negative selection has limited the occurrence of the other potential types of S alleles. The least complex scheme leading to the evolution of these different alleles and not involving selection for or against particular genotypes can be proposed. Independent mutations generated the 6.9-kb Hind III fragment and the 5.8-kb Pvu II fragment and the combination of these in a single allele was the result of a subsequent recombinational event (Fig. 10). The common polymorphism seen with Eco RI showed no linkage disequilibrium with any of the other RFLPs despite cosegregating with the Bam HI polymorphism, and locating the site of

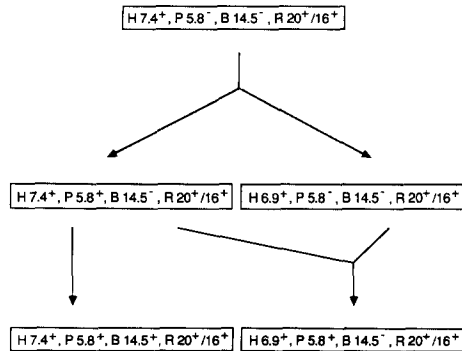


FIGURE 10. Proposed scheme of evolution of the 10 common alleles of CR1 identified by the combination of different RFLPs. B14.5⁺ and B14.5⁻ represent the RFLP defined by the presence or absence of the 14.5-kb Bam HI fragment hybridizing with CR1-4 and corresponding to the S and F allotypes, respectively. H7.4⁺ and H6.9⁺ represent the RFLP defined by the allelic 7.4- and 6.9-kb Hind III fragments which correlate with the high and low CR1 number on erythrocytes, respectively. P5.8⁺ and P5.8⁻ represent the RFLP defined by the presence and or absence of the 5.8-kb Pvu II fragment. R20⁺/16⁺ represents the RFLP defined by the allelic 20- and 16-kb Eco RI fragments and indicates that either of the fragments can be associated with each of the five alleles defined by the other restriction enzymes. The relative positions of the loci defined by the polymorphic fragments are not known. In this least complex scheme for the evolution of the CR1 alleles, independent mutations generated the alleles containing the 5.8-kb Pvu II fragment and the 6.9-kb Hind III fragment. From the allele containing the 5.8-kb Pvu II fragment, evolved the allele having the 14.5-kb Bam HI fragment associated with the S allotype. A recombinational event involving the alleles containing the 6.9-kb Hind III and the 5.8-kb Pvu II fragments, respectively, yielded the allele containing both fragments.

this polymorphism in the CR1 gene will determine whether distance or recombinational hot spots account for this finding. As has been shown (23–25) with other complement proteins, these RFLPs in CR1 may be associated with additional allotypes that can be visualized by the use of analytical procedures that separate proteins on the basis of charge rather than size. Furthermore, the relatively high number and frequency of CR1 alleles makes this locus a useful marker system (26) for the long arm of chromosome 1 to which the CR1 gene maps (27).

Summary

Human CR1 exhibits an unusual form of polymorphism in which allotypic variants differ in the molecular weight of their respective polypeptide chains. To address mechanisms involved in the generation of the CR1 allotypes, DNA from individuals having the F allotype (250,000 M_r), the S allotype (290,000 M_r), and the F' allotype (210,000 M_r) was digested by restriction enzymes, and Southern blots were hybridized with CR1 cDNA and genomic probes. With the use of Bam HI and Sac I, an additional restriction fragment was observed in 20 of 21 individuals having the S allotype with no associated loss of other restriction fragments. Southern blot analysis with a noncoding genomic probe derived from the S allotype-specific Bam HI fragment showed hybridization to this fragment and to two other fragments that were also present in FF individuals. Thus, an intervening sequence may be repeated twice in the F allele and three times in

the S allele. A restriction fragment length polymorphism (RFLP) unique to two individuals expressing the F' allotype was seen with Eco RV, but the absence of persons homozygous for this rare allotype prevented further comparisons with the F and S allotypes. Analysis of the CR1 transcripts associated with the three CR1 allotypes indicated that these differed by 1.3–1.5 kb and had the same rank order as the corresponding allotypes. Taken together, these findings suggest that the S allele was generated from the F allele by the acquisition of additional sequences, the coding portion of which may correspond to a long homologous repeat of ~1.4 kb that has been identified in CR1 cDNA.

We saw two other RFLPs with Hind III and Pvu II that were in linkage disequilibrium with the Bam HI–Sac I RFLPs associated with the S allotype, and a third polymorphism was seen with Eco RI that was not in linkage disequilibrium with the other polymorphisms. Thus, 10 commonly occurring CR1 alleles can be defined, making this locus a useful marker for the long arm of chromosome 1 to which the CR1 gene maps.

Received for publication 7 July 1986.

References

1. Fearon, D. T., and W. W. Wong. 1983. Complement ligand-receptor interactions that mediate biological responses. *Annu. Rev. Immunol.* 1:243.
2. 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.
3. 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.
4. Iida, K. and V. Nussenzweig. 1981. Complement receptor is an inhibitor of the complement cascade. *J. Exp. Med.* 153:1138.
5. 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.
6. Klickstein, L. B., W. W. Wong, J. A. Smith, C. Morton, D. T. Fearon and J. H. Weis. 1985. Identification of long homologous repeats in human CR1. *Complement.* 2:44. (Abstr.)
7. 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.
8. Kristensen, T., R. A. Wetzel 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.
9. 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.
10. 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.
11. Dykman, T. R., J. L. Cole, K. Iida and J. P. Atkinson. 1983. Polymorphism of human erythrocyte C3b/C4b receptor. *Proc. Natl. Acad. Sci. USA.* 80:1698.
12. 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 erythematosus. *J. Exp. Med.* 159:691.

13. Dykman, T. R., J. A. Hatch, M. S. Aqua and J. P. Atkinson. 1985. Polymorphism of the C3b/C4b (CR1) receptor: characterization of a fourth allele. *J. Immunol.* 134:1787.
14. Seya, T., V. M. Holers and J. P. Atkinson. 1985. Purification and functional analysis of the polymorphic variants of the C3b/C4b receptor (CR1) and comparison with H, C4-binding protein (C4 bp) and decay accelerating factor (DAF). *J. Immunol.* 135:2661.
15. 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.
16. Wilson, J. G., E. E. Murphy, W. W. Wong, L. B. Klickstein, J. H. Weis and D. T. Fearon. 1986. Identification of a restriction fragment length polymorphism by a CR1 cDNA that correlates with the number of CR1 on erythrocytes. *J. Exp. Med.* 164:50.
17. 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.
18. Maniatis, T., E. F. Fritsch, J. Sambrook. 1982. *Molecular Cloning: A Laboratory manual.* Cold Spring Harbor Laboratory, Cold Spring Harbor, NY.
19. Chirgwin, J. M., A. E. Przybyla, R. J. MacDonald and W. J. Rutter. 1979. Isolation of biologically active ribonucleic acid from sources enriched in ribonuclease. *Biochemistry.* 18:5294.
20. Thomas, P. S. 1980. Hybridization of denatured RNA and small DNA fragments transferred to nitrocellulose. *Proc. Natl. Acad. Sci. USA.* 77:5201.
21. Changelian, P., R. M. Jack, L. A. Collins and D. T. Fearon. 1985. PMA induces the ligand-independent internalization of CR1 on human neutrophils. *J. Immunol.* 134:1851.
22. 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.
23. Abbal, M., M. Thomsen, A. Cambon-Thomsen, J. Archambeau, M. Calot and D. Fathallah. 1985. Two subtypes of Bf F by isoelectric-focusing: differential linkage to other HLA markers. *Hum. Genet.* 69:181.
24. Campbell R. D. and D. R. Bentley. 1985. The structure and genetics of the C2 and factor B genes. *Immunol. Rev.* 87:19.
25. Whitehead, A. S., D. E. Woods, E. Fleischnick, J. E. Chin, E. J. Yunis, A. J. Katz, P. S. Gerald, C. A. Alper and H. R. Colten. 1984. DNA polymorphism of the C4 genes: a new marker for analysis of the major histocompatibility complex. *New Engl. J. Med.* 310:88.
26. Botstein, D., R. L. White, M. Skolnick and R. W. Davis. 1980. Construction of a genetic linkage map in man using restriction fragment length polymorphisms. *Am. J. Hum. Genet.* 32:314.
27. Weis, J. H., C. C. Morton, G. A. P. Bruns, J. J. Weis, L. B. Klickstein, W. W. Wong, and D. T. Fearon. 1986. 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.* In press.