Proc. Natl. Acad. Sci. USA
Vol. 84, pp. 2459–2463, April 1987
Immunology

Human complement C3b/C4b receptor (CR1) mRNA polymorphism that correlates with the CR1 allelic molecular weight polymorphism

(membrane protein/unequal crossing-over)

V. MICHAEL HOLERS, DAVID D. CHAPLIN, JOSEPH F. LEYKAM, BARBARA A. GRUNER, VIJAYA KUMAR, AND JOHN P. ATKINSON

Howard Hughes Medical Institute Laboratories and Department of Medicine, Divisions of Rheumatology and Allergy and Clinical Immunology, Washington University School of Medicine, St. Louis, MO 63110

Communicated by William H. Daughaday, December 24, 1986 (received for review October 23, 1986)

ABSTRACT **The human C3b/C4b receptor (CR1) is a M_r ≈200,000 single-chain integral membrane glycoprotein of human erythrocytes and leukocytes. It functions both as a receptor for C3b- and C4b-coated ligands and as a regulator of complement activation. Prior structural studies have defined an unusual molecular weight allelic polymorphism in which the allelic products differ in molecular weight by as much as 90,000. On peripheral blood cells there is codominant expression of CR1 gene products of M_r 190,000 (A), 220,000 (B), 160,000 (C), and 250,000 (D). Results of prior biosynthetic and tryptic peptide mapping experiments have suggested that the most likely basis for the allelic molecular weight differences is at the polypeptide level. In order to define further the molecular basis for these molecular weight differences, human CR1 was purified to homogeneity, tryptic peptide fragments were isolated by HPLC and sequenced, oligonucleotide probes were prepared, and a CR1 cDNA was identified. A subclone of this CR1 cDNA was used as a probe of RNA blots of Epstein–Barr virus-transformed cell lines expressing the allelic variants. Each allelic variant encodes two distinct transcripts. A mRNA size polymorphism was identified that correlated with the gene product molecular weight polymorphism. This finding, in addition to a prior report of several homologous repeats in CR1, is consistent with the hypothesis that the molecular weight polymorphism is determined at the genomic level and may have been generated by unequal crossing-over.**

The human C3b/C4b receptor (CR1) (reviewed in refs. 1 and 2) is a single-chain integral membrane glycoprotein of M_r ≈200,000 that is expressed on cells of immunological (T and B lymphocytes, monocytes, mast cells, polymorphonuclear cells) and nonimmunological (primate erythrocytes, glomerular podocytes) lineages. Functions mediated by CR1 include phagocytosis of erythrocytes and bacteria, endocytosis of small immune complexes, and generation of second signals. Erythrocyte CR1 binds complement component C3b- or C4b-bearing immune complexes and may transport them to the mononuclear phagocytic system in the liver and spleen. CR1 is also a cofactor for factor I-mediated cleavage of C3b and C4b and accelerates the decay of the classical and alternative-pathway C3 convertases.

Family studies have determined that CR1 is genetically linked in humans to two serum proteins, C4-binding protein and H, that have similar but not identical cofactor and decay-accelerating activities (3, 4). These three proteins have similar amino acid compositions and at least part of the structure of each of the three proteins consists of consensus repeats of ≈60 amino acids (5–8). Initial identification of a partial CR1 cDNA has allowed localization of this complement regulatory gene complex to the q32 region of chromosome 1 (9). More recently, the complement receptor for C3d (CR2) was shown to be structurally related to CR1 (10).

An interesting and unusual allelic polymorphism of CR1 based on molecular weight differences has been described (11–14). The four allelic variants have molecular weights of 190,000 (CR1-A), 220,000 (CR1-B), 160,000 (CR1-C), and 250,000 (CR1-D) and are found with a gene frequency of 0.83, 0.16, 0.01, and 0.002, respectively. The structural basis for this polymorphism is incompletely understood but appears to be due to peptide differences rather than posttranslational modifications (15). This conclusion is based on results (15) that demonstrated the following: (*i*) deglycosylation with endoglycosidase F decreases the molecular weight by ≈20,000 for all four variants, (*ii*) biosynthetic labeling of all four allotypic CR1 variants in Epstein–Barr virus (EBV)-transformed cell lines identifies pro-CR1 forms that retain the molecular weight differences of the mature CR1 molecules, and (*iii*) CR1 contains only N-linked carbohydrates and no sulfate, phosphate, or lipids.

Methods to purify CR1 to homogeneity at the 1- to 10-nM level have been described by ourselves (16) and others (17, 18). The NH_2 terminus of CR1 is blocked (16, 17). In order to obtain sequence of internal tryptic peptides, a method was developed to rapidly isolate and sequence individual peptides. A 2.5-kilobase (kb) cDNA clone for CR1 was identified and characterized. The molecular nature of the CR1 allelic molecular weight polymorphism was assessed by an analysis of mRNA of cells expressing the CR1 phenotypes.

MATERIALS AND METHODS

Purification of CR1 and Tryptic Peptide Fragments. Four nanomoles of erythrocyte CR1 (16) was reduced in 0.1 M NH_4HCO_3/1% $NaDodSO_4$/20 mM dithiothreitol at 37°C for 90 min and then alkylated in the dark by the addition of 50 mM iodoacetic acid at 37°C for 60 min. After precipitation overnight at −20°C with a 10-fold excess (vol/vol) of methanol and a 2% (wt/wt) amount of L-1-tosylamido-2-phenylethyl chloromethyl ketone-treated trypsin (TPCK-trypsin), the preparation was centrifuged at 25,000 × *g* at 4°C for 30 min. The supernatant was removed, the tube was dried under N_2, and CR1 was resolubilized in 0.5 ml of 0.1 M NH_4HCO_3/0.001% $NaDodSO_4$/2 mM $CaCl_2$. TPCK-trypsin (2%, wt/wt) was added and then again at 4 hr. After incubation at 37°C for a total of 16 hr, the sample was lyophilized.

A "two-dimensional" HPLC separation of tryptic fragments was performed, using a Waters liquid chromatograph.

The publication costs of this article were defrayed in part by page charge payment. This article must therefore be hereby marked "*advertisement*" in accordance with 18 U.S.C. §1734 solely to indicate this fact.

Abbreviation: EBV, Epstein–Barr virus.

Table 1. Oligonucleotide probe and pUCR1-4 nucleotide sequences

Oligonucleotide	Sequence
7P12	Tryptic peptide: ValPheGluLeuValGlyGluPro
	Oligonucleotide: 5′ GTITTYGARYTIGTIGGIGRCC 3′
	Nucleotide: 5′ GTGTTTGAGCTTGTGGGTGAGCC 3′
	Translated: ValPheGluLeuValGlyGluPro
WO-1	Oligonucleotide: 5′ TTTTGTTAGGTATAATGCACTGAGGGGCCGGGCCGC 3′
	Nucleotide: 3′ AAACAATCCATATTACGTGACTCCCCGCCCCGGCG 5′
9P1	Tryptic peptide: ValHisGlnProProProAspValLeuHisAlaGluArg
	Oligonucleotide: 5′ GTGCATCAGCCCCCCCCTGATGTGCTGCATGCTGAG 3′
WO-2	Oligonucleotide: 5′ TAATTGGAATCCTTCATCACAAACAAAATCCACTTT 3′

Included are oligonucleotide sequences mapped to positions on pUCR1-4 (Fig. 2). For 7P12, the nucleotide sequence of pUCR1-4 (Fig. 2). For 7P12, the nucleotide sequence of pUCR1-4 and translated protein sequence are compared to the CR1-7P12 tryptic peptide sequence and oligonucleotide probe sequence. Correlation of the WO-1 oligonucleotide probe sequence to the nucleotide sequence of pUCR1-4 is shown as well as the 9P1 and WO-2 oligonucleotide probe sequences. I, deoxyinosine; R, A or G; Y, C or T.

The initial separation was on a TSK SP-5PW sulfopropyl column (7.5 × 75 mm; Toya Soda, Tokyo), using a gradient of 20 mM NH_4OAc (pH 3.0) to 500 mM NH_4OAc (pH 9.0). Eluate absorption was monitored at 280 nm. Eighteen pools of fractions from this column eluate were then lyophilized and redissolved in 0.1% trifluoroacetic acid. The separation in the second dimension was on a Vydac C_{18} reversed-phase HPLC column (4.6 × 250 mm; Separations Group, Hesperia, CA), using a 0–60% acetonitrile gradient in 0.1% trifluoroacetic acid over 3 hr. Eluate absorption was monitored at 214 nm. Fractions containing apparent tryptic fragment peaks were lyophilized and stored. Tryptic peptides were sequenced on an Applied Biosystems (Foster City, CA) 470A sequencer.

Oligonucleotides and cDNA Construction. Oligonucleotides corresponding to the sequences of tryptic fragments CR1-7P12 and CR1-9P1 (see Fig. 1, and Table 1) were constructed by the phosphoramidite method (19) on an Applied Biosystems oligonucleotide synthesizer. Oligonucleotides WO-1 and WO-2 (Table 1), derived from sequences presented by Klickstein and coworkers (8, 9), were synthesized in an antisense direction. The sequence for oligonucleotide 9P1 was derived using codon preference (20). Oligonucleotide 7P12 included a mixture of all possible codons, except for deoxyinosines at four positions. Oligonucleotides were end-labeled using [γ-^{32}P]ATP (Amersham) and polynucleotide kinase (New England Biolabs).

Total RNA was obtained according to Chirgwin *et al.* (21) from HL-60 promyelocytic cells after induction with dimethyl sulfoxide for 72 hr. Poly(A)$^+$ mRNA was isolated by passage over an oligo(dT) column (Pharmacia P-L Biochemicals). cDNA was prepared using the RNase H/DNA polymerase I method (22). cDNAs size-selected by agarose gel electrophoresis to be >1.5 kb were cloned in λgt11 (Stratagene Systems, San Diego, CA). Approximately 10^5 unamplified recombinant phage were screened. Hybridizations were performed in 6× SSC (1× = 0.15 M NaCl/0.015 M sodium citrate)/0.05 M sodium phosphate, pH 6.5/5× Denhardt's solution (1× = 0.02% Ficoll 400/0.02% polyvinylpyrrolidone/0.02% bovine serum albumin)/20% (vol/vol) formamide/5% dextran sulfate/denatured herring sperm DNA (0.1 mg/ml) at 42°C for oligonucleotide probes 9P1, WO-1, and WO-2 and in 6× SSC/5× Denhardt's solution/denatured herring sperm DNA (0.1 mg/ml) for probe 7P12. Blots were washed in 2× SSC/0.1% $NaDodSO_4$ at 30°C, and autoradiographs were developed using Kodak XAR-5 film.

Nucleotide Sequence and RNA and DNA Blot Analysis of Cell Lines. The CR1 cDNA was subcloned into pUC-19 and M13mp19 according to standard methods (23). Nucleotide sequencing was performed by the dideoxy method (24).

RNA blot analysis was performed following electrophoresis in either 0.9% agarose/0.6% formaldehyde (Fig. 3) or 0.7% agarose/0.6% formaldehyde (Fig. 4) in a buffer containing 0.02 M 3-(*N*-morpholino)propanesulfonic acid, 5 mM sodium acetate, and 1 mM EDTA (disodium salt). Transfer to nylon membrane (Micron Separations, Honeyoye Falls, NY) was done at 4°C, using a Bio-Rad electrotransfer apparatus. Genomic Southern blot analysis was performed by standard methods (25). cDNA probes were labeled using either a nick-translation kit (Bethesda Research Laboratories) or the hexanucleotide-priming method (26). Hybridization was in 6× SSC/50% formamide, 0.1% $NaDodSO_4$/1× Denhardt's solution/12.5% dextran sulfate/10 mM Tris·HCl, pH 7.6, at 42°C. Blots were washed in 0.2× SSC/0.1% $NaDodSO_4$ at 56°C before autoradiography.

HL-60 and EBV-transformed cell lines were maintained as reported (15). Identification of the surface CR1 phenotype of EBV-transformed lines was verified by surface iodination and immunoprecipitation (15).

RESULTS

Purification and Amino Acid Sequence of CR1 Tryptic Fragments. CR1 tryptic peptides were separated by ion-exchange HPLC on a sulfopropyl column followed by reversed-phase HPLC. This allowed the separation of tryptic

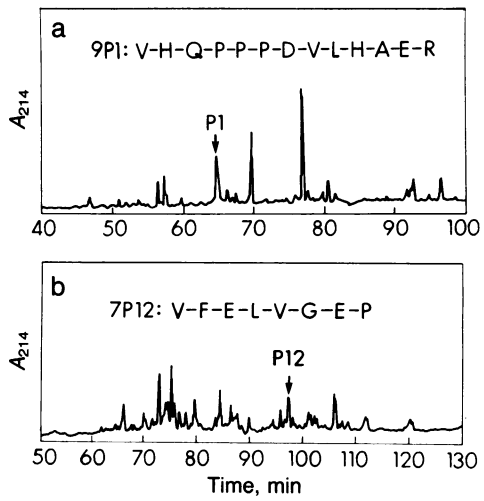

FIG. 1. Two chromatograms of the second-dimension reversed-phase HPLC separation of CR1 tryptic fragment. Arrows indicate peaks from which the tryptic peptide sequences (standard one-letter amino acid symbols) in each chromatogram were obtained. Only those areas of the chromatogram which contain peaks are included.

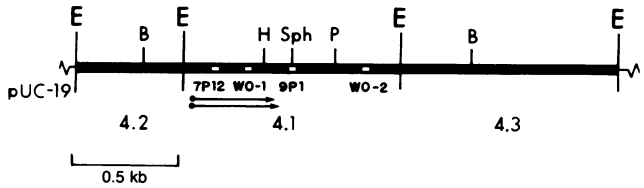

FIG. 2. Restriction map of pUCR1-4 insert. Correlation with the sequences 7P12 and WO-1 is in Table 1. Location and extent of *Eco*RI subclones pUCR1-4.1, pUCR1-4.2, and pUCR1-4.3 are indicated. Arrows denote the regions of the initial nucleotide sequence analysis that were used to identify sites of apparent hybridization of probes 7P12 and WO-1. Open blocks represent sites of hybridization of the four oligonucleotides. Restriction enzyme cleavage sites: E, *Eco*RI; B, *Bam*HI; H, *Hin*dIII; P, *Pst* I; Sph, *Sph* I.

peaks with single NH_2-terminal sequences in 10 of 11 cases, at levels of 5–500 pmol. Fig. 1 shows the reversed-phase HPLC chromatograms from 2 of the 18 second-dimension separations.

Identification of CR1 cDNAs. By use of oligonucleotide probes as described in *Materials and Methods*, four cDNAs were identified and then purified. One of these, λH6CR1-4, was found to have an insert of length 2.5 kb and was chosen for detailed characterization. The cDNA insert of λH6CR1-4 contained two internal *Eco*RI sites and was subcloned intact into pUC19 after partial *Eco*RI digestion of the phage DNA. The resulting plasmid was designated pUCR1-4. The three individual *Eco*RI cDNA fragments of λH6CR1-4 were also cloned into pUC19 and were designated pUCR1-4.1, pUCR1-4.2, and pUCR1-4.3 (Fig. 2). Inserts of these three *Eco*RI subclones each hybridized to specific mRNA species from dimethyl sulfoxide-induced HL-60 cells (positive for surface expression of CR1) but not from HeLa and HSB-2 (both negative for CR1 surface expression) (Fig. 3*a*). Nucleotide sequencing identified in subclone 4.1 the protein-coding sequence of tryptic peptide CR1-7P12 and the apparent hybridizing sequence to oligonucleotide WO-1 (Table 1). To further confirm the identity of pUCR1-4, genomic Southern blot analysis was performed using the insert of pUCR1-4 as a probe (data not shown). This revealed a complex pattern similar if not identical to those in prior reports (9, 27).

RNA Blot Analysis of HL-60 Cells and CR1 Polymorphic Variants. Upon induction with dimethyl sulfoxide or vitamin D, the HL-60 cell line increases its surface expression of CR1 by ≈200-fold (28). This increase correlates with a coordinate increase of biosynthetically labeled pro-CR1 (28). Analysis of mRNA from uninduced and dimethyl sulfoxide-induced HL-60 cells (Fig. 3*b*) showed that this increase of CR1 protein synthesis also correlates with a concomitant increase in the steady-state level of CR1 mRNA as detected with this cDNA probe.

In order to better define the mechanism of the generation of the CR1 polymorphic variants, mRNA was isolated from EBV-transformed cell lines expressing the AA (EB19), BB (EB22), AB (EB1922), AD (EB1925), and AC (EB1916) gene products. Blot analysis (Fig. 4) indicated that specific bands are detected for each allelic variant. For the A (M_r 190,000) variant, the specific band is at 8.6 kb; for the B (M_r 220,000) variant, at 11.6 kb; and for the C (M_r 160,000) variant, at 7.3 kb. For the D variant, a faint band at 12.8 kb is visualized. For each allelic variant there is a correlation between the relative size of the surface CR1 gene product and the relative size of the specific mRNA.

Analysis of Fig. 4 also indicates that two distinct bands are clearly visualized for the AA (HL-60 and EB19, lanes 1 and 2) and BB (EB22, lane 3) homozygous phenotypes. Two bands are also present in the upper region of the blot for the AD phenotype (EB1925, lane 5), and these probably represent the two transcripts for the D allele. Though a single new band is visualized in the AC (EB1916, lane 6) phenotype, the presumed upper band of two transcripts for the C allele would comigrate with the lower A transcript, as is apparent in the AB (EB1922, lane 4) phenotype, where again only three bands are visualized. These data are consistent with the presence of two transcripts for each allele, as is directly shown in the case of the A, B, and D alleles.

The experiment shown in Fig. 4 was repeated three times, and each time the same pattern was visualized. Table 2 summarizes data from these experiments. Part 1 presents size differences between the two transcripts for each of the four individual alleles and assumes that the upper band of the C allele overlies the lower band of the A allele. In part 2, the size of the lower band of one allele is compared to the lower band

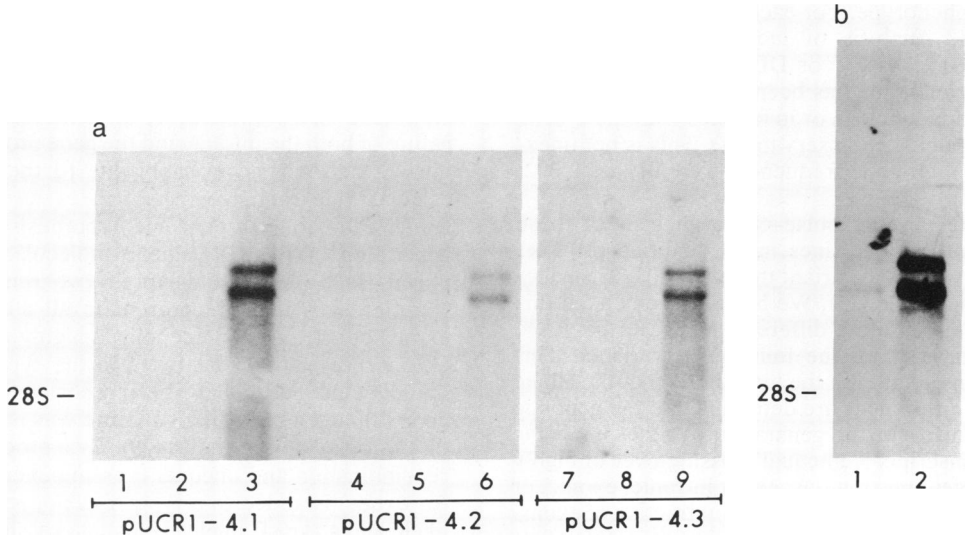

FIG. 3. (*a*) RNA blot analysis of *Eco*RI subclones of λH6CR1-4. Inserts of subclones pUCR1-4.1, -4.2, and -4.3 were hybridized with 5 μg of poly(A^+) RNA from HeLa (lanes 1, 4, and 7), HSB-2 (lanes 2, 5, and 8), or HL-60 (lanes 3, 6, and 9) cell lines. All lanes were from a single blot. (*b*) RNA blot analysis of 5 μg of poly(A^+) RNA from HL-60 cells that had been maintained under standard culture conditions (lane 1) or in the presence of 1% (vol/vol) dimethyl sulfoxide (lane 2) for 72 hr; the insert of pUCR1-4 was labeled and used as probe. RNA samples were electrophoresed in 0.9% agarose/0.6% formaldehyde gel and then transferred to nylon. Autoradiographic exposure time was 16 hr. Position of 28S rRNA is indicated.

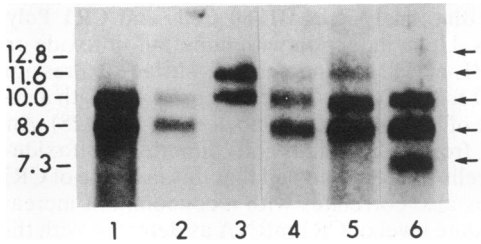

FIG. 4. RNA blot analysis of human cell lines that express the different polymorphic forms of CR1. The *Eco*RI insert of pUCR1-4.1 was labeled and used as probe of poly(A)$^+$ RNA obtained from the following cell lines: HL-60, 1 μg of RNA (lane 1); EB19, 10 μg (lane 2); EB22, 5 μg (lane 3); EB1922, 15 μg (lane 4); EB1925, 15 μg (lane 5); EB1916, 10 μg (lane 6). All lanes are from a single 0.7% agarose/0.6% formaldehyde gel. Autoradiogram was exposed for 24 hr for lanes 1, 2, 3, and 6 and 72 hr for lanes 4 and 5. Sizes of the hybridizing mRNAs (in kb) are at left, and arrows at right identify the five bands that were resolved.

of the next largest allele. These data point out the similarity of the size differences between mRNA of each allele.

DISCUSSION

Human CR1 is a multifunctional surface glycoprotein of M_r ≈200,000 with both receptor and complement regulatory functions. Because the NH_2 terminus of CR1 is blocked (16, 17), a method was developed to obtain internal protein sequence. Tryptic digestion of 4 nmol of CR1 gave more than 100 peptides, and sequential use of a sulfopropyl column and reversed-phase HPLC resulted in excellent separation and adequate recoveries for protein sequence analysis.

By use of oligonucleotide probes, four cDNAs were identified. One, of length 2.5 kb and designated λH6CR1-4, was further characterized and verified as specific for CR1 by means of DNA and RNA blot hybridization and nucleotide sequence analysis. Using the pUCR1-4.1 *Eco*RI insert as a probe, further information regarding the nature of the CR1 allelic molecular weight polymorphism was sought by blot analysis of poly(A)$^+$ RNA from cell lines expressing the polymorphic variants. As shown in Fig. 4, the size of CR1 transcripts bears a distinct relationship to the surface CR1 molecular weight phenotype. For each allele a specific band can be visualized. Analysis of individuals or cell lines expressing a homozygous CC or DD phenotype would be informative; however, none has been identified in over 300 randomly screened individuals or in families with the C or D alleles (J.P.A., unpublished observations). This is not unexpected in view of their gene frequencies of 0.01 and 0.002, respectively.

Biosynthetic labeling and pulse–chase analysis of these same EBV-transformed cell lines have identified pro-CR1 forms for each polymorphic variant that are ≈20 kDa smaller than the surface molecule (15). We have now extended this observation to demonstrate a different-sized CR1 poly(A)$^+$ mRNA species for each of the four allelic variants. This finding is most consistent with the hypothesis that the allelic molecular weight differences are determined at the genomic level. One mechanism for the generation of genes encoding different-sized transcripts is unequal crossing-over of highly homologous CR1 genomic regions, either intronic, exonic, or both, resulting in the overall addition or subtraction of exons. The presence of two apparent transcripts for each allele complicates this analysis. However, the data presented in Fig. 4 and summarized in Table 2 are quite consistent with the hypothesis of unequal crossing-over as a mechanism to account for the four CR1 alleles and, in fact, lend potential insight into the genomic organization of the allelic variants. This suggestion has previously been hypothesized based on both protein (29) and nucleotide (8) analyses and will be further addressed below.

Table 2. Summary of CR1 allelic mRNA size differences

Allele(s)	Size difference, kb (mean ± SD)
Part 1. Allelic transcripts	
A	1.65 ± 0.17
B	1.85 ± 0.24
C	1.57 ± 0.43
D	1.40 ± 0.17
Part 2. Smaller transcripts of each allele	
C vs. A	1.57 ± 0.43
A vs. B	1.65 ± 0.17
B vs. D	1.85 ± 0.24

Data were derived from four experiments done as for Fig. 4.

The first question concerns the relationship between the two transcripts for the A, B, and D alleles (probably present for the C allele as well). While minor CR1 bands are present on human erythrocytes (11), on HL-60 cells and B cells there is no evidence for another form of CR1 other than the CR1 surface product for each allele. HL-60 cells do not secrete CR1 (J.P.A., unpublished data), and there is no evidence that B cells make a secreted or otherwise altered form of CR1. The simplest explanation for the presence of two transcripts and a single apparent gene product is that a difference resides in the untranslated 5′ or 3′ areas of the mRNA. The similarity in the calculated size difference between the two transcripts for each of the four alleles (Table 2, part 1) also suggests that the basis for the size difference is in a region of the gene that is not critical to the size of the translated gene product and which is not significantly altered in the generation of the allelic variants.

If one assumes, then, that each allele has two transcripts that vary in their untranslated regions, the mRNA differences that might account for the gene product differences become much easier to analyze. Table 2, part 2, is a summary of these size differences. Each allelic variant is ≈1.6 kb different in size from the next. The presence of an ≈1.6-kb difference between each allele strongly suggests that the addition or subtraction of this fragment is the basis for the molecular weight difference between the allelic gene products, which differ in molecular weight by increments of 30,000 [M_r 160,000 (C), 190,000 (A), 220,000 (B), and 250,000 (D)]. Sixteen hundred bases of mRNA codes for a polypeptide of M_r 55,000–60,000; however, the similarity of the repeating nature of both the mRNA and the gene product indicates that this discrepancy is probably due to the methods used to calculate apparent sizes.

Fig. 5 illustrates a plausible hypothesis to account for the generation of the CR1 allelic molecular weight polymorphism. Each allele would express two transcripts due to 5′ or 3′ differences that are present in all alleles. An exon (or exons) of ≈1.6 kb and encoding a highly repetitive region of CR1 is present at least twice in the type A allele. Highly homologous repeats have been reported to be present in this gene (8), and a repeating structure was also predicted based on tryptic mapping studies (29). By a process of homologous recombination and unequal crossing-over, an ≈1.6-kb exon (or exons) is added to a type A allele, resulting in a type B allele and a type C allele. Unequal crossing-over in a type B allele would result in a type D allele and another type A allele. The presence of a higher number of B alleles than C alleles in the population suggests there may be a selective disadvantage to the C allelic product. Studies of the C allele in a population of patients with systemic lupus erythematosus (SLE) suggest that a high expression of the C allele is

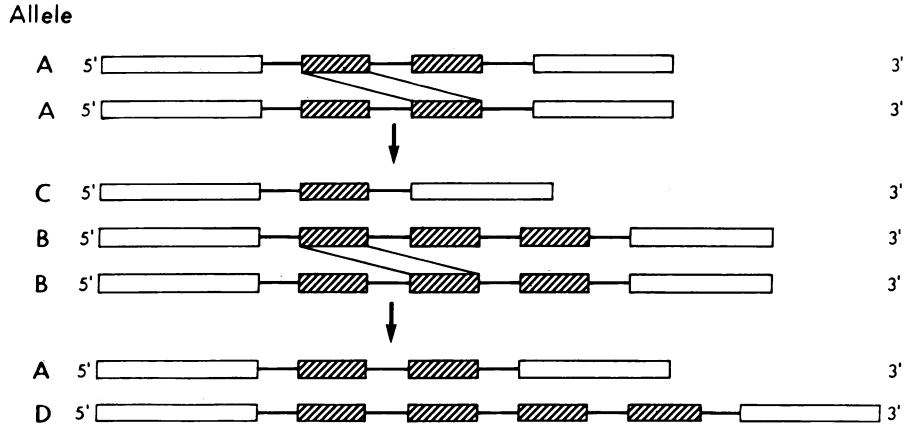

FIG. 5. Diagram of a hypothesis to account for the generation of the CR1 polymorphism. Open blocks represent genomic regions 5′ and 3′ of a postulated highly repetitive region containing ≈1.6 kb of an exon or exons (hatched blocks). Unequal crossing-over, shown in step 1 between two highly repetitive genomic regions in an A allele, would result in B and C alleles (step 2). Likewise, further unequal crossing-over of the B allele would generate the D allele (step 3).

associated with SLE and may represent one type of a selective disadvantage of this allele (13).

Proof of this hypothesis for the generation of the CR1 allelic variants will require more detailed sequencing studies and mapping of the variants. Other mechanisms, such as alternative RNA splicing of a single heterogeneous nuclear CR1 transcript or generation of tandem repeats, are not directly excluded by the data. However, initial genomic mapping studies have identified restriction fragment length polymorphisms related to the AA and BB phenotypes (9, 29).

Other aspects of the structure of CR1 deserve further consideration. The presence of numerous ≈60 amino acid repeats in CR1 (8) as well as in other complement regulatory proteins and non-complement binding proteins (5–7) suggests that this unit has been duplicated and then dispersed during the evolution of these genes. Duplication and unequal crossing-over are important mechanisms in the evolution of many genes and gene families—for example, collagen (30), immunoglobulins (31, 32), hemoglobin (33), transferrin (34), fibrinogen (35), and haptoglobin (36)—and undoubtedly played an important role in the development of this complement regulatory family. That the ≈60 amino acid consensus repeats are usually only ≈30% homologous, both within a protein and when compared to another protein in this family (7), indicates that the genetic events leading to their development are relatively old. Although largely composed of this type of consensus repeat, CR1 has two additional unique structural features, the molecular weight polymorphism and the presence of highly (>90%) homologous long repeats (≈800 bp) encoding numerous ≈60 amino acid consensus repeats. It is likely that these two features are related, and the striking homology of these long repeats suggests that these developed secondary to relatively recent genetic events. The genomic organization of this protein and characterization of the structure of CR1 in other species should provide further insight into the evolution of this gene family.

We thank Jeffrey Milbrandt, Dennis Hourcade, and Douglas Lublin for their insightful comments and Lorraine Whiteley and Pat Stewart for excellent secretarial assistance. This work was supported by National Institutes of Health Grant 5R01-A119642, the Monsanto Corp., and the Lottie Caroline Hardy Trust Fund.

1. Fearon, D. T. (1983) *Springer Semin. Immunopathol.* **6**, 159–172.
2. Dykman, T. R., Holers, V. M. & Atkinson, J. P. (1985) in *Microbiology*, ed. Leive, L. (Am. Soc. Microbiol., Washington, DC), pp. 19–22.
3. Rodriguez de Cordoba, S., Lublin, D. M., Rubinstein, P. & Atkinson, J. P. (1985) *J. Exp. Med.* **161**, 1189–1195.
4. Holers, V. M., Cole, J. L., Lublin, D. M., Seya, T. & Atkinson, J. P. (1985) *Immunol. Today* **6**, 188–192.
5. Chung, L. P., Bentley, D. R. & Reid, K. B. M. (1985) *Biochem. J.* **230**, 133–141.
6. Kristensen, T., Wetsel, R. A. & Tack, B. F. (1986) *J. Immunol.* **136**, 3407–3411.
7. Reid, K. B. M., Bentley, D. R., Campbell, R. D., Chung, L. P., Sim, R. B., Kristensen, T. & Tack, B. F. (1986) *Immunol. Today* **7**, 230–234.
8. Klickstein, L. B., Wong, W. W., Smith, J. A., Weis, J. H. & Fearon, D. T. (1985) *Complement* **2**, 44 (abstr.).
9. Wong, W. W., Klickstein, L. B., Smith, J. A., Weis, J. H. & Fearon, D. T. (1985) *Proc. Natl. Acad. Sci. USA* **82**, 7711–7715.
10. Weiss, J. J., Fearon, D. T., Klickstein, L. B., Wong, W. W., Richards, S. A., de Bruyn Kops, A., Smith, J. A. & Weis, J. H. (1986) *Proc. Natl. Acad. Sci. USA* **83**, 5639–5643.
11. Dykman, T. R., Cole, J. L., Iida, K. & Atkinson, J. P. (1983) *Proc. Natl. Acad. Sci. USA* **80**, 1698–1702.
12. Wong, W. W., Wilson, J. C. & Fearon, D. T. (1983) *J. Clin. Invest.* **72**, 685–693.
13. Dykman, T. R., Hatch, J. A. & Atkinson, J. P. (1984) *J. Exp. Med.* **159**, 691–703.
14. Dykman, T. R., Hatch, J. A., Aqua, M. S. & Atkinson, J. P. (1985) *J. Immunol.* **134**, 1787–1789.
15. Lublin, D. M., Griffith, R. C. & Atkinson, J. P. (1986) *J. Biol. Chem.* **261**, 5736–5744.
16. Holers, V. M., Seya, T., Brown, E., O'Shea, J. J. & Atkinson, J. P. (1986) *Complement* **3**, 63–78.
17. Wong, W., Jack, R. M., Smith, J. A., Kennedy, C. A. & Fearon, D. T. (1985) *J. Immunol. Methods* **82**, 303–313.
18. Sim, R. B. (1985) *Biochem. J.* **232**, 883–889.
19. Matteucci, M. D. & Caruthers, M. H. (1981) *J. Am. Chem. Soc.* **103**, 3185–3191.
20. Lathe, R. (1985) *J. Mol. Biol.* **183**, 1–12.
21. Chirgwin, J. M., Przybyla, A. E., MacDonald, R. J. & Rutter, W. J. (1979) *Biochemistry* **18**, 5294–5299.
22. Gubler, U. & Hoffman, B. J. (1983) *Gene* **25**, 263–269.
23. Maniatis, T., Fritsch, E. F. & Sambrook, J. (1982) *Molecular Cloning: A Laboratory Manual* (Cold Spring Harbor Laboratory, Cold Spring Harbor, NY).
24. Sanger, F., Nicklen, S. & Coulson, A. R. (1977) *Proc. Natl. Acad. Sci. USA* **74**, 5463–5467.
25. Southern, E. M. (1975) *J. Mol. Biol.* **98**, 503–517.
26. Feinberg, A. P. & Vogelstein, B. (1983) *Anal. Biochem.* **132**, 6–13.
27. Wong, W. W., Kennedy, C. A., Klickstein, L. B., Bonaccio, E., Weis, J. H. & Fearon, D. T. (1986) *Fed. Proc. Fed. Am. Soc. Exp. Biol.* **45**, 865 (abstr.).
28. Atkinson, J. P. & Jones, E. A. (1984) *J. Clin. Invest.* **74**, 1649–1657.
29. Nickells, M., Seya, T., Holers, V. M. & Atkinson, J. P. (1986) *Mol. Immunol.* **23**, 661–668.
30. Wozney, J., Hanahan, D., Tate, V., Boedtker, H. & Doty, P. (1981) *Nature (London)* **294**, 129–135.
31. Blomberg, B., Traunecker, A., Eisen, H. & Tonegawa, S. (1981) *Proc. Natl. Acad. Sci. USA* **78**, 3765–3769.
32. Heiter, P. A., Hollis, G. F., Korsmeyer, S. J., Waldmann, T. A. & Leder, P. (1981) *Nature (London)* **294**, 536–540.
33. Shen, S. H., Slightom, J. L. & Smithies, O. (1981) *Cell* **26**, 191–203.
34. Park, J., Schaeffer, E., Sodoli, A., Baralle, F. E., Cohen, G. N. & Zakin, M. M. (1985) *Proc. Natl. Acad. Sci. USA* **82**, 3149–3153.
35. Kant, J. A., Fornace, A. J., Saxe, D., Simon, M. I., McBride, O. W. & Crabtree, G. R. (1985) *Proc. Natl. Acad. Sci. USA* **82**, 2344–2348.
36. Maeda, N., Yang, F., Barnett, D. R., Bowman, B. H. & Smithies, O. (1984) *Nature (London)* **309**, 131–135.