

IDENTIFICATION OF AN ALTERNATIVE POLYADENYLATION
SITE IN THE HUMAN C3b/C4b RECEPTOR (COMPLEMENT
RECEPTOR TYPE 1) TRANSCRIPTIONAL UNIT AND
PREDICTION OF A SECRETED FORM OF
COMPLEMENT RECEPTOR TYPE 1

BY DENNIS HOURCADE, DAWN R. MIESNER, JOHN P. ATKINSON,
AND V. MICHAEL HOLERS

*From the Howard Hughes Medical Institute Laboratories and Department of Medicine, Division of
Rheumatology, Washington University School of Medicine, St. Louis, Missouri 63110*

The human C3b/C4b receptor, or complement receptor type one (CR1)¹ is a single chain membrane glycoprotein found on a variety of cell types (reviewed in reference 1). It is one of a family of proteins that interacts with the two complement proteins, C3b and C4b, which bind to immune complexes and to foreign particles. Over 85% of the CR1 on circulating cells is located on the surface of erythrocytes, where it can mediate the binding, processing, and transport of C3b-coated immune complexes. In addition, CR1 on phagocytic cells promotes endocytosis of small complexes and phagocytosis of larger particles. CR1 also expresses regulatory activity in that it acts as a cofactor for the factor I-mediated proteolytic inactivation of C3b and C4b and accelerates the decay of the classical and alternative pathway C3 convertases.

The proteins that bind C3b/C4b have recently become the focus of intense interest, in part because many of them (CR1, H, C4bp, CR2, and DAF) are found at a single locus on human chromosome 1 (2-6). They also share a common structural motif of tandemly repeated domains of 60-70 amino acids (7). These domains, known as short consensus repeats (SCRs), contain a number of invariant and other frequently conserved residues at specific positions.

Four polymorphic variants of human CR1 have been identified at the protein level (8-11). Their reduced forms are of *M_r* 220,000 (CR1-A), 250,000 (CR1-B), 190,000 (CR1-C), and 280,000 (CR1-D). In addition to classical genetic studies (8-12), two lines of molecular evidence suggest that each allotypic variant is generated from a different CR1 allele: (a) CR1 messenger RNAs from the different variants demonstrate size increments of ~1,400 bp (13, 14); and (b) restriction fragment-length polymorphisms of CR1 genomic sequences indicate chromosomal differences among the variants in the CR1-encoding region (14).

Recently, Klickstein et al. (15) reported cDNA sequences that encode the COOH

This work was supported in part by National Institutes of Health grants 53104 and 53095.

¹ *Abbreviations used in this paper:* CR1, complement receptor type 1; LHR, long homologous repeat; SCR, short consensus repeat.

terminal 76% of the CR1-A polymorphic variant. They showed that nearly all of this portion of the molecule is composed of SCRs and that the COOH terminus contains a transmembrane segment and a cytoplasmic domain. In addition they found that 21 of 23 SCRs in this region are organized in three long homologous repeats (LHRs) of seven SCRs each.

We have isolated CR1 clones from cDNA libraries of DMSO-induced HL-60 cells (13). First, we report here the amino acid sequence of the NH₂ terminal 28% of the major polymorphic form of CR1 which we derived from one of these cDNAs. Together, with the similarly derived sequence of Klickstein et al. (15), this completes the primary structure of the mature protein. Second, although the NH₂ terminus of CR1 is composed of a fourth LHR, we find that the first two NH₂-terminal SCRs are markedly divergent when compared with the corresponding SCRs of the other LHRs. We hypothesize that this segment is instrumental in cofactor activity, decay accelerating activity, and/or C3b/C4b binding, and we propose a model for the organization of CR1 that is based on differences in the degree of homology within the repeated sequences. Third, we identify an alternative polyadenylation site in the CR1 transcriptional unit, and we predict the synthesis of a secreted product. Fourth, we report a CR1-like genomic sequence that is highly homologous to the alternative polyadenylation site and the surrounding exons of CR1.

Materials and Methods

The isolation and characterization of pUCR1-4 has been described previously (13). The CR1-4 insert carried by that plasmid was removed by partial digestion of pUCR1-4 with Eco RI, isolated by standard methods (16), and joined to the sequencing vector pBSKS (Stratagene Cloning Systems, La Jolla, CA). A series of derivatives was constructed and DNA sequencing was performed using the dideoxy method of Sanger et al. (17). Analysis of the sequence data was performed on a personal computer (AT; IBM Instruments, Inc., Danbury, CT) using the MicroGenie (Beckman Instruments, Inc., Palo Alto, CA) program.

Cosmid libraries were constructed as outlined in reference 16, using the human cell line EB-19 (18) as a DNA source, the pTCF vector (19), and an electrophoresis apparatus (Bullseye; Hoeffer Scientific Instruments, San Francisco, CA) to fractionate by size genomic DNA partially digested with enzyme Sau 3A. The λ genomic library was that of Maniatis and colleagues (20), constructed with the Charon 4A vector (21) and human fetal liver DNA partially digested with Hae III and Alu I. It was obtained through the American Type Culture Collection, Rockville, MD.

The growth of cell lines, RNA isolation, and Northern blot construction and hybridization have been described (13). The DNA probes were electroeluted from polyacrylamide gels and then labeled by the oligonucleotide-priming method (22). At least 0.5 μ g DNA (10^8 cpm) was used for each blot. Blots were reused after brief boiling and slow cooling in 2X SSC (16) solution and after checking for residual hybridization by autoradiography. Southern blots were prepared as described in reference 16.

Results and Discussion

CR1 Partial cDNA Sequence and Organization of CR1. We have previously reported the isolation and characterization of a partial CR1 cDNA (13). It encodes several known CR1 peptides and is homologous to CR1 messenger RNAs on Northern blots. We have now obtained the complete nucleotide sequence of that insert. It consists of an open reading frame that extends 1,679 bp followed by an apparently untranslated region of 697 bp (Fig. 1). Included in the open reading frame are the sequences of eight peptides of the CR1 protein (13, 15, 22a).

The first 16 amino acids of the open reading frame bear a strong resemblance to a typical eukaryotic signal sequence (23), including a possible NH₂/hydrophobic boundary (immediately following Ser), a hydrophobic region (Leu-Leu-Ala-Val-Val-Val-Leu-Leu-Ala-Leu), and a typical COOH region (Pro-Val-Ala-Trp-Gly-Gln). We anticipate the initial methionine codon occurs a short distance upstream of this coding region. According to the “(-3, -1)” rule (23), the putative signal peptide would be cleaved between Gly and Gln, resulting in a terminal Gln residue. Since the NH₂ terminus of the mature CR1 protein is resistant to standard biochemical sequencing procedures (24, 25), it is likely that the Gln is modified to pyroglutamic acid (26).

The remainder of the open reading frame consists of 8.5 SCRs, which are characteristic of the family of proteins that interact with C3b/C4b. These repeats (see Fig. 2) are 60–70 amino acids in length and include invariant residues (four cysteines, a glycine, and a tryptophan) as well as a number of frequently conserved ones (7, 15).

The open reading frame ends with a stop codon (TGA) at a possible 5' splice sequence (CAGG/GTGAGT) (27). This site is followed by an apparently untranslated region of 697 bp with no features of SCRs discernable in any reading frame. The sequence ends with an AATAAA polyadenylation signal (28), a sequence (CTTT-GACTGC) similar to the proposed polyadenylation consensus sequence (TTTT-CACTGC) of Benoist et al. (29), and a nine-base poly(A) tail.

Our sequence, comprising the NH₂ terminal 28% of CR1, includes at its 3' end a precise overlap of 248 base pairs with the 5' end of the sequence of Klickstein et al. (15) (Fig. 1). Merging the sequences of our two groups leads to a CR1 protein composed of 1,998 amino acids organized into 30 SCRs followed by a transmembrane segment and cytoplasmic region (Fig. 3 A).

While each SCR in the C3b/C4b-binding protein family exhibits, typically, 20–35% amino acid sequence homology with any other SCR in the family, the SCRs of CR1 exhibit an additional internal homology. As first described by Klickstein et al. (15), CR1 can be organized in tandem LHRs of seven SCRs each (Fig. 3 B). Each LHR is 65–90% homologous to any other LHR. Our NH₂-terminal sequence extends this model through LHR-A.

With completion of the primary structure, it is apparent that the SCRs of CR1 can also be organized in another form (Fig. 3 C). This model incorporates the homologous repetition featured in the LHR model (Fig. 3 B) while using different degrees of intramolecular homology as a physical basis for assigning separate regions. The most dramatic example is region II, composed of 16 SCRs in LHRs A, B, and C (SCR-3 through SCR-18). Within this region, the seven SCR repetition is duplicated nearly precisely; for example, SCR 3–9 is 99% homologous to SCR 10–16, and SCR 10–11 is 99% homologous to SCR 17–18 (Fig. 3 C). Adjacent to this lies region III, which extends 10 SCRs (SCR-19 through SCR-28). Within this region the seven SCR repetition is also seen in highly homologous degree; thus, SCR 19–21 is 91% homologous to SCR 26–28. Region II and region III have been divided between SCR-18 and SCR-19 because the degree of homology in the seven SCR repeat unit changes at this juncture. SCR 12–18 is only 67% homologous to SCR 19–25.

Region I, consisting of SCR-1 and SCR-2, which is the NH₂ terminus of the protein, is 61% homologous to SCR-8 and SCR-9 in region II, 61% homologous to SCR-15 and SCR-16 in region II, and 59% homologous to SCR-22 and SCR-23

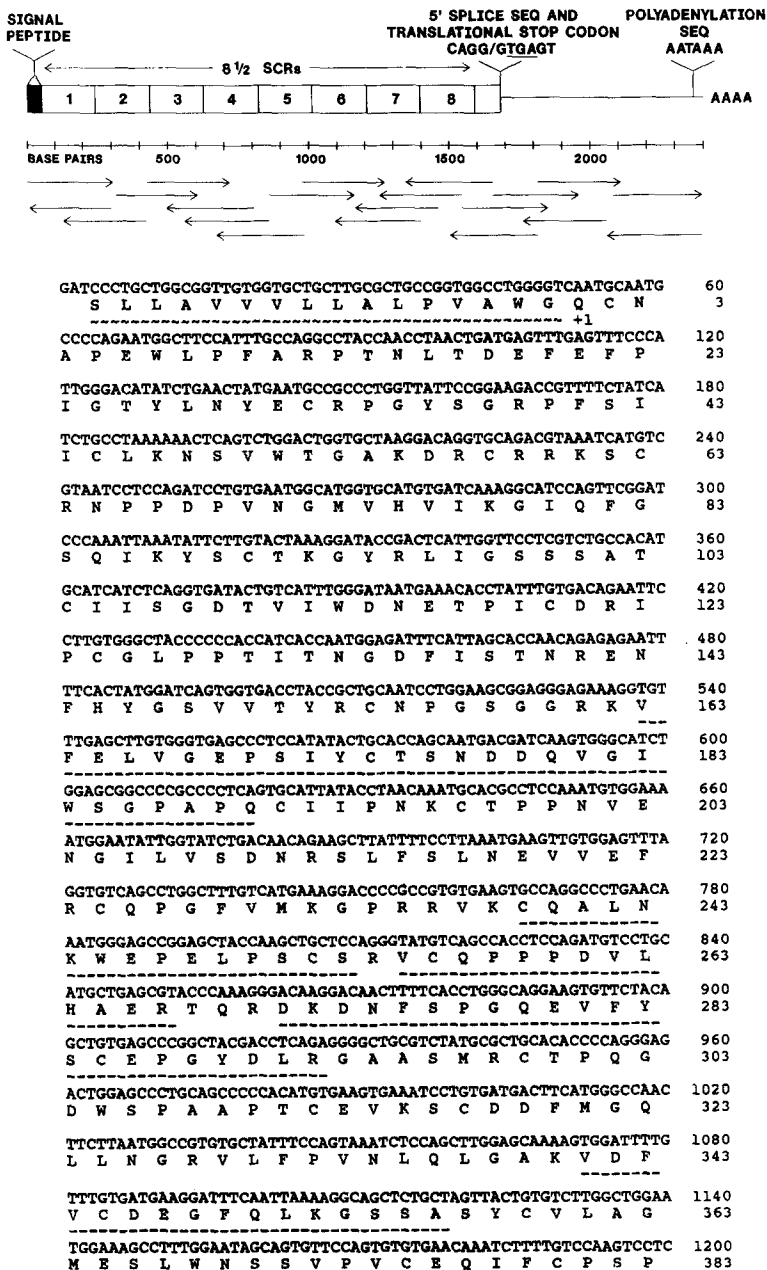


FIGURE 1. Figure continued on facing page.

of region III. Region IV, the last two SCRs in CR1 (29 and 30), is not a part of the seven SCR repetition. It is 26% homologous to SCR-22 and SCR-23, 23% homologous to SCR-15 and SCR-16, 23% homologous to SCR-8 and SCR-9, and 26% homologous to SCR-1 and SCR-2.

```

CAGTTATTCCCTAATGGGAGACACACAGGAAAACCTCTGGAAGTCTTCCCTTTGGAAAAG 1260
P V I P N G R H T G K P L E V F P F G K 403
-----
CAGTAAATTACACATGCGACCCCAACCCAGACAGAGGGACGAGCTTCGACCTCATTGGAG 1320
A V N Y T C D P H P D R G T S F D L I G 423
-----
AGAGCACCATCCGCTGCACAAGTGACCCCTCAAGGGAATGGGGTTTGGAGCAGCCCTGCC 1380
E S T I R C T S D P Q G N G V W S S P A 443
-----
CTGGCTGTGGAATTTCTGGGTCACTGTCAAGCCCAGATCAATTTCTGTTGCCAAGTTGA 1440
P R C G I L G H C Q A P D H F L F A K L 463
-----
AAACCCAAACCAATGCATCTGACTTCCCATTTGGGACATCTTTAAAGTACGAATGCCGTC 1500
K T Q T N A S D F P I G T S L K Y E C R 483
-----
CTGAGTACTACGGGAGGCCATTCTCTATCACATGTCTAGATAACCTGGTCTGGTCAAGTC 1560
P E Y Y G R P F S I T C L D N L V W S S 503
-----
CCAAAGATGCTCTGTAACGTAATCATGTAATACTCCCTCCAGATCCAGTGAATGGCATGG 1620
P K D V C K R K S C K T P P D P V N G M 523
-----
TGCATGTGATCACAGACATCCAGTTGGATCCAGAATCAACTATTCTTGTACTACAGGGT 1680
V H V I T D I Q V G S R I N Y S C T T G 543
-----
GAGTTGGCAGCAACATCTCTTGGTTTAAAGAGTCCAGCACAGCGATAGTACTTTCTAGCC 1740
*
ACATCTCAGCAAGGAAACTAGGCTATTGCCACCTGCTCTTAAAGAGGCTTGAACACAGGTG 1800
-----
TTAACTCCTGATTGAAATGAACAAAGATAGGAGAAGATTAGGGGGAAAATCTGTATCCTT 1860
-----
GCTGGAACCAGGGCAGTGCACATATAAGAGTATGCTGTTCACTGGATGGGAAAGAAAA 1920
-----
AACTTAGAAGTGTAGTAGTCAAAGCACACAAAACCCCTAACCCAGAGTAGACATTGCTG 1980
-----
GAAGAAAGGGAAGACCATGTAGCAGCTGTGTGAGAGAATGAATCTTAATGATAACAGCAT 2040
-----
GATCCCTTGCTAGGGCTGCCATCAAAAAGTACAGGCCCTTCCCTCGTTTTATTGACTTCGC 2100
-----
AGATGTTATGCTTTTTACAAATTGAACGCTTGTGGAAACGCTGTGTAAGCATGTTTCGTCG 2160
-----
GCATCATTATCCAACAGCGTGTGTTGACTTCGTGTCTCTGTGTAGCATTTTGATTATTC 2220
-----
TCACAGTATCCAGATGTTTTATTATTATCATGCTGTGATAGTGATCTGTATCAGTG 2280
-----
ATCTTTGATGTTACTATTGTATTGTTGGGGTCCCTACGAACTGCACCCATATAAGACA 2340
-----
GAAAACCTAATCAATAAATGTGCGTGTCTTACTGCAAAAAAAAAA 2385

```

FIGURE 1. Complete nucleotide sequence of CR1 cDNA insert and derived amino acid sequence. Indicated are the signal peptide region (*wavy line*), known CR1 peptide sequences (*broken line*), the 5' donor splice sequence (*overlined*), and the AATAAA polyadenylation signal (*overlined*). Arrows mark the region that overlaps the 5' end of the partial CR1 sequence described by Klickstein et al. (15). Also shown is a schematic of the insert and a summary of the sequencing strategy used. These sequence data have been submitted to the EMBL/GenBank Data Libraries under the accession number Y00812.

Q	N	A	P	E	W	L	P	F	A	R	P	T	N	L	T	D	E	F	E	F	P	I	G	T	L	N	Y	E	R	P	G	Y	S	G	R	P	F	S	I	I	G	L	K	N	S	V	M	T	G	A	K	D	R	R															
K	S	R	N	P	P	D	P	V	N	G	M	V	H	V	I	K	G	I	Q	F	G	S	Q	I	K	Y	S	C	T	K	Q	Y	R	L	I	S	S	S	A	T	C	I	I	S	G	D	T	V	I	N	D	N	E	T	P	I	C	R											
I	P	C	L	P	P	T	I	T	N	G	D	F	I	S	T	N	R	E	N	F	H	Y	G	S	V	T	T	Y	N	C	H	P	G	S	Q	G	R	K	V	F	E	L	V	G	E	P	S	I	Y	C	T	S	N	D	D	Q	V	G	I	N	S	G	P	A	P	C	I	I	P
N	K	C	T	P	P	N	V	E	N	G	I	L	V	S	D	N	R	S	L	F	S	L	N	E	V	V	E	F	N	C	O	P	G	F	V	M	K	E	P	R	R	V	K	C	Q	A	L	N	K	W	E	P	E	L	P	S	C	S											
R	V	C	P	P	P	D	V	L	H	A	E	R	T	G	R	D	K	D	N	F	S	P	G	O	E	V	F	Y	C	E	P	Q	Y	D	L	R	G	A	A	S	M	R	C	T	P	Q	G	D	N	S	P	A	A	P	T	G	E	V											
K	S	C	D	F	M	G	Q	L	L	N	G	R	V	L	F	P	V	N	L	Q	L	G	A	K	V	D	F	C	D	E	G	F	Q	L	K	S	S	A	S	Y	C	V	L	A	G	M	E	S	L	M	N	S	S	V	P	V	G	E	Q										
I	F	C	S	P	P	P	I	P	N	G	R	H	T	G	K	L	E	V	F	F	P	G	K	A	V	N	T	C	D	P	H	P	D	R	G	T	S	F	D	L	I	Q	E	S	T	I	R	C	T	S	D	P	Q	G	N	G	V	M	S	S	P	A	P	R	C	I	L		
G	M	C	A	A	P	D	H	F	A	K	L	K	T	O	T	N	A	S	D	F	I	G	T	S	L	K	Y	E	C	R	P	E	Y	Y	G	R	P	F	S	I	T	G	L	D	N	L	V	M	S	S	P	K	D	V	E	K	R												
K	S	C	K	T	P	P	D	P	V	N	G	M	T	D	I	Q	V	G	S	R	I	N	Y	S	C	T	T	G																																									

FIGURE 2. Alignment of the SCRs in the amino acid sequence derived from the CR1 cDNA insert. The invariant residues are boxed and a consensus sequence, which indicates both the invariant and the frequently conserved residues, is shown below.

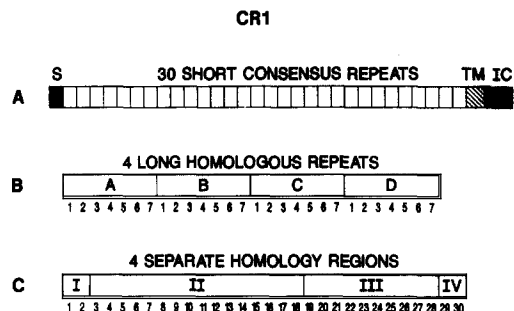


FIGURE 3. Schematic representation of CR1. (S) Signal peptide; (TM) transmembrane spanning region; and (IC) intracytoplasmic domain. Numbers refer to SCRs in the different LHRs (B) and homology regions (C).

The complex homologies among the SCRs in CR1 suggest specific evolutionary relationships. The arrangement of the first 28 SCRs into LHRs implies that this region arose by duplication of a common seven SCR ancestral unit (15). Once multiple LHRs were established, however, additional events would have been required to maintain the strict repetition in regions II and III. Highly homologous sequences in CR1, such as those seen in regions II and III, could be established by duplication and then maintained by unequal crossover and/or gene conversion (13, 15). Unequal crossover in region I and/or III could lead to the polymorphic variants seen in CR1 (13–15). Region I apparently diverged from within the LHR organization, avoiding the process of tandem evolution so apparent in the adjacent areas. Region IV is not part of the LHR organization and could have been a contemporary of the proposed seven SCR ancestral unit or it could have appeared afterwards.

The divergence of sequence found in region I is of particular interest because, by analogy with H, another member of the C3b/C4b-binding superfamily, this region is the likely location of several functional sites. The evidence suggests that it is the NH₂ terminal five or six SCRs (out of total of 20 SCRs) of H that carry C3b-binding domains as well as cofactor activity (30, 31). In addition, both H and C4bp are thought to be highly elongated, semi-rigid structures (32–34), and electron microscopic analysis has shown that C4b attaches near the tips of the C4bp tentacles (33). Thus, at one extreme, the NH₂-terminal region of CR1 could be involved in cofactor activity, decay accelerating activity, and C3b/C4b binding while the remainder of the protein would form an arm protruding from the cell surface. Experiments with proteolytic fragments of C4bp have shown, however, that binding and cofactor activity may be assigned to different internal SCRs (35). Therefore, the active sites in CR1 may be more evenly distributed; for example, monovalent functions could be mediated at the NH₂ terminus in region I and near the cell membrane in region IV, while multivalent functions could be mediated by the internal regions II and III. It is also possible that each separate homology region constitutes a separate functional domain. In any case, the unique role of the NH₂-terminal region in the function of CR1 postulated here would be a selective force instrumental in maintaining the divergence of region I from the adjacent regions.

Alternative Polyadenylation in the CR1 Transcriptional Unit. The open reading frame found in our derived sequence terminates at a stop codon nested within a 5' donor splice sequence (CAGG/GTGAGT) and the cDNA ends with a polyadenylation signal and a poly(A) tail. Since the composite sequence of Klickstein et al. (15) extends

the open reading frame well beyond this point, it is possible that the end of our open reading frame corresponds to an exon/intron junction in the CR1 gene. By this model, our cDNA clone would have been derived from an alternatively processed CR1 transcript, truncated at an alternative polyadenylation site located in the adjacent intron (Fig. 4).

The most straightforward means of testing this hypothesis is to compare this part of the CR1 gene with the cDNA. If correct, there must be a corresponding genomic sequence that extends from the coding region for the NH₂-terminal half of the SCR-9, through the putative exon/intron junction, and the putative intron sequence to the proposed polyadenylation site where divergence from the cDNA must occur. To this end we generated a 586-bp DNA restriction fragment that lies entirely within the proposed intron sequence (cDNA bp 1,799–2,385; Fig. 1). Genomic Southern hybridization using this fragment as probe yielded a pattern indicating at least two highly related genomic copies and several distantly related sequences (Fig. 5).

Screening four cosmid libraries and two λ genomic libraries with our 586-bp probe, we isolated three clones. These clones defined two nonoverlapping genomic regions. Further analysis suggests that together they represent the genomic regions that bear the most homology to our probe (Fig. 5).

DNA sequencing of one of these genomic regions, found on a λ genomic clone, yielded a segment with 99% homology to part of the cDNA sequence (Fig. 6), including a potential exon encoding the NH₂-terminal half of SCR-9 followed by the putative intron in our cDNA clone, divergence at the proposed polyadenylation site, and a potential exon encoding the COOH-terminal half of SCR-9 found in non-truncated cDNAs (15). Two mismatches in the translated region, however, found in the COOH-terminal half of the SCR, are consistent with the published CR1 SCR-16 sequence, which differs from SCR-9 only by these same base pairs. A third mismatch, found in the NH₂-terminal half of the SCR (Fig. 6), is not found in the known SCR-9 or SCR-16 cDNA sequence. It is most likely a polymorphic variation, since this clone was isolated from a different genetic source than the cDNAs. Thus, although we cannot rule out the possibility of a separate gene, this first genomic region most likely carries the two exons encoding CR1 SCR-16 and the accompanying intron.

The second genomic region, found on two different cosmids, was also sequenced (Fig. 6). It, too, contains extensive homology to our cDNA, including a potential exon encoding a sequence similar to the NH₂-terminal half of SCR-9/SCR-16, and

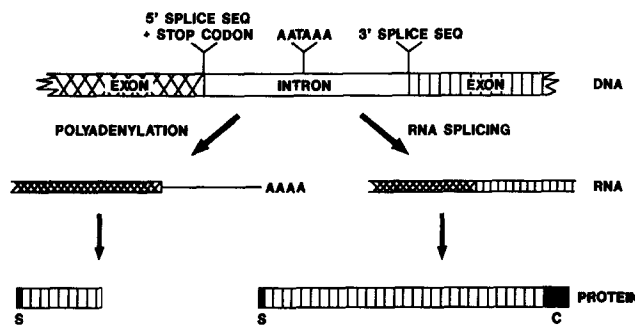


FIGURE 4. Proposed model of alternative polyadenylation and RNA splicing in the CR1 transcriptional unit and the predicted polypeptide products. (S) signal peptide; (C) transmembrane region and cytoplasmic domain.

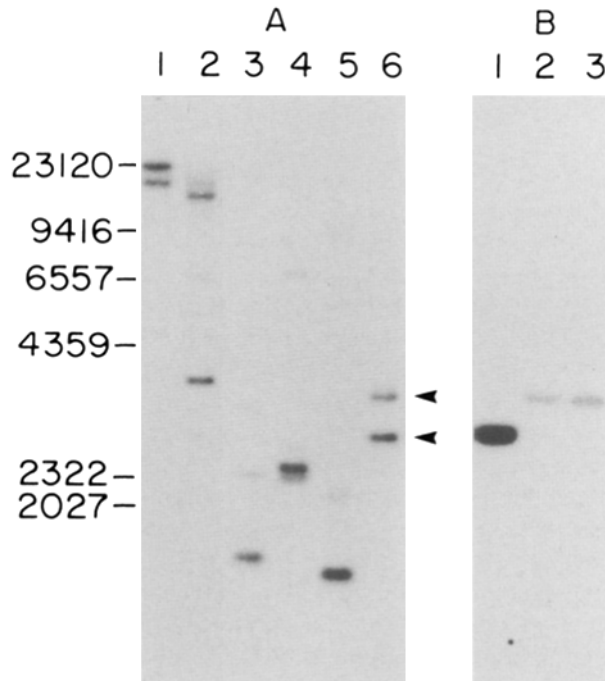


FIGURE 5. (A) Southern blot of EB19 genomic DNA using a 584-bp probe. 5 μ g DNA was cut with Eco RI (lane 1), Eco RV (lane 2), Dra I (lane 3), Ssp I (lane 4), Hind III (lane 5), and Bam HI (lane 6). All lanes were from the same 1% agarose gel. Exposure was overnight. (B) Southern blot of cloned genomic DNA. λ genomic clone 1 (lane 1), cosmid 4 (lane 2), and cosmid 5.1 (lane 3) were cut with Bam HI and run on a similar gel as in A. Southern blot was hybridized to the same probe and exposure was overnight. Comparison with size markers (shown on left of A) in each gel showed that Bam HI fragments from the clones in B were of the same mobility as the major bands in genomic Bam HI fragments in A, lane 6 (see arrowheads).

the correspondence continues through the apparently untranslated region with divergence at the proposed polyadenylation site. It also contains a potential exon encoding a sequence similar to the COOH-terminal half of SCR-9/SCR-16. The homologies, however, averaged 95% at the nucleotide level. Additional sequencing (unpublished data) revealed a number of possible exons that together exhibit similar homology to most of the NH₂-terminal coding region of CR1. Because the homology at the amino acid level averages 90%, we conclude that this sequence is not part of the CR1 gene, but from a highly related genomic region. It is likely to be the same region observed by Wong et al. (36), since it exhibits a similar restriction pattern.

The genomic sequences we have found lend strong support to our proposal of alternative polyadenylation in the CR1 transcriptional unit. As already described, the cDNA of CR1 is composed of long homologous repeats of seven SCRs with SCR-16 differing from SCR-9 by only 2 bp. Previous work suggests that this homology extends to the genomic level (14). Thus any comparison using the SCR-16 coding region should also be valid for the SCR-9 coding region. In our CR1 genomic clone (as well as our CR1-like genomic clone), the sequence that encodes the NH₂-terminal half of SCR-16 is interrupted by an exon/intron junction at the same site as the SCR-9 sequences in our cDNA, continues with sequence nearly identical to the putative intron in the cDNA, and diverges from the cDNA sequence at the proposed polyadenylation site. This structure is predicted for SCR-9 coding region by the alternative polyadenylation hypothesis (Fig. 4). No other RNA processing would be necessary to generate the truncated cDNA.

Northern Blot Analysis. Several DNA fragments were generated from our cDNA


```

CR1 GENOMIC SEQUENCE          CATTTCCTTTCCACAG/GTAAATCATGTA
TRUNCATED cDNA                CAAAGATGCTGTAAAC/-----
CR1-LIKE GENOMIC SEQUENCE     -----/---G-----

30  AAATCCTCCAGATCCAGTGAATGGCATGGTGCATGTGATCACAGACATCCATGTTGGAT
    -----T-----G-----

90  CCAGAATCAACTATTCTTGTACTACAGG/GTGAGTTGGCAGCAACATCTCTGGTTTAAAG
    -----/-----CC--

149 AGTTCAGCACAGCGATAGTACTTTCTAGCCACATCTCAGCAAGGAACTAGGCTATTGC
    -----A--C--C-----G-----G--

209 CTACCTGCTCTTAAGAGGCTTGAACACAGGTGTTAACTCCTGATTGAAATGAACAAAGAT
    ( )-----G-----T-----A-----
    ( )-----

269 AGGAGAAGATTAGGGGAAAACTGTATCCTTGCTGGAAACCAGGGCAGTGCACATATAA
    -----A-----

329 AGAGTATGCTGTTCAGTGGATGGGAAGGAAAAAATTAGAAGTGTAGTAGTCAAAGCAC
    ( )-----A-----C-----G--A--
    (-T-----C-----

389 ACAAAACAACCTAACCCAGAGTAGACATTGCTGGAAGAAAGGGAAGCATTGTAGCAGCT
    -----GG-----

449 GTGTGAGAGAATGAATCTTAATGATAACAGCATGATCCCTTGCTAGGGCTGCCATCAAAA
    -----CA-----G-----

509 AGTACAGGCCTTCCTCGTTTTATTGTACTTCGCAGATGTTATGCTTTTTACAAATGAAC
    -----A-----A-----G

569 GCTTGTGGGAACGCTG  TGTAAAGCATGTTGCTCGGCATCATTATCCAACAGCGTGTGT
    -----C-----CAAC----- ( )A-T-----T-----

628 GACTTCGTGCTCTGTGTAGCATTTTATTCTCAGATATCCAGATGTTTTCATTA
    -----G-----A-----G-----

688 TTATCATGTCTGTGATAGTATCTGTATCAGTATCTTTGATGTTACTATTGTCAATTGT
    -----T-----T-----

748 TTGGGTCCTACGAACTGCACCCATATAAGACAGAAACTTAATCAATAAATGTGCGTGTG
    -----A----- ( )-----A--

808 CTTTGACTGCTCCATGGACTAGACATTCCCCTTCTGTCTCCC...
    -----AAAAAAAAA-----G-----...

1  CR1 GENOMIC SEQUENCE          ...TTTCATTTTTGCTTTAG/GCACCGACTC
NON-TRUNCATED cDNA SCR-9      ...AACTATTCTTGTACTACAGG/-----
CR1-LIKE GENOMIC SEQUENCE     ...-----/-----

31  ATGGTCACTCATCTGCTGAATGTATCCTCTCAGGCAATACTGCCATTGGAGCAGGAAG
    -----G-----G-----
    -----G-----T-----

91  CCGCCAATTTGTCAAC/GTGAATTGAA...
    -----/GAATTCCTTG
    --A-----/-----...
    
```

FIGURE 6. Comparison of portions of the CR1 genomic sequence with the CR1 cDNA and the CR1-like genomic sequences. Relative differences between the CR1 clone and the cDNAs and the CR1-like clone are noted. Parenthesis indicate a missing base. Diagonal slash indicates junctions between exons and introns. The sequence of the nontruncated cDNA is taken from reference 15. Sequencing of the genomic clones was performed with the aid of synthetic oligonucleotide primers generated from the previously determined cDNA sequences. These sequence data have been submitted to the EMBL/GenBank Data Libraries under the accession number Y00812.

clone and used to probe poly(A)-containing RNA isolated from HL-60 cells and from EBV-transformed lymphocytes. The 586-bp probe derived from the untranslated portion of the cDNA and used in the Southern blots shown in Fig. 5 was also homologous to several RNA species ranging in size from 3.0 to 4.4 kb (Fig. 7 B). As expected, it did not hybridize to the longer RNA species (7.3-11.6 kb, depending on the allotype) believed to encode the full-length CR1 receptor (11, 12, 17). That is because the intron sequence would be removed in the splicing that joins the RNA

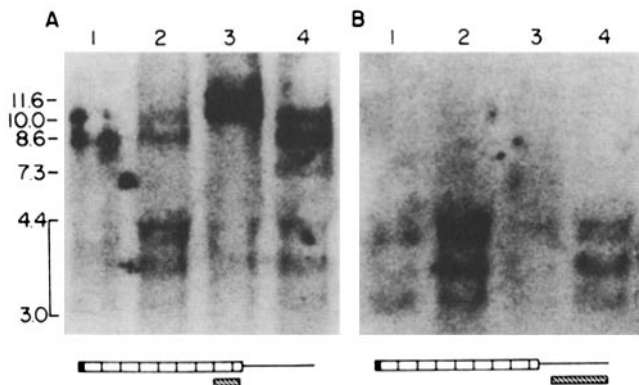


FIGURE 7. Northern blot analysis of cell lines that express CR1. The probe used in *A* was a 257-bp DNA fragment and the probe used in *B* was a 586-bp fragment also used in Fig. 5. They were each derived from the regions of the CR1 insert that are shown. The source of RNAs is as described in reference 13. (1) 4 μ g DMSO-induced HL-60 (CR1-A homozygote) RNA; (2) 15 μ g EB 19 (CR1-A homozygote) RNA; (3) 15 μ g EB22 (CR1-B homozygote) RNA; (4) 10 μ g EB1916 (CR1-A/CR1-C heterozygote) RNA. EB19, EB22, and EB1916 refer to EBV-transformed B lymphocyte cell lines derived from individuals expressing the indicated CR1 phenotype. The same blot was used for each probe and exposure was 6 d. Sizes of the hybridizing mRNAs (in kb) are at left.

segments encoding the two halves of SCR-9 (Fig. 4). In contrast, probes derived from the translated region are homologous to both full-length and truncated CR1 mRNA (Fig. 7 *A*).

The Northern blot results suggest that the CR1 transcriptional unit could produce a number of truncated RNA species. The largest of these species, 4.4 kb, could be an artifact because it migrates close to the 28S RNA band. Even so, several distinct transcripts remain. Our sequence data indicates that one of these is processed at SCR-9. Given the sequence similarity, one might expect another alternative polyadenylation site in SCR-16, yielding a second truncated messenger, 1,350 bases greater in length than the first one. Remaining messengers could be the result of additional alternative polyadenylation sites in the same introns or multiple transcriptional start sites. In addition, it is possible that the CR1-like genomic sequence described above could be the source of one or more of these transcripts, since the probes used in the Northern blots also hybridize to the CR1-like genomic region. Further work is required to understand the source of each of the short transcripts seen on the Northern blots.

Our Northern analysis yielded similar 3.0–4.4-kb species whether the RNA source carried the CR1-A polymorphism, the CR1-B polymorphism, or both the CR1-A and CR1-C polymorphisms, although the CR1-B transcripts were faint (Fig. 7). This result suggests that the CR1-A and the CR1-B alleles may carry the same alternative processing sites but does not preclude the possibility that CR1-C is lacking some or all of these sites.

In our previous work (13), hybridization of poly(A)-containing mRNA to probes encoding the CR1-translated region did reveal short species. At that time their relation to the CR1 gene was not known and it was possible that they represented hy-

bridization to other members of the gene family. Only recently has it become clear that CR1 nucleotide sequences do not hybridize at high stringency to the mRNAs of DAF and CR2 (3, 37). In a different report (17), hybridization of similar probes to tonsillar RNA apparently yielded no shorter species. This could be due to the difference in cell types.

Prediction of a Secreted Form of CR1. The polypeptide predicted from the cDNA clone is a signal sequence followed by the NH₂-terminal 8.5 SCRs. The first eight SCRs should retain the same functional capacity as they do in CR1 since their structure would probably not change. The last half-SCR segment, 35 amino acids in length, would not retain the conformation found in the full repeat, since it is believed that the cysteines at position 4 and 46 and the cysteines at positions 32 and 57 are linked by their side chains in the repeated domains, as in β_2 glycoprotein I (7, 38). The terminal-half SCR could be a candidate for proteolytic cleavage at the arginine or lysine residues that immediately precede it.

Since a short form of CR1 would lack the transmembrane sequence and cytoplasmic anchor found at the COOH terminus of the receptor (15), it would likely be secreted. It is possible that a secreted CR1 form would express complement-mediating activity that surface-bound forms could not.

Although a secreted form related to CR1 has been reported (39), it is much larger in size than any predicted from our work. There have been other studies, however, that have shown the presence of biosynthetically-labeled C3b-binding proteins of approximately the size appropriate for a 60-kD form in HL-60 cells (see Fig. 4 of reference 40). Alternatively, it is possible that the secreted form is not expressed at maximal levels in cells that are producing membrane CR1 and will only be found in other cell types, other tissues, or during specific stages of B lymphocyte development or differentiation.

Regulation of the CR1 Transcriptional Unit by Selective Polyadenylation. The use of two different processing events at an intron within SCR-2 of LHR-B appears to govern the production of two classes of CR1 messenger RNAs. Polyadenylation results in the shorter transcripts that could direct the synthesis of secreted CR1, while splicing is required for the production of longer transcripts that direct the synthesis of the CR1 receptor. In general, it appears that the addition of poly(A) to 3' ends of nuclear RNA occurs more rapidly than RNA splicing (41-44), so it is the control of polyadenylation at this site that probably determines the processing pathway used.

The selection of alternative polyadenylation sites within a transcriptional unit resulting in differential gene expression occurs in the regulation of the late transcriptional unit of adenovirus 2 (45) and in the tissue-specific expression of calcitonin and calcitonin gene-related peptide (46). Moreover, it also appears in the developmental progression from expression of surface IgM in B lymphocytes to the secretion of IgM in mature plasma cells (47). In each of these cases, as well as that of CR1, *trans*-acting factors could be involved in the selection of polyadenylation sites. The alternative polyadenylation site in CR1 could be used to regulate the ratio of full-length messenger to truncated messenger.

Evolution of the C3b/C4b-binding Protein Multigene Family. As mentioned, CR1 is a member of a "superfamily" of C3b/C4b-binding proteins, all of which possess tandemly repeated domains like the CR1 SCRs (7). Although some of the genes that encode these proteins are located on different chromosomes, CR1 is located in a

cluster at q3.2 on human chromosome 1 (3) along with C4bp, H, DAF, and CR2 (2-6). Members of this "immediate" family each manifest cofactor and/or decay-accelerating activity as well as C3b/C4b-binding capacity and include both surface and secreted proteins (7).

Evidence is accumulating that many of the SCRs in the superfamily are each encoded by a single exon (7). We report exceptions to that rule in CR1 (SCR-9 and SCR-16) and the SCR-9/SCR-16 homology in the CR1-like sequence. Furthermore, we find that the sequence of a second CR1 cDNA indicates the presence of another "split" SCR (SCR-20) in CR1 and examination of the CR1-like genomic clone reveals split exons encoding the SCR-2 and SCR-6/SCR-13 homologies (unpublished data). All of these divisions occur at the same position in the consensus sequence. In addition, murine H and C4bp genes each code an SCR split at the same site (48, 49). It is likely that all of these split SCRs are related through a common ancestor. Another split SCR is seen in the human haptoglobin sequence (50). In that case the intervening sequence appears to occur at a different position in the consensus sequence and therefore, is likely to have arisen through an independent event.

Two simple models could account for the split SCRs. A primordial exon encoding an entire SCR was interrupted through transposition by an intervening sequence. Alternatively, the primordial SCR was encoded by two separate but adjacent exons. Intron deletion might have led to composite ("unsplit") SCRs.

It will be of interest to learn whether other members of the C3b/C4b superfamily carry similarly positioned introns with alternative polyadenylation sites. It has already been seen in other members of the superfamily (DAF and H) that alternative RNA splicing can result in a transcript that would encode a second form (51-55).

It has been suggested, based on alternative polyadenylation in the production of surface and secreted forms of IgM from a single transcriptional unit, that regulated polyadenylation sites could have been instrumental in the evolution of secreted antibodies from surface receptors (56). It now appears possible that alternative polyadenylation could similarly have been instrumental in the evolution of the family of C3b/C4b binding proteins from a primitive receptor gene to the modern genes that encode secreted C3b/C4b-binding proteins as well as genes that produce both secreted and surface forms.

Summary

The human C3b/C4b receptor or complement receptor type one (CR1) is an ~200-kD single chain membrane glycoprotein of human peripheral blood cells that mediates the binding, processing, and transport of C3b-bearing immune complexes and regulates the activity of the complement cascade. Analysis of partial cDNA clones has shown that the COOH terminus is composed predominantly of three tandemly repeated regions of 450 amino acids each (15). In this report, we present a cDNA sequence that encodes the NH₂ terminus of CR1. It appears to have been derived from an alternatively processed transcript, caused by polyadenylation occurring at a site within an intron in the CR1 transcriptional unit. The resulting truncated messenger carries an open reading frame that would produce a short, secreted CR1 form. We present genomic sequences and Northern blots which support this hypothesis and we propose that the NH₂-terminal end of CR1 is a likely location for active

sites. In addition, we report evidence for a CR1-like sequence in the human genome and we present a model for the organization of CR1.

We thank Stanley Korsmeyer and David Chaplin for their helpful reviews of the manuscript, Lisa Westfield for synthesis of oligonucleotide primers, and Pat Parvin and Lorraine Whitely for excellent secretarial assistance.

Received for publication 11 August 1987 and in revised form 8 July 1988.

References

1. Ross, G. D., and M. E. Medof. 1985. Membrane complement receptors specific for bound fragments of C3. *Adv. Immunol.* 37:217.
2. DeCordoba, S. R., D. M. Lublin, P. Rubinstein, and J. P. Atkinson. 1985. Human genes for three complement components that regulate the activation of C3 are tightly linked. *J. Exp. Med.* 161:1189.
3. 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 C3d/Epstein Barr virus receptor map to human chromosome 1. *J. Immunol.* 138:312.
4. Lublin, D. M., R. S. Lemons, M. M. LeBeau, V. M. Holers, M. L. Tykocinski, M. E. Medof, and J. P. Atkinson. 1987. The gene encoding decay-accelerating factor (DAF) is located in the complement-regulatory locus on the long arm of chromosome 1. *J. Exp. Med.* 165:1731.
5. Rey-Campos, J., P. Rubinstein, and S. Rodriguez de Cordoba. 1988. A physical map of the human regulator of complement activation gene cluster linking the complement genes CR1, CR2, DAF, and C4BP. *J. Exp. Med.* 167:664.
6. Carroll, M. C., E. M. Aliquot, P. J. Katzman, L. B. Klickstein, J. A. Smith, and D. T. Fearon. 1988. Organization of the genes encoding complement receptors type 1 and 2, decay-accelerating factor, and C4-binding protein in the RCA locus on human chromosome 1. *J. Exp. Med.* 167:1271.
7. Kristensen, T., P. D'Eustachio, R. T. Ogata, L. Ping Chung, K. B. M. Reid, and B. F. Tack. 1987. The superfamily of C3b/C4b-binding proteins. *Fed. Proc.* 46:2463.
8. 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.
9. Wong, W. W., J. C. Wilson, and D. T. Fearon. 1983. Genetic regulation of a structural polymorphism of human C3b receptor. *J. Clin. Invest.* 72:685.
10. 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.
11. 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.
12. Van Dyne, S., V. M. Holers, D. M. Lublin, and J. P. Atkinson. 1987. The polymorphism of the C3b/C4b receptor in the normal population and in patients with systemic lupus erythematosus. *Clin. Exp. Immunol.* 68:570.
13. Holers, V. M., D. D. Chaplin, J. F. Leykam, B. A. Gruner, V. Kumar, and J. P. Atkinson. 1987. Human complement C3b/C4b receptor (CR1) mRNA polymorphism that correlates with the CR1 allelic molecular weight polymorphism. *Proc. Natl. Acad. Sci. USA.* 84:2459.
14. 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 polymor-

- phisms 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.
15. Klickstein, L. B., W. W. Wong, J. A. Smith, J. H. Weis, J. G. Wilson, and D. T. Fearon. 1987. Human C3b/C4b receptor (CR1). Demonstration of long homologous repeating domains that are composed of the short consensus repeats characteristic of C3b/C4b binding proteins. *J. Exp. Med.* 165:1095.
 16. Maniatis, T., E. F. Fritsch, and J. Sambrook. 1982. *Molecular Cloning: A Laboratory Manual*. Cold Spring Harbor Laboratory, Cold Spring Harbor, NY. 545 pp.
 17. Sanger, F., S. Nickles, and A. R. Coulson. 1977. DNA sequencing with chain-terminating inhibitors. *Proc. Natl. Acad. Sci. USA.* 74:5463.
 18. Lublin, D. M., R. C. Griffith, and J. P. Atkinson. 1986. Influence of glycosylation on allelic and cell-specific Mr variation, receptor processing, and ligand binding of the human complement C3b/C4b receptor. *J. Biol. Chem.* 261:5736.
 19. Grosveld, F. G., T. Lund, E. J. Murray, A. L. Mellor, H. H. M. Dahl, and R. A. Flavell. 1982. The construction of cosmid libraries which can be used to transform eukaryotic cells. *Nucleic Acids Res.* 10:6715.
 20. Lawn, R. M., E. F. Fritsch, R. C. Parker, G. Blake, and T. Maniatis. 1978. The isolation and characterization of linked δ and β -globin genes from a cloned library of human DNA. *Cell.* 15:1157.
 21. Blattner, F. R., B. G. Williams, A. E. Blechl, K. Denniston-Thompson, H. Faber, L. A. Furlong, D. J. Grunwald, D. O. Kiefer, D. D. Moore, J. W. Schumm, E. L. Sheldon, and O. Smithies. Charon phages: safer derivatives of bacteriophage lambda for DNA cloning. 1977. *Science (Wash. DC).* 196:161.
 22. Feinberg, A. P., and B. Vogelstein. 1983. A technique for radiolabeling DNA restriction endonuclease fragments to high specific activity. *Anal. Biochem.* 132:6.
 - 22a. 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.
 23. Von Heijne, G. 1985. Signal sequences. The limits of variation. *J. Mol. Biol.* 184:99.
 24. Holers, V. M., T. Seya, E. Brown, J. J. O'Shea, and J. P. Atkinson. 1986. Structural and functional studies on the human C3b/C4b receptor (CR1) purified by affinity chromatography using a monoclonal antibody. *Complement.* 3:63.
 25. Wong, 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.
 26. Podell, D. N., and G. N. Abraham. 1978. A technique for the removal of pyroglutamic acid from the amino terminus of proteins using calf liver pyroglutamate amino peptidase. *Biochem. Biophys. Res. Commun.* 81:176.
 27. Mount, S. M. 1982. A catalogue of splice junction sequences. *Nucleic Acids Res.* 10:59.
 28. Proudfoot, N. J., and G. G. Brownlee. 1976. 3' non-coding region sequences in eukaryotic mRNA. *Nature (Lond.).* 263:211.
 29. Benoist, C., K. O'Hare, R. Breathnach, and P. Chambon. 1980. The ovalbumin gene-sequence of putative control regions. *Nucleic Acids Res.* 8:127.
 30. Hong, V., T. Kinoshita, Y. Dohi, and K. Inoue. 1982. Effect of trypsinization on the activity of human factor H. *J. Immunol.* 129:647.
 31. Alsenz, J., T. F. Schulz, J. D. Lambris, R. B. Sim, and M. P. Dierich. 1985. Structural and functional analysis of the complement component Factor H with the use of different enzymes and monoclonal antibodies to Factor H. *Biochem. J.* 232:841.
 32. DiScipio, R. G., and T. E. Hugli. 1982. Circular dichroism studies of human Factor H. A regulatory component of the complement system. *Biochim. Biophys. Acta.* 709:58.

33. Dahlback, B., C. A. Smith, and H. J. Muller-Eberhard. 1983. Visualization of human C4b-binding protein and its complexes with vitamin K-dependent protein S and complement protein C4b. *Proc. Natl. Acad. Sci. USA.* 80:3461.
34. Perkins, S. J., L. P. Chung, and K. B. M. Reid. 1986. Unusual ultrastructure of complement-component-C4-binding protein of human complement by synchrotron x-ray scattering and hydrodynamic analysis. *Biochem. J.* 233:799.
35. Chung, L. P., and K. B. M. Reid. 1985. Structural and functional studies on C4b-binding protein, a regulatory component of the human complement system. *Biosci. Rep.* 5:855.
36. Wong, W., J. Cahill, C. Kennedy, E. Bonaccio, J. Wilson, L. Klickstein, L. Rabson, and D. Fearon. 1987. Analysis of genomic polymorphisms in the human CR1 gene. *Complement.* 4:240.
37. Medof, M. E., D. M. Lublin, V. M. Holers, D. J. Ayers, R. R. Getty, J. F. Leykam, J. P. Atkinson, and M. L. Tykocinski. 1987. Cloning and characterization of cDNAs encoding the complete sequence of decay-accelerating factor of human complement. *Proc. Natl. Acad. Sci. USA.* 84:2007.
38. 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.
39. 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.
40. Atkinson, J. P., and E. A. Jones. 1984. Biosynthesis of the human C3b/C4b receptor during differentiation of the HL-60 cell line. Identification and characterization of a precursor molecule. *J. Clin. Invest.* 74:1649.
41. Nevins, J. R., and J. E. Darnell Jr. 1978. Steps in the processing of Ad2 mRNA: Poly(A)⁺ nuclear sequences are conserved and poly(A) addition precedes splicing. *Cell.* 15:1477.
42. Lai, C. J., R. Dhar, and G. Khoury. 1978. Mapping the spliced and unspliced late lytic SV40 RNAs. *Cell.* 14:971.
43. Schibler, U., K. B. Marcu, and R. P. Perry. 1978. The synthesis and processing of the messenger RNAs specifying heavy and light chain immunoglobulins in MPC-11 cells. *Cell.* 15:1495.
44. Gilmore-Hebert, M., and R. Wall. 1979. Nuclear RNA precursors in the processing pathway to MOPC 21K light chain messenger RNA. *J. Mol. Biol.* 135:879.
45. Nevins, J. R., and M. C. Wilson. 1981. Regulation of adenovirus-2 gene expression at the level of transcriptional termination and RNA processing. *Nature (Lond.).* 290:113.
46. Rosenfeld, M. G., S. G. Amara, and R. M. Evans. 1984. Alternative RNA processing: determining neuronal phenotype. *Science (Wash. DC).* 225:1315.
47. Mather, E. L., K. J. Nelson, J. Haimovich, and R. P. Perry. 1984. Mode of regulation of immunoglobulin μ - and S-chain expression varies during β -lymphocyte maturation. *Cell.* 36:329.
48. Barnum, S., J. Kenney, T. Kristensen, D. Noack, M. Seldon, P. D'Eustachio, D. Chaplin, and B. Tack. 1987. Chromosomal location and structure of the mouse C4BP gene. *Complement.* 4:131.
49. Vik, D. P., J. B. Keeney, S. Bronson, B. Westlund, T. Kristensen, D. D. Chaplin, and B. F. Tack. 1987. Analysis of the murine Factor H gene and related DNA. *Complement.* 4:235.
50. Maeda, N., F. Yang, D. R. Barnett, B. H. Bowman, and O. Smithies. 1984. Duplication within the haptoglobin Hp² gene. *Nature (Lond.).* 309:131.
51. Caras, I. W., M. A. Davitz, L. Rhee, G. Weddell, D. W. Martin, and V. Nussenzweig. 1987. Cloning of decay-accelerating factor suggests novel use of splicing to generate two proteins. *Nature (Lond.).* 325:545.
52. Schultz, T. F., W. Schwable, K. K. Stanley, E. Weiss, and M. P. Dierich. 1986. Human complement factor H: isolation of cDNA clones and partial cDNA sequence of the 38-

- kDa tryptic fragment containing the binding site for C3b. *Eur. J. Immunol.* 16:1351.
53. 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.
 54. Ripoché, J., A. J. Day, B. Moffatt, and R. B. Sim. 1987. Messenger RNA coding for a truncated form of human complement factor H. *Biochem. Soc. Trans.* 15:651.
 55. Ripoché, J., A. J. Day, T. J. R. Harris, and R. B. Sim. 1988. The complete amino acid sequence of human complement factor H. *Biochem. J.* 249:593.
 56. Early, P., J. Rogers, M. Davis, K. Calame, M. Bond, R. Wall, and L. Hood. 1980. Two mRNAs can be produced from a single immunoglobulin μ gene by alternative RNA processing pathways. *Cell.* 20:313.