Proc. Natl. Acad. Sci. USA
Vol. 84, pp. 9194–9198, December 1987
Immunology

Molecular cloning of the cDNA encoding the Epstein–Barr virus/C3d receptor (complement receptor type 2) of human B lymphocytes

(nucleotide sequence/amino acid sequence/internal repeat/viral envelope gp350/220)

MARGARET D. MOORE, NEIL R. COOPER, BRIAN F. TACK, AND GLEN R. NEMEROW

Department of Immunology, Scripps Clinic and Research Foundation, 10666 North Torrey Pines Road, La Jolla, CA 92037

Communicated by Hans J. Müller-Eberhard, September 2, 1987

ABSTRACT **Complementary DNA clones for complement receptor type 2 (CR2), the B-lymphocyte membrane protein that serves as the receptor for Epstein–Barr virus and the C3d complement fragment, were obtained by screening a λgt11 library generated from Raji B lymphoblastoid cell mRNA. A 4.2-kilobase (kb) clone, representing the entire coding sequence of the protein plus untranslated 5′ and 3′ nucleotide sequences was obtained and sequenced. The 4.2-kb clone, which contains all but about 500 base pairs (bp) of the 5′ untranslated region of the full-length CR2 mRNA, consists of 63 bp of 5′ untranslated nucleotide sequence followed successively by a start codon, a 20-amino acid hydrophobic signal peptide, 1005 amino acids having a repeating motif, a 28-amino acid probable transmembrane domain, and a 34-amino acid cytoplasmic tail. The deduced amino acid sequence of the protein indicates that the extracellular domain consists entirely of 16 tandemly arranged repeating elements, each 60–75 amino acids in length, which are identified by multiple conserved residues. This repeating motif also occurs in the C3b/C4b receptor, several complement proteins, and a number of noncomplement proteins. In CR2, the 16 repeats occur in four clusters of four repeats each. Approximately 10% of the deduced amino acid sequence, including the amino and carboxyl termini, was confirmed by amino acid sequencing of tryptic peptides derived from purified CR2. The nucleotide and derived amino acid sequence of CR2 and related studies are presented here.**

A M_r 145,000 B-lymphocyte membrane glycoprotein, designated complement receptor type 2 (CR2), serves as the receptor for Epstein–Barr virus (EBV) and the C3d and C3dg fragments of the third component of complement (1–5). An EBV viral envelope protein termed gp350/220 mediates the binding of EBV to CR2 (6, 7). Gp350/220 and C3d, the natural ligands, exhibit two regions of primary sequence similarity, a finding that suggests that common domains in these two proteins mediates binding to CR2. In addition to its role in permitting EBV infection, CR2 has been implicated in triggering B-cell activation (7–11); consistent with this finding, CR2 has been found to be phosphorylated upon treatment of B cells with phorbol myristate or anti-Ig (12).

Previous studies have shown that the mature M_r 145,000 CR2 molecule consists of a M_r 111,000 polypeptide chain and multiple N-linked oligosaccharides (13). A previous study (14) reported the isolation of a partial CR2 cDNA clone and indicated that CR2 is highly similar to CR1, the C3b/C4b receptor; however, relatively limited nucleotide and protein sequence information was presented in that publication. The genes encoding both CR1 and CR2 have been mapped to human chromosome 1, band q32 (15).

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.

To further define the structural and functional properties of CR2, cDNA clones encoding CR2 were isolated, and the complete amino acid sequence of the protein* was deduced. Sequence analysis of tryptic peptides derived from purified CR2 allowed verification of the amino and carboxyl termini of the receptor.

MATERIALS AND METHODS

CR2 Purification and Analysis. CR2 was isolated from detergent lysates of Raji B lymphoblastoid cells (provided by A. Theofilopoulos, Scripps Clinic, La Jolla, CA) by immunoaffinity chromatography as earlier published (3). Radiolabeled CR2 was subjected to complete trypsin digestion, and the resulting peptides were purified by reverse-phase HPLC as previously described (16). Amino acid sequence analysis was done on an Applied Biosystems (model 470A) gas-phase sequencer, or on a Beckman spinning-cup sequencer (model 890 M).

Construction and Screening of the cDNA Library. A cDNA library was constructed in λgt11 (Stratagene, San Diego, CA) as previously described (17) using mRNA isolated from Raji B lymphoblastoid cells. Complementary DNA was synthesized with a commercially available synthesis kit (cDNA synthesis system, Amersham). Packaging of recombinant cDNA clones with phage coat proteins was done as recommended by the manufacturer (Gigapack, Stratagene). The library consisted of 2.5×10^6 recombinants.

The cDNA library was initially probed with a ^{32}P-kinase labeled 39-mer oligonucleotide probe (14). High-density screening was done by hybridization on replicative nitrocellulose filters as described (18, 19). Positive phage plaques were replated and screened at plaque densities decreasing by a factor of 10 until uniformly positive hybridization signals were obtained.

Candidate CR2 cDNA clones were isolated from 85-mm plate lysates using affinity chromatography with anti-λ phage antibody (LambdaSorb, Promega Biotec, Madison, WI). Insert sizes were determined by digestion with *Eco*RI followed by agarose gel electrophoresis. Large-scale preparation of λ phage was done as described (20). The purified phage were digested with *Eco*RI, and the cDNA inserts were isolated by preparative gel electrophoresis.

Characterization and Sequencing of cDNA Clones. Nick-translated (nick-translation kit, Bethesda Research Labora-

Abbreviations: CR2, complement receptor type 2 (C3d receptor); CR1, complement receptor type 1 (C3b/C4b receptor); EBV, Epstein–Barr virus.

*The sequence reported in this paper is being deposited in the EMBL/GenBank data base (Bolt, Beranek, and Newman Laboratories, Cambridge, MA, and Eur. Mol. Biol. Lab., Heidelberg) (accession no. J03565).

tories) ^{32}P-labeled cDNA inserts were used to probe RNA (21) or Southern blots as described (22).

The isolated CR2 inserts were ligated into the *Eco*RI site of plasmid Bluescript (Stratagene) (23). Sequencing was done by the Sanger dideoxy chain-termination method (24) modified for double-stranded DNA as described by the manufacturer. Sequencing of the longer inserts was performed by the exonuclease III and mung bean nuclease method (25) modified for plasmid Bluescript.

DIAGON software analysis (ref. 26; release 2.1) was done on the completed amino acid and nucleotide sequence of CR2.

RESULTS AND DISCUSSION

Isolation and Characterization of λCR2 cDNA Clones. Approximately 5×10^5 recombinant λ phage were screened with a CR2 oligonucleotide probe, and two clones, designated λA11 and λF11, were identified. Both clones contained 1.2-kb inserts as indicated by digestion with *Eco*RI (Fig. 1*A*). The insert from λA11 was isolated, nick-translated with ^{32}P, and used to probe RNA blots derived from either T or B cells. As shown in Fig. 1*A*, the λA11 insert hybridized to a single 4.7-kb mRNA species from the Raji B cell line but did not recognize mRNA from the human T-cell-derived HSB-2 cell line. These results are in agreement with the reported size of CR2 mRNA (5 kb) derived from tonsil B cells and with the absence of detectable CR2 RNA in the HSB-2 cell line (14).

The CR2 cDNA insert from clone λA11 was nick-translated and used to reprobe the Raji library. Two additional clones, designated λE41 and λL11, were isolated by hybridization screening. As shown in Fig. 1*B*, the L11 clone contained a 1.2-kb insert, whereas the E41 clone contained three *Eco*RI-derived fragments of 1.8, 1.6, and 0.8 kb, respectively. The 1.8-kb fragment of E41 and the 1.2-kb insert of L11 hybridized to the A11 insert on Southern blot analysis (Fig. 1*B*). Further analysis of the λE41 clone by *Eco*RI digestion and ^{32}P-end-labeling followed by polyacrylamide gel electrophoresis failed to detect additional small (<500 bp) insert fragments. A summary of the alignment and restriction mapping of the λCR2 clones is shown in Fig. 1*C*.

Nucleotide and Deduced Amino Acid Sequence of CR2. The nucleotide sequence of CR2 cDNA was obtained by sequencing a series of nested deletions of the three *Eco*RI fragments of the λE41 clone, as well as by sequencing subclones generated by endonuclease restriction digestion (Fig. 1*C*). The entire sequence of both the λA11 and λE41 cDNA clones was determined; ≈90% of the coding region of the λE41 CR2 clone was sequenced on both strands.

The complete nucleotide and derived amino acid sequence of the E41 CR2 clone is shown in Fig. 2. The protein-coding sequence is preceded by 63 bp of untranslated nucleotide sequence followed by a start codon and a 20-amino acid hydrophobic signal peptide. The sequence surrounding the start codon is consistent with Kozak's rules for initiation (27). The predicted amino terminus of the native protein is isoleucine, as determined by Von Heijne's rules for signal-peptide cleavage (28). A tryptic peptide derived from the amino terminus of purified CR2, corresponding to residues 1–12 of the deduced sequence (Fig. 2) was identified. The mature protein contains 1067 amino acids, the extracellular domain of which is made up of 16 tandem repeat elements of 60–75 amino acids (see below). A putative 28-amino acid transmembrane anchor (residues 1005–1033), ascertained by

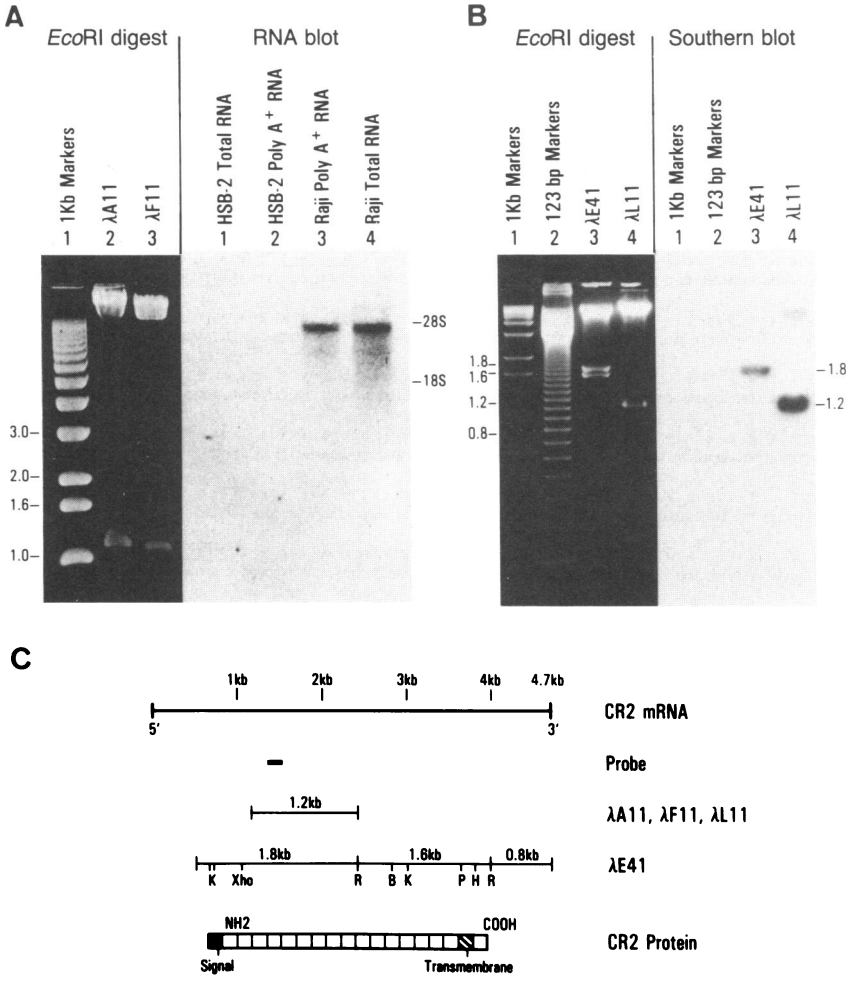
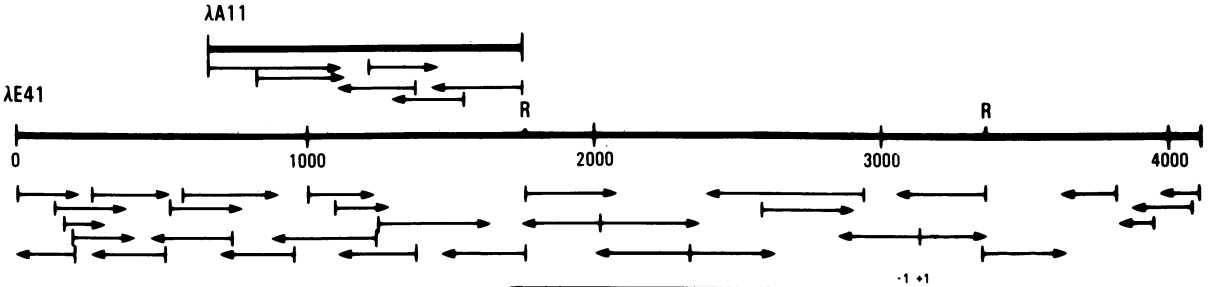


FIG. 1. Characterization of λCR2 cDNA clones. Partial λCR2 clones λA11 and λF11 cDNA were digested with *Eco*RI and electrophoresed on a 1.0% agarose gel that was stained with ethidium bromide (*A Left*). The nick-translated 1.2-kb insert from λA11 was used to probe RNA blots of total (20 μg) or poly(A)$^+$-selected (2 μg) Raji or HSB-2 cell RNA (*A Right*). Hybridization was done at 65°C in 6× SSC (1× SSC = 0.15 M NaCl/0.015 M sodium citrate) for 18 hr. Filters were washed with 0.1× SSC at 65°C. λCR2 cDNA clones E41 and L11 were digested with *Eco*RI and electrophoresed on a 1.5% gel (*B Left*). A Southern blot of this gel (*B Right*) was probed with ^{32}P-nick translated insert from λA11. Hybridization was done at 65°C and 6× SSC. Blots were washed with 0.1× SSC at 65°C. An alignment of the restriction map of the probe and CR2 cDNA clones, as well as the deduced protein, are indicated in relation to the predicted CR2 mRNA (C).



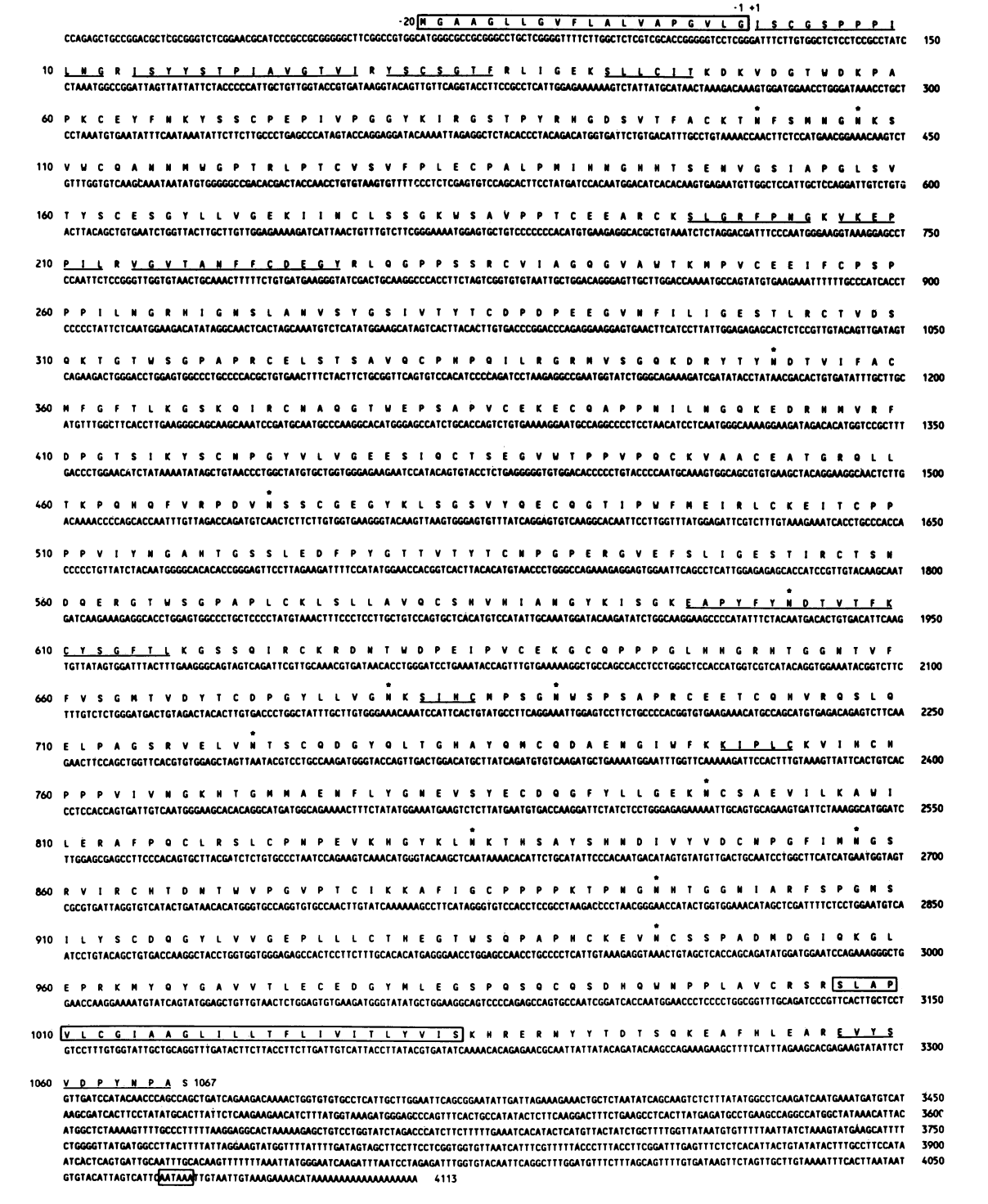


FIG. 2. Nucleotide and derived amino acid sequence of CR2. The position and orientation of the subclones of CR2 cDNA that were sequenced are indicated by arrows at the beginning of the sequence. CR2 nucleotides are numbered at right, whereas the amino acid residues are indicated by numbers at left. Positions of the signal peptide, transmembrane domain, and polyadenylylation sequence are indicated by the boxed residues. The locations of tryptic peptides derived from purified CR2, which were sequenced, are indicated by underlining. Potential sites for N-linked glycosylation are indicated by *.

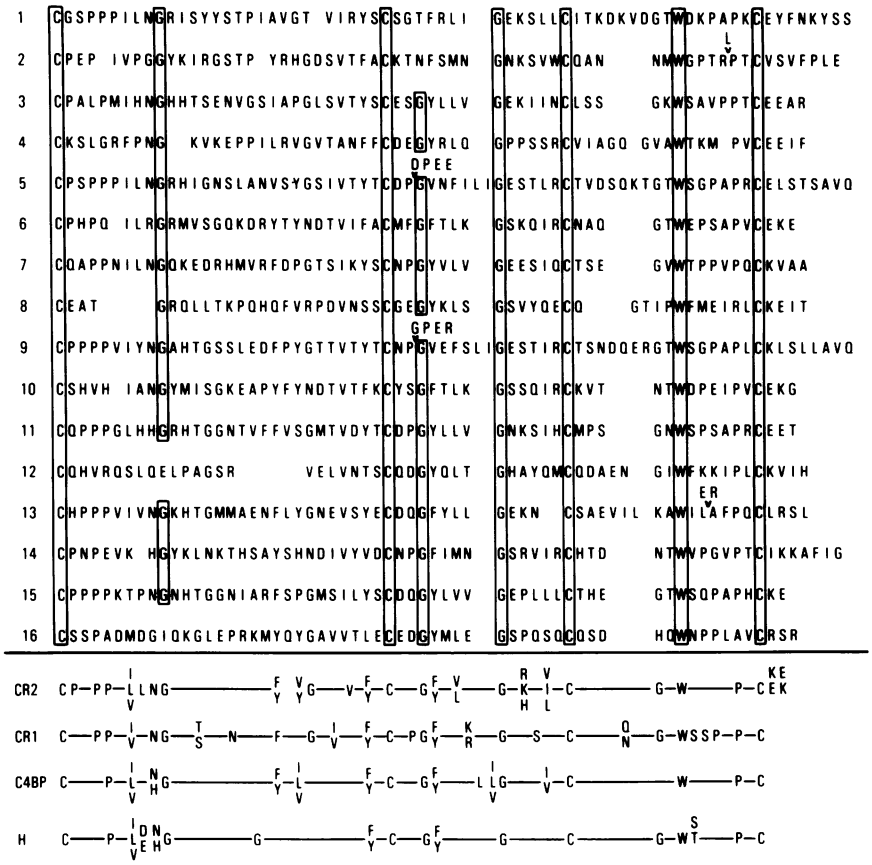


FIG. 3. Alignment of the 16 tandem repeating elements of CR2 and similarity with other C3 binding proteins (*Upper*). Highly conserved residues are indicated by the boxed amino acids, whereas other conserved residues are indicated by the consensus shown at bottom. CR2 consensus sequence is compared with the consensus of other C3 binding proteins.

computer-generated hydropathy analysis, is located just before the carboxyl terminus of the protein, indicating that the receptor is oriented in the membrane of B cells in an amino terminus (out)/carboxyl terminus (in) configuration, which is similar to many, but not all, eukaryotic membrane receptors (29). The probable cytoplasmic domain of CR2 consists of 34 residues, which include several potential sites for serine and threonine phosphorylation by protein kinases (30). A CR2 tryptic peptide conforming to that predicted from analysis of the λE41 clone was identified in these studies. Approximately 10% of the deduced amino acid sequence was confirmed by amino acid sequencing of tryptic peptides (Fig. 2).

The deduced 1067 amino acids of CR2 correspond to a protein of M_r 116,974; this size is close to the expected size of the nonglycosylated molecule (13). The receptor also contains 13 potential sites for N-linked glycosylation (Fig. 2).

The remaining 3′ 751 nucleotides of λE41 CR2 clone are untranslated and contain a polyadenylylation signal.

Internal Repeating Structure of CR2 and Similarity to Other Proteins. As shown in Fig. 3, the extracellular domain of CR2, excluding the signal, is composed of 16 tandemly arranged internal repeating units of 60–75 amino acids each. The presence of consensus repeating elements in other C3 binding proteins has been previously reported (31). This repeat motif structure contains four highly conserved cysteine residues, two conserved glycines, and one tryptophan residue, as well as other conserved residues as summarized in Fig. 3. Dot matrix analyses of the amino acid sequence shown in Fig. 4 show that CR2 exhibits an additional level of repeating structure because the 16 repeats occur in clusters of four: 1–4, 5–8, 9–12, and 13–16; of these repeat clusters 5–8 and 9–12 exhibit a higher degree of identity (51%) than the others (35–40%) (Fig. 4). The CR2 tandem repeating elements show varying degrees of similarity with other C3 and/or C4 binding proteins, including the C3b/C4b receptor (complement receptor type 1) (32), C4 binding protein (C4bp) (33), factor H (31), and factor B (34, 35). Because a number of other proteins that presumably lack C3 binding activity, such as the IL-2 receptor (36) and haptoglobin (37), also contain homologous repeating structures, the conserved residues within the repeating element probably do not play a direct role in ligand binding. More likely is that these residues facilitate the formation of independent folding domains necessary for the generation of an extended molecule as has been proposed for C4bp (31, 38) and CR1 (32).

As noted earlier, CR2 binds C3d and the gp350/220 glycoprotein of EBV. A region of amino acid sequence similarity in C3d and gp350/220 may well mediate the binding of these diverse proteins to CR2 (6, 7, 39). It will be of interest to determine whether CR2 interacts with C3d and gp350/220 via a specific sequence of amino acids and, if so, whether this sequence is found in a single repeating element. Furthermore, CR2 mediates internalization of EBV into B cells and signal transduction leading to B-cell activation. Analyses of the structural basis for these various functional activities will be facilitated by the elucidation of the complete sequence and domain structure of the molecule.

The authors express their gratitude to Dr. Georg Widera for his help in construction of the library and to Drs. Rick Wetzel and Martin F. E. Siaw for HPLC peptide purification. The authors also thank Cam Nguyen and Virginia Keivens-Schwend for excellent technical assistance and Bonnie Weier for preparation of the manuscript. Publication 4995-IMM. This project was funded by Public Health Services Grants: AI17354, CA14692, CA36204 and the PEW Scholars Award for Biomedical Research.

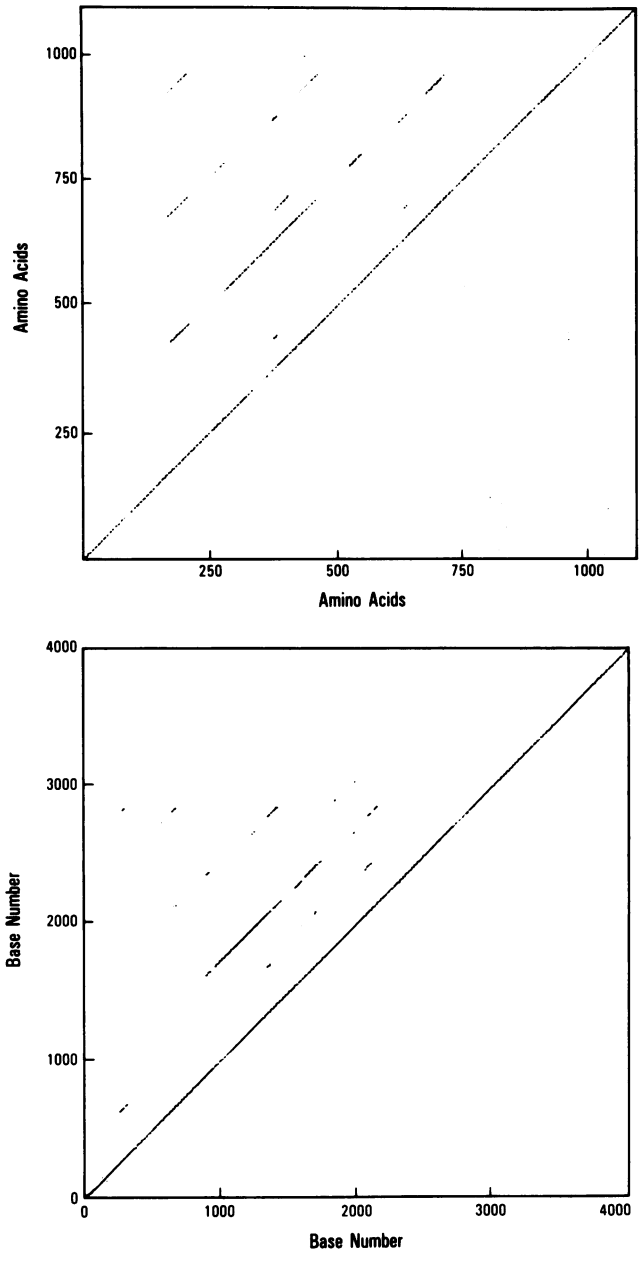


FIG. 4. DIAGON graphics analysis of the nucleotide and deduced amino acid sequence of CR2. To reveal repetitive sequences within the molecule, the entire sequence of CR2 was compared with itself and within itself. A dot was plotted when there was a score of 708 in an amino acid-score matrix for a span of 59 amino acids (*Upper*). For nucleotide analysis (*Lower*), a dot indicates when there was at least a 40 bp/91 bp match. The diagonal line bisecting the origin represents the identity of the sequence with itself. An additional parallel line of identity indicates similarity between two groups of four repeating elements, 5–8 and 9–12.

1. Fingeroth, J. D., Weis, J. J., Tedder, T. F., Strominger, J. L., Biro, P. A. & Fearon, D. T. (1984) *Proc. Natl. Acad. Sci. USA* **81,** 4510–4514.
2. Nemerow, G., Wolfert, R., McNaughton, M. & Cooper, N. R. (1985) *J. Virol.* **55,** 347–351.
3. Nemerow, G., Siaw, M. F. E. & Cooper, N. R. (1986) *J. Virol.* **58,** 709–712.
4. Mold, C., Cooper, N. R. & Nemerow, G. R. (1986) *J. Immunol.* **136,** 4140–4145.
5. Frade, R., Barel, M., Ehlin-Henriksson, B. & Klein, G. (1985) *Proc. Natl. Acad. Sci. USA* **82,** 1490–1493.
6. Nemerow, G. R., Mold, C., Keivens-Schwend, V., Tollefson, V. & Cooper, N. R. (1987) *J. Virol.* **61,** 1416–1420.
7. Tanner, J., Weis, J., Fearon, D., Whang, Y. & Kieff, E. (1987) *Cell* **50,** 203–213.
8. Frade, R., Crevon, M. C., Barel, M., Vazquez, A., Krikorian, L., Charriaut, C. & Galanaud, P. (1985) *Eur. J. Immunol.* **15,** 73–76.
9. Wilson, B., Platt, J. & Kay, N. (1985) *Blood* **66,** 824–829.
10. Melchers, F., Erdei, A., Schulz, T. & Dierich, M. P. (1985) *Nature (London)* **317,** 264–265.
11. Hutt-Fletcher, L. M. (1987) *J. Virol.* **651,** 774–781.
12. Changelian, P. S. & Fearon, D. T. (1986) *J. Exp. Med.* **163,** 101–115.
13. Weis, J. J. & Fearon, D. T. (1985) *J. Biol. Chem.,* **260,** 13824–13830.
14. Weis, 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.
15. Weis, J. H., Morton, C. C., Bruns, G. A. P., Weis, J. J., Klickstein, L. B., Wong, W. W. & Fearon, D. T. (1987) *J. Immunol.* **138,** 312–315.
16. Kristensen, T., Wetsel, R. A. & Tack, B. F. (1986) *J. Immunol.* **136,** 3407–3411.
17. Young, R. A. & Davis, R. W. (1983) *Science* **222,** 778–782.
18. Benton, W. D. & Davis, R. W. (1977) *Science* **196,** 180–182.
19. Maniatis, T., Fritsch, E. F. & Sambrook, J. (1982) *Molecular Cloning: A Laboratory Manual* (Cold Spring Harbor Laboratory, Cold Spring Harbor, NY).
20. Maniatis, T., Hardison, E., Lacy, E., Lauer, J., O'Connell, C., Quon, D., Sim, D. K. & Efstratiadis, A. (1978) *Cell* **15,** 678–701.
21. Thomas, P. S. (1980) *Proc. Natl. Acad. Sci. USA* **77,** 5201–5202.
22. Southern, E. M. (1975) *J. Mol. Biol.* **98,** 503–517.
23. Fernandez, J. M., Short, J. M., Renshaw, M., Huse, W. D. & Sorge, J. A. (1987) *Gene,* in press.
24. Sanger, F., Nicklen, S. & Coulson, A. R. (1977) *Proc. Natl. Acad. Sci. USA* **74,** 5463–5467.
25. Guo, L. H., Yang, R. C. A. & Wu, R. (1980) *Nucleic Acids Res.* **11,** 5521–5540.
26. Staden, R. (1982) *Nucleic Acids Res.* **10,** 2951–2961.
27. Kozak, M. (1986) *Cell* **44,** 283–292.
28. Von Heijne, G. (1983) *Eur. J. Biochem.* **133,** 17–21.
29. Wickner, L. & Lodish, H. (1985) *Science* **230,** 400–407.
30. Feramisco, J. R., Glass, D. B. & Krebs, E. (1980) *J. Biol. Chem.* **255,** 4240–4245.
31. 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,** 230–233.
32. Klickstein, L. B., Wong, W. W., Smith, J. A., Weis, J. H., Wilson, J. G. & Fearon, D. T. (1987) *J. Exp. Med.* **165,** 1095–1112.
33. Chung, L. P., Bently, D. R. & Reid, K. B. M. (1985) *Biochem. J.* **230,** 133–141.
34. Mole, J. E., Anderson, J. K., Davison, E. A. & Woods, D. E. (1984) *J. Biol. Chem.* **259,** 3407–3412.
35. Morley, B. J. & Campbell, R. D. (1984) *EMBO J.* **3,** 153–157.
36. Leonard, W. J., Depper, F. M., Kanehisa, M., Kronke, M., Peffer, J., Svetlik, S. P. B., Sullivan, M. & Greene, W. C. (1985) *Science* **230,** 633–639.
37. Kurosky, A., Barnett, D. R., Lee, T. H., Touchstone, B., Hay, R. E., Arnott, M. S., Bowman, B. H. & Fitch, W. M. (1980) *Proc. Natl. Acad. Sci. USA* **77,** 3388–3392.
38. Dahlbeck, B., Smith, C. A. & Müller-Eberhard, H. J. (1983) *Proc. Natl. Acad. Sci. USA* **80,** 3461–3465.
39. Lambris, J. D., Ganu, V. S., Hirani, S. & Müller-Eberhard, H. J. (1985) *Proc. Natl. Acad. Sci. USA* **82,** 4235–4239.