Brief Definitive Report

ANALYSIS OF TWO cDNA CLONES ENCODING THE B LYMPHOCYTE ANTIGEN CD20 (B1, Bp35), A TYPE III INTEGRAL MEMBRANE PROTEIN

By IVAN STAMENKOVIC and BRIAN SEED

From the Department of Genetics, Harvard Medical School, and the Department of Molecular Biology, Massachusetts General Hospital, Boston, Massachusetts 02114

Recent studies have suggested that the pan-B cell antigen CD20 (B1, Bp35) plays an important role in B cell activation. mAbs to CD20 induce different cellular responses depending on the antibody used and the stage of differentiation or activation of the target B cells. The mAb 1F5 activates resting B cells by initiating the transition from the G_0 to the G_1 phase of the cell cycle, and induces dense tonsillar B cells to proliferate (1, 2). Other anti-CD20 mAbs such as B1 have been shown to block B cell activation (3) and both 1F5 and B1 can inhibit B cell differentiation (2). Recently it has been suggested that phosphorylation and internalization of CD20 may be necessary steps for B cell entry into the G_1 phase of the cell cycle (4, 5). Here we report the isolation and expression of two CD20 cDNA clones using a recently described technique (6).

Materials and Methods

Preparation of cDNA Library and Recovery of cDNA Clones by Panning. Poly(A)$^+$ RNA was prepared from the human Burkitt cell line Daudi by oligo(dT) cellulose chromatography of total RNA isolated by procedures described previously (6). cDNA preparation and expression library construction were carried out as described (6, 7).

Anti-CD20 mAbs 1F5, 2H7, B1, L27, G28-2, 93-1B3, B-C1, and NU-B2 were obtained from the International Leukocyte Typing Workshop (4). Purified mAbs were used at a concentration of 1 μg/ml and ascites were used at a dilution of 1:1,000. Panning was done as described (6) except that in the first round of screening, eight 10-cm dishes of 50% confluent COS cells were transfected by the DEAE-Dextran method (7). Subsequent screening cycles were performed by spheroplast fusion as described (6).

Immunoprecipitation, Sequencing, RNA and DNA Blot Hybridization. The B cell line RAJI was metabolically labeled with [^{35}S]methionine and [^{35}S]cysteine (200 μCi/ml) in cysteine/methionine-free RPMI medium (Gibco Laboratories, Grand Island, NY) supplemented with antibiotics and 5% dialyzed FCS for 12 h at 37°C. COS cells transfected by the DEAE-Dextran method were similarly labeled 36 h posttransfection. The labeled cells were lysed with 1% NP-40, 400 mM NaCl, 10 mM EDTA, and 10 mM ATP in PBS (8). After removing debris and nuclei by centrifugation (13,000 *g*, 10 min), the lysate was precleared with protein A-agarose beads (Calbiochem-Behring Corp., La Jolla, CA) and with nonspecific isotype-matched antibody (mouse IgG2a, Coulter Immunology, Hialeah, FL) for 4 h. The resulting lysates were incubated with B1 mAb (Coulter Immunology), 50 μg/ml, at 4°C overnight. Fresh protein A-agarose beads were added for 1 h, washed five times with lysis buffer and eluted. The eluate was electrophoresed through 12.5% polyacrylamide gels.

DNA and RNA blot analysis and hybridization probe preparation were carried out as described previously (6, 7). Sequencing was done by the dideoxy method (6, 7).

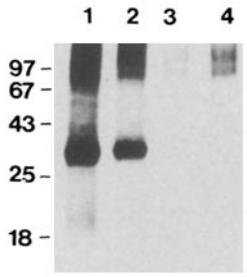

FIGURE 1. Immunoprecipitation of CD20 from (lane *1*) transfected COS cells; (lane *2*) Raji cell line; (lane *3*) Raji cells with nonspecific isotype-matched antibody; (lane *4*) transfected COS cells with nonspecific isotype matched antibody. Molecular mass markers are shown to the left.

B Cell Purification. Cells were gently teased from tonsils in sterile PBS. T cells were depleted by anti-CD3/complement lysis; cells were incubated with CD3 mAb in PBS at 4°C for 45 min. After two washes in PBS, the cells were incubated with rabbit complement (Pel-freez Biologicals, Rogers, AR) at 37°C for 45 min. Monocytes were removed by two 30-min rounds of panning on tissue culture dishes at 37°C. 80–85% of the recovered cells were $sIgM^+$.

Results and Discussion

Two cDNA clones, bearing inserts of 1.5 (CD20.4) and 1.0 kb (CD20.6), were isolated from a Daudi cell cDNA library by panning with a panel of mAbs against CD20. COS cells transfected with either clone reacted with all members of the panel of antibodies. Immunoprecipitation of the cDNA-encoded protein from transfected COS cells showed a single 33-kD band, slightly smaller than that obtained from Raji cells (Fig. 1).

The two cDNA inserts differed only in the 3′ untranslated region. The insert in clone CD20.6 has a short poly(A) tail and lacks a consensus polyadenylation signal, while the insert in CD20.4 lacks a poly(A) tail and extends 431 bp beyond the 3′ terminus in CD20.6 (Fig. 2 *A*). While this manuscript was in revision, Einfeld et al. (9) and Tedder et al. (10) have reported the isolation of cDNA clones from phage and plasmid libraries respectively.

RNA blot analysis showed that three transcripts of 3.8, 3.0, and 1.5 kb were present in B cells but absent from other cell types (Fig. 3), in agreement with the known pattern of antibody reactivity (1, 2, 4, 5). It appears likely that the CD20.6 clone is derived from the 1.5-kb transcript or possibly from an even shorter, undetectable species. Because the CD20.4 clone lacks a $poly(A)^+$ tail, its source cannot be inferred at present.

DNA blot analysis showed that the CD20 genomic sequences are not rearranged during development and are not amplified in the cell lines examined. A restriction fragment length polymorphism was observed in a DNA sample obtained from placenta (Fig. 4).

The amino acid sequence predicted by the cDNA contains 297 residues and has a molecular mass of 33,097 daltons. The sequence contains three major hydrophobic regions spanning residues 51–103, 117–141, and 183–203 (Fig. 1). Two other notable characteristics are the absence of an NH_2-terminal signal peptide and the presence of a highly charged COOH-terminal domain. A polyclonal anti-CD20 antibody that recognizes the last 18 residues of the COOH terminus reacts with lysates of cells expressing CD20 but not with intact cells (9), suggesting that the CD20 COOH terminus is located within the cytoplasm. Since there is no NH_2-terminal signal

A

```
   1 AAAGACAAACTGCACCCACTGAACTCCGCAGCTAGCATCCAAATCAGCCCTTGAGATTTGAGGCCTTGGAGACTCAGGAGTTTTGAGAGC

  91 AAAATGACAACACCCAGAAATTCAGTAAATGGGACTTTCCCGGCAGAGCCAATGAAAGGCCCTATTGCTATGCAATCTGGTCCAAAACCA
   1    MetThrThrProArgAsnSerValAsnGlyThrPheProAlaGluProMetLysGlyProIleAlaMetGlnSerGlyProLysPro
                               ---CHO---

 181 CTCTTCAGGAGGATGTCTTCACTGGTGGGCCCCACGCAAAGCTTCTTCATGAGGGAATCTAAGACTTTGGGGGCTGTCCAGATTATGAAT
  30 LeuPheArgArgMetSerSerLeuValGlyProThrGlnSerPhePheMetArgGluSerLysThrLeuGlyAlaValGlnIleMetAsn
                                                    ==========================

 271 GGGCTCTTCCACATTGCCCTGGGGGGTCTTCTGATGATCCCAGCAGGGATCTATGCACCCATCTGTGTGACTGTGTGGTACCCTCTCTGG
  60 GlyLeuPheHisIleAlaLeuGlyGlyLeuLeuMetIleProAlaGlyIleTyrAlaProIleCysValThrValTrpTyrProLeuTrp
     ==========================================================================================

 361 GGAGGCATTATGTATATTATTTCCGGATCACTCCTGGCAGCAACGGAGAAAAACTCCAGGAAGTGTTTGGTCAAAGGAAAAATGATAATG
  90 GlyGlyIleMetTyrIleIleSerGlySerLeuLeuAlaAlaThrGluLysAsnSerArgLysCysLeuValLysGlyLysMetIleMet
     ========================================                                  =========

 451 AATTCATTGAGCCTCTTTGCTGCCATTTCTGGAATGATTCTTTCAATCATGGACATACTTAATATTAAAATTTCCCATTTTTTAAAAATG
 120 AsnSerLeuSerLeuPheAlaAlaIleSerGlyMetIleLeuSerIleMetAspIleLeuAsnIleLysIleSerHisPheLeuLysMet
     ================================================================================

 541 GAGAGTCTGAATTTTATTAGAGCTCACACACCATATATTAACATATACAACTGTGAACCAGCTAATCCCTCTGAGAAAAACTCCCCATCT
 150 GluSerLeuAsnPheIleArgAlaHisThrProTyrIleAsnIleTyrAsnCysGluProAlaAsnProSerGluLysAsnSerProSer

 631 ACCCAATACTGTTACAGCATACAATCTCTGTTCTTGGGCATTTTGTCAGTGATGCTGATCTTTGCCTTCTTCCAGGAACTTGTAATAGCT
 180 ThrGlnTyrCysTyrSerIleGlnSerLeuPheLeuGlyIleLeuSerValMetLeuIlePheAlaPhePheGlnGluLeuValIleAla
             ============================================================================

 721 GGCATCGTTGAGAATGAATGGAAAAGAACGTGCTCCAGACCCAAATCTAACATAGTTCTCCTGTCAGCAGAAGAAAAAAAGAACAGACT
 210 GlyIleValGluAsnGluTrpLysArgThrCysSerArgProLysSerAsnIleValLeuLeuSerAlaGluGluLysLysGluGlnThr

 811 ATTGAAATAAAAGAAGAAGTGGTTGGGCTAACTGAAACATCTTCCCAACCAAAGAATGAAGAAGACATTGAAATTATTCCAATCCAAGAA
 240 IleGluIleLysGluGluValValGlyLeuThrGluThrSerSerGlnProLysAsnGluGluAspIleGluIleIleProIleGlnGlu

 901 GAGGAAGAAGAAGAAACAGAGACGAACTTTCCAGAACCTCCCCAAGATCAGGAATCCTCACCAATAGAAAATGACAGCTCTCCTTAAGTG
 270 GluGluGluGluGluThrGluThrAsnPheProGluProProGlnAspGlnGluSerSerProIleGluAsnAspSerSerProEnd 297
                                                                  ---CHO---

 991 ATTTCTTCTGTTTTCTGTTTCCTTTTTTAAACATTAGTGTTCATAGCTTCCAAGAGACATGCTGACTTTCATTTCTTGAGGTACTCTGCA
                                                  *

1081 CATACGCACCACATCTCTATCTGGCCTTTGCATGGAGTGACCATAGCTCCTTCTCTCTTACATTGAATGTAGAGAATGTAGCCATTGTAG

1171 CAGCTTGTGTTGTCACGCTTCTTCTTTTGAGCAACTTTCTTACACTGAAGAAAGGCAGAATGAGTGCTTCAGAATGTGATTTCCTACTAA

1261 CCTGTTCCTTGGATAGGCTTTTTAGTATAGTATTTTTTTTTGTCATTTTCTCCATCAGCAACCAGGGAGACTGCACCTGATGGAAAAGAT

1351 ATATGACTGCTTCATGACATTCCTAAACTATCTTTTTTTTATTCCACATCTACGTTTTTGGTGGAGTCCCTTTTTATCATCCTTAAAACA

1441 ATGATGCAAAAGGGCTTTAGAGCACAATGGATCT 1474
```

B

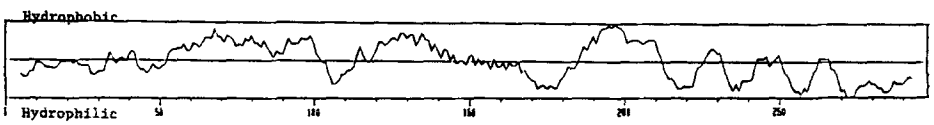

FIGURE 2. (*A*) Sequence of the CD20.4 cDNA. The sites of potential *N*-linked glycosylation are denoted by the symbol-CHO-; the hydrophobic regions are underscored. The site of the poly(A)$^+$ tail in clone CD20.6 is denoted by an asterisk. (*B*) Hydrophobicity profile of the amino acid sequence in A.

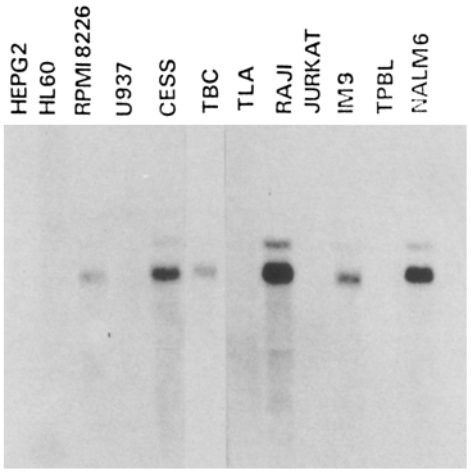

FIGURE 3. RNA blot analysis. 20 μg of total RNA were denatured, electrophoresed through agarose, transferred to nylon membranes, hybridized with CD20 cDNA, and exposed for 12 h. RNA sources included the pre-B leukemia Nalm-6, the B-LBL lines IM-9 and CESS, the BL Raji, the myeloma RPMI 8226, the T cell leukemia Jurkat, the myeloid lines HL60 and U937, tonsillar B cells (TBC), lymphokine-activated T cells (TLA) and peripheral blood T cells (TPBL).

peptide, it is likely that the NH_2 terminus is also intracellular and that the first hydrophobic region acts as an internal membrane insertion signal. The first hydrophobic region is composed of 53 residues and is therefore long enough to span the membrane twice if organized as an α helix. Because there are two remaining hydrophobic regions, the intracellular localization of the COOH terminus requires that the first hydrophobic domain exit the membrane on the intracellular side. Alternatively, the COOH-terminal antibody may only recognize epitopes exposed by detergent treatment allowing the COOH terminus to be extracellular and forcing the first hydrophobic domain to exit the membrane on the extracellular side. The sequence contains two potential *N*-glycosylation sites (Asn-Xaa-Ser/Thr) at position

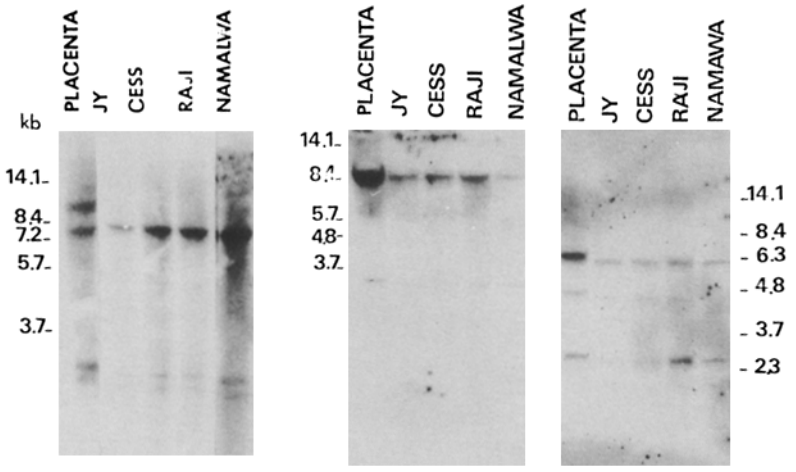

FIGURE 4. DNA blot analysis. DNA samples were digested with Eco RI (a), Hind III (b), and BamHI (c), electrophoresed through agarose, transferred to nylon, hybridized with CD20 cDNA, and exposed for 48 h. DNA sources included placenta, the B-LBL lines CESS and JY and the BL lines Raji and Namalwa. Molecular mass markers are shown for each blot.

9 and 293, but neither of these are expected to be used if they are located in intracellular domains of the molecule. If the COOH terminus is intracellular, the only extracellular domain would lie between residues 142 and 182. This region is rich in serine and threonine residues which might support *O*-glycosylation.

Immunoprecipitation of transfected COS cells with the anti-CD20 mAb B1 gave a single 33-kD form (Fig. 2), suggesting that CD20 is not glycosylated in these cells. In contrast, 35- and 37-kD species have been reported in certain B cell subsets and lines (4, 5). It is possible that proteolysis, *O*-linked glycosylation, or other posttranslational modifications may account for the difference in mass. An alternative explanation may be that alternate initiation sites are used in different B cell subpopulations (9). The CD20 cDNA contains two NH_2-terminal ATG codons separated by 16 residues. Neither ATG is embedded in the initiation consensus sequence proposed by Kozak (11). The use of the first or the second ATG would yield protein molecular masses of 33.1 and 30.8 kD, respectively. The immunoprecipitation result shown in Fig. 2 suggests that only one ATG is used in both COS and Raji cells. However, using a different immunoprecipitation protocol, we have detected a major 33-kD band and a minor 31-kD band in COS cells (data not shown). Further investigation will be necessary to clarify these results.

Comparison of the peptide sequence with the sequences in the National Biomedical Research Foundation database detected no significant homology to other proteins by the FASTP rapid sequence alignment algorithm. Because the bulk of the protein appears to be confined to the interior of the membrane and the cell, it seems plausible that it may play a role in transducing signals from other transmembrane proteins to the cell interior. Consistent with this role is the relatively hydrophilic nature of the hydrophobic regions, which might allow hydrogen bond interactions with the transmembrane portions of other proteins.

Summary

Two cDNA clones encoding the pan-B cell CD20 antigen were isolated from a COS cell expression library. The two clones bear identical coding sequences and differ only in the length of the 3′ untranslated region. The predicted CD20 sequence is 297 residues long and contains three hydrophobic domains, one of which is long enough to span the membrane twice. COS cells transfected with either CD20 clone express an immunoreactive protein of 33 kD.

We thank Jeff Ledbetter for sharing sequence and protein data, Mary Valentine for the immunoprecipitation protocol, Ed Clark for valuable discussions, David Simmons for gifts of RNA, and Sandro Aruffo and Andy Peterson for valuable advice and critical review of the manuscript.

Received for publication 13 October 1987 and in revised form 5 April 1988.

References

1. Clark, E. A., G. Shu, and J. A. Ledbetter. 1985. Role of Bp35 cell surface polypeptide in human B cell activation. *Proc. Natl. Acad. Sci. USA.* 82:1766.
2. Golay, J. T., E. A. Clark, and P. C. Beverley. 1985. The CD20 (Bp35) antigen is involved in activation of B cells from the G_0 to the G_1 phase. *J. Immunol.* 135:3795.
3. Tedder, T. F., A. W. Boyd, A. S. Freedman, L. M. Nadler, and S. F. Schlossman. 1985.

The B cell surface molecule B1 is functionally linked with B cell activation and differentiation. *J. Immunol.* 135:973.

4. Valentine, M. A., G. Shu, L. Gaur, J. A. Ledbetter, and E. A. Clark. 1987. Structure and function of the B cell specific 35-37kDa CD20 protein. *In* Leukocyte Typing III. A. McMichael, et al., editors. Oxford University Press, Oxford. 440.
5. Valentine, M. A., L. Gaur, R. Torres, and E. A. Clark. 1987. Expression of the human B cell surface protein CD20: alteration by phorbol myristate acetate. *Proc. Natl. Acad. Sci. USA.* 84:8085.
6. Seed, B., and A. Aruffo. 1987. Molecular cloning of the human CD2 antigen, the T cell erythrocyte receptor by a rapid immunoselection procedure. *Proc. Natl. Acad. Sci. USA.* 84:3365.
7. Aruffo, A., and B. Seed. 1987. Molecular cloning of a CD28 cDNA by a high efficiency COS cell expression system. *Proc. Natl. Acad. Sci. USA.* 84:8573.
8. Pletz, G. A., B. Gallis, and M. Peterlin. 1987. Monoclonal antibody immunoprecipitation of cell membrane glycoproteins. *Anal. Biochem.* 167:239.
9. Einfeld, D. A., J. P. Brown, M. A. Valentine, E. A. Clark, and J. A. Ledbetter. 1988. Molecular cloning of the human B cell CD20 receptor predicts a hydrophobic protein with multiple transmembrane domains. *EMBO (Eur. Mol. Biol. Organ.) J.* 7:711.
10. Tedder, T. F., M. Streuli, S. F. Schlossman, and H. Saito. 1988. Isolation and structure of a cDNA encoding the B1 (CD20) cell-surface antigen of human B lymphocytes. *Proc. Natl. Acad. Sci. USA.* 85:208.
11. Kozak, M. 1984.Compilation and analysis of sequences upstream from the translational start site in eukaryotic mRNAs. *Nucleic Acids Res.* 12:857.