Brief Definitive Report

CD19, THE EARLIEST DIFFERENTIATION ANTIGEN OF THE B CELL LINEAGE, BEARS THREE EXTRACELLULAR IMMUNOGLOBULIN-LIKE DOMAINS AND AN EPSTEIN-BARR VIRUS-RELATED CYTOPLASMIC TAIL

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

Arising from progenitor cells in the bone marrow, B cells undergo differentiation through a series of stages characterized by the orderly rearrangement and expression of Ig genes (1, 2). However, the very earliest stages of commitment to the B lineage are poorly understood. Before the appearance of cytoplasmic IgM, committed pre-B lymphocytes can be distinguished by the appearance of class II MHC antigens, followed by the B cell-specific 95-kD surface glycoprotein CD19 (3). All resting B cells display CD19 antigens, and CD19 expression persists upon activation (4), but is lost upon further differentiation to Ig-secreting plasma cells (4). Almost all early B cell malignancies show CD19 expression (3). The observation that CD19 undergoes antibody-induced internalization suggests that it may provide a suitable target for immunotoxin-mediated treatment of aggressive forms of B cell lymphomas and leukemias that respond poorly to conventional chemotherapy (5).

Although the physiological role of CD19 is not at present known, anti-CD19 mAbs block the activation of mature B cells elicited by anti-Ig antibodies (6), and one anti-CD19 mAb, B43, induces pre-B cells to proliferate directly (5). This report describes the isolation and expression of a cDNA clone encoding CD19. The extracellular portion of CD19 is organized in Ig-like domains, and the intracellular portion bears significant relatedness to an Epstein-Barr virus protein of unknown function. Lesser, but apparently significant homology was found with a segment of the EBV envelope protein and the *int-1* oncogene.

Materials and Methods

Monoclonal Antibodies. The B cell panel of mAb was obtained from the Third International Leukocyte Typing Workshop. Anti-CD19 mAb included HD37, B4, BU12, SJ25-C1, and HD2/37. mAb B43 was a kind gift from Dr. Fatih Uckun (University of Minnesota, Minneapolis, MN).

cDNA Library Construction and Cloning. A cDNA library was constructed from the Burkitt lymphoma cell line Daudi as described (7, 8), was introduced into COS cells by the DEAE-Dextran method (7, 8), and was enriched for CD19-encoding cDNAs by panning (7, 8). After three rounds of introduction into COS cells and panning, plasmid DNA was prepared from single colony isolates, transfected into COS cells, and scored for CD19 expression.

RNA Blots, DNA Blots, and Sequencing. RNA and DNA blot hybridizations and sequencing were performed as described (7, 8).

J. Exp. Med. © The Rockefeller University Press · 0022-1007/88/09/1205/06 $2.00
Volume 168 September 1988 1205-1210

Results and Discussion

To isolate a cDNA clone encoding CD19, an expression library was constructed from the Burkitt lymphoma line Daudi, introduced into COS cells by the DEAE-Dextran method, and subjected to three rounds of panning and reintroduction into *Escherichia coli* as described (7, 8). After the third round of panning, one of eight randomly picked colonies yielded DNA that when transfected into COS cells gave a positive indirect immunofluorescence reaction with anti-CD19 mAbs. COS cells transfected with the cDNA clone reacted with all anti-CD19 mAbs tested but with no other mAbs from a large B cell panel (data not shown). The positive plasmid contained an insert of ~1.9 kb.

RNA blot hybridization analysis revealed a single RNA species of 2 kb whose expression was restricted to B cell lines (Fig. 1). The highest abundance of the message was observed in the pre-B cell line Nalm-6, followed by the Burkitt lymphoma line Raji. Weaker expression was found in the B-lymphoblastoid lines IM-9 and CESS, while no expression was observed in the plasmacytoma line RPMI 8226, consistent with the observation that CD19 expression is lost upon terminal differentiation to plasma cells (3, 4). Peripheral T cells, the T cell leukemia line Jurkat, the promyelocytic leukemia line HL60, the promonocytic leukemia line U937, and the hepatoma line HepG2 were all negative for CD19 expression.

DNA blot hybridization analysis showed a simple pattern consistent with a single copy gene (data not shown).

The nucleotide sequence of the cDNA insert consists of 1,920 nucleotides terminating in a poly(A) tail 16 nucleotides downstream of the consensus polyadenylation signal (AATAAA) (Fig. 2 *a*). An open reading frame, starting at an ATG embedded in an initiation consensus sequence (9), encodes a protein of 467 amino acids with a predicted molecular weight (M_r) of 51,799. The first methionine is followed by 19 predominantly hydrophobic amino acids. Excision of these residues, at a site corresponding to the signal peptide cleavage rules predicted by von Heijne (10), yields a mature protein of M_r 49,300. The resulting extracellular domain would consist of 271 residues containing 5 potential *N*-linked glycosylation sites (Asn-Xaa-Ser/Thr). The predicted extracellular domain is followed by 28 predominantly hydrophobic amino acids, with the exception of one arginine residue, consistent with a trans-

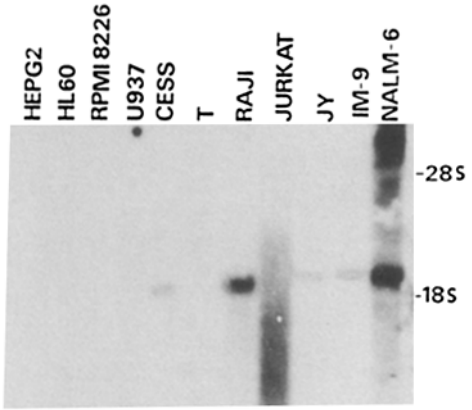

FIGURE 1. RNA blot hybridization analysis. 20 μg of each RNA sample was electrophoresed through a 1% agarose gel, denatured, transferred to a nylon filter, and hybridized to a ^{32}P-labeled CD19 cDNA probe. Sources of RNA are indicated. NALM-6 is a pre-B leukemia, IM-9 a B lymphoblastoid line, JY and CESS EBV-transformed B lymphoblastoid lines, Raji a Burkitt lymphoma, RPMI 8226 a plasmcytoma, Jurkat a T cell leukemia; T represents lymphokine-activated peripheral T cells, HL60 and U937 are myeloid leukemia cell lines and HepG2 a hepatoblastoma.

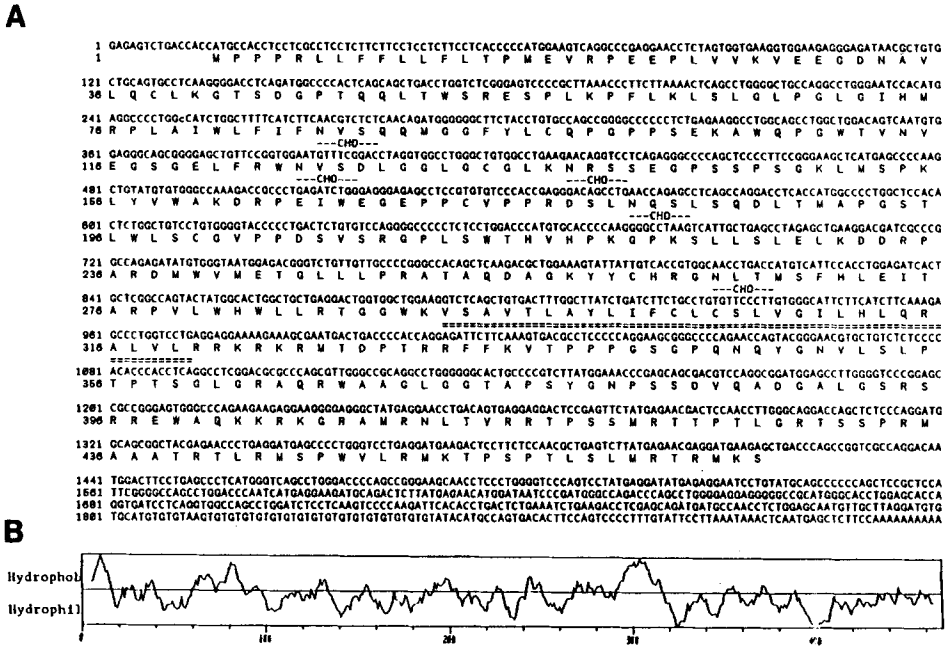

FIGURE 2. (*A*) Sequence of the CD19 cDNA. The sites of potential *N*-linked glycosylation are denoted by the symbol-CHO-; the predicted transmembrane region is underscored. (*B*) Hydropathicity profile of the amino acid sequence in *A*.

membrane region. Although uncommon, charged residues have been observed in the transmembrane domains of a few surface proteins, including the α and β chains of the T cell antigen receptor complex (11) and IgE Fc receptors (12). The predicted cytoplasmic domain is composed of 148 residues and contains a large number of serine and threonine residues (16 and 17, respectively) providing potential phosphorylation sites. The cytoplasmic domain is also arginine and proline rich (24 and 15 residues, respectively).

The predicted amino acid sequence was compared with the National Biomedical Research Foundation (NBRF) database using the FASTP algorithm. Significant relatedness was found with members of the Ig family, two proteins encoded by Epstein-Barr virus, and the transforming protein *int-1*. The Ig homology arises from the extracellular domain which is organized into three contiguous Ig-like domains (13). Domains 1 and 3 have intercysteine distances of 59 and 61 residues, respectively, only slightly shorter than the 62–67 residue intercysteine spacing typical of V region domains. The sequence surrounding the second cysteine in domains 1 and 3 matches the V region consensus D-X-G/A-X-Y-C. Domain 2 has an intercysteine distance of only 39 residues, and the sequence around the second cysteine more closely resembles that from C region domains (13).

Quantitative sequence comparisons were performed with the ALIGN program of the NBRF Protein Identification Resource (14). A score for the test sequences >3 SD above the mean for an ensemble of randomly permuted sequences is considered significant (13). Optimal alignment of CD19 domains 1 and 3 gave a score 4.61

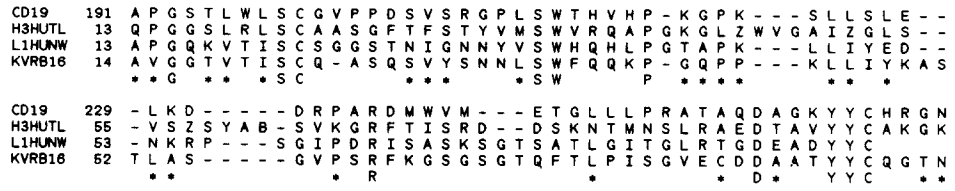

FIGURE 3. Optimal alignment of the third Ig-like domain of CD19 (residues 191–265) with human Ig heavy chain V III region, (NBRF code H3HUTL) human Ig λ chain V I region (L1HUNW), and rabbit κ chain V region (KVRB16) by the ALIGN program of the Protein Identification Resource (NBRF) (28). Asterisks denote conserved residues between CD19 and at least one V region. Residues conserved among all five proteins are denoted by the corresponding letter.

SD above the mean. When the two domains were compared with V regions in the NBRF database, domain 3 showed greater homology to various V regions than did domain 1 (Fig. 3). The greatest similarity was found with a rabbit V/κ region, for which a score 9.99 SD above the mean was found, corresponding to a probability of 8×10^{-24} for an equally good or better match occurring between two proteins of identical composition.

Surprisingly, significant homology was found between the cytoplasmic domain of CD19 and the EBV proteins BSLF-1 (15) and gp 220/350 (16), and the transforming protein *int-1* (17). BSLF-1 is a protein of unknown function, while gp 220/350 plays a role in the attachment of EBV to cell membrane (16). *int-1* is implicated in viral mammary tumorigenesis and is expressed in the developing murine nervous system (17). The ALIGN score for optimal alignment of CD19 and BSLF-1, gp 220/350 and *int-1* was 8.42, 4.91, and 7.60 SD above the mean, respectively (Fig. 4).

The functional significance of the similarities between CD19 and EBV proteins and *int-1* is at present unclear. However, the degree of relatedness to BSLF-1 appears to eliminate coincidence as a plausible explanation. The CD19-related sequences in BSLF-1 may have arisen adventitiously, for example as a result of viral capture

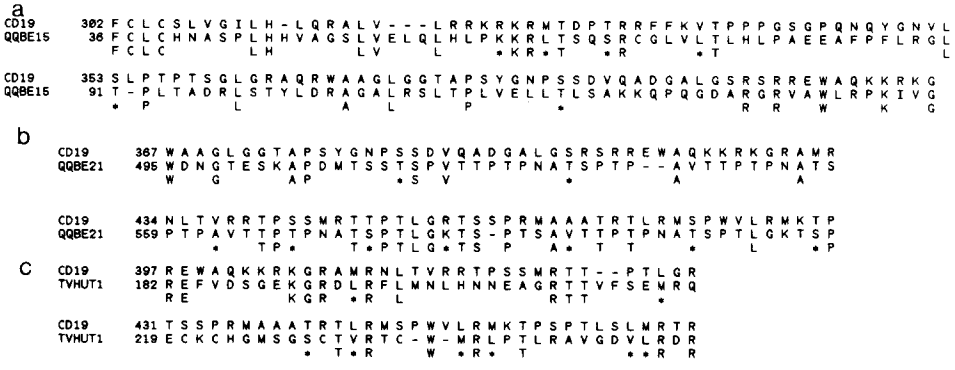

FIGURE 4. Optimal alignment between segments of the cytoplasmic domain of CD19 and the EBV protein BSLF1 (NBRF code QQBE15) (*a*), the EBV protein gp350/220 (NBRF code QQBE21) (*b*), and the *int-1* oncogene (NBRF code TVHUT1). Conserved residues are displayed beneath the aligned sequences; asterisks denote closely related residues.

of cellular sequences, or they may serve some viral role. In the latter case, either capture of cellular sequences or convergent evolution might account for the observed homology. It will be interesting to see if the limits of homology correspond to exon boundaries in the genomic DNA.

Summary

The isolation and expression of a full-length cDNA clone encoding the B cell–specific glycoprotein CD19 is reported. The sequence of the cDNA predicts a glycosylated integral membrane protein with a precursor molecular weight of 51.8×10^3 and an extracellular domain organized into three contiguous Ig-like sub-domains. The cytoplasmic domain bears significant relatedness to two proteins encoded by the Epstein-Barr virus and the *int-1* oncogene. CD19 transcripts are restricted to members of the B cell lineage, being most abundant in pre-B cell lines and least abundant in plasmacytomas.

Received for publication 16 May 1988 and in revised form 6 July 1988.

References

1. Huber, B. T. 1982. B cell differentiation antigens as probes for functional B cell subsets. *Immunol. Rev.* 64:57.
2. Coffman, R. L. 1982. Surface antigen expression and immunoglobulin gene rearrangement during mouse pre-B cell development. *Immunol. Rev.* 69:5.
3. Nadler, L. M., K. C. Anderson, G. Marti, M. P. Bates, E. Park, J. F. Daley, and S. F. Schlossman. 1983. B4, a human B lymphocyte-associated antigen expressed on normal, mitogen activated, and malignant B lymphocytes. *J. Immunol.* 131:244.
4. Freedman, A. S., A. W. Boyd, D. C. Fisher, S. F. Schlossman, and L. M. Nadler. 1986. Changes with in vitro activation of the B cell panel antigens. *In* Leukocyte Typing II. E. L. Reinherz, B. F. Haynes, L. M. Nadler, and I. D. Bernstein, editors. Springer-Verlag, Berlin. 443–454.
5. Uckun, F. M., W. Jaszcz, J. L. Ambrus, A. S. Fauci, K. Gajl-Peczalska, C. W. Song, M. R. Wick, D. E. Myers, K. Waddick, and J. A. Ledbetter. 1988. Detailed studies on expression and function of CD19 surface determinant by using B43 monoclonal antibody and the clinical potential of anti-CD19 immunotoxins. *Blood.* 7:13.
6. Pezzutto, A., G. Dorken, P. S. Rabinovitch, J. A. Ledbetter, G. Moldenhauer, and E. A. Clark. 1987. CD19 monoclonal antibody HD37 inhibits anti-immunoglobulin-induced B cell activation and proliferation. *J. Immunol.* 138:2793.
7. Seed, B., and A. Aruffo. 1987. Molecular cloning of the CD2 antigen, the T cell erythrocyte receptor, by a rapid immunoselection procedure. *Proc. Natl. Acad. Sci. USA.* 84:3365.
8. 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.
9. Kozak, M. 1984. Compilation and analysis upstream from the translational start site in eukaryotic mRNAs. *Nucleic Acids Res.* 12:857.
10. von Heijne, G. 1986. A new method for predicting sequence cleavage sites. *Nucleic Acids Res.* 14:4683.
11. Chien, Y-h, D. M. Becker, T. Lindsten, M. Okamura, D. I. Cohen, and M. M. Davis. 1984. A third type of murine T-cell receptor gene. *Nature (Lond.).* 312:31.
12. Kinet, J.-P., H. Metzger, J. Hakimi, and J. Kochan. 1987. A cDNA presumptively coding for the α subunit of the receptor with high affinity for immunoglobulin E. *Biochemistry.* 26:4605.

13. Williams, A. F. 1987. A year in the life of the immunoglobulin superfamily. *Immunol. Today.* 8:298.
14. Dayhoff, M. O., W. C. Barker, and L. T. Hunt. 1983. Establishing homologies in protein sequences. *Methods Enzymol.* 91:524.
15. Baer, R., A. T. Bankier, M. D. Biggin, P. L. Deininger, P. J. Farrell, T. J. Gibson, G. Hatfull, G. S. Hudson, S. C. Satchwell, C. Seguin, P. S. Tuffnell, and B. G. Barrell. 1984. DNA sequence and expression of the B95-8 Epstein-Barr virus genome. *Nature (Lond.).* 310:207.
16. Beisel, C., J. Tanner, T. Matsuo, D. Thorley-Lawson, F. Kezdy, and E. Kieff. 1985. Two major outer envelope glycoproteins of Epstein-Barr virus are encoded by the same gene. *J. Virol.* 54:665.
17. van Ooyen, A., V. Kwee, and R. Nusse. 1985. The nucleotide sequence of the human int-1 mammary oncogene; evolutionary conservation of coding and noncoding sequences. *EMBO (Eur. Mol. Biol. Organ.) J.* 4:2905.