The EMBO Journal vol.7 no.3 pp.711 - 717, 1988

Molecular cloning of the human B cell CD20 receptor predicts a hydrophobic protein with multiple transmembrane domains

D.A.Einfeld[1], J.P.Brown, M.A.Valentine[2], E.A.Clark[2] and J.A.Ledbetter

ONCOGEN, 3005 First Avenue, Seattle, WA 98121, and [2]Department of Microbiology and Immunology, University of Washington, Seattle, WA 98195, USA

[1]Present address: Cellular and Molecular Biology Program, University of Alabama, Birmingham, AL 35294, USA

Communicated by N.A.Mitchison

CD20 is an antigen expressed on normal and malignant human B cells that is thought to function as a receptor during B cell activation. Here we report the isolation of a CD20-specific cDNA clone from a λgt11 library using a polyclonal antiserum raised against purified CD20 antigen. Additional cDNA clones were then isolated from a λgt10 library. Alignment of the sequences of overlapping λ clones reveal a single consensus sequence except for a divergence that preceded the first methionine within the open reading frame. Normal B cells and B cell lines contain a prominent 2.6 kb mRNA and a lower level of a 3.3 kb mRNA. An oligonucleotide derived from one of the divergent sequences hybridized to the 3.3 kb mRNA only, indicating that the two mRNA species are derived from an alternative splicing mechanism. The predicted amino acid sequence of CD20 reveals three major hydrophobic regions of ~53, 25 and 20 amino acids. CD20 lacks an NH_2-terminal signal peptide and contains a highly charged COOH-terminal domain. Although CD20 is immunoprecipitated as a doublet of 33 and 35 kd proteins from B cells, *in vitro* translation of CD20 cDNA produced a single 33 kd protein that was specifically immunoprecipitated with monoclonal CD20 antibodies. CD20 was strongly phosphorylated on resting B cells after CDw40 stimulation, suggesting that CD20 may be functionally regulated by a protein kinase(s).

Key words: CD20/cell activation/membrane phosphoprotein/cDNA cloning

Introduction

The CD20 antigen is a hydrophobic phosphoprotein of ~35 kd that is expressed on mature B cells and almost all B cell lymphomas and chronic lymphocytic leukemias (Oettgen *et al.*, 1983; Stashenko *et al.*, 1980). Surface expression of CD20 is high on resting B cells (~1.5×10^5 molecules/cell) in peripheral blood or lymphoid organs, and is even higher (4-fold) on activated germinal center B cells (Ledbetter and Clark, 1986). A role for CD20 as a receptor in B cell activation has been proposed from studies showing stimulation of a G_0 to G_1 cell cycle transition of resting B cells by the 1F5 CD20 monoclonal antibody (mAb) (Clark *et al.*, 1985; Gollay *et al.*, 1985). 1F5 synergizes with the 12 kd B cell growth factor in driving cells through the cell cycle (Clark and Ledbetter, 1986; Gollay *et al.*, 1985). Other effects of 1F5 binding to resting B cells include induction of c-*myc* transcription (Smeland *et al.*, 1985) and increased expression of class II histocompatibility antigens (Clark and Shu, 1987). A number of the effects of 1F5 are similar to anti-immunoglobulin (Ig) immobilized on beads, but the effects are additive (Clark *et al.*, 1985) and the greatest response occurs when anti-Ig is added after pre-incubation of B cells with 1F5 (Clark and Shu, 1987). While the mechanism of activation of lymphocytes by anti-CD20 mAb is not known, CD20-mediated activation does not involve increased intracellular Ca^{2+} concentration (Smeland *et al.*, 1987; Rabinovitch *et al.*, 1987).

Another anti-CD20 mAb, B1, specifically blocks B cell proliferation when cells are exposed to various mitogenic stimuli or infected with EBV (Tedder *et al.*, 1985). Unlike the effects of 1F5, which seem to occur early in activation and involve the G_0 to G_1 transition, the inhibitory effects of B1 appear to occur at a later stage (Tedder *et al.*, 1986). Both 1F5 and B1, however, inhibited *in vitro* differentiation of B cells to immunoglobulin secretion in response to pokeweed mitogen (Gollay *et al.*, 1985).

CD20 is phosphorylated at serines, primarily, and at threonines, but not at tyrosines (Oettgen *et al.*, 1983). Increased phosphorylation of CD20 on dense tonsillar B cells occurs after treatment with phorbol myristate acetate (Valentine and Clark, 1988) and on *in vivo* activated (buoyant tonsillar) B cells by stimulation with B cell growth factor (Valentine and Clark, 1988).

As a basis for better understanding how the CD20 molecule functions, we sought to isolate CD20-specific cDNA. We report here the successful cloning and sequencing of the entire open reading frame for CD20. The predicted amino acid sequence shows that CD20 contains multiple hydrophobic domains and probably spans the membrane four times. CD20 shows no major sequence homology to other known proteins.

Results

Screening 4.5×10^5 recombinant plaques in a λgt11 library of Daudi cDNA with the polyclonal antiserum led to the isolation of one clone, gt11.13a, which hybridized specifically on Northern blots to mRNA present in Daudi cells (B cell), but not in Jurkat (T cell) or LMCF-7 (breast carcinoma) cells (data not shown). Clone gt11.13a is 337 bp in length. Its sequence contains multiple stop codons in all reading frames of either orientation. The orientation of the clone relative to the putative CD20 mRNA was determined by probing Northern blots with two oligonucleotides complementary to the opposite strands of gt11.13a. Only one of these oligonucleotides hybridized to Daudi mRNA; the message was the same size as that detected with the cDNA clone.

©IRL Press Limited, Oxford, England

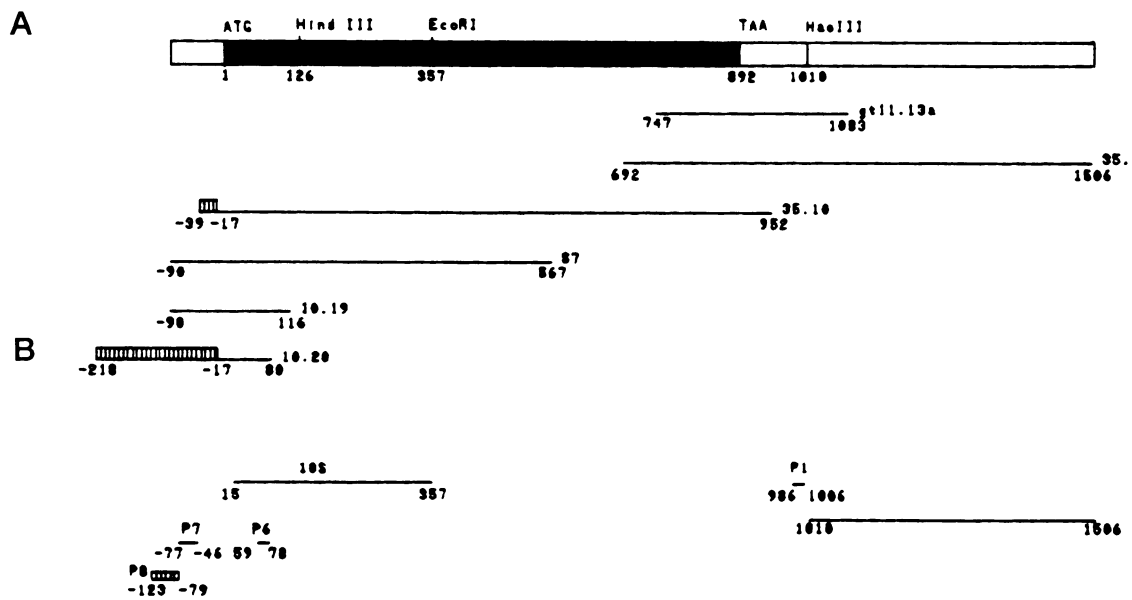

Fig. 1. (**A**) cDNA clones (—) aligned on the basis of overlapping sequences. Nucleotides are numbered with the A of the first ATG in the open reading frame being position 1. The coding (■) and untranslated (□) regions of the cDNA are illustrated at the top. The 5′ divergent sequences present in clones 35.10 and 10.20 (▨) are indicated. (**B**) Oligonucleotides and DNA fragments used as probes on Northern blots. P1, P6, P7 and P8 are synthetic oligonucleotides. P8 recognizes part of the alternate sequence shown in Part A (▨). 10S was subcloned from 35.10 as an *Eco*RI−*Eco*RI fragment though the 5′ end is not a true *Eco*RI site, having the sequence GAAATTC. The *Hae*III−*Eco*RI fragment extends from the *Hae*III site at position 1010 to the *Eco*RI site introduced on the end of the cDNA molecule by the synthetic linker described in Materials and methods.

```
      AACAAACTGCACCCACTGAACTCCGCAGCTAGCATCCAAATCAGCCCTTGAGATTTGAGGCCTTGGAGACTCAGGAGTTTTGAGAGCAAAATGACAACACCCAGAAATTCAGTAAATGGG   30
                                                                                     MetThrThrProArgAsnSerValAsnGly

      ACTTTCCTGGCAGAGCCAATGAAAGGCCCTATTGCTATGCAATCTGGTCCAAAACCACTCTTCAGGAGGATGTCTTCACTGGTGGGCCCCACGCAAAGCTTCTTCATGAGGGAATCTAAG  150
11    ThrPheLeuAlaGluProMetLysGlyProIleAlaMetGlnSerGlyProLysProLeuPheArgArgMetSerSerLeuValGlyProThrGlnSerPhePheMetArgGluSerLys

      ACTTTGGGGGCTGTCCAGATTATGAATGGGCTCTTCCACATTGCCCTGGGGGGTCTTCTGATGATCCCAGCAGGGATCTATGCACCCATCTGTGTGACTGTGTGGTACCCTCTCTGGGGA  270
51    ThrLeuGlyAlaValGlnIleMetAsnGlyLeuPheHisIleAlaLeuGlyGlyLeuLeuMetIleProAlaGlyIleTyrAlaProIleCysValThrValTrpTyrProLeuTrpGly

      GGCATTATGTATATTATTTCCGGATCACTCCTGGCAGCAACGGAGAAAAACTCCAGGAAGTGTTTGGTCAAAGGAAAAATGATAATGAATTCATTGAGCCTCTTTGCTGCCATTTCTGGA  390
91    GlyIleMetTyrIleIleSerGlySerLeuLeuAlaAlaThrGluLysAsnSerArgLysCysLeuValLysGlyLysMetIleMetAsnSerLeuSerLeuPheAlaAlaIleSerGly

      ATGATTCTTTCAATCATGGACATACTTAATATTAAAATTTCCCATTTTTTAAAAATGGAGAGTCTGAATTTTATTAGAGCTCACACACCATATATTAACATATACAACTGTGAACCAGCT  510
131   MetIleLeuSerIleMetAspIleLeuAsnIleLysIleSerHisPheLeuLysMetGluSerLeuAsnPheIleArgAlaHisThrProTyrIleAsnIleTyrAsnCysGluProAla

      AATCCCTCTGAGAAAAACTCCCCATCTACCCAATACTGTTACAGCATACAATCTCTGTTCTTGGGCATTTTGTCAGTGATGCTGATCTTTGCCTTCTTCCAGGAACTTGTAATAGCTGGC  630
171   AsnProSerGluLysAsnSerProSerThrGlnTyrCysTyrSerIleGlnSerLeuPheLeuGlyIleLeuSerValMetLeuIlePheAlaPhePheGlnGluLeuValIleAlaGly

      ATCGTTGAGAATGAATGGAAAAGAACGTGCTCCAGACCCAAATCTAACATAGTTCTCCTGTCAGCAGAAGAAAAAAAAGAACAGACTATTGAAATAAAAGAAGAAGTGGTTGGGCTAACT  750
211   IleValGluAsnGluTrpLysArgThrCysSerArgProLysSerAsnIleValLeuLeuSerAlaGluGluLysLysGluGlnThrIleGluIleLysGluGluValValGlyLeuThr

      GAAACATCTTCCCAACCAAAGAATGAAGAAGACATTGAAATTATTCCAATCCAAGAAGAGGAAGAAGAAGAAACAGAGACGAACTTTCCAGAACCTCCCCAAGATCAGGAATCCTCACCA  870
251   GluThrSerSerGlnProLysAsnGluGluAspIleGluIleIleProIleGlnGluGluGluGluGluGluThrGluThrAsnPheProGluProProGlnAspGlnGluSerSerPro

      ATAGAAAATGACAGCTCTCCTTAAGTGATTTCTTCTGTTTTCTGTTTCCTTTTTTAAACATTAGTGTTCATAGCTTCCAAGAGACATGCTGACTTTCATTTCTTGAGGTACTCTGCACAT  990
291   IleGluAsnAspSerSerPro

      ACGCACCACATCTCTATCTGGCCTTTGCATGGAGTGACCATAGCTCCTTCTCTCTTACATTGAATGTAGAGAATGTAGCCATTGTAGCAGCTTGTGTTGTCACGCTTCTTCTTTTGAGCA 1110

      ACTTTCTTACACTGAAGAAAGGCAGAATGAGTGCTTCAGAATGTGATTTCCTACTAACCTGTTCCTTGGATAGGCTTTTTAGTATAGTATTTTTTTTTGTCATTTTCTCCATCAGCAACC 1230

      AGGGAGACTGCACCTGATGGAAAAGATATATGACTGCTTCATGACATTCCTAAACTATCTTTTTTTTATTCCACATCTACGTTTTTGGTGGAGTCCCTTTTGCATCATTGTTTTAAGGAT 1350

      GATAAAAAAAAAATAACAACTAGGGACAATACAGAACCCATTCCATTTATCTTTCTACAGGGCTGACATTGTGGCACATTCTTAGAGTTACCACACCCCATGAGGGAAGCTCTAAATAGC 1470

      CAACACCCATCTGTTTTTTGTAAAAACAGCATAGCTT                                                                                    1507


      ACTGACAATAAGCAATTAAATAAATAAGAACTCAGCAGTAGGCCTTGCCTCAGATCCAAGGTCACTCGGAAGAGGCCATGTCTACCCTCAATGACACTCATGGAGGAAATGCTGAGAGAA  -98

      GCATTCAGATGCATGACACAAGGTAAGACTGCCAAAAATCTTGTTCTTGCTCTCCTCATTTTGTTATTTGTTTTATTTTTAGGAGTTTTGAGAGCAAAATGACAACACCCAGAAATTCAG   22
                                                                                            MetThrThrProArgAsnSer
```

Fig. 2. CD20 cDNA and predicted protein sequences. (**A**) S7 consensus sequence. Numbers at the right and left margins indicate the positions of the last nucleotide and first amino acid of each row, respectively. The numbering system assigns the A of the initiating ATG as nucleotide 1 and the met encoded by this codon as the first amino acid. The regions labelled A, B, and C refer to the hydrophobic domains (refer to Figure 6). Shown at the bottom is the sequence of the 5′ end of clone 10.20.

Additional cDNA clones were obtained by screening λgt10 libraries using clone gt11.13a or fragments of subsequent clones (Figure 1A). Clones 35.7 and 35.10 were obtained by screening a λgt10 library with clone gt11.13a. Use of a 342 bp fragment of 35.10, designated 10S, led to isolation of clone S7 from the same library. Clones 10.19 and 10.20 came from a λgt10 library of cDNA primed with oligonucleotide P6 (Figure 1B). Four additional clones from this library (not shown) have been sequenced and are identical to 10.19 at the 5′ end, varying only in the location of their 3′ termini.

Alignment of the sequences of the overlapping λ clones yields a single consensus sequence except near the 5′ terminus of the cDNA. The sequences of clones S7, 10.19 and the four additional clones form one consensus sequence for the 5′ end of the cDNA. Interestingly, all six clones terminate at the same base. Clone 10.20 diverges from the S7 consensus sequence beginning at the 72nd base pair from the 5′ end of S7 and moving in the 5′ direction. The sequence of 35.10 matches that of 10.20, though it does not extend as far as 5′. The consensus sequence based on the S7 5′ sequence exhibits an open reading frame of 930 bp (Figure 2A). Using the 10.20, 5′ sequence gives a 963 bp open reading frame which terminates at the same 3′ position (Figure 2B). The point of divergence between the two sequences, though lying within the open reading frames, precedes the first methionine.

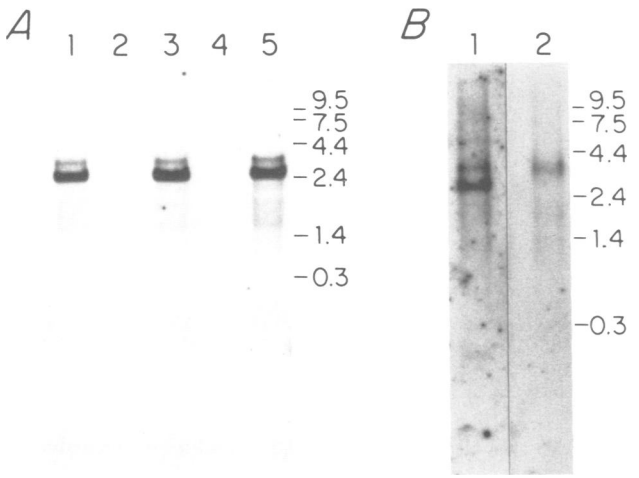

Fig. 3. (**A**) Northern blot of poly(A)$^+$ RNA from DHL-10 (B lymphoma), **lane 1**; HSB-2 (T-lineage ALL), **lane 2**; 616 (EBV-transformed B cell), **lane 3**; HPB-ALL (T-lineage ALL), **lane 4**; and T-51 (EBV-transformed B cell), **lane 5**, probed with nick-translated 10S. The blot was washed in 1 × SSC/0.1% SDS at 42°C followed by 0.1 × SSC/0.1% SDS at room temperature and autoradiographed for 24 h. The position of the RNA ladder (BRL) is shown by arrows. (**B**) Northern blot of poly(A)$^+$ RNA from the buoyant fraction of human tonsillar B cells (Ledbetter and Clark, 1986) probed with kinased oligonucleotides P7 (**lane 1**) and P8 (**lane 2**). Blots were washed in 2 × SSC at 37°C and autoradiographed for 40 h.

Northern blot analysis of a number of lymphocyte cell lines shows that expression of the CD20 antigen correlates with the presence of mRNA detected with our cDNA clones. The Burkitt's lymphoma lines Ramos and Daudi, the EBV-transformed lines T-51 and 616, and the B cell lymphoma derived DHL-10, all of which bind anti-CD20 mAbs, react with a CD20 cDNA probe in Northern blots. The probe hybridizes to two major mRNA species of 2.6 and 3.3 kb in these cell lines (Figure 3A), whereas neither mRNA species is detected in the IgG myeloma 8226 and the T-lineage ALL lines CEM, 8402, HPB-ALL, HSB-2, Jurkat and Molt-4, all of which do not express CD20.

The presence of two mRNA species and the occurrence of two 5′ terminal cDNA sequences led us to probe Northern blots with oligonucleotides specific for the divergent sequences. Oligonucleotide P7 is complementary to bases −77 through −46 of the S7 sequence and oligonucleotide P8 is complementary to bases −123 through −79 of the 10.20 sequence (Figure 1B). P7 binds to both the 2.6 and 3.3 kb mRNA species whereas P8 binds only to the 3.3 kb message (Figure 3B). Northern blots probed with gt11.13a, with a DNA fragment extending from the *Hae*III site at 1010 through the end of clone 35.7, with DNA fragment 10S (bp 16−357), as well as Northern blots probed with oligonucleotides P6 and P1 (Figure 1B) all show both the 2.6 and 3.3 kb messages. Thus these two mRNA species are very closely homologous, if not identical, over a span of 1500 bp. The only difference we have detected is in their ability to bind oligonucleotide P8. One explanation of this result is that the P8 sequence is part of an intron which has been spliced out of the 2.6 kb molecule. The binding pattern of P7 would be explained by the location of its complementary sequence 5′ of the deleted region. Removal of the P8 sequence by RNA splicing would result in the P7 sequence being moved within closer proximity of the initiation methionine. Consistent with this hypothesis, the sequence at the point of divergence between S7 and 10.20 is homologous to consensus sequences for RNA splice sites (Kozak, 1984; Padgett *et al.*, 1986). The sequence of 10.20 matches the consensus sequence for the 3′ end of an intron at 11 of 14 bases (Figure

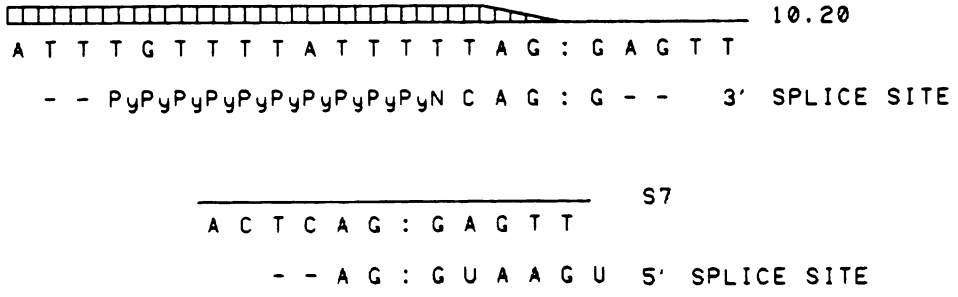

Fig. 4. Sequences of clones 10.20 and S7 in the regions of their divergence are aligned above consensus sequences for 5′ and 3′ RNA splice sites, respectively (Padgett *et al.*, 1986). The consensus sequences are for unspliced RNA. Y, pyrimidine (C or T); N, any base.

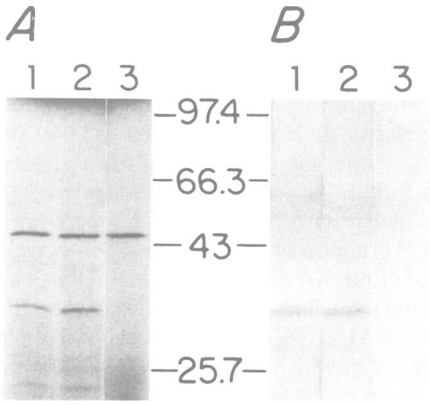

Fig. 5. *In vitro* translation of CD20. (**A**) 5 μl of *in vitro* translation product from reticulocyte lysate incubated alone (**lane 3**) or with RNA made from the S7 cDNA construct (**lane 1**), or with RNA made from the 35.10 cDNA construct (**lane 2**) were analyzed by SDS−PAGE (12%). The gel was dried and autoradiographed overnight. The positions of mol. wt markers (in kds) are indicated. (**B**) Immunoprecipitations were performed as described in the methods section using the CD20 mAbs 2H7 (**lane 1**), or 1F5 (**lane 2**), or the anti-CD28 mAb 9.3 (**lane 3**). Samples were analyzed on a 12% gel that was then fixed with 20% methanol: 8% acetic acid, rinsed in water for 30 min, soaked in 1.6% sodium salicylate for 30 min, dried and autoradiographed for 8 days.

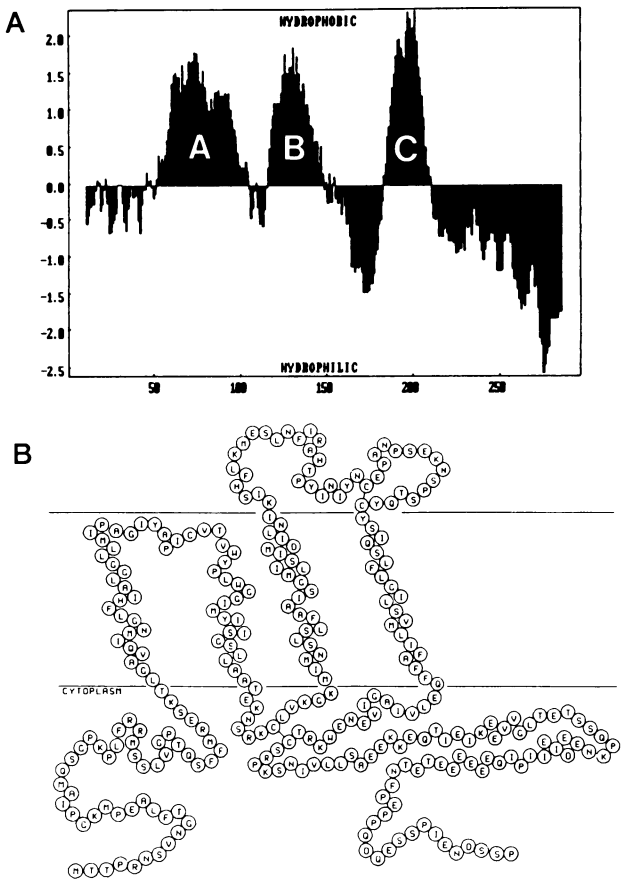

Fig. 6. (**A**) Hydropathy plot of the CD20 protein. The algorithm (Kyte and Doolittle, 1982) was applied using a window of 20 residues. (**B**) Proposed model of CD20 orientation in the cell membrane. Amino acid residues are indicated using the one letter code.

4). The S7 sequence contains the consensus AG dinucleotide at what would be the 5′ side of a 5′ splicing site.

As noted above, the two cDNA sequences differ only 5′ of the first in-frame ATG codon of the open reading frame and so predict the same protein product. Initiation at the first ATG, which is preceded by an A at the −3 position, would yield a protein of 297 amino acids and a mol. wt of 33 093 daltons. SDS−PAGE analysis of the CD20 antigen shows a doublet of 33 and 35 kd (Oettgen *et al.*, 1983; Figure 7). To clarify the relationship between our cDNA clones and the CD20 doublet we utilized CD20 cDNA to produce RNA for *in vitro* translation. Two constructs were made by inserting into *Eco*RI−*Hin*cII cut SP65 plasmid a 1.5 kb *Hin*dIII−*Hin*cII fragment of clone 35.10 (extending from the *Hin*dIII site at 126 to a *Hin*cII site in λgt10 ~650 bp beyond the 3′ end of the 35.10 insert) together with an *Eco*RI−*Hin*dIII fragment from either 35.10 or S7 (which extends from the 5′ *Eco*R1 end to the *Hin*dIII site at 126). *In vitro* translation of RNA produced from each of these constructs yielded a major band at 33 kd (Figure 5A). This band was not present in the translation performed in the absence of foreign RNA. In addition this band was precipitated specifically by the anti-CD20 mAbs 2H7 and 1F5 (Figure 5B). CD20 produced by *in vitro* translation of our cDNA occurs as a single species rather than a doublet.

The predicted amino acid sequence of CD20 shows no significant homology to sequences in the Genbank50 or PIR12 databanks. Analysis of the hydropathy of CD20 using the algorithm of Kyte and Doolittle (Kyte and Doolittle, 1982) reveals three major hydrophobic stretches of ~53, 25 and 20 amino acids (A, B and C, Figure 6A). The lengths of hydrophobic regions B and C are consistent with that of transmembrane spanning segments. Hydrophobic stretch A is large enough to cross the membrane twice. Other major features of the primary structure of CD20 are the absence of an NH_2-terminal signal peptide and the presence of a highly charged C-terminal domain. The 85 residues at the C-terminus comprise 25 negatively charged residues, 22 glutamic and three aspartic acids, and eight positively charged residues, six lysines and two arginines. A proposed model of the orientation of CD20 in the cell membrane is shown in Figure 6B.

It has previously been shown that anti-CD20 and anti-CDw40 mAbs in combination induce activation and proliferation of purified B cells without requiring any additional signals (Clark and Ledbetter, 1986). The effect of the anti-CDw40 mAb G28-5 on the phosphorylation of CD20 was examined. When anti-CDw40 mAb is added to tonsillar B cells during metabolic labelling with ^{32}P, phosphorylation of CD20 was increased at least 11-fold (Figure 7). FACS analysis of surface CD20 on cells incubated under these conditions showed an immediate and sustained drop in both the percentage of positive cells and the mode of fluorescence intensity (data not shown), indicating that the remaining surface molecules were hyperphosphorylated.

Discussion

We report here the isolation of a cDNA clone for the human B cell CD20 antigen from a λgt11 library using a polyclonal anti-CD20 antiserum and use of this clone to obtain additional cDNA clones from a λgt10 library. The predicted amino acid sequence contains 297 residues and has a mol. wt of 33 097. There are three hydrophobic regions located at residues 51−102, 117−141 and 183−203. The hydrophobic nature of CD20 has been noted previously (Oettgen

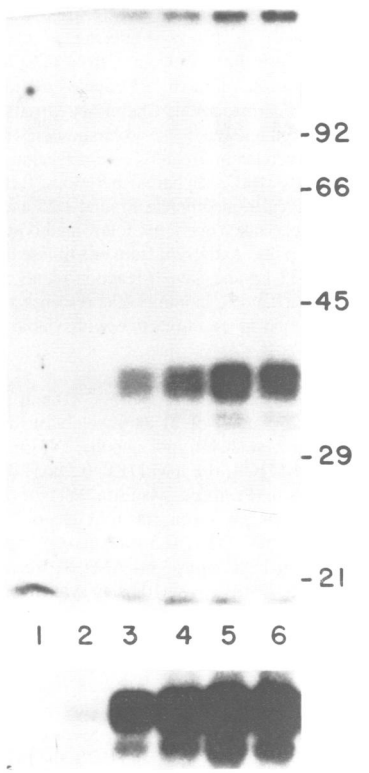

Fig. 7. Fresh tonsillar B cells were metabolically labelled for 5 h with ^{32}P as described (Oettgen *et al.*, 1983). Cells (10^7) were then incubated with 2 μg of anti-CDw40 (G28-5) for the last 10 min, 1, 2, and 3 h of labelling (**lanes 3–6**, respectively) or with medium alone (**lane 2**). The cells were then immunoprecipitated as described (Valentine *et al.*, 1988). mAbs 1F5 (**lanes 2–6**) or MOPC (**lane 1**) were added to cells prior to lysis. A longer exposure is included to show the medium control (bottom panel).

et al., 1983). We found no significant sequence homology of CD20 with other known proteins. CD20 cDNA has been independently isolated using an expression system in COS cells (I.Stamenkovic and B.Seed, personal communication). Comparison of sequences between our laboratories shows complete agreement in the open reading frame.

We have proposed a model of the configuration of CD20 (Figure 6B) in which hydrophobic domain A spans the membrane twice, with the protein bending back on itself in the region of Pro^{73}, Gly^{75} and Pro^{79} which lie near the middle of this domain. An additional proline occurs at position 87. There are no charged residues in the region bounded by these prolines, so we propose that the protein does not exit from the membrane in this region. Hydrophobic regions B and C are shown as transmembrane segments.

As to the polarity of the protein, the C-terminus is placed in the cytoplasm because the polyclonal anti-CD20 antiserum which reacted with the λgt11 clone containing the 28 C-terminal amino acid residues and could immunoprecipitate CD20, did not bind to intact cells expressing CD20 (data not shown). The NH_2-terminus has also been placed within the cell assuming that hydrophobic region A acts as an internal membrane insertion signal in the absence of an NH_2-terminal signal peptide. The implication is that the only part of the protein exposed on the surface of the cell is a 41 amino acid stretch (142–182) lying between hydrophobic regions B and C. It will be possible to test this aspect of the model using antibodies made against peptides constructed from the amino acid sequence of CD20. All anti-CD20 mAbs cross-block one another (Clark and Einfeld, 1986) indicating either a limited immunogenicity or a limited exposure on the cell surface.

The third aspect of the proposed model involves the disulfide bonds. The CD20 sequence contains five cysteine residues. There is no evidence from SDS–PAGE analysis that CD20 forms intermolecular disulfide bonds with itself or any other molecule. Three of the cysteines occur within separate hydrophilic regions of the protein. Two of these, Cys^{111} and Cys^{220}, are both predicted to lie in the cytoplasm and could form one intramolecular bond. Cys^{167}, which is in the putative extracellular domain, is shown bonded to Cys^{183}, which lies near the end of hydrophobic region C. Alternatively Cys^{167} may form a disulfide bond to Cys^{81}. In either case the unlinked cysteine may be sequestered in the membrane. The presence of intra-chain disulfide bonds has been suggested previously from differences of CD20 migration in SDS–PAGE under reducing and non-reducing conditions (Valentine and Clark, 1988).

Other membrane proteins which have been cloned and shown to have multiple membrane spanning regions are primarily of two families. Rhodopsin (Nathans and Hogness, 1983), β-2-adrenergic receptor (Dixon *et al.*, 1986), and the muscarinic acetylcholine receptor (Kubo *et al.*, 1986) transit signals through GTP-binding (G) proteins. Each of these glycoproteins, with apparent mol. wts around 70 kd, has seven hydrophobic stretches. These consist of 20–24 uncharged and mostly nonpolar amino acids and the proteins are thought to span the membrane seven times. They lack NH_2-terminal signal sequences and appear to have glycosylated extracellular NH_2-termini. A second more heterogeneous group, consisting of proteins which act as ion channels, includes the $(Ca^{2+} + Mg^{2+})$ATPase (MacLennan *et al.*, 1985), the catalytic subunit of the $(Na^+ + K^+)$ATPase (Sanger *et al.*, 1980), sodium channels (Noda *et al.*, 1986), erythrocyte band 3 (Kopito and Lodish, 1985), and the nicotinic acetylcholine receptor (Ballivet *et al.*, 1982; Devillers-Thery *et al.*, 1983; Furutani *et al.*, 1983; Noda *et al.*, 1983). Band 3 (100 kd) has 10 hydrophobic regions and is thought to span the membrane 12 times. The hydrophobic regions of Ca^{2+}- and Na^+-ATPases (100–110 kd) have been proposed to cross the membrane eight and seven to nine times, respectively via their hydrophobic regions. Though larger in size (250 kd), sodium channel proteins, with four internally repeated units, and the acetylcholine receptor, with five subunits, are composites of simpler transmembrane structures. Each repeat in sodium channels is thought to cross the membrane six times while the subunits of the acetylcholine receptor are each thought to cross four to six times. One feature of these proteins that has been highlighted is the presence of amphipathic helices which could be involved in forming an ion channel (Finer-Moore and Stroud, 1984). Another feature is the predominance of negatively charged amino acids which would direct cation flow. Hydrophobic region B of CD20 contains a single charged residue, Asp^{137}, and six polar residues. The distribution of these amino acids could give this transmembrane helix amphipathic character. Region A lacks amphipathicity except for a very short stretch containing Gln^{56}, Asn^{59}, and His^{63}. Region C appears very hydrophobic having only four uncharged polar amino acids, three of which are clustered together. CD20 also has a predominantly negatively charged carboxy terminal domain with 25 acidic

residues versus eight basic residues. Thus the CD20 molecule appears to have some similarity to ion channels.

Other proteins with multiple membrane spanning regions are the glucose transporter (Mueckler *et al.*, 1985), the lactose carrier protein (Foster *et al.*, 1983) and the Epstein–Barr virus (EBV)-encoded membrane protein that is expressed in latently infected cells (Feenewald *et al.*, 1984; Liebowitz *et al.*, 1986). The transport molecules lactose carrier and glucose transporter both appear to cross the membrane 12 times. The EBV protein is more similar to CD20. The amino terminus is followed by six hydrophobic regions which could each span the membrane and the C-terminal 200 amino acids constitute an acidic hydrophilic domain containing 37 aspartic acids and six glutamic acids but only two arginines and one lysine. Both the NH_2- and C-terminal domains appear to be located in the cytoplasm giving very little extracellular exposure to this protein which is capable of transforming established cell lines (Wang *et al.*, 1985).

The expression of CD20 appears to be transcriptionally controlled since neither the 2.6 nor 3.3 kb mRNAs are detected in non-expressing cells. The presence of two mRNA species may be the result of RNA splicing, since the only difference we have detected between the mRNAs correlates with the divergence of the cDNA sequences at a consensus RNA splice site. RNA was isolated from whole cells but both mRNA species are present in the poly(A)$^+$ fraction. The role of the two mRNA species in CD20 expression and its regulation remains to be determined.

Immunoprecipitation of CD20 from B cells using mAbs gives a 33/35 kd doublet. The relationship between the two molecules of the CD20 doublet is not clear. While the ATG encoding Met^{23} is preceded by a G at the −3 position, showing some homology to initiation consensus sequences, and would initiate a protein 2.3 kd smaller than a protein starting with Met^1, the ATG encoding Met^1 should be favored since it lies closer to the 5′ end. *In vitro* translation yielded a single protein migrating at 33 kd using RNA derived from cDNA constructs containing either the S7 on the 35.10 divergent sequences. This suggests that the doublet does not arise by alternative initiation. However, there is no evidence that CD20 is glycosylated, and dephosphorylation of CD20 using alkaline phosphatase does not cause an interconversion of the two species (Valentine *et al.*, unpublished data). Thus future studies will be required to determine the relationship between the 33 and the 35 kd bands.

CDw40 mAb G28-5 delivers a signal to B cells that cooperates with 1F5 CD20 stimulation to drive cell division (Clark and Ledbetter, 1986; Ledbetter *et al.*, 1987). Because of the functional interaction between these receptors, we examined whether stimulation with anti-CDw40 would affect phosphorylation of CD20. There was an 11-fold increase in CD20 phosphorylation within the first three hours of CDw40 stimulation (Figure 7), suggesting that CD20 is functionally regulated by phosphorylation. This is also supported by observations that CD20 phosphorylation is also increased by phorbol-12-myristate-13-acetate (an activator of protein kinase C) and by stimulation with anti-Ig or 12 kd B cell growth factor (Valentine and Clark, 1988).

Materials and methods

Antibodies

Monoclonal CD20-specific antibodies 1F5 and 2H7 have been previously described (Clark *et al.*, 1985; Clark and Shu, 1987; Ledbetter and Clark, 1986). Anti-CDw40 mAb G28-5 and anti-CD28 mAb 9.3 were produced and purified as previously described (Ledbetter *et al.*, 1985, 1987). Polyclonal anti-CD20 antiserum was obtained from mice immunized with the CD20 antigen immobilized on mAb 1F5 coupled to cyanogen bromide activated Sepharose 4B (Pharmacia Fine Chemicals, Upsala, Sweden). The antigen came from the B cell line Daudi after solubilization with 0.5% NP-40. The immunogen was suspended in Freund's complete adjuvant and injected into mice intraperitoneally (100 μl sepharose per mouse). Boosts of 40 and 35 μl sepharose were given in incomplete Freund's 25 and 32 days later, respectively. Serum samples taken one week following boosts were analyzed by reactivity on Western blots. Antiserum from one mouse bound to a single B cell-specific band of 33 kd and gave a reaction identical to anti-CD20 mAb 1F5 in immunoprecipitations. In immunofluorescence assays, the serum from this mouse was found to be unreactive with viable cells.

Cloning in λ

cDNA was synthesized from 10 μg poly(A)$^+$ Daudi RNA using an oligo(dT) primer and 250 units of avian myeloblastosis virus reverse transcriptase (Boehringer Mannheim, Indianapolis, IN) [in 50 mM Tris pH 8.5, 40 mM KCl, 8 mM $MgCl_2$, 0.8 mM DTT, 0.5 mM deoxyribonucleotides and 400 units RNAs in (Promega, Madison, WI) for 60 min at 45°C]. The second strand was synthesized in 100 mM Hepes pH 7.1, 10 mM $MgCl_2$, 70 mM KCl, 25 mM DTT, 0.5 mM deoxyribonucleotides, and 20 units Klenow (BRL) at 15°C overnight. After S1 treatment the cDNA was dG-tailed, ligated into *Eco*RI-cut λgt10 using synthetic oligonucleotides with the sequence $AATTC_{12}$ and packaged *in vitro*. *Escherichia coli*, strain C600 Hfl$^+$, were infected with phage and grown on thirty 150 mm dishes at a density which gave distinct plaques. A phage pellet was collected by centrifugation (90 min at 23 000 r.p.m. in SW28 rotor) of an SM buffer (Maniatis *et al.*, 1982) overlay of the plates. λ DNA was purified by incubation of phage at 70°C for 5 min in the presence of 0.2% SDS, 0.2 M Tris pH 8.0, 20 mM EDTA (pH 8.0), and 0.2% diethyl pyrocarbonate; and removal of protein by precipitation in 0.45 M potassium acetate. Following ethanol precipitation the DNA was digested with *Eco*RI, reprecipitated and applied to a 10–40% sucrose step gradient (13 steps) that was centrifuged at 26 000 r.p.m. for 6.5 h at 15°C in an SW28 rotor. The three uppermost fractions, free of λ arms when analyzed by electrophoresis on an 0.7% agarose gel, were pooled, precipitated, ligated into *Eco*RI-cut λgt11 (Yanisch-Perron *et al.*, 1985), and packaged *in vitro*. Plaque lifts of phage infected *E.coli* strain Y1090 onto nitrocellulose filters were exposed to chloroform vapors for 15 min, rinsed in 10 mM Tris pH 8.0, 150 mM NaCl and 0.05% Tween 20 (TBST) and incubated in TBST with 5% dry milk powder for 60 min at room temperature (or overnight at 4°C). Filters were then incubated for 90 min at room temperature in TBST/5% dry milk containing anti-CD20 antiserum at a 1:1000 dilution. After rinsing in TBST, filters were incubated in TBST/5% dry milk with alkaline phosphatase conjugated goat anti-mouse Ig (Promega) at 1:7500 for 30 min, rinsed in TBST and developed using 5-bromo-4-chloro-3-indolyl phosphate and nitro blue tetrazolium (Leary *et al.*, 1983). Plaques representing positive clones were picked and replated twice or until all plaques were positive.

RNA, Northern blots and DNA sequencing

Total RNA was prepared by the guanidinium thiocyanate/CsCl method (Maniatis *et al.*, 1982). Poly(A)$^+$ RNA was selected using oligo (dT) chromatography. Northern blots were performed according to a standard procedure (Maniatis *et al.*, 1982) using agarose-formaldehyde gels. After transfer to nylon (Amersham, Arlington Heights, IL) the RNA was irradiated with UV light and pre-hybridized for at least 4 h at 42°C in 5 × SSC, 5 × Denhardt's, 0.5% SDS, 50% formamide and tRNA for nick-translated fragments; or at 45°C in 5 × SSC, 10 × Denhardt's, 50 mM sodium phosphate pH 7.2, and 7% SDS for oligonucleotides. Probes were hybridized in these solutions for at least 16 h. Nick translations were according to Maniatis (Maniatis *et al.*, 1982). Oligonucleotides were synthesized on a Model 380A DNA synthesizer (Applied Biosystems, Foster City, CA) using the solid-phase phosphoramidite method. Oligonucleotides were labelled with 10 units T4 polynucleotide kinase (New England Biolabs, Beverly, MA) in a 20 μl volume containing 60 mM Tris pH 7.6, 5 mM 2-mercaptoethanol, 10 mM $MgCl_2$ and 150 μCi [γ-^{32}P]ATP (50 pmol) for 30 min at 37°C.

cDNA clones were sequenced by the dideoxy-chain termination method (Sanger *et al.*, 1977, 1980) after subcloning into either bacteriophage M13 derivatives (Webb *et al.*, 1987) for single-stranded DNA, or pEMBL18 plasmid (Chen and Seeburg, 1985; Dente *et al.*, 1983).

In vitro translation of CD20

RNA was synthesized from 2 μg linearized SP65 plasmid DNA (Promega) containing a CD20 cDNA insert, in a volume of 100 μl containing 0.5 mM

ribonucleotide triphosphates, 15 U of SP6 RNA polymerase (Promega), 1 mM dithiothreitol and 100 units RNasin (Promega). The reaction was incubated at 37°C for 60 min. For *in vitro* translation, 5 μg RNA was added to nuclease-treated rabbit reticulocyte lysate (Promega) supplemented with 180 U RNasin, 20 nM methionine-free amino acids and 300 μCi [^{35}S]methionine for 60 min at 30°C. For immunoprecipitations the translation mixture was diluted with an equal volume of 20 mM Tris pH 8.0, 100 mM NaCl, 1 mM EDTA and 1% NP-40. 100 μl of this lysate was incubated at 4°C with 6 μg mAb for 90 min and then 10 min after addition of 400 μl of a 2% Staph A suspension. The bacteria were washed four times with 1 ml of buffer containing 0.5% NP-40 and eluted with SDS–PAGE sample buffer.

CD20 phosphorylation

Resting B cells from tonsils were labelled with ^{32}P for 5 h as described (Oettgen *et al.*, 1983). After stimulation with G28-5 anti-CDw40, CD20 was immunoprecipitated by addition of 1F5 prior to cell lysis as described (Valentine and Clark, 1988).

Acknowledgements

We thank Dr T.Rose for helpful discussions and advice and C.Chapman for preparation of the manuscript. This work was supported by Oncogen Corporation and by Public Health Service Grants RR00166 and GM37905 from the National Institutes of Health.

References

Ballivet,M., Patrick,J., Lee,J. and Heinmann,S. (1982) *Proc. Natl. Acad. Sci. USA*, **79**, 4466–4470.

Chen,E.Y. and Seeburg,P.H. (1985) *DNA*, **4**, 165–170.

Clark,E.A., Shu,G. and Ledbetter,J.A. (1985) *Proc. Natl. Acad. Sci. USA*, **82**, 1766–1770.

Clark,E.A. and Ledbetter,J.A. (1986) *Proc. Natl. Acad. Sci. USA*, **83**, 4484–4498.

Clark,E.A. and Einfeld,D. (1986) In Reinherz,E.L. *et al.* (eds), *Leukocyte Typing II.* Springer-Verlag, Berlin.

Clark,E.A. and Shu,G. (1987) *J. Immunol.*, **138**, 720–725.

Dente,L., Cesareni,G. and Cortese,R. (1983) *Nucleic Acids Res.*, **11**, 1645–1655.

Devillers-Thery,A., Giraudeat,J., Betaboulet,M. and Changeux,J.-P. (1983) *Proc. Natl. Acad. Sci. USA*, **80**, 2067–2071.

Dixon,R.A.F., Kobilka,B.K., Strader,D.J., Benovic,J.L., Dohlman,H.G., Frielle,T., Bolanowski,M.A., Bennett,C.D., Rands,E., Diehl,R.E., Mumford,R.A., Slater,E.E., Sigal,I.S., Caron,M.G., Lefkowitz,R.J. and Strader,C.D. (1986) *Nature*, **321**, 75–79.

Feenewald,S., van Saten,V. and Kieff,E. (1984) *J. Virol.*, **51**, 411–419.

Finer-Moore,J. and Stroud,R.M. (1984) *Proc. Natl. Acad. Sci. USA*, **81**, 155–159.

Foster,D.L., Boublik,M. and Kaback,H.R. (1983) *J. Biol. Chem.*, **258**, 31–34.

Furutani,Y., Hirose,T., Miyata,T. and Numa,S. (1983) *Nature*, **302**, 528–532.

Gollay,J., Clark,E.A. and Beverley,P.C.L. (1985) *J. Immunol.*, **135**, 3795–3801.

Kopito,R.R. and Lodish,H.F. (1985) *Nature*, **316**, 234–238.

Kozak,M. (1984) *Nucleic Acids Res.*, **12**, 857–872.

Kubo,T., Fukuda,K., Mikami,A., Maeda,A., Takashi,H., Mishinia,M., Haga,T., Haga,K., Ichiyama,A., Kangawa,K., Kojima,M., Matsuo,H., Hirose,T. and Numa,S. (1986) *Nature*, **323**, 411–416.

Kyte,J. and Doolittle,F. (1982) *J. Mol. Biol.*, **157**, 105–132.

Leary,J.L., Brigati,D.J. and Ward,D.C. (1983) *Proc. Natl. Acad. Sci. USA*, **80**, 4045–4049.

Ledbetter,J.A., Martin,P.J., Spooner,C.E., Wofsy,D., Tsu,T.T., Beatty,P.G. and Gladstone,P. (1985) *J. Immunol.*, **135**, 2331–2336.

Ledbetter,J.A. and Clark,E.A. (1986) *Human Immunology*, **15**, 30–43.

Ledbetter,J.A., Shu,G., Gallagher,M. and Clark,E.A. (1987) *J. Immunol.*, **138**, 788–794.

Liebowitz,D., Wang,D. and Kieff,E. (1986) *J. Virol.*, **58**, 233–237.

MacLennan,D.H., Brandl,C.J., Korczak,B. and Green,N.M. (1985) *Nature*, **316**, 696–700.

Maniatis,T., Fritsch,E.F. and Sambrook,J. (1982) *Molecular Cloning: A Laboratory Manual*. Cold Spring Harbor Laboratory Press, Cold Spring Harbor, New York.

Mueckler,M., Caruso,C., Baldwin,S.A., Panico,M., Blench,I., Morris,H.R., Allard,W.J., Lienhard,G.E. and Lodish,H.F. (1985) *Science*, **229**, 941–945.

Nathans,J. and Hogness,D.S. (1983) *Cell*, **34**, 807–814.

Noda,M., Takahaski,H., Tanabe,T., Toyosata,M., Kikyotani,S., Furutani,Y., Hirsoe,T., Miyata,T. and Numa,S. (1983) *Nature*, **302**, 528–533.

Noda,M., Ikeda,T., Kayano,T., Suzuki,H., Tkeshima,H., Kuraskaski,M., Takahaski,H. and Numa,S. (1986) *Nature*, **320**, 188–192.

Oettgen,H.C., Bayard,P.J., Van Ewijk,W., Nadler,L.M. and Terhorst,C.P. (1983) *Hybridoma*, **2**, 17–28.

Padgett,R.A., Grabowski,P.J., Koharsak,M.M., Seiler,S. and Sharp,P.A. (1986) *Annu. Rev. Biochem.*, **55**, 1119–1150.

Rabinovitch,P.S., Clark,E.A., Pezzutto,A., Ledbetter,J.A. and Draves,K.E. (1987) In McMichael,A.J. (ed.), *Leukocyte Typing III.* Oxford University Press, Oxford, pp. 435–439.

Sanger,F., Nicklen,S. and Coulson,A.R. (1977) *Proc. Natl. Acad. Sci. USA*, **74**, 5463–5467.

Sanger,F., Coulson,A.R., Barrell,B.G., Smith,A.J.H. and Roe,B.A. (1980) *J. Mol. Biol.*, **143**, 161–178.

Shull,G.E., Schwartz,A. and Lingrel,J.B. (1985) *Nature*, **316**, 691–695.

Smeland,E., Godal,T., Rudd,E., Beiski,K., Funderud,S., Clark,E.A., Pfeifer-Ohlsson,S. and Ohlsson,R. (1985) *Proc. Natl. Acad. Sci. USA*, **82**, 6255–6259.

Smeland,E.B., Beiske,K., Ohlsson,R., Holte,H., Rudd,E., Blomhoff,H.K., Jefferis,R. and Godal,T. (1987) *J. Immunol.*, **138**, 3179–3184.

Stashenko,P., Nadler,L.M., Hardy,R. and Schlossman,S.F. (1980) *J. Immunol.*, **125**, 1678–1685.

Tedder,T.F., Boyd,A.W., Freedman,A.S., Nadler,L.M. and Schlossman,S.F. (1985) *J. Immunol.*, **135**, 973–979.

Tedder,T.F., Forsgren,A., Boyd,A.W., Nadler,L.M. and Schlossman,S.F. (1986) *Eur. J. Immunol.*, **16**, 881–887.

Valentine,M.A. and Clark,E.A. (1988). *Proc. Natl. Acad. Sci. USA*, in press.

Webb,N.R., Rose,T.M., Malik,N., Marquardt,H., Shoyab,M., Todaro,G.J. and Lee,D.C. (1987) *DNA*, **6**, 71–79.

Yanisch-Perron,C., Vieira,J. and Messing,J. (1985) *Gene*, **33**, 103–119.

Young,R.A. and Davis,R.W. (1983) *Proc. Natl. Acad. Sci. USA*, **80**, 1194–1198.

Received on December 4, 1987