Structure of the Mouse CD72 (Lyb-2) Gene and Its Alternatively Spliced Transcripts[1]

Han Ying, Emiko Nakayama,[2] William H. Robinson, and Jane R. Parnes[3]

Division of Immunology and Rheumatology, Department of Medicine, Stanford University School of Medicine, Stanford, CA 94305

The complete sequence of the CD72 gene from the C57L mouse, including the 5′ and 3′ flanking sequences, is reported. The gene spans 6830 base pairs and includes nine exons surrounding eight introns. It does not have an obvious TATAA box, so it belongs to a group of genes with TATA-less promoters that are regulated during mammalian immunodifferentiation. cDNA sequence comparisons among $CD72^a$, $CD72^b$, and $CD72^c$ alleles have demonstrated two distinct seven amino acid insertion/deletions among these allelic variants. Based on our genomic sequence studies as well as PCR analyses, we found that different strains of mice can alternatively or exclusively use either of two AG sites surrounding the 21-bp insertion/deletions as 3′ splice sites in an allele-specific manner. Other alternative splicing events, such as exon skipping, also contribute to CD72 polymorphism. In mouse splenic B cells there are allele-specific distributions of CD72 mRNAs that contain sequences from both exon 3 and exon 4, from either exon 3 or exon 4, or from neither exon 3 nor 4. It is unclear what the in vivo function might be of the proteins encoded by the mRNA forms lacking these exon sequences. *The Journal of Immunology,* 1995, 154: 2743–2752.

CD72 (Lyb-2) is a 45-kDa type II transmembrane protein expressed predominantly on B-lineage cells in both mouse and human (1–5). Expression of this protein is lost at the terminally differentiated plasma cell stage (2, 4–6). Although the precise function of CD72 is unknown, studies have shown that anti-mouse-CD72 mAb induces an increase in the metabolism of phosphatidylinositol in purified small splenic B cells (7). It also induces an increase in MHC class II expression on B cells (8–9) and mobilization of small amounts of cytoplasmic free Ca^{2+} in those cells (9). Anti-mouse-$CD72^a$ mAb can induce proliferation of B cells and can synergize with IL-4 in the induction of Ag-specific B cells (9–14). Anti-CD72 mAb 10.1.D2 and its monovalent Fab′ fragment could directly transform a fraction of small resting B cells into blasts (10, 11, 13). Two mAbs specific for the mouse $CD72^a$ allele have been shown to inhibit the Ab response by splenic B cells to T cell-dependent Ags but not to T cell-independent Ags (10, 15, 16). It was recently discovered that CD5, a 67-kDa transmembrane protein expressed on all T cells as well as on a subgroup of B cells (CD5 or $B1^a$ cells), is a natural ligand of CD72 (17–19). This finding suggests that CD72 and CD5 may play an important role in B and T cell cross-talk, and their interaction may have significant influence upon T and B cell activation and proliferation.

Conventional antisera and mAbs have identified five serologic specificities defining four allelic forms of mouse CD72: $CD72^a$, $CD72^b$, $CD72^c$, and $CD72^d$ (3). We have cloned the cDNAs encoding mouse $CD72^a$, $CD72^b$, and $CD72^c$ (4, 20), as well as the human homologue (5), and demonstrated that CD72 alleles are highly conserved in their cytoplasmic domain but exhibit a high degree of polymorphism in their extracellular domains (20). cDNA sequence comparisons showed two distinct seven-amino acid insertion/deletions among these allelic variants. One short form of CD72 cDNA lacking sequence from exon 3 and exon 4 was isolated both from C57BL/6 mouse spleen RNA and from the CH12.LX B cell lymphoma line. Another short-form lacking sequence from exon 3 was also isolated from C57BL/6 mouse spleen RNA (20). In this paper, we show the complete genomic sequence of the mouse CD72 gene as well as its intron/exon structure. We

Received for publication April 27, 1994. Accepted for publication December 8, 1994.

The costs of publication of this article were defrayed in part by the payment of page charges. This article must therefore be hereby marked *advertisement* in accordance with 18 U.S.C. Section 1734 solely to indicate this fact.

[1] This work was supported by National Institutes of Health Grant AI30155 to J.R.P. W.H.R. is supported by the Molecular and Cellular Immunobiology Training Grant AI07290 from the National Institutes of Health.

[2] Deceased.

[3] Address correspondence and reprint requests to Dr. Jane R. Parnes, Division of Immunology and Rheumatology, Medical School Lab, Surge Building P-306, Stanford University Medical Center, Stanford, CA 94305-5487.

0022-1767/95/$02.00

also demonstrate that the two seven-amino acid insertion/deletions each result from alternative usage of two distinct 3′ RNA splice sites lying 21 bp apart. In addition, we show that there are a variety of different-size forms of CD72 mRNA resulting from alternative splicing patterns that either include or exclude sequences from exons 3 and 4, which encode part of the cytoplasmic tail and the transmembrane region, respectively. Finally, we show that the distribution pattern of these various mRNA species is also allele-specific.

Materials and Methods

Mice and cell lines

DBA/2, C57L, BALB/c, C57BL/6, AKR, and SJL mice were purchased from The Jackson Laboratory (Bar Harbor, ME). The pre-B cell line L1.2 was provided by Dr. I. Weissman (Stanford University).

Genomic screening and subcloning

A genomic library from MboI partially digested DNA from the C57L mouse pre-B cell line L1.2 was prepared in the vector Lambda Fix according to the manufacturer's instructions (Stratagene, La Jolla, CA). Clones were screened as described (21) using a mouse CD72^a cDNA (C57L) probe labeled with [α-^{32}P] dCTP (Amersham, Arlington Heights, IL) by nick translation. A positive clone was isolated and the 15.2-kb insert containing the whole CD72 gene was subcloned into Bluescript KS$^+$ (Stratagene, La Jolla, CA) at *Hin*dIII and *Sal*I restriction enzyme sites. The integrity of the CD72 gene was demonstrated by its ability to yield allele-specific surface expression of CD72^a protein after transfection into B cell lines (data not shown) as well as by its sequence. It was further characterized by restriction enzyme mapping, digested into smaller fragments and subcloned into M13 mp18, M13 mp19, and Bluescript vectors for sequence analysis.

CD72 gene sequence analysis

Nucleotide sequences were determined by the method of Sanger and coworkers (22). Single-stranded or double-stranded DNA templates were generated from subclones of CD72 in M13 mp18, M13 mp19, and Bluescript vectors. Sequencing primers included T7, T3, M13 -20, and M13 -40 primers and a series of synthetic oligonucleotides that correspond to known mouse CD72 sequences. Sequences were compiled by first using Microgenie and then the GCG sequence analysis program (Genetic Computing Group, Madison, WI). The majority of the genomic fragments were sequenced on both strands, but a portion of the 5′ and 3′ ends were sequenced on one strand on at least two independent subclones.

Sequence analysis of PCR products

A genomic fragment containing sequences from the 3′ end of exon 6, the whole of intron 6, and the 5′ end of exon 7 was obtained from DBA/2 (CD72^a) mouse tail genomic DNA by PCR using oligo 5 (representing sequence near the 3′ end of exon 6) and oligo 6 (representing sequence near the 5′ end of exon 7) as PCR primers. Each 100-μl reaction included 100 ng of genomic DNA, 50 ng of each oligo, 10 μl of 10X Taq polymerase reaction buffer (Stratagene, La Jolla, CA), 5 μl dNTPs (5 mM each), and 2.5 units of Taq polymerase (Stratagene, La Jolla, CA) (Table I). The reactions were stopped after 30 cycles. PCR products were gel purified and sequenced directly using oligo 6 as a primer. A genomic fragment containing sequences from the 3′ end of exon 7, the whole of intron 7, and the 5′ end of exon 8 was obtained from AKR mouse (CD72^c) tail DNA by PCR using oligo 7 (representing sequence from the 3′ end of exon 7) and oligo 8 (representing sequence from the 5′ end of exon 8) as primers. The PCR products were subcloned into the Bluescript SK$^-$ vector at the *Eco*RV site. Double-stranded DNA was prepared and sequenced using oligo 8 as a primer. The complete sequences of oligonucleotides used in these PCR experiments are listed below (5′ to 3′). Numbers in parentheses indicate positions of those oligonucleotides in the CD72 cDNA (according to the numbering in Robinson et al. (20)). In each case, the 5′ GGATCC does not derive from the cDNA sequence but was added for subcloning.

Table I. *PCR thermal cycle conditions*

Cycle	Denaturation	Annealing	Polymerization
First cycle	5′ at 94°C	2′ at 50°C	2′ at 72°C
Subsequent*	1′30″ at 94°C	2′ at 50°C	2′ at 72°C
Last cycle	1′ at 94°C	2′ at 50°C	10′ at 72°C

* Cycles 2–29.

oligo 5 (806–828) GGATCCTACTGCACATCTCTGTCCTCC

oligo 6 (911–895) GGATCCTAAGCCGCTGGGCAGAG

oligo 7 (909–930) GGATCCTTAGAGGAGTTGCTAGATCGT

oligo 8 (1063–1045) GGATCCGGTGAAGCTCTGCACACT

Analysis of alternative splicing by gel electrophoresis of PCR products

Total spleen RNA from different strains of mice was prepared by the RNAzol method according to the manufacturer's instructions (CINNA/BIOTECX Laboratories International Inc., Friendswood, TX), reverse transcribed into cDNA, and used as PCR templates. PCR was performed using synthesized oligonucleotides as primers; the 5′ primer in each PCR reaction was end-labeled with [γ-^{32}P] ATP (Amersham). In each PCR reaction, 50,000 cpm of the end-labeled 5′ oligo (supplemented with 50 ng of the same cold oligo) and 50 ng of the 3′ oligo were added. cDNA reverse-transcribed from 1 μg of splenic RNA was used as template. PCR conditions were the same as above. PCR products were analyzed by electrophoresis on a denaturing polyacrylamide gel (6% acrylamide, 6 M urea). Control reverse transcription and PCR reactions using total RNA from the thymoma line 1010, which does not express CD72, did not yield detectable PCR products (data not shown). The sequences of the oligonucleotides used in the PCR analyses are the same as above plus additional oligonucleotides listed below (5′ to 3′). Numbers in parentheses indicate positions of those oligonucleotides in the CD72 cDNA (according to the numbering in Robinson et al. (20)):

oligo 9 (201–218) CCTGCACTAGCGGACAA

oligo 10 (543–524) ATGTGTCCTGGCTCGAGATC

oligo S1 (433–417) GGTGGCTTCCCAAATCCTGG

It should be noted that calculation of predicted sizes of PCR products takes into account the added sequence at the 5′ ends of oligos 5–8.

Purification of splenic B cells

Fresh spleen cells from different strains of mice were suspended in HBSS plus 2% FCS and incubated in 100-mm × 20 mm-tissue culture dishes (Corning, Inc., Corning, NY) coated with goat anti-mouse IgM antiserum (Cappel, West Chester, PA) for 30 min. Unbound cells were removed by one wash with PBS and bound cells were then collected using Pasteur pipettes. The purity of B cells prepared in this manner is greater than 95%.

Results

Organization of the mouse CD72 gene

To clone the mouse CD72 gene we constructed and screened a genomic library from the C57L mouse (CD72^a) pre-B cell line L1.2 using a CD72^a cDNA as a probe.[4] One positive clone with an insert of 15.2 kb contained the full-length

[4] The accession number of our CD72 gene sequence is L35772 at the sequence data base at the National Center for Genome Resources.

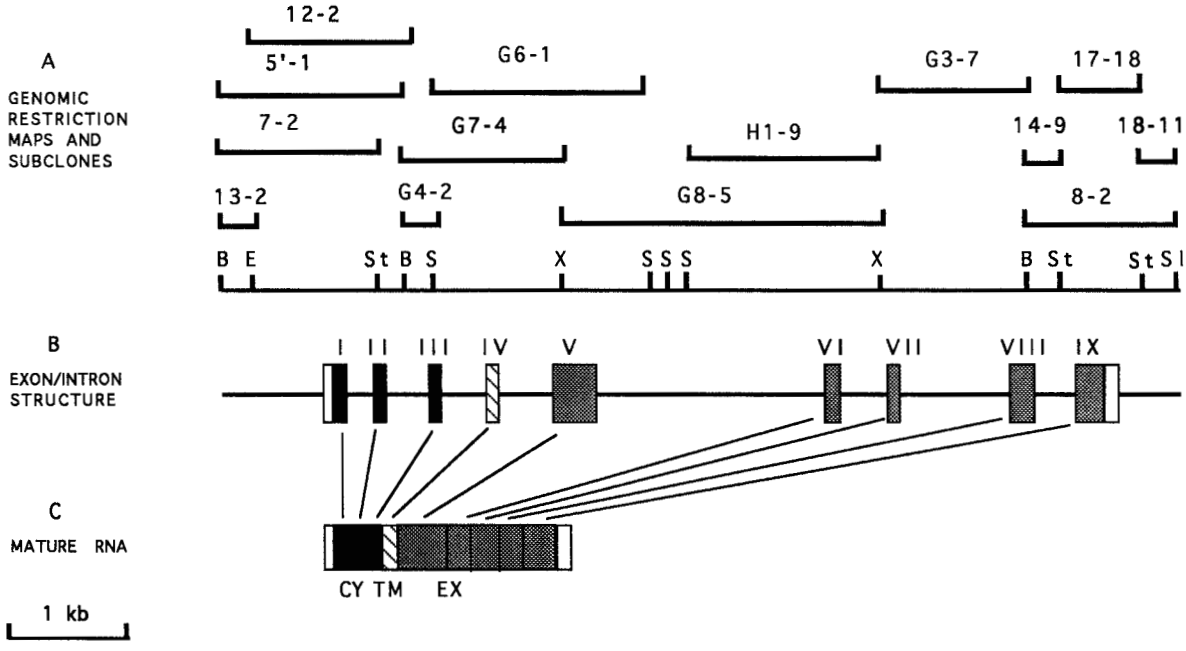

FIGURE 1. Sequencing strategy and organization of the mouse CD72[a] gene. *A*) Restriction enzyme map of the CD72[a] gene from the C57L mouse pre-B cell line L1.2. The restriction enzyme recognition sites are: B: *Bam*HI; E: *Eco*RV; St: *Stu*I; S: *Sac*I; Sl: *Sal*I; X: *Xho*I. The bars above represent subclones made by inserting the indicated genomic fragments into M13 mp18, M13 mp19, or Bluescript vectors. DNA sequences were obtained by using single-stranded or double-stranded DNA as templates and M13 -20 primer, M13 -40 primer, T7 and T3 primers or CD72 sequence-specific oligonucleotides as primers. *B*) The intron/exon structure of the CD72[a] gene is shown. Exons are indicated by boxes and numbered with roman numerals. Exons I to III encode the cytoplasmic domain (CY); exon IV encodes the transmembrane domain (TM); exons V through IX encode the extracellular domains (EX). The open box at the beginning of exon I represents the 5′ untranslated region and that at the end of exon IX represents the 3′ untranslated region. *C*) The structure of the mature mRNA and the contribution of the various exons to the encoded protein structure are illustrated.

CD72 gene as shown by its ability to yield allele-specific surface expression of CD72[a] protein after transfection into B cell lines (data not shown) and by its sequence analysis. Subclones were sequenced as shown in Figure 1. Exon sequences and exon/intron borders were identified by comparison to CD72 cDNA sequences (4, 20).

The mouse CD72[a] gene spans 6830 bp from the beginning of the first exon (for the current analysis the 5′ end of the cDNA isolated from the C57L mouse is considered as the 5′ end of the first exon [4]) to the 3′ end of the last exon (i.e., the poly[A] addition site [4]) (Fig. 2). It is composed of 9 exons surrounding 8 introns. The exon sequences are identical to the published cDNA sequence from the C57L mouse. Exons 1–3 encode the cytoplasmic domain. Exon 4 encodes the transmembrane domain. Exons 5–9 encode the extracellular domain. All introns are bounded by the canonical GT-AG splicing sequences. Notably, intron 6 alternatively uses two different AG sites at the beginning of exon 7 as 3′ splice sites, and intron 7 uses two different AG sites at the beginning of exon 8 as 3′ splice sites. These alternative splicing patterns will be discussed later.

The DNA sequence of approximately 1 kb of the 5′ flanking sequence was determined (Fig. 2). There is no obvious consensus TATAA box in the 5′ flanking region, nor is there any obvious consensus initiator sequence. We sequenced about 600 bp of the 5′ flanking sequences of CD72[b] (derived from C57BL/6 mice) and CD72[c] alleles (derived from SJL mice) (data not shown). We also sequenced the 5′ flanking sequence of the human CD72 gene up to 1 kb upstream of the translation start site (data not shown). We did not find any TATAA consensus or initiator consensus in those sequences. Preliminary S1 mapping and primer extension analyses suggest multiple transcription start sites of the gene (data not shown). Comparing the human and mouse 5′ flanking sequences, we found several identical fragments. These include the sequence 5′ TTCCTC 3′ beginning at nucleotide 965 of the mouse gene; its reverse complement precisely matches the consensus sequence recognized by the B cell- and macrophage-specific transcription factor PU.1 (5′ GAGGAA 3′) (23). The study of the role in CD72 transcription regulation of the putative PU.1 recognition sequence and other fragments homologous between mouse and human CD72 5′ flanking sequences is in progress.

3′ untranslated sequences are thought to influence polyadenylation and mRNA processing (24, 25). The CD72 cDNA poly(A) tail is not preceded by the canonical

poly(A) addition signal AATAAA. However, there are several related sequences located at the appropriate location (e.g., AATATT or TTTAAA ending, respectively 12 or 6 bp upstream of the poly[A] tail) (4). The sequence TGTGTGTC is found 15 bp downstream of the polyadenylation site. This sequence corresponds fairly well to a consensus sequence, YGTGTTYY (Y = pyrimidine), which has been shown to be required for efficient formation of the 3′ terminus of mRNA and is found in most mammalian genes within 50 bp downstream of the conserved AATAAA sequence (25).

Several repeat sequences are present in the CD72 gene. In the 5′ flanking region the sequence CAGGAGGAGG GGGCTCTGGCTGGA is present as three consecutive direct repeats from nucleotide 778–849 (in the sequence from C57L mice). In the 5′ flanking sequence from C57BL/6 ($CD72^b$) and from SJL ($CD72^c$) mice, this 24-bp sequence is present as two consecutive direct repeats. In intron 5, from nucleotide 4749–4765 (Fig. 2), there is a stretch of G repeated 16 times in the $CD72^a$ gene. Two series of AG repeats (nucleotides 4770–4782 and 4816–4841) are also found in intron 5, just downstream of the G repeat. The role of these repeats, if any, in CD72 regulation is unknown.

Alternative usage of two 3′ splice sites at the intron 6/exon 7 junction

We have previously isolated a cDNA clone from the DBA/2 mouse ($CD72^a$) which differs from that of the original $CD72^a$ cDNA clone from the C57L mouse only by a 21-bp insertion of nucleotides 870–890 in the cDNA (20). This insertion is also absent in our previously described $CD72^b$ and $CD72^c$ cDNA clones (Fig. 3*A* and Ref. (20)). In the C57L mouse genomic sequence, the 21-bp insertion is located at the junction of intron 6 and exon 7 (Figs. 2 and 3*B*). It is flanked by an AG at its 5′ end and ends with an AG at its 3′ end. Each of these AGs could potentially represent a 3′ splice site, because each is preceded by a stretch of pyrimidines and each sequence context matches

FIGURE 2. Nucleotide sequence of the mouse CD72 gene from C57L mice. The 9 exons and 8 introns together with 5′ and 3′ flanking sequences are shown. Exons are line-shifted. Protein coding sequences are indicated by bold letters. The underlined or overscored sequences in exons indicate the locations of the oligonucleotides used for PCR analyses, and the 5′ to 3′ orientations of those oligonucleotides are indicated by arrows. In the 5′ flanking sequence, a 24-bp sequence repeated three times consecutively from nucleotide 778 to 849 is underlined, as is a TTCCTC sequence, representing a PU.1 box on the complementary strand, beginning at nucleotide 965. The indicated 5′ boundary of the first exon is based on the start of the cDNA sequence from C57L (4). The 3′ end of exon 9 corresponds to the location of the poly(A) tail at the end of the cDNA sequence from C57L.

A

```
                                exon7                    exon8

                   945                                                          992

CD72a(DBA/2)       ATA CAG ... ATG AGC AAG AAG TGG AGG CAT GAC TAT GAC TCT CAA AGC CGA
CD72a(C57L)        --- --- ... --- --- --- --- --- --- --- --- --- --- --- --- --- ---
CD72b(BALB/c)      --- --- ... --- --- --- --- --- --- --G --- -C- --- --- --- --- ---
CD72c(AKR)         --- --- GAG --- --- --- --- --- --. ... ... ... ... ... ... ..- TC-
```

B

```
                          intron 7             exon 8

CD72a(C57L)     GTGTGTCCCAACTCTGTTCACCAGGCATGACTATGACTCTCAAAGCCGATATTGTGACAA
CD72c(AKR)      ----A-----------------CT-TC--CA-C-T-----------TC-----------T
```

FIGURE 3. Comparison of the nucleotide sequences of mouse CD72 cDNA and gene at the 5′ end of exon 7. *A*) The cDNA sequences corresponding to the junction of sequences derived from exons 6 and 7 are compared among DBA/2 (CD72^a), C57L (CD72^a), BALB/c (CD72^b), and AKR (CD72^c) mice. Numbers above the sequence are locations of the first and the last nucleotide on each line based on the numbering of the previously published cDNA sequence from DBA/2 (20). Identical nucleotides are indicated by hyphens, and gaps in the alignment are indicated by dots. Sequence derived from exon 7 is underlined. The cDNA from DBA/2 has a 21-bp insertion from nucleotide 870–890 (indicated by bold letters) which is located at the immediate 5′ end of the sequence from exon 7. *B*) Genomic sequences (part of intron 6 and exon 7) from DBA/2 and C57L mice are compared. The sequence from DBA/2 was obtained as described in *Materials and Methods*. The two potential 3′ splice site AGs are indicated by bold letters. The underlined sequence is the 21-bp insertion in the cDNA sequence from DBA/2. Identical nucleotides are indicated by hyphens.

very well with a typical 3′ splice site (-[Py]$_n$NCAG-) (26, 27). From the genomic sequence we predict that C57L mice could potentially use either AG as a 3′ splice site. To determine whether DBA/2 mice have a similar genomic organization at the intron 6/exon 7 junction, an 800-bp DNA fragment spanning intron 6 and containing a portion of the 3′ end of exon 6 and the 5′ end of exon 7 was isolated by PCR using genomic DNA from the DBA/2 mouse as a template and oligo 5 from within exon 6 and oligo 6 from within exon 7 as primers (Fig. 2 and *Materials and Methods*). The PCR product was directly sequenced using oligo 6 as a primer. We compared the sequences at the intron 6/exon 7 junction between the DBA/2 and the C57L mouse (Fig. 3*B*), and they are identical within the region examined. This suggests that DBA/2 mice may also be capable of using both AGs as 3′ splice sites. If this is true, two different forms of mRNA should be produced in both DBA/2 and C57L mice. To test whether both forms of mRNA can be produced in any one strain or whether this alternative splicing is allele-specific, total spleen RNA from six strains representing three different alleles was prepared, reverse transcribed, and analyzed by PCR using oligo 5 and oligo 6 as primers. PCR products were analyzed on a polyacrylamide gel (Fig. 4). According to the positions of the primers in CD72 cDNA, a 117-bp fragment will be generated if the 5′ AG is used, and a 96-bp fragment will be generated if the 3′ AG is used (taking into account the linker sequences at the ends of the two oligonucleotides). Figure 4 shows that all

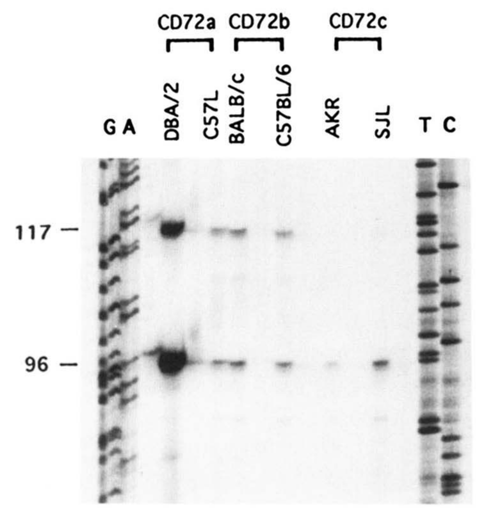

FIGURE 4. PCR analysis of the alternative usage of two 3′ splice sites at the intron 6/exon 7 border. Total spleen RNA was prepared from six mouse strains representing three CD72 alleles, reverse transcribed, and analyzed by PCR using oligo 6 and end-labeled oligo 5 as primers. PCR products were analyzed by polyacrylamide gel electrophoresis (see *Materials and Methods*). Sequencing reactions using a CD72^b cDNA clone from C57BL/6 as a template and oligo 5 as a primer were used as size markers and are shown on the left (G and A reactions) and right (T and C reactions). PCR products from spleen RNA from DBA/2 (CD72^a), C57L (CD72^a), BALB/C (CD72^b), C57BL/6 (CD72^b), AKR (CD72^c), and SJL (CD72^c) mice are indicated. The numbers in the left margin indicate the sizes of the bands at those positions.

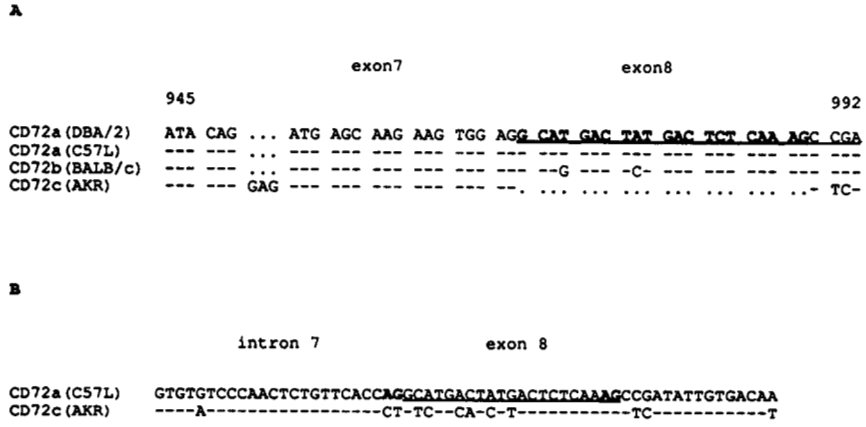

FIGURE 5. Comparison of the nucleotide sequences of mouse CD72 cDNA and gene at the 5′ end of exon 8. *A*) The cDNA sequences corresponding to the junction of exons 7 and 8 are compared among DBA/2 ($CD72^a$), C57L ($CD72^a$), BALB/c ($CD72^b$), and AKR ($CD72^c$) mice. Numbers on the top of the sequences represent locations of the first and last nucleotide shown relative to the previously published cDNA sequence from DBA/2 (20). Identical nucleotides are indicated by hyphens, and gaps in the alignment are indicated by dots. Exon 8 sequences are underlined. The cDNA from AKR has a 21-bp deletion of nucleotides 966–986 (the sequence indicated by bold letters in DBA/2), at the immediate 5′ end of exon 8, and a 3-bp insertion between nucleotides 947 and 948 (relative to the cDNA from DBA/2). *B*) The genomic sequences (part of intron 7 and exon 8) from C57L and AKR are compared. The sequence from AKR was obtained as described in *Materials and Methods*. The potential 3′ splice site AGs are indicated by bold letters. The underlined sequence is the 21-bp deletion in the cDNA from AKR. Identical nucleotides are indicated by hyphens.

strains examined have both alternatively spliced products. $CD72^a$ strains appear to use both sites equally well, consistent with the genomic sequence analysis (Fig. 3*B*). The pattern of usage of the two splice sites seems to be the same for $CD72^b$ strains as for $CD72^a$ strains, so it is likely that the $CD72^b$ gene sequence also contains these two consensus 3′ splice sites. In contrast, the $CD72^c$ strains use the 3′ AG far more than the 5′ AG as a 3′ splice site. It is therefore likely that a polymorphism between the $CD72^c$ allele and other alleles results in the 5′ AG being less favorable as a 3′ splice site. These results were further confirmed by an additional set of PCR reactions using oligo 5 from within exon 6 and oligo 8 from within exon 8 as primers (see below and Fig. 7).

To confirm that the PCR bands in Figure 4 are CD72-specific, we did control PCR using cDNA reverse transcribed from total RNA from the mouse thymoma line 1010 (which does not express CD72) as the template and oligo 5 and oligo 6 as primers. This PCR reaction did not yield any detectable PCR product (data not shown), demonstrating the specificity of the observed bands from spleen RNA for CD72. Similar control reverse transcription and PCR reactions were also performed for experiments described below, using total RNA from the cell line 1010 (see Figs. 6, 7, and 8). Again, these PCR reactions did not yield any detectable PCR products (data not shown).

Alternative usage of two 3′ splice sites at the intron 7/exon 8 junction

We have previously shown (20) that the cDNA from a $CD72^c$ strain (AKR), as compared with those from both $CD72^a$ and $CD72^b$ strains, has a 3-bp insertion of a GAG between cDNA positions 950 and 951 and a 21-bp deletion from cDNA positions 966 to 986 (Fig. 5*A*). As shown in Figure 5*B*, this 21-bp deletion is flanked by an AG at its 5′ end and ends with another AG at its 3′ end in the genomic sequence from C57L mice. The 5′ AG is preceded by CTCTGTTCACC. This sequence compares well with the 3′ splice-site consensus -$[Py]_n$NCAG-, in that the AG is preceded by a polypyrimidine tract and there is a cytidine immmediately 5′ of the AG. The 5′AG should therefore be an efficient 3′ splice site, However, the 3′ AG is preceded by ATGACTCTCAA which does not have as many pyrimidines as the sequence preceding the 5′ AG and has an adenine rather than a cytidine immediately 5′ of AG. The 3′ AG should therefore be a less favorable 3′ splice site than the 5′ AG. To examine the sequence from $CD72^c$ mice at this junction, a genomic fragment spanning intron 7 was isolated by PCR from AKR genomic DNA using oligo 7 from within exon 7 and oligo 8 from within exon 8 as primers, and the intron 7/exon 8-junction fragment was sequenced using oligo 8 as a primer (see *Materials and Methods*). In contrast to C57L mice ($CD72^a$), AKR mice do not have the 5′ AG, which is instead replaced by CT (Fig. 5*B*) and hence, cannot be used as a 3′ splice site. In addition to other sequence differences, two purines within the 21-bp sequence from the C57L mouse located, respectively, 11 and 9 bases 5′ of the 3′ AG, are replaced by two pyrimidines in the sequence derived from the AKR mouse (A to C, G to T). These changes make the pyrimidine tract better in the sequence from AKR mice than in the sequence from C57L mice, and thus should

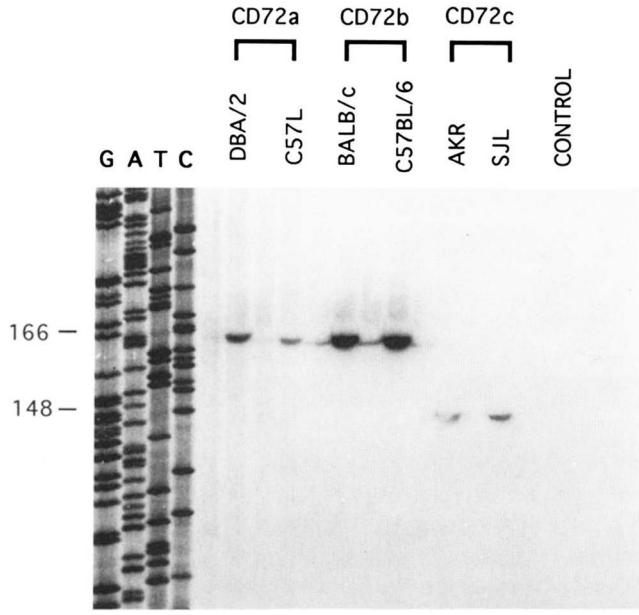

FIGURE 6. PCR analysis of the alternative usage of two 3′ splice sites at the intron 7/exon 8 border. Total spleen RNA was prepared from six mouse strains representing three CD72 alleles, reverse transcribed, and analyzed by PCR using oligo 8 and end-labeled oligo 7 as primers. PCR products were analyzed by polyacrylamide gel electrophoresis (see *Materials and Methods*). *Lanes G, A, T,* and *C* are sequencing reactions performed using a C57BL/6 $CD72^b$ cDNA clone as a template and oligo 5 as a primer to serve as size markers. PCR products from cDNAs from DBA/2 ($CD72^a$), C57L ($CD72^a$), BALB/c ($CD72^b$), C57BL/6 ($CD72^b$), AKR ($CD72^c$), and SJL ($CD72^c$) are as indicated. Control PCR was performed identically to all the others except that no template was added. The numbers in the left margin indicate the sizes of the bands at those positions.

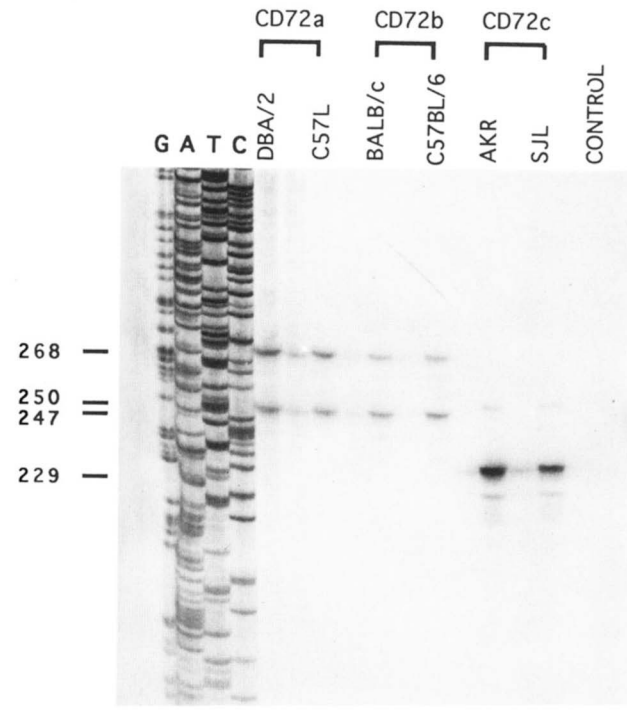

FIGURE 7. Additional analysis of 3′ splice site usage in mRNA corresponding to exons 7 and 8. Total spleen RNA was obtained from DBA/2, C57L, BALB/c, C57BL/6, AKR, and SJL mice, reversed transcribed and analyzed by PCR using oligo 8 and end-labeled oligo 5 as primers. The PCR products were analyzed by polyacrylamide gel electrophoresis. *Lanes* are labeled according to the origin of the template. Control PCR was performed identically to all the other samples except that no template was added. Sequencing reactions from a $CD72^b$ cDNA clone using oligo 5 as a primer were used as size markers (*lanes G, A, T,* and *C*). The numbers in the left margin indicate the sizes of the bands at the indicated positions.

make the 3′ AG in the sequence from AKR mice a potentially more efficient 3′ splice site. We thus predict that $CD72^a$ strains should mostly use the 5′ AG as a 3′ splice site while $CD72^c$ strains can use only the 3′ AG as the 3′ splice site (because it is the only one present, at least in AKR mice). To confirm this hypothesis, PCR analysis was performed using reverse-transcribed cDNA from total spleen RNA as a template and oligo 7 and oligo 8 as primers. The PCR products were electrophoresed on a polyacrylamide gel. In $CD72^a$ strains oligo 7 and oligo 8 will generate a 166-bp fragment if the 5′ AG is used as a splice site or a 145-bp fragment if the 3′ AG is used. In $CD72^c$ strains, the primers will generate a 148-bp fragment and no larger fragment if only the 3′ AG is used. Figure 6 clearly shows that both $CD72^a$ and $CD72^b$ use only the 5′ AG as a 3′ splice site. On the other hand, $CD72^c$ strains use only the 3′ AG as the 3′ splice site.

To further confirm the results from the above two sets of experiments regarding alternative splicing at the intron 6/exon 7 and intron 7/exon 8 junctions, we performed another PCR using oligo 5 and oligo 8 as primers (Figs. 2, 7, and *Materials and Methods*). If the interpretation from both sets of experiments is correct, we should expect to see two bands in the PCR products from $CD72^a$ and $CD72^b$ mice. For example, in DBA/2 mice the size of the upper band should be 268 bp, representing sequences including a portion of exon 6, the 21-bp insertion at the junction of intron 6/exon 7, exon 7, the 21-bp insertion at the junction of intron 7/exon 8 and a portion of exon 8. The size of the lower band should be 247 bp, representing the same sequence as the upper band except that the 21-bp insertion at the junction of intron 6/exon 7 is not included. Similarly, we should be able to see two bands in the PCR products from $CD72^c$ mice. The upper (minor) band, which includes the 21-bp insertion at the intron 6/exon 7 junction but not the 21-bp insertion at the intron 7/exon 8 junction, should be 250 bp. Because the cDNA from $CD72^c$ has a 3-bp insertion between nucleotide 950 and 951 in the cDNA sequence, the upper band from $CD72^c$ is 3 bp larger than the lower band from $CD72^a$ and $CD72^b$. The lower

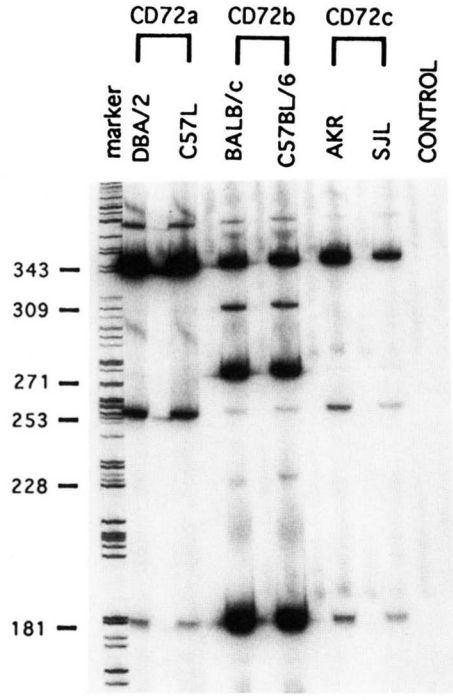

FIGURE 8. Alternative splicing patterns of mRNAs encoded in exons 3 and 4. To analyze the alternative usage of sequence encoded in exon 3 and exon 4 of the CD72 gene, oligo 10 from within exon 5 and end-labeled oligo 9, which is located at the immediate 3′ end of exon 2, were used for PCR analysis. Total RNA from purified spleen B cells of the indicated strain was reverse transcribed and used as a PCR template. PCR products were analyzed by electrophoresis on a polyacrylamide gel. The marker lane is a sequencing reaction to serve as a size marker and only the G reaction is loaded. *Lanes* are labeled according to the template of origin. The control lane is a PCR reaction performed identically to all the others except that no template was added. The numbers in the left margin indicate the sizes of the bands at the indicated positions.

(major) band from $CD72^c$, which represents sequences including neither of the two 21-bp insertions, should be 229 bp. Our results in Figure 7 correspond to these predictions and are consistent with the results in Figures 4 and 6.

Exon skipping in CD72 mRNA

In addition to full-length cDNA clones, we isolated cDNA clones missing either 162 bp (nucleotides 223 to 384) or 72 bp (nucleotides 223 to 294) from C57BL/6 mice ($CD72^b$) (20). Based on the genomic sequence, we know that the first of these cDNAs lacks precisely the sequences from both exon 3 (which encodes part of the cytoplasmic tail) and exon 4 (which encodes the transmembrane domain), whereas the second cDNA lacks the sequence from exon 3. We therefore questioned whether such exon-skipping events are unique to one CD72 allele or are generalizable. We also questioned whether there might be a form of mRNA lacking only the sequence from exon 4. We designed a PCR analysis using cDNA reverse transcribed from purified splenic B cells from six different strains as templates. Oligo 10, from within exon 5, and end-labeled oligo 9, from within exon 2, were used as primers (Fig. 2 and *Materials and Methods*). The PCR products were analyzed on a polyacrylamide gel. Theoretically, the full-length mRNA will generate a 343-bp PCR fragment; the exon 3^- mRNA will yield a 271-bp band; the exon 4^- mRNA will yield a 253-bp band and the exon 3^-4^- mRNA will give rise to a 181-bp band. Figure 8 shows that all these bands are present. Because we have not previously isolated an exon 4^- cDNA clone, we proved that the 253-bp band indeed represents such a transcript by cutting it out, then eluting, subcloning, and sequencing the DNA. As shown in Figure 8 and Table II, the distribution of these alternatively spliced forms of CD72 mRNA varies according to the allele. In $CD72^a$ and $CD72^c$ strains of mice, the full-length mRNA (i.e., including sequence from all exons) predominates. In these strains, there is very little of the exon 3^- form (<1%), approximately 10 to 20% of the mRNA is exon 4^- and 5 to 10% of the mRNA is exon 3^-4^-. In contrast, in $CD72^b$ strains only about 20% of the mRNA is full length, about 25% is exon 3^-, less than 1% is exon 4^- and 40% is exon 3^-4^-. We performed a similar PCR analysis using the same primers but with cDNA reverse transcribed from total spleen RNA, as opposed to purified B cells as templates. The results were identical to those in Figure 8 (data not shown).

Table II. *The relative level of expression of exon 3/exon 4 alternatively spliced CD72 gene products*[a]

Mouse Strain	Full Length	309-bp band	exon 3-	exon 4-	exon 3- exon 4-
DBA/2 ($CD72^a$)	70%		<1%	20%	5%
C57L ($CD72^a$)	75%		<1%	20%	5%
BALB/c ($CD72^b$)	20%	5%	25%	<1%	40%
C57BL/6 ($CD72^b$)	20%	5%	25%	<1%	40%
AKR ($CD72^c$)	75%		<1%	10%	10%
SJL ($CD72^c$)	75%		<1%	10%	10%

[a] The autoradiograph shown in Figrue 8*A* was scanned using a GS300 Transmittance/Reflectance Scanning Densitometer (Hoefer Scientific Instrument, San Francisco, CA). The relative integration value is the integration value of each band divided by the total integration values of all bands in each PCR reaction. The total relative value is less than 100% because the total integration value includes the integration value of nonspecific bands.

In the analysis shown in Figure 8, an additional 309-bp band is present only in $CD72^b$ strains (Fig. 8). When we performed a similar PCR analysis using as primers oligo 9 from within exon 2 and oligo S1 from within exon 5, but closer to the 5′ end of exon 5 than oligo 10 (Fig. 2 and *Materials and Methods*), the four bands corresponding to full-length exon 3^-, exon 4^- and exon 3^-4^- forms were present. However, we did not see a band corresponding to the 309-bp band mentioned above (our unpublished observation). We tried to isolate the 309-bp fragment by running the PCR products (using cold oligo 9 and oligo 10 as primers) on an 8% polyacrylamide gel but could not detect it, despite easy visualization of the bands representing all

the other alternatively spliced transcripts. We are therefore unsure whether the 309-bp band is nonspecific or whether it might represent another infrequent alternatively spliced form with the unique splicing occurring within the sequence of exon 5.

Discussion

We have isolated and sequenced the mouse CD72 gene from the L1.2 pre-B cell line (derived from C57L mice, $CD72^a$). The genomic structure was further examined to delineate the functional regions of the protein, because introns frequently mark the boundaries of structural and functional domains. The CD72 gene consists of 9 exons spanning 6830 bp. We sequenced approximately 1 kb of 5′ flanking region of the $CD72^a$ gene (from C57L mice) and about 600 bp of 5′ flanking regions of the CD72 gene from $CD72^b$ (from C57BL/6 mice) and $CD72^c$ (from AKR mice). There is no obvious TATAA box in the 5′ flanking sequence of any of these three CD72 alleles. Similarly, we have not seen a TATAA box in the 5′ flanking region of the human CD72 gene (H. Ying, I. von Hoegen, and J. R. Parnes, unpublished results). We are currently performing transcription start site analysis as well as transient reporter gene expression assays to locate the promoter of the CD72 gene.

We also demonstrate in this paper that alternative splicing is an important facet of CD72 polymorphism. The alternative usage of two 3′ splice sites at the intron 6/exon 7 junction site and the allele-specific usage of the distinct 3′ splice sites at the junction of intron 7/exon 8 both give rise to seven amino acids insertion/deletions within the distal extracellular region of the CD72 protein (residues 271 to 360), the region that is the most polymorphic among alleles with respect to amino acid substitutions (20). Alternative splicing often generates proteins varying in specific domains, thus allowing the fine modulation of protein function (26, 27). Because we have isolated cDNAs with or without these insertion/deletions and have expressed these as surface proteins in L cells (20), we believe that these various mRNA forms result in expressed protein in B cells. Whether these insertion/deletions modify CD72 function in some way is still unknown. We have identified these alternative mRNA forms in mouse splenic B cells. We do not yet know whether the alternative splicing events might be specific for distinct stages in B cell development or activation. Notably, the polymorphism generated by these alternative splicing events is localized to the most highly polymorphic region of the CD72 protein. We have previously shown that the membrane distal region (residues 271–361) of the extracellular domain contains a large number of amino acid substitutions, resulting in only 75 to 88% amino acid identity in pairwise comparisons among the $CD72^a$, $CD72^b$, and $CD72^c$ alleles. In contrast, there are 93 to 99% identical amino acids in such comparisons of the sequences of the cytoplasmic tail, transmembrane region, and membrane proximal region (residues 117–270) of the extracellular domain (20). Whether the high degree polymorphism of the membrane distal region of the extracellular domain plays any role in the function of CD72 is unknown.

In addition to alternative splicing at the 5′ ends of exons 7 and 8, we also show here that alternative splicing can lead to forms of CD72 mRNA lacking sequence from exons 3 and/or 4. The full-length CD72 mRNA and the mRNA species lacking sequence from exon 3 (encoding part of the cytoplasmic tail), exon 4 (encoding the transmembrane domain), or both were detected by PCR from total spleen RNA (data not shown) and RNA from purified splenic B cells (Fig. 8). Their distribution pattern is allele specific. The exon skipping in each case retains the protein coding frame (exon 3 is 72 bp and exon 4 is 90 bp). We have isolated cDNA clones missing the sequence from exon 3 or from both exon 3 and exon 4. When transfected into L cells, the exon 3^- cDNA generates a protein product that can be detected on the cell surface using immunofluorescence staining (W. H. Robinson and J. R. Parnes, unpublished observations). No surface expression is detected in L cells transfected with exon 3^-4^- cDNA (W. H. Robinson, J.-F. Chang, and J. R. Parnes, unpublished observations). We were also unable to detect CD72 protein in the supernatant of L cells transfected with the exon 3^-4^- cDNA (W. H. Robinson and J. R. Parnes, unpublished observations). However, the exon 3^-4^- mRNA can be detected by RT-PCR in L cells transfected with exon 3^-4^- cDNA (H. Ying and J. R. Parnes, unpublished results). CD72 lacks an amino-terminal signal peptide and relies on its transmembrane region (which is encoded by exon 4 and represents the only substantial hydrophobic stretch of sequence within the CD72 protein) to serve as a noncleaved signal peptide. Thus, it is unlikely that the exon 3^-4^- or exon 4^- forms of protein could be secreted or surface expressed. We do not yet know whether the latter two forms of mRNA encode intracellularly localized CD72 protein and what the function of such forms of CD72 might be.

Acknowledgments

We thank Dr. H. Neuman de Vegvar for valuable computer technical advice. We also thank Dr. J.-F. Chang for the preparation of purified splenic B cells and G. Frank for the preparation and characterization of some CD72 genomic subclones.

References

1. Sato, H., and E. A. Boyce. 1976. A new alloantigen expressed. selectively on B cells: the Lyb-2 system. *Immunogenetics 3:565.*
2. Shen, F., M. Spanodis, and E. A. Boyse. 1977. Multiple alleles of the Lyb-2 locus. *Immunogenetics 5:481.*
3. Tung, J., J. Michaelson, H. Sato, E. S. Vitetta, and E. A. Boyse. 1977. Properties of the Lyb-2 molecule. *Immunogenetics 5:485.*
4. Nakayama, E., I. von Hoegen, and J. R. Parnes. 1989. Sequence of the Lyb-2 B-cell differentiation antigen defines a gene superfamily of receptors with inverted membrane orientation. *Proc. Natl. Acad. Sci. USA 86:1352.*

5. Von Hoegen, I., E. Nakayama, and J. R. Parnes. 1990. Identification of a human protein homologous to the mouse Lyb-2 B cell differentiation antigen and sequence of the corresponding cDNA. *J. Immunol. 144: 4870.*
6. Yakura, H., F. Shen, E. A. Boyse, and L. Tang. 1980. The Lyb-2 phenotype of hemolytic PFC. *J. Immunol. 10:603.*
7. Grupp, S. A., J. A. Harmony, A. R. Baluyut, and B. Subbarao. 1987. Early events in B-cell activation: anti-Lyb-2, but not BSF-1, induces a phosphatidylinositol response in murine B cells. *Cell. Immunol. 110:131.*
8. Polla, B. S., J. Ohara, W. E. Paul, N. Nabavi, A. Myer, H.-C. Liou, F.-W. Shen, S. Gillis, J. V. Bonventre, and L. H. Glimcher. 1988. Differential induction of class II gene expression in murine pre-B-cell lines by B-cell stimulatory factor-1 and by antibodies to B-cell surface antigens. *J. Mol. Cell. Immunol. 3:363.*
9. Subbarao, B., J. Morris, and A. R. Baluyut. 1988. Properties of anti-Lyb-2-mediated B-cell activation and the relationship between Lyb-2 molecules and receptors for B-cell stimulatory factor-1 on murine Blymphocytes. *Cell. Immunol. 112:329.*
10. Subbarao, B., and D. E. Mosier. 1982. Lyb antigens and their role in B lymphocyte activation. *Immunol. Rev. 69:81.*
11. Subbarao, B., and D. E. Mosier. 1983. Induction of B lymphocyte proliferation by monoclonal anti-Lyb 2 antibody. *J. Immunol. 130: 2033.*
12. Subbarao, B., and F. Melchers. 1984. The action of an Lyb2.1-specific monoclonal antibody in soluble or immobilized form on resting and activated B cells. *Curr. Top. Microbiol. Immunol. 113:72.*
13. Subbarao, B., and D. E. Mosier. 1984. Activation of B lymphocytes by monovalent anti-Lyb-2 antibodies. *J. Exp. Med. 159:1796.*
14. Yakura, H., I. Kawabata, T. Ashida, F. W. Shen, and M. Katagiri. 1986. A role for Lyb-2 in B cell activation mediated by a B cell stimulatory factor. *J. Immunol. 137:1475.*
15. Yakura, H., F. W. Shen, M. Kaemmer, and E. A. Boyse. 1981. Lyb-2 system of mouse B cells: evidence for a role in the generation of antibody-forming cells. *J. Exp. Med. 153:129.*
16. Yakura, H., F. W. Shen, E. Bourcet, and E. A. Boyse. 1982. Evidence that Lyb-2 is critical to specific activation of B cells before they become responsive to T cell and other signals. *J. Exp. Med. 155:1309.*
17. Van de Velde, H., I. von Hoegen, W. Luo, J. R. Parnes, and K. Thielemans. 1991. The B-cell surface protein CD72/Lyb-2 is the ligand for CD5. *Nature 351:662.*
18. Luo, W., H. Van de Velde, I. von Hoegen, J. R. Parnes, and K. Thielemans. 1992. Ly-1 (CD5), a membrane glycoprotein of mouse T lymphocytes and a subset of B cells, is a natural ligand of the B cell surface protein Lyb-2 (CD72). *J. Immunol. 148:1630.*
19. Van de Velde, H., I. Von Hoegen, W. Luo, J. R. Parnes, and K. Thielemans. 1992. The T/B cell antigen, CD5, and the B-cell surface protein, CD72, form a pair of interacting receptors. *Ann. NY Acad. Sci. 651:443.*
20. Robinson, W. H., H. Ying, M. C. Miceli, and J. R. Parnes. 1992. Extensive polymorphism in the extracellular domain of the mouse B cell differentiation antigen Lyb-2/CD72. *J. Immunol. 149:880.*
21. Sambrook, J., E. F. Fritsch, and T. Maniatis. 1989. *Molecular Cloning*, 2nd Ed., Cold Spring Harbor Laboratory Press, Plainview, NY. p 2.108.
22. Sanger, F., S. Nicklen, and A. A. Coulson. 1977. DNA sequencing with chain-terminating inhibitors. *Proc. Natl. Acad. Sci. USA 74: 5463.*
23. Klemsz, M. J., S. R. McKercher, A. Celada, C. Van Beveren, and R. A. Maki. 1990. The macrophage and B cell-specific transcription factor PU.1 is related to the ets oncogene. *Cell 61:113.*
24. Proudfoot, N. J., and G. G. Brownlee. 1976. 3′ non-coding region sequences in eukaryotic messenger RNA. *Nature 263:211.*
25. McLauchlan, J., D. Gaffney, J. L. Whitton, and J. B. Clements. 1985. The consensus sequence YGTGTTYY located downstream from the AATAAA signal is required for efficient formation of mRNA 3′ termini. *Nucleic Acids Res. 13:1347.*
26. Smith, C. W., J. G. Patton, and B. Nadal-Ginard. 1989. Alternative splicing in the control of gene expression. *Ann. Rev. Genet. 23:527.*
27. McKeown, M. 1992. Alternative mRNA splicing. *Ann. Rev. Cell. Biol. 8:133.*