MOLECULAR AND CELLULAR BIOLOGY, Nov. 1988, p. 4889–4895
0270-7306/88/114889-07$02.00/0

Vol. 8, No. 11

Organization of the *Ly-5* Gene

YUMIKO SAGA,* JWU-SHENG TUNG,† FUNG-WIN SHEN,‡ THOMSON C. PANCOAST, AND EDWARD A. BOYSE

Memorial Sloan-Kettering Cancer Center, New York, New York 10021

Received 19 April 1988/Accepted 16 August 1988

A single *Ly-5* gene is known to generate a variety of transmembrane glycoprotein isoforms that distinguish various cell lineages and stages of differentiation within the hematopoietic developmental compartment of the mouse. Systems homologous to Ly-5 are known in rats and in humans. The complete exon-intron organization of the *Ly-5* gene is described in this report. The *Ly-5* gene occupies about 120 kilobases of chromosome 1 and comprises 34 exons, of which 32 (Ex-3 to Ex-34) are protein coding. Ex-1, Ex-2, and parts of Ex-3 and Ex-34 are untranslated. In all cDNA clones examined, either Ex-1 or Ex-2 was represented, but not both, implying that Ex-1 and Ex-2 in *Ly-5* mRNA may be mutually exclusive. Primer extension and S1 nuclease protection mapping were used to identify initiation (cap) sites for transcription. The finding of putative cap sites for Ex-1 and Ex-2, and of corresponding TATA-like sequences, suggests the presence of two promoters. In both Ex-1$^+$ and Ex-2$^+$ cDNA clones the next exon is Ex-3, which has a translation-initiating codon. The intron between Ex-3 and Ex-4 is unusually long, about 50 kilobases. Evidence is given that Ex-5, like Ex-6 and Ex-7 (studied previously), is another alternative exon that is selectively programmed, alone or together with Ex-6 or Ex-7 or both, to generate actual or potential Ly-5 isoforms by alternative splicing.

Interest in the mouse Ly-5 system (5, 12, 19) and its counterparts in the rat (L-CA system) (23, 28) and in humans (LCA/T200 system) (11, 21) centers particularly on the expression of these transmembrane glycoproteins by most or all hematopoietic cells as a range of molecular isoforms that typify these cells according to their lineage or stage of differentiation from the hematopoietic stem cell (12, 17, 26). It is known, and further confirmed in the present report, that the various glycoprotein isoforms which each of these systems comprises are generated from a single gene by selective usage of certain exons. Elucidation of the hematopoietic functions of these isoforms should be aided by detailed description of *Ly-5* genetic mechanisms. Attention has focused mainly on the smallest isoform, T200 (M_r 200,000; T lymphoid cells), and the largest, B220 (M_r 220,000; B lymphoid cells); several other Ly-5 isoforms of intermediate M_r are as yet less fully defined. Isoforms T200 and B220 differ in protein composition (25), resulting from use of particular exons in the generation of B220 but not T200 (15, 24). Investigations of Ly-5 function have so far been confined largely to properties of immune cells. For example, methods that involve blocking of function by Ly-5 antibody in vitro have been used to implicate Ly-5 in the lysis of target cells by natural killer cells (4, 10) and cytolytic T lymphocytes (6, 7) and in the action of macrophage factor on B lymphocytes during B-cell differentiation leading to antibody production (29). These functional studies cannot be fully reviewed here, but suffice it to say that they are consistent with the involvement of Ly-5 in cellular interregulation of the signal-receptor kind. Phylogenetic considerations may suggest a similar role for the pertinent isoforms in properties of nonlymphoid hematopoietic cell lineages.

* Corresponding author.
† Present address: Merck Sharpe & Dohme Research Laboratories, Rahway, NJ 07065.
‡ Present address: Showa University Research Institute for Biomedicine in Florida, St Petersburg, FL 33716.

MATERIALS AND METHODS

Screening of a genomic library for the *Ly-5* gene. A genomic library which had been made for other purposes from cells of the (B6 × A)F_1 T-cell leukemia variant EARAD1-β2M$^-$ in EMBL4 vector (14) was screened with a 4.7-kilobase (kb) *Xho*I insert of the *Ly-5* cDNA clone pLy-5-68 (18).

Sequencing of DNA. Genomic DNA fragments containing exon sequences detected by Southern blotting with cDNA probes were subcloned in vectors M13mp18 and M13mp19 and sequenced by the method of Sanger et al. (16).

Preparation of RNA. Total RNA was prepared from the cell lines mentioned herein by using guanidinium isothiocyanate followed by centrifugation through a cushion of CsCl (14). Poly(A)$^+$ RNA was then isolated by oligo(dT)-cellulose chromatography.

S1 nuclease protection mapping. Probes for S1 mapping were prepared from genomic DNA clones. End-labeled probe DNA (2×10^5 cpm) was hybridized with total RNA (30 μg) at various temperatures and digested with S1 nuclease (at 37°C for 30 min) (14). Protected fragments were analyzed on an 8% polyacrylamide–7 M urea sequencing gel.

Primer extension. Synthetic oligonucleotide primer, end labeled with polynucleotide kinase, was coprecipitated with 2 μg of poly(A)$^+$ RNA or 30 μg of total RNA, dissolved in 5.5 μl of 10 mM Tris hydrochloride (pH 8.0)–1 mM EDTA plus 4.5 μl of 2 M KCl, heated at 65°C for 5 min, and then hybridized at 50°C for 1 h. The reaction mixture was adjusted to 60 μl, containing 50 mM Tris (pH 8.3), 140 mM KCl, 10 mM $MgCl_2$, 5 mM dithiothreitol, actinomycin D (50 μg/ml), bovine serum albumin (100 μg/ml), 0.16 mM deoxynucleoside triphosphates, and 10 U of reverse transcriptase, and incubated at 42°C for 1 h. The reaction was stopped by addition of EDTA (to 20 mM) followed by extraction with phenol-chloroform and precipitation with ethanol. The final pellet was dissolved in formamide dye and electrophoresed on sequencing gel.

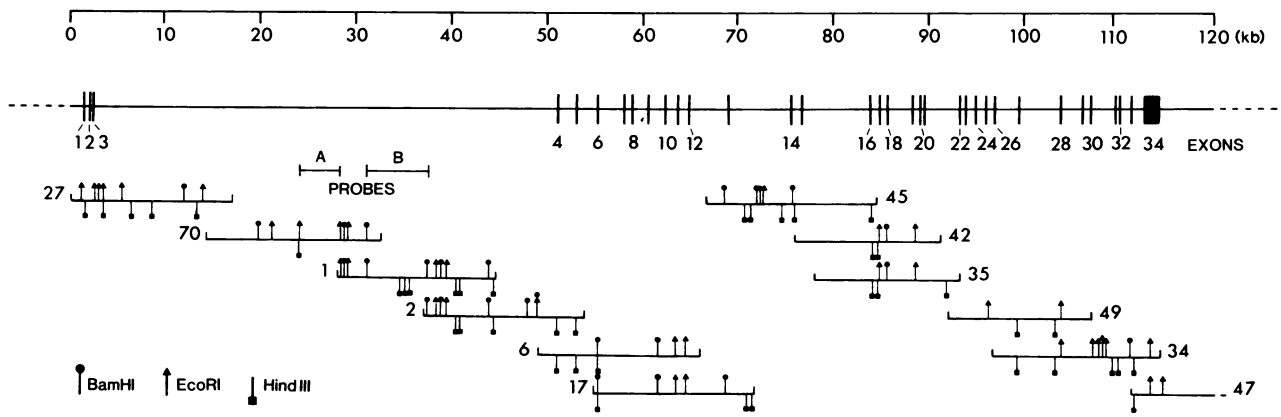

FIG. 1. Molecular map of *Ly-5* and restriction maps of recombinant bacteriophages. DNA fragments A and B, representing parts of the long intron 3 (between Ex-3 and Ex-4) and prepared from clones 70 and 1, respectively, provided probes for Southern blotting (see Fig. 4).

RESULTS

Cloning of the complete *Ly-5* gene. The genomic library constructed in EMBL4 was screened with a 4.7-kb insert of the T-cell cDNA clone pLy-5-68 (18). Six overlapping clones (clones 45, 42, 35, 49, 34, and 47), spanning ca. 60 kb of genomic DNA, were isolated and subjected to restriction mapping (Fig. 1). These six clones contained all of the sequences represented in the T-cell *Ly-5* cDNA clone pLy-5-T4 (15), which is the longest cDNA clone available, except for the initial 5′ 270 base pairs (bp). We therefore began gene walking, with probes derived from genomic DNA. Clones 17, 6, and 2 were thus isolated, but the initial 170 bp representing pLy-5-T4 was still not found. Accordingly, Southern blotting was used to identify genomic DNA fragments containing the missing initial 5′ region. Cellular DNA was digested with *Eco*RI or *Hin*dIII, separated on an agarose gel, and Southern blotted. The blots were hybridized with a 5′ cDNA probe representing the 170-bp region. Hybridizing bands of ca. 1.6 kb (*Eco*RI) and ca. 2.1 kb (*Hin*dIII) were identified. Genomic fragments in these size ranges were enriched by fractionation in an agarose gel and ligated with a pUC vector to make sublibraries. By screening with the 5′ cDNA probe, we isolated a clone (H2) which contains a 2.1-kb *Hin*dIII fragment. This insert was then used to screen the original genomic library and yielded clone 27, but clone 27 did not overlap clone 2. After walking from the 3′ end of clone 27 and the 5′ end of clone 2, we finally obtained clones 1 and 70, which overlap. The total length of the *Ly-5* gene is ca. 120 kb.

Exon-intron organization. The exon-intron organization of the *Ly-5* gene was determined by Southern blotting with T-cell and B-cell cDNA probes and subsequent sequencing (Fig. 1 and 2). Exon-intron boundaries for the first eight exons (Ex-1 to Ex-8) have already been reported (15). Although the first two exons (Ex-1 and Ex-2, untranslated), as well as Ex-3, which includes a signal peptide sequence, are clustered within a region of ca. 1 kb, the next exon (Ex-4), encoding the seven N-terminal amino acids of the mature polypeptide, is located after an unusually long intron of ca. 50 kb. The next three exons (Ex-5, Ex-6, and Ex-7) are not seen in the T-cell cDNA clone pLy-5-T4 because they are not represented in isoform T200 of T cells but are used for the larger isoform B220 of B cells and probably for some T-cell isoforms (1, 21). Ex-5 was not identified previously, because our B-cell cDNA clones (pLy-5-B11 and pLy-5-B15), derived by primer extension, did not contain Ex-5 (15). However, using the B-cell cDNA clone, mLC-1, constructed by Thomas et al. (24), we found Ex-5 in the expected region between Ex-4 and Ex-6.

Northern (RNA) blotting confirmed the selective usage of Ex-5 in B cells. The presence of Ex-5 accords with expression of isoform B220 (I.29 and 70Z-3 B cells) but not isoform T200 (ISL-57 and ASL1 T cells), as is the case with Ex-6 and Ex-7 (15) (Fig. 3). According to alternative usage of the three exons Ex-5, Ex-6, and Ex-7, eight different isoforms may exist, and all of these have been observed as transcripts in at least one of the mouse, rat, and human systems (1, 13, 15, 21, 24; Saga, unpublished observations). Because of these extra exons (Ex-5, Ex-6, and Ex-7), the extracellular region of B220 protein is 139 amino acids (aa) larger (541 aa) than the extracellular region of T200 protein (402 aa). The detailed structures of B220 and T200 deduced from the DNA sequence and exon organization are given elsewhere (J. S. Tung, Y. Saga, and E. A. Boyse, Immunogenetics, in press).

The Ly-5 transmembrane region (22 aa) relates to the single exon Ex-17. Ex-18 to Ex-34 encode an unusually extensive cytoplasmic region (705 aa).

The 34 exons vary in size, but most are between 50 and 200 bp in length. The final exon (Ex-34) is the largest and has a termination codon and 3′ untranslated region.

Splice junctions follow the AG/GT rule. Intron sizes are notably varied. The smallest is 81 bp, which is consistent with an estimated minimal functional size for effective splicing (27). The largest intron identified (ca. 50 kb) lies between Ex-3 (signal peptide) and Ex-4 (N terminal). The sizes of other introns range from 0.6 to 6.6 kb (average, ca. 2.1 kb).

Because the original genomic library had been prepared, for other purposes, from a T-cell-variant leukemia cell line (EARAD1-β2M$^-$), Southern blotting was used to determine whether the unusually long intron 3 (between Ex-3 and Ex-4) is truly a normal feature of the *Ly-5* gene, as distinct from being a peculiarity of EARAD1-β2M$^-$ cells or a cloning artifact. For this purpose, we prepared probes A and B from genomic DNA clones 70 and 1 (Fig. 1) and hybridized them with DNA of B6 or SJL/J mouse liver cells (see below), T cells (ASL1), B cells (70Z-3), and EARAD1-β2M$^-$ cells. The same expected sizes and numbers of bands were uniformly observed for DNA from all these sources, except for SJL/J liver cells (Fig. 4). Thus, the large size of intron 3 is a normal feature of the *Ly-5* gene.

The distinctive *Bam*HI band for SJL/J liver, 8 kb versus 6 kb for B6 liver, confirms that the genomic DNA clones

	Consensus sequence (Acceptor) TTTTTTTTTTTTTTTTNTAG / G CCCCCCCCCCCCCCCC C		Consensus sequence (Donor) CA G / GTAAGT A G	Total size (bp) Exon	Intron
	-50	-20 -1			
Ex-1		5' UT---TGG	GTAAGA aaattattgatgac	-114	
Ex-2		5' UT...AAA	GTATGC gttcttttgtttta	-84	81
Ex-3	gctttacagagacaaacttcaagagagata	ACCATTATTTTGCCTTTCAG GGAGA.....ATG	GG...ACA G GTAAGC	acacccatattcat	112 -50 k
Ex-4	gtatgctttaaatctcaaaatttctaaaag	TGTGTTTATTTTTTTTTCAG	GG CAA ACA CCT AC...GAT G GTAAGA	attaatatttgact	24 -4 k
Ex-5	gggacatgattgatatattccttctcttct	TCCATCCATTAACATAACAG	AA CTG AGC ACA AC...CCA G GTGACC	actaagcctggccc	129 -3 k
Ex-6	ctgccagccacaggtgtgatcaacataacc	ATCTCACTTTCTAACCTCAG	GT GTG TTA TCC AC...CCA G GTTTGG	gggctctttagaca	147 -2.2 k
Ex-7	cagcatgtggcatgatggagtgatagtgtt	GGTGACCCTTCTGATTGCAG	GT GTG CCA GGG GA...TCA G GTGTGA	ccattatttagcat	141 720
Ex-8	gagagaatgaaactaattaacttttaccgt	TTTTGTTCATTTCCTTGCAG	GT GCC AGC CTC AC...ATA G GTGACA	atcttaccctcagg	72 1307
Ex-9	aaaataatttctcaatgattttcctctttt	TTTTTTAATCTCAATTACAG	CT ACC ACA ACG AA...TGT G GTAAGT	ttgtcgacccagca	24 1839
Ex-10	cattcattctcgatgatgattgtactaatt	TCCTTTTTTTTTTTTTTCCAG	CT GCC ATG TTT GG...CCA G GTGAAT	gtcaatttctctct	216 -1 k
Ex-11	ctccacaccaggaaccatcacctaagagca	CCTCTCCGTTTCCTCCACAG	GG ACT GAC AAG TT...CCA G GTACGA	tgctgggcagagaa	141 -1 k
Ex-12	atttactatgtggcatttcatatactatgt	GCAATACTTACTTCTTTCAG	AA AAT AAT ACA AA...GGG A GTAAGT	atatcgtttatgtt	159 -3.5 k
Ex-13	tcttttcagtgtcttatttttcccattata	ACTTTTTCTGTTTTTAACAG	TT CCA GAS ACG CC...TCA G GTAATG	taaaattccactag	135 -6.6 k
Ex-14	gatcctgactttcagaaacttactaactca	GGCTTTCTTTTTAAATCTAG	AA AAA TGT AAA AG...GAT C GTAAGT	ttttggctttaata	156 -1.1 k
Ex-15	aaactttacatgtcaaaatattaagataac	AAATGTCTCTTATTTTATAG	GT CCG GAC AAG GT...T CTG GTAAGG	tcatgtgttcctta	209 -6.5 k
Ex-16	cacatatagtagaaaaccaaattcttttct	TTTTCTTAACTGAAATTTAG	GTC TCT TTT CAC A...AAT A GTAAGT	catcctggcctatt	61 -0.9 k
Ex-17	gcaaagtatttgaaataataagctcttttt	GTGTCACTTTCTATTTCTAG	TT AAT GCT AAA GC...CC AG GTAAGA	gacagttaattctg	109 -0.8 k
Ex-18	ttgcaaagacagaaactttttattttatttt	ATTTTATTTTATTTTCACAG	C AAT TTA GAT GAA...AGG G GTGAGT	atgaataattttgg	35 -2.6 k
Ex-19	tggacagacagacagatgaaacacaggtct	CACCTCTAACTTGTTTACAG	AT GAT GAA AAG CA...T CAG GTATGA	gttgtccttgacag	110 -0.8 k
Ex-20	tccttctttttgggaatcttctgtaacccaa	AGAATATTTGTATCTTTCAG	AGC ATT CCA CGG G...CCC T GTGAGT	ggactttggaactg	91 85
Ex-21	accgtatggctactctttagtttagtatta	ATTTCTTTTTCTTTAATCAG	AT GAT TAT AAC CG...T GAT GTAAGT	aatgtatgtgctgg	77 -5 k
Ex-22	aatgattaagtgttccataaatgactgttc	TCTTTGTGTGTTTTGATCAG	GGC TTC AAG GAA C...CAA G GTAATT	cattttacatcttg	37 87
Ex-23	ttgttttatagcacttctcgaggattggat	AACTTGTACTTAACTCCTAG	GG CCC CAT GAG AC...C AGG GTAAGA	atcccagaaggctg	98 -1.2 k
Ex-24	gaactgtattacaatagactagagtcatta	ATGATTCTTTTAAACTGTAG	AAC AAG TGC GCA G...A CAT GTGAGT	ttgtctcattctat	126 -1 k
Ex-25	agtttgttttgaaagaaactttctcaaaga	ATGCCTTCTCTCTTGAGTAG	AAA AAA GAA AAA G...GC AG GTAAAG	actctgtagggcaa	158 -1 k
Ex-26	ctgagacatctggtatacctcacaccttac	TTTTCTGTTTGTTCCCTTAG	T GCT GGT GTT GGG...G GAG GTACAG	tcttagtttatgtg	136 -1.7 k
Ex-27	atagggctctgcaatacattttagttatgt	GTGTTGTTTGATTTTTGCAG	GCA CAG TAT ATC C...C CAG GTAGCA	atgggtcacagcaa	150 -4 k
Ex-28	tttaggaaatattggatttttaaaggaggct	TTCTTCTTTATTTTTCTTAG	AGA CTT CCA TTC A...CCA T GTATGT	acctgacctttctt	91 -1.5 k
Ex-29	tcctcagtccttaggacttctttggtaaga	CTTCGTTTCTTGTTCCATAG	AT GAC TTT AAC AG...G ATG GTAGGT	acccacactccaaa	134 -0.6 k
Ex-30	atctttctttgatggtagcactaacaaatt	CCATCTTTGTGATTTTTTAG	AGT TAC TGG AAA C...C CAG GTTTGT	gctttgaagaatgc	135 -2.3 k
Ex-31	aagacaacacctaggtaaaacatggcaagt	CATGTTATTTTTTTATCTAG	GAA GTC TGT GCT C...C AAG GTAGGT	aaacactcaaaaag	123 250
Ex-32	actggaccgaagttcatttgaactaatgtt	CACTTCACCTCGGATTGCAG	AGG AAG GAG CCC C...GC AG GTTAGG	aagctgatggggtg	176 -1.5 k
Ex-33	tgagttaattcatgttctacttttaaaaa	ATCCCTTTTTCTTTTGACAG	A GAT GGA TCC CAG...C TAT GTAAGT	gtccctcgatgtca	136 -1.2 k
Ex-34	tgggtgtaagaacgttctctatgtaagttc	ATGCTGCTTTCTTTCACCAG	GAG...TAG..ATAATAAACTTTC- -AACTATTATCTTC (TTTGCTTATGTG...)		1364

FIG. 2. Exon-intron boundaries.

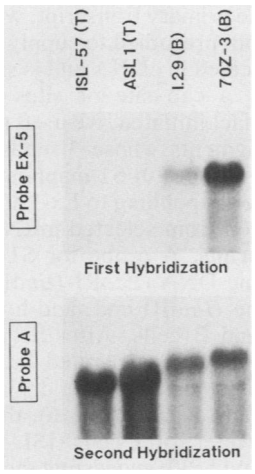

FIG. 3. Northern blots illustrating selective expression of Ex-5 in B-cell mRNA but not T-cell mRNA. The genomic probe containing Ex-5 (0.6-kb *Hin*dIII-*Pvu*II fragment subcloned from genomic clone 6 [Fig. 1]) is seen to hybridize with mRNA of B cells but not of T cells (upper panel). After removal of probe Ex-5, probe A (2.3-kb *Bam*HI fragment of cDNA clone pLy-5-T4 representing the common exons Ex-13 to Ex-32) is seen to hybridize with mRNA of both T and B cells (lower panel).

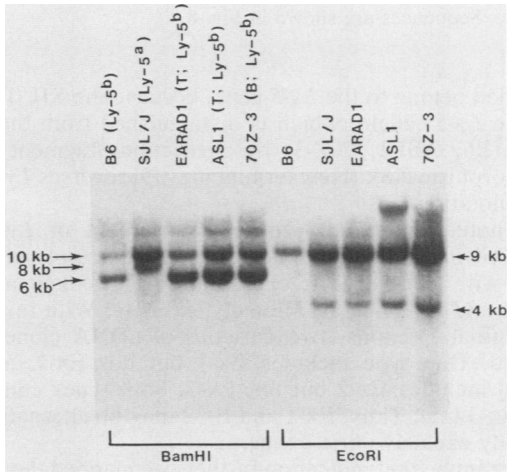

FIG. 4. Southern blots indicating that the long intron 3 (between Ex-3 and Ex-4) is a normal feature of the *Ly-5* gene. Liver DNA was obtained from B6 (*Ly-5*^{b}) and SJL/J (*Ly-5*^{a}) mice. The locations of probes A and B, used in combination for this hybridization, are indicated in Fig. 1. The *Bam*HI restriction fragment length polymorphism distinction, 6 kb (*Ly-5*^{b}) versus 8 kb (*Ly-5*^{a}), signifies that the genomic clones from which probes A and B were derived belong to the *Ly-5* gene.

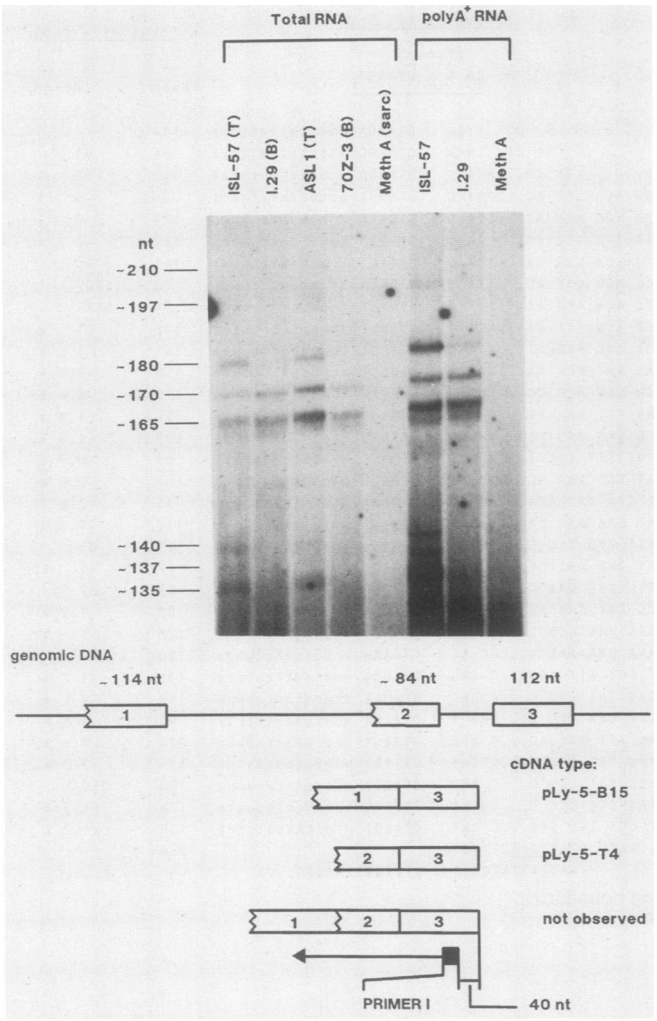

FIG. 5. Primer extension data suggesting that usage of Ex-1 or Ex-2 is mutually exclusive. The largest band (ca. 210 nt) is still too small for a transcript containing both Ex-1 and Ex-2 (>250 nt needed). Sequences are shown in Fig. 8.

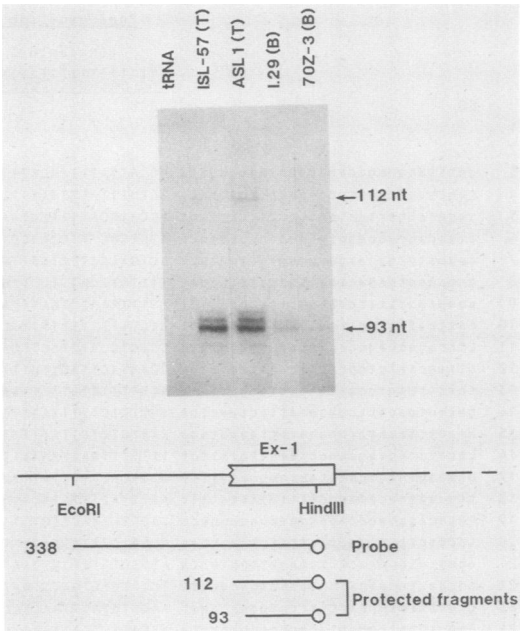

FIG. 6. S1 nuclease protection mapping suggesting two cap sites at which transcription from Ex-1 may be initiated. The two protected fragments (112 and 93 nt) are far more intense with T cells than with B cells; the 112-nt band of B cells is scarcely visible. These two bands correspond to products of ca. 197 and ca. 180 nt, respectively, extended from primer I (Fig. 5). The end-labeled (○) probe and the protected fragments are diagramed. Sequences are given in Fig. 8.

identified belong to the *Ly-5* gene, because the SJL/J mouse has the *Ly-5*^{a} allele, which is distinguished from the *Ly-5*^{b} allele (B6, ASL1, 70Z-3) by restriction fragment length polymorphism (18). (New terminology [9] reverses *Ly-5*^{a} and *Ly-5*^{b} notation.)

As noted before (18), no rearrangement of *Ly-5* was observed in comparing DNA of Ly-5 nonexpressor origin (liver) with DNA of Ly-5 expressor origin (leukemia cells).

Ex-1 and Ex-2 may be mutually exclusive. With respect to their initial 5′ exons, two varieties of cDNA clones were isolated. One type includes Ex-1 but not Ex-2, and the second includes Ex-2 but not Ex-1, both types continuing with Ex-3 (15). Thus, Ex-1 and Ex-2 may be alternative and mutually exclusive first exons.

To examine that indication further, we mapped the 5′ ends of *Ly-5* transcripts by primer extension. A primer relating to Ex-3 (Fig. 5, primer I) was prepared and hybridized with mRNA or total RNA from T cells (ISL-57 and ASL1) and B cells (70Z-3 and I.29). After reverse transcription, primer-extended products were separated in a sequencing gel (Fig. 5). Three common bands were observed for all four cell populations, and T cells yielded additional bands (discussed below) not seen with B cells. The largest band, ca. 210 nucleotides, is smaller than would be expected if Ex-1, Ex-2, and Ex-3 were all used, because that would require at least 250 nucleotides. Therefore, we conclude that Ex-1 and Ex-2, at least in these cell lines, are mutually exclusive.

Cap sites and promoters. Processed transcripts having either Ex-1 or Ex-2 might arise from one primary transcript, by alternative splicing, or from two distinct primary transcripts. For a single primary transcript, we would expect one more common exon, presumed to supply a donor splice site, which was not detected in cDNA and is situated upstream of Ex-1. To identify a cap site or sites to indicate where transcription is in fact initiated, we used two methods. First, genomic DNA fragments whose 3′ ends are in Ex-1 or in Ex-2 were used as probes in S1 mapping. Second, oligonucleotide primers corresponding to Ex-1 and Ex-2 were tested for primer extension from selected mRNAs.

(i) Ex-1$^+$ transcripts. A probe for S1 mapping was prepared from genomic DNA (*Eco*RI-*Hin*dIII; 338 nucleotides [nt]), labeled at the *Hin*dIII end, and hybridized with total RNA of T cells and B cells. After S1 nuclease digestion, protected fragments were separated on a sequencing gel. Two protected fragments, of 112 and 93 nt, were observed with both T cells and B cells (Fig. 6), the intensity of both bands being far greater for T cells (ISL-57 and ASL1) than for B cells (707-3 and I.29), suggesting that the Ex-1$^+$ species of mRNA is much more abundant in T cells.

The contiguous genomic sequences upstream of the 93- and 112-nt protected fragments do not resemble an acceptor splice site (see Fig. 8), because they are not significantly rich in pyrimidine (C+T), which is a characteristic feature of 3′ splice acceptor regions (2, 3).

The conclusion that the upstream regions in question are

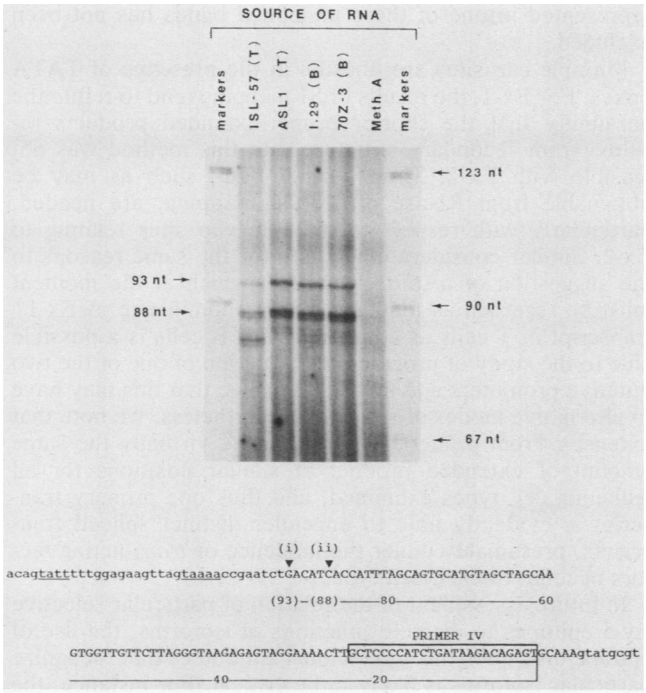

FIG. 7. Primer extension data suggesting two cap sites (i and ii) at which transcription from Ex-2 may be initiated. Results are similar for all T- and B-cell lines shown. Two major products (93 and 88 nt) extended from primer IV correspond to products of ca. 170 and ca. 165 nt, respectively, extended from primer I (Fig. 5). TATA-like sequences, TAAA and TATT, are observed upstream from indicated potential cap sites. A more extensive sequence is given in Fig. 8.

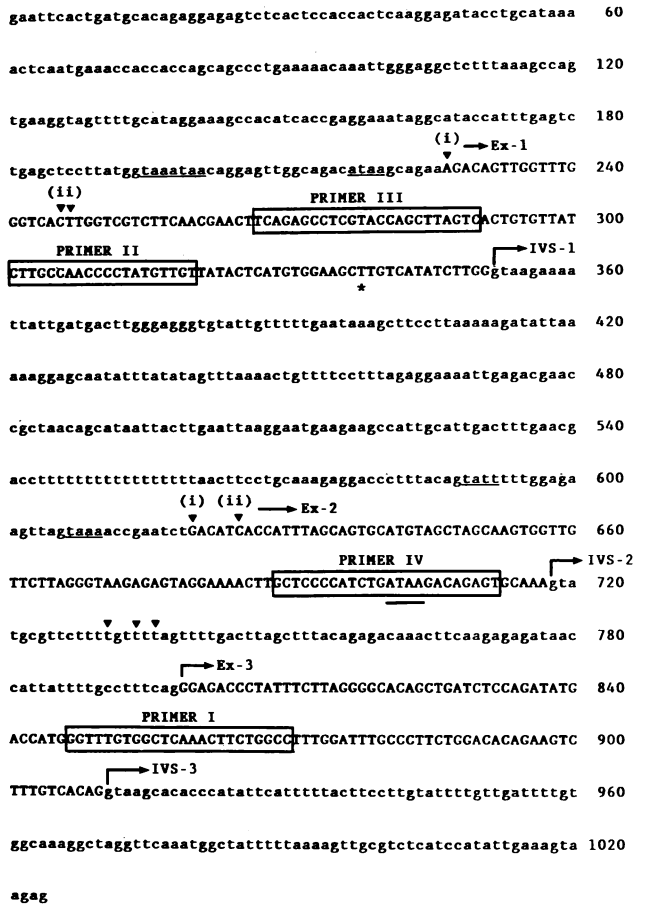

FIG. 8. Genomic DNA sequence including Ex-1, Ex-2, and Ex-3. Interpretations of Fig. 5 to 7 are illustrated here. Putative cap sites for Ex-1 (i and ii), Ex-2 (i and ii), and Ex-3 are indicated (▽). Corresponding TATA-like sequences are underlined. The 3′ end of the S1 probe is indicated by an asterisk.

more likely to be involved in initiation of transcription than in splicing was also supported by primer extension, from primers II and III (see Fig. 8), which yielded products corresponding to the protected fragments of 112 and 93 nt observed in S1 mapping (Fig. 6). The presence of two cap sites is also suggested by extension from primer I of Ex-3 (Fig. 5). The bands of ca. 197 and ca. 180 nt in Fig. 5 are consistent with initiation of Ex-1$^+$ transcription at potential Ex-1 cap sites i and ii, respectively (see Fig. 8).

Sequences upstream of these two potential Ex-1 cap sites were searched for characteristic promoter sequences (CAAT and TATA). TAAATAA and ATAA sequences were found about 30 nt upstream from each potential cap site. Although no typical CAAT sequence was found, several similar sequences are observed within 200 nt upstream from the putative cap sites (see Fig. 8). The location of TATA-like sequences (about 30 nt upstream) is appropriate for the position of a TATA box as a eucaryotic promoter (8).

(ii) Ex-2$^+$ transcripts. The above approaches could also be applied to Ex-2, but since no restriction site appropriate to S1 mapping was available for Ex-2, only primer extension, with primer IV, was feasible. Results similar to these for Ex-1 were obtained with all cells tested, again featuring two main extended products, of 93 and 88 nt (Fig. 7), representing two potential cap sites. Since the longest product was only 14 nt longer than the longest cDNA clone so far obtained (24), it is unlikely that there can be an additional unidentified upstream exon containing a cap site. These two bands presumably correspond to the ca. 170- and ca. 165-nt products extended from primer I of Ex-3 (Fig. 5) and represent putative cap sites i and ii of Ex-2 (Fig. 7 and 8). TATA-like sequences are observed, TAAA at 13 and 18 nt upstream, and TATT at 31 and 36 nt upstream, from the potential cap sites (Fig. 7).

Clearly, two putative promoters could account for alternative initiation of transcription from Ex-1 and Ex-2, respectively.

DISCUSSION

Study of Ly-5 isoformism has so far been founded mainly on the lymphocytic branch of hematopoiesis, leaving all the other Ly-5$^+$ lineages of blood cells relatively unexplored. The hypothesis that Ly-5 isoforms participate in intercellular regulation involving the many functionally diversified cell lineages derived from the hematopoietic stem cell, which is the first ontogenetic appearance of Ly-5 expression, can be viewed in phylogenetic perspective. We envisage an origin of *Ly-5* far back in evolution in a primordial hematopoietic cell and rudimentary function and as yet only elementary potential for successive program diversification. In such a context, the evolution of Ly-5 isoforms from a single *Ly-5* gene can be viewed as a device to propagate and compartmentalize a primitive Ly-5 function among the evolving incipient branches of hematopoiesis, comprising an increasing variety of progenitor and terminal cell types and controlling cytokines. In that view, the constitutive exons (consti-

tutive in the sense of being common to all Ly-5 isoforms) may represent conserved essential Ly-5 function, and the optional exons (used for some, but not all, isoforms) would supply the basis for a range of distinctive but allied functions typifying different hematopoietic cell lineages.

As noted above, varied usage of the three optional exons so far identified (Ex-5, Ex-6, and Ex-7), representing an outer part of the extracellular domain of Ly-5 glycoprotein, provides eight possible combinations; evidence for all of these is present in one or another of the mouse Ly-5 and homologous rat and human systems. Alternative exon usage may suffice to account for all isoform variation; however, it may not do so, particularly since the isoforms of nonlymphoid hematopoietic lineages remain to be adequately defined and because some Ly-5 isoforms might be distinguished by features other than Ly-5 protein composition, such as differential glycosylation.

Present data do not preclude the existence of further optional exons; just one of these would greatly amplify the potentiality of combinatorial variation. As noted, a 10-kb section of the 120-kb *Ly-5* gene is the site of the three optional exons Ex-5, Ex-6, and Ex-7 and is the first place in which to look for optional exons that may have escaped detection, because the mouse, rat, and human cDNA libraries searched so far have come mostly from lymphoid cell lines. Other lineages may use other exons. In fact, S1 mapping gives some slender evidence of a further optional sequence situated between Ex-4 and Ex-8 and used by Pu-5 macrophages (14) but evidently not Ex-5, Ex-6, or Ex-7, because probes for these three exons have shown no relevant hybridization in Northern blotting of Pu-5 cells (Saga, unpublished observations). It is an open question whether nonlymphoid Ly-5$^+$ cell lineages, exemplified by Pu-5, may use exons not used by lymphoid lineages and encoded between Ex-4 and Ex-8.

Little can be said about how the known optional exons (Ex-5, Ex-6, and Ex-7) are differentially employed by alternative splicing. Comparison of donor and acceptor sites reveals no obvious difference between optional and constitutive exons, but it may be significant that the C+T content of regions upstream of all three optional exons is low in comparison with the constitutive exons. For isoforms generated from a single primary transcript, a requirement for a *trans*-acting factor is implied.

The present data suggest the presence of two Ly-5 promoters, one upstream from Ex-1 and the other upstream from Ex-2, yielding the observed Ex-1$^+$ and Ex-2$^+$ alternative transcripts. A question then arises regarding whether all transcripts indicated by primer extension from Ex-3 (common to all described transcripts) can be similarly explained. Recapitulating the data in Fig. 5, Ex-1$^+$ transcript was much more abundant in T cells than in B cells, and bands of 197 and 180 nt were consistent with initiation of transcription at potential cap site i or ii, respectively. (The largest product [210-nt band] was very faint, and whether it represents a cap site is quite uncertain.) Bands of 170 and 165 nt accorded with putative cap sites i and ii of Ex-2. However, the remaining bands (140, 137, and 135 nt of T cells but not B cells) had no corresponding bands among Ex-1$^+$ or Ex-2$^+$ transcripts (Fig. 6 and 7). One explanation is that a third promoter allows initiation of transcription 135 to 140 nt upstream from primer I and accounts for an Ex-1$^-$ Ex-2$^-$ transcript typical of T cells; a TATA-like sequence was observed ca. 30 nt upstream of putative cap sites (Fig. 8). However, the possibility that an additional 5′ exon is represented in one of these undefined bands has not been excluded.

Multiple cap sites are unusual in the presence of TATA boxes. For Ex-1, the results of S1 mapping tend to refute the possibility that the shorter primer-extended products resulted from secondary structure, but this method was not feasible with Ex-2. Thus, further data, such as may be obtainable from RNase protection mapping, are needed, particularly with respect to multiple cap sites relating to Ex-2. Similar considerations apply for the same reasons to the suggestion of a third promoter, which at the moment must be regarded as tenuous. The predominance of Ex-1$^+$ transcript in T cells as compared with B cells is a possible clue to the study of programmed selection of one of the two putative promoters and to any relevance that this may have to alternative modes of splicing. Nevertheless, we note that extension from primer IV of Ex-2 gave virtually the same amount of extended product at similar positions for all leukemia cell types examined, and thus one primary transcript is evidently able to engender distinct spliced transcripts, presumably under the influence of *trans*-acting factors peculiar to each cell type.

In future assessment of the relation of particular selective Ly-5 epitopes to discrete functions of isoforms, the use of epitope-discriminating monoclonal antibodies that recognize particular isoforms is likely to be critical. For instance, the monoclonal antibody OX-22, which distinguishes two functionally discrete sets of T cells and can inhibit graft-versus-host reactions, may recognize an Ex-6-encoded Ly-5 peptide (1, 20). Also, in the human LCA/T200 system, the monoclonal antibody 2H4, which divides the T4$^+$ T-cell set into 2H4$^+$ suppressor inducer and 2H4$^-$ helper subsets, may recognize an epitope encoded by the Ex-5 region (22). In this manner, the prospect of relating particular *Ly-5* exons to particular epitopes and domains, and hence to particular isoforms, is encouraging.

ACKNOWLEDGMENTS

We thank Matthew Thomas for generously providing the B-cell *Ly-5* cDNA clone mLC-1.

This work was supported in part by Public Health Service grants CA-39827 and AI-21840 from the National Institutes of Health. E.A.B. is American Cancer Society Research Professor of Cell Surface Immunogenetics.

LITERATURE CITED

1. **Barclay, A. N., D. I. Jackson, A. C. Willis, and A. F. Williams.** 1987. Lymphocyte specific heterogeneity in the rat leucocyte common antigen (T200) is due to differences in polypeptide sequences near the NH_2-terminus. EMBO J. **6**:1259–1264.
2. **Breathnach, R., and P. Chambon.** 1981. Organization and expression of eucaryotic split genes coding for proteins. Annu. Rev. Biochem. **50**:349–383.
3. **Green, M. R.** 1986. Pre-mRNA splicing. Annu. Rev. Genet. **20**: 671–708.
4. **Kasai, M., J. C. Leclerc, F. W. Shen, and H. Cantor.** 1979. Identification of Ly-5 on the surface of 'natural killer' cells in normal and athymic inbred mouse strains. Immunogenetics **8**: 153–159.
5. **Komuro, K., K. Itakura, E. A. Boyse, and M. John.** 1975. Ly-5: a new T-lymphocyte antigen system. Immunogenetics **1**:452–456.
6. **Lefrancois, L., and M. J. Bevan.** 1985. Functional modifications of cytotoxic T-lymphocyte T200 glycoprotein recognized by monoclonal antibodies. Nature (London) **314**:449–452.
7. **Lefrancois, L., L. Puddington, C. E. Machamer, and M. J. Bevan.** 1985. Acquisition of cytotoxic T lymphocyte-specific carbohydrate differentiation antigens. J. Exp. Med. **162**:1275–

1293.

8. **Maniatis, T., S. Goodbourn, and J. A. Fisher.** 1987. Regulation of inducible and tissue-specific gene expression. Science **236:** 1237–1245.
9. **Morse, H. C., III, F. W. Shen, and U. Hammerling.** 1987. Genetic nomenclature for loci controlling mouse lymphocyte antigens. Immunogenetics **25:**71–78.
10. **Newman, W.** 1982. Selective blockade of human natural killer cells by a monoclonal antibody. Proc. Natl. Acad. Sci. USA **79:** 3858–3862.
11. **Omary, M. B., I. S. Trowbridge, and H. A. Battifora.** 1980. Human homologue of murine T200 glycoprotein. J. Exp. Med. **152:**842–852.
12. **Omary, M. B., I. S. Trowbridge, and M. P. Scheid.** 1980. T200 cell surface glycoprotein of the mouse: polymorphism defined by the Ly-5 system of alloantigens. J. Exp. Med. **151:**1311–1316.
13. **Ralph, S. J., M. L. Thomas, C. C. Morton, and I. S. Trowbridge.** 1987. Structural variants of human T200 glycoprotein (leukocyte-common antigen). EMBO J. **6:**1251–1257.
14. **Saga, Y., J. S. Tung, F. W. Shen, and E. A. Boyse.** 1986. Sequences of Ly-5 cDNA: isoform-related diversity of Ly-5 mRNA. Proc. Natl. Acad. Sci. USA **83:**6940–6944. (Erratum, **84:**1991, 1987.)
15. **Saga, Y., J. S. Tung, F. W. Shen, and E. A. Boyse.** 1987. Alternative use of 5′ exons in the specifiation of Ly-5 isoforms distinguishing hematopoietic cell lineages. Proc. Natl. Acad. Sci. USA **84:**5364–5368.
16. **Sanger, F., S. Nicklen, and A. R. Coulson.** 1977. DNA sequencing with chain-terminating inhibitors. Proc. Natl. Acad. Sci. USA **74:**5463–5467.
17. **Scheid, M. P., P. W. Kincard, and T. M. Dexter.** 1982. A multipotential hematopoietic cell line, p. 145–148. *In* E. Vitetta and C. F. Fox (ed.), B and T cell tumors: biological and clinical aspects, vol. 14. Academic Press, Inc., New York.
18. **Shen, F. W., Y. Saga, G. Litman, G. Freeman, J. S. Tung, H. Cantor, and E. A. Boyse.** 1985. Cloning of Ly-5 cDNA. Proc. Natl. Acad. Sci. USA **82:**7360–7363.
19. **Shen, F. W., J. S. Tung, and E. A. Boyse.** 1986. Further definition of the Ly-5 system. Immunogenetics **24:**146–149.
20. **Spickett, G. P., M. R. Brandon, D. W. Mason, A. F. Williams, and G. R. Woollett.** 1983. MRC OX-22, a monoclonal antibody that labels a new subset of T lymphocytes and reacts with the high molecular weight form of the leukocyte-common antigen. J. Exp. Med. **158:**795–810.
21. **Streuli, M., L. R. Hall, Y. Saga, S. F. Schlossman, and H. Saito.** 1987. Differential usage of three exons generates at least five different mRNAs encoding human leukocyte common antigens. J. Exp. Med. **166:**1548–1566.
22. **Streuli, M., T. Matsuyama, C. Morimoto, S. F. Schlossman, and H. Saito.** 1987. Identification of the sequence required for expression of the 2H4 epitope on the human leukocyte common antigens. J. Exp. Med. **166:**1567–1572.
23. **Thomas, M. L., A. N. Barclay, J. Gagnon, and A. F. Williams.** 1985. Evidence from cDNA clones that the rat leukocyte-common antigen (T200) spans the lipid bilayer and contains a cytoplasmic domain of 80,000 Mr. Cell **41:**83–93.
24. **Thomas, M. L., P. J. Reynolds, A. Chain, Y. Ben-Neriah, and I. S. Trowbridge.** 1987. B-cell variant of mouse T200 (Ly-5): evidence for alternative mRNA splicing. Proc. Natl. Acad. Sci. USA **84:**5360–5363.
25. **Tung, J. S., M. C. Deere, and E. A. Boyse.** 1984. Evidence that Ly-5 products of T and B cells differ in protein structure. Immunogenetics **19:**149–154.
26. **Tung, J. S., M. P. Scheid, M. A. Peirotti, U. Hammerling, and E. A. Boyse.** 1981. Structural features and selective expression of three Ly-5$^+$ cell-surface molecules. Immunogenetics **14:**101–106.
27. **Upholt, W. B., and L. J. Sandell.** 1986. Exon/intron organization of the chicken type II procollagen gene: intron size distribution suggests a minimal intron size. Proc. Natl. Acad. Sci. USA **83:** 2325–2329.
28. **Woollett, G. R., A. N. Barclay, M. Puklavec, and A. F. Williams.** 1985. Molecular and antigenic heterogeneity of the rat leucocyte-common antigen from thymocytes and T and B lymphocytes. Eur. J. Immunol. **15:**168–173.
29. **Yakura, H., F. W. Shen, E. Bourcet, and E. A. Boyse.** 1983. On the function of Ly-5 in the regulation of antigen-driven B cell differentiation: comparison and contrast with Lyb-2. J. Exp. Med. **157:**1077–1088.