Alternative use of 5' exons in the specification of Ly-5 isoforms distinguishing hematopoietic cell lineages

(hematopoietic differentiation/Ly-5 molecular isoforms/alternative splicing/S1 nuclease mapping)

Yumiko Saga, Jwu-Sheng Tung, Fung-Win Shen, and Edward A. Boyse

Memorial Sloan–Kettering Cancer Center, New York, NY 10021

ABSTRACT Previous inferences that Ly-5 glycoprotein isoforms of murine hematopoietic cells are generated by alternative splicing of primary transcripts of a single *Ly-5* gene are supported by the present study. A cDNA library was prepared from B cells by extension from primer representing a known T-cell cDNA sequence. Three different Ly-5 clones from this library included sequences missing in T-cell cDNA clones. From the constitution of cDNA clones and of the *Ly-5* gene, and from S1 nuclease mapping, it is concluded that at least two exons, provisionally numbered Ex-6(B) and Ex-7(B), in the 5'-proximal region are mainly represented in mRNA of the B-cell lines examined but not of the T-cell lines examined. Also, exons 1 and 2 appear to be used alternatively in different species of B-cell mRNA and probably also in different species of T-cell mRNA.

Interest in the Ly-5 system of the mouse (1), independently described as the T200 system (2) before it was known that Ly-5 and T200 refer to the same system, stems largely from representation of the products as multiple glycoprotein isoforms of the cell-surface defining stages and lineages of hematopoietic development (3–5). Two of these isoforms, a 200-kDa isoform from T cells and a 220-kDa isoform from B cells, are conveniently named T200 and B220 (6). Evidence from protein composition, the constitution of cDNA clones and details of genomic organization (7, 8) have suggested that the larger B220 isoform is encoded by a longer mRNA containing extra sequences shortly following a leader peptide sequence and that T200 and B220 mRNAs are generated by alternative splicing of the primary transcript of a single Ly-5/T200 gene. That conclusion is supported by evidence presented here that B220 cDNA clones include extra sequences at the predicted position, representing at least two exons absent from T200 mRNA in consequence of alternative splicing.

MATERIALS AND METHODS

Preparation of RNA. Total RNA was prepared from the cell lines named using guanidinium isothiocyanate followed by centrifugation through a cushion of CsCl (9). Poly(A)$^+$ RNA was then isolated by oligo(dT)-cellulose chromatography (10).

Construction of a Primer-Extension B-Cell cDNA Library. Antisense oligonucleotide primer comprising 18 nucleotides (nt) was synthesized according to a sequence 151–168 nt downstream from the putative interpolation site of extra sequences used by B cells (8). Primer was hybridized with 10 μg of poly(A)$^+$ RNA prepared from the B-cell leukemia I.29. The first strand was synthesized by reverse transcriptase at 42°C (11) and second-strand synthesis was by RNase H and

DNA polymerase I according to Gubler and Hoffman (12). Double-stranded DNA fragments, blunt-ended by RNase A and DNA polymerase Klenow fragment, were ligated into the bacterial alkaline phosphatase-treated *Sma* I site of pUC13 vector (Pharmacia).

S1 Nuclease Protection Mapping. Probes for S1 nuclease mapping were prepared from clones pLy5-B15 and -B51 (see below). End-labeled probe DNA (2×10^5 cpm) was hybridized with total RNA (30 μg) at various temperatures and digested with S1 nuclease (37°C for 30 min) (13). Protected fragments (PFs) were analyzed on 6% polyacrylamide/7 M urea sequencing gel.

RNA Transfer Blotting. Electrophoresis and blotting of mRNAs and hybridization with nick-translated ^{32}P-labeled probes were done as described (13).

Construction and Screening of a λ Genomic Library. This library was made from cells of the T-cell leukemia variant EARAD1-β2M$^-$ in EMBL4 vector as described (8).

Sequencing of DNA. DNA was sequenced according to Sanger *et al.* (14) on fragments subcloned in M13mp18 and M13mp19.

RESULTS AND DISCUSSION

B-Cell Ly-5 cDNA Has Sequences Not Present in T-Cell Ly-5 cDNA. By screening the B-cell primer-extension library with a T-cell Ly-5 cDNA probe, three varieties of B-cell cDNA were obtained, represented by clones pLy-5-B11, -B12, and -B15. Fig. 1 illustrates that all three clones have sequences that are lacking in the T-cell cDNA clones and are located as inferred from S1 nuclease mapping (8).

All three B-cell clones share a common B-cell distinctive sequence upstream of position 196 of the T-cell cDNA clone T4 [the present notation T4 (pLy-5-T4) replaces the former notation pLy-5-R4 (8)]. Upstream of this common B-cell sequence (141 nt; Fig. 1, box E) the three B-cell clones differ (Fig. 1)—B11 is completed by a sequence (194 nt) identical to T4 (boxes C and B). B15 has a longer B-cell sequence (141 and 147 nt; boxes E and D), followed by a shorter sequence shared with T4 (135 nt; box C), and is completed by a 91-nt sequence (box A) not shared by any other T- or B-cell cDNA clone. B12 is completed by a sequence that contains stop codons in the same reading frame, suggesting an intron (confirmed by genomic sequencing; see below).

Reading frames for the B-cell distinctive sequences of B11 and longer B15 clones (equivalent to 47 and 96 amino acids, respectively) are in phase with T4. The predicted 96-amino acid region contains two potential N-glycosylation sites and is rich in serine and threonine residues (potential sites for O-linked glycosylation). This region contains no cysteines for disulfide linkage and is relatively hydrophilic, allowing that this region is exposed and may be an epitope for antigen known to be unique to the B220 Ly-5 isoform (6).

Abbreviations: nt, nucleotide(s); UT, untranslated; PF, protected fragment.

a

5′
92 151   286   primer
                                          ⌐⌐   pLy-5-
   ▯B▯   C   ┤- - - - - - - - - - - ▨F▨▨   -T4

   ▯B▯   C   ┤- - - - - - - ≡E≡▨F▨   -B11

1   91   287   433   574
▨A▨ - - - - C ▨/D/▨≡E≡▨F▨   -B15

         ▭ intron   ≡E≡▨F▨   -B12

b

A   CGAACTTCAGAGCCTCGTACCAGCTTAGTCACTGTGTTATCTTGCCAACCCCTATGTTGTTATACTCATGTGGAAGCTTGTCATATCTTGG   91

B   TTGTTCTTAGGGTTAGAGAGTAGGAAACTTGCTCCCCATCTGATAAGACAGAGTGCAAA   150

C
GGAGACCCTATTTCTTAGGGGCACAGCTGATCTCCAGATATGACCATG GGT TTG TGG CTC AAA CTT CTG GCC TTT GGA TTT GCC CTT CTG   240
                                            met gly leu trp leu lys leu leu ala phe gly phe ala leu leu
                                            (-23)   (-9)
GAC ACA GAA GTC TTT GTC ACA GGG CAA ACA CCT ACA CCC AGT GAT G   286
asp thr glu val phe val thr gly gln thr pro thr pro ser asp g
    (1)   (7)

D
GT GTG TTA TCC ACG CTG CTG CCT CAC CTG TCC CCA CAG CCT GAC TCG CAG ACG CCC TCT GCC GGA GGA GCT GAC ACT   363
ly val leu ser thr leu leu pro his leu ser pro gln pro asp ser gln thr pro ser ala gly gly ala asp thr
                                                                                              (33)
CAG ACA TTC AGC AGC CAA GCT GAC AAT CCC ACA CTC ACG CCT GCT CCC GGC GGC GGG ACT GAC CCA CCA G   433
gln thr phe ser ser gln ala asp asn pro thr leu thr pro ala pro gly gly gly thr asp pro pro g
                                                                                        (56)

E
GT GTG CCA GGG GAG AGG ACT GTA CCG GGG ACC ATT CCT GCA GAC ACA GCC TTT CCT GTT GAT ACC CCC AGC CTT GCA   510
ly val pro gly glu arg thr val pro gly thr ile pro ala asp thr ala phe pro val asp thr pro ser leu ala
                                                                                              (82)
CGC AAC AGC TCT GCT GCC TCA CCT ACA CAC ACC TCC AAT GTC AGC ACC ACA GAT ATC TCT TCA G   574
arg asn ser ser ala ala ser pro thr his thr ser asn val ser thr thr asp ile ser ser g
                                                                            (103)

F
GT GCC AGC CTC ACA ACT CTT ACA CCA TCC ACT CTG GGC CTT GCA AGC ACT GAC CCT CCA AGC ACA ACC ATA GCT ACC   651
ly ala ser leu thr thr leu thr pro ser thr leu gly leu ala ser thr asp pro pro ser thr thr ile ala thr
                                                                                              (129)
ACA ACG AAG CAA ACA TGT GCT GCC ATG TTT GGG AAC ATT ACT GTG AAT TAC ACC TAT GAA TCT AGT AAT CAG ACT TTT   729
thr thr lys gln thr cys ala ala met phe gly asn ile thr val asn tyr thr tyr glu ser ser asn gln thr phe
                                                                                              (155)
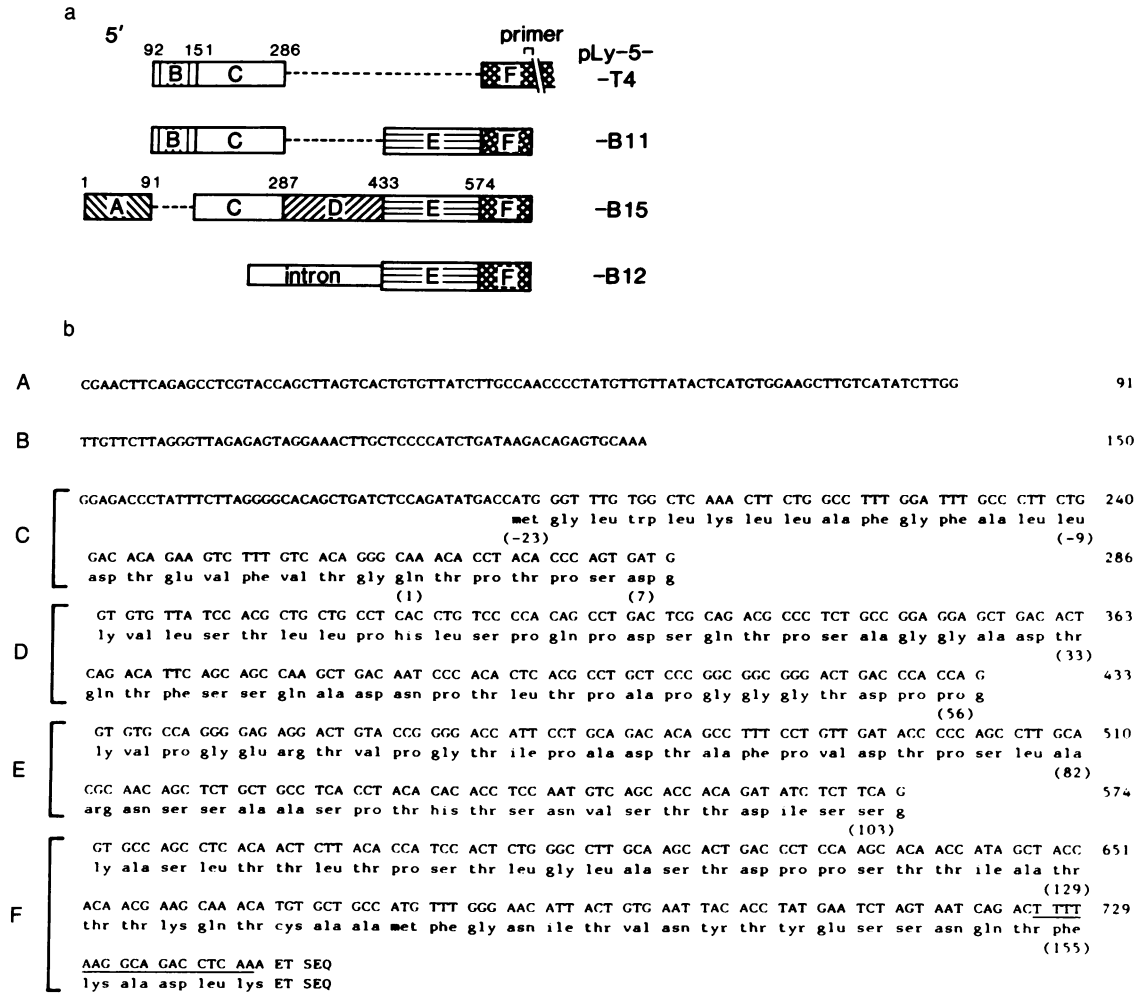AAG GCA GAC CTC AAA ET SEQ
lys ala asp leu lys ET SEQ

FIG. 1. (*a*) Diagrammatic comparison of sequences of the three B-cell Ly-5 cDNA clones (B11, B15, B12) and the T-cell cDNA clone pLy-5-T4 (formerly pLy-5-R4) (8). (*b*) Composite 5′ nucleotide and predicted amino acid sequences of B-cell Ly-5 cDNA clones. Primer sequences for B-cell extended cDNA library construction are underlined. Nucleotides are numbered on the right. Amino acids are numbered beneath in parentheses.

B-Cell-Distinctive Ly-5 mRNA Sequences Are Encoded by Exons Lacking in T-Cell mRNA.

A genomic library in EMBL4, screened with a T-cell Ly-5 cDNA probe, yielded 13 overlapping clones spanning ≈120 kilobases (kb) and inferred to comprise a single *Ly-5* gene from which diverse Ly-5 isoforms are generated (8). The B-cell-distinctive cDNA sequences noted above also imply an origin of Ly-5 isoforms by alternative splicing of primary Ly-5 transcripts.

Fig. 2 shows exon–intron organization of the 5′ end of the *Ly-5* gene, including the first 8 exons that comprise the 5′ region of T4.

The first exon represented in T4 cDNA is Ex-2 (Fig. 1, box B) and belongs to the 5′ untranslated (UT) region. The next is Ex-3, composing the rest of the UT region and the leader sequence (15). Ex-4 contains the first seven N-terminal amino acids common to all Ly-5 isoforms (15). These exons form the region denoted by box C (Fig. 1). The next T4 exon, Ex-8, is ≈10 kb downstream from Ex-4.

To verify the prediction from cDNA sequences that B-cell-distinctive sequences should lie between Ex-4 and Ex-8, a B-cell cDNA fragment containing these sequences was used to probe DNA from the λ genomic clones in Southern blotting. Two exons, Ex-6(B) and Ex-7(B) (Fig. 1, boxes D and E), were identified by sequencing genomic fragments thus selected and are seen to lie between Ex-4 and Ex-8. [A further upstream B-cell-distinctive exon, called Ex-5(B) in our provisional notation (129 nt), is predicted by the complete sequence of a cDNA clone isolated from the pre-B-cell line
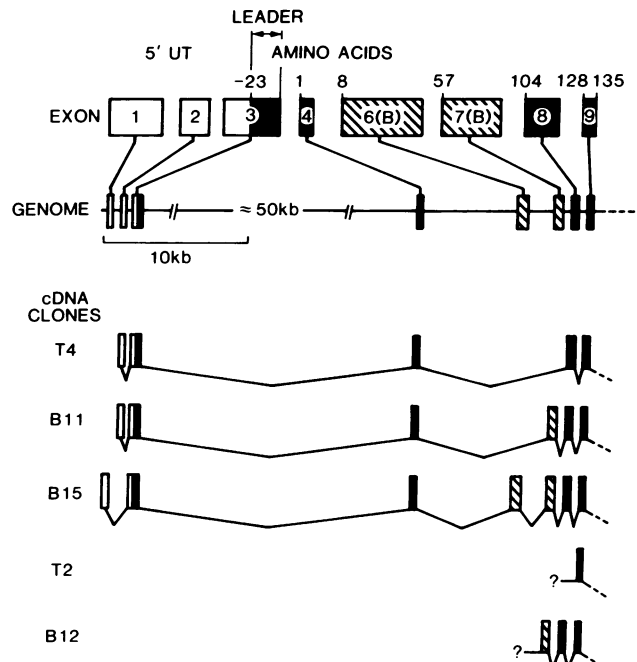
LEADER

5′ UT ┌─┤AMINO ACIDS
                    -23 1 8   57   104 128 135
EXON ▯1▯ ▯2▯ ▮3▮ ▮4▮ ▨6(B)▨ ▨7(B)▨ ▮8▮ ▯9▯

GENOME ▮▯▮ —//— ≈50kb —//—   ▮   ▨▯▮▮▮ ----
       └──────┘
        10kb

cDNA
CLONES

T4   ▯▯   ▮   ▮▮

B11 ▯▯   ▮▮▮

B15 ▯   ▮   ▨▮▮▮

T2   ?▮

B12   ?▨▮▮▮

FIG. 2. Organization and processing patterns of the 5′ proximal region of the *Ly-5* gene (exons not to scale).
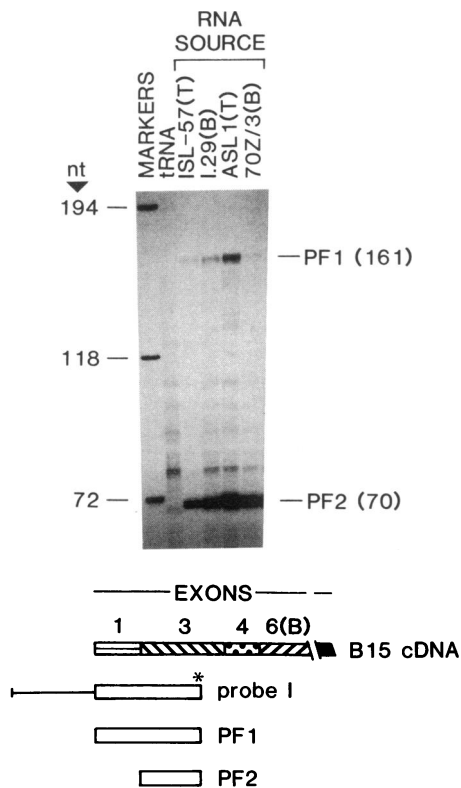
FIG. 3. S1 nuclease mapping with probe I (Ex-1 and Ex-3). A complete PF (PF1) of 161 nt and a partial PF (PF2) of 70 nt are seen with both T and B cells, suggesting that T-cell Ly-5 transcripts, like B-cell transcripts, may use either Ex-1 or Ex-2.

70Z/3 by Thomas *et al.* (16). The remaining sequences reported here and by Thomas *et al.* (16) are not expected to reveal differences accounting for Ly-5 alloantigenic variation because the cells of origin, I/St in the case of I.29 B cells (present report), and C57BL/6 and DBA/2 in the case of 70Z/3 cells (16), have the same Ly-5 type (Ly-5.1, *Ly-5ᵃ*) (17)].

The exon common to all three B-cell cDNA clones (box E) is Ex-7(B). The 5' sequence peculiar to B12 (Fig. 1) is seen to be an intron that abuts Ex-7(B) upstream. In addition to

inclusion versus exclusion of Ex-6(B), it appears that there is at least one additional site for alternative splicing in the 5' UT region. Thus, clone B15 begins with Ex-1, whereas clones B11 and also T4 begin with Ex-2. S1 nuclease mapping was used to find out whether some Ly-5 mRNA species of T cells also may contain Ex-1 in place of Ex-2. For this purpose, a probe with its 5' end in Ex-3 and incorporating all of Ex-1 (probe I, Fig. 3) was prepared from B15.

Fig. 3 shows that a protected fragment (PF1, 161 nt) was observed with mRNA of T cells (leukemias ISL-57 and ASL1) as well as B cells (leukemias I-29 and pre-B 70Z/3), suggesting that some T-cell mRNA may use Ex-1 in place of Ex-2. A second PF (PF2, 70 nt) is probably protected by T4 mRNA and B11-like mRNA, which incorporate only the Ex-3 region of the probe.

Thus, tentatively, T cells or sublineages of T cells, like B cells, may use either Ex-1 or Ex-2, and in both cases transcription may begin at alternative sites. Preliminary data from primer extension suggest alternative initiation sites for both T- and B-cell transcripts and that the presence of Ex-1 and of Ex-2 are mutually exclusive. Alternative use of Ex-1 and Ex-2 would seem to generate diversity of sequence in the 5' UT region without altering the reading frame.

In Both T- and B-Cell Lineages, Ly-5 mRNAs Are Diverse, Using Alternative 5' Exons. The diversity of B-cell transcripts inferred from the composition of clones B11 and B15 was studied further by S1 nuclease mapping. Probe II [767 nt, from the 5' end to the *Eco*RV site in Ex-7(B)] was prepared from clone B51 (see legend of Fig. 4), end-labeled, hybridized with total B-cell RNA (I.29, 70Z/3), and digested with S1 nuclease.

Fig. 4 shows four PFs. The longest (PF1, 531 nt) is evidently protected by mRNA of B15 type [Ex-1, -3, -4, -6(B), and -7(B)]. PF2 (414 nt) is probably protected by mRNA containing Ex-3, -4, -6(B), and -7(B), without Ex-1; because Ex-1 is missing in these RNAs (RNA type 1 predicted), we surmise that they use Ex-2 instead of Ex-1, which T4 and B11 use. PF3 (278 nt) appears to be protected by RNAs (RNA type 2 predicted) that contain Ex-6(B) and -7(B) and presumably relates to a third B-cell exon, provisionally called Ex-5(B), noted by Thomas *et al.* (16) (see above), which is not represented in our cDNA clones. PF4 (131 nt) is probably protected by mRNA representing B11, containing only Ex-7(B) among B-cell specific sequences, and/or B12, which contains the upstream intron sequence.
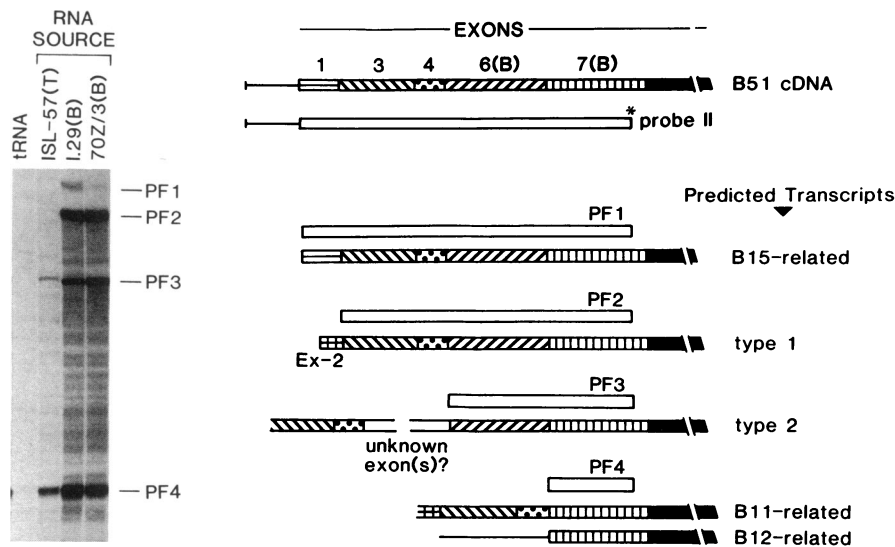

FIG. 4. S1 nuclease mapping with probe II [Ex-1, -3, -4, -6(B), and -7(B)], suggesting alternative use of 5' exons in both T and B cells. Probe II was prepared from clone B51, which is identical to B15 but has an added 5' sequence of 261 nt of which 235 nt from the 5' end are apparently unrelated to Ly-5 and probably represent a cloning artifact. PF1, 531 nt; PF2, 414 nt; PF3, 278 nt; PF4, 131 nt.
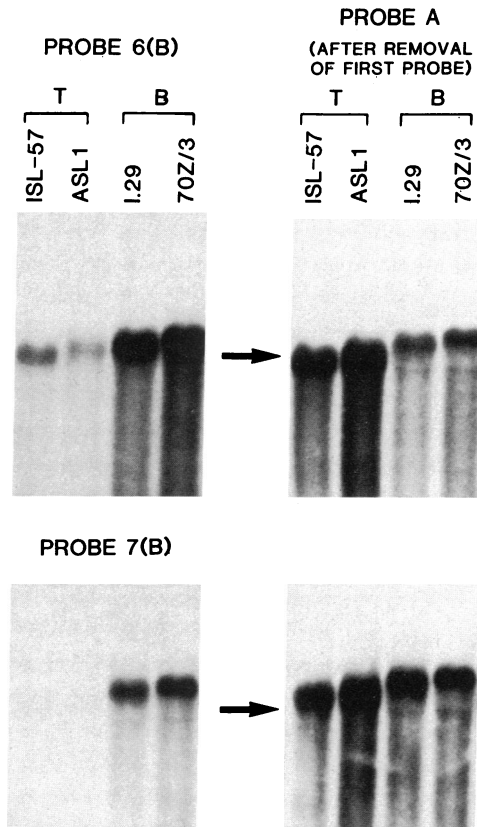
Genetics: Saga *et al.*

Proc. Natl. Acad. Sci. USA 84 (1987)   5367

FIG. 5. RNA transfer blotting with probes 6(B) and 7(B) (*Left*) prepared from genomic clones and containing Ex-6(B) and Ex-7(B), respectively, followed by probe A (*Right*) prepared from cDNA clone pLy-5-T4 as a 2.3-kb *Bam*HI fragment and representing a sequence common to both T- and B-cell cDNA. Ly-5 RNA of I-29 cells appears slightly smaller than RNA of 70Z/3 cells, in keeping with the observation (unpublished) that I-29 cells express slightly smaller and diverse Ly-5 isoforms.

Thus B-cell mRNAs are diverse in their 5' coding regions. Whether each transcript is functional remains to be seen; there appears to be a single Ly-5 isoform (220 kDa) in 70Z/3 pre-B cells but I.29 B leukemia cells appear to yield more than one Ly-5 isoform, all of somewhat smaller molecular mass (accord-

ing to Laemmli-type concentration-gradient NaDodSO$_4$/PAGE; unpublished observation), which might imply more than one functional transcript in the I.29 cell population.

Since the 5' end of probe II (Fig. 4) is located within Ex-7(B), T-cell mRNA should not protect this probe and should yield no PFs. Accordingly, total RNA of ISL-57 T cells yielded neither PF1 nor PF2. However, small amounts of PF3 and PF4 were observed (Fig. 4). This finding, that a B-cell-specific probe can be marginally protected by T-cell mRNA, called for further review of T-cell transcripts. Therefore, we prepared two genomic probes—probe 6(B) containing Ex-6(B) and probe 7(B) containing Ex-7(B)—for RNA transfer blotting (Fig. 5). Both probes hybridized strongly with B-cell RNA as expected. However, probe 6(B), and to a lesser extent probe 7(B), hybridized with RNA of T cells (ISL-57 and ASL1), suggesting that some species of T-cell RNA may include Ex-6(B) and/or Ex-7(B), although most T-cell Ly-5 RNA does not, as concluded previously (8). After complete removal of probes 6(B) and 7(B), the same blots were hybridized with probe A, a 2.3-kb *Bam*HI fragment of cDNA clone pLy-5-T4, which is common to T and B cells. Comparing the hybridization patterns of probes 6(B) and A in Fig. 5, it is clear that probe A gives a broader band, probably signifying more than one RNA species, in contrast to probe 6(B), which gives only the upper part of that band. The same may apply to Ex-7(B) in lesser degree according to the lesser intensity of the hybridization bands evident in Fig. 5. Thus, on present evidence PF3 and PF4 generated from T cells (as shown in Fig. 4) may be construed to represent minor species of T-cell RNA that include Ex-6(B) and/or Ex-7(B).

Possible Relationship of Genomic Sequence to Alternative Modes of Processing of Primary Transcripts. In this context, we searched regions flanking intron–exon boundaries of the 33 Ly-5 exons identified, which include the sequences of all present cDNA clones, to find sequences that might distinguish constitutive exons (used for all Ly-5 isoforms) from conditional exons (used for some Ly-5 isoforms). The first eight exons shown in Fig. 6 are flanked by consensus splicing junction sequences and follow the AG/GT rule (18). All six exons of the coding region terminate within a codon, splitting the first and second nucleotides. We note that a stretch of similar sequences [GG(T/A)GTGAT] (doubly underlined) occurs in the upstream intron of Ex-6(B) (−31 to −38) and Ex-7(B) (−27 to −34) but not of the other 31 exons. The last nucleotide of the upstream intron and first six 5' nucleotides

FIG. 6. Exon–intron boundary sequences of the first eight Ly-5 exons. Aligned boundary sequences show close similarities to consensus sequences for both donor and acceptor splicing sites (18). Doubly underlined and heavily underlined sequences are unique to Ex-6(B) and Ex-7(B).

(heavily underlined) of the splicing acceptor sites of Ex-6(B) (G/GTGTGT) and Ex-7(B) (G/GTGTGC), but not of the acceptor sites of the 31 other exons, resemble the consensus sequence of the splicing donor site, G/GT(G/A)AGT (18). Possibly these junction sequences peculiar to Ex-6(B) and Ex-7(B) may be recognized as splicing donor sites in the T-cell T200 program but not in the B-cell B220 programs, the succeeding exon being spliced out only in T cells.

1. Komuro, K., Itakura, K., Boyse, E. A. & John, M. (1975) *Immunogenetics* 1, 452–456.
2. Trowbridge, I. S., Ralph, P. & Bevan, M. J. (1975) *Proc. Natl. Acad. Sci. USA* 72, 157–161.
3. Michaelson, J., Scheid, M. P. & Boyse, E. A. (1979) *Immunogenetics* 9, 193–197.
4. Tung, J.-S., Scheid, M. P., Pierotti, M. A., Hammerling, U. & Boyse, E. A. (1981) *Immunogenetics* 14, 101–106.
5. Tung, J.-S., Scheid, M. P. & Palladino, M. A. (1983) *Immunogenetics* 17, 649–654.
6. Coffman, R. L. & Weissman, I. L. (1981) *Nature (London)* 289, 681–683.
7. Shen, F.-W., Saga, Y., Litman, G., Freeman, G., Tung, J.-S., Cantor, H. & Boyse, E. A. (1985) *Proc. Natl. Acad. Sci. USA* 82, 7360–7363.
8. Saga, Y., Tung, J.-S., Shen, F.-W. & Boyse, E. A. (1986) *Proc. Natl. Acad. Sci. USA* 83, 6940–6944, and correction (1987) 84, 1991.
9. Chirgwin, J. M., Przybyla, A. E., McDonald, R. J. & Rutter, W. J. (1979) *Biochemistry* 18, 5294–5299.
10. Aviv, H. & Leder, P. (1972) *Proc. Natl. Acad. Sci. USA* 69, 1408–1412.
11. Davis, M. M., Cohen, D. I., Nielsen, E. A., Steinmetz, M., Paul, W. E. & Hood, L. (1984) *Proc. Natl. Acad. Sci. USA* 81, 2194–2198.
12. Gubler, U. & Hoffman, B. J. (1983) *Gene* 25, 263–269.
13. Maniatis, T., Fritsch, E. F. & Sambrook, J. (1982) *Molecular Cloning: A Laboratory Manual* (Cold Spring Harbor Laboratory, Cold Spring Harbor, NY).
14. Sanger, F., Nicklen, S. & Coulson, A. R. (1977) *Proc. Natl. Acad. Sci. USA* 74, 5463–5467.
15. Barclay, A. N., Jackson, D. I., Willis, A. C. & Williams, A. F. (1987) *EMBO J.* 6, 1259–1264.
16. Thomas, M. L., Reynolds, P. J., Chain, A., Ben-Neriah, Y. & Trowbridge, I. S. (1987) *Proc. Natl. Acad. Sci. USA* 84, 5360–5363.
17. Shen, F. W., Tung, J.-S. & Boyse, E. A. (1986) *Immunogenetics* 24, 146–149.
18. Mount, S. M. (1982) *Nucleic Acids Res.* 10, 459–472.