Proc. Natl. Acad. Sci. USA
Vol. 83, pp. 6940–6944, September 1986
Genetics

Sequences of Ly-5 cDNA: Isoform-related diversity of Ly-5 mRNA

(restriction maps/S1 nuclease protection/alternative splicing/protein domains)

YUMIKO SAGA, JWU-SHENG TUNG, FUNG-WIN SHEN, AND EDWARD A. BOYSE

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

Contributed by Edward A. Boyse, May 30, 1986

ABSTRACT The Ly-5 system of the mouse is expressed exclusively by hematopoietic cells and comprises a series of glycoprotein isoforms that typify different hematopoietic cell lineages. The 200-kDa isoform of T cells and the 220-kDa isoform of B cells are known to differ in peptide composition. The complete 1152 amino acid sequence of the 200-kDa isoform protein deduced from cDNA sequence appears to comprise a leader sequence of some 30 residues, an external N-terminal domain of 370 residues, a probably single transmembrane domain of 22 residues, and an unusually large cytoplasmic domain of 730 residues. Both the external and cytoplasmic domains include regions of internal homology suggestive of evolution from a smaller ancestral gene. RNA transfer blotting has previously shown that B-cell mRNA for Ly-5 is larger than T-cell mRNA. S1 nuclease protection mapping with Ly-5 cDNA probes suggests that this difference can be ascribed to interpolation of an extra B-cell sequence located at the 5′ end of B-cell mRNA, probably immediately following the leader sequence. From restriction mapping of overlapping Ly-5 genomic clones spanning 60 kilobases it is concluded that Ly-5 isoforms are generated by differential processing of transcripts of a single gene, rather than from a family of linked *Ly-5* genes.

Salient features of the Ly-5 system of the mouse are restriction of its expression to the hematopoietic compartment of development (1) and the generation of discrete Ly-5 glycoprotein isoforms (2–4), two of which are known to differ in protein composition (5), that characterize different hematopoietic cell lineages. It has been suggested that functions of Ly-5 observed in the lymphocyte lineage may represent one category of a set of allied intercellular regulatory mechanisms in which Ly-5 isoforms are involved throughout the hematopoietic compartment (6). The present report is concerned mainly with the genetic mode of origin of Ly-5 isoforms.

Ly-5, originally defined by intraspecies alloantigens, is deemed to be the same system as T200, originally defined by monomorphic antigen recognized by rat antiserum. The rat system L-CA (leukocyte-common antigen) is deemed to represent the rat homologue of Ly-5/T200.

MATERIALS AND METHODS

Rescreening of the C1.Ly1-T1 cDNA Library. A sublibrary of the original C1.Ly1-T1 cDNA library with insert size 4–7 kilobases was screened with the 5′ fragment C (probe 68.1) of pLy-5-68 (7).

DNA Sequencing. This was conducted according to Sanger *et al.* (8) on fragments subcloned in M13mp18 and mp19 (9). Subcloning utilized both blunt-end and sticky-end cloning strategies, and the sequences were confirmed on both strands.

Preparation of RNA. Total RNA was prepared (7) from the cells and cell lines named using guanidium isothiocyanate followed by centrifugation through a cushion of $CsCl_2$.

S1 Nuclease Protection Mapping. Probes for S1 mapping were prepared from pLy-5-68 subcloned in pBR322 vector. As described by Maniatis *et al.* (10), 3′- or 5′-end-labeled probe DNA (1×10^5 cpm) was hybridized with total RNA (30 μg) at various temperatures and digested with S1 nuclease (37°C for 30 min), after which DNA·RNA hybrids were electrophoresed on 1.5% agarose gels or 6% polyacrylamide/7 M urea gels.

Construction and Screening of a λ Genomic Library. This genomic library was made from EARAD1-β2M-negative cells, a (B6 × A)F_1 T-cell leukemia variant with mutations in both *B2m* alleles. Cellular DNA was partially digested with *Sau*3A and inserted into λ vector EMBL.4 according to Frischauf *et al.* (11). About 1×10^6 plaques were screened with pLy-5-68 cDNA.

RESULTS AND DISCUSSION

Nucleotide and Predicted Amino Acid Sequences for Ly-5 cDNA. The partial sequence of the 5′ region of pLy-5-68 cDNA (7) indicates that this clone is incomplete at the 5′ end. The original T-cell cDNA library was therefore rescreened with a 5′ fragment of the original clone pLy-5-68. Of nine positive clones (pLy-5-R1 to -R9), four with inserts longer than pLy-5-68 were selected (pLy-5-R1 to -R4) and restriction mapped (Fig. 1). These four clones are of two types. R3 and R4 are identical in sequence to pLy-5-68 but have 111 and 132 more nucleotides, respectively, at their 5′ ends. R1 and R2, which are identical, differed from pLy-5-68 in sequence in the 5′ region, displaying an *Ava* II site in that region, and are 663 nucleotides longer than pLy-5-68 at the 5′ end.

Nucleotide and predicted amino acid sequences are given for clones pLy-5-R2 and pLy-5-R4 in Fig. 2. pLy-5-R4 (4590 nucleotides) appears to include the complete coding region. There are two ATG triplets (positions 100–102 and 106–108) near the beginning of the same reading frame. The second ATG (positions 106–108) should be the initiation codon because the sequence ACCATGG agrees with the consensus sequences suggested for eukaryotic initiation sites (12). This initiation codon (positions 106–108) marks an open reading frame coding for a protein of 1152 amino acids terminating at the stop codon TAG (positions 3562–3564). There are two potential polyadenylylation signals, AATAAA hexamers, positions 4347–4352 and 4567–45472, in the long 3′-untranslated region, and polyadenylylation begins 18 nucleotides downstream from the second AATAAA. The first polyadenylylation signal is apparently also used because we have observed a third type of Ly-5 cDNA, clone pLy-5-R5, to which the poly(A) tail was attached 12 nucleotides downstream. The sequence of pLy-5-R5 was identical to R4 except for a shorter 3′ end.

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

Abbreviations: L-CA, leukocyte-common antigen; PF, protected fragment (in S1 nuclease protection mapping); Mφ, macrophages.

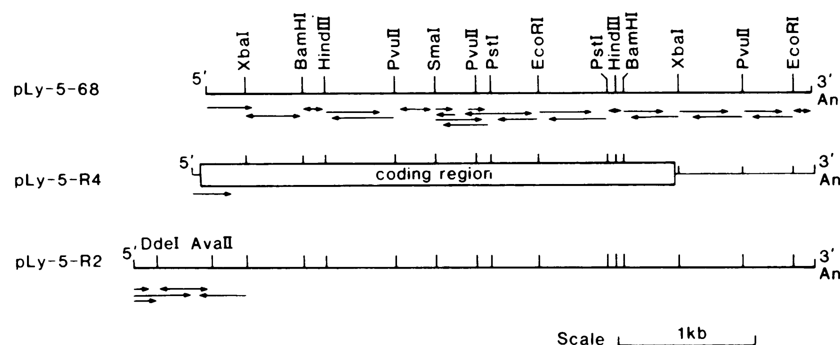

FIG. 1. Restriction maps and sequencing strategies for Ly-5 cDNAs.

Computer-aided hydrophobicity analyses suggest that the first 30 or so amino acids represent a leader peptide, and the 22 amino acids, positions 426–447 (numbered from the initiation codon) represent a transmembrane segment. The portion ≈44 kDa proximal to the N terminus is presumably extracellular and 14 theoretical N-linked glycosylation sites are predicted in this region from the sequence Asn-Xaa-Thr(Ser). There are only four theoretical N-glycosylation sites in the unusually large C-terminal cytoplasmic region of ≈83 kDa. In total, there are 30 cysteines, 18 of them evidently extracellular. These 18 cysteines are probably involved in intramolecular disulfide bonding, as has been suggested for cysteines similarly placed in rat L-CA, the rat homologue of mouse Ly-5/T200 (13). Ly-5 is known to be phosphorylated exclusively at serine residues (14, 15). Several potential phosphorylation sites are found in the cytoplasmic region of Ly-5 protein. The region including amino acids 844–861 is a candidate for phosphorylation by glycogen synthase kinase 5 (16), and the regions of residues 453–458 and 817–823 are both potential sequences for phosphorylation by protein kinase type C (17).

The molecular size of the predicted protein is ≈127 kDa (without leader peptide), which is ≈33 kDa less than the previously estimated 160 kDa of the T-cell Ly-5 isoform devoid of N-linked carbohydrates (5). This discrepancy may reflect inaccuracy of molecular sizes estimated by NaDodSO$_4$/PAGE; the reported molecular size for T-cell Ly-5/T200 ranged from 175 to 200 kDa. Also, phosphate side chains and/or O-linked carbohydrate may increase the apparent size in NaDodSO$_4$/PAGE.

By computer analysis, certain nucleotide and amino acid sequences within the extracellular and cytoplasmic domains of pLy-5-R4 are repetitious, suggesting evolution by duplication of a smaller ancestral gene. Thus in the extracellular domain there is 38% homology between nucleotide regions 20–130 and 570–680 and 37% homology between regions 290–420 and 1030–1160. Similarly, the cytoplasmic protein coding domain appears divisible into two roughly equal parts that include stretches of 100–200 nucleotides of 30–50% homology (nucleotides 1590–1710 vs. 2460–2580, 1770–1950 vs. 2690–2870, and 2230–2350 vs. 3170–3290).

The R2 sequence from position 798 to the 3′ end is identical to the R4 sequence from position 267 to the 3′ end and thus R2 is 531 nucleotides longer at its 5′ end. There would be 16 stop codons in the first 798 nucleotides of R2 if the reading frame were the same as for R4 and 20 or 22 stop codons if either of the two alternative frames were used, making it unlikely that R2 represents a functional message that could account for the second Ly-5 product defined in T cells by serological and biochemical criteria (18).

Nevertheless, two Ly-5 RNA species represented by R2 and R4 cDNAs are revealed in T cells by S1 mapping. Thus with probe 68.1a (comprising 288 nucleotides from the 5′ end to the first downstream *Xba* I site of pLy-5-68, positions 133–420 of R4), two prominent protected fragments (PFs) are seen (see Fig. 5*A* and below). One is a complete PF of 288 nucleotides representing R4. The other is a 153 nucleotide PF representing R2. The length of each PF corresponds with the homologous sequences between R2 and R4 (Fig. 2).

Comparison Between Ly-5 and Rat L-CA. Computer-aided comparison of protein and nucleotide sequences of pLy-5-R4 (Fig. 2) with rat L-CA (13) substantiates the homology of mouse pLy-5-R4 with rat L-CA. Overall amino acid homology with respect to the extracellular domain of Ly-5, where antigens accessible on intact cells should be situated, is 47%, and nucleotide homology is 69% (both allowing small gaps). The presumed transmembrane domains are identical in amino acid sequence and also in nucleotide sequence except for the silent substitution of cytosine for thymidine in L-CA at position 1429. Overall amino acid and nucleotide homologies for the cytoplasmic domain are 87 and 89%, respectively. The incomplete noncoding 3′ sequence reported for the L-CA cDNA clone pLC-1 (13) has 262 nucleotides, and the first 60 nucleotides of this clone downstream of the stop codon show 80% homology with the equivalent region of pLy-5-R4. However, the remaining sequence (≈200 nucleotides) of pLC-1 shows no homology with Ly-5-R4. But the sequence near the 3′ end of another L-CA cDNA clone, pLC-2 (105 nucleotides) (13), shows ≈70% homology with the equivalent region of pLy-5-R4, including a stretch of 22 perfectly matched nucleotides that contains the second AATAAA hexamer of Ly-5-R4. Thus sequences near the beginning (≈60 nucleotides) and the end (≈100 nucleotides) of the untranslated 3′ region of pLy-5-R4 (1029 nucleotides) show 70–80% homology with those of L-CA cDNA. The middle portion (≈870 nucleotides) of this region, however, is probably quite different in pLy-5-R4 as compared with L-CA.

The positions of 15 of the 18 cysteine residues in the extracellular domain of Ly-5 protein match those of L-CA protein and the number of matchable cysteines may be higher because N-terminal sequencing of L-CA protein is not complete, implying that similar conformations can be reached by intramolecular disulfide bonding. The three potential phosphorylation sites in the cytoplasmic domain, noted above, are conserved in L-CA and Ly-5.

A Difference in the 5′ Region of Ly-5 RNA that Distinguishes T Cells from B Cells and Macrophages. In RNA transfer blotting with the pLy-5-68 probe, B-cell Ly-5 RNA (≈5 kb) is ≈0.3 kb longer than T-cell Ly-5 RNA (7), which accords with the difference in protein size of the respective Ly-5 isoforms of tunicamycin-treated cells (190 kDa and 160 kDa, respectively), which are known to differ in peptide composition (5). Similarly, RNA transfer blotting has distinguished mRNAs of cytotoxic T lymphocytes and helper T cells (19). Appropriate Ly-5 cDNA clones of B cells and macrophages (Mϕ) are not yet available, but S1 mapping gives further information on the respective mRNAs.

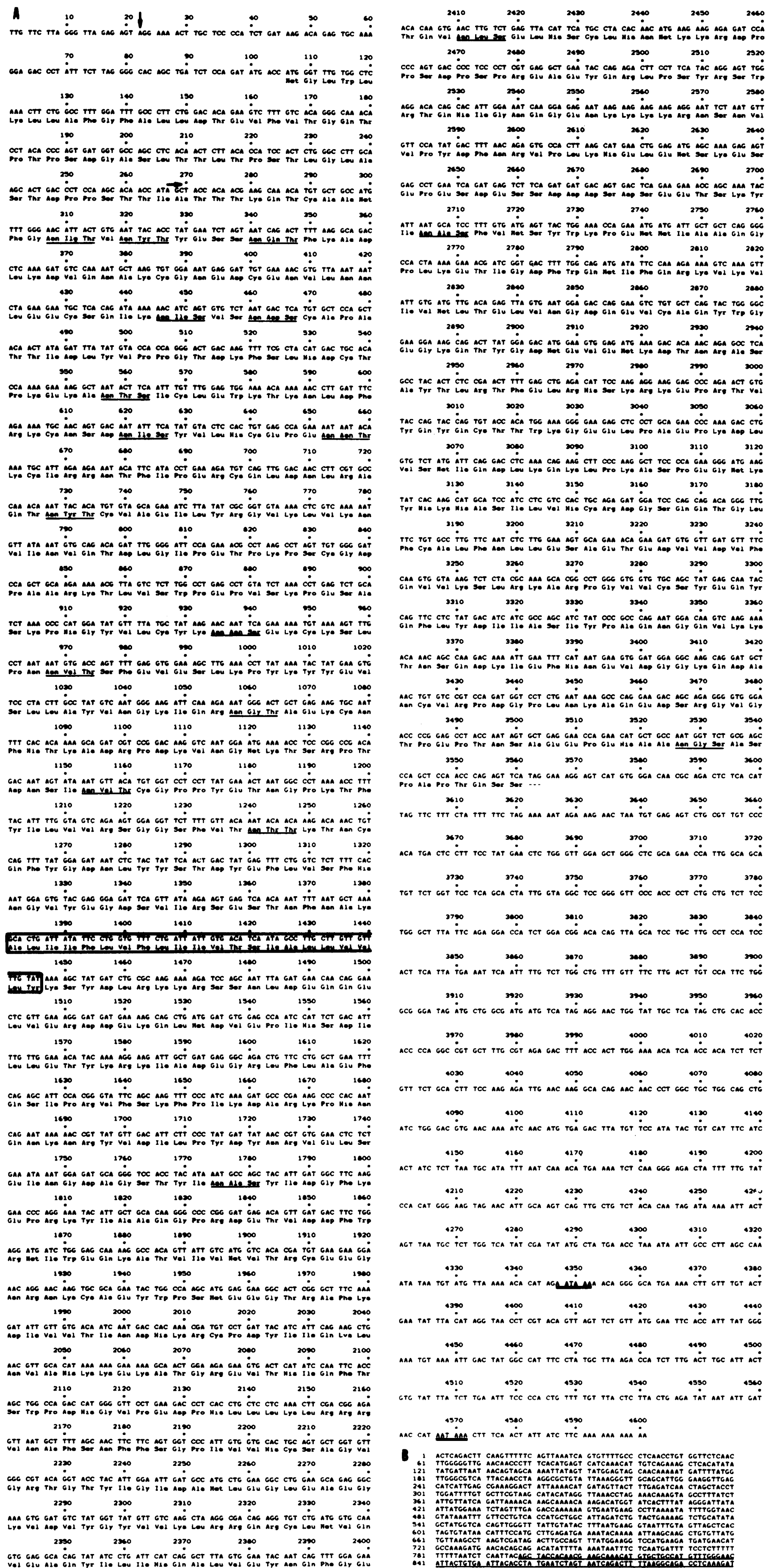

FIG. 2. DNA and predicted amino acid sequences of Ly-5 cDNA clones. (*A*) pLy-5-R4 with the deduced amino acid sequence. The open box indicates the putative transmembrane region. Theoretical N-glycosylation sites are underlined. Two poly(A) sites are doubly underlined. The vertical arrow marks the 5′ end of pLy-5-R3. The horizontal arrow marks the start of the sequence that matches R2. (*B*) pLy-5-R2 (only the first 900 nucleotides are shown). The portion underlined and the subsequent remaining sequence (not shown) to the 3′ end is identical to pLy-5-R4.

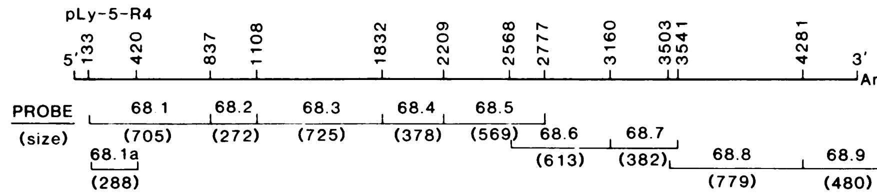

FIG. 3. Nine cDNA probes for S1 nuclease protection mapping. cDNA probes were generated by restriction enzyme digestion of pLy-5-68 and inserted into pBR322. Positions related to pLy-5-R4 and sizes (excluding vector sequences) of probes (68.1–68.9) are indicated. Vector sequence (not shown) should not be protected by RNA during S1 digestion.

Probes for S1 mapping were prepared from subclones of pLy-5-68. The nine probes obtained (68.1–68.9) spanned the length of pLy-5-68 from 5' to 3' (Fig. 3). Each was end-labeled and hybridized with total RNA of leukemia ISL-57 T cells (isoform 200 kDa), B-cell leukemia I.29 (isoform 220 kDa), and the PU-5 Mϕ cell line (isoform 205 kDa) at optimal temperature determined for each probe. After S1 digestion, DNA·RNA hybrids were electrophoresed on 1.5% agarose gels.

Probe 68.9 was unsatisfactory, perhaps because of interference by a poly(A) tail. Probes 68.2–68.8 did not distinguish between the three sources of RNA; Fig. 4*A* is typical, showing one common PF for all three cell types. However, with probe 68.1 (705 nucleotides, from the 5' end to the first downstream *Bam*HI site of pLy-5-68), at its optimal hybridization temperature of 52°C, the PF for RNA of I.29 and PU-5 cells was shorter than that of ISL-57, which gave an expected complete PF of ≈700 nucleotides (Fig. 4*B*), suggesting that Ly-5 mRNAs of B cells and Mϕ differ in sequence from T-cell mRNA. This mRNA distinction of T cells from B cells and Mϕ was verified with RNA from normal spleen cells, which include all three cell types and yielded both PFs, and from thymocytes, which yielded only the same PF as ISL-57 (Fig. 4*B*). To determine the exact point of difference in sequence, a shorter probe 68.1a (288 nucleotides, from the 5' end of pLy-5-68 to the first downstream *Xba* I site) was derived, and PFs were analyzed on 6% polyacrylamide/7 M urea DNA sequencing gels. At 52°C (Fig. 5*B*), results with 68.1a and 68.1 were similar (i.e., ISL-57 RNAs gave a complete PF of 288 nucleotides whereas I.29 and PU-5 RNAs gave a shorter PF of 230 nucleotides). I.29 and PU-5 cells also yielded a 288 PF, faint at 52°C but intense at 42°C (Fig. 5*A*).

These data may signify two possibilities: First, that B-cell Ly-5 mRNA differs in sequence from T-cell mRNA by an extra sequence interpolated at about 58 nucleotide downstream from the 5' end of probe 68.1a—i.e., immediately following the presumed pLy-5-R4 leader sequence. This extra sequence would form an internal loop during DNA·RNA duplex formation. Hybridization at lower temperature (e.g., 42°C) may enhance duplex bonding in the vicinity of the loop and thus protect the probe from digestion during digestion of the loop, thus accounting for the complete 288 PF seen. At higher temperature (e.g., 52°C), duplex bonding in the vicinity of the loop may not be complete, thus allowing cutting through by S1 nuclease to give PFs of 230 nucleotides, as observed, and 58 nucleotides (not seen

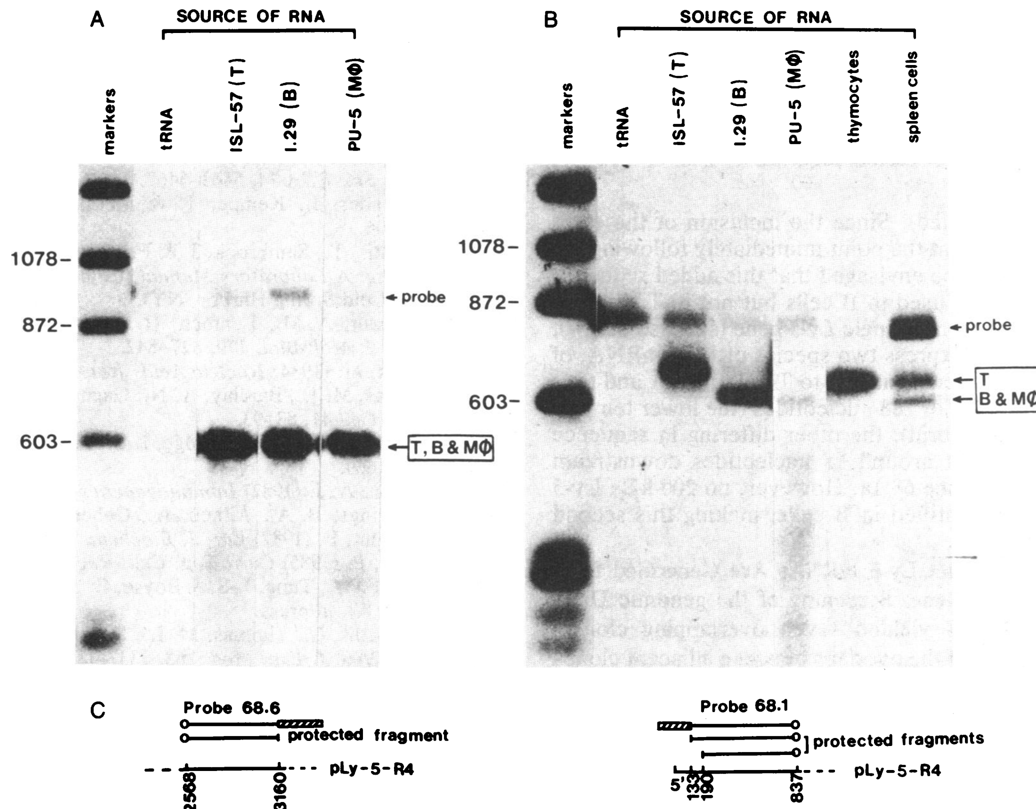

FIG. 4. Comparison of Ly-5 RNA from T cells, B cells, and Mϕ by S1 protection mapping. No difference was observed with probes 68.2–68.8, as illustrated here for probe 68.6 (*A*). With probe 68.1 (*B*) the PF for B cells and Mϕ was shorter than that for T cells. Spleen cells include all three cell types. The structure of the end-labeled (○) probes (hatched box represents vector sequence) and of PFs are diagramed (*C*).

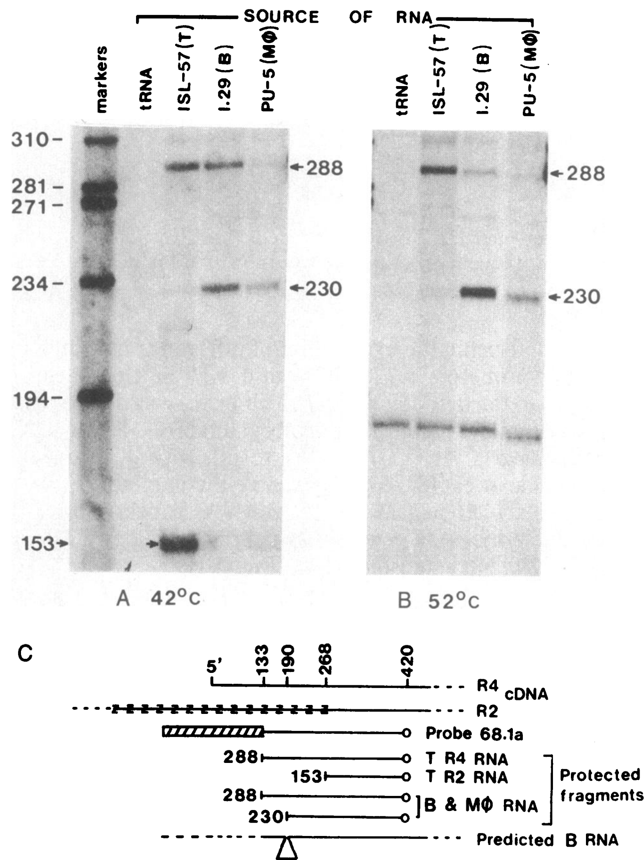

FIG. 5. S1 nuclease protection with the shorter probe 68.1a. (*A*) A complete PF of 288 nucleotides (R4-type RNA) and a 153 nucleotide PF (R2-type RNA) were seen for T cells. Two PFs (288 and 230 nucleotides, respectively) were seen for B cells and Mϕ. Hybridization temperature, 42°C. (*B*) The intensity of the 288-nucleotide PF decreased when hybridization was carried out at 52°C. The 153-nucleotide PF was not seen in T cells at 52°C probably because the low G+C content (37%) of this PF is unfavorable to the stability of hybrids at the higher temperature. Hybridization temperature, 52°C. (*C*) The structure of the end-labeled (○) probes (hatched box represents vector sequences), and of PFs are diagramed.

because not radiolabeled). Since the inclusion of the extra sequence is predicted at the point immediately following the signal peptide, it can be envisaged that this added sequence represents an exon(s) used in B cells but not in T cells via alternative splicing from a single *Ly-5* gene (see also below). Second, that B cells express two species of Ly-5 mRNA, of which one has a sequence similar to T-cell mRNA and thus gives the complete PF of 288 nucleotides (the lower temperature favoring this hybrid), the other differing in sequence from T-cell mRNA at around 58 nucleotides downstream from the 5′ end of probe 68.1a. However, no 200-kDa Ly-5 isoform has been identified in B cells, making this second possibility less likely.

Evidence that Distinct Ly-5 mRNAs Are Generated from One *Ly-5* Structural Gene. Screening of the genomic DNA library with pLy-5-68 yielded seven overlapping clones. Restriction mapping of the overlaps between adjacent clones has so far shown no discordant restriction site, a result that is consistent with a single *Ly-5* gene spanning at least 60 kilobases. The Ly-5 genomic map thus constructed showed 10 *Eco*RI sites. Cellular DNA was, therefore, digested with *Eco*RI and analyzed by Southern blotting with three cDNA fragments representing the entire sequence of pLy-5-68 as probes. The sizes and map positions of the hybridized fragments were as expected. These data are consistent with the generation of Ly-5 protein isoforms by differential processing of the primary transcript.

Further Comment. Thomas *et al.* (13) pointed out that the protein sequence of L-CA in the rat resembles no other known proteins. Thus the close homology of rat L-CA and Ly-5 sequences emphasizes once again that L-CA is the rat homologue of mouse Ly-5. As noted (13), despite lack of sequence homology with the receptor for epidermal growth factor, L-CA and Ly-5 share with the epidermal growth factor receptor the exceptional feature of an exceedingly large cytoplasmic domain, although the L-CA and Ly-5 cytoplasmic domains are even larger and show internal homology that the cytoplasmic domain of the epidermal growth factor receptor does not.

This work was supported in part by Grant CA-39827 from the National Cancer Institute. E.A.B. is American Cancer Society Research Professor of Cell Surface Immunogenetics. Y.S. is the recipient of a fellowship from the Cancer Research Institute.

1. Scheid, M. P. & Triglia, D. (1979) *Immunogenetics (N.Y.)* **9**, 423–433.
2. Michaelson, J., Scheid, M. P. & Boyse, E. A. (1979) *Immunogenetics (N.Y.)* **9**, 193–197.
3. Tung, J.-S., Scheid, M. P., Pierotti, M. A., Hammerling, U. & Boyse, E. A. (1981) *Immunogenetics (N.Y.)* **14**, 101–106.
4. Tung, J.-S., Scheid, M. P. & Palladino, M. A. (1983) *Immunogenetics (N.Y.)* **17**, 649–654.
5. Tung, J.-S., Deere, M. C. & Boyse, E. A. (1984) *Immunogenetics (N.Y.)* **19**, 149–154.
6. Yakura, H., Shen, F.-W., Bourcet, E. & Boyse, E. A. (1983) *J. Exp. Med.* **157**, 1077–1088.
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. Sanger, F., Nicklen, S. & Coulson, A. R. (1977) *Proc. Natl. Acad. Sci. USA* **74**, 5463–5467.
9. Norrander, J., Kempe, T. & Messing, J. (1983) *Gene* **26**, 101–106.
10. Maniatis, T., Sambrook, J. & Fritsch, E. G. (1982) *Molecular Cloning: A Laboratory Manual* (Cold Spring Harbor Laboratory, Cold Spring Harbor, NY).
11. Frischauf, A. M., Lehrach, H., Poustka, A. & Murray, N. (1983) *J. Mol. Biol.* **170**, 827–842.
12. Kozak, M. (1984) *Nucleic Acids Res.* **12**, 857–872.
13. Thomas, M. L., Barclay, A. N., Gagnon, J. & Williams, A. F. (1985) *Cell* **41**, 83–93.
14. Omary, M. D. & Trowbridge, I. S. (1980) *J. Biol. Chem.* **255**, 1662–1669.
15. Watson, A. J. (1982) *Immunogenetics (N.Y.)* **16**, 485–490.
16. Hemmings, B. A., Aitken, A., Cohen, P., Raymond, M. & Hofmann, F. (1982) *Eur. J. Biochem.* **127**, 473–481.
17. Cohen, P. (1985) *Curr. Top. Cell. Reg.* **27**, 23–37.
18. Shen, F.-W., Tung, J.-S. & Boyse, E. A. (1986) *Immunogenetics (N.Y.)*, in press.
19. Lefrancois, L., Thomas, M. L., Bevan, M. J. & Trowbridge, I. S. (1986) *J. Exp. Med.* **163**, 1337–1342.