A NEW MEMBER OF THE IMMUNOGLOBULIN SUPERFAMILY THAT HAS A CYTOPLASMIC REGION HOMOLOGOUS TO THE LEUKOCYTE COMMON ANTIGEN

By MICHEL STREULI, NEIL X. KRUEGER, LAURIE R. HALL, STUART F. SCHLOSSMAN, and HARUO SAITO

From the Division of Tumor Immunology, Dana-Farber Cancer Institute; and the Departments of Pathology, Medicine, and Biological Chemistry and Molecular Pharmacology, Harvard Medical School, Boston, Massachusetts 02115

Leukocyte common antigens (LCAs[1]; also called T200, CD45, and in the mouse, Ly-5) are a family of high molecular weight (180–220 kD) cell surface glycoproteins found on all leukocytes (1). While the exact function of the LCA molecules is not known, they are thought to play important roles in the regulation of a variety of immune responses (1), including induction of suppressor activity (2). The cytoplasmic regions of the LCAs are composed of two homologous domains of ~300 amino acids each (3–6). Moreover, these domains are highly conserved between rodent and human LCAs (89% identity). These observations suggested that the cytoplasmic region of LCAs has an important functional role that poses a strong selective force against evolutional drift. If this is the case, then it is possible that structurally similar cytoplasmic domains are used in other molecules with similar function. In this report, we describe a gene, called *LAR*, that encodes a protein homologous to LCAs. Furthermore, the LAR protein has an extracellular region homologous to both the Ig-like and non-Ig-like domains of the neural-cell adhesion molecule (N-CAM) (7).

Materials and Methods

Isolation of Genomic and cDNA Clones. A human placenta genomic DNA library (8) was hybridized to the ^{32}P-labeled nick-translated 3.2-kb Xba I fragment isolated from the mouse LCA cDNA clone pLY-5-68 (6), in the presence of 4× SSC, 50% (vol/vol) formamide, and 10% (wt/vol) sodium dextran sulfate, at 28°C. The filters were washed at 44°C in 0.1× SSC, 0.1% SDS. Positive clones were then hybridized to the human LCA cDNA clones, LCA.6 and LCA.2 (4), under stringent conditions (hybridization at 42°C and washing at 65°C). Those clones that did not hybridize to the human LCA probes were further characterized by restriction mapping and nucleotide sequence determination. The 1-kb Sac I fragment (probe 1) of the LAR1 clone was used as a hybridization probe to screen a human tonsil lymphocyte cDNA library (4) using stringent hybridization and washing conditions. cDNA clones thus isolated were in turn used as hybridization probes to screen the same cDNA library in order to isolate cDNA clones that contain additional 5′ sequences.

This work was supported by grant AI-26598 and a Biomedical Research Support Grant, both from the National Institutes of Health, and an Investigator Award from the Cancer Research Institute (H. Saito). M. Streuli is a recipient of a postdoctoral fellowship from the Swiss National Science Foundation.

[1] *Abbreviations used in this paper:* LCA, leukocyte common antigen; N-CAM, neural-cell adhesion molecule.

Nucleotide Sequence Determination. Nucleotide sequences were determined by the method of Maxam and Gilbert (9).

Other Methods. Restriction site mapping, subcloning, nick translation, and Southern and Northern blot hybridization, are according to Maniatis et al. (10).

Results

Isolation of a Human Gene Homologous to the Leukocyte Common Antigen Gene. Six human genomic DNA clones that hybridized to LCA cDNA probes under relaxed conditions but not under a stringent condition were isolated from a human genomic DNA library. These genomic DNA clones were characterized by restriction mapping, Southern blot hybridization, and nucleotide sequence determination. Although five of the genomic clones contained short stretches of nucleotide sequences fortuitously identical to various portions of the mouse LCA cDNA, the sixth clone, named LAR1 for LCA Related, contained several open reading frames, each homologous to contiguous segments of the LCAs. Fig. 1 *A* shows the restriction map of a portion of the genomic DNA clone, LAR1. To ascertain if the *LAR* gene is expressed in lymphoid cells, poly(A)$^+$ RNAs isolated from several human B and T cell lines were examined by Northern RNA blot hybridization using probe 1 (Fig. 1 *A*). Fig. 2 *A* shows that several T cell lines express LAR mRNA, which is ~8 kb in length, while B cells express very little or no LAR mRNA. To determine the primary structure of the protein encoded by the *LAR* gene, we isolated LAR cDNA clones from a tonsil lymphocyte cDNA library using probe 1. After several rounds of cDNA walking, we isolated overlapping cDNA clones that spanned a total of 7.7 kb (Fig. 1 *B*). The total nucleotide sequences of these cDNAs showed that the LAR mRNA has a long open reading frame that would encode a protein of 1,897 amino acids (Fig. 3 *A*).

The deduced LAR amino acid sequence has typical features of membrane glyco-

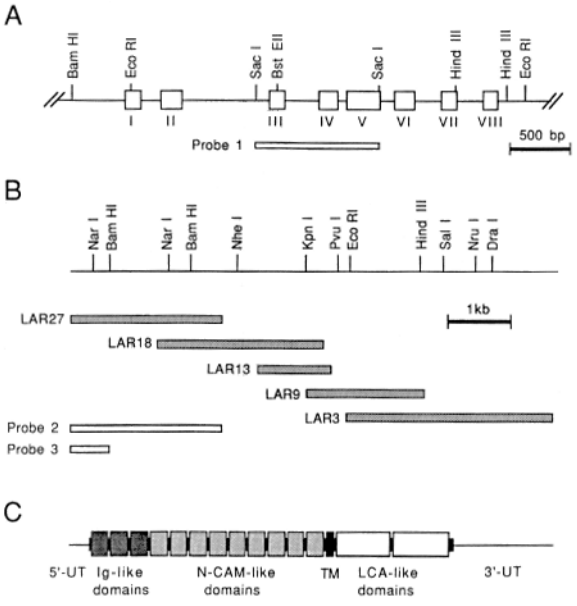

FIGURE 1. (*A*) Restriction map of a portion of the human genomic DNA clone LAR1 containing exons encoding a part of the cytoplasmic region. Boxes represent exons. The position of hybridization probe 1 is indicated by an open box. (*B*) Restriction map of overlapping human LAR cDNAs. Shaded boxes indicate the lengths of characterized LAR cDNA clones. Open boxes indicate cDNA probes used. (*C*) A schematic model of the structure of the LAR mRNA is shown. Thin lines indicate 5′ and 3′ untranslated sequences (*UT*), while the thick bar indicates the translated region. The open boxes superimposed onto the thick bar represent LCA-like domains. The shaded boxes indicate Ig-like and N-CAM-like domains. The filled box indicates the transmembrane segment.

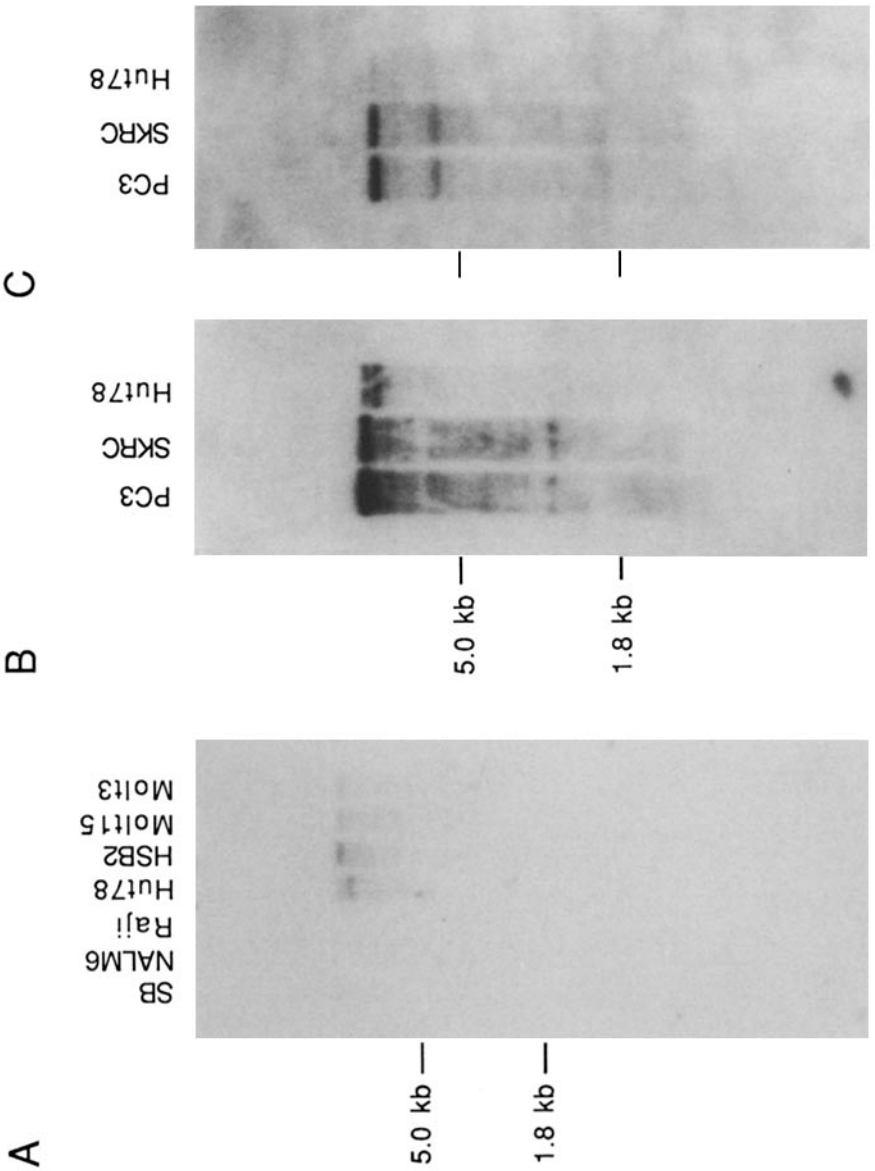

FIGURE 2. RNA blot analyses of poly(A)$^+$ RNA isolated from various cells. Probes 1–3, shown in Fig. 1, were used for the blots shown in *A–C*, respectively. RNA was extracted from human B cell lines (SB, NALM6, and Raji), T cell lines (Hut78, HSB2, Molt15, and Molt3), a prostate cell line (PC3), and a kidney cell line (SKRC). 2 μg of poly(A)$^+$ RNA was used for each line.

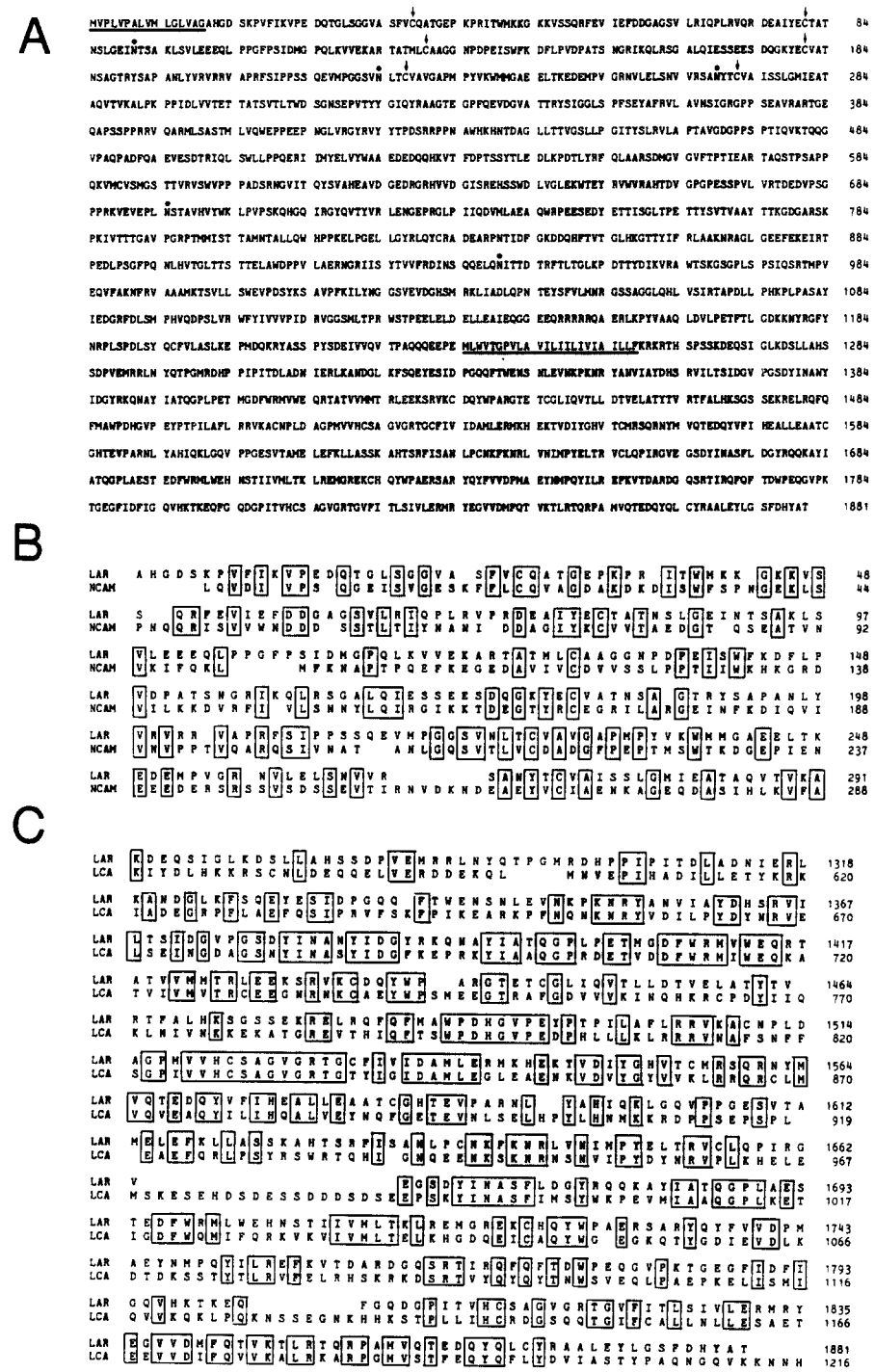

FIGURE 3. (*A*) Deduced amino acid sequence of the human LAR protein. Nucleotide sequences of the human LAR cDNA clones (Fig. 1 *B*) were determined by the method of Maxam and Gilbert (9). The numbers on the right show amino acid positions in the predicted mature form of

proteins. The NH_2-terminal sequence is highly hydrophobic, and is probably a signal peptide. Analysis using the algorithm of von Heijne (11) predicted that the most likely NH_2 terminus of the mature protein is the seventeenth amino acid, alanine. The alanine was, therefore, assigned the amino acid position 1. The LAR sequence has a second highly hydrophobic stretch (amino acid positions 1235–1258), which is likely to be a transmembrane peptide. Thus, the mature LAR protein would have a 1,234 amino acid extracellular region and a 623 amino acid cytoplasmic region, connected by a 24 amino acid transmembrane peptide. In Fig. 1 *C*, the structure of the LAR cDNA is schematically shown. The extracellular region has five potential *N*-linked glycosylation sites (indicated by dots in Fig. 3 *A*). The cytoplasmic region of the LAR protein has 38% amino acid identity to the cytoplasmic region of the LCA protein (Fig. 3 *C*).

Similarity between LCA and LAR exists also at the level of gene organization. Eight exons, arbitrarily called exons I through VIII, were identified in ~3.5 kb of the LAR1 genomic clone (Fig. 1 *A*). These exons encode a major portion of the cytoplasmic region of the LAR protein. The exon-intron organizations of the *LAR* gene segment and the corresponding region of the *LCA* gene (12) are very similar. For example, LAR exons III, IV, VII, and VIII precisely correspond to LCA exons 23, 24, 29, and 30, respectively (data not shown).

The LAR Gene Belongs to the Ig Superfamily. In contrast to the cytoplasmic regions, the extracellular regions of the LAR and LCA proteins have no significant sequence similarity. However, the NH_2-terminus sequence of LAR protein consists of three repeat units of Ig-like domains. While any member of the Ig superfamily can be aligned with this segment of the LAR protein sequence, the N-CAM sequence had the highest degree (27%) of identity (Fig. 3 *B*). N-CAM has five Ig-like domains at the NH_2 terminus, which are followed by a further segment of sequence before the transmembrane sequence (7). This extra, non-Ig-like sequence can be split into two domains, N-CAM(vi) and N-CAM(vii), of ~100 amino acids that show weak similarity to each other. In LAR there are nine similar but non-Ig-like domains, LAR(iv) through LAR(xii), after the first three Ig-like domains. The nine domains, which are ~100 amino acids in size, show convincing sequence similarity (between 16 and 33%) to N-CAM(vi) and somewhat weaker similarity to N-CAM(vii) (Fig. 4). Thus, the entire LAR extracellular sequence is made up of a total of 12 domains, which are related to N-CAM sequence.

Expression of the LAR Gene. To examine the expression of LAR mRNA, 14 cell lines (four lung, three breast, two colon, two melanoma, and one each of kidney, prostate, and myelomonocyte cell lines) were examined by dot blot hybridization using a LAR cDNA probe. The results (data not shown) demonstrated that all 14 cell lines express the LAR mRNA to various extents. Two nonlymphoid cell lines,

---

the LAR protein. Putative signal and transmembrane peptides are underlined. Cystein residues thought to be involved in intradomain disulphide linkages are indicated by arrows. The potential *N*-glycosylation sites (N-X-S or N-X-T) are indicated by black dots. The standard one-letter code was used. (*B*) A possible alignment of the three Ig-like domain sequences of the human LAR protein and the first three Ig-like sequences of the mouse N-CAM (13). The residues common to both proteins are boxed. The numbers on the right indicate amino acid positions as above. (*C*) A possible alignment of the two LCA-like domains of the human LAR protein and the corresponding region of the human LCA protein. These sequence data have been submitted to the EMBL/GenBank Data Libraries under the accession number Y00815.

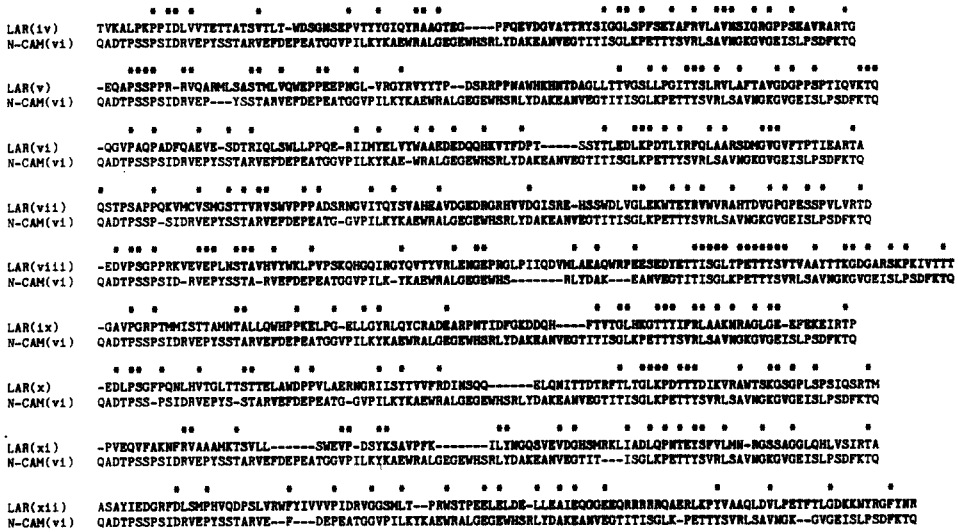

FIGURE 4. Possible alignments of the nine non-Ig-like domains of the human LAR protein and one of the non-Ig-like domains of the chicken N-CAM sequence (7). Identical amino acids are indicated by asterisks. The amino acid positions of each domain are: LAR(iv), 288–383; LAR(v), 384–482; LAR(vi), 483–576; LAR(vii), 577–678; LAR(viii), 679–791; LAR(ix), 792–885; LAR(x), 886–982; LAR(xi), 983–1070; LAR(xii), 1081–1186; and N-CAM(vi), 477–577.

namely, the prostate cell line PC3 and kidney cell line SKRC, express relatively large amounts of the 8-kb LAR mRNA (Fig. 2 *B*). Therefore, the *LAR* gene is expressed in a broad range of cell types.

The human T cell line Hut78 expresses two LAR mRNAs of slightly different lengths (Fig. 2, *A* and *B*). Because the differential splicing of the exons encoding peptides near the NH_2 terminus of the LCA molecules generates multiple forms of LCA mRNAs (4, 5, 14), it was of interest to see if these LAR mRNAs are also generated by differential splicing. To test this possibility, Northern blot analysis was performed using probe 3 (Fig. 1), which is derived from the very 5′ end of the LAR cDNA and includes the 5′ untranslated region as well as the first Ig-like domain. Fig. 2 *C* demonstrates that the LAR mRNA of PC3 and SKRC cells, and the larger LAR mRNA of Hut78 cells hybridize to probe 3, but that the smaller Hut78 LAR mRNA does not. This result demonstrates that the two LAR mRNAs are different in the 5′ regions, although this does not prove that these two mRNA are generated by differential splicing.

Discussion

What are the possible functions of the Ig-like domains and LCA-like domains of the LAR protein? The most basic function of the Ig-like domains seems to be their capacity to form dimeric structures with other Ig-like domains (15). This adhesive property can be either homophilic or heterophilic. Thus, the three Ig-like domains of the LAR protein may also interact with other Ig-like domains. If the LAR protein has homophilic properties like N-CAM, then LAR may function as a cell adhesion molecule. The fact that all of the LAR extracellular sequence can be aligned

to N-CAM may support this hypothesis. The function of the cytoplasmic regions of the LAR and LCA molecules is less clear. One possible clue, however, is the finding that LCAs are associated with the cytoskeletal component fodrin (16), which is implicated in the control of exocytosis (17). This suggests that the function of the conserved cytoplasmic domains of the LCA and LAR proteins might be to associate with fodrin. Association of transmembrane proteins and the cytoskeleton is probably essential for processes such as cell motility, cell–cell recognition, phagocytosis, endocytosis, and exocytosis.

Evolution of genes by assembly of functionally independent domains is frequently observed. The *LAR* gene presents an interesting case that brings together hitherto unrelated molecules, namely, the LCAs and the Ig superfamily. Further characterization of the *LAR* gene and its product will help to understand the function of the LCA molecules.

Summary

A human gene (*LAR*) that hybridizes to mouse leukocyte common antigen cDNA under relaxed hybridization conditions was isolated. The *LAR* gene is expressed in a broad range of cells, including T lymphocytes, kidney, and prostate cells. The structure of the protein encoded by the *LAR* gene was deduced by determining the nucleotide sequences of a 7.7-kb LAR cDNA. The putative LAR protein is composed of a 1,234 amino acid extracellular region, a 24 amino acid transmembrane segment, and a 623 amino acid cytoplasmic region. The cytoplasmic region contains two homologous domains that have extensive sequence similarity to the cytoplasmic region of the leukocyte common antigens. The NH_2-terminal region of the extracellular segment of the LAR protein contains three tandem Ig-like domains and nine non-Ig-like domains. Among the known Ig-like proteins, the LAR protein has the highest degree of similarity to neural-cell adhesion molecule. The non-Ig-like domains of the LAR protein are also similar to the non-Ig-like domains of neural-cell adhesion molecule.

We thank Drs. L. Klickstein and D. T. Fearon for the human tonsillar cDNA library, Dr. P. Leder for the human genomic DNA library, and Ali Lashgari for technical help. We are particularly grateful to the reviewer who pointed out the similarity between non-Ig-like segments of the N-CAM and LAR proteins.

Received for publication 9 June 1988 and in revised form 20 July 1988.

References

1. McMichael, A. J. 1987. Leukocyte Typing III. Oxford University Press, Oxford. 1050 pp.
2. Rudd, C. E., C. Morimoto, L. L. Wong, and S. F. Schlossman. 1987. The subdivision of the T4 (CD4) subset on the basis of the differential expression of L-C/T200 antigens. *J. Exp. Med.* 166:1758.
3. 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 M_r. *Cell.* 41:83.
4. 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.

5. Ralph, S. J., M. L. Thomas, C. C. Morton, and I. S. Trowbridge. 1987. Structural variants of human T200 glycoprotein (leukocyte common antigen). *EMBO (Eur. Mol. Biol. Organ.) J.* 6:1251.
6. 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.
7. Cunningham, B. A., J. J. Hemperly, B. A. Murray, E. A. Prediger, R. Brackenbury, and G. M. Edelman. 1987. Neural cell adhesion molecule: structure, immunoglobulin-like domains, cell surface modulation, and alternative RNA splicing. *Science (Wash. DC).* 236:799.
8. Ravetch, J. V., U. Siebenlist, S. Korsmeyer, T. Waldmann, and P. Leder. 1981. Structure of the human immunoglobulin mu locus. *Cell.* 27:538.
9. Maxam, A., and W. Gilbert. 1980. Sequencing end-labeled DNA with base-specific chemical cleavages. *Methods Enzymol.* 65:499.
10. Maniatis, T., E. F. Fritsch, and J. Sambrook. 1982. Molecular Cloning: A Laboratory Manual. Cold Spring Harbor Laboratory, Cold Spring Harbor, NY. 545 pp.
11. von Heijne, G. 1986. A new method for predicting signal sequence cleavage sites. *Nucleic Acids Res.* 14:4683.
12. Hall, R. L., M. Streuli, S. F. Schlossman, and H. Saito. 1988. Complete exon-intron organization of the human leukocyte common antigen (CD45) gene. *J. Immunol.* In press.
13. Barthels, D., M.-J. Santoni, W. Wille, C. Ruppert, J.-C. Chaix, M.-R. Hirsch, J. C. Fontecilla-Camps, and C. Goridis. 1987. Isolation and nucleotide sequence of mouse NCAM cDNA that codes for a Mr 79000 polypeptide without a membrane-spanning region. *EMBO (Eur. Mol. Biol. Organ.) J.* 6:907.
14. Saga, Y., J.-S. Tung, F.-W. Shen, and E. A. Boyse. 1987. Alternative use of 5′ exons in the specification of Ly-5 isoforms distinguishing hematopoietic cell lineages. *Proc. Natl. Acad. Sci. USA.* 84:5364.
15. Williams, A. F., and A. N. Barclay. 1988. The immunoglobulin superfamily: domains for cell surface recognition. *Annu. Rev. Immunol.* 6:381.
16. Suchard, S. J., and L. Y. W. Bourguignon. 1987. Further characterization of a fodrin-containing transmembrane complex from mouse T-lymphoma cells. *Biochem. Biophys. Acta.* 896:35.
17. Perrin, D., O. K. Langley, and D. Aunis. 1987. Reorganization of alpha-fodrin induced by stimulation in secretory cells. *Nature (Lond.).* 326:498.