Proc. Natl. Acad. Sci. USA
Vol. 84, pp. 204–208, January 1987
Immunology

Molecular cloning of Ly-1, a membrane glycoprotein of mouse T lymphocytes and a subset of B cells: Molecular homology to its human counterpart Leu-1/T1 (CD5)

(lymphocyte differentiation antigen/immunoglobulin gene superfamily/Ly-1 B-cell lineage)

HUEI-JEN S. HUANG*†, NANCY H. JONES‡, JACK L. STROMINGER§, AND LEONARD A. HERZENBERG*

*Department of Genetics, Stanford University, Stanford, CA 94305-1234; ‡Dana Farber Cancer Institute, 44 Binney Street, Boston, MA 02115; and §Department of Biochemistry and Molecular Biology, Harvard University, 7 Divinity Avenue, Cambridge, MA 02138

Contributed by Leonard A. Herzenberg, August 13, 1986

ABSTRACT **We report the isolation of cDNA clones of the mouse lymphocyte differentiation antigen Ly-1. One of these cDNA clones was confirmed to be full-length by DNA sequencing and by expression of Ly-1 by L cells transfected with this clone. Analysis of the predicted amino acid sequence indicated that the Ly-1 polypeptide is synthesized with a 23 amino acid leader and that the mature protein consists of an amino-terminal region of 347 amino acids, a transmembrane sequence of 30 residues, and a carboxyl-terminal region of 94 amino acids. The amino-terminal region appears to be divided into two subregions by a threonine- and proline-rich sequence of 23 amino acids that is highly conserved between Ly-1 and its human homologue Leu-1 (CD5) in position and amino acid composition. The first amino-terminal subregion of 111 amino acids is predicted to be arranged in a β-pleated sheet structure of six strands. The entire amino-terminal region is rich in cysteine, with all of its 22 cysteine residues conserved between Ly-1 and Leu-1. The carboxyl-terminal region has no cysteines. Ly-1 and Leu-1 are 63% identical, with a gradient of identical residues from 43% for the first amino-terminal subregion to 58% for the second amino-terminal subregion and 90% for the carboxyl-terminal region. The predicted secondary structure of the first amino-terminal subregion and identities of certain conserved residues among most members of the immunoglobulin gene superfamily suggest that Ly-1 and Leu-1 are distant members of this family.**

Ly-1 (formerly Lyt-1), a lymphocyte differentiation antigen with a molecular weight of 67,000, is expressed on nearly all murine thymocytes (1, 2). Cytotoxic-depletion studies originally suggested that Ly-1 was selectively expressed on helper T cells (3). Sensitive analyses, such as those using the fluorescence-activated cell sorter (FACS), have shown that Ly-1 is a pan-T marker present at a higher level on helper T cells than on suppressor or cytotoxic T cells (1, 2). Levels of Ly-1 within the range of Ly-1 levels on cytotoxic/suppressor T cells have been found on some B-cell tumors (4) and on a distinct subpopulation and lineage of B cells in normal animals (5, 6).

Little is known about the role of Ly-1 in lymphocyte function and/or differentiation. Antibodies against Ly-1 molecules can augment alloantigen- or mitogen-induced lymphocyte proliferation, suggesting a possible role for Ly-1 in regulating T-cell proliferation (7, 8). The enhancing effect of anti-Ly-1 antibodies on T-cell proliferation is associated with the increased secretion of interleukin 2 (IL-2) and the increased expression of the IL-2 receptor. Similar enhancing properties of anti-Ly-1 antibodies were observed for the human and rat Ly-1 homologs (9–11).

To further our understanding of the structure and function of Ly-1 and its differential expression on different subpopulations of lymphocytes, we undertook the molecular cloning of the Ly-1 gene. Originally, we planned to isolate the Ly-1 gene by a gene-transfer/hybridization-subtraction method which one of our laboratories, and others, have used successfully to clone the human Leu-2 gene (12, 13). While this work was in progress, however, a putative human Leu-1 cDNA was isolated (39). Using this clone as a probe, we isolated a number of putative Ly-1 cDNA clones. Here we describe the isolation and characterization of a full-length Ly-1 cDNA clone, present the predicted amino acid sequence and secondary structure, and compare the sequence of mouse Ly-1 with that of human Leu-1 (39).

METHODS

Genomic Transfection. Thymidine kinase (TK)-deficient L cells (LTK^- cells) were cotransfected with 20 μg of BALB/cN mouse liver or kidney genomic DNA and 1 μg of pBR322 containing the chicken TK gene, by a standard calcium phosphate precipitation procedure (14, 15). Plasmids containing cDNA inserts were transfected similarly. Primary genomic Ly-1 transfectants were analyzed and cloned by use of a FACS as described (16). To obtain secondary genomic Leu-1 transfectants, cellular DNA from primary Leu-1 transfectants (established by P. Kavathas) was used in the transfection. To obtain genomic Ly-1- and Leu-1-amplified lines, we selected the brightest 0.5% of the stained cells for several cycles.

Antibodies. The fluorescein-conjugated monoclonal anti-Ly-1 antibody 53-7.1 (1) and monoclonal anti-Leu-1 antibody L17F12 were provided by the Becton Dickinson Monoclonal Antibody Center (Mountain View, CA).

Screening of Mouse cDNA Libraries. A C57BL/6 thymocyte cDNA library in the phage vector λgt10 was plated on *Escherichia coli* C600/Hfl and probed with a nick-translated Leu-1 cDNA clone (pT1-1). Plaque hybridization was performed as described (17), with low-stringency washing in 6× SSPE with 0.1% $NaDodSO_4$ and then 4× SSPE with 0.1% $NaDodSO_4$ at 65°C. (1× SSPE = 180 mM NaCl/10 mM sodium phosphate, pH 7.7/1 mM EDTA.) Three mouse T-cell cDNA libraries, pCD-C5, MD13-10, and E1, which were constructed in an Okayama–Berg expression vector, were screened with a Ly-1 probe. Colony hybridization was performed as described (17).

Nucleotide Sequencing. DNA sequence was determined by the dideoxynucleotide termination method (18) after subclon-

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: FACS, fluorescence-activated cell sorter; kb, kilobase(s); bp, base pair(s).

†To whom reprint requests should be addressed.

ing restriction endonuclease fragments into M13 mp18 and mp19 phage vectors.

DNA and RNA Hybridization. DNA, digested with restriction endonucleases, was analyzed by the Southern blot technique (19). RNA was prepared from cell lines and tissue by the guanidinium thiocyanate procedure (20). Poly(A)$^+$ RNA was obtained by chromatography on oligo(dT)-cellulose (21). RNA was analyzed by blot hybridization (22).

Computer Analysis. Nucleotide and amino acid sequence analyses were done by using the Beckman MicroGenie programs. The National Biomedical Research Foundation protein data bank¶ was searched by the program described by Beckman MicroGenie and by the program described by Lipman and Pearson (23). The predicted secondary structure of Ly-1 was determined using the program PLANS (24).

RESULTS

Establishment of Ly-1-Amplified and Leu-1-Amplified Genomic Transfectants. Genomic transfectants that have spontaneously amplified the human Leu-2 or mouse Lyt-2 gene can be isolated by several cycles of FACS selection for the brightest fraction of fluorescein-conjugated antibody-stained cells (25). We used a similar method to obtain genomic transfectants that have amplified Leu-1 or Ly-1. We have established three such Ly-1-amplified lines (S1A5.1, K1A7.1, and K1B7.5) and three Leu-1-amplified lines (N1-1, P2-2, and Q1-3). The mean fluorescence of these cell clones is 10–20 times greater than that of the original transfectants (data not shown). These amplified transfectants were used in helping to identify Leu-1 and Ly-1 cDNA molecular clones. In corroboration of these observations at the protein level, hybridization of cDNA probes to RNA and genomic DNA blots indicated elevation of both mRNA levels and gene copy number for Leu-1 and Ly-1 in the transfectants compared to thymocytes, T-cell lines, and normal tissues (see below).

Isolation and Characterization of Ly-1 cDNA Clones. A putative Leu-1 cDNA clone isolated with synthetic oligonucleotide probes was confirmed to contain a sequence coding for the Leu-1 gene by finding that it hybridized specifically and intensely with both RNA and DNA from three Leu-1-amplified transfectants (39). We then screened 500,000 plaques of a mouse thymocyte cDNA library with this clone under conditions of low-stringency hybridization. The cDNA insert from one of the isolated clones (GL1-4.2.1) was used as a hybridization probe to analyze genomic DNA from two Ly-1-amplified genomic transfectants (S1A5.1 and K1A7.1), L cells, and a genomic Lyt-2 transfectant (L2) on a Southern blot. This insert gave very strong hybridization bands with DNA from the Ly-1-amplified transfectants and bands corresponding to the same size, but at single-copy intensity, with DNA from L cells and the Lyt-2 transfectant (Fig. 1*a*). This cDNA also hybridized strongly with RNA from the Ly-1-amplified transfectants, but not with that from L cells (Fig. 1*b*). In addition, this probe hybridized with RNA from mouse thymus and a transformed Ly-1 B-cell line (NFS-5.3) that expresses a high level of Ly-1. These data strongly suggest that the cDNA clone GL1-4.2.1 contains a sequence encoding at least a portion of the Ly-1 gene.

Since the insert of GL1-4.2.1 is only 1.4 kb long, whereas the Ly-1 mRNA both of thymocytes and of the Ly-1 B-cell line NFS-5.3 is 2.2 kb, we screened mouse cDNA library pCD-C5 using GL1-4.2.1 as a probe in order to find a full-length insert. Of 60,000 colonies screened, we obtained 16 positive clones. Restriction mapping revealed that these clones contained cDNA inserts having the same or similar 3′ ends but variable 5′ ends. The longest insert (2.0 kb) of clone C5-6.1 was subjected to nucleotide sequencing, but its inability to transfect for Ly-1 showed it was less than full-length. Thus, we used the 190-base-pair (bp) 5′ fragment of C5-6.1 to screen another two mouse pCD libraries, MD13-10 and E1. One isolated clone (MD-10) contains a 2.1-kb insert, which restriction mapping showed had an additional 100 bp at its 5′ end. Nucleotide sequencing (see below) strongly suggested that MD-10 was a full-length cDNA clone. This was confirmed by stable expression of Ly-1 by L cells that had been transfected with MD-10 (Fig. 2).

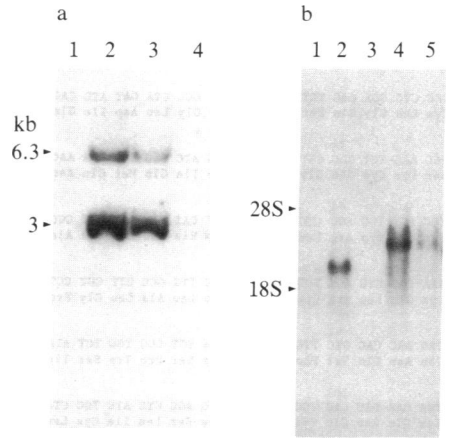

FIG. 1. (*a*) Autoradiogram of Southern blot showing that the Ly-1 gene is greatly amplified in the genomic DNA obtained from two Ly-1-amplified transfectants. DNA (15 μg per lane) from L cells (lane 4), the two Ly-1 transfectants K1A7.1 (lane 3) and S1A5.1 (lane 2), and a Lyt-2 transfectant (lane 1) was digested with *Hin*dIII, electrophoresed in a 0.8% agarose gel, transferred to nitrocellulose filter paper, and hybridized with a ^{32}P-labeled insert of GL1-4.2.1. kb, Kilobases. (*b*) Autoradiogram of blot of RNA from different sources hybridized to a nick-translated Ly-1 cDNA (GL1-4.2.1). Ten micrograms of poly(A)$^+$ RNA from thymocytes (lane 1), Ly-1$^+$ B-cell line NFS-5.3 (lane 2), and L cells (lane 3) and 2 μg of poly(A)$^+$ RNA from the Ly-1-amplified transfectants S1A5.1 (lane 4) and K1A7.1 (lane 5) were loaded. Positions of 28S and 18S rRNAs are indicated at left.

Nucleotide Sequence of Ly-1 cDNA and the Deduced Protein Sequence. The partial restriction endonuclease map, DNA sequencing strategy, nucleotide sequence, and predicted amino acid sequence of Ly-1 cDNA are shown in Fig. 3. The longest open reading frame starts at the seventh nucleotide, with a methionine codon. This reading frame extends 1482 bases, encoding a polypeptide of 494 amino acids. Additionally, this nucleotide sequence contains a 5′ untranslated sequence of 6 bases and a 3′ untranslated sequence of 457

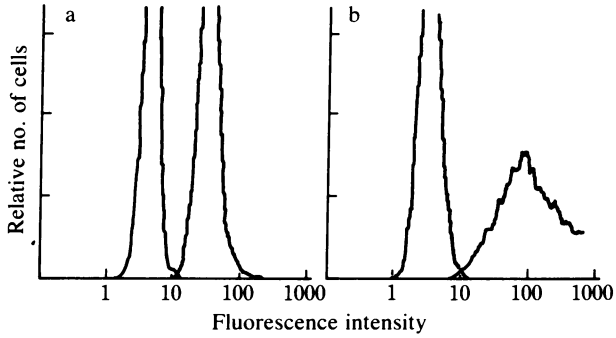

FIG. 2. FACS fluorescence-intensity histograms of two Ly-1 cDNA transfectants, MD10-1.7 (*a*) and MD10-4.1 (*b*). In each panel, the right histogram shows cells stained with anti-Ly-1 antibody; the left shows the same cells stained with anti-Lyt-2 antibody (isotype control, rat IgG2a).

¶Protein Identification Resource, National Biomedical Research Foundation (Washington, DC); release no. 8, February 1986.

a

```
                -20                                                 -12
TTA TCC ATG GAC TCC CAC GAA GTG CTG CTG GCT GCC ACA TAC CTG CTG GGA ACG CTG GCG
       Met Asp Ser His Glu Val Leu Leu Ala Ala Thr Tyr Leu Leu Gly Thr Leu Ala

    -4          1                                               11
GCT TTC TGC CTC GGA CAG TCT GGA AGG GGT GGC CTA GAT ATC CAG GTG ATG CTA AGT GGC
Ala Phe Cys Leu Gly Gln Ser Gly Arg Gly Gly Leu Asp Ile Gln Val Met Leu Ser Gly

                21                                              31
TCC AAT TCC AAG TGT CAG GGT CAA GTG GAG ATC CAG ATG GAA AAC AAG TGG AAA ACA GTG
Ser Asn Ser Lys Cys Gln Gly Gln Val Glu Ile Gln Met Glu Asn Lys Trp Lys Thr Val

                41                                              51
TGC AGT TCC AGT TGG AGG CTG AGC CAG GAC CAT TCG AAA AAT GCC CAG CAG GCC TCT GCA
Cys Ser Ser Ser Trp Arg Leu Ser Gln Asp His Ser Lys Asn Ala Gln Gln Ala Ser Ala

                61                                              71
GTG TGC AAA CAG CTG AGA TGT GGT GAC CCC TTG GCC CTT GGT CCT TTC CCT TCA TTG AAC
Val Cys Lys Gln Leu Arg Cys Gly Asp Pro Leu Ala Leu Gly Pro Phe Pro Ser Leu Asn

                81                                              91
AGA CCC CAG AAC CAG GTC TTC TGC CAA GGA TCT CCG TGG TCT ATA TCC AAC TGC AAC AAC***
Arg Pro Gln Asn Gln Val Phe Cys Gln Gly Ser Pro Trp Ser Ile Ser Asn Cys Asn Asn

               101                                             111
ACA AGT TCA CAA GAC CAG TGC CTT CCG CTG AGC CTG ATC TGC CTA GAG CCC CAG AGA ACG
Thr Ser Ser Gln Asp Gln Cys Leu Pro Leu Ser Leu Ile Cys Leu Glu Pro Gln Arg Thr

               121                                             131
ACA CCT CCA CCC ACA ACC ACC CCA CCC ACC ACC GTG CCG GAG CCC ACA GCT CCT CCC AGA
Thr Pro Pro Pro Thr Thr Thr Pro Pro Thr Thr Val Pro Glu Pro Thr Ala Pro Pro Arg

               141                                             151                ***
TTG CAG CTG GTG CCA GGA CAC GAA GGC CTG AGG TGC ACA GGT GTG GTG GAA TTC TAC AAT
Leu Gln Leu Val Pro Gly His Glu Gly Leu Arg Cys Thr Gly Val Val Glu Phe Tyr Asn

               161                                             171
GGC AGC TGG GGT GGC ACC ATC CTC TAC AAG GCC AAG GAC AGG CCC CTG GGC CTG GGG AAC
Gly Ser Trp Gly Gly Thr Ile Leu Tyr Lys Ala Lys Asp Arg Pro Leu Gly Leu Gly Asn

               181                                             191
CTC ATC TGT AAG TCT CTG CAG TGT GGC TCT TTC TTG ACA CAT CTG TCC GGG ACA GAG GCA
Leu Ile Cys Lys Ser Leu Gln Cys Gly Ser Phe Leu Thr His Leu Ser Gly Thr Glu Ala

               201                                             211
GCA GGG ACA CCA GCT CCT GCA GAG CTG AGG GAC CCC AGG CCC TTG CCA ATT CGA TGG GAG
Ala Gly Thr Pro Ala Pro Ala Glu Leu Arg Asp Pro Arg Pro Leu Pro Ile Arg Trp Glu

          ***  221                                             231
GCC CCG AAC GGG AGC TGT GTC TCA CTA CAG CAG TGC TTC CAG AAA ACA ACG GCC CAG GAG
Ala Pro Asn Gly Ser Cys Val Ser Leu Gln Gln Cys Phe Gln Lys Thr Thr Ala Gln Glu

               241                                             251
GGC GGC CAG GCG CTC ACC GTC ATC TGC TCT GAT TTC CAG CCC AAG GTT CAG AGC CGC CTG
Gly Gly Gln Ala Leu Thr Val Ile Cys Ser Asp Phe Gln Pro Lys Val Gln Ser Arg Leu

               261                                             271
GTC GGG GGC AGC AGT GTG TGT GAG GGC ATC GCT GAA GTG CGC CAG AGA TCA CAG TGG GAG
Val Gly Gly Ser Ser Val Cys Glu Gly Ile Ala Glu Val Arg Gln Arg Ser Gln Trp Glu

                                281                                             291
GCC CTG TGT GAC AGT TCT GCA GCC AGG GGC CGG GGA CGG TGG GAA GAG CTA TGT CGG GAG
Ala Leu Cys Asp Ser Ser Ala Ala Arg Gly Arg Gly Arg Trp Glu Glu Leu Cys Arg Glu

                               301                                             311
CAG CAG TGT GGC GAC CTC ATC TCC TTC CAC ACG GTG GAT GCT GAC AAA ACC TCC CCC GGG
Gln Gln Cys Gly Asp Leu Ile Ser Phe His Thr Val Asp Ala Asp Lys Thr Ser Pro Gly

                               321                                             331
TTC CTC TGT GCC CAG GAG AAG CTG TCT CAG TGT TAC CAT CTC CAG AAA AAA AAA CAC TGC
Phe Leu Cys Ala Gln Glu Lys Leu Ser Gln Cys Tyr His Leu Gln Lys Lys Lys His Cys

                               341                                             351
AAC AAG AGG GTA TTC GTC ACA TGC CAA GAC CCA AAC CCA GCG GGC CTG GCC CCA GGC ACT
Asn Lys Arg Val Phe Val Thr Cys Gln Asp Pro Asn Pro Ala Gly Leu Ala Pro Gly Thr

                               361                                             371
GTG GCA AGC ATC ATC CTG ACC CTT GTA CTC CTG GTG GTG CTG CTG GCC ATG TGT GGT CCT
Val Ala Ser Ile Ile Leu Thr Leu Val Leu Leu Val Val Leu Leu Ala Met Cys Gly Pro

                               381                                             391
CTG GTC TAC AAG AAG CTG GTG AAG AAA TTT CGT CAG AAA AAG CAG CGT CAG TGG ATT GGC
Leu Val Tyr Lys Lys Leu Val Lys Lys Phe Arg Gln Lys Lys Gln Arg Gln Trp Ile Gly

                          401  ***                                411
CCC ACA GGA GTG AAC CAG AAC ATG TCT TTC CAT CGA AGC CAC ACA GCA ACT GTG CGG TCC
Pro Thr Gly Val Asn Gln Asn Met Ser Phe His Arg Ser His Thr Ala Thr Val Arg Ser

                               421                                             431
CAA GTT GAG AAC CCA ACA GCC TCT CAT GTG GAC AAT GAA TAC AGC CAG CCT CCC AGG AAC
Gln Val Glu Asn Pro Thr Ala Ser His Val Asp Asn Glu Tyr Ser Gln Pro Pro Arg Asn

                               441                                             451
TCT CAC CTG TCT GCT TAT CCA GCT CTG GAA GGA GCC CTA CAC CGA TCT TCT ACA CAA CCT
Ser His Leu Ser Ala Tyr Pro Ala Leu Glu Gly Ala Leu His Arg Ser Ser Thr Gln Pro

                    ***        461                                             471
GAC AAC TCC TCT GAC AGT GAC TAT GAC CTG CAA GTG GCT CAG AGA CTG TAA AAG CGC CAG
Asp Asn Ser Ser Asp Ser Asp Tyr Asp Leu Gln Val Ala Gln Arg Leu

TGTCTACGCTTGGACATTTGTCCTTTGTACCTCACTTGAGGTCATCAATGAAGGCCAAGAGCATCCCCTCCATCCTGAGG

AAGCCTTTACCACTTTGAGTGCTGCCCACTTCCACTTCAGCCAGAACCTGGACAGAAGCCAAGGTCCTCTGGACAAGATC

GCGGGCCCTGCAGTGGCAGGCCAGCTCCCATATCCCCTATGGCCACGGCAGCCTGACTGCTTCCTGCAAGAGCCAACACT

CAGCAGACAGAGCCTACGAGCAACCAATCCTTGCAGAGACTCCAGTATCCAGGAGTCCCAATCCTCTAGGAGCCACTGTG

ACCAGTGGGGTTGGCTTTCTCCTTGGACTCTGACGCCATTGAACAAATGAGGAGGCGATACACCAAGTGGGTACTTACAA

GGACTTATTTCTCAAGTGTATTTTTTAATATATTACTTGTAATAA
```

b

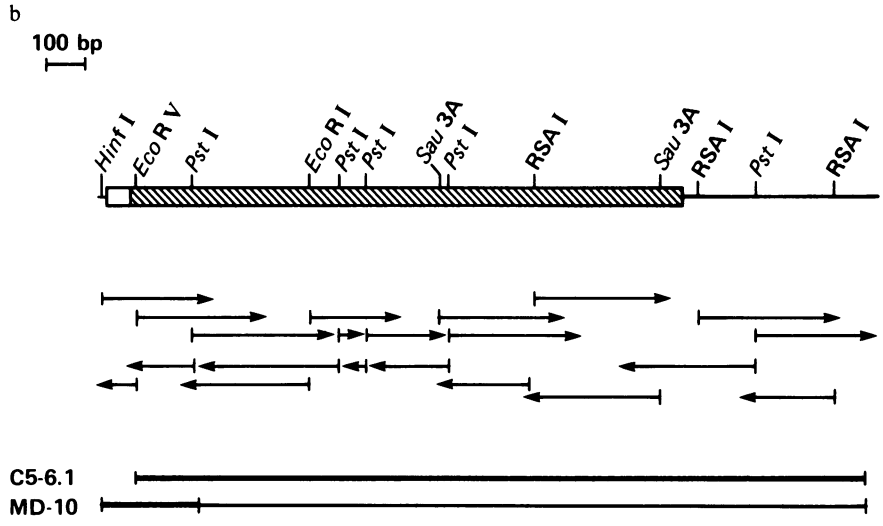

FIG. 3. (*a*) Nucleotide and amino acid sequence derived from Ly-1 cDNA clones C5-6.1 and MD-10. The transmembrane region is underlined. Possible N-linked glycosylation sites are marked by asterisks. The threonine/proline-rich region is indicated by a broken underline. The 23 amino acid leader peptide is numbered −23 to −1; the mature peptide is numbered 1 to 471. A poly(dA) tract of more than 50 bases is observed after the last nucleotide shown in the figure. (*b*) Restriction endonuclease map of Ly-1 cDNA clones. The protein-coding region is boxed. The hatched region represents the mature polypeptide after cleavage of the leader peptide. DNA sequencing strategy (direction and extent of sequence) is presented below the map. Below the sequencing strategy are the two cDNA clones used in this sequencing: the regions that were sequenced are indicated by thick lines.

bases. We do not find the polyadenylylation signal AATAAA in the 3′ untranslated region but observe a poly(dA) tract of more than 50 bases at the 3′ end.

The first 23 amino acids (from methionine to glycine) have a hydrophobicity profile characteristic of a leader sequence. Alignment of the deduced amino acid sequence of Ly-1 with that predicted from the Leu-1 nucleotide sequence and that of purified Leu-1 protein segments suggests that mature Ly-1 polypeptide begins at glutamine-24 which is defined as residue 1 (Fig. 3). Analysis of the hydrophobicity plot also reveals a region (residues 348–377) of strong hydrophobicity followed by a hydrophilic sequence (residues 378–391). We predict that this region is a transmembrane sequence. Thus, the mature polypeptide consists of an amino-terminal region of 347 amino acids, a transmembrane domain of 30 residues, and a carboxyl-terminal region of 94 amino acids. Three potential N-linked glycosylation sites are present at positions 95, 155, and 218. After the transmembrane region, there are two possible N-linked glycosylation sites at positions 402 and 457. Given that the predicted molecular weight of the mature Ly-1 polypeptide (471 amino acids) is 51,500 and the apparent molecular weight of Ly-1 is 67,000, N-linked and/or O-linked glycosylation could account for the difference of 15,500.

We predict (with help and advice from I. Kuntz and F. Cohen, University of California, San Francisco, and R. Kautz of Stanford University) that the amino-terminal region

can be further divided into two subregions by a hinge-like sequence (residues 112–134). This sequence is rich in proline and threonine residues, which are also present frequently in the hinge regions of immunoglobulin molecules (26–29). The first subregion (residues 1–111) appears to consist of a single domain with six β-strand structures, perhaps arranged in a barrel-like configuration. The amino-terminal region is very cysteine-rich. There are 8 cysteines in the first 111 residues (7.2%). There are no cysteines in the hinge region, but the second amino-terminal subregion is also cysteine-rich, with 14 cysteines out of 213 amino acids (6.6%). In contrast, the carboxyl-terminal region of 94 amino acids has no cysteines. This presumably cytoplasmic domain is rich in charged amino acids.

Comparison of Ly-1 and Leu-1. Both mature Ly-1 and Leu-1 proteins consist of 471 amino acids, with 63% identity after computer alignment of both sequences. There is a gradient of increasing homology along the two molecules: the first amino-terminal subregion has 43% amino acid identity, the second amino-terminal subregion has 58%, and the carboxyl-terminal region has 90% identity (Fig. 4). Two of three potential N-linked glycosylation sites in the amino-terminal region, as well as the two in the carboxyl-terminal region, are conserved between Ly-1 and Leu-1. Further, all 23 cysteines in each protein are conserved in very similar positions in the two sequences (Fig. 4).

DISCUSSION

In this paper we predict the protein structural outlines of the mouse lymphocyte differentiation antigen Ly-1 based on the isolation and DNA sequence of a full-length cDNA clone. Leu-1- and Ly-1-amplified genomic transfectants we generated helped to confirm that a putative Leu-1 cDNA clone in fact partially encoded Leu-1 protein. Finding that a candidate Ly-1 cDNA clone selected with this cDNA hybridized specifically and intensely with RNA from the two genomic Ly-1 transfectants and from a Ly-1 B-cell line that expresses a high level of Ly-1 on the cell surface, as well as less intensely with RNA from mouse thymus, provides evidence that this clone contains a Ly-1 cDNA insert.

With the aid of this cDNA clone and two iterative sequencing and library-screening steps, we isolated a Ly-1 cDNA clone, MD-10, that proved to be full-length both by DNA sequencing and transfection of this clone into L cells. First, the size of the cDNA insert (2.1 kb) of MD-10 is about the same as that of Ly-1 message. Second, the complete DNA sequence has a long open reading frame that could code for 494 amino acids with a methionine codon at nucleotide position 7, followed by a hydrophobic leader peptide of 23 amino acids typical of cell surface proteins. Alignment of the predicted amino acid sequence of Ly-1 with that of Leu-1 and with the amino acid sequence of purified Leu-1 protein segments strongly supports our prediction. This is further strengthened by the expression of Ly-1 by L cells transfected with MD-10.

Using the protein homology search program of Lipman and Pearson (23), we found that Ly-1 has no extensive homology with any sequence in the data bases searched. Nevertheless, we noted that the size (111 amino acids) and the predicted structure (six-β-strand structure) of the first amino-terminal domain are similar to those of an immunoglobulin domain. We found a cysteine at position 20 as well as at position 83, in good agreement with the distance between the cysteine residues that form disulfide bonds in immunoglobulins (Fig. 4). The positions for a number of amino acid residues found in Ly-1 (and Leu-1) support the idea that Ly-1 (and Leu-1) are members, albeit quite distantly related, of the immunoglobulin gene superfamily. Tryptophan, located at position 32, is highly conserved among immunoglobulins (33). Some so-

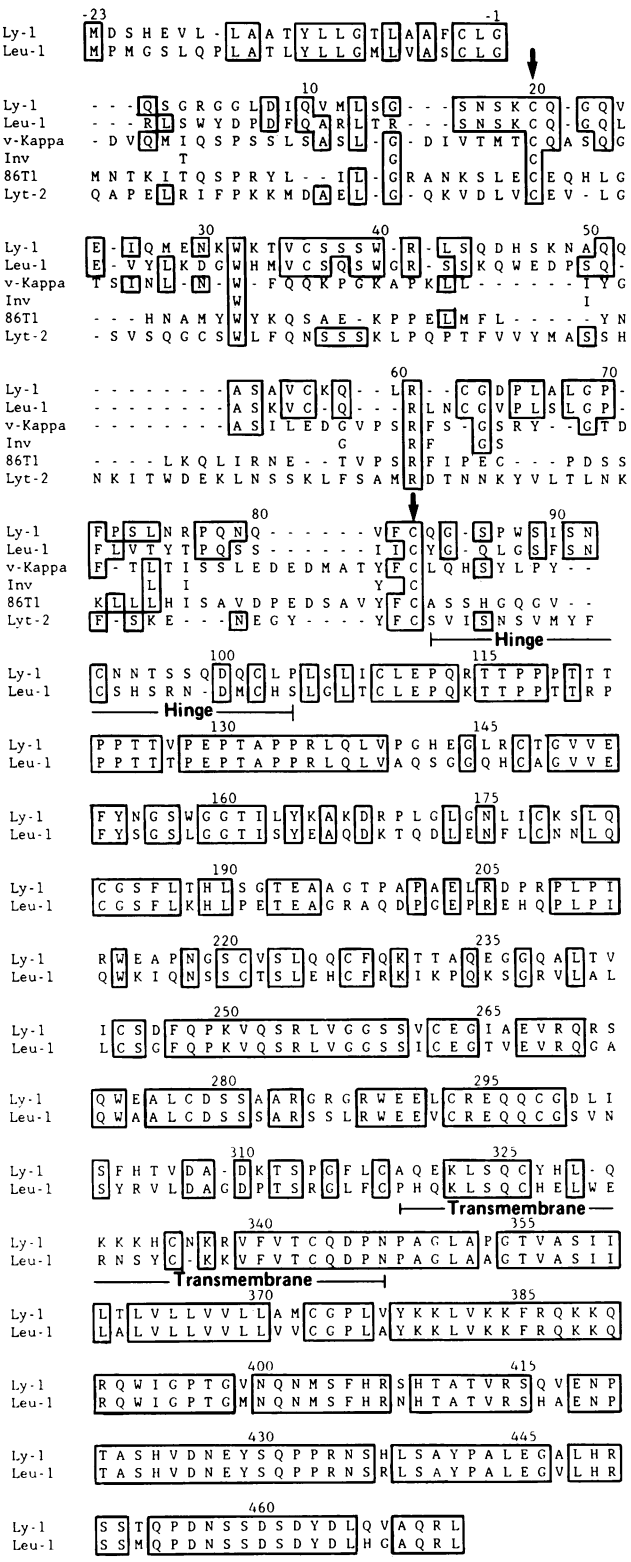

FIG. 4. Amino acid sequence alignment of Ly-1 and Leu-1 and comparison of the first amino-terminal subregion of Ly-1 and Leu-1 with a mouse κ light chain of immunoglobulin J606 (30), a mouse T-cell antigen receptor β-chain (86T1) (31), and the immunoglobulin-related domain of mouse Lyt-2 (32). The invariant (Inv) residues in the light chain variable region are included. Homologies of these proteins to Ly-1 and/or Leu-1 as well as between Ly-1 and Leu-1 are boxed. Numbering above the sequence is that of Ly-1. Cysteines that could form an immunoglobulin-interchain disulfide bond are marked by arrows. Standard one-letter amino acid abbreviations are used.

called invariant residues of the light-chain variable region are present at the appropriate positions in the first amino-

terminal domain of Ly-1 (arginine at position 61, glycine at position 63, and phenylalanine at position 71). All these residues mentioned above are conserved between Ly-1 and Leu-1. In addition, in Ly-1 but not Leu-1, we find glycine at position 15, leucine at position 74, and phenylalanine at position 82, conserved residues in the immunoglobulin family. In spite of these similarities, this subregion is unlike an immunoglobulin domain in having 8 cysteines, all of which are present also in Leu-1.

We noted that the insulin receptor (34), epidermal growth factor receptor (35), and low density lipoprotein receptor (36) all have a cysteine-rich domain, and so we compared the cysteine distribution pattern of Ly-1 with these proteins but found no similar pattern or amino acid homology.

The strong homology (90% identity) of the carboxyl-terminal region of Ly-1 and Leu-1 suggests a functional importance of this region. There is a potential tyrosine phosphorylation site at amino acid position 429 in both Ly-1 and Leu-1. The sequence surrounding this tyrosine residue (Asp-Asn-Glu-Tyr-Ser-Gln-Pro) is similar to the sequence (Asp-Asn-Glu-Tyr-Thr-Ala-Arg) surrounding the tyrosine autophosphorylation site encoded by the protooncogene c-*src* and the oncogenes *src*, *yes*, and *fgr* (37). There are also two potential threonine phosphorylation sites (His-Arg-Xaa-Xaa-Thr) at positions 410 and 453 of Ly-1. Given the evidence that Ly-1 acts as a receptor in regulating T-cell proliferation (7, 8), and given that phosphorylation of tyrosine and serine/threonine may play important physiological roles for several growth factor receptors and protooncogenes and oncogene products (37, 38), it is possible that the ligand of Ly-1 could cause phosphorylation of the carboxyl-terminal region of Ly-1.

We thank especially Dr. J. Tso for stimulating and helpful discussions. We thank Drs. H. Nakauchi, P. Kavathas, M. Tagawa, I. Kuntz, D. Tarlinton, R. Kautz, L. Herzenberg, M. Davis, H. Schulman, F. Cohen, N. Tsurushita, and D. Guinta for helpful discussions; Drs. H. Gershenfeld and I. Weissman for providing the mouse thymocyte cDNA library; Drs. K. Arai and T. Yokota (DNAX Research Institute, Palo Alto, CA) for providing three pCD libraries; Berri Benstead for excellent technical help; and L. Lloyd for excellent secretarial help. This work was supported in part by National Institutes of Health Grants GM17367, CA04681, and GM28428.

1. Ledbetter, J. A. & Herzenberg, L. A. (1979) *Immunol. Rev.* **47**, 63–90.
2. Ledbetter, J. A., Rouse, R. V., Micklem, H. S. & Herzenberg, L. A. (1980) *J. Exp. Med.* **152**, 280–295.
3. Cantor, H. & Boyse, E. A. (1975) *J. Exp. Med.* **141**, 1376–1389.
4. Lanier, L. L., Warner, N. L., Ledbetter, J. A. & Herzenberg, L. A. (1981) *J. Exp. Med.* **153**, 998–1003.
5. Hardy, R. R., Hayakawa, K., Herzenberg, L. A., Morse, H. C., III, Davidson, W. F. & Herzenberg, L. A. (1984) *Curr. Top. Microbiol. Immunol.* **113**, 231–236.
6. Hayakawa, K., Hardy, R. R., Herzenberg, L. A. & Herzenberg, L. A. (1985) *J. Exp. Med.* **161**, 1554–1568.
7. Hollander, N., Pillemer, E. & Weissman, I. (1981) *Proc. Natl. Acad. Sci. USA* **78**, 1148–1151.
8. Loydberg, L. & Shevach, E. M. (1985) *Eur. J. Immunol.* **15**, 1007–1013.
9. Ledbetter, J. A., Martin, P. J., Spooner, C. E., Wofsy, D., Tsu, T. T., Beatty, P. G. & Gladstone, P. (1985) *J. Immunol.* **135**, 2331–2336.
10. Dallman, M. S., Tomas, M. L. & Green, J. R. (1984) *Eur. J. Immunol.* **14**, 260–267.
11. Thomas, Y., Glickman, E., Demartino, J., Wang, J., Goldstein, G. & Chess, L. (1984) *J. Immunol.* **133**, 724–728.
12. Kavathas, P., Sukhatme, V. P., Herzenberg, L. A. & Parnes, J. R. (1984) *Proc. Natl. Acad. Sci. USA* **81**, 7688–7692.
13. Littman, D. R., Thomas, Y., Maddon, P. J., Chess, L. & Axel, R. (1985) *Cell* **40**, 237–246.
14. Graham, I. L., Bacchetti, S. R. & McKinnon, R. (1980) in *Introduction of Macromolecules into Viable Mammalian Cells*, eds. Baserga, R., Croce, C. & Rovera, G. (Liss, New York), pp. 2–25.
15. Wigler, M., Sweet, R., Sim, G. K., Wold, B., Pellicer, A., Lacy, E., Maniatis, T., Silverstein, S. & Axel, R. (1979) *Cell* **16**, 777–785.
16. Kavathas, P. & Herzenberg, L. A. (1983) *Proc. Natl. Acad. Sci. USA* **80**, 524–528.
17. Maniatis, T., Fritsch, E. F. & Sambrook, J. (1982) *Molecular Cloning: A Laboratory Manual* (Cold Spring Harbor Laboratory, Cold Spring Harbor, NY), pp. 309–362.
18. Sanger, F., Nicklen, S. & Coulson, A. (1977) *Proc. Natl. Acad. Sci. USA* **74**, 5463–5467.
19. Southern, E. M. (1975) *J. Mol. Biol.* **98**, 503–517.
20. Chirgwin, J. M., Przybyla, A. E., MacDonald, R. J. & Rutter, W. J. (1979) *Biochemistry* **18**, 5294–5299.
21. Aviv, H. & Leder, P. (1972) *Proc. Natl. Acad. Sci. USA* **69**, 1408–1412.
22. Thomas, P. S. (1980) *Proc. Natl. Acad. Sci. USA* **77**, 5201–5205.
23. Lipman, D. J. & Pearson, W. R. (1985) *Science* **227**, 1435–1441.
24. Cohen, F. E., Abarbanel, R. M., Kuntz, I. D. & Fletterick, R. J. (1986) *Biochemistry* **25**, 266–275.
25. Kavathas, P. & Herzenberg, L. A. (1983) *Nature (London)* **306**, 385–386.
26. Yamawaki-Kataoka, Y., Miyata, T. & Honjo, T. (1981) *Nucleic Acids Res.* **9**, 1365–1381.
27. Wels, J., Word, C. J., Rimm, D., Der-Balan, G. P., Martinez, H. M., Tucker, P. W. & Blattner, F. R. (1984) *EMBO J.* **3**, 2041–2046.
28. Honjo, T., Obata, M., Yamawaki-Kataoka, Y., Kataoka, T., Kawakami, T., Takahashi, N. & Mano, Y. (1979) *Cell* **18**, 559–568.
29. Dangl, J. L. (1986) Dissertation (Stanford University, Stanford, CA).
30. Kabat, E. A., Wu, T. T., Bilofsky, H., Reid-Miller, M. & Perry, H. (1983) *Sequences of Proteins of Immunological Interest* (U.S. Dept. of Health and Human Services, Washington, DC).
31. Hedrick, S. M., Nielsen, E. A., Kavaler, J., Cohen, D. I. & Davis, M. M. (1984) *Nature (London)* **308**, 155–158.
32. Nakauchi, H., Nolan, G. P., Hsu, C., Huang, H.-J. S., Kavathas, P. & Herzenberg, L. A. (1985) *Proc. Natl. Acad. Sci. USA* **82**, 5126–5130.
33. Williams, A. F., Barclay, A. N., Clark, M. J. & Gagnon, J. (1985) in *Proceedings of the Sigrid Juselius Symposium: Gene Expression during Normal and Malignant Differentiation*, eds. Andersson, L. C., Gahmberg, C. G. & Ekblom, P. (Academic, New York), pp. 125–138.
34. Ebina, Y., Ellis, L., Jarnagin, K., Edery, M., Graf, L., Clauser, E., Ou, J.-H., Masiary, F., Kan, Y. M., Goldfine, J. D., Roth, R. A. & Ruter, W. J. (1985) *Cell* **40**, 747–758.
35. Ullrich, A., Coussens, L., Hayflick, J. S., Dull, T. T., Gray, A., Tom, A. W., Lee, J., Yarden, Y., Libermann, T. A., Schlessinger, J., Downward, J., Mayes, E. L. V., Whittle, N., Waterfield, M. D. & Seeburg, P. H. (1984) *Nature (London)* **309**, 418–425.
36. Yamamoto, T., Davis, C. G., Brown, M. S., Schneider, W. J., Casey, M. L., Goldstein, J. L. & Russel, D. W. (1984) *Cell* **39**, 27–38.
37. Hunter, T. & Cooper, J. A. (1985) *Annu. Rev. Biochem.* **54**, 897–930.
38. Lin, C. R., Chen, W. S., Lugar, C. S., Carpenter, C. D., Gill, G. N., Evans, R. M. & Rosenfeld, M. G. (1986) *Cell* **44**, 839–848.
39. Jones, N. H., Clabby, M. L., Dialynas, D. P., Huang, H.-J. S., Herzenberg, L. A. & Strominger, J. L. (1986) *Nature (London)* **323**, 346–349.