

---

**Isolation and characterization of the gene for the murine T cell differentiation antigen and immunoglobulin-related molecule, *Lyt-2***

---

Hiromitsu Nakauchi<sup>1</sup>, Masatoshi Tagawa, Garry P.Nolan and Leonard A.Herzenberg<sup>2</sup>

---

Department of Genetics, Stanford University, Stanford, CA 94305, USA

---

Received December 16, 1986; Revised and Accepted April 20, 1987

Accession number Y00157

---

**ABSTRACT**

We present here the sequence of the 5310 base pair Hind III-cleaved genomic DNA segment that includes the gene for the *Lyt-2*, a murine differentiation antigen expressed on most immature T lymphocytes as well as the cytotoxic suppressor T cell subset. We also present the complete intron/exon structure of *Lyt-2*. There are five exons: a fused leader and immunoglobulin variable region like exon, a hinge region exon, a transmembrane exon and two alternatively spliced intracytoplasmic exons (alternative splicing of these exons yields the 38 kDa alpha and 34 kDa alpha' *Lyt-2* polypeptides). The promoter region contains a "TATA" box and sequences homologous to the putative immunoglobulin transcriptional control elements *cd/pd*. S1 protection analysis reveals that thymocytes, T cells from lymph nodes, and a *Lyt-2* transfectant obtained by introduction of total genomic DNA have the same initiation site. In the 3' region, there is a polyadenylation signal sequence after a 700 bp long 3' untranslated region.

**INTRODUCTION**

The murine T lymphocyte differentiation antigen *Lyt-2* is a membrane glycoprotein co-expressed with *Lyt-3* on immature T cells as well as a distinct subset of T cells: cytotoxic and suppressor cells (1-3). The postulated function of the *Lyt-2,3* complex, based on blocking experiments with monoclonal antibodies against *Lyt-2*, is to serve as an accessory molecule to aid in the binding of cytotoxic T cells (CTL) to non-polymorphic region(s) of the major histocompatibility complex class I molecules expressed on all target cells (4-9). The human homolog *Leu-2* has similar functional characteristics and tissue distribution as *Lyt-2* (10-12). In thymocytes, the *Lyt-2* determinant is found on two polypeptides of 38kDa (alpha) and 34kDa (alpha'), whereas lymph node T cells express only the 38 kDa polypeptide (13-15). This represents a maturation phenotype for *Lyt-2* expressing cells. We, and others (16-18) have reported the sequence of two *Lyt-2* cDNAs which arise from differential splicing. In addition, we have formally proven that the differential splicing of the primary transcript is responsible for the two differently sized *Lyt-2* molecules by utilizing cDNA expression vectors (17). We have also reported that the *Lyt-2* molecule is homologous to *Leu-2* and belongs to

the immunoglobulin gene super family (16). In this paper, we report the complete sequence, the intron/exon organization and some control elements of the *Lyt-2.2* gene.

### MATERIALS AND METHODS

**Lyt-2 genomic cloning:** An Mbo I partial genomic library of B10.A liver DNA, in lamda J1 phage (19), was screened with the insert of pLY2C-1 (16) as a hybridization probe. Plaque hybridization was performed under the conditions described previously (16). Positive clones were picked and DNA was prepared by standard methods. The DNA was cleaved with several restriction enzymes and electrophoresed on a 0.7of L(TK<sup>-</sup>) cells with *Lyt-2* genomic clones was performed as described (20).

**DNA Sequence Analysis:** The isolated 5.3-Kbp Hind III fragment digested with Sau3AI, Alu I, and HaeIII was shotgun cloned into mp18 and mp19 phage vector. Single stranded phage DNA was sequenced by the dideoxynucleotide chain termination technique of Sanger, et al. (21) as modified by Messing (22). Forced subcloning of genomic fragments with Aha III, Bam HI, Eco RI, Hind III, Pst I, Xba I, Xho I, was also used to construct the sequence. Analysis of the sequence data was performed using BIONET (NIH Grant 1U41 RRO01681-01) and STADEN (23,24) programs on a VAX 780 computer, and the Beckman MicroGenie sequence analysis program.

**RNA Extraction and S1 Nuclease Protection Assay:** Total RNA and poly A<sup>+</sup> RNA was obtained as described (25). <sup>32</sup>P-labeled single strand probe was synthesized from M13 mp template based on Burke's procedure (26). Annealing of probe fragments with RNAs and S1 nuclease (Pharmacia P-L Biochemicals) digestions were performed as described by Favalloro, et al. (27). The <sup>32</sup>P-labeled probes made by this procedure were also used for plaque hybridization.

### RESULTS

#### Isolation of genomic DNA encoding *Lyt-2.2*

We screened an Mbo I partial genomic library of B10.A liver DNA with the insert of the previously characterized partial *Lyt-2* cDNA clone (pLY2C-1) and isolated a lambda j1 phage (L-6CCA) containing a 15-Kbp insert. Hybridization with pLY2C-1 as well as the full length cDNA clone, pLY2C-22 (16), suggested that the *Lyt-2* gene is located on a single 5.3-Kb Hind III fragment. This 5.3-kb Hind III fragment was subcloned into the unique Hind III site of pBR322 (p6CCA). In order to establish that the insert of p6CCA contains the entire functional *Lyt-2* gene, DNA-mediated gene transfer was performed. About 25% of TK<sup>+</sup>

(thymidine kinase) L cells cotransfected with either the L-6CCA or p6CCA DNA were shown to be positive for *Lyt-2* by FACS (Fluorescence Activated Cell Sorter) analysis (16). The presence of *Lyt-2* molecule was confirmed by immunoprecipitation, showing both 38Kd and 34Kd polypeptides (16). Thus, *Lyt-2* alpha and alpha' chains can be expressed as homo- and heterodimers in the absence of *Lyt-3*.

#### Nucleotide sequence and exon/intron structure of the *Lyt-2.2* gene

Comparison of the p6CCA sequence with that of full length cDNA (pLY2C-22) (16) reveals the intron/exon boundaries shown in Figure 1a and Table I. The *Lyt-2* gene has five exons (Figure 1b): a fused leader and immunoglobulin variable region like exon, a hinge region exon, a transmembrane exon and two intracytoplasmic exons. Thus, each exon corresponds approximately to a functional domain of the molecule; this is similar to other members of the Immunoglobulin superfamily (28), except that *Lyt-2* does not have a separate leader exon as do all other sequenced immunoglobulin super family members. We have previously shown that the 38 kDa alpha *Lyt-2* peptide is derived from the splicing of all five exons to form an alpha-encoding mRNA and that the

34 kDa alpha' *Lyt-2* peptide is formed by the splicing of the transmembrane donor site directly to the fifth exon, thus deleting exon 4 from the mRNA (17). Loss of this 31 bp exon effects a frameshift and results in the 34 Kd alpha' *Lyt-2* peptide.

Nucleotides at the intron/exon boundaries which conform to consensus splice donor and acceptor sequences (29) are shown in Table I. Only the transmembrane donor, which alternately splices to the fourth or fifth exon, does not have the AG motif on the exon side of the boundary. We note that the mRNA splicing between exons 1 and 2, 2 and 3, 3 and 4, and 3 and 5 occurs between the first and second nucleotides of a codon triplet, as with most immunoglobulin and Class I MHC splice junctions (30). Splicing between exon 4 and 5 joins the second and third nucleotides of a triplet, similar to an IgM membrane-bound exon. (31)

#### Determination of 5' end of *Lyt-2* message and putative transcriptional regulatory elements:

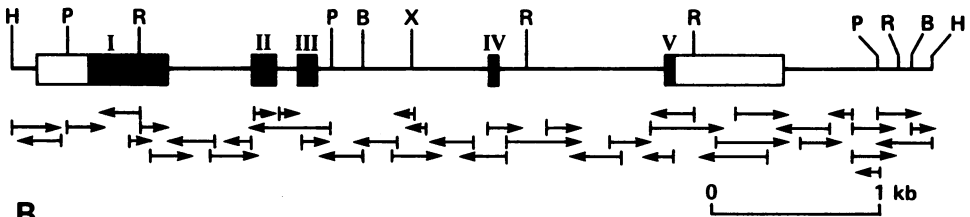
A Hinf I-Hinf I fragment (507-559) was used to prime a Hind III- Eco RI (1-745) DNA fragment (which spans the ATG start codon and putative promoter region) cloned into mp19, in M13 primer extension, to produce a <sup>32</sup>P-labeled single-strand probe (probe structures in Figure 2a). This probe was hybridized to RNA from thymus, lymph nodes and total genomic *Lyt-2* transfectants and then subjected to S1 analysis (Figure 2b). All RNAs protected a 430 bp band (data for lymph nodes not shown). Thymus RNA protected several other bands which, since they were not consistently found in several separate RNA preparations and S1 experiments, are considered to be due to partial RNA degradation. We also identified a 270

A

AAGCTTGGGACCCAGCAAGCTATCTACAAGAAAAAGAGGAGACTGGCCA 54  
 TCGCTCCGCTCATGCTCTGCTGACTCATATTTTTTTTATTAAGATTAACT 108  
 TTTATCTTTCTTTTTCAGCCACTCTGCAGACAAAGAAATCTGAAGCTCGCT 162  
 TCGAAGTCTTTCTCTTCCGCAATCTGAGTTTCAATCTGCCACTTGGCTCT 216  
 GCAAGGTCATCTCACTCGACTCCGAGCCCTTAAAGGCTGGTGACACT 270  
 CTTTCTGGGAGCTTTGGGTCACATCATATCTGAGCAGCTGGAAATCTGCTCT 324  
 TCGAGCTCTGCTAAAGAGCAGTTCCCGACCTCAGCCCTCCCGACCGCA 378  
 CCTCTCGGCGCTCTCTCGGCGCCCTCCCTAGAGCCCTAGCTTGAACATAAG 432  
 METAlaSerProLeuThrArgPheLeuSerL 486  
 TCGTCTGCTGGAGCACCACCTGCGCTCAGCCTGACCGCCTTCTCTGCGC  
 euAnLeuLeuLeuLeuGluSerIleIleLeuGlySerGlyGluAlaLysP 540  
 TGAACCTGCTGCTGGGAGCTGATGATTCTTGGGACTGGAGAAGCTAAGC  
 roClnAlaProGluLeuArgIlePheProLysLysMETAsPAlaGluLeuGlyC 594  
 CACAGCCACCGCAACTCCGAATCTTCCAAAGAAAATGACCGCCGAACTGGCT  
 InLysValAspLeuValCysGluValLeuGlySerValSerIleGlyCysSerF 648  
 AGAAGTGGACCTGATGTAAGTCTTGGGCTCCGTTTCCGAAAGGATGCTCT  
 rpLeuPheGlnAsnSerSerLysLeuProGlnProThrPheValTyrMETA 702  
 GCTTCTCGAAGACCTGAGCTCAACTCCCGAGCCAGCTGCTTTATATGCT  
 laSerSerHisAsnLysIleThrTrpAspGluLysLeuAsnSerSerLysLeuP 756  
 CTTGATCCCAAGCAAGATAAGCTGGAGAGAGCTGATTCCTGCAACTGT  
 heSerAlaMETArgAspThrAsnAsnLysTyrValLeuThrLeuAsnLysPheS 810  
 TTTTCCGATGAGGCAAGCAATAAAGTACCTTCTCACCTCGAAGCACTTCA  
 erLysGluAsnGluGlyTyrPheCysSerValIleSerAsnSerValMET 864  
 GCAAGAAAACCAAGGCTACTATTTCTGCTCCTCAGCAGCACTCGCTGATG  
 yPheSerSerValValProValLeuGlnLys 918  
 ACTTCAGTTCTCTGCTGCGACTCTCAGCAAAATTTGGAAGTGCAGCTCCCG  
 TCTGCTGGGTTTTAGGGCTCAGACTGAAACATTTCCAGCCAGATCCCTCTC 972  
 TTAAGTCTCTTGGGCGCTTGGAGGTATTCTATACCTGCTGCGCAATTAGAC 1026  
 ACCAAAAAGTTGAGCTGAGCTGGCTGGGACTGCGAAGTATTTTAAAGCC 1080  
 CCGTGAATCGGCTGAGCTGCGTGCAGCTGTGATCTCGCCGCTTGGCTGCC 1134  
 AGDAGTAAATTTGGTATTGACTCAGCTTGGCTGGTGGAGCCGACGCTCC 1188  
 TTAGATGATCTAGGCACATGGCCCAATCACTCACTGCTCCGCGCTGGCCA 1242  
 GCAAGGAGTAGCCAGCAGCTATCCAGGGCTCTCTTCTCTCTAGTAAAG 1296  
 CTAGCAAGGGCTCCTCTTCTGGTAAAGCTAAGTGGAAAGGGCTGCTGATC 1350  
 alAsnSerThr 1404  
 CACAATTCGAGCTGACCCGAGCCACTATTCTTTTGGAGTGAAGCTACTACT  
 ThrThrLysProValLeuArgThrProSerProValHisProThrGlyThrSer 1458  
 ACTACAGGCACTGCTGGCACTGCTCACTCTGCTGCAAGTCCAGGCACTCT  
 GlnProGlnArgProGluAspCysArgProArgGlySerV 1512  
 CAGCCCGAGAGCAGAAAGATTCTGGCCCGCTGGCTGAGCTACTGCTCTCAA  
 GCSAATGAGTAAAGCAGTACCTGGGCGAAAATCTGATCTTGGAGGGCAG 1566  
 ACCTGGACGGGAGCGCTGCTGGGCGAGCTTGGGCGCTCCGACAGGCTTTC 1620  
 allLysGlyThrGlyLeuAspPheAla 1674  
 TAAATTCGATTCATGGGACTCTCACTGCAAGGGCAGCGGATGGACTCGCC  
 CysAspIleTyrIleTrpAlaProLeuAlaGlyIleCysValAlaLeuLeuLeu 1728  
 TGTGATTAATGATCTGGGACCTTGGCCGATCTGGGCGGCTTCTGCTGCT  
 SerLeuIleIleThrLeuIleCysTyrHisA 1782  
 TCTTGATGATGACTCTGATCTGCTAGCAGCAATGTTCTGGGAGTCCGCCCA  
 TCGACGGCCAGCCGCGCCGAGCCGACCGCCGCGGCAAGAGAGTGGTTGCTGCT 1836  
 CAGGAGGAGCTCTGAGCTTTGCTGCTCTCGACAGCACTTCAAGGAGATTACT 1890  
 TTTAGTTGGGAAAAGGGGAGGGGCTTCCAGCTGACTATAGGACCACTGTA 1944  
 CTGGGTTCTTTCTTCAAGCCAGGACAGGTATACGACCCCACTGCTCTTCCA 1998  
 CCGCTTGGCACTTTTCAGAGAGCCGCTTCTGAGGATCCCAAGAGAGCC 2052  
 AAATCAGCACTAGCAGCCAGCACTAAGTAAAGCAGCTTCAAGGAGGCTTAC 2106  
 ATACCGTCAGACAGGCGCCAGAGGCTGAGTCAACCACTAGCTGCGCTTGCAG 2160  
 ATTTGCTTAATCCGACCTCAGAAACCACTCAGCTGCTGAGCTATGAGTAG 2214  
 TTACAGCCAGGAGCTCAAGATTCGAGAGACACTCAGTACTGATGATTAGAT 2268  
 GCGCTCCCTCCGCAAGAGGTTTCTTTCAATGTTCTGACTATGCTAGT 2322  
 ACTACTAGGAGCTAGGATAACCCGAACTTCTGATGCTGCTCTGCTGCT 2376  
 CCTCTATTTGGAGCTCTGGGCAAGGGCAGTGGCCAGCAAAATTTAAA 2430  
 ATGTTTTCTTTATGATATTATGATATTATTTATGATTTTAAAG 2484

CGAGGCTTCTCTGCTGAGCCCTGGCTGCTGCTGAGCTCAAGCTTACAGCAGGC 2538  
 TGGGCTCAAACTCAGAAAAGCCGCTGCTTCTGCTTCAATCTTGGATATAA 2592  
 AGCGTGGTGCACAGCCGCTGCTCAATTTTTTATTTTAACTGAACTTACG 2646  
 GAAGAGTAGGCAAGTAGTCTTCACTCAAGCTTCAAGACTCTGAGACTCA 2700  
 AAGTCTCGACAGGAAAGCTTAGAAGCCGATCTCGGAGCTTATCTATCTC 2754  
 rgSerArgLysArgValCysLysCysProAr 2808  
 CCTTCTCTCTCTGCTAGCGGACCGAAAAGCTGTTGCAAACTCCGAGGTGACT  
 AAGCTCTGCTAGTCTGCTGCTGCGGCGGCTGAGTCTGCTGCTGCTGCTGCT 2862  
 AGCTTTGGAACTTCTCTAGACAGAGAGCTCAAGCAGCTGCTGCTGCTGCTG 2916  
 AGGGACAGGACTCAGCCAGACTCTGAAGTACAGAGCAGAAAGCTGGAG 2970  
 GCTGCTGAAATCTTCTGAGAAAAGGACAGCAGACAGCTGCTGAGTGGAC 3024  
 AGCAGGACGACTGAGTGGAGGGACTAGCTGAAGCGCAAGGCTTTCAGTAT 3078  
 TTTGAGGACTGAGTCACTGCTGACTGCTTCTGCAAGCAATTAATCTGCTTA 3132  
 CTTTTCTGAATCTCAATTTGAAATGTAAATAGAAAAGCCGCTCTCCGCA 3186  
 CTAAACCCCGAGCTAGAAATGACTTCCCTAGACACTCAGGCTCAAGCT 3240  
 CAGCTTAAAGCAGAGGCTTCTAGAGGGCACTGAGCTCAGCTCTGCTGCTG 3294  
 GAGTAAAGCTCAGCAGCTGCTGCTGCTGAGGCTGACTCAGCTCTCTGGAT 3348  
 TCTCAGTTTGGCTCAGGATAGACTCAGCCAGCCGCTGACTTAAACACA 3402  
 GCTTAATCTCAGCTCGGACTCAGAGCTCAGAGCCACTAGTGGCAGTAG 3456  
 GCTACTCAGCTGCTGAGGACTTCTAGAGGCTGCTGAGCTCAGCCAGGCTCTTAAACTCT 3510  
 CTGAGCAGTGCAGCCCTCTTGGTGCCTTGGAGCAAGAAAGGCTGGC 3564  
 TGAGTAACTTCACTTAATGCTTAAAGCAAAATCTCAAGCCCACTAAAAAC 3618  
 TGAGTGTGCTGCTGCGCGCGCTGCGCGCGCTGCTGCTGCTGCTGCTGCT 3672  
 TAAAGAGTTCAGCCACATGACTGACTGCTGCTGCTGCTGCTGCTGCTGCT 3726  
 gProLeuValArgIleGlnGlyLysProArgProSerLysIleVal\* 3780  
 CAGCCCGCTCAGTCAAGCAGCAAGGCAAGCCAGCTCTCAGAAAATGCTGA  
 AAAAGCCAGCCGAGGAGCTAGCACTAGACTGACTCAGAGACTCTCTGCT 3834  
 GCAAGCAGCCGCGCTCTTCTGAACTTCTGCTGCTGCTGCTGCTGCTGCT 3888  
 TGTATTTTGTAGTGGGCTGATGGGCAAGCTCTTCTTTCTTAAATTC 3942  
 TCTTTCAGCAAAAGCTGCTGATGATGATGATGATGATGATGATGATGATG 3996  
 AACACAGAGTAGTCTGCTGCTGCGGCTGCTGCTGCTGCTGCTGCTGCTG 4050  
 CACTTGATATAACTCTGCTGCGCCATATAGAGAGAGCTGAAGTACAGA 4104  
 GGTGAAAAGCAGGAGTCTGAGCCGAGCAAGCAAGCTAGAGAGCCCT 4158  
 CACTGACTGCTGGGAGTCTGAGCCCGCTGCGCTGCTGCTGCTGCTGCTGCT 4212  
 GCTCTTTAGTATAGTACTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 4266  
 AATCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 4320  
 CAGGAGAACTGCTTGGGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 4374  
 GGGAAATCTGGGACTGCTGAGGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 4428  
 TCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 4482  
 CAGACCTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 4536  
 GTAGGTTAACTGAAACAGATGACAGAGAGGCGCTGATCTGAACTG 4590  
 AATGCTGAGTCAAGCTGGGACTTTTTATAACAGAAAATTCAGGGGGGA 4644  
 TATCTCAGCAGCAGAGTCACTGAAATTTACAGATACAGAGAGTACTT 4698  
 GGAAGAAATTTGAGGAACTAGATAAATTTTAAACAGAGATTACAGGCT 4752  
 GCTTAGATCACTGAGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 4806  
 CAGCAGGGGCTAAATCTTGGGAGCCCTCAAGAAATGCTGCTGCTGCTGCT 4860  
 CCTAAATCTAGGCTGCTTAAATCTTCTGCTGCTGCTGCTGCTGCTGCT 4914  
 TCGAGGAGTATCTCACTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 4968  
 TTTTATCTGAGGAACTCCGAGCTTCTGCTGAGCAGCAGTGAAGTAAAG 5022  
 CTGTTTTATGATGCTTAATGACTGCTGCTGCTGCTGCTGCTGCTGCTGCT 5076  
 ATCTGGGACTGCTGAGAGATGGGGGCTGCTGCTGCTGCTGCTGCTGCTGCT 5130  
 CTATGTCAGATGAACTTAAAGGCACTCAAGTAAAGTAACTGCTTAAAGA 5184  
 GCGCAGTGGATCAGTACTAGACTAAGCCGAGGAGTGGAAAATTAATTAAG 5238  
 TTTAGAGGAAAGCCAGCTCAGAGGAAATTTCTGCAAACTGCTTCTGCTG 5292  
 TGGCAGCTTCCAGCTT

Poly A



**B**

Figure 1: Nucleotide sequence and genomic organization of the Lyt-2.2 gene

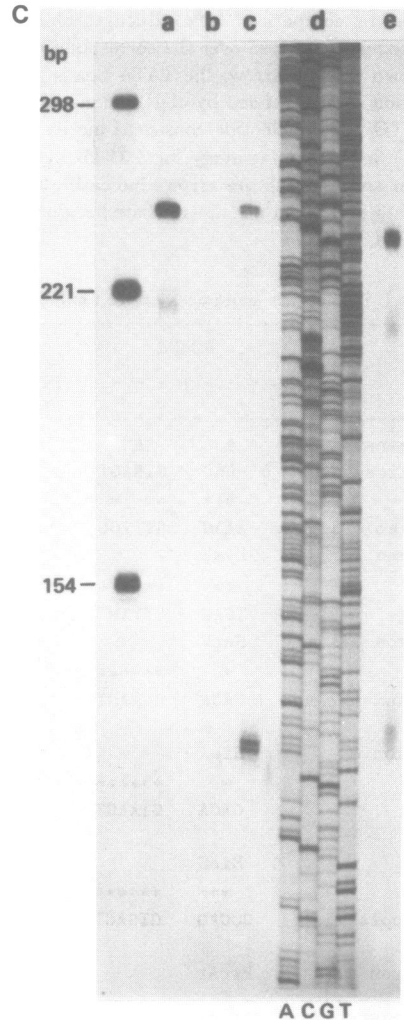
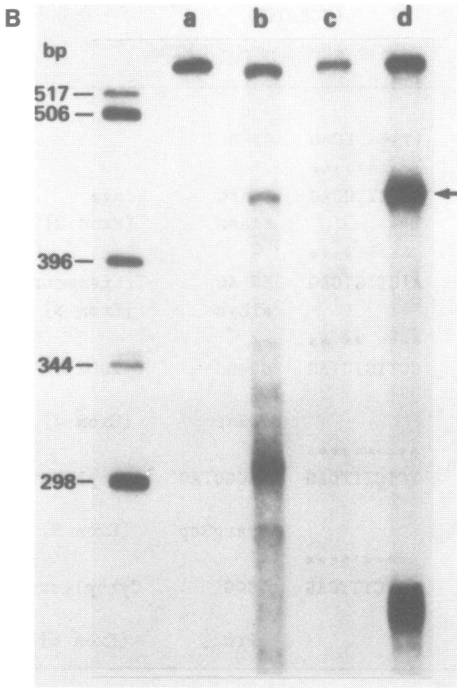
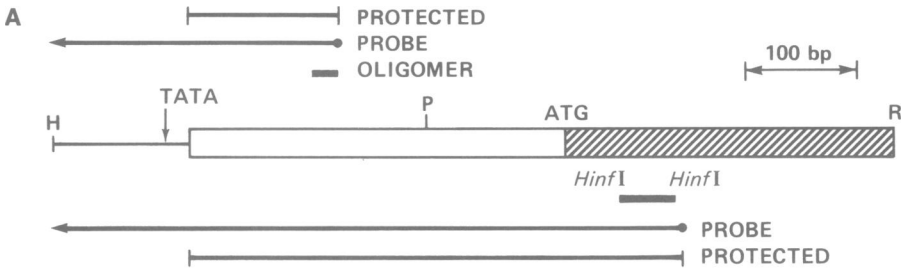
A) The nucleotide sequence of the 5310 base pair genomic Lyt-2.2 is shown. Exonic regions, derived by comparison to a full-length alpha cDNA clone, are noted by the predicted amino acid sequence written over the corresponding nucleic acid sequence. Start site of transcription is shown with an arrow; the TATA box is underlined and cd/pd-related sequences are boxed. The poly A site defined by cDNA sequencing is indicated with an arrow; the AATTAAG and CACTG pentanucleotide consensus motifs are indicated by a broken line above the sequence.

B) Sequencing strategy for 5.31-kb Lyt-2 gene. A schematic of the genomic structure is shown and beneath are arrows indicating the direction of sequencing. Black and open boxes indicate each exon and untranslated regions, respectively. H; Hind III, P; Pst I, R; EcoRI, B; BamH I, X; Xba I

Table I. Nucleotide sequences around splice junctions of the Lyt-2 gene

	DONOR			ACCEPTOR		
	EXON	INTRON	INTRON	EXON	EXON	
Consensus splice	A 5'-CAG ***	A GTGAGT ** *	... ..	(Y)G ICAG *****	G GT-3' *	
Variable-like [Exon 1]	AAAG LysV ***	GTTTGG ***	... ..	GCTTTTCAG **** **	TGAAC alAsn *	Hinge [Exon 2]
Hinge [Exon 2]	TCAG SerV *	GTTAGT *****	... ..	ATCTTTCAG **** ** **	TGAAG alLys **	Transmembrane [Exon 3]
Transmembrane 1 [Exon 3]	CACA HisA *	GTAAGT *****	... ..	CCTTGTCTAG *****	GGAGC rgSer **	Cytoplasmic [Exon 4]
1'	CACA HisS ***	GTAAGT *****	... ..	TCTCTTCAG *****	GCCGCTAG erArgStp *	Cytoplasmic [Exon 5]
Cytoplasmic 2 [Exon 4]	CCCAG ProAr	GTGAGT	... ..	TCTCTTCAG	GCCG gPro	Cytoplasmic [Exon 5]

Y, unspecified pyrimidine nucleotide  
 N, Unspecified nucleotide  
 \*, Identity to consensus splice sequence



bp protected band in the total genomic transfectant, possibly due to aberrant transcription initiation. To delineate the initiation site, a synthetic oligomer (CTCAAGGGTCCGGGG) complementary to nucleotides 228 to 242 was used to prime the M13 vector containing the Hind III to Pst I fragment for preparation of single-strand probe. Total RNA from thymocytes and genomic transfectants both protect the oligonucleotide-primed probe to give a 5' initiation site at nucleotides 123 and 124 in the genomic sequence, thus localizing at least one transcription initiation event to this site (Figure 2c). Position 123/124 is compatible with the 430 bp protected band in the localization experiment described above. Using this initiation site as a landmark, we searched and identified a "TATA" box (TATTAA) 29bp upstream. We found no other "TATA" box sequence or ATG start codon between this initiation site and the methionine start codon defined as the translational initiation point for *Lyt-2*. Additionally, no sequence corresponding to a presumptive "CAAT" box could be found.

Since it is clear that *Lyt-2* is evolutionarily related to immunoglobulin molecules, we searched our genomic sequence for DNA motifs homologous to the defined, though putative, immunoglobulin regulatory elements *cd/pd* (32,33). A nucleotide stretch bearing homology to the *cd* consensus sequence and the SV40 enhancer is found beginning at position 308 in Figure 1a with the alignment of homology shown in Table 2a. Downstream of this sequence, starting at position 326 in Figure 1a, there is a DNA motif with homology to *pd* of immunoglobulins, shown in Table 2b.

#### Determination of 3' end of *Lyt-2* RNA:

Sequencing of alpha cDNA clone 30-3 localized a poly A addition site to nucleotide 4476

#### Figure 2: S1 nuclease analysis of transcriptional initiation of *Lyt-2.2*

A) Probes used for S1 analysis in Figure 2b and 2c. The *Hinf* I fragment indicated was used to prime the Hind III(H)- Eco RI(R) fragment cloned into mp19 to produce a homogeneously labeled probe. The synthetic oligomer listed in the text was used as primer of the Hind III(H)-Pst I(P) fragment cloned into mp19 to precisely define the initiation site. Open box and shaded box indicate 5' untranslated region and exon I, respectively.

B) S1 analysis using *Hinf* I-primed single-strand probe. Probe alone, with no S1 nuclease added to reaction mix (lane a). Probe hybridized with 3 $\mu$ g of poly A<sup>+</sup> thymus RNA (b) 3 $\mu$ g of polyA<sup>+</sup> L cell RNA (c) and 15 $\mu$ g of RNA from L cells transfected with total mouse genomic DNA and selected for amplified *Lyt-2.2* expression (d). The fragment size was determined with *Hinf* I digested pBR322 size marker.

C) Delineation of 5'-most mRNA initiation site using synthetic oligomer. Probe alone, with no S1 nuclease added to reaction mix (lane a) Probe hybridized with 80 $\mu$ g of L cell RNA (b) 15 $\mu$ g of RNA from amplified *Lyt-2.2* L cell transfectant(c) and 80 $\mu$ g of total thymus RNA (e). Sequence of Hind III-PstI mp19 probe used in this S1 analysis primed with the same oligomer (d).

Table II. *Lyt-2* promoter region has homology to immunoglobulin regulatory sequences and SV40 enhancer.

A)		
cd CONSENSUS		ATGCAAATNA
<i>Lyt-2</i> (308)		TAGGAAATCA
SV40 enhancer		ATGCAAAGCA
B)		
pd CONSENSUS	TGCAG/CCTGTGNCAG	
LYT-2 (326)	TGCAG CTGGCTAAAG	

A) The immunoglobulin cd consensus sequence (31) aligned with a sequence from the *Lyt-2* promoter starting at position 308 in the genomic sequence, and core region of the SV40 enhancer.

B) The immunoglobulin pd consensus sequence (31) aligned with a sequence from the *Lyt-2* promoter starting at position 326 in the genomic sequence.

in the genomic sequence. We identified a consensus polyadenylation signal AATTAAAA at position 4456 and two consensus pentanucleotide sequences CACTG 5' (position 4433) and 3' (4475) to the polyadenylation signal. In order to test for the possibility of other poly A sites 3' to this site because of the presence of RNA species migrating at 3.0-kb (16), we examined 23 newly isolated, uncharacterized *Lyt-2* cDNA clones with two single strand probes derived from M13 vector with the following inserts: Xho I - Pst I (4517-4999) and Stu I - Hind III (4873-5305). No cDNA clones hybridized with these probes. However, 17 out of 23 clones were positive when hybridized with the Sac I - Pst I (4194-4999) probe. Based on these results, we conclude that *Lyt-2* has one poly-A site which begins polyadenylation at residue 4476. We found no poly-A sites 3' to this.

## DISCUSSION

We have cloned and analyzed the gene encoding the *Lyt-2.2* murine glycoprotein. Combining the results of Southern blot analysis of cloned phage DNAs with full length *Lyt-2* cDNA, and transfection experiments, we conclude that the 5.3-Kb Hind III fragment is sufficient to encode the entire *Lyt-2* gene. This was confirmed by sequence analysis of this DNA fragment and previously isolated *Lyt-2* cDNAs. The *Lyt-2* gene has five exons: a fused leader and immunoglobulin variable region-like exon, a hinge region exon, a transmembrane exon and two intracytoplasmic exons. The organization of exons corresponding to functional domains of the protein is similar to other members of the immunoglobulin gene superfamily (28). *Lyt-2* is exceptional in that the leader exon is fused to the variable region-like exon, presumably by deletion of the intron since the human and rat homologues both have been shown to have a



---

separate leader exon (34,35). Sequences at the boundaries of introns and exons predicted by alignment of the alpha and alpha' cDNAs to the genomic sequence conform to consensus splice sequences for mammalian mRNAs (29). Splicing between the first four exons of *Lyt-2* occurs between the first and second nucleotides of the codon, representing what we postulate to be, at least for the variable region-like exon 1, a surviving evolutionary splicing pattern linking *Lyt-2* to immunoglobulin genes.

The S1 protection study revealed that the *Lyt-2* gene has an unusually long 5' untranslated region when compared with most other mRNAs studied (36). There is no ATG codon nor "TATA" box-like sequence after the putative initiation site and before the methionine codon of the leader peptide. Preliminary analysis shows that *Lyt-2* and its human homologue *Leu-2* show significant homology in their 5' untranslated regions, including a cd-like sequenced (unpublished observation); this may indicate that the structure of the 5' noncoding region participates in the regulated expression of these genes.

In the promoter region of *Lyt-2*, there are sequences similar to cd and pd related elements which are considered to be a transcriptional regulatory elements in Ig genes (32,33). The position of these sequences relative to the ATG is similar to that of mouse  $V_k$ , 3'Ck, or human  $V_H$  (32), except that in *Lyt-2*, these sequences locate 3' to the mRNA initiation site because of an unusually long 5' non coding region in *Lyt-2*. Considering that *Lyt-2* is a member of immunoglobulin superfamily, and linked to the immunoglobulin kappa loci, it will be of interest to know whether these sequences have a similar function, if any, to cd or pd sequences of immunoglobulin genes. On the basis of transfections with *Lyt-2* cDNA expression vectors to which we linked the *Lyt-2* genomic promoter, we have shown that these constructs allow expression of *Lyt-2* (17). However, S1 analysis of these cDNA transfectants and L cells transfected with the isolated 5.3-kb Hind III fragment shows both probes (see Figure 2a) to be completely protected, indicating transcription initiates from upstream of the 5' Hind III site. Whether this indicates that sequences in the genome upstream of the Hind III site are necessary to correctly initiate transcription, i.e. the start site as defined in total DNA transfectants and thymus, or that vector sequences are influencing transcription initiation remains to be determined.

We and others have shown that two types of mRNA arise by differential RNA splicing of exon 4 (17,18). We have formally proven by transfection with alpha and alpha' cDNA expression vectors (17) that the two molecular forms of *Lyt-2* are due to differential splicing and not post-translational modification (although we have shown that there are slight differences in glycosylation between the two chains). Alternative splicing similar to what we see with

the *Lyt-2* gene is known to occur in several other genes, such as proprotachykinin in bovine (37), alpha A-crystallin in mouse (38), r-fibrinogen in rat (39), and myelin basic protein in mouse (40), however the mechanism which produces this alternative splicing is not understood. Since Solnick reported the importance of the secondary structure of the primary transcript as a mechanism for differential splicing (41), we searched for inverted repeats or repetitive sequences in the vicinity of exon 4. There are two long inverted repeats: the first at position 2966-2980 and its inverted homologue at 1796-1783 ( $\Delta G = -25.2$  Kcal/mol), and the second at 3594-3612 ( $\Delta G = -22.0$  Kcal/mol) with its inverted homologue at 2165-2148. Since a number of other factors probably play crucial roles in the alternative splicing process, we are unable to assign at present the relative importance these inverted repeats have to the differential splicing of *Lyt-2*.

During the completion of this manuscript, Liaw, et al. (42) published a partial nucleotide sequence of the *Lyt-2.2* gene.

#### ACKNOWLEDGMENTS

The authors would like to thank Dr. Y-h. Chen for the genomic library, Dr. Lawrence Korn for the use of MicroGenie, Dr. Vernon Oi for the oligonucleotide, Dr. Herman Oppermann for some sequence data, and Drs. Naoya Tsurushita and J. Tso for helpful discussions. We also thank Maria Chiara Rattazzi and Berri Benstead for technical help, and Linda Lloyd for secretarial assistance. This work was supported in part by grants GM-17367 and CA-04681 from the National Institutes of Health.

<sup>1</sup>Present address: Department of Immunology, Juntendo University, Tokyo 113, Japan

<sup>2</sup>To whom correspondence should be addressed

#### REFERENCES

1. Cantor, H. and Boyse, E.A. (1975) *J. Exp. Med.* **141**, 1376-1389.
2. Cantor, H. and Boyse, E.A. (1977) *Immunol. Rev.* **33**, 105-124.
3. Ledbetter, J.A., Rouse, R.V., Micklem, H.S. and Herzenberg, L.A. (1980) *J. Exp. Med.* **152**, 280-295.
4. Landegren, U., Ramstedt, U., Axberg, I., Ullberg, M., Jondal, M. and Wigzell, H. (1982) *J. Exp. Med.* **155**, 1579-1584.
5. Ledbetter, J.A. and Seaman, W.E. (1982) *Immunol. Rev.* **68**, 197-218.
6. Marts, E., Heagy, W. and Gromkowski, S.H. (1983) *Immunol. Rev.* **72**, 73-96.
7. Nakayama, E., Shiku, H., Stockbert, E., Oettgen, H.F. and Old, L.J. (1979) *Proc. Natl. Acad. Sci. USA* **76**, 1977-1981.
8. Shinohara, N. and Sachs, D.H. (1979) *J. Exp. Med.* **150**, 432-444.
9. Swain, S.L. (1983) *Immunol. Rev.* **74**, 129-142.

10. Bernard, A., Bounsell, L. and Hill, C. (1984) In Bernard, A., Bounsell, L., Dausset, J., Milstein, C., and Schlossman, S.F. (eds), *Leucocyte Typing*, Springer-Verlag, Berlin, p. 48.
11. Evans, R.L., Wall, D.W., Platsoucas, C.D., Siegal, F.P., Fikrig, S.M., Testa, C.H. and Good, R.A. (1981) *Proc. Natl. Acad. Sci. USA* **78**, 544-548.
12. Reinherz, E.L. and Schlossman, S.F. (1980) *Cell* **19**, 821-827.
13. Jay, G., Palladino, M.A., Khoury, G. and Old, L.J. (1982) *Proc. Natl. Acad. Sci.* **79**, 2654-2657.
14. Ledbetter, J.A., Seaman, W.E., Tsu, T.T. and Herzenberg, L.A. (1981) *J. Exp. Med.* **153**, 1503-1516.
15. Murray, B.J., Mercer, W., McKenzie, I.F.C. and Walker, I.D. (1985) *Immunogenetics* **21**, 519-527.
16. Nakauchi, H., Nolan, G.P., Hsu, C., Huang, H.S., Kavathas, P. and Herzenberg, L.A. (1985) *Proc. Natl. Acad. Sci. USA* **82**, 5126-5130.
17. Tagawa, M., Nakauchi, H., Herzenberg, L.A. and Nolan, G.P. (1986) *Proc. Natl. Acad. Sci. USA* **83**, 3422-3426.
18. Zamoyska, R., Vollmer, A.C., Sizer, K.C., Liaw, C.W. and Parnes, J.R. (1985) *Cell* **43**, 153-163.
19. Chien, Y.-h., Gascoigne, N.R.J., Kavaler, J., Lee, N.E., and Davis, M.M. (1984) *Nature* **309**, 322-326.
20. Wigler, M., Sweet, R., Sim, G.K., Wold, B., Pellicer, A., Lacy, E., Maniatis, T., Silverstein, S. and Axel, R. (1979) *Cell* **16**, 777-785.
21. Sanger, F., Nicklen, S. and Coulson, A.R. (1977) *Proc. Natl. Acad. Sci. USA* **74**, 5463-5467.
22. Messing, J. (1983) *Methods Enzymol.* **101**, 20-78.
23. Staden, R. (1982) *Nucleic Acid Res.* **10**, 2951-2961.
24. Staden, R. and McLachlan, A.D. (1982) *Nucleic Acid Res.* **10**, 141-156.
25. Chirgwin, J.M., Przybyla, A.E., MacDonald, R.J. and Rutter, W.J. (1979) *Biochemistry* **18**, 5294-5299.
26. Burke, J.F. (1984) *Gene* **30**, 63-68.
27. Favaloro, J., Treisman, R. and Kamen, R. (1980) *Methods Enzymol.* **65**, 718-749.
28. Hood, L., Kronenberg, M. and Hunkapiller, T. (1985) *Cell* **40**, 225-229.
29. Breathnach, R. and Chambon, P. (1981) *Ann. Rev. Biochem.* **50**, 349-383.
30. Sharp, P.A. (1981) *Cell* **23**, 643-646.
31. Richards, J.E., Gilliam, A.C., Shen, A., Tucker, P.W. and Blattner, F.R. (1983) *Nature* **306**, 483-487.
32. Falkner, F.G. and Zachau, H.G. (1984) *Nature* **310**, 71-74.
33. Parslow, T.G., Blair, D.L., Murphy, W.J. and Granner, D.K. (1984) *Proc. Natl. Acad. Sci. USA* **81**, 2650-2654.
34. Johnson, P., Gagnon, J., Barclay, A.N. and Williams, A.F. (1985) *EMBO J.* **4**, 2539-2545.
35. Littman, D.R., Thomas, Y., Maddon, P.J., Chess, L. and Axel, R. (1985) *Cell* **40**, 237-246.
36. Kozak, M. (1984) *Nucleic Acid Res.* **12**, 857-872.
37. Nawa, H., Kotani, H. and Nakanishi, S. (1984) *Nature* **312**, 729-734.
38. King, C.R. and Piatigorsky, J. (1983) *Cell* **32**, 707-712.
39. Crabtree, G.R. and Kant, J.A. (1982) *Cell* **31**, 159-166.
40. Takahashi, N., Roach, A., Teplow, D.B., Prusiner, S.B. and Hood, L. (1985) *Cell* **42**, 139-148.
41. Solnick, D. (1985) *Cell* **43**, 667-676.
42. Liaw, C.W., Zamoyska, R. and Parnes, J.R. (1986) *J. Immunol.* **137**, 1037-1043.