P. cynomolgi (B) was used. One selection event involving the pooling of eggs from the five most susceptible females in a group of 50 examined was followed by four isofemale line–selection steps. Seven of the ten isofemale lines examined at the fourth isofemale line–selection step were virtually fully susceptible to *P. cynomolgi* (B). These two strains of *P. cynomolgi* were used because the G3 stock was highly refractory to the NIH strain and highly susceptible to the Lond strain of the parasite. Therefore, selection for refractoriness was felt to be more stringent if Lond was used and selection for susceptibility more stringent with *P. cynomolgi* (B). By our definition, an infected mosquito was a female with normal or encapsulated oocysts (or both) on the gut.

5. P. Gotz and A. Vey, *Parasitology* **70**, 77 (1975).
6. C. C. Chen and B. R. Laurence, *Int. J. Parasitol.* **15**, 421 (1985).

7. We thank P. Graves for valuable advice and encouragement during the course of this work.

9 June 1986; accepted 15 September 1986

Isolation and Sequence of L3T4 Complementary DNA Clones: Expression in T Cells and Brain

BÉATRICE TOURVIEILLE, SCOTT D. GORMAN, ELIZABETH H. FIELD, TIM HUNKAPILLER, JANE R. PARNES

T lymphocytes express on their surface not only a specific receptor for antigen and major histocompatibility complex proteins, but also a number of additional glycoproteins that are thought to play accessory roles in the processes of recognition and signal transduction. L3T4 is one such T-cell surface protein that is expressed on most mouse thymocytes and on mature mouse T cells that recognize class II (Ia) major histocompatibility complex proteins. Such cells are predominantly of the helper/inducer phenotype. In this study, complementary DNA clones encoding L3T4 were isolated and sequenced. The predicted protein sequence shows that L3T4 is a member of the immunoglobulin gene superfamily. It is encoded by a single gene that does not require rearrangement prior to expression. Although the protein has not previously been demonstrated on nonhematopoietic cells, two messenger RNA species specific for L3T4 are found in brain. The minor species comigrates with the L3T4 transcript in T cells, whereas the major species is 1 kilobase smaller.

MATURE MOUSE T LYMPHOCYTES can be divided into two subsets by their expression of the alternative T-cell differentiation antigens L3T4 (CD4 in humans) and Lyt-2 (CD8 in humans). The L3T4 subset consists predominantly of helper/inducer T cells and correlates best with recognition by T cells of class II (Ia) major histocompatibility complex (MHC) molecules (*1–3*). The Lyt-2 subset is made up primarily of cytotoxic and suppressor T cells and correlates best with recognition of class I (H-2K, D, or L) MHC molecules (*3, 4*). Monoclonal antibodies specific for L3T4 or Lyt-2 inhibit the functional activity (cytotoxicity, proliferation, lymphokine release) of T cells that bear these proteins (*1–6*). It has been postulated that L3T4 and Lyt-2 play a role in increasing the avidity of the interaction between T cells and antigen-presenting cells or target cells, perhaps by binding to nonpolymorphic regions of class II and class I MHC proteins, respectively (*2, 3, 6, 7*). It has alternatively been postulated that monoclonal antibodies specific for L3T4 (or CD4) inhibit function by directly transmitting a negative signal to the T cell (*8*). Although Lyt-2 and CD8 are normally known to be expressed only on particular subsets of thymocytes, T cells, and natural killer cells, CD4 has also been shown to be expressed by normal cells of the monocyte/macrophage and Langerhans lineages in both humans and rats (*9*). However, none of these proteins has been reported to be expressed on normal nonhematopoietic cells.

The genes encoding human CD8, CD4, and mouse Lyt-2 have been recently cloned, and their predicted amino acid sequences have revealed that they are evolutionarily related to immunoglobulin (Ig) variable (V) regions (*10–14*). We have now cloned the complementary DNA (cDNA) encoding mouse L3T4 and show that it too is a member of the Ig gene superfamily. We further show that the gene is expressed not only in T-lineage cells but also in brain and that the size of the mRNA in brain is different from that in T cells.

To isolate mouse L3T4 cDNA clones we screened a C57BL/Ka mouse thymocyte cDNA library with a full-length human CD4 cDNA clone (*12*) used as probe. Two mouse clones that hybridized to the human clone were isolated. The nucleotide sequence of the one (pcL3T4-C7) with the longer insert (1.3 kb) was determined (Fig. 1). Because this clone did not contain the 5′ end, it was used as a probe to isolate from the same library an additional cDNA clone (pcL3T4-14) that extended farther in the 5′ direction. The nucleotide sequence of the 5′ untranslated region, the leader, and the first ten amino acids of the mature protein were therefore determined from pcL3T4-14 (Fig. 1). The nucleotide sequence shown in Fig. 1 predicts a mature protein of 435 amino acids (predicted molecular size 48,853 daltons), with 372 amino acids external to the cell, a 25–amino acid hydrophobic transmembrane region, and a 38–amino acid highly basic cytoplasmic domain. The mature protein sequence is preceded by a 22–amino acid hydrophobic leader or signal peptide as is typically found at the NH_2-terminus of cell surface and secreted proteins. The point of cleavage of this leader was predicted by comparison with other published leader sequences.

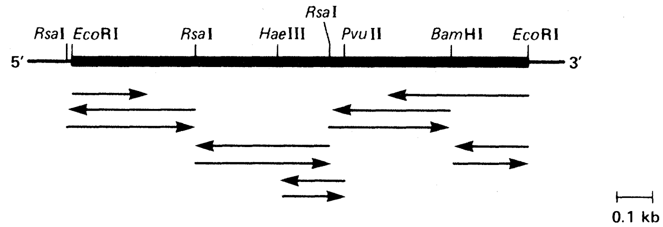
As expected, the nucleotide sequence of the L3T4 cDNA was homologous to that of the human CD4 clone with which it was selected, and the encoded protein was also closely related (Fig. 2A). The most highly conserved region was the cytoplasmic domain (79% at the amino acid level), which may play a role in signal transduction. In contrast, the external portions of L3T4 and CD4 contained only 55% identical residues. This latter finding is similar to our previous results comparing the mouse (Lyt-2) and human (CD8) sequences of the alternative T-cell differentiation marker (*13*). The mouse L3T4 protein has four predicted *N*-linked glycosylation sites (Asn-X-Thr or Asn-X-Ser) at residues 165, 276, 301, and 370, as compared to only two in the human CD4 (*12*).

We and others observed previously that CD8 and Lyt-2 have NH_2-terminal external domains that are homologous to the Ig light chain V regions (*11, 13, 14*). A similar relation has been found for human CD4 (*12*). We therefore searched a series of data banks with the L3T4 sequence to see whether similar or additional homologies could be found. These computer comparisons indicated that L3T4 is also a member of the Ig gene superfamily. The NH_2-terminal domain of the mature protein (90 to 101 amino acids, depending on where one arbitrarily sets the border) is homologous to Ig V regions, with the greatest similarity being to light chain V regions, especially κ (up to 35%) (Fig. 2B). This domain of L3T4 has the two cysteines (residues 20 and 90) that form the characteristic disulfide loop of Ig-like homology units, as well as the structurally important tryptophan (residue 32) that is always found 12 to 15 residues downstream from the first cysteine of the disulfide

B. Tourvieille, S. D. Gorman, E. H. Field, J. R. Parnes, Department of Medicine, Division of Immunology, Stanford University Medical Center, Stanford, CA 94305.
T. Hunkapiller, Division of Biology, California Institute of Technology, Pasadena, CA 91125.

loop. In addition to the similarity in amino acid sequence, computer predictions of potential β-sheet structure and hydrophobicity plots were extremely similar for the NH_2-terminal domain of L3T4 and for Ig light and heavy chain V regions. Twelve residues beyond the second cysteine of the V-like domain is a sequence (amino acids 102 to 110) that bears some similarity to Ig λ J (joining) segments (Fig. 2C). The placement of this sequence relative to the V-like domain suggests that it may be evolutionarily related to Ig J segments, although it is missing the central Gly-X-Gly, which is highly characteristic of Ig J sequences.

We found another region of the L3T4 protein between amino acids 140 and 175 (V′) that appears more weakly related to the COOH-terminal portion of Ig heavy and light chain V regions, including the second cysteine of the Ig disulfide loop (Fig. 2D). This sequence is most closely related to the NH_2-terminal V-like sequence of L3T4, suggesting an internal duplication. Although the significance of this short region of homology is unclear, it is followed by a sequence (amino acids 176 to 185) that is even more similar to λ light chain J segments than the sequence following the NH_2-terminal V-like domain (Fig. 2E). It is therefore possible that this region of the protein also evolved from an ancestor of Ig V regions but that it suffered a major deletion of its NH_2-terminal portion during the course of evolution. Even though both L3T4 and Lyt-2 are homologous to κ variable regions, they show little homology to each other except for the conserved residues present in most members of the Ig gene superfamily. Both this sequence divergence and the fact that L3T4 has undergone an internal duplication resulting in two V-like homology units (in contrast to the single such unit in Lyt-2) suggest that L3T4 and Lyt-2 have very different evolutionary histories after splitting off from κ. The remainder of the external protein (connecting peptide) bears no significant homology to other known proteins except to the human CD4.




```
                                                   |--> *  L
 -22                                               Met Cys Arg Ala Ile Ser Leu Arg Arg Leu    -13
   1  GCTCAGATTCCCAACCAACAAGAGCTCAAGGAGACCACC     ATG TGC CGA GCC ATC TCT CTT AGG CGC TTG     69

                                                    |-->V
 -12  Leu Leu Leu Leu Leu Gln Leu Ser Gln Leu Leu Ala Val Thr Gln Gly Lys Thr Leu Val     8
  70  CTG CTG CTG CTG CTG CAG CTG TCA CAA CTC CTA GCT GTC ACT CAA GGG AAG ACG CTG GTG   129

                                                *
   9  Leu Gly Lys Glu Gly Glu Ser Ala Glu Leu Pro Cys Glu Ser Ser Gln Lys Lys Ile Thr    28
 130  CTG GGG AAG GAA GGG GAA TCA GCA GAA CTG CCC TGC GAG AGT TCC CAG AAG AAG ATC ACA   189

  29  Val Phe Thr Trp Lys Phe Ser Asp Gln Arg Lys Ile Leu Gly Gln His Gly Lys Gly Val    48
 190  GTC TTC ACC TGG AAG TTC TCT GAC CAG AGG AAG ATT CTG GGG CAG CAT GGC AAA GGT GTA   249

  49  Leu Ile Arg Gly Gly Ser Pro Ser Gln Phe Asp Arg Phe Asp Ser Lys Lys Gly Ala Trp    68
 250  TTA ATT AGA GGA GGT TCG CCT TCG CAG TTT GAT CGT TTT GAT TCC AAA AAA GGG GCA TGG   309

  69  Glu Lys Gly Ser Phe Pro Leu Ile Ile Asn Lys Leu Lys Met Glu Asp Ser Gln Thr Tyr    88
 310  GAG AAA GGA TCG TTT CCT CTC ATC ATC AAT AAA CTT AAG ATG GAA GAC TCT CAG ACT TAT   369

          *                                                  |-->J
  89  Ile Cys Glu Leu Glu Asn Arg Lys Glu Glu Val Glu Leu Trp Val Phe Lys Val Thr Phe   108
 370  ATC TGT GAG CTG GAG AAC AGG AAA GAG GAG GTG GAG TTG TGG GTG TTC AAA GTG ACC TTC   429

         <--|
 109  Ser Pro Gly Thr Ser Leu Leu Gln Gly Gln Ser Leu Thr Leu Thr Leu Asp Ser Asn Ser   128
 430  AGT CCG GGT ACC AGC CTG TTG CAA GGG CAG AGC CTG ACC CTG ACC TTG GAT AGC AAC TCT   489

                                  *           |-->V′
 129  Lys Val Ser Asn Pro Leu Thr Glu Cys Lys His Lys Lys Gly Lys Val Val Ser Gly Ser   148
 490  AAG GTC TCT AAC CCC TTG ACA GAG TGC AAA CAC AAA AAG GGT AAA GTT GTC AGT GGT TCC   549

                                                                      •   *
 149  Lys Val Leu Ser Met Ser Asn Leu Arg Val Gln Asp Ser Asp Phe Trp Asn Cys Thr Val   168
 550  AAA GTT CTC TCC ATG TCC AAC CTA AGG GTT CAG GAC AGC GAC TTC TGG AAC TGC ACC GTG   609

                              |-->J′                                 |-->CP
 169  Thr Leu Asp Gln Lys Lys Asn Trp Phe Gly Met Thr Leu Ser Val Leu Gly Phe Gln Ser   188
 610  ACC CTG GAC CAG AAA AAG AAC TGG TTC GGC ATG ACA CTC TCA GTG CTG GGT TTT CAG AGC   669

 189  Thr Ala Ile Thr Ala Tyr Lys Ser Glu Gly Glu Ser Ala Glu Phe Ser Phe Pro Leu Asn   208
 670  ACA GCT ATC ACG GCC TAT AAG AGT GAG GGA GAG TCA GCG GAG TTC TCC TTC CCA CTC AAC   729

 209  Phe Ala Glu Glu Asn Gly Trp Gly Glu Leu Met Trp Lys Ala Glu Lys Asp Ser Phe Phe   228
 730  TTT GCA GAG GAA AAC GGG TGG GGA GAG CTG ATG TGG AAG GCA GAG AAG GAT TCT TTC TTC   789

 229  Gln Pro Trp Ile Ser Phe Ser Ile Lys Asn Lys Glu Val Ser Val Gln Lys Ser Thr Lys   248
 790  CAG CCC TGG ATC TCC TTC TCC ATA AAG AAC AAA GAG GTG TCC GTA CAA AAG TCC ACC AAA   849

 249  Asp Leu Lys Leu Gln Leu Lys Glu Thr Leu Pro Leu Thr Leu Lys Ile Pro Gln Val Ser   268
 850  GAC CTC AAG CTC CAG CTG AAG GAA ACG CTC CCA CTC ACC CTC AAG ATA CCC CAG GTC TCG   909

                              •
 269  Leu Gln Phe Ala Gly Ser Gly Asn Leu Thr Leu Thr Leu Asp Lys Gly Thr Leu His Gln   288
 910  CTT CAG TTT GCT GGT TCT GGC AAC CTG ACT CTG ACT CTG GAC AAA GGG ACA CTG CAT CAG   969

                                                  •                   *
 289  Glu Val Asn Leu Val Val Met Lys Val Ala Gln Leu Asn Asn Thr Leu Thr Cys Glu Val   308
 970  GAA GTG AAC CTG GTG GTG ATG AAA GTG GCT CAG CTC AAC AAT ACT TTG ACC TGT GAG GTG  1029

 309  Met Gly Pro Thr Ser Pro Lys Met Arg Leu Thr Leu Lys Gln Glu Asn Gln Glu Ala Arg   328
1030  ATG GGA CCT ACC TCT CCC AAG ATG AGA CTG ACC CTG AAG CAG GAG AAC CAG GAG GCC AGG  1089

                                                                              *
 329  Val Ser Glu Glu Gln Lys Val Val Gln Val Val Ala Pro Glu Thr Gly Leu Trp Gln Cys   348
1090  GTC TCT GAG GAG CAG AAA GTA GTT CAA GTG GTG GCC CCT GAG ACA GGG CTG TGG CAG TGT  1149

 349  Leu Leu Ser Glu Gly Asp Lys Val Lys Met Asp Ser Arg Ile Gln Val Leu Ser Arg Gly   368
1150  CTA CTG AGT GAA GGT GAT AAG GTC AAG ATG GAC TCC AGG ATC CAG GTT TTA TCC AGA GGG  1209

       •          |--> TM      *
 369  Val Asn Gln Thr Val Phe Leu Ala Cys Val Leu Gly Gly Ser Phe Gly Phe Leu Gly Phe   388
1210  GTG AAC CAG ACA GTG TTC CTG GCT TGC GTG CTG GGT GGC TCC TTC GGC TTT CTG GGT TTC  1269

              *           *   *       |-->  *  CY
 389  Leu Gly Leu Cys Ile Leu Cys Cys Val Arg Cys Arg His Gln Gln Arg Gln Ala Ala Arg   408
1270  CTT GGG CTC TGC ATC CTC TGC TGT GTC AGG TGC CGG CAC CAA CAG CGC CAG GCA GCA CGA  1329

                                                      *       *
 409  Met Ser Gln Ile Lys Arg Leu Leu Ser Glu Lys Lys Thr Cys Gln Cys Pro His Arg Met   428
1330  ATG TCT CAG ATC AAG AGG CTC CTC AGT GAG AAG AAG ACC TGC CAG TGC CCC CAC CGG ATG  1389

 429  Gln Lys Ser His Asn Leu Ile Trm                                                   435
1390  CAG AAG AGC CAT AAT CTC ATC TGA GGCC                                             1417
```

Fig. 1. Nucleotide and deduced amino acid sequence of L3T4 cDNA. The insert of cDNA clone pcL3T4-C7 was sequenced by the dideoxynucleotide chain termination method (*22*) with M13 vectors mp18 and mp19 (*23*) according to the strategy shown at the top. The closed box indicates the cDNA insert while the thinner lines indicate phage arms. The Eco RI sites at the 5′ and 3′ ends of the insert are the result of synthetic linker addition. The sequence of the 5′ untranslated region, leader, and first ten amino acids of the mature protein was determined similarly from M13 mp18 and mp19 subclones of a Sac I (5′) to Hinf I (3′) fragment of an overlapping cDNA clone, pcL3T4-14. This clone was found to be identical in sequence to pcL3T4-C7 in the region of overlap. The number of the first and last amino acid or nucleotide in each line is indicated in the left and right margins, respectively. Horizontal arrows indicate the start (or end) of the predicted protein domains. Cysteine residues are marked by asterisks and potential *N*-linked glycosylation sites by closed circles. Abbreviations: L, leader sequence (signal peptide); V and V′, sequences homologous to immunoglobulin variable regions; J and J′, sequences homologous to immunoglobulin joining segments; CP, connecting peptide; TM, transmembrane region; and CY, cytoplasmic tail.

The transmembrane region of L3T4 contains no charged residues and is only 54% homologous to the human CD4 transmembrane domain. In contrast, the transmembrane domain was the most highly conserved region (79%) between mouse Lyt-2 and human CD8 (*13*). Analysis of the L3T4 transmembrane region by a hydrophobic moment plot (*15*) predicts that this sequence penetrates the membrane as a single α-helix unassociated with any other transmembrane region. This result is consistent with biochemical data suggesting that L3T4 is a monomeric protein (*1, 2*). The predicted cytoplasmic domain of L3T4 is highly charged, with an excess of 12 basic residues. This is similar to previous observations for Lyt-2, CD8, and CD4 (*11–13*). There is no significant homology between the cytoplasmic domain of L3T4 and that of other sequenced proteins (except its human homolog).

The pattern of expression of L3T4 mRNA in a variety of mouse cell lines and tissues was examined by Northern blot analysis in order to correlate mRNA expression with the known pattern of expression of the protein (Fig. 3A). The insert of L3T4 cDNA clone pcL3T4-C7 hybridized to a single major mRNA species of 3.7 kb in thymus, spleen, and lymph node. This same hybridizing mRNA was also seen in two thymoma cell lines known to express the protein (VL3/1 and KKT2). In contrast, no hybridizing mRNA was detectable in several other thymoma cell lines that do not express surface L3T4 (1112, MBL2, and R1.1R/TLIII 7X.6), in two B-cell lines (Bal 17 and 225), in an undifferentiated teratocarcinoma cell line (F9), in an Ia^+ macrophage line (WEHI-3), or in liver. Although not obvious in Fig. 3A, a small amount of the 3.7-kb mRNA was detectable in kidney on the original autoradiograph. It is possible that this mRNA represents contamination of the kidney RNA with RNA from $L3T4^+$ blood cells, but we cannot exclude the possibility of a low level of expression by endogenous renal cells. The thymus lane (lane 1) and, to a lesser extent, the VL3/1 lane (lane 10) in Fig. 3A are greatly overexposed and suggest the presence of a number of larger and smaller mRNA species. Several of these are believed to be artifacts related to compressions in the regions of the 28*S* and 18*S* ribosomal RNA (rRNA), possibly combined with a low level of degradation. These bands are not visible in repeated blots with different RNA preparations (Fig. 3B). However, the two largest bands are present on other blots and are probably partially spliced precursors, since both nuclear and cytoplasmic RNA are present.

Of particular interest was the finding that normal mouse brain contained a small amount of a 2.7-kb L3T4 transcript that was not convincingly detectable in any of the other tissues or cell lines examined by using multiple blots and different RNA preparations. However, we cannot exclude the possibility that small amounts of this transcript are present in other tissues at levels below that found in brain. Because an mRNA species comigrating with the brain 2.7-kb transcript was not clearly detectable in thymus, spleen, or lymph node, it is highly unlikely that the expression of this mRNA in brain is a result of contaminating $L3T4^+$ blood cells. This 2.7-kb transcript is found in the $poly(A)^+$ fraction of brain RNA and not in the $poly(A)^-$ fraction (Fig. 3C). We do not know which cells in the brain express this mRNA and whether they are of neuronal, glial, or possibly macrophage origin.

A

```
                 ↦V
L3T4    1   VTQGKTLVLGKEGESAELPCESSQKKITVFTWKFSDQRKILGQHGKGVLIRGGSPSQFDRFDSKKGAWEKGSFPLIINKL   80
CD4     1   A***NKV****K*DTV**T*TA****SIQ*H**N*N*I****NQ*-SF*TK*P*KLN-**A**RRSL*DQ*N*****KN*   78

                                        ↦J    ↤                                   ↦V'
L3T4   81   KMEDSQTYICELENRKEEVELWVFKVTFSPGTSLLQGQSLTLTLDSNSKVSNPLTECKHKKGKVVSGSKVLSMSNLRVQD  160
CD4    79   *I***D*****V*DQ****Q*L**GL*ANSD*H***********E*PP-G*S*SVQ*RSPR**NIQ*G*T**V*Q*EL**  157

                             ↦J'       ↦CP
L3T4  161   SDFWNCTVTLDQKKNWFGMTLSVLGFQSTAITAYKSEGESAEFSFPLNFAEEN--GWGELMWKAEKDSFFQPWISFSIKN  238
CD4   158   *GT*T***LQN***VE*KIDIV**A**KASSIV**K***QV******A*TV*KLT*S***W*Q**RA*SSKS**T*DL**  237

L3T4  239   KEVSVQKSTKDLKLQLKETLPLTLKIPQVSLQFAGSGNLTLTLD-K-GTLHQEVNLVVMKVAQLNNTLTCEVMGPTSPKM  316
CD4   238   *****KRV*Q*P***MGKK***H*TL**ALP*Y********A*EA*T*K**********RAT**QKN*****W******L  317

                                                                                 ↦TM
L3T4  317   RLTLKQENQEARVSEEQKVVQVVAPETGLWQCLLSEGDKVKMDSRIQVL---SRGVNQTVFLACVLGGSFGFLGFLGLCI  393
CD4   318   M*S**L**K**K**KRE*A*W*LN**A*M******DSGQ*LLE*N*K**PTW*TP*-*PMA*-I****VA*L*L*I**G*  395

                ↦CY
L3T4  394   LCCVRCRHQQRQAARMSQIKRLLSEKKTCQCPHRMQKSHNLI                                        435
CD4   396   FF*****RR***E*********************F**TCSP*                                        437
```

B

```
                        10        20             30        40        50             60        70        80        90
L3T4 V      ---VTQGKTLVLGKEGESAELPCESSQKK---ITVFTWKFSDQRKILGQHGKGVLIRG-----GSPSQFDRFDSKKGAWEKGSFPLIINKLKMEDSQTYIC
Vκ          DIKM**SPSSMYASL**RVTIS*KA**DI---NSYL**F--Q*KP--*KSP*TL*Y*ANRLVD*V**---**---S*SGSGQD*S*T*SS*EY**MGI*Y*
Vλ          QSAL**-PPSAS*SL*Q*VTIS*TGTSSDVGGYNYVS*Y--Q*HA--*KAP*VIIYEVNKRPS*V*D---**---S*SKSGNTAS*TVSG*QA**EAD*Y*
```

C

```
            102     110
L3T4 J      WVF----KVTFSP
λ J1        ***GGGT*L*VLG
λ J2        Y**GGGT***VLG
λ J3        FI*GSGT***VLG
κ J1        *T*GGGT*LEIK-
```

D

```
L3T4 V'     KKGKVVSGSKVLSMSNLRVQDSDFWNCTVT (140-169)
MOPC 47A    ISRDNSE*ILY*Q*NT**A***ATYY*ARD (71-100)
Gal         ISRDNAKN*LY*Q*NS***E*TALYY*ARG (70-99)
L3T4 V      ***AWEK**FP*IINK*KME**QTYI*ELE (64-93)
```

E

```
            176      185
L3T4 J'     W-FGM-T-LSVLG
λ J1        *V**GG*K*T***
λ J2        YV**GG*KVT***
λ J3        FI**SG*KVT***
κ J1        *T**GG*K*EIK-
```

Fig. 2. Homology of L3T4 to human CD4 and to immunoglobulin V regions. (A) The amino acid sequence of L3T4 is compared with that of human CD4. Predicted protein domains of L3T4 are labeled as in Fig. 1. The numbers of the first and last amino acids in each line are indicated in the left and right margins, respectively. (B) The sequence of the NH_2-terminal domain of the mature L3T4 protein is compared with a mouse Ig κ (V-T1) V region (*24*) and a human Ig λ (Mcg) V region (*25*). The numbers above refer to the amino acid positions in L3T4. (C) The amino acid sequence of a segment of L3T4 (residues 102 to 110) homologous to Ig J segments (L3T4 J) is compared with the sequences of mouse λ J1, J2, and J3 (*26*) and mouse κ J1 (*27*). (D) The amino acid sequence of a segment of L3T4 (V′) homologous to the COOH-terminal portion of Ig V regions is shown with the sequences of the homologous portion of two Ig heavy chain V regions, mouse MOPC 47A (*28*) and human GAL (*29*), and the homologous portion of the L3T4 NH_2-terminal V-like region (L3T4 V). The numbers in parentheses indicate the residues of each protein shown. (E) The sequence of a second J-like sequence of L3T4, J′ (residues 176 to 185) is compared with the sequences of three mouse λ J segments (*26*) and one mouse κ J segment (*27*). In all of these comparisons, asterisks indicate identical residues and dashes indicate gaps in the alignment.

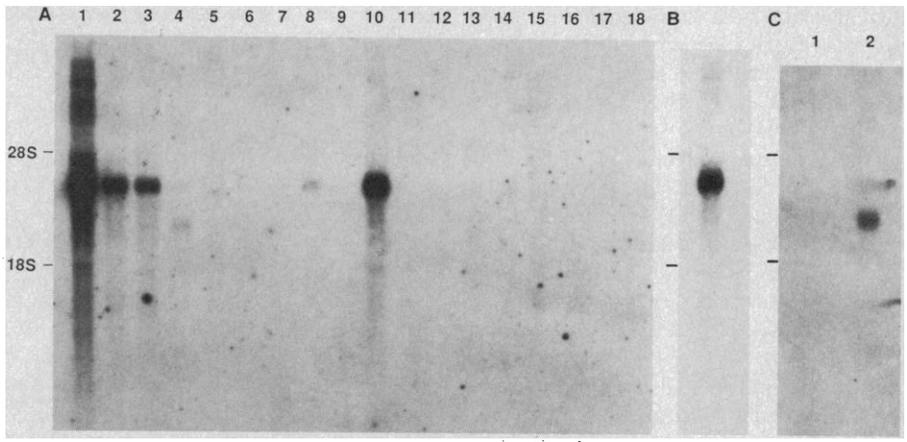


Fig. 3. Northern blot analysis of L3T4 mRNA expression. Total RNA was isolated from tissues and cell lines by the procedure of Chirgwin *et al.* (*30*). (A) Each RNA sample (10 μg) was subjected to electrophoresis through a 1.5% agarose gel containing 2.2*M* formaldehyde. The RNA was transferred to nitrocellulose (*31*) and the blot was hybridized as described (*13*). The probe consisted of a 1.3-kb insert of cDNA clone pcL3T4-C7 labeled with ^{32}P by random hexamer priming (*32*). The blot was washed as described (*13*) and exposed for 9 days at −70°C with an intensifying screen. RNA's were from the following sources: 1, thymus; 2, lymph node; 3, spleen; 4, brain; 5, kidney; 6, liver; 7, rat glial cell line C6 (A. Perlman, Stanford); 8, thymoma cell line KKT2 (I. Weissman, Stanford); 9, no sample; 10, thymoma cell line VL3/1 (I. Weissman, Stanford); 11, thymoma cell line 1112 (J. Allison, University of California, Berkeley); 12, thymoma cell line MBL2 (I. Weissman, Stanford); 13, thymoma cell line R1.1R/TLIII 7X.6 (R. Hyman, Salk Institute); 14, B-cell line Bal 17 (I. Weissman, Stanford); 15, B-cell line 225 (J. Allison, University of California, Berkeley); 16, F9 teratocarcinoma cell line; 17, macrophage cell line WEHI-3 (P. Jones, Stanford); and 18, WEHI-3 after 48 hours of induction with a γ-interferon–containing cell-free supernatant from concanavalin A–treated (48 hours) mouse spleen cells (*33*). All tissues and cell lines were of mouse origin except for the rat glial cell line C6. (B) Total thymus RNA (15 μg) was subjected to electrophoresis, blotted, and hybridized as in (A) except that the RNA was prepared independently. Exposure was for 5 hours at −70°C with an intensifying screen. (C) Total brain RNA isolated as in (A), but independently, was separated into poly(A)$^-$ (lane 1) and poly(A)$^+$ (lane 2) fractions by passage through and elution from an oligo(dT)-cellulose column (*34*) two times. Each fraction (9 μg) was subjected to electrophoresis, blotted, and hybridized as in (A). Exposure was for 18 hours at −70°C with an intensifying screen. The migration positions of 28*S* and 18*S* rRNA are marked in the left margin of each autoradiograph and were determined by ethidium bromide staining.

Since we have previously found that the mouse L3T4 cDNA cross-hybridizes with a single rat gene on genomic Southern blots (*16*), we examined a rat glial cell line, C6, for expression of RNA hybridizing to the mouse L3T4 cDNA. No cross-hybridizing mRNA was detectable (Fig. 3A). However results with a single cell line do not rule out the possibility that glial cells are the source of the mouse 2.7-kb transcript.

The L3T4 mRNA contains a very long 3′ untranslated region (estimated at 1.6 to 1.7 kb), so it was possible that the difference in mRNA lengths resulted from two alternative polyadenylation sites. If so, the encoded protein would be the same in both cases. However, S1 nuclease mapping indicates that the two mRNA's differ in protein coding sequence at the 5′ end (*17*). Since we have only detected a single gene for mouse L3T4, it seems most likely that the difference in the two mRNA's is the result of alternative mRNA splicing patterns. The typical 3.7-kb L3T4 transcript is also detectable in brain, but at a lower level. We do not know whether the same cells in brain express both mRNA species. It is possible that the small amount of 3.7-kb transcript in brain results from contamination with blood cells.

The homology of L3T4 to Ig V regions and the expression of an alternative mRNA form in brain led us to examine whether this gene rearranges in cells that express it. We therefore examined genomic DNA from mouse liver (nonexpressing), VL3/1 (a thymoma cell line expressing the 3.7-kb mRNA), and mouse brain (3.7-kb and 2.7-kb mRNA's) by Southern blot (*18*) analysis

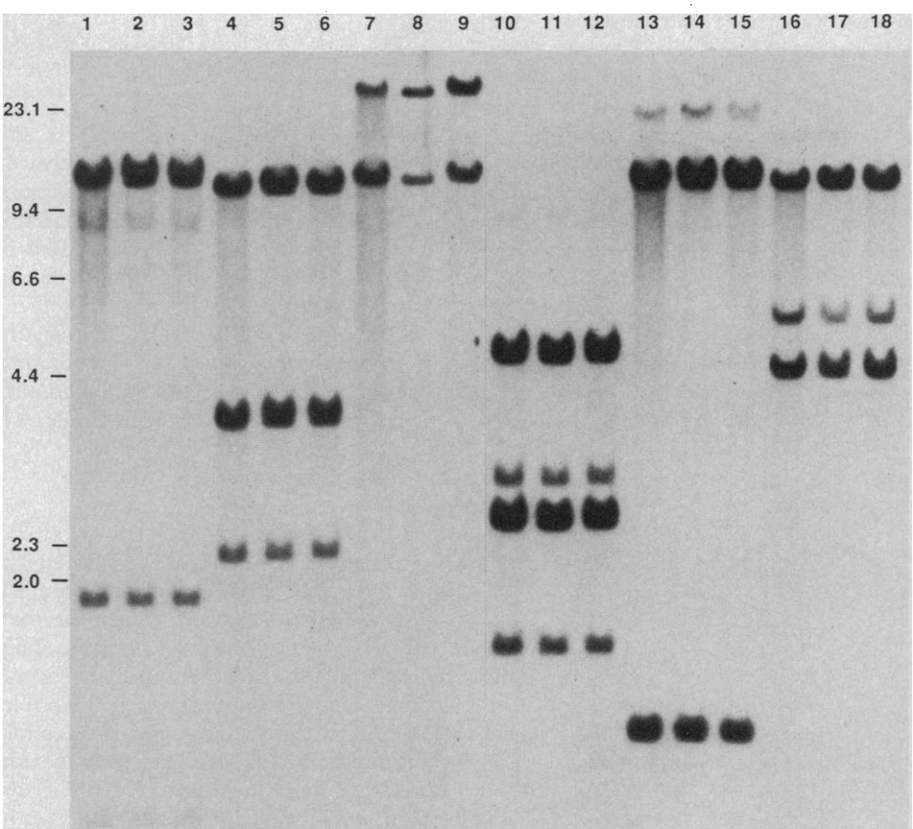


Fig. 4. Southern blot analysis of L3T4 gene in thymoma cell line, liver, and brain. Genomic DNA (8 μg) was digested with the indicated restriction endonuclease and subjected to electrophoresis through a 0.8% agarose gel. The DNA was transferred to nitrocellulose by the method of Southern (*18*) and the blot was hybridized, as previously described (*13*), to the insert of L3T4 cDNA clone pcL3T4-C7 labeled as described in the legend to Fig. 3. The blot was washed as described (*13*) and exposed overnight at −70°C with an intensifying screen. Enzymes used were (lanes 1 to 3) Bam HI; (lanes 4 to 6) Bgl II; (lanes 7 to 9) Hind III; (lanes 10 to 12) Pvu II; (lanes 13 to 15) Eco RI; and (lanes 16 to 18) Xba I. Sources of DNA were L3T4$^+$ mouse thymoma cell line VL3/1 (lanes 1, 4, 7, 10, 13, and 16), mouse liver (lanes 2, 5, 8, 11, 14, and 17), and mouse brain (lanes 3, 6, 9, 12, 15, and 18). Size markers are indicated in the left margin and represent Hind III fragments of phage λ run in parallel and stained with ethidium bromide.

after digestion with six different restriction enzymes (Fig. 4). We could detect no rearrangement in the L3T4 gene with the complete insert of pcL3T4-C7 as a probe. On the basis of the data from VL3/1 we conclude that no rearrangement is required for expression in T-lineage cells. Although we found no evidence for rearrangement in total brain DNA, we cannot rule out the possibility that a rearrangement occurs in fewer than 5% of brain cells, and that these are the cells that produce the smaller mRNA.

The expression of L3T4 mRNA in mouse brain is intriguing given the high frequency of central nervous system involvement in human acquired immune deficiency syndrome (AIDS) (*19, 20*); the presence of human T-cell lymphotropic virus type III (HTLV-III), the retrovirus responsible for AIDS, in the brain of affected individuals (*19*); and the demonstrated role of human CD4 as a cellular receptor for HTLV-III (*21*). These observations suggest that CD4 is probably expressed in human brain at the protein level. The large evolutionary divergence that we have found between the external domains of L3T4 and CD4 may explain why HTLV-III does not infect mouse T cells. In any event, the tissue-specific expression of an alternative form of L3T4 mRNA suggests a yet undefined role for this molecule in brain.

REFERENCES AND NOTES

1. D. P. Dialynas *et al.*, *J. Immunol.* **131**, 2445 (1983).
2. D. P. Dialynas *et al.*, *Immunol. Rev.* **74**, 29 (1983).
3. S. L. Swain, *ibid.*, p. 129.
4. M. Okada and C. S. Henney, *J. Immunol.* **125**, 300 (1980); S. L. Swain, *Proc. Natl. Acad. Sci. U.S.A.* **78**, 7101 (1981).
5. S. L. Swain, D. P. Dialynas, F. W. Fitch, M. English, *J. Immunol.* **132**, 1118 (1984); E. Nakayama, H. Shiku, E. Stockert, H. F. Oettgen, L. J. Old, *Proc. Natl. Acad. Sci. U.S.A.* **76**, 1977 (1979); N. Hollander, E. Pillemer, I. L. Weissman, *J. Exp. Med.* **152**, 674 (1980).
6. H. R. MacDonald, A. L. Glasebrook, C. Bron, A. Kelso, J.-C. Cerottini, *Immunol. Rev.* **68**, 89 (1982).
7. J. L. Greenstein, J. Kappler, P. Marrack, S. J. Burakoff, *J. Exp. Med.* **159**, 1213 (1984).
8. T. R. Malek, C. Chan, L. H. Glimcher, R. N. Germain, E. M. Shevach, *J. Immunol.* **135**, 1826 (1985); P. Wassmer, C. Chan, L. Logdberg, E. M. Shevach, *ibid.*, p. 7688; I. Bank and L. Chess, *J. Exp. Med.* **162**, 1294 (1985).
9. G. S. Wood, N. L. Warner, R. A. Warnke, *J. Immunol.* **131**, 212 (1983); R. A. Moscicki, E. P. Amento, S. M. Krane, J. T. Kurnick, R. B. Colvin, *ibid.*, p. 743; W. A. Jeffries, J. R. Green, A. F. Williams, *J. Exp. Med.* **162**, 117 (1985).
10. P. Kavathas, V. P. Sukhatme, L. A. Herzenberg, J. R. Parnes, *Proc. Natl. Acad. Sci. U.S.A.* **81**, 7688 (1984).
11. D. R. Littman, Y. Thomas, P. J. Maddon, L. Chess, R. Axel, *Cell* **40**, 237 (1985); V. P. Sukhatme, K. C. Sizer, A. C. Vollmer, T. Hunkapiller, J. R. Parnes, *ibid.*, p. 591.
12. P. J. Maddon *et al.*, *ibid.* **42**, 93 (1985).
13. R. Zamoyska, A. C. Vollmer, K. C. Sizer, C. W. Liaw, J. R. Parnes, *ibid.* **43**, 153 (1985).
14. H. Nakauchi, G. P. Nolan, C. Hsu, H. S. Huang, P. Kavathas, L. A. Herzenberg, *Proc. Natl. Acad. Sci. U.S.A.* **82**, 5126 (1985).
15. D. Eisenberg, E. Schwarz, M. Komaromy, R. Wall, *J. Mol. Biol.* **179**, 125 (1984).
16. E. H. Field, B. Tourvieille, J. R. Parnes, unpublished results.
17. S. D. Gorman, B. Tourvieille, J. R. Parnes, unpublished results.
18. E. M. Southern, *J. Mol. Biol.* **98**, 503 (1975).
19. G. M. Shaw *et al.*, *Science* **227**, 177 (1985); J. A. Levy *et al.*, *Lancet* **1985-II**, 586 (1985); D. D. Ho *et al.*, *N. Engl. J. Med.* **313**, 1493 (1985).
20. W. D. Snider *et al.*, *Ann. Neurol.* **14**, 403 (1983); A. L. Belman *et al.*, *ibid.* **18**, 560 (1985); C. A. Carne *et al.*, *Lancet* **1985-II**, 1206 (1985).
21. A. G. Dalgleish *et al.*, *Nature (London)* **312**, 763 (1984); D. Klatzmann *et al.*, *ibid.*, p. 767; J. S. McDougal *et al.*, *J. Immunol.* **135**, 3151 (1985); J. A. Levy *et al.*, *Ann. Intern. Med.* **103**, 694 (1985); J. S. McDougal *et al.*, *Science* **231**, 382 (1986).
22. F. Sanger, S. Nicklen, A. R. Coulson, *Proc. Natl. Acad. Sci. U.S.A.* **74**, 5463 (1977).
23. J. Messing, *Methods Enzymol.* **101**, 20 (1983).
24. W. Altenburger, M. Steinmetz, H. G. Zachau, *Nature (London)* **287**, 603 (1980); S. Bodary and B. Mach, *EMBO J.* **1**, 719 (1982).
25. J. W. Fett and H. F. Deutsch, *Biochemistry* **13**, 4102 (1974).
26. B. Blomberg, A. Traunecker, H. Eisen, S. Tonegawa, *Proc. Natl. Acad. Sci. U.S.A.* **78**, 3765 (1981); B. Blomberg and S. Tonegawa, *ibid.* **79**, 530 (1982).
27. E. E. Max, J. G. Seidman, P. Leder, *ibid.* **76**, 3450 (1979).
28. E. A. Robinson and E. Appella, *J. Biol. Chem.* **254**, 11418 (1979).
29. S. Watanabe, H. U. Barnikol, J. Horn, J. Bertram, N. Hilschmann, *Z. Physiol. Chem.* **354**, 1505 (1973).
30. J. M. Chirgwin, A. E. Przybyla, R. J. MacDonald, W. J. Rutter, *Biochemistry* **18**, 5294 (1979).
31. P. S. Thomas, *Proc. Natl. Acad. Sci. U.S.A.* **77**, 5201 (1980).
32. A. P. Feinberg and B. Vogelstein, *Anal. Biochem.* **132**, 6 (1983).
33. D. Paulnock-King, K. Sizer, Y. R. Freund, P. P. Jones, J. R. Parnes, *J. Immunol.* **135**, 632 (1985).
34. H. Aviv and P. Leder, *Proc. Natl. Acad. Sci. U.S.A.* **69**, 1408 (1972).

35. We thank D. O'Neill for editorial assistance. Supported by NIH grants AI11313 and GM34991; by the Ministère des Affaires Etrangères-France, the Ligue Nationale Française contre le Cancer, and the Philippe Foundation (B.T.); by NCI postdoctoral fellowship 1 F32 CA07877-01 (S.D.G.); by a Bank of America–Giannini Foundation postdoctoral fellowship (E.H.F.); and by a John A. and George L. Hartford faculty fellowship award (J.R.P.).

6 June 1986; accepted 25 August 1986

URF6, Last Unidentified Reading Frame of Human mtDNA, Codes for an NADH Dehydrogenase Subunit

ANNE CHOMYN, MICHAEL W. J. CLEETER, C. IAN RAGAN, MARCIA RILEY, RUSSELL F. DOOLITTLE, GIUSEPPE ATTARDI

The polypeptide encoded in URF6, the last unassigned reading frame of human mitochondrial DNA, has been identified with antibodies to peptides predicted from the DNA sequence. Antibodies prepared against highly purified respiratory chain NADH dehydrogenase from beef heart or against the cytoplasmically synthesized 49-kilodalton iron-sulfur subunit isolated from this enzyme complex, when added to a deoxycholate or a Triton X-100 mitochondrial lysate of HeLa cells, specifically precipitated the URF6 product together with the six other URF products previously identified as subunits of NADH dehydrogenase. These results strongly point to the URF6 product as being another subunit of this enzyme complex. Thus, almost 60% of the protein coding capacity of mammalian mitochondrial DNA is utilized for the assembly of the first enzyme complex of the respiratory chain. The absence of such information in yeast mitochondrial DNA dramatizes the variability in gene content of different mitochondrial genomes.

SINCE THEIR INITIAL DISCOVERY IN human mitochondrial DNA (mtDNA) (*1*), the eight so-called unidentified reading frames (URF's), which together represent about 60% of the protein coding capacity of the mitochondrial genome of animal cells (*2*), have raised considerable interest. Particularly intriguing has been the observation that these reading frames, with the exception of one, have no homology to any of the identified and unidentified reading frames of mtDNA of *Saccharomyces cerevisiae* (*3*) and *Schizosaccharomyces pombe* (*4*), although several of them occur in mtDNA of other lower eukaryotic cells (*5–14*). Recently, direct evidence obtained by the use of antibodies to synthetic peptides predicted from the DNA sequence and by analysis of protease fingerprints has shown that seven of the eight URF's are expressed in HeLa cells (*15–17*).

The smallest of the URF's (URFA6L), a 207-nucleotide (nt) reading frame overlapping out of phase the NH_2-terminal portion of the adenosinetriphosphatase (ATPase) subunit 6 gene has been identified as the animal equivalent of the recently discovered yeast H^+-ATPase subunit 8 gene (*18, 19*). The functional significance of the other URF's has been, on the contrary, elusive. Recently, however, immunoprecipitation experiments with antibodies to purified, rotenone-sensitive NADH-ubiquinone oxido-

A. Chomyn and G. Attardi, Division of Biology, California Institute of Technology, Pasadena, CA 91125.
M. W. J. Cleeter and C. I. Ragan, Department of Biochemistry, University of Southampton, Southampton SO9 3TU, England.
M. Riley and R. F. Doolittle, Department of Chemistry, University of California at San Diego, La Jolla, CA 92093.