Proc. Natl. Acad. Sci. USA
Vol. 82, pp. 3819–3823, June 1985
Immunology

Isolation and characterization of mouse *Thy-1* genomic clones

(cosmid clones/DNA sequence/transmembrane segment)

HSIU-CHING CHANG*, TETSUNORI SEKI†, TETSUYA MORIUCHI*, AND JACK SILVER†

†Cellular and Molecular Biology Unit, Hospital for Joint Diseases, New York, NY 10003; and *Department of Microbiology and Public Health, Michigan State University, East Lansing, MI 48823

Communicated by Anton Lang, February 11, 1985

ABSTRACT **The mouse *Thy-1.2* gene was isolated from a C57Bl/6 cosmid library and its nucleotide sequence was determined from an 8-kilobase-long *Eco*RI fragment. The predicted amino acid sequence indicates that the mouse Thy-1 molecule contains a 19 amino acid leader peptide and the 112 amino acids reported previously from protein sequence analysis, plus 31 extra amino acids at the carboxyl terminus. These 31 amino acids contain a stretch of 20 amino acids, at positions 124–143, which is highly hydrophobic. RNA transfer blot analysis of RNA from mouse tissues indicates that the sequence coding for these 31 amino acids is present on poly(A)-containing RNA of brain and thymus tissues. This hydrophobic segment very likely provides the basis for integration of Thy-1 within the plasma membrane. The entire coding sequence of *Thy-1* is distributed among three exons, encoding amino acid residues −19 to 8, −7 to 106, and 107 to 143, respectively. Comparison of the mouse and rat *Thy-1* genes shows that both have a similar gene organization and encode a highly conserved transmembrane segment.**

The Thy-1 antigen was originally defined in mice as a cell surface alloantigen of thymus and brain with two allelic forms, *Thy-1.1* and *Thy-1.2* (1). Subsequently, the *Thy-1.1* determinant was identified in rats (2). In both species, Thy-1 is present in large amounts on the surface of thymus and brain cells, with about 600,000 molecules per cell on rat thymocytes and equivalent amounts on brain cells. Thus, Thy-1 is probably the most abundant surface glycoprotein of both cell types (3). In rodents, Thy-1 is also expressed in small amounts on fibroblasts (4), epidermal cells (5), mammary glands (6), and immature skeletal muscle cells (7). In many of these tissues the level of Thy-1 expression changes dramatically during cell differentiation.

The molecules expressing the Thy-1 antigenic determinant have been isolated from rat and mouse brain cells (8, 9). Thy-1 has been shown to have a molecular weight of 17,500, one-third of the molecule being carbohydrate and the remainder being a polypeptide of 111 amino acids in rat (10) and 112 amino acids in mouse (11).

Glycoproteins that are structurally related to rat and mouse brain Thy-1 have been purified from the brain of humans (12), dogs (13), chickens (14), and frogs (15), and even from invertebrate squid (11). Although the expression of Thy-1 seems to be conserved in neuronal cells and fibroblasts, its expression on lymphoid cells varies remarkably between species (16). For example, Thy-1 is present on peripheral T cells of mice but absent from T cells of man and rats. Furthermore, of these three species, only rats express Thy-1 in bone marrow cells (17). Thus, Thy-1 represents an interesting model for studies of differential tissue expression of membrane molecules.

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.

The Thy-1 molecule was previously thought to have an unusual mode of integration in the membrane. Although Thy-1 can be heavily labeled in the thymocyte membrane by an affinity label that reacts only with the hydrophobic region of membrane proteins (18), biochemical studies have failed to demonstrate any extended sequence of hydrophobic amino acids that might function as a transmembrane segment. However, based on the observation that the COOH-terminal peptide displays hydrophobic properties, it has been proposed that Thy-1 is anchored to the membrane through a hydrophobic glycolipid attached to the carboxyl-terminal cysteine residue at position 111 (10). We have recently isolated rat *Thy-1* cDNA and genomic clones and the results indicate that Thy-1 is in fact a molecule of 142 amino acids, 31 amino acids longer than proposed previously (19). Within these extra 31 amino acids there is a hydrophobic segment of 20 amino acids, which appears to function as a transmembrane component anchoring Thy-1 to the cell surface.

In this study, we report the isolation of mouse *Thy-1* genomic clones from a C57Bl/6 cosmid library and the DNA sequence analysis of the mouse *Thy-1* gene. In addition, the predicted amino acid sequences and organization of the mouse and rat genes are compared.

MATERIALS AND METHODS

DNA Cloning. A cosmid library of mouse C57Bl/6 DNA was constructed in the cosmid vector c2RB as described (20) with minor modification. The mouse C57Bl/6 DNA was partially digested with *Sau*3A to an average size of about 40 kilobases (kb) and then dephosphorylated before ligation with *Bam*HI- and *Sma* I-digested c2RB DNA. A high ATP concentration was used during the ligation reaction to prevent the joining of the *Sma* I blunt ends. The ligated DNA was packaged *in vitro* and used to infect *Escherichia coli* strain 1046. The mouse library was screened by using the insert of the rat *Thy-1* cDNA clone (21) as a hybridization probe. The DNA of putative Thy-1-positive clones was purified by using a minipreparation procedure (see below). The DNA was cleaved with *Eco*RI and other appropriate restriction enzymes and electrophoresed on a 0.6% agarose gel for Southern blot analyses. Confirmed Thy-1-positive cosmid clones were then digested with *Eco*RI and subcloned into the *Eco*RI site of pBR322. The subcloned DNA was used to transform *E. coli* strain HB101 and the colonies were screened with the original *Thy-1* probe. The general procedures for the growth, screening, and analysis of cosmid libraries, as well as nick-translation and Southern blot analyses, were performed as described (22, 23).

DNA Preparation. The preparation of DNA from putative *Thy-1* clones was performed by using the alkaline-extract method (24) with some modifications. A single bacterial colony was cultured (40 ml) overnight without any amplification. The bacteria were pelleted and resuspended in 50 mM

Abbreviations: kb, kilobase(s); bp, base pair(s); UT, untranslated.

glucose/10 mM EDTA/25 mM Tris·HCl, pH 8.0, without lysozyme. The cells were lysed by the addition of 2 vol of 0.2 M NaOH/1% $NaDodSO_4$, and the cellular DNA was precipitated by addition of 0.5 vol of 3 M potassium acetate (pH 5.2). The plasmid DNA remaining in the supernatant was precipitated with 0.6 vol of isopropanol. The DNA was resuspended in 10 mM Tris·HCl/1 mM EDTA, pH 8.0, treated with DNase-free RNase, phenol extracted, and then precipitated with 2.5 vol of ethanol. The DNA isolated by this procedure can be used for endonuclease restriction enzyme digestion and ligation for subcloning. Large-scale DNA preparation was performed as described above, except that CsCl gradient centrifugation was used for purification of the plasmid DNA prior to RNase treatment and phenol extraction.

DNA Sequence Analysis. The Thy-1-positive subclones pcT108 and pcT34 were subjected to nucleotide sequence analyses. The entire genomic clone was cleaved at selected sites with restriction endonucleases and fragments corresponding to the insert were purified by agarose gel electrophoresis and electroelution or electrophoresis on low-melting-point agarose (Bethesda Research Laboratories) and extraction with 0.3 M NaOAC (pH 5.2)-saturated phenol. Fragments were dephosphorylated with calf intestinal alkaline phosphatase and end-labeled with T4 polynucleotide kinase. Only one of the fragments was end labeled by filling in with DNA polymerase I. Double-end-labeled fragments were digested with restriction enzymes, separated on agarose gel or low-melting-point agarose, and eluted from gels. The appropriate fragments were then subjected to partial chemical degradation sequence analysis as described (25).

RNA Analysis. Total RNA from mouse tissues was isolated as described (26). The poly(A) fraction was purified on an oligo(dT) column (27) and RNA transfer blot analysis was performed as described (28). Mouse poly(A)-RNA was electrophoresed on a 1.2% formaldehyde/agarose gel, blotted onto a nitrocellulose filter, and hybridized with the ^{32}P-labeled fragment Pv3 in 50% formamide at 42°C.

RESULTS

Isolation and Characterization of Mouse *Thy-1* Genomic Clones. Mouse C57Bl/6 DNA, partially digested with *Sau*3A to an average size of 40 kb, was used to construct a cosmid library in the vector c2RB (20). The library was screened with a ^{32}P-labeled insert of the rat thymocyte *Thy-1* cDNA clone pT64 (21). After screening 175,000 clones, 3 independent clones, CT10, CT34, and CT5F, were isolated. Southern blot analysis of *Eco*RI-digested DNA using rat *Thy-1* as a probe showed that each clone had only one hybridizing band but only clone CT10 had an 8-kb *Eco*RI-hybridizing fragment. This corresponded to the size of the *Thy-1* hybridizing fragment obtained by Southern blot analysis of *Eco*RI-digested mouse cellular DNA. Partial restriction enzyme mapping of all 3 clones indicated that they were overlapping and that clone CT10 shared only 4 and 2 kb of its 3′ end with clones CT34 and CT5F, respectively (Fig. 1*A*). The total insert of all 3 clones spans ≈75 kb of the mouse genome. The *Thy-1*-hybridizing *Eco*RI fragments of clones CT34 and CT10 were subcloned into the *Eco*RI site of pBR322 and the resultant subclones, pcT34 and pcT108, were analyzed by restriction enzyme mapping (Fig. 1*B*) and subjected to DNA sequence analysis by using the protocol illustrated in Fig. 1*C*.

Structure of the Mouse *Thy-1* Gene. The nucleotide sequence of the mouse *Thy-1* gene and the predicted protein sequence of mouse *Thy-1* are shown in Fig. 2. The intron–exon organization of the gene was established by comparison of its sequence with the nucleotide sequence of a rat *Thy-1* cDNA clone, which contains the entire coding sequence of the mature protein (19), and by comparison of the

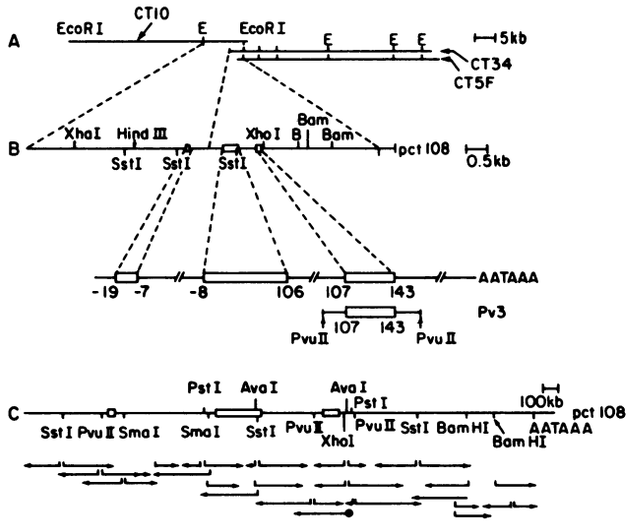

FIG. 1. Restriction enzyme maps of *Thy-1* genomic clones. (*A*) Partial *Eco*RI digestion maps of *Thy-1*-containing clones CT10, CT34, and CT5F are shown. Clone CT10 shares only 4 and 2 kb of the 3′ end of the *Thy-1* gene with CT34 and CT5F, respectively, making the total insert of these clones span 75 kb of DNA along the mouse genome. (*B*) Partial restriction map of the *Thy-1*-containing genomic clone pcT108 and the intron–exon organization of the mouse *Thy-1* gene. The boxed areas denote the coding regions of *Thy-1*. The amino acids were numbered relative to the amino terminus of the protein sequence as determined. (*C*) The strategy for sequencing the *Thy-1* gene. The pBR322 clones containing the *Eco*RI insert of *Thy-1* were cleaved at selected sites with restriction endonucleases as indicated. The fragments were isolated and 5′ end-labeled with T4 polynucleotide kinase or 3′ end-labeled by filling in with DNA polymerase I. Double-end-labeled fragments were digested with restriction enzymes, separated on the agarose gel, eluted, and then subjected to DNA sequencing by the method of Maxam and Gilbert (25). The extent of sequence determined from each fragment is indicated by the length of the arrow. The solid circle and vertical lines at one end of each arrow indicate labeling at the 3′ or 5′ end, respectively. The DNA sequence was completely determined by sequencing either both complementary strands or the same strand at least twice, starting with different cleavage sites.

predicted protein sequence with the protein sequence reported previously for mouse *Thy-1* (11). The coding sequence is divided into three exons by introns with sizes of 590 base pairs (bp) and 386 bp, respectively. The first coding exon [actually the second exon of the gene since the 5′ untranslated (UT) region is interrupted by a large intron, see below] encodes the first 12 amino acids of the signal peptide, whereas the second coding exon encodes the remaining 7 amino acids of the signal peptide and amino acid residues 1–106 of the rat Thy-1 protein. The third coding exon encodes the remainder of the Thy-1 protein, amino acids 107–143, followed by a termination codon. The 1106 nucleotides between the termination codon and the first poly(A) signal are believed to be part of the 3′ UT region based on analogy with the rat *Thy-1* gene, which also has a long 3′ UT region of 1056 bp (19). Thus, the mouse *Thy-1* gene is organized identically to the rat *Thy-1* gene, with the only significant difference being in the size of the introns (Fig. 3). The predicted protein sequence indicates that the gene we have cloned is the *Thy-1.2* gene since the only difference between the two mouse allelic forms, *Thy-1.1* and *Thy-1.2*, is amino acid residue 89, which in *Thy-1.1* is arginine and in *Thy-1.2* is glutamine (11). This is consistent with the *Thy-1.2* designation of C57Bl/6 mice based on serologic determinations.

The mouse *Thy-1* gene has an intron within the 5′ UT region based on alignment of its sequence with the nucleotide

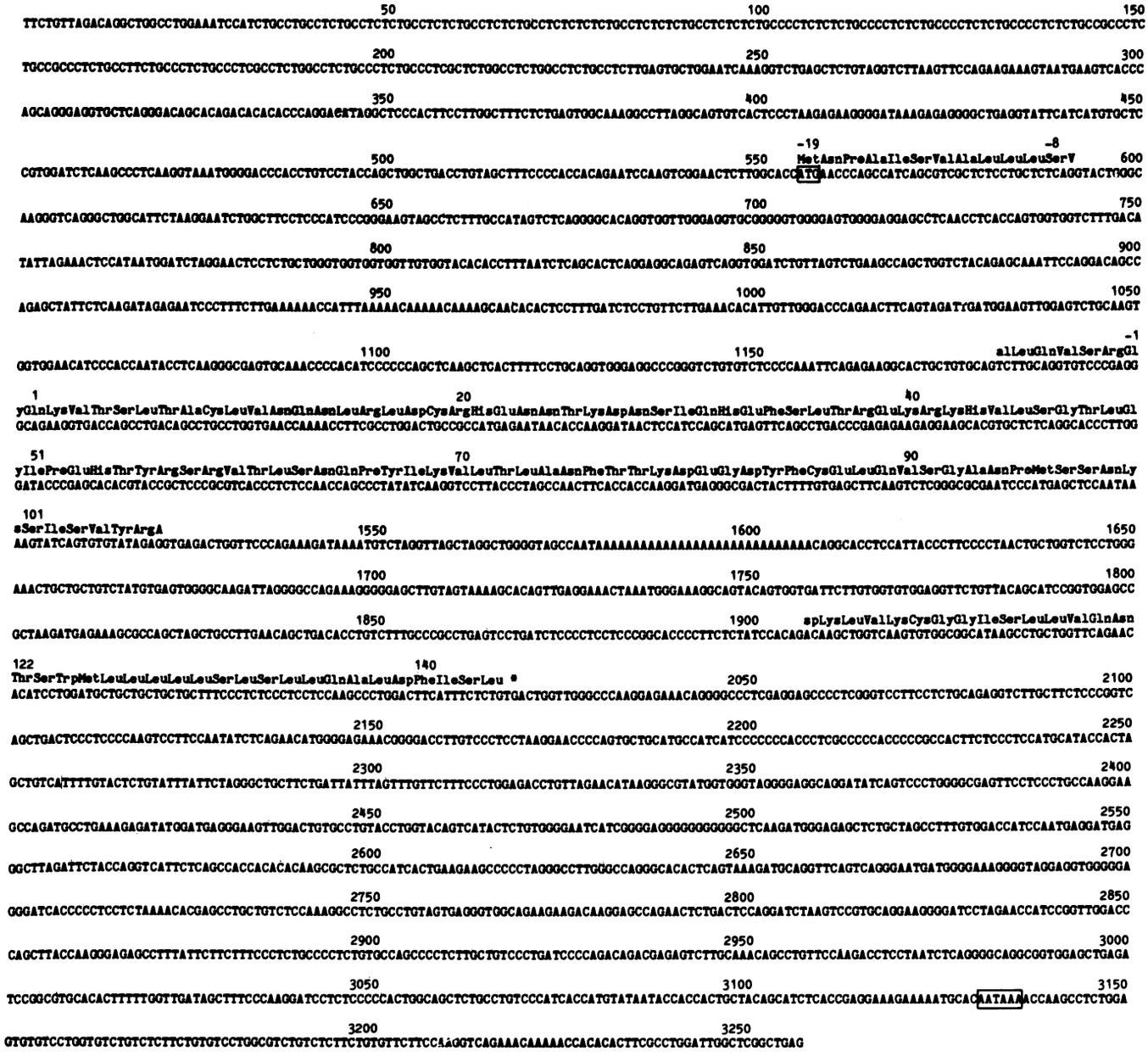

FIG. 2. Sequence of the mouse *Thy-1.2* gene. The nucleotide sequence and predicted protein sequence (three-letter amino acid code) of the *Thy-1* gene are shown. The initiation codon and the A-A-T-A-A-A sequence are boxed. The termination codon is indicated by an asterisk.

sequence of a previously defined rat *Thy-1* cDNA clone, pT64 (21), which contains part of the 5′ UT region, and the rat *Thy-1* genomic sequence (Fig. 4). The size of this intron can be estimated by hybridizing a nearly full-length mouse *Thy-1* cDNA clone (a gift from M. Davis, Stanford) to a Southern blot of *Eco*RI–*Hin*dIII-digested DNA from clone pcT108 (Fig. 1*B*). Both fragments of the *Thy-1* insert hybridize, indicating that the remainder of the 5′ UT region extends beyond the *Hin*dIII site of the insert, which is ≈1.5 kb 5′ to the initiation codon (Fig. 1*B*). Furthermore, the 1-kb *Sst* I fragment of the insert (Fig. 1*B*) fails to hybridize to this nearly full-length cDNA clone, whereas the *Xha* I–*Hin*dIII fragment hybridizes strongly (data not shown). Thus, the size of the intron interrupting the 5′ UT region is at least 1.5 kb in length, with the remainder of the 5′ UT lying beyond the *Hin*dIII site. The lack of restriction enzyme sites within a large segment of

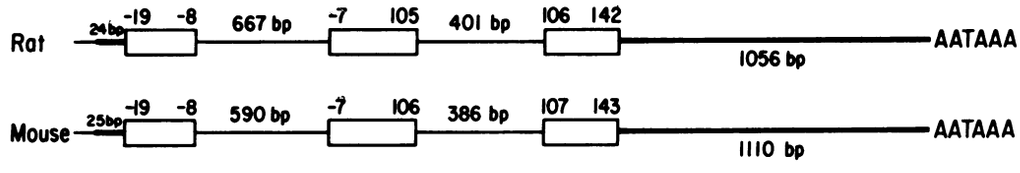

FIG. 3. Comparison of the gene organization of mouse and rat *Thy-1*. Coding regions are indicated by the boxed areas. 5′ and 3′ UT regions that are part of exons are indicated by thick lines and introns are indicated by thin lines.

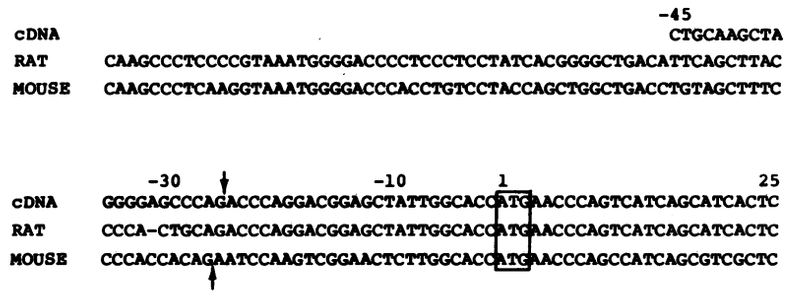

FIG. 4. Comparison of the 5′ nucleotide sequences of rat cDNA and mouse and rat genomic DNA. The sequence of the rat cDNA (top line), starting from position −45 upstream from the initiation codon, ATG, is compared to the sequences of rat and mouse genomic clones (middle line and bottom line, respectively). The initiation codon is boxed, and the putative splicing points are marked by arrows.

this intron, presumably due to the presence of highly repetitive sequences, has made it difficult to sequence this region.

The Nucleotide Sequence of *Thy-1* Predicts the Presence of a Hydrophobic Segment at the Carboxyl Terminus. The predicted protein sequence of the mouse *Thy-1* gene indicates the presence of an additional 31 amino acids at the carboxyl terminus previously not detected by conventional protein sequence analysis. RNA transfer blot analysis indicates that the DNA sequence encoding these extra 31 amino acids is also present in the mature mRNA. When a blot of poly(A)-RNA from mouse brain, liver, spleen, and thymus is hybridized with a *Pvu* II fragment, Pv3, which codes for amino acids 107–143 (Fig. 1*B*), only one species of RNA is detectable (Fig. 5). No other band was observed when the same blot was hybridized 2 months later with the entire insert of the *Thy-1* clone pcT108. We have observed previously that a rat *Thy-1* cDNA and a genomic clone also possess an additional 31 amino acids not described previously (19). These 31 amino acids are highly conserved in both rat and mouse (Fig. 6) and show only 9.7% divergence (3 of 31 residues), in contrast to the remainder of the molecule, which is 19.1% (25 of 131 residues) divergent. Within these 31 amino acids there is a highly hydrophobic stretch of 20 amino acids starting at the tryptophan residue at position 124 and ending with the leucine residue at the carboxyl end. It is intriguing to note that this is a sufficient number of amino acids to traverse the lipid bilayer of the membrane and may, therefore, form the molecular basis by which Thy-1 integrates within the membrane.

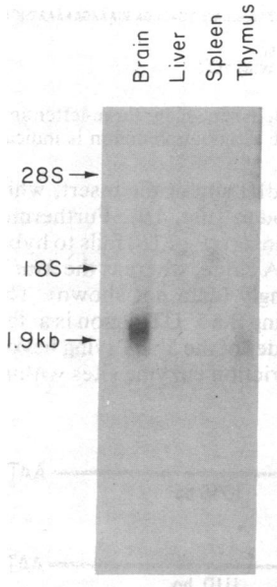

FIG. 5. RNA transfer blot analyses of RNA from mouse tissues. Poly(A)-containing RNA was isolated from mouse brain, liver, spleen, and thymus and subjected to RNA transfer blot analysis.

DISCUSSION

We have isolated three cosmid clones spanning 75 kb of the C57Bl/6 mouse genome and containing the *Thy-1.2* gene. The *Thy-1* gene is ≈4.6 kb long. By comparing the nucleotide sequence of the mouse *Thy-1* gene to previously isolated rat *Thy-1* cDNA and genomic clones (19) we have been able to deduce the intron–exon organization. This comparison reveals that the coding sequence is divided into three exons, one encoding the first 12 amino acids of the signal peptide, a second encoding the remaining 7 amino acids of the signal peptide and 106 amino acids of the mature protein, and a third

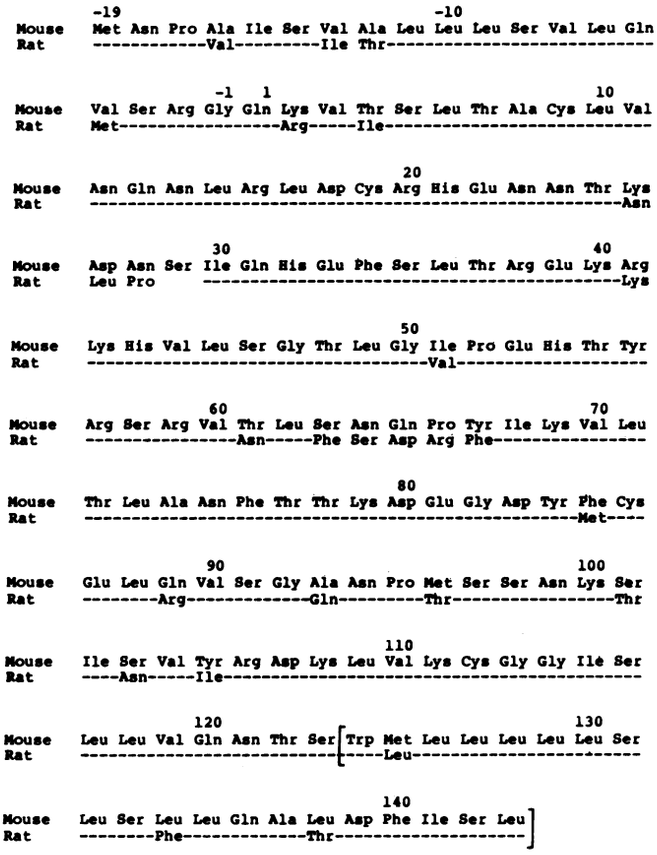

FIG. 6. Comparison of the predicted amino acid sequences of mouse and rat *Thy-1*. Only sequence differences are indicated. The 20 amino acid hydrophobic segment that represents a putative transmembrane region is bracketed.

encoding the remaining 31 amino acids, which contain a transmembrane-like hydrophobic segment. By analogy to the rat *Thy-1* gene, we have inferred that the last coding exon also contains a very long (1110 bp) 3′ UT region. The three exons containing the protein coding sequence are separated by two moderate-sized introns of 590 bp and 386 bp. In contrast, the intron interrupting the 5′ UT region is very large. Comparisons of the 5′ ends of the rat and mouse cDNA and genomic clones (Fig. 4) and Southern blotting of the mouse *Thy-1* gene and hybridization with a nearly full-length mouse *Thy-1* cDNA clone indicate that the 5′ UT region is interrupted by an intron that begins 25 bp upstream from the initiation codon and is ≈1.5 kb in length. This is consistent with the size of the 5′ intron of the rat *Thy-1* gene, where we have found a promoter-like "TATA box" ≈1.5 kb upstream from the initiation codon (unpublished data). The lack of restriction enzyme sites within a large portion of this intron, presumably due to the presence of repeating sequences, has made it difficult to sequence the entire intron. The organization of the mouse *Thy-1* gene is identical to that of the rat *Thy-1* gene except for differences in the size of the introns (Fig. 3). Furthermore, the *Thy-1* gene is organized in a fashion similar to that of other membrane proteins to which it is structurally homologous—i.e., histocompatibility antigens and membrane-bound immunoglobulin (29, 30), in which the exons encoding the signal peptide and a hydrophobic transmembrane segment are separated by introns from the major exon(s) encoding the domain(s) of the protein.

Previous studies by other investigators using conventional biochemical techniques have suggested that the mouse Thy-1 molecule is only 112 amino acids in length and lacks a hydrophobic protein segment normally found in other membrane proteins and whose function is to anchor the molecule in the plasma membrane (11). However, these investigators have proposed, based on studies indicating unusual hydrophobic properties for the protein and some of its peptide fragments, that Thy-1 is anchored to the cell surface through a hydrophobic glycolipid attached to the carboxyl-terminal cysteine residue at position 112. We now report that both the rat and mouse *Thy-1* genes in fact encode a molecule of 142 and 143 amino acids, respectively, 31 amino acids longer than reported previously. Furthermore, the nucleotide sequence of a rat *Thy-1* cDNA clone (19) and RNA transfer blot analysis using a probe encoding primarily this extra protein segment (Fig. 5) indicate that these 31 additional amino acids are also present in mature poly(A)-RNA. In addition, biosynthetic experiments indicate that at least a portion of this 31 amino acid segment (up to the tryptophan residue at protein 124) is translated and is part of the mature Thy-1 molecule present at the membrane (19). We believe that the differences observed between the protein and DNA data may be due to proteolytic digestion of the Thy-1 molecule during purification. Within the extra amino acids predicted from the DNA data there is an extremely hydrophobic stretch of 20 amino acids starting at position 124 and ending with the carboxyl-terminal leucine residue that resembles a transmembrane segment. The length of this segment is sufficient to traverse the lipid bilayer of the membrane but not sufficient to contain an intracytoplasmic tail. Accordingly, there are no basic amino acids at the carboxyl end characteristically found at the beginning of intracytoplasmic tails. The hydrophobic segment very likely provides the means by which Thy-1 is attached to the cell membrane, making it analogous to the mode of attachment of other membrane glycoproteins. However, the lack of an intracytoplasmic portion is quite unusual and may represent an important functional difference.

This work was supported by a grant from the National Institutes of Health.

1. Reif, A. E. & Allen, J. M. (1964) *J. Exp. Med.* **120,** 413–433.
2. Douglas, T. C. (1972) *J. Exp. Med.* **126,** 1054–1062.
3. Acton, R. T., Morris, R. J. & Williams, A. F. (1974) *Eur. J. Immunol.* **4,** 598–602.
4. Stern, P. L. (1973) *Nature (London)* **246,** 76–78.
5. Scheid, M., Boyse, E. A., Carswell, E. A. & Old, L. J. (1972) *J. Exp. Med.* **135,** 938–955.
6. Lennon, V. A., Unger, M. & Dulbecco, R. (1978) *Proc. Natl. Acad. Sci. USA* **75,** 6093–6097.
7. Lesley, J. F. & Lennon, V. A. (1977) *Nature (London)* **268,** 163–165.
8. Letarte-Muirhead, M., Barclay, A. N. & Williams, A. F. (1975) *Biochem. J.* **151,** 685–697.
9. Trowbridge, I. S., Weissman, I. L. & Bevan, M. J. (1975) *Nature (London)* **256,** 652–654.
10. Campbell, D. G., Gagnon, J., Reid, K. B. M. & Williams, A. F. (1981) *Biochem. J.* **195,** 15–30.
11. Williams, A. F. & Gagnon, J. (1982) *Science* **216,** 696–703.
12. Cotmore, S. F., Crowhurst, S. A. & Waterfield, M. D. (1981) *Eur. J. Immunol.* **11,** 597–603.
13. McKenzie, J. L. & Fabre, J. W. (1981) *Transplantation* **31,** 275–282.
14. Rostas, J. A. P., Shevenan, T. A., Sinclair, C. M. & Jeffrey, P. L. (1983) *Biochem. J.* **213,** 143–152.
15. Mansour, M. H. & Cooper, E. L. (1984) *J. Immunol.* **132,** 2515–2523.
16. Dalchau, R. & Fabre, J. W. (1979) *J. Exp. Med.* **149,** 576–582.
17. Hunt, S. V., Mason, D. W. & Williams, A. F. (1977) *Eur. J. Immunol.* **7,** 817–823.
18. Owen, M., Knott, J. C. A. & Crumpton, M. J. (1980) *Biochemistry* **19,** 3092–3099.
19. Seki, T., Moriuchi, T., Chang, H.-C., Denome, R., Pleogh, H. & Silver, J. (1984) *Science* **227,** 649–651.
20. Bates, P. L. & Swift, R. A. (1983) *Gene* **26,** 137–146.
21. Moriuchi, T., Chang, H.-C., Denome, R. & Silver, J. (1983) *Nature (London)* **301,** 80–82.
22. Maniatis, T., Fritsch, E. & Sambrook, J. (1982) *Molecular Cloning: A Laboratory Manual* (Cold Spring Harbor Laboratory, Cold Spring Harbor, NY).
23. Grosveld, T., Murray, E. J., Mellor, A. L., Dahl, H. H. M. & Flavell, R. A. (1982) *Nucleic Acids Res.* **10,** 6715–6731.
24. Birnboim, H. C. & Doly, J. (1979) *Nucleic Acid* **7,** 1513–1523.
25. Maxam, A. M. & Gilbert, W. (1977) *Proc. Natl. Acad. Sci. USA* **74,** 560–564.
26. Chirgwin, J. M., Przybyla, A. E., MacDonald, R. J. & Rutter, W. J. (1979) *Biochemistry* **18,** 5294–5299.
27. Aviv, H. & Leder, P. (1972) *Proc. Natl. Acad. Sci. USA* **69,** 1408–1413.
28. Lehrach, H., Diamond, D., Wozney, J. M. & Boedtker, H. (1977) *Biochemistry* **16,** 4734–4742.
29. Steinmetz, M., Freling, J., Fisher, D., Hunkapiller, T., Pereira, D., Weissman, S. M., Uehara, H., Nathenson, S. & Hood, L. (1979) *Cell* **24,** 125–134.
30. Larhammar, D., Schenning, L., Gustafsson, K., Wiman, K., Claesson, L., Rask, L. & Peterson, P. A. (1982) *Proc. Natl. Acad. Sci. USA* **79,** 3687–3691.