Structure of the murine Thy-1 gene

Vincent Giguère, Ken-Ichi Isobe¹ and Frank Grosveld

Laboratory of Gene Structure and Expression, National Institute for Medical Research, The Ridgeway, Mill Hill, London NW7 1AA, and ¹Department of Zoology, University College London, Gower Street, London WC1E 6BT, UK

Communicated by A. F. Williams

We have cloned the murine Thy-1.1 (AKR) and Thy-1.2 (*Balb*/c) genes. The complete exon/intron structure and the nucleotide sequence of the Thy-1.2 gene was determined. The gene contains four exons and three intervening sequences. The complete transcriptional unit gives rise to a tissue and developmental stage-specific mRNA of 1850 bp. The 5' end of the gene has multiple initiation sites and a non-TATA box promoter. The 3' end shows a single polyadenylation site after a very long untranslated region.

Key words: Thy-1/gene structure

Introduction

The Thy-1 antigen is a small glycoprotein found in large amounts at the cell surface of brain cells and rodent thymocytes, and in lesser quantity on a number of other tissues (Campbell et al., 1981). In many of these tissues, the level of expression of Thy-1 has been shown to vary during differentiation (Owen and Raff, 1970; Schachner and Hämmerling, 1974; Bash and Berman, 1982). In the mouse, Thy-1 occurs in two allotypic forms referred to as Thy-1.1 (AKR) and Thy-1.2 (Balb/c and others) (Reif and Allen, 1964) encoded by a gene that maps on chromosome 9 (Blankenhorn and Douglas, 1972). Structural and sequence homologies between the Thy-1 glycoprotein and immunoglobulin (Ig) domains and other members of the Ig superfamily (Cohen et al., 1981; McNicholas et al., 1982; Mostov et al., 1984; Clark et al., 1985) have led to the suggestion that Thy-1 may represent the primordial Ig domain (Williams and Gagnon, 1982) lending support to the argument that the immune system may have evolved from the nervous system. Thus, the study of the Thy-1 gene is interesting, both from the point of view of its structure and the mechanisms which control its tissue-specific expression.

We describe here the isolation of the mouse Thy-1.1 and Thy-1.2 genes. We present the complete exon/intron organization of the mouse Thy-1.2 gene, including the structure of its 5' end exon and flanking promoter region. The results show that the genomic organization of the Thy-1 gene is similar to that of the other members of the Ig superfamily. The exceptions are the 5' end exon and the promoter region which are different from those of the Ig superfamily and from the majority of previously studied eukaryotic genes.

Results

Isolation of the genomic fragment encoding the mouse Thy-1.2 gene

To isolate the mouse Thy-1.2 gene, we used a rat Thy-1 cDNA (Moriuchi et al., 1983) to screen a mouse cosmid library con-

structed from a partial *MboI* digestion of WEHI 3B cell line DNA ligated to the vector pTCF (Grosveld *et al.*, 1982b). Approximately 4 x 10⁵ recombinants were screened, and we isolated one clone that hybridized strongly to the rat Thy-1 cDNA probe. This clone, pMT1, contains a 40 kb insert which was mapped using conventional restriction enzyme double digests. Only the *Eco*RI restriction map is shown in Figure 1. Hybridization with the rat cDNA showed that the mouse Thy-1.2 gene was located on a single 8.2-kb *Eco*RI fragment. To facilitate sequence analysis and study of the expression of the mouse Thy-1.2 gene, the 8.2-kb *Eco*RI fragment was subcloned into the expression vector pBSV (Grosveld *et al.*, 1982a), giving rise to the plasmid pBST1, to allow expression from the Thy-1 promoter in the presence of an SV40 enhancer.

We then established that the 8.2-kb EcoRI genomic fragment contains the entire functional Thy-1.2 gene by sequence analysis (see below) and DNA-mediated gene transfer. Human HeLa cells were transfected with the subclone pBST1, collected 2 days later and assayed for the presence of the Thy-1.2 antigen on the cell surface using a monoclonal antibody against Thy-1.2 (Lake, 1979). The HeLa cells show 30% of the binding seen with the positive control BW5174, the T-cell lymphoma cell line (Goldsby et al., 1977, data not shown). The results indicate that the HeLa cells express the mouse Thy-1.2 antigen very efficiently on the cell surface when transformed with pBST1. If we assume a transformation efficiency of 20-40% (Banerji et al., 1981), the transformed HeLa cells actually express the Thy-1 antigen at the same level as the positive control BW5174. Since it is known that the introduced gene is usually transcribed from its own promoter in this system (Banerji et al., 1981), it is very likely that the 8.2-kb EcoRI fragment contains the entire transcriptional unit of the mouse Thy-1.2 gene.

Nucleotide sequence and complete exon/intron structure of the mouse Thy-1.2 gene.

To study the intron/exon organization of the Thy-1.2 gene we obtained additional cDNA clones from a rat thymocyte cDNA library (Clark et al., 1985) because the pT64 cDNA (Moriuchi et al., 1983) represents only part of the Thy-1 mRNA transcript. One of the new cDNA clones, pT2, with the longest insert (1600 bp) was selected for complete sequence analysis. At the same time we isolated all the mouse Thy-1.2 genomic fragments from pBST1 which hybridized to pT2. The strategy used to sequence these fragments, which contain the entire transcriptional unit as well as 600 bp of 5'-flanking sequences, is shown in Figure 1. Comparison of the genomic pBST1 sequence with the sequence obtained from the pT2 cDNA permitted the precise identification of the exon/intron boundaries; all the splice sites are of the type 1 as in the other genes of the Ig superfamily (Sharp, 1981) and in agreement with the consensus sequence at exon/intron boundaries (Benoist et al., 1980). The 3' end of the pT2 sequence agrees with a polyadenylation signal at position 6122 (Figure 2) and represents the 3' end of the mRNA (see below). Lastly, in contrast to the previously reported structure of the gene (Evans



Fig. 1. Organization and sequencing strategy of the genomic clone pMT1 containing the *Balb/c* mouse Thy-1.2 gene. Only the restriction sites used for sequencing are marked. The arrows indicate the direction and extent of the fragments sequenced. Exons are depicted as boxes, introns and flanking regions as lines. 5'UT, 5'-untranslated region; SS, signal sequence peptide; E, extracellular domain; TM, transmembrane-like domain; 3'UT, 3'-untranslated region.

et al., 1984; Seki *et al.*, 1985), the first 27 bp at the 5' end of the pT2 sequence indicate the existence of at least one more exon 2.6 kb upstream to bring the total number to four exons (see 5' end analysis below).

The exon/intron organization of the mouse Thy-1.2 gene corresponds approximatively to the functional domains of the protein, and is similar to the other members of the Ig superfamily (Hood *et al.*, 1985). The exception is the 5'-untranslated region of the mRNA which is encoded by two exons, exon 1 and part of exon 2 which also contains the signal peptide. Exon 3 codes for the external domain of the protein. Exon 4 codes for a transmembrane-like domain and an unusually long 3'-untranslated region.

3' End analysis of the mouse Thy-1.2 gene

On the basis of their first analysis of a rat Thy-1 cDNA, Moriuchi et al. (1983) proposed that the Thy-1 gene could be subject to differential polyadenylation. For a complete S1 analysis of the 3' end of the Thy-1 mRNA, we used two different probes derived from genomic DNA. The first one (Figure 3B) is a 344-bp Hinfl fragment extending from nucleotide 5932 to 6276 (see Figure 2 for numbering) and covers the polyadenylation signal AATAAA located at position 6122. The second is an exon/intron border probe which corresponds to an 805-bp fragment extending from a BstEII site located at position 4255 in the third exon to an ApaI site situated in the fourth exon (Figure 3C), and covers the proposed polyadenylation signal AATAAA at position 4543 (Moriuchi et al., 1983). These probes were hybridized to total RNA from BW5147 cells, the heteroduplexes were digested with S1 nuclease, and the length of the protected fragments was determined on a denaturing 5% polyacrylamide gel. As shown on Figure 3A, the Hinfl probe generates two protected fragments of ~ 210 nucleotides, which would situate the polyadenylation site $\sim 12 - 15$ bp downstream the polyadenylation signal at position 6122. This is in agreement with the sequence of the rat Thy-1 cDNA clone pT2 (Figure 4B), which puts the polyadenylation site 12 bp from the sequence AATAAA.

A single protected fragment of 310 nucleotides in length is observed when the BstEII/ApaI probe is used (Figure 3A, lane 4), which corresponds exactly to the border between exon 3 and intron 4. We did not detect any bands at 300 nucleotides, corresponding to the poly(A) site in the third exon as suggested by Moriuchi et al. (1983). Our results are in agreement with the finding that only one species of mRNA of ~1850 nucleotides is detected with a mouse Thy-1 probe on a Northern blot of total RNA from BW5147 cells (Figure 5) or brain and thymus mRNA (not shown). Interestingly, the same mRNA is developmentally regulated in F9 embryonal carcinoma cells. Figure 5 (lanes 2 and 3) shows that the Thy-1 mRNA is undetectable before differentiation but switched on after differentiation of the F9 cells with retinoic acid (Hogan et al., 1981). Consequently it appears that the same transcriptional unit is utilized in different tissues and in early development since it is active in F9 cells, which are thought to represent early embryonic cell types giving rise to parietal and visceral endoderm (Hogan et al., 1981).

5' End analysis of the mouse Thy-1.2 gene

The sequence analysis revealed that the only sequence which matches the first 27 bp of the rat Thy-1 cDNA clone pT2, is located 2.6 kb upstream from the exon containing the initiation codon methionine (Figures 2 and 4A). In order to define the precise 5' end of the mouse Thy-1 mRNA, we have used a 5'-labelled 31-nucleotide Hinfl/HhaI fragment which is part of the putative first exon as a primer for extension. The result of this analysis (Figure 6A) shows three major extension products; the longest corresponding to an A residue at position 594 in Figure 2, the two other bands correspond to a G and an A residue at position 595 and 597, respectively. Overexposure of the film to the gel reveals additional weak bands corresponding to 5' end positions indicted by open triangles in Figure 2. To exclude the possibility that the extended products seen by primer extension correspond to strong stops for the reverse transcriptase, we confirmed the 5' end with an S1 nuclease protection experiment using total RNA from BW5147 cells. A 520-nucleotide PstI fragment

	10	20	30	40	50	60	70	80	90	100	
0	GTTTCCCTAACAGGGG	CCAAGTGTTTT	GAGTAACTGA	AGGTGGGCATG	GATGCCTAGGA	AGCGAAACAA	GCCAAGCAG	ATGCACCCCT	TGCCTTTCTT	CCGA	
100	AGGGCTGCAGTAGCAT	rggaaaagat g	GAAAACAACC	AATCCATTCCC	CTTTGCTGATA	TAACAGGCTC	CAAAGCCAA	AACCTGTCAC	TGGAGGCTCA	AGAG	
200	CAGATCTCCAGCCAAC	GAGGCAAAGGA	ATGGOGGAAG	CTGC AGGGCCT	CCCTCTGGTT	ATCCAGGCTI	CTGAAGGTT	CAAGCAAAGA	AAGGGTTACA	ACCT	
300	TAAAAGGAGAGCGTCC	CCGGGGTATGG	GTAGAAGACT	GCTCCACCCCG	ACCCCCAGGG	TCCCTAACCG	TCTTTTCCT	TGGGCGAGTC	AGCCCAATCA	CAGG	
4 00	ACTGAGAGTGCCTCTT	TAGTAGCAGO	AAGCCACTTC	GGACACCCAAA	TGGAACACCI	CCAGTCAGCC	CTCGCCGAC	CACCCCACCC	CCTCCATCCT	TTTC	
500	CCTCAGCCTCCGATTO	GCTGAATCTA	GAGTCCCTCC	CTGCTCCCCC	TCTCTCCCCA			GCTTCAGCGC	TGGGTGCAGC	AACT	EULT
600	GGAGGCGTTGGCGCA	CTAGGAGGAGG	CTGCAGCTAG	GGGAGTCTAGG	TGAGAGCAGG	CCGACGGGGAG	GGACCCGCA	CATGCAAGGA		GCGG	501
700	GATGCAAGCCTTCCCC	CAGCTACAGTI	TTGGGAAAGG	ATACTAAGGCG	CTCCTATATG	GGGGGCGCGGG	AACTTGGGG	AAAGAAGGTG	CTCCCAAGTC	GAGG	
800	TGGGAGAGGAAGGCAG	GTGCGGGGTCA	CGGGCTTTCT	CCCTGCTAACG	GACGCTTTCG	AAGAGTGGGT	GCCGGAGGA	GAACCATGAG	GAAGGACATC	AAGG	
900	ACATCAAGGACAGCC	TTGGTCCCCA	AGCTCAAATC	GCTTTAGTGGT	GCGAATAGAG	GGAGGAGGTG	GGTGGCAAA	CTGGAGGGAG	CCCCCAACGG	GTGA	
1000	CCTCGTGGCTGGCTGG	GTGCGGGGTA	GTAGGTAAGA	AAACCGCAATG	TTGCGGGAGG	GGACTGGGTG	GCAGGCGCG	GGGGAGGGGA	AGCTAGAAA	GAT	
1100	GCGAGGGAGCGGAGGG	GGGAGGGAGC	GGGGGGAATCT	CAACTGGTAGA	GGAAAGTTAA	AATGAGGAAA	TAGCATCAG	GGTGGGGTTA	CCAAGCCGG	GCCT	
1200	CAGGGAAAGGGCGCAA	AGTTTGTCTG	GGTGTGGGCT	TAGGTGGGCTG	GGTATGAGAT	TCGGGGGCGCC	GAAAACACT	GCTGCGCCTCI	GCCAAATCA	CGCT	
1 300	ACCCCTGTATCTAGT	CTGCTAGGCT	TCTCCAGCCC	CAGCCCCAATT	CTTTTCTCAG	TGTCCCCTTC	CCTCCCCTG	AATCTCAAGCO	CACACTCCC	гсст	
1 4 00	CCATAACCCACTGTT	TCAAATCTAA	GTCATTTGCC	ACCCAACAACC	ATCAGGAGGC	GGAAGCAGAC	GGGAGGAGT	TTGAGATCAA	TTGGGCTAC	ATCA	
1500	CGAGTTCCAAGCTCAC	CCAAGGCTTCT	TAAGGAGACC	TTGTCTCTAAA	ATTAATTAAT	TAATTAATTA	ATAGTCCCC	TTTCTCTGCC.	CAGAACCTTO	GGA	
1600	TCTGGCTCCTGGTCGC	AGCTCCCCC	ACCCCAGGCT	GACATTCACTG	CCATAGCCCA	TCCGGAAATC	CTAGTCTAT	FTCCCCATGGA	TCTTGAACTO	GCAG	
1700	AGAGAATGGCAGAGTG	GCCCGCCCTG	TGCAAAGGAT	GTTCCTAGCCT	AGGTGGAGCT	CGCGAACTCG	CAGACTGTG	CTCTCTTGGG	CAAGGACAGG	CTA	
1800	GACAGCCTGCCGGTG1	GTTGAGCTAG	GGCACTGTGG	GG		About 1	.4 Kb				
							-19 Met	AsnProAlsIl	leSerValAla	iLeu	
3300	CACCTGTCCTACC	CAGCTGGTTGA	CCTGTAGCTT	TCCCCACCACA	GAATCCAAGT	CGGAACTCTT	GGCACCATG	AACCCAGCCAT	CAGCGTCGCT	CTC	CC
3400	LeuLeuSerV CTGCTCTCAG <u>GT</u> ACTC	GGCAAGGGTC	AGGGCTGGCA	TTCTAAGGAAT	••••••	about	0.6 КЪ				33
4100		ACCTCAAGGGG	GAGTGCAAAC	CCCACATCCCC	CCAGCTCAAG	TCTACTTTTC	CTGCAGGTG	GAGGCCCCGG	GTCTGTGTCI	rccc	
4 200	CAAATTCAGAGAAGGG	CACTGCTGTGC	alLeuGl C <u>AG</u> TCTTGCA	nValSerArgG GGTGTCCCGAG	-1 1 lyGlnLysVa GGCAGAAGGT	lThrSerLeu GACCAGCCTG	ThrAlaCys ACAGCCTGC	10 LeuValAsnGl CTGGTGAACCA	InAsnLeuArg	Leu CTG	
	20 AspCysArgHisGlu/	AsnAsnThrLy	sAspAsnSer	30 IleGlnHisGl	uPheSerLeu	ThrArgGluL	40 ysArgLysH	isValLeuSer	GlyThrLeuG	50 31yI	
4300	GACTGCCGCCATGAGA	ATAACACCAA	GGATAACTCC	ATCCAGCATGA	GTTCAGCCTG	ACCCGAGAGA 70	AGAGGAAGC	ACGTGCTCTCA	IGGCACCCTTG 80	GGA	F
44 00	TACCOGAGCACACGTA	VrArgSerArg ACCGCTCCCGC 90	GTCACCCTCT	erAsnGlnPro CCAACCAGCCC 1	TyrlleLysV TATATCAAGG 00	alLeuThrLe TCCTTACCCT	uAlaAsnPh AGCCAACTT	eThrThrLysA CACCACCAAGG	ABPG1uG1yAe ATGAGGGCGA	арТу СТА	-
4500	CTTTTGTGAGCTTCA/	AGTCTCGGGGCG	CGAATCCCAT	GAGCTCCAATA	AAAGTATCAG	TGTGTATAGA	A G <u>GT</u> GAGACT(GTTCCCAGAG	AGATAAAATG	TCT	
4 600	AGGTTAGCTAGGCTGC	GGTAGCCAAT	*******	AAAAAACGGGC	ACCTCCATTA	CCCTTCCCCT	AACTGCTGG	TCTCCTGGGAA	ACTGCTGCTG	TCT	
4700	ATGTGAGTGGGGCAAC	ATTAGGGG	Abo	ut 0.16 Kb.	c	CGCCTGAGTC	CTGATCTCC		CACCCCTTCT	CTA	
4900	BpLysLeu TCCACAGACAAGCTGG	/slLysCysGl STCAAGTGTGG	yGlyIleSer CGGCATAAGC	LeuLeuValG1 CTGCTGGTTCA	nAsnThrSer GAACACATCC	TrpMetLeuL TGGATGCTGC	euLeuLeuL TGCTGCTGC	euSerLeuSer TTTCCCTCTCC	LeuLeuGlnA	alaL CCC	ТИ
	140 14 euAspPheIleSerLe	13 PuTER									1 1 1 1
5000	TAGACTTCATTTCTCI	GTGACTGGTT	GGGCCCAAGG.	AGAAACAGGAG	CCCTCGAGGT	CCTTCCTCTG	CAGAGGTCT	IGCTTCTCCCG	GTCAGCTGAC	TCC	
5100	CTCCCCAAGTCCTTCA	LAATATCTCAG	AACATGGGGA	GAAACGGGGGAC	CTTGTCCCTC	CTAAGGAACC	CCAGTGCTG	CATGCCATCAT	CCCCCCCACC	CTC	
5200	GCCCCCACCCCCGCCA	CTTCTCCCTC	CATGCATACC	ACTAGCTGTCA	TTTTGTACTC	TGTATTTATT	CTAGGGCTG	CTTCTGATTAT	TTAGTTTGTT	CTT	
5300	TCCTTGGAGACCTGTI	AGAACATAAG	GGCGTATGGT	GGGTAGGGGAG	GCAGGATATC	AGTCCCTGGG	GCGAGTTCC'	CCCTGCCAAC	CAAGCCAGAT	GCC	
5400	TGAAAGAGATATGGAT	GAGGGAAGTT	GGACTGTGCC'	TGCACCTGGTA	CAGTCACACT	CTGTTGAAAG	AATCATCGG	GAGGGGGGGGGG	GGCTCAAGAG	iggg	
5500	AGAGCTCTGCTGAGTC	CTTTGTGGACC	ATCCAATGAG	GATGAGGGCTT	AGATTCTACC	AGGTCATTCT	CAGCCACCA	CACACAAGCGC	TCTGCCATCA	CTG	3'UT
5600	AAGAAGCCCCCTAGGG	CCTTGGGCCA	GGGCACACTC.	AGTAAAGATGC	AGGTTCAGTC	AGGGAATGAT	GGGGAAAGG	GTAGGAGGTG	GGGGGAGGGAT	CAC	
5700	CCCCTCCTCTAAAACA	CGAGCCTGCT	GCCTCCAAAG	GCCTCTGCCTG	TAGTGAGGGT	GGCAGAAGAA	GACAAGGAG	CAGAACTCTG	ACTCCAGGAT	CTA	
5800	AGTCCGTGCAGGAAGG	GGATCCTAGA	ACCATCTGGT	TGGACCCGGCT	TACCAAGGGA	GAGCCTTTAT	TCTTCTTTT	CCTTGCCCCT	CTGTGCCAGO	ccc	
5900	TCTTGCTGTCCCTGAT	CUCCCCAGACA	GCGAGAGTCT	IGCAACCTGCC	TCTTCCAAGA	CTCCTAATC	TCAGGGGGCA	GCGGTGGAGT	GAGATCCGGC	GTG	
6000	CACACTTTTTGGAAGA	TAGCTTTCCC	AAGGATCCTC	TUCCCCACTGG	CAGCTCTGCC	TGTCCCATCA	CCATGTGTA	ATACCACCACT	GCTACAGCGI	CTC	
6100	ACCUACGAAAGAAAAC	TGUACAATAA	AAGCCAAGCCT	UTGGAGTGTGT	CCTGGTGTCT	GICTCTTCTG	TGTCCTGGC	JTCTGTCTCTT	CTGTGTTCTI	CCA	
6200	AGGTCAGAAACAAAAA	UCACACACTT	CAACCTAGAT	JUCTCGGCTGA	GCACTTCTGT	GTGCAGAAGG	TCCAACCGA	CTCTGGGTACC	CGCCCTCCC	TAT	
6300	TUCCTTGCCTCCTGTC	TCCCGCTTTT	TATAGCTCCC	ratgctgggct	TUTUTGGAGA	GTGAAATCTT	TGCCCAAAT	CAATGCGCATI	CTCTCTGCTC	GAGT	
6400	CATCTGGCGACAGCAG	TTTGAGTTCA	UUUUGUCAAACI	AUATGGGCCCA	GUTATGTAG						

Fig. 2. DNA sequence of the mouse Thy-1.2 gene. The amino acid sequence is shown above the DNA sequence. The splice signal dinucleotides GT and AG and the polyadenylation signal AATAAA are underlined. The location of the mRNA cap sites as determined by S1 nuclease and primer extension analysis are indicated by circles and triangles, respectively. Open symbols indicate weak bands on the autoradiographs.

(Figure 6C) was 5' end-labelled with polynucleotide kinase, hybridized and treated with S1 nuclease. In agreement with the primer extension, more than one band is observed on the gel (Figure 6B). The longest protected fragment of 36 nucleotides corresponds to a G residue at position 592 while other bands correspond to the same positions as mapped by the primer exten-



Fig. 3. Mapping the polyadenylation site at the 3' end of the Thy-1 mRNA. Total BW5147 cells RNA (25 μ g, lanes 2 and 4) or yeast tRNA (10 μ g, lanes 1 and 3) was annealed to 3' end-labelled probes and RNA/DNA hybrids were digested with S1 nuclease (Weaver and Weissman, 1979). The autoradiogram is shown in panel A. The *Hinfl* 3' end probe shown on panel B was used in lanes 1 and 2, the *BstEII*/Apal exon 3/intron 4 junction probe shown on panel C was used in lanes 3 and 4. The markers were pBR322 x *Hinfl*. P indicates the positions of the input probes and arrows the positions of the protected fragments.

sion analysis. It should be noted that the characteristic TATA (and CAAT) box usually present in eukaryotic promoters are absent in the region upstream of these putative cap sites, instead, it contains an A-rich sequence at -26 and several CG-rich areas as found in other non-TATA box promoters, in particular, the hexanucleotide CCTTGGTG at -34 which is also found in the mouse dihydrofolate reductase (DHFR) promoter (Crouse *et al.*, 1982).

To show that these sequences can initiate transcription, the 520-bp PstI fragment containing the Thy-1 gene 5' end was cloned in front of the chloramphenicol transferase (CAT) gene with HindIII linkers to yield the plasmid pBr.T1.CAT. In addition to pBr.T1.CAT, we made a further construct (pBr.T1.SV.CAT) by inserting the SV40 enhancer element at the 3' side of the CAT gene (Figure 7). Both of these constructs and the control plasmids pSV2CAT and pSV1CAT (Gorman et al., 1982) were introduced into L-cells by calcium phosphate precipitation (Wigler et al., 1979) and in BW5147 cells (not shown) with DEAE dextran (Deans et al., 1984). After 60 h the cells were harvested and analyzed for CAT activity (Gorman et al., 1982). The L-cell result (Figure 8) shows that the Thy-1 promoter (pBr.T1.CAT) is comparable with the SV40 promoter (pSV1.CAT) without enhancer (lanes 2 and 4, Figure 8). The Thy-1 promoter is stimulated to efficient expression by the insertion of the SV40 enhancer 3' to the CAT gene (pBr.T1.SV.CAT, lane 3), and is comparable with the SV40 promoter plus enhancer (pSV2.CAT, lane 1). A similar result is obtained with the BW5174 cells at a much lower efficiency due to a very inefficient rate of transformation. The stimulated transcription with the SV40 enhancer is in agreement with the observed efficient expression of the whole Thy-1 gene (see above) in the expression plasmid pBSV which also cotains an SV40 enhancer (Grosveld et al., 1982a).

Comparison of the mouse Thy-1.1 gene and Thy-1.2 gene

Southern blot analysis of several different mouse strains showed restriction fragment length polymorphisms (RFLPs) for the Thy-1 gene, most notably between the Thy-1.2 (*Balb/c*) gene and the Thy-1.1 (AKR/j) gene. The product of these genes has been reported to differ by a single amino acid at position 89 (Figure 2; Cotmore *et al.*, 1981; Williams and Gagnon, 1982). However, of the fragments generated by treatment with restriction enzymes tested, only the *Eco*RI fragments were very similar (data not shown). To analyze this in more detail, the Thy-1.1 gene from AKR/j mice was cloned as an 8.2-kb *Eco*RI fragment in λ phage

Α							
	Cap	Intron 1					
Mouse Thy-1.2 gene	 GCAGCAACTGGAGGCGTTGGCGACTAGGAGGAGGCTGC-AGCTAGGGGAGTCT. 	AGAATCCAAGTCGG					
Rat Thy-1 cDNA	AGGAGGCTGCAAGCTAGGGGAGCCC	AGACCCAGGACGG					
В							
	Polyadenylation s	ite					
Mouse Thy-1.2 gene	AGGAAAGAAAACTGCAC <u>AATAAAA</u> CCAAGCCTCTGGAGTGTGTCCT						
Rat Thy-1 cDNA	AAGGAAGAGAACTGCAC <u>AATAAAA</u> CAAAGCCTGTGGAAAAAAAAAAAAAAA						

Fig. 4. Comparison of the mouse Thy-1.2 genomic DNA sequence with the rat Thy-1 cDNA clone pT2. (A) exon I of the mouse Thy-1.2 gene as compared with the 5' end of the rat cDNA. (B) sequences polyadenylation signal at the 3' end of the cDNA and the mouse Thy-1.2 gene. (|) indicates nucleotide differences.



Fig. 5. Northern blot analysis of the Thy-1 mRNA. Lane 1, 25 μ g total BW5147 cell RNA. Lanes 2 and 3, 10μ g poly(A)⁺ RNA obtained from undifferentiated and differentiated F9 cells, respectively. The probe used to hybridize these blots is the *BstEII/ApaI* restriction fragment shown in Figure 3, panel C. Size markers are mouse ribosomal 28S and 18S RNA and 4S tRNA.

641 (Murray et al., 1977). Twelve positive in 10⁵ plaques were identified and the 8.2-kb insert was subcloned in the expression vector pBSV. As expected, sequence analysis showed a single base pair substitution $(A \rightarrow G)$ at the amino acid 89 position changing the Gln to Arg, to account for the two different allotypes of the murine Thy-1 antigen. In addition to this, both sequence and restriction enzyme mapping showed additional differences between the genes (Figure 9); a 30-bp and 200-bp deletion/insertion at ~ 1100 bp and 800 bp upstream from the CAP site, respectively, a 160-bp insertion/deletion in the first intron, a 13-bp insertion/deletion in the third intron and, finally, an 11-bp insertion in the last exon. Moreover, a number of restriction sites (KpnI, NcoI, XbaI) have changed in the first intron. All of these changes have been confirmed by genomic blotting data (not shown). Increstingly, when the Thy-1.1 gene from AKR/j is compared with the partial sequence of the Thy-1.2 gene from C57.BL/6 (Seki et al., 1985), the homology is higher than the homology found between the Thy-1.2 gene from C57.BL/6 and the Thy-1.2 gene from Balb/c.

Discussion

We have cloned the genes encoding the two allotypic forms of the mouse Thy-1 antigen and the structure of the mouse Thy-1.2



Fig. 6. Mapping the sites of transcription initiation by primer extension and S1 nuclease analysis. Panel A shows the result of the primer extension analysis using a 31-nucleotide primer complementary to the position 612-643 (see Figure 2) and 25 μ g of total RNA from BW5147 cells (lane 2) or 10 μ g yeast tRNA (lane 1). Panel B shows the result of the S1 nuclease analysis. Total RNA from BW5147 cells (25 μ g, lane 2) or yeast tRNA (10 μ g, lane 1) was annealed to a 5' end-labelled probe and subjected to S1 nuclease digestion. Panel C shows the *PstI* 5' end probe used for the S1 nuclease the position of the input primer or the input probe. M indicates the pBR322 x *Hinf*1 marker fragments. Arrows indicate the major extended product or protected fragment in the respective experiments.

gene has been determined. The gene consists of four exons, which correspond approximately to the functional domains of the protein. The transcription unit of the mouse gene spans 6 kb of genomic DNA of which the coding and non-coding regions total 1700 bp after processing. This is consistent with a size of 1850 nucleotides determined for the Thy-1 mRNA on Northern blots (Figure 5), assuming a poly(A) tail of ~ 150 nucleotides (Sawiki *et al.*, 1977). Using primer extension and S1 nuclease mapping assays, we have defined the borders of the transcriptional unit. The Thy-1 mRNA has multiple sites of transcription initiation at the 5' end, and a single polyadenylation site at the 3' end of the gene. This last finding contradicts Moriuchi *et al.* (1983), indicating a cloning artefact rather than their suggestion that the Thy-1 gene is subject to differential polyadenylation.



Fig. 7. Construction of CAT expression vectors containing the 5' region of the Thy-1 gene and the SV40 enhancer. pBR CAT was constructed as described (Walker *et al.*, 1983). The 520-bp *PstI* fragment containing the Thy-1 promoter was obtained from pBSTI. The *PvuII/NcoI* restriction fragment containing the SV40 enhancer was obtained from pSV2-CAT (Gorman *et al.*, 1982). Arrows indicate the direction of transcription. Enzymes used: H, *Hind*III; B, *Bam*HI; P, *PstI*; N, *NcoI*; Pv, *PvuII*; RT, reverse transcriptase.



Fig. 8. Assay of CAT activity in mouse L-cells transfected with CAT expression vectors. Cells were transfected with 10 μ g of DNA and assayed 60 h later for CAT activity. Lane 1, pSV2-CAT; lane 2, pSV1-CAT; lane 3, pBRT1-SV-CAT; lane 4, pBRT1-CAT. CM, chloramphenicol; Ac1, 1-acetylchloramphenicol; Ac3, 3-acetyl-chloramphenicol.

Although the physiological role of Thy-1 is still not clear, the analysis of its exon/intron structure fully confirms its homology with the Ig superfamily of proteins (Williams and Gagnon, 1982), the different functional domains of the protein are encoded on separate exons with the exception of the 5' leader sequence of the gene which is divided over two exons. This property is shared by β 2-microglobulin, the only other published protein in the Ig superfamily that contains a single extra cellular domain. Thy-1 might therefore be more closely related to β 2-microglobulin than was suggested by Hood *et al.* (1985). Comparison of the Thy-1 genes from different mouse strains shows that the Thy-1.1 gene is more similar to the C57BL/6 Thy-1.2 (Seki *et al.*, 1985) than is the *Balb*/c Thy-1.2 gene. At present we do not know whether

this indicates a high degree of polymorphism of the Thy-1 gene or whether this represent a very limited number of variant genes propagated by inbreeding.

The mouse Thy-1 glycoprotein is expressed in a developmental stage- and tissue-specific manner, its highest levels being found at the cell surface of thymocytes and brain cells. Recently, reintroduction of a cloned Thy-1.2 gene into different cell lines showed that the resulting transformants express up to 50-fold more cell surface Thy-1.2 antigen when the cell lines transformed already expressed the Thy-1 antigen at its cell surface (Evans et al., 1984). Although no data on the levels of the Thy-1 mRNA were presented, these results suggest that the Thy-1 gene might be subject to cell-specific transcriptional control. Thus, analysis of the mouse Thy-1 gene structure is a crucial first step toward studying the mechanisms which regulates its expression. Interestingly, sequencing analysis reveals that the Thy-1 promoter is unusual. The immediate 5'-flanking region of the putative cap site contains a relatively high G/C content (68%) and lacks the conserved TATA and CAAT boxes normally found in the majority of eukaryotic promoters (Corden et al., 1980). Few promoters which share these unusual characteristics have been described. They include viral promoters (Baker et al., 1979; Baker and Ziff, 1981) the human β -tubulin gene (Lee *et al.*, 1983) as well as the genes encoding different 'housekeeping' enzymes (Crouse et al., 1982; Melton et al., 1984; Reynolds et al., 1984; Valerio et al., 1985). For most of these genes, the absence of a conserved TATA box produces mRNAs whose 5' termini are heterogeneous, a phenomenon also observed for the Thy-1 mRNA. Short regions of homology have been observed between certain of these promoters, especially with the sequence CCGCCC present in the SV40 early transcriptional unit (Reynolds et al., 1984; Melton et al., 1984). However, only one strong region of homology could be detected when the Thy-1 promoter was compared with the above-mentioned genes at position -34 with the corresponding region of the mouse DHFR gene (Crouse et al., 1982). This makes the Thy-1 promoter unique and deletion mutant studies are in progress to identify sequences responsible for the correct initiation of the Thy-1 mRNA per se.



Fig. 9. Comparison of the Thy-1 locus in different mouse strains. Open boxes indicate exons. Different size triangles show different size DNA insertions and () indicates the presence of a restriction site. N, NcoI; K, KpnI.

In addition, preliminary results suggest that more than just the 5' end of the gene is required to obtain tissue-specific expression (Giguere, unpublished).

Previous studies (Stern *et al.*, 1975) have shown that the Thy-1 antigen can be detected on differentiated embryonal carcinoma cells. This observation is confirmed for F9 EC cells at the level of Thy-1 mRNA (Figure 5). At present we cannot distinguish between the two possible mechanisms for the increase of the mRNA; either stabilization of Thy-1 mRNA (which was rapidly turned over before differentiation) or alternatively, a transcriptional activation. Either way, since F9 EC cells can be transformed by DNA-mediated gene transfer and give rise to regulated expression of exogenous genes (Rosenthal *et al.*, 1984), this should provide an interesting system to study the early expression of Thy-1.

Materials and methods

Isolation of Thy-1 genomic clones

The murine Thy-1.2 gene was isolated from a cosmid pTCF library of a *Balb/c* WEHI 3B (Warner *et al.*, 1969) cell line DNA as described by Grosveld *et al.* (1982b). The library was screened with a Thy-1.2 cDNA insert as a probe (Moriuchi *et al.*, 1983). One positive clone was isolated and mapped by standard double restriction enzyme digests. The murine Thy-1.1 gene was cloned from AKR DNA which was enriched for the presence of the 8.2-kb Thy-1 *EcoRI* fragment by preparative gel electrophoresis. The purified material was cloned into λ 641 (Murray *et al.*, 1977) to yield a partial library of 10⁵ recombinant phage plaques. Twelve positives were picked after hybridization with a *Bst*EII-*ApaI* Thy-1.2 third exon probe. The 8.2-kb Thy-1 containing *EcoRI* fragments were isolated by gel electrophoresis and recloned in the expression vector pBSV (Grosveld *et al.*, 1982a) by standard procedures.

DNA-mediated transfection and CAT assay

The recombinant DNA construct, either with pBSV or the CAT plasmids were introduced into eukaryotic cells by calcium phosphate co-precipitation (Wigler *et al.*, 1979) or by DEAE-dextran as described by Deans *et al.* (1984). After transfection of HeLa cells with the pBST1 (Thy-1 pBSV) the cells were harvested after 2 days in culture (α MEM 10% NBCS) and analyzed for the presence of the Thy-1 antigen on the cell surface with a monoclonal anti-Thy-1 antibody by standard procedures. After transfection with the CAT gene constructs the mouse L-cells were grown for 72 h in α MEM 10% NBCS and prepared for CAT assays as described by Gorman *et al.* (1982).

Analysis of the Thy-1.2 mRNA

Total RNA was isolated from cells in culture by the guanidium/caesium chloride method (Glisin *et al.*, 1974; Ullrich *et al.*, 1977). Northern blot analysis was performed as described by Thomas (1983). S1 nuclease analysis was performed as described (Berk and Sharp, 1977; Weaver and Weissman, 1979) using probes end labelled by T4 polynucleotide kinase or reverse transcriptase to a specific activity $> 10^7$ c.p.m./µg.

For primer extension analysis, a 474-bp Xbal/BstEII restriction fragment was purified on agarose gel, cut with *Hinfl*, end-labelled by T4 polynucleotide kinase, recut by *Hha*I and repurified on an 8% polyacrylamide-7 M urea gel. The singlestranded 31 nucleotide long *Hinfl-Hha*I primer was eluted from the gel and its purity confirmed by sequencing. We suspended 25 μ g of total RNA and ~0.2 pmol of ³²P-labelled primer in 10 μ I of buffer containing 400 mM NaCl, 10 mM Pipes pH 6.4. The mixture was sealed in a glass capillary, denatured at 90°C for 2 min, and allowed to hybridize overnight at 50°C. To initiate primer extension, the glass capillary was broken and its contents delivered into reactive mixture containing 50 mM Tris HCl pH 8.2, 10 mM DTT, 6 mM MgCl₂, 0.5 mM of each dNTP, 2.5 μ g actinomycin D and 10 units of reverse transcriptase. The reaction was incubated at 41°C for 1 h and the extended products analyzed by electrophoresis on 8% polyacrylamide-7 M urea gel followed by autoradiography.

Acknowledgements

We are grateful to Roger James and Arnon Rosenthal for their assistance and to Cora O'Carroll for her speedy preparation of the manuscript. V.G. was supported by the Medical Research Council of Canada. K-I.I. was supported by a visiting I.C.R.F. fellowship. The work was supported by the Medical Research Council UK.

References

- Baker, C.C. and Ziff, E.B. (1981) J. Mol. Biol., 149, 189-221.
- Baker, C.C., Heresse, J., Courtois, G., Galibert, F. and Ziff, E. (1979) Cell, 18, 569-580.
- Banerji, J., Rusconi, S. and Schafner, W. (1981) Cell, 27, 299-303.
- Bash, R. and Berman, S. (1982) Eur. J. Immunol., 12, 359-364.
- Benoist, C., O'Hare, K., Breathnach, R. and Chambon, P. (1980) Nucleic Acids Rs., 8, 127-142.
- Berk, A. and Sharp, P. (1977) Cell, 12, 721-732.
- Blankenhorn, E. and Douglas, T. (1972) J. Hered., 63, 259-263.
- Campbell, D.G., Gagnon, J., Reid, K.B.M. and Williams, A.F. (1981) *Biochem.* J., 195, 1-30.
- Clark, M.J., Gagnon, J., Williams, A.F. and Barclay, A.N. (1985) EMBO J., 4, 113-118.
- Cohen, F.E., Novotny, J., Sternberg, M.J.E., Campbell, D.G. and Williams, A.F. (1981) *Biochem. J.*, 195, 31-40.
- Corden, J., Wasylyk, A., Buchwalder, A., Sassone-Corsi, P., Kedinger, C. and Chambon, P. (1980) Science (Wash.), 209, 1406-1414.
- Cotmore, S.F., Crowhurst, S.A. and Waterfield, M.D. (1981) Eur. J. Immunol., 11, 597-603.
- Crouse, G.F., Simonsen, C.C., McEwan, R.N. and Schimke, R.T. (1982) J. Biol. Chem., 257, 7887-7897.
- Deans, R.J., Denis, K.A., Taylor, A. and Wall, R. (1984) Proc. Natl. Acad. Sci. USA, 81, 1292-1296.
- Evans, G.A., Ingraham, H.A., Lervis, K., Cunningham, K., Seki, T., Moriuchi, T., Chang, H.C., Silver, J. and Hyman, R. (1984) Proc. Natl. Acad. Sci. USA, 81, 5532-5534.
- Glisin, V., Crkvenjakov, R. and Byus, C. (1974) Biochemistry (Wash.), 13, 2633-2637.
- Goldsby, R., Osborne, B., Simpson, E. and Herzenberg, R. (1977) Nature, 267, 707-708.
- Gorman, C., Moffat, L. and Howard, B. (1982) Mol. Cell. Biol., 2, 1044-1051.
- Grosveld,G.C., deBoer,E., Shewmaker,C.K. and Flavell,R.A. (1982a) Nature, 295, 120-126.
- Grosveld, F.G., Lund, T., Mellor, A.L. and Flavell, R.A. (1982b) Nucleic Acids Res., 10, 6715-6732.
- Hogan, B.L., Taylor, A. and Adamson, E. (1981) Nature, 291, 235-237.
- Hood, L., Kronenberg, M. and Hunkapiller, T. (1985) Cell, 40, 225-229.
- Lake, P. (1979) Eur. J. Immunol., 9, 875-879.
- Lee, M.G.-S., Lewis, S.A., Wilde, C.D. and Cowan, N.J. (1983) Cell, 33, 477-487.
- McNicholas, J., Steinmetz, M., Hunkapiller, T., Jones, P. and Hood, L. (1982) Science (Wash.), 218, 2339-1232.
- Melton, D.W., Konecki, D.S., Brennand, J. and Caskey, C.T. (1984) Proc. Natl. Acad. Sci. USA, 81, 2147-2151.

Moriuchi, T., Chang, H.-C., Denome, R. and Silver, J. (1983) Nature, 301, 80-82.

Mostov, K.E., Friedlander, M. and Blobel, G. (1984) Nature, 308, 37-43.

Murray, N.E., Brammar, W.J. and Murray, K. (1977) Mol. Gen. Genet., 150, 53-61.

- Owen, J. and Raff, M. (1970) J. Exp. Med., 132, 1216-1232.
- Reif, A. and Allen, J. (1964) J. Exp. Med., 120, 413-433.
- Reynolds, G.A., Basu, S.K., Osborne, T.F., Chin, D.J., Gel, G., Brown, M.S., Goldstein, J.L. and Luskey, K.L. (1984) Cell, 38, 275-285.
- Rosenthal, A., Wright, S., Cedar, H., Flavell, R.A. and Grosveld, F.G. (1984) Nature, 310, 415-418.
- Sawiki, S.G., Jelinek, W. and Darnell, J.E. (1977) J. Mol. Biol., 113, 219-235.
- Schachner, M. and Hämmerling, U. (1974) Brain Res., 73, 362-371.
- Seki, T., Chang, J.-C., Meriuchi, T., Denome, R., Ploegh, H. and Silver, J. (1985) Science (Wash.), 227, 649-651.
- Sharp, P.A. (1981) Cell, 23, 643-646.
- Stern, P.L., Martin, G.R. and Evans, M.J. (1975) Cell, 6, 445-465.
- Thomas, P.S. (1983) Methods Enzymol., 100, 255-265.
- Ullrich, A., Shine, J., Chirgwin, J., Pictet, R., Tischer, E., Rutter, W.J. and Goodman, H.M. (1977) Science (Wash.), 196, 1313-1319.
- Valerio, D., Duyvesteyn, M.G.C., Dekker, B..M., Weeda, G., Berkvens, T.M., van der Boom, L., van Ormonde, H. and van der Eb, A.J. (1985) *EMBO J.*, 4, 437-443.
- Walker, M.D., Edlund, T., Boulet, A.M. and Rutter, W.J. (1983) Nature, 306, 557-561.
- Warner, N.L., Moore, M.A.S. and Metcalf, D. (1969) J. Natl. Cancer Inst., 43, 963-982.
- Weaver, R. and Weissman, S. (1979) Nucleic Acids Res., 7, 1175-1193.
- Wigler, M., Sweet, R., Sim, G.K., Wld, B., Pellicer, A., Lacy, E., Maniatis, T., Silverstein, S. and Axel, R. (1979) Cell, 16, 777-785.
- Williams, A.F. and Gagnon, J. (1982) Science (Wash.), 216, 696-703.

Received on 10 April 1985; revised on 29 May 1985