The EMBO Journal vol.6 no.7 pp.1975-1980, 1987

Identification of a testis-specific gene from the mouse t-complex next to a CpG-rich island

Gudrun A.Rappold[1,2], Lisa Stubbs[1,2], Siegfried Labeit[1], Radomir B.Crkvenjakov[1,3] and Hans Lehrach[1,2]

[1]European Molecular Biology Laboratory, Postfach 10.2209, 6900 Heidelberg, FRG, and [3]Genetic Engineering Center, PO Box 794, Beograd, Yugoslavia

[2]Present address: ICRF, PO Box 123, Lincoln's Inn Fields, London, WC2A 3PX, UK

Communicated by H.Lehrach

We have used an approach based on the observation of CpG-rich regions near the 5′ end of many genes to screen a panel of cosmids derived from the t-complex and tested candidate sequences for evidence of transcription in a number of different mouse tissues. One gene so identified is expressed specifically in testicular germ cells and maps to a subregion of the t-complex also containing loci involved in transmission ratio distortion and male sterility. The transcript is first detected during the pachytene stage of the first meiotic division, but is expressed in highest levels in the later haploid spermatogenic stages. Sequence analysis verified the existence of a CpG-rich element on the 5′ end of the gene and predicts a unique protein species with no significant homologies to those previously determined.

Key words: testis-specific gene/t-complex/CpG-rich region

Introduction

The proximal half of mouse chromosome 17, which contains the t-complex, is one of the genetically best analyzed regions of a mammalian genome. Its extensive genetic characterization has lead to the mapping of over 20 different genes within this 15 cM big region (t lethals, T, Fu, qk, tf, genes involved in transmission distortion and sterility) (Klein and Hammerberg, 1977; Silver, 1981, 1985; Washburn and Eicher, 1983). In addition, ENU mutagenesis has led to the identification of new t-complex mutations whose characterization is still in progress (Shedlovsky *et al.*, 1986). It is therefore likely that genes isolated from this segment of chromosome 17 by molecular means can be correlated with already exisiting t-complex mutations.

The t, or 'mutant' form of genes lying within this region are generally inherited as haplotypes, due to strong suppression of recombination between t and wild-type chromosomes in heterozygous mice. Males carrying two different or identical t-complex chromosomes are always sterile, and embryos homozygous for a given t-haplotype are often unable to complete normal development; but despite these obvious selective disadvantages, t haplotypes are carried in large numbers in wild mouse populations. This is due to a striking transmission ratio distortion in heterozygous males, which may pass t-bearing chromosomes to more than 90% of their offspring. Transmission ratio distortion has been shown recently to involve the cumulative action of three distorter/sterility genes (Tcd-1, Tcd-2, Tcd-3) acting on a responder gene (Tcr) (Lyon, 1986).

To identify genes within the t-complex, we have used an approach based on the observation of CpG-rich regions located 5′ to many mammalian genes (reviewed by Bird, 1986). CpG-rich regions were identified by cleavage with restriction endonucleases containing at least one, and preferentially two CpG doublets in their recognition sequences. Due to the low frequency of CpGs in mammalian DNA, sites for these enzymes occur rarely in the mammalin genome. Sites are however commonly found within the CpG-rich 'island' sequences (Brown and Bird, 1986).

In our analysis we identified potential CpG islands within cosmids isolated by homology to t-complex derived probes ('microclones') originally isolated by microdissection of the proximal half of chromosome 17 (Röhme *et al.*, 1984). Sequences neighboring rare cutting sites (*Sac*II, *Bss*HII, *Mlu*I, *Sma*I) were tested for evidence of transcription in a number of selected mouse tissues.

We present here the characterization of one t-complex gene isolated using this approach. DNA sequences near a rare-cutter cluster in cosmid 117/11 detect an ~0.8-kb highly abundant, testis-specific transcript. The transcript is present in RNA derived from highly enriched populations of spermatogenic cells, first appearing during the pachytene stage of meiosis, and is present in greatest abundance in later spermatogenic stages. The significance of its testis-specific expression is further enhanced by the mapping of cosmid 117/11 sequences to an ~1 cM region which has been shown to contain the gene for the distorter/sterility gene, Tcd-3 (Lyon, 1986). The sequence of the cDNA clone homologous to transcribed cosmid sequences predicts a unique protein species with little or no sequence homology to those which have been described thus far.

Results

Tissue-specific expression of 117c3

Cosmid 117/11 was chosen for study because of the presence of a cluster of rate-cutting enzyme sites. Another cosmid, 117/8 overlaps with 117/11 and contains neighboring sequences extending another 30 kb (Figure 1). One *Eco*RI fragment, present only in 117/11 and containing restriction sites of the endonucleases *Sac*II, *Bss*hII and *Sma*I, was tested for homology with RNA transcribed in several mouse tissues. These preliminary hybridizations with the genomic fragment detected an abundant RNA species, specifically in testicular tissue. Using the transcribed cosmid fragment, we isolated a cDNA clone from a mouse testis cDNA library (kindly provided by K.Willison). As can be seen in Figure 2a and b, a highly abundant transcript of ~0.8 kb was detected with the cDNA clone 117c3 in RNA derived from testicular tissue. A trace of mRNA was also found in brain after prolonged exposure, as has been described for several other testis-specific transcripts (Wolgemuth *et al.*, 1986), although the significance of this low level of expression is not understood.

The relationship between the cDNA and the genomic cosmid clone is shown in Figure 1. The length of the sequenced cDNA is 732 bases, and consists of at least three exons, which is based on a comparison between the hybridization patterns of labelled

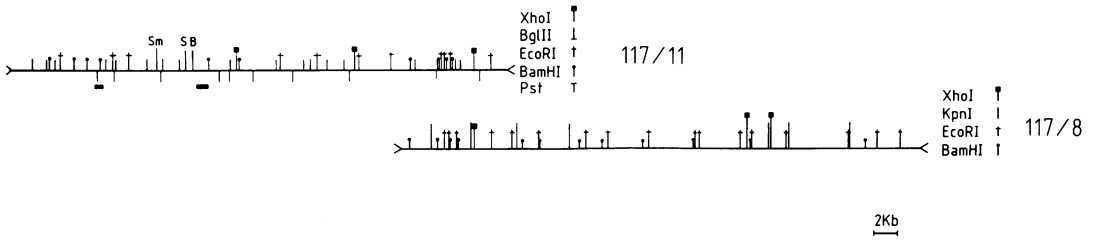

Fig. 1. Restriction maps of cosmids 117/11 and 117/8. Cosmids 117/11 and 117/8 were independently isolated by homology with microclone Tu117 from a 129 SV/Slcp library. Restriction maps were determined by the method of Rackwitz *et al.* (1985). The two cosmids overlap by 10 kb, and together cover ~72 kb of genomic DNA. The bars underlying the map of 117/11 represent regions of homology with the cDNA clone 117c3. Enzymes which cut rarely are *Sac*II (S), *Bss*hII (B) and *Sma* (Sm).

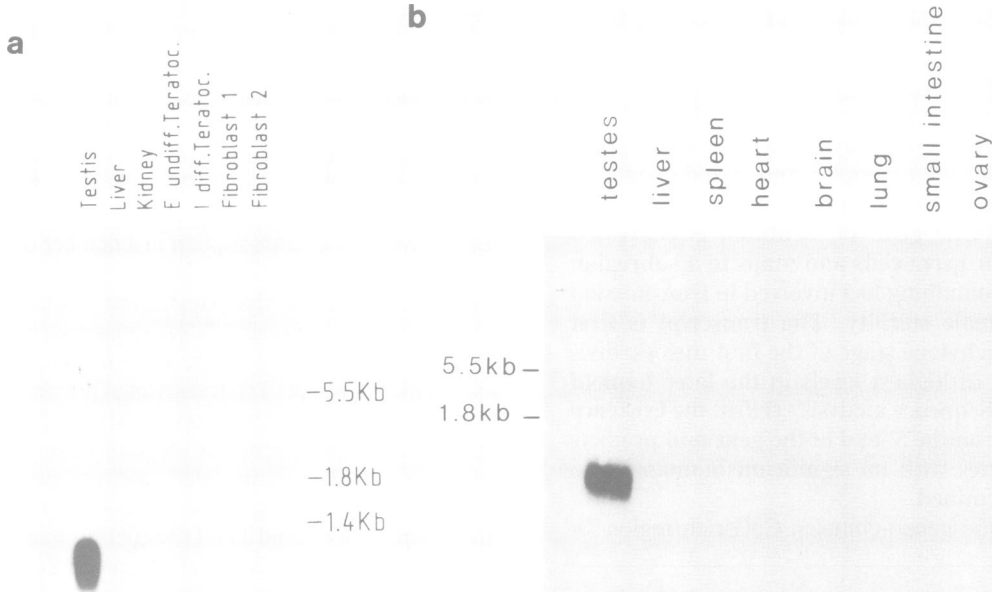

Fig. 2a,b. Tissue-specific expression of clone 117c3. Northern blots of RNA isolated from a variety of mouse tissues were hybridized with labelled 117c3 cDNA. 10 μg each of total RNA derived from tissues, or from cultured cells (fibroblasts from two different sources, F9 teratocarcinoma, differentiated and undifferentiated lines) were used.

cDNA to *Eco*RI-digested genomic DNA and the cosmids 117/11 and 117/8 (data not shown). The two widely separated fragments from cosmids 117/11 (represented by bars in Figure 1) were determined by hybridizing the cDNA to different single and double digests of cosmid DNA. The sizes of the *Eco*RI fragments from the cosmid which hybridized with the cDNA correspond exactly to the *Eco*RI bands hybridizing with genomic DNA, but a third *Eco*RI fragment detected in the genomic digest has no counterpart with either of the two cosmids. This suggests that a third hybridizing fragment is located downstream of the cloned segments in genomic DNA.

Sequence of the mouse cDNA clone 117c3

DNA sequences of cDNA 117c3 were obtained from both strands, resulting in 732 bases sequenced. Sequence analysis revealed a long open reading frame from position 94 to 603, which predicts a polypeptide product as shown in Figure 3. Alternatively, ATG codons occurring at bp 178 and 340 might be used *in vivo* for initiation of protein translation. The transcriptional orientation was concluded by the polyadenylation signal AAATAAA at bp 700, followed by a poly(A) tail at bp 715–732, and inferred by a CG cluster at the 5′ end. Figure 4 shows the CpG and (as a comparison) the GpC distribution of the gene. The CpG content of the first 200 base pairs at the 5′ end is ~10 times higher than the following bases and thus verifies the existence of an HTF-like structure around the 5′ end of the gene. We speculate that the HTF island overlaps with the first exon (the exact size of which we have not determined, but may be very small) and may extend even further, since sites for the two restriction enzymes *Sac*II and *Bss*hII apparently lie within the first intron, at the 5′ site.

The cDNA clone 117c3 predicts a primary translation product of 170 amino acids with an estimated molecular mass of 21 000 daltons (21 kd). Comparing the DNA sequence and the predicted

```
        10                  30                  50
TTCCCAGCTTCTGAACACTTCTGAGACCAAGTACATGGAGAAGCAGAGGAAGCGGGAGGC
---------+---------+---------+---------+---------+---------+

        70                  90                 110
AGAGGGCGGAGGCGTCCCGGTGAACCGATCAAGATGGAGCGGCGAGGCCAATTGCGGAAG
---------+---------+---------+---------+---------+---------+
                                  MetGluArgArgGlyGlnLeuArgLys

       130                 150                 170
ACGCCCACCGGCCAAACGCATCAATCCCCGGTGTCTAAGAGAGAAAGGAAGCCTAGCATG
---------+---------+---------+---------+---------+---------+
ThrProThrGlyGlnThrHisGlnSerProValSerLysArgGluArgLysProSerMet

       190                 210                 230
TTCGAGAAGGAGTCATATGCACAGATCTTAAGAGAAAGACTGAGAGAGTCTTTTCATGAT
---------+---------+---------+---------+---------+---------+
PheGluLysGluSerTyrAlaGlnIleLeuArgGluArgLeuArgGluSerPheHisAsp

       250                 270                 290
GTTCAGTACGTGGAACCTCCGTTTGATGACTCAATTGCTGATGTAGGCAAAGAATGGAAA
---------+---------+---------+---------+---------+---------+
ValGlnTyrValGluProProPheAspAspSerIleAlaAspValGlyLysGluTrpLys

       310                 330                 350
AGTGCCCTGGCAAAATTAAAGTTTGCTAATTCATACAGAATGGAGCCACTGAAGAAATTT
---------+---------+---------+---------+---------+---------+
SerAlaLeuAlaLysLeuLysPheAlaAsnSerTyrArgMetGluProLeuLysLysPhe

       370                 390                 410
CAAGCACATTTGGTAGAAACTAAAGTCCAGCAGATATTAAAGGACAGTCTTAAAGATGTC
---------+---------+---------+---------+---------+---------+
GlnAlaHisLeuValGluThrLysValGlnGlnIleLeuLysAspSerLeuLysAspVal

       430                 450                 470
AAATATGATGACAAAGCCCCTCATTTGTCACTTGAATTGGCAGATCAATATTGGCAGCAG
---------+---------+---------+---------+---------+---------+
LysTyrAspAspLysAlaProHisLeuSerLeuGluLeuAlaAspGlnTyrTrpGlnGln

       490                 510                 530
TCAAAAAATTTCATACCATGGATATATTATTATACAAGTATTATTTATTCAAAwGACTGG
---------+---------+---------+---------+---------+---------+
SerLysAsnPheIleProTrpIleTyrTyrTyrThrSerIleIleTyrSerLysAspTrp

       550                 570                 590
TCAAGCAATAAATATTGCCAGCAGATGGATCTGGGATGTGGCATGGGACAACTGGGTAGA
---------+---------+---------+---------+---------+---------+
SerSerAsnLysTyrCysGlnGlnMetAspLeuGlyCysGlyMetGlyGlnLeuGlyArg

       610                 630                 650
AGCTAAACATGAAACAGAGTCTTACGTGGTATTGGCCTTGGTGTTTGCTCTCTATTGTga
---------+---------+---------+---------+---------+---------+
SerEnd

       670                 690                 710
atagcttaggaccagcattttcacccccccatccttcaaaataaatgatatatacaaaaaa
---------+---------+---------+---------+---------+---------+

       730
aaaaaaaaaaa
---------+--
```

Fig. 3. Nucleotide sequence and the deduced amino acid sequence of 117c3. Numbers above the sequence refer to the nucleotide position. The sequence of the predicted primary translation product is also shown, assuming initiation at the first occurring ATG at bp 94.

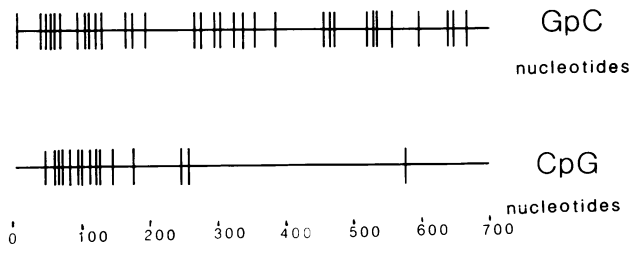

Fig. 4. CpG and GpC distribution in the 117c3 gene. Each vertical line represents a CpG doublet (lower map) or a GpC doublet (upper map). Note that there is a CpG cluster at the 5′ end of the gene.

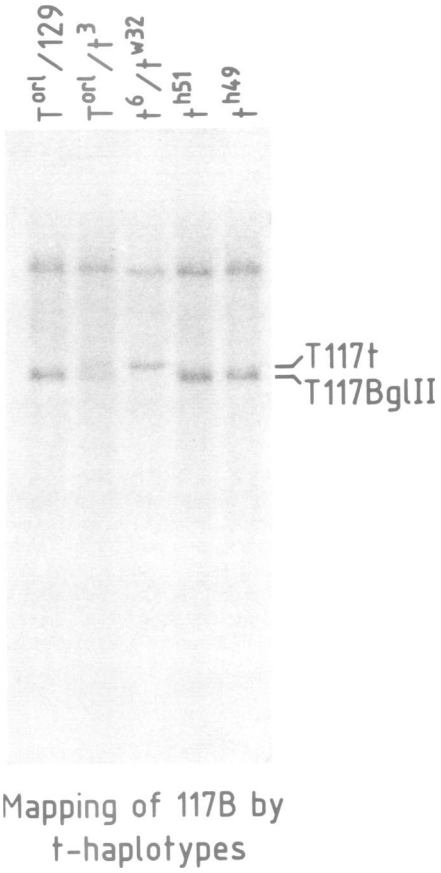

Fig. 5. Southern blot hybridization of t-complex probe 117B to mouse DNA carrying different partial t-haplotypes. Hybridization of 117B to *Bgl*II digested DNA of different t-hapolotypes. 117B is a 3.6-kb RI fragment from the overlapping cosmid 117/8. It detects an RFLP between a partial and a complete t-haplotype form of chromosome 17. The upper 9.4-kb band T117(t) is t-specific, the lower 9.0-kb band T117 (*Bgl*II) occurs in wild-type DNA. The band common to all tracks is due to an internal *Bgl*II site.

protein sequence with those already entered in the EMBL and Genebank data bases reveals no significant homologies with proteins thus far characterized.

Genetic mapping of 117c3

The cosmid 117/11 was first shown to be located within the region of chromosome 17 corresponding to the t-complex using an interspecies backcross system (Robert *et al.*, 1985). A genomic fragment derived from cosmid 117/11 detects a restriction fragment length polymorphism between *Mus musculus* and *M. spretus* DNA with the enzyme *Bam*HI. Comparing the distribution of *musculus* or *spretus* alleles for this fragment in a panel of backcross progeny with known t-complex markers, we were able to position 117c3 to a region of chromosome 17 proximal to the α-globin Hba-ps4 pseudogene (data not shown), which is close to the marker *tufted*.

In order to map 117c3 sequences more precisely, and possibly to relate them to male-specific fertility and transmission ratio distortion genes already genetically mapped on chromosome 17, we took advantage of a restriction fragment length polymorphism detected by a fragment from cosmid 117/8 between DNA from t and wild-type mice, to map the sequence using DNA from a panel of mice representing partial t-haplotypes (Fox *et al.*, 1985). Partial t-haplotypes arise by rare recombination events between the t and wild-type chromosomes in heterozygous mice. This results in hybrid forms of chromosome 17 that carry t-chromosome along varying fractions of the entire length of the complex, and wild-type alleles beyond the point of recombination. We selected a panel of partial t-haplotypes which overlap in such a way that a polymorphic probe can be quite precisely mapped to a subregion of the t-complex (Figure 5). A 3.6-kb *Eco*RI fragment from cosmid 117/8 detects an RFLP between 129 and tDNA restricted with the enzyme *Bgl*II. The 9.4-kb band T117(t) is t-specific and the length of the wild-type-specific fragment is 9.0 kb. The presence of a 9.4-kb long t-specific band in t^3 and t^6 DNA and the absence of this band in t^{h49} and t^{h51} places this probe in the middle region of the t-complex next to cloned markers 122 and 66C (Fox *et al.*, 1985) (Figures 5 and 6). This localization is of special interst, since the distorter gene,

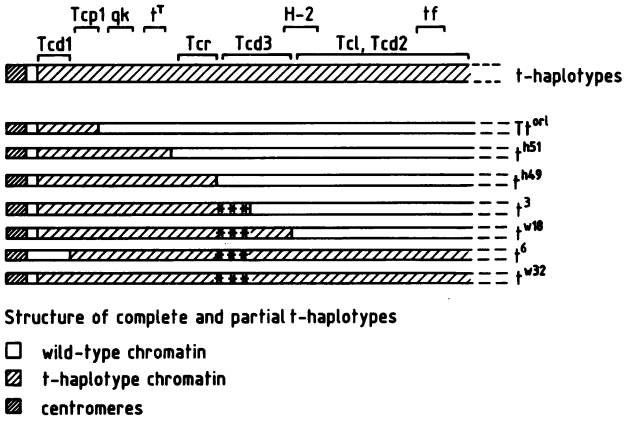

Fig. 6. Schematic structure of partial t-haplotypes used in genetic localization. Wild-type sequences are drawn white and t-haplotype chromatin hatched. The centromere is shown to the left. The genetic factors associated with t-haploytpes are named according to a proposal by Lyon (1986) and called t-complex sterility distorter genes 1, 2 and 3 (Tcd1, Tcd2, Tcd3), t-complex responder (to distortion genes) (Tcr) and t-complex lethality (Tcl). The genetic markers above stand for Testes protein 1 (Tcp 1), quaked (qk), Brachyury (t^T), tufted (tf) and the histocompatibility region (H2).

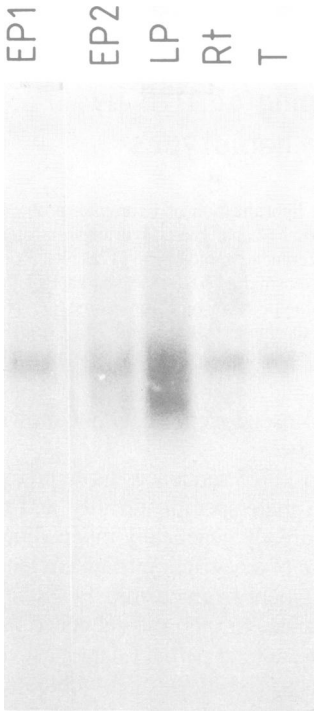

Fig. 7. Different populations of spermatocyte development. 117c3 was hybridized to 5 μg total RNA isolated from each of four cell fractions enriched for spermatocytes of specific meiotic stages. EP1 and EP2 are early prophase; LP, late prophase; Rt, round spermatid fraction; and T, random testes tissue.

Tcd-3, has recently been mapped to this segment of t-complex DNA (Lyon, 1986).

Localization and timing of expression of 117c3 in mouse testis

Relevant to the possible role of 117c3 sequences in transmission distortion, or male fertility in general, is the question of specifically where and when the gene is expressed within testis tissue. Although 90% of the cells in normal adult testis are germ cells in varying stages of spermatogenic development, transcription

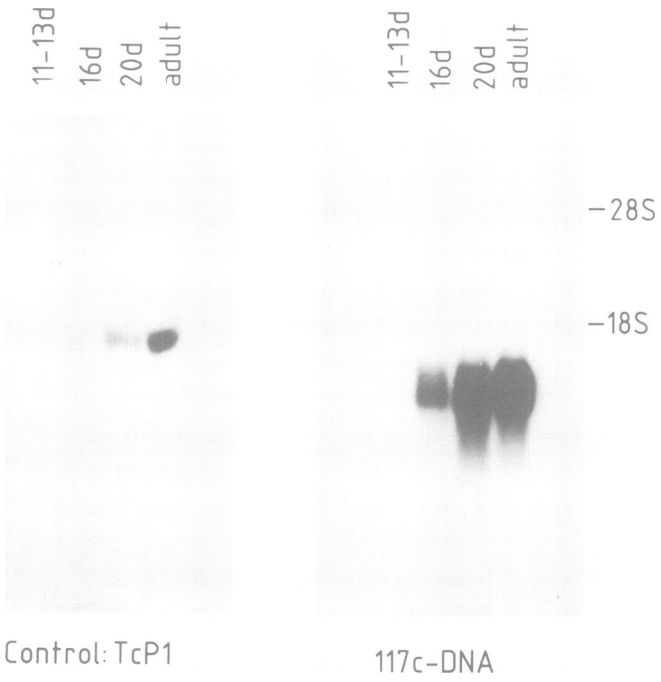

Fig. 8. Expression of 117c3 DNA in mouse testes of different developmental stages. 2 μg of total RNA from testis of 11–13, 16, 20-day-old and adult mice were transferred to Genescreen membrane and hybridized with the 117c3 cDNA sequence as probe. As a control the same filter was hybridized with another testis-specific cDNA clone, Tcp1.

may still be due to the presence of any of a variety of somatic cells comprising the testis structure. In order to approach this question, we probed RNA derived from highly enriched spermatocyte preparations, pooled in groups according to their stage in spermatogenic development (Materials and methods). Though none of these fractions are pure, either with regard to spermatogenic stage or complete absence of somatic cells, the level of somatic cell contamination is quite low (always $<1\%$). In addition the fractionation scheme is based on striking size changes of cells as they advance through spermatogenesis (Bellvé *et al.*, 1979). Therefore any specific type of somatic cell will be restricted to one or few fractions.

As can be seen in Figure 7, all cell fractions contain the RNA species detected with the 117c3 cDNA, and in approximately equal quantities. This fact makes it highly likely that it is testicular germ cells, and not somatic cells, that express the gene; no apparent stage specificity can be detected from the analysis of these fractionated cells. However, the earliest stage from which cells can be obtained relatively free of spermatocytes in later stages is late pachytene cells (LP fraction), and it is impossible to judge from these results whether germ cells of all stages express the gene, or if transcription begins later in meiotic prophase.

In order to address this question, we isolated RNA from testes of mice before and during the onset of puberty, throughout a period when the first wave of male germ cells enters and completes meiosis (Bellvé *et al.*, 1979). Since the timing of spermatogenesis is quite constant between individuals in a species, it is possible, by examining testicular RNA from mice of a particular age, to determine the earliest stage at which transcription occurs. Figure 8 shows the results of such an analysis with the 117c3 probe, and as a control, the Tcp-1 cDNA (Willison *et al.*, 1986) which is known to be expressed primarily during the later haploid stages. Significant transcription of 117c3 sequences is

first detected at 16 days after birth, when the first wave of spermatocytes has advanced to the pachytene stage; no later cells are present in the testes of these pre-pubertal mice. The transcript is present in even larger quantities as development ensues, with levels of RNA in 20-day-old mice nearly equal to those observed in fully mature adults (Figure 8).

Discussion

In order to identify potential t-complex genes and possibly correlate them with already mapped and well studied t-complex mutations, we took advantage of the observation that many mammalian genes contain CpG-rich sequences near their 5′ ends. These DNA segments can often be recognized by their susceptibility to cutting by restriction enzymes, such as *Sac*II and *Bss*HII, which cut rarely within mammalian genomes due to the presence of one or more CpG doublets in their recognition sites (Brown and Bird, 1986).

By screening cosmids derived from microdissected t-complex probes for the presence of sites for such rare-cutting enzymes, we identified a sequence transcribed in high abundance specifically within testicular tissue. The isolation of a testis-specific gene from t-complex sequences is of special interest, because many of the unique features associated with the biology of t-haplotype mice are related to genes involved in male-specific transmission ratio distortion and sterility (Lyon *et al.*, 1986). The sequence of a cDNA clone, isolated from a mouse testis library (Dudley *et al.*, 1984) using the transcribed cosmid sequence as a probe, predicts a unique protein species of 21 kd; neither the cDNA nor the predicted protein sequence show significant homologies to others thus far present in available sequence banks. The sequence does however verify the existence of a CpG-rich region near to the 5′ end of the gene. As expected for a sequence partly overlapping an 'HTF island' (Bird, 1986), the frequency of CpG sequences in the first 200 bp is equal to or higher than the GpC frequency (Figure 4). In contrast to this, the 3′ end of the cDNA shows the expected deficiency of CpG sequences. In spite of its tissue-specific expression, this gene might therefore belong to the class of genes overlapping with HTF islands, usually expected to be associated with housekeeping functions. Other exceptions to this generalization have, however, been described (Bird, 1986).

Genetic mapping of the identified gene, an essential step in the possible correlation of the transcribed region with geneticaly localized mutations was complicated by the low frequency of polymorphisms found within the 70-kb cloned genomic region including the transcribed sequence. While no polymorphic DNA fragments allowing the mapping by recombinant inbred strains could be identified, mapping was possible using a *M. musculus*−*M. spretus* backcross system (Robert *et al.*, 1985), localizing the gene to the more proximal part of the t-complex; Since genes involved in the transmission distortion/sterility phenomena associated with the mutant from of the t-complex can only be localized using partial t-haplotypes (Fox *et al.*, 1985; Lyon, 1986), localization of the transcribed clone relative to the available partial t-haplotypes was likely to be most informative.

Such an analysis, possible because of a polymorphism detected between t and wild-type DNA by one fragment within the overlapping cosmid 117/8, placed the gene between the recombination breakpoints of t^{h49} and t^3, a segment of ~1 cM located within the central portion of the t-complex, between the two large inversions characteristic for mutant t-haplotypes (Herrmann *et al.*, 1986). The mapping of a testis-specific gene to this particular region is especially interesting because of the recent identification of the distorter locus tcd3 within the same genetic segment (Lyon, 1986). A male-specific sterility factor, tcs3, has also been mapped within this interval, but Lyon has suggested that tcd3 and tcs3 might represent separate manifestations of the activities of a single gene. The location of the 117c3 transcription unit raises the intriguing possibility that it may be related to either or both of these genetically identified functions.

We have shown that 117c3 sequences are transcribed specifically within germ cells, and not the somatic cells of the testis. The transcript is first detected as the first wave of spermatocytes enters the pachytene stage but increases in abundance as the cells advance to later spermatogeneic stages, with the level of transcription nearly as high in 20-day-old animals as it is in fully mature adults. This suggests that although expression of 117c3 sequences may begin before the end of the first meiotic division, the transcript continues to be made in haploid cells. This suggestion is strengthened by the observation that the RNA species is detected in a cell fraction containing only premeiotic spermatocytes and post-meiotic round spermatids; since the latter cell type is present in the testis 11 days after birth (Bellvé *et al.*, 1979), and since no transcript is detected in animals of that age, the transcript observed in RNA from the mixed cell fraction must be contributed by the haploid spermatids.

The mechanisms of transmission ratio distortion and t-complex male sterility remain completely unknown, but most proposed mechanisms of transmission ratio distortion in mouse involve events occurring after the first meiotic division (Lyon, 1986) and so are compatible with the notion of gene expression in the later, haploid stages. Further elucidation of the possible relationship between 117c3 sequences and aspects of male fertility affected by t-complex genes will require a close comparison of the structure and activities of the transcript in t and wild-type males. Preliminary studies have detected no obvious differences in the size or abundance of the RNA species in testes from t-heterozygotes with very high levels of distortion and wild-type males (data not shown), but differences between t and wild-type genes may well lie in subtle changes of DNA and protein sequence. Alternatively, 117c3 may represent another, previously uncharacterized, testis-specific function lying near the distorter/sterility gene(s). Experiments are currently underway to investigate further the role of 117c3 sequences in spermatogenic development, and their possible relationship with t-complex genes involved in male-specific sterility and transmission ratio distortion.

Materials and methods

Cosmid clone

Cosmids 117/11 and 117/8 were isolated from a library of the 129 SV/Slcp genome using a probe resulting from the microdissection of the proximal region of mouse chromosome 17 (Röhme *et al.*, 1985).

To prepare the DNA of the probe fragments, a gel slice from a low melting point agarose gel (BRL) was cut out, 100 mM NaCl EDTA were added and the sample melted by incubating for 2 min at 75°C, then transferred to 37°C. Agarase (5−10 units; Calbiochem) was added and the sample was incubated at 37°C for 2 h. After phenol and ether extraction the DNA was ethanol precipitated and dissolved in TE. The DNA was labelled by oligo-priming (Feinberg and Vogelstein, 1984) by using 10−40 ng of template. Specific activity of the probes was 10^8-10^9 c.p.m./μg. Restriction mapping of cosmids was carried out according to Rackwitz *et al.* (1985), using a computer program package for restriction map analysis and manipulation described by Zehetner and Lehrach (1986).

Genomic DNA isolation and analysis

High mol. wt DNA for genomic Southern blot analysis from adult organs of mice was isolated by standard techniques. DNA was digested, electrophoresed and blotted onto nylon membrane (Gene Screen, NEN) and bound by u.v.-crosslinking. This latter procedure enabled the re-use of the filters up to 10 times and more

without significant loss of signal. Hybridization conditions were essentially as described by Church and Gilbert (1984).

Northern analysis

Total RNA was isolated following the procedure of Cathola *et al.* (1983). Tissues were homogenized in 5 M guanidine monothiocyanate followed by direct precipitation of RNA from the guanidinium by 4 M LiCl. RNA was electrophoresed through formaldehyde agarose gels (Lehrach *et al.*, 1977; Maniatis *et al.*, 1982), transferred in 20 × SSC to a nylon-membrane (Gene Screen, NEN). Filters were baked at 80°C under vacuum, and the RNA cross-linked to the membrane (Church and Gilbert, 1984). Hybridization was carried out essentially as described for genomic Southerns.

Preparation of spermatocyte cell fractions

Cell fractions enriched for spermatocytes in specific meiotic stages were prepared by unit gravity sedimentation of cells through BSA gradients, as described by Chandley *et al.* (1977). Fractionated cells were pooled after microscopic examination, into four major groups. The LP (late prophase) fraction consisted of 90–95% pachytene spermatocytes, with symplasts of post-meiotic round spermatids as the major contaminant; as we prepare this fraction, somatic cell contamination is negligible (<0.5%). Fraction EP (early prophase) 2 contains pachytene and zygotene spermatocytes, spermatid symplasts and a small number of spermatogonia; fraction EP 1 contains spermatocytes in earlier stages of meiotic prophase, symplasts and a large number of premeiotic spermatogonia. Somatic cell contamination is also highest in this fraction, although still being <1%. The Rtid fraction is comprised mainly of postmeiotic round spermatids (80%), the remaining 20% of the population being composed of spermatocytes which have not yet entered meiotic prophase.

Screening of cDNA libraries

We used a mouse testis cDNA library (kindly provided by Keith Willison; Dudley *et al.*, 1984). The cDNA library was plated at high density on MSI membrane (Micron Sep. Inc.) and the master filter stored at −70°C. Two replica filters were used for hybridizations. The positive colony area was cut out from the master filter and used for rescreening.

The plasmid DNA was fixed to the filter by denaturation in 0.5 M NaOH/1.5 M NaCl for 10 min. The filter was neutralized in 50 mM NaPi pH 7.2 for 2 min, and the cell debris removed by gentle rubbing. The filter was washed by soaking in 50 mM NaPi pH 7.2, air dried, baked for 20 min and the DNA cross-linked with a filter by u.v. treatment (see also Herrmann *et al.*, 1987). Hybridization conditions were according to Church and Gilbert (1984).

DNA sequence analysis

To facilitate sequence analysis, 117c3 was subcloned as a blunt end fragment into the *Sma*I site of mp18 (Norrander *et al.*, 1983) and a pair of clones with opposite orientation were selected by the c-test as described in Messing (1983). Ordered deletion libraries were prepared for each orientation and six clones for each orientation with different deletions were selected for sequencing, which covered the whole insert on both strands. Sequencing reactions were performed by the thiophosphate methods as described by Labeit *et al.* (1986,1987).

Acknowledgements

We would like to thank our colleagues Denise Barlow, Maja Bucan, Günter Zehetner, Bernhard Herrmann, Richard Meeham and Chris Sander for material and many helpful discussions, Keith Willison for providing the mouse testis cDNA library, Wendy Moses for preparing the manuscript and Adrian Bird and Howard Cooke for introducing us to the concept of HTF islands. G.R. was supported by a grant from the Deutsche Forschungsgemeinschaft (LE 496/6-1) to H.L., L.S. by a NIH fellowship and R.C. by a Humboldt fellowship.

References

Bellvé,A.R. (1979) In Finn,C.A. (ed.), *Oxford Reviews of Reproductive Biology*. Oxford University Press, London, Vol. 1, pp. 159–261.
Bird,A.P. (1986) *Nature*, **321**, 209–213.
Brown,W.R.A. and Bird,A.P. (1986) *Nature*, **322**, 477–481.
Cathola,G., Savouret,J.F., Mendez,B., West,B.L., Karin,M., Martial,J.A. and Baxter,J.D. (1983) *DNA*, **4**, 329–335.
Chandley,A.C., Hotta,Y. and Stern,H. (1977) *Chromosoma*, **62**, 243–253.
Church,G.H. and Gilbert,W. (1984) *Proc. Natl. Acad. Sci. USA*, **81**, 1991–1995.
Dudley,K., Potter,J., Lyon,M.F., Willison,K.R. (1984) *Nucleic Acids Res.*, **12**, 4281–4293.
Feinberg,A.P. and Vogelstein,B. (1984) *Anal. Biochem.*, **137**, 266–267.
Frischauf,A.M. (1985) *Trends Genet.*, **1**, 100–103.
Fox,H.S., Martin,G.R., Lyon,M.F., Herrmann,B., Frischauf,A.M., Lehrach,H. and Silver,L.M. (1985) *Cell*, **40**, 63–69.
Herrmann,B., Bucan,M., Mains,P.E., Frischauf,A.M., Silver,L.M. and Lehrach,H. (1986) *Cell*, **44**, 469–476.
Herrmann,B., Barlow,D. and Lehrach, H. (1987) *Cell*, **48**, 813–825.
Ishii,S., Merlino,G.T. and Pastan,J. (1985b) *Science*, **230**, 1378–1380.
Kettenah,N.P. and Hartl,D.L. (1976) *Science*, **193**, 1020–1021.
Klein,J. and Hammerberg,C. (1977) *Immunol. Rev.*, **33**, 71–104.
Labeit,S., Lehrach,H. and Goody,R.S. (1986) *DNA*, **5**, 173–177.
Labeit,S., Lehrach,H. and Goody,R.S. (1987) *Methods Enzymol.*, in press.
Lehrach,H., Diamond,D., Wozney,J.M. and Boedtker,H. (1977) *Biochemistry*, **16**, 4743–4747.
Lyon,M.F., Evans,E.P., Jarvis,S.E. and Sayers,J. (1979) *Nature*, **279**, 38–42.
Lyon,M.F. (1986) *Cell*, **44**, 357–363.
Maniatis,T., Fritsch,E.F. and Sambrook,J. (1982) *Molecular Cloning. A Laboratory Mannual*. Cold Spring Harbor Laboratory, New York.
Messing,J. (1983) *Methods Enzymol.*, **101**, 20–78.
Norrander,J., Kempe,T. and Messing,J. (1983) *Gene*, **26**, 101–106.
Rackwitz,H.R., Zehetner,G., Murialdo,H., Delius,H., Chai,J.H., Poustka,A., Frischauf,A. and Lehrach,H. (1985) *Gene*, **40**, 259–266.
Robert,B., Barton,P., Minty,A., Daubas,P., Weydert,A., Bonhomme,F., Catalan,J., Chazottes,D., Guenet,J.-L. and Buckingham,M. (1985) *Nature*, **314**, 181–183.
Röhme,D., Fox,H., Herrmann,B., Frischauf,A.-M., Edström,J.E., Mains,P., Silver,L.M. and Lehrach,H. (1984) *Cell*, **36**, 783–788.
Shedlovsky,A. *et al.* (1986) *Genet. Res. Camb.*, **47**, 135–142.
Silver,L.M. (1981) *Cell*, **27**, 239–240.
Silver,L.M. (1985) *Annu. Rev. Genet.*, **19**, 179–208.
Washburn,L.L. and Eicher,E.M. (1983) *Nature*, **303**, 338–340.
Willison,K.R., Dudley,K., Potter,J. (1986) *Cell*, **44**, 727–738.
Wolgemuth,D.J., Engelmyer,E., Duggal,R.N., Gizang-Ginsberg,E., Mutter,G.L., Penzetto,C., Vicinao,C. and Zakeri,Z.F. (1986) *EMBO J.*, **5**, 1229–1235.
Zehetner,G. and Lehrach,H. (1986) *Nucleic Acids Res.*, **14**, 335–349.

Received on February 20, 1987; revised on April 8, 1987