Development 111, 561-571 (1991)
Printed in Great Britain © The Company of Biologists Limited 1991



# Isolation and characterisation of a testis-expressed developmentally regulated gene from the distal inversion of the mouse *t*-complex

NICHOLAS D. MAZARAKIS[1,*], DANIEL NELKI[1], MARY F. LYON[2], S. RUDDY[2], E. P. EVANS[3], PAUL FREEMONT[4] and KEITH DUDLEY[1,†]

[1]*Biomolecular Sciences Division, King's College London, Campden Hill, Kensington W8 7AH, UK*
[2]*MRC Radiobiology Unit, Chilton, Didcot, Oxon, UK*
[3]*William Dunn School of Pathology, South Parks Road, Oxford, UK*
[4]*ICRF, Lincoln's Inn Fields, London, UK*

* Present address: Laboratory of Neurobiology, NIMR, The Ridgeway, Mill Hill, London NW7 1AA, UK
† To whom correspondence should be addressed

## Summary

**We differentially screened a pool of mouse testis clones in order to identify genes important in germ cell development. One of the isolated clones was found to be expressed only in the male germ line where it is first detected at around the pachytene spermatocyte stage. This gene maps to a subregion of the *t*-complex in the distal inversion near, but not within, the $t^{w18}$ and the $t^{h20}$ deletions. A comparison of the *t* and wild forms of the gene reveals a high degree of sequence conservation. This gene is associated with a CpG-rich island at its 5′ end. It encodes a novel protein with extensive α-helical structure indicative of coiled-coil interactions.**

Key words: testis-specific gene, *t*-complex, distal inversion, CpG island, mouse.

## Introduction

Spermatogenesis, the differentiative pathway by which diploid germ cells develop into haploid spermatozoa has been extensively studied in the mouse (for reviews see Hecht, 1988; Willison and Ashworth, 1987). It can be subdivided into three phases: a premeiotic phase starting shortly after birth when mitotic divisions increase the numbers of diploid spermatocytes, a meiotic phase when two successive divisions give rise to round spermatids and a postmeiotic phase (spermiogenesis) when extensive morphogenetic changes culminate in the production of mature sperm (Bellvé, 1979).

The *t*-complex is a naturally occurring variant, which occupies approximately 15 cMorgan of the proximal region of mouse chromosome 17, from a point near the centromere to a point between the major histocompatibility complex (MHC) and the *Pgk-2* locus, and which represents about 1 % of the genome. Alternative forms of the complex are termed *t*-haplotypes. The *t*-complex results in a range of phenotypic abnormalities affecting embryonic development and male fertility (Silver, 1985). Males homozygous for semilethal or heterozygous for two complementing lethal haplotypes are sterile, the former having very few sperm, and the latter having normal numbers of sperm that are unable to fertilise eggs (Bennett, 1975; Dooher and Bennett, 1977). Males heterozygous for a single haplotype (+/*t*) show transmission ratio distortion in that they can transmit the *t*-carrying chromosome to up to 95% or more of their offspring. One possible explanation for such non-Mendelian transmission is that the haploid products of meiosis are not functionally equivalent, a consequence of gene expression during or after meiosis. A number of studies indicate that transmission ratio distortion is caused by the functional inactivation of wild-type sperm by their *t* counterparts in heterozygous animals (Seitz and Bennett, 1985; Olds-Clarke and Peitz, 1986), an effect that could be achieved by the selective passage of message or protein between germ cells.

At the DNA level, it is now known that the region of mouse chromosome 17 where the *t*-complex maps has undergone two major non-overlapping inversions, a proximal and a distal (Hermann *et al.* 1986; Artzt *et al.* 1982) and also two smaller inversions, a centromeric and a middle (Hammer *et al.* 1989). These rearrangements are responsible for the low rate of recombination between wild-type and *t*-chromosomes in the region of the *t*-complex. Rare recombination events generate partial *t*-haplotypes comprising proximal, central or distal portions of the *t*-complex (Fox *et al.* 1985), which are useful tools for the genetic analysis of this region (Lyon and Meredith, 1964).

Lyon (1984) proposed a model explaining transmission ratio distortion in terms of a single *t*-responder locus (*Tcr*) and three *t*-distorter loci (*Tcd 1,2,3*), a model analogous to that of the segregation distortion (SD) system in *Drosophila* (Hartl and Hairaizumi, 1976). In this model, the products of the distorter loci interact with the responder locus in *cis* or *trans* and in an additive way leading to the 'poisoning' of the wild *Tcr* allele thus raising the transmission of sperm carrying the *Tcr* on the *t*-chromosome in $+/t$ heterozygotes. Making use of partial haplotype analysis, the responder and distorter elements have been mapped within the *t*-complex (Fox *et al.* 1985) while a fourth distorter locus (Tcd-4) has also been identified (Silver and Remis, 1987). Expanding on this model, Lyon (1986) proposed that *Tcd* loci are equivalent to sterility loci *Tcs* responsible for sterility in $t/t$ homozygous males while recessive *t*-lethal genes confer a selective advantage by eliminating such sterile males from the population. Candidate genes for distorter/sterility loci (*Tcp-1* Willison *et al.* 1986; *117c3* Rappold *et al.* 1987; *tctex-1* Lader *et al.* 1989) and for the responder locus (*T66* Schimenti *et al.* 1988) have now been cloned.

In this paper, we describe a differential screening experiment that led to the isolation of a testis-specific cDNA clone that is expressed in a developmentally regulated manner in the mouse germline. This gene was mapped to the *t*-complex and has several features that suggest that it might play an important role in sperm development and function.

## Materials and methods

### *Mice*

CBA/Ca, C57BL/6 and TO mice were purchased from Harlan Olac Limited. All *t*-haplotype mice were bred by Dr M. Lyon at Harwell. BXD recombinant inbred mice were purchased from the Jackson Laboratories.

### *Cell lines*

Cell line R44–1 was obtained from Dr L. Stubbs (ICRF) with the permission of Dr F. Ruddle. The cell line Southern presented in Fig. 4 was also a gift from Dr L. Stubbs.

### *Differential screen*

Colonies selected from a CBA/Ca testis cDNA library (Dudley *et al.* 1984) and stored in microtiter plates at −120°C were transferred in duplicate onto Whatman 541 paper by means of a manual replicator (Taub and Thompson, 1982). The filters were probed with randomly primed cDNA probes made from 18S sucrose gradient fractionated testis poly(A)$^+$ RNA derived from 2- and 3-week CBA/Ca mice. Poly(A)$^+$ RNA was prepared using oligo (dT)-trisacryl cellulose affinity chromatography columns (LKB) by the procedure of Chirgwin *et al.* (1979). Probe construction was according to Dudley *et al.* (1984). M-MLV reverse transcriptase was obtained from BRL.

Selected clones were colony-purified, and plasmid DNA from each was digested with *Eco*RI and *Hin*dIII, electrophoresed on agarose gels blotted onto Hybond nylon filters and rehybridized with the 3 week probe. All clones selected for further analysis had positive hybridizing inserts.

### *Northern blotting*

RNA was prepared as described by Dudley *et al.* (1984) and analysed on denaturing formaldehyde gels as described by Maniatis *et al.* (1982). RNA was blotted onto nitrocellulose and hybridized according to Dudley *et al.* (1984).

### *Southern blotting*

High molecular weight genomic DNA was prepared from mouse tissues as reported by Silver *et al.* (1983). 10 μg of DNA was digested to completion with specific restriction enzymes (NBL, BRL), electrophoresed on agarose gels and blotted on Hybond nylon filters (Amersham) according to Southern (1975). Filters were baked for 1 h and u.v. crosslinked according to the manufacturers' specifications. Hybridization conditions were 50 % formamide, 5×SSC, 5×Denhardts, 1 % SDS, 200 μg ml$^{-1}$ denatured, sonicated salmon sperm DNA at 42 °C for 18–24 h.

### *Probe construction*

Probes for Southern and northern hybridization were prepared using the random oligonucleotide priming method (Feinberg and Vogelstein, 1984). cDNA inserts were purified using the Geneclean Kit (Stratech).

### *Library screening*

The *t* haplotype mouse testis cDNA library (prepared from $t^6/t^{w1}$ haplotype mice and kindly provided by Keith Willison) was screened at high density using duplicate Hybond N filters (Amersham) which were hybridized at high stringency according to the manufacturers' specifications. Single colonies were obtained by two further rounds of screening.

### *Sequencing*

All sequencing, except where stated, was done by the dideoxy chain termination method (Sanger *et al.* 1977) using the Sequenase enzyme and Kit (United States Biochemicals).

Clone 46 was derived from a cDNA library constructed from testes RNA of CBA/Ca mice using *Eco*RI and *Hin*dIII linkers in the pUC9 vector (Dudley *et al.* 1984). Its entire insert was subcloned into M13 mp18 and 19 vectors and sequenced from each direction using the M13 universal primer. Using a single central *Pst*I site, *Eco*RI–*Pst*I and *Hin*dIII–*Pst*I fragments were isolated and subcloned in the same vectors and sequenced using the universal M13 primer. The sequencing of this clone was completed from both strands using synthetic oligonucleotide primers.

The insert of clone pBs13 was subcloned into the pBS SK vector and sequenced using T3, T7 and synthetic primers. Due to the presence of two severe compressions at the 5′ of the sequence, which proved refractile to dideoxy sequencing using dITPs and deaza-analogues or the thermostable *Taq* polymerase enzyme, a 5′ *Eco*RI–*Sau*3A fragment was end labelled and sequenced with the chemical method of sequencing (Maxam and Gilbert, 1977).

### *Hybrid selection*

Hybrid selection was performed as described by Willison *et al.* (1986) except that the DNA was denatured in 0.4 N NaOH prior to application to the filters.

### In situ *hybridization*

*In situ* hybridization to metaphase spreads was carried out as described by Lyon *et al.* (1986, 1988).

### *Computing*

The PBs13 amino acid sequence was used to search the

protein data banks PIR 20 (National Biomedial Research Foundation) and OWL8.1 (Akrigg *et al.* 1988) by the method of Collins *et al.* (1988) using the Smith and Waterman (1981) algorithm with a PAM 250 and 100 matrix (Dayhoff, 1978) and a gap penalty of 8. Further DNA sequence analysis was performed using the GENBANK data base of the Microgenie DNA analysis programmes (1988 Beckman, updates). Further analysis was carried out on the VAX cluster using the 'Wisconsin' DNA package and the EMBL and GENBANK data banks (Daresbury, England).

## Results

### *Differential screen*

Previously we described the preparation of a testis cDNA library made in the plasmid vector pUC9 (Dudley *et al.* 1984). Clones expressed more abundantly in the testis than in the liver or exclusively in the testis were identified by screening against randomly primed cDNA probes made from liver and testis poly(A)$^+$ RNA. Of 600 clones originally identified as 'testis specific', 259 were further screened with randomly primed cDNA probes made from the 18S poly(A)$^+$ RNA isolated from the testis of 2-week and 3-week-old mice, in order to distinguish expression in premeiotic (2 week) and meiotic or postmeiotic stages (3 week). Twelve clones that reproducibly showed preferential hybridization to the 3 week probe were selected, colony purified and studied further. DNA from plasmid preparations of these clones was used to make probes to screen Northern blots made from total testis RNA from 2-week and 3-week-old mice as well as kidney, liver and brain RNA. Six of the clones hybridized only to RNA isolated from the testis of 3-week-old mice and not to RNA from 2-week-old testis samples or other tissues, and hence were regarded as showing meiotic or postmeiotic expression specific to the germ cells (eg. clone 46, Fig. 1A). The remaining six clones detected transcripts in all the tissue RNAs examined (eg. clone 1, Fig. 1B). The clones that showed the testis-specific postmeiotic pattern of expression were used as probes on Southern blots of mouse and human genomic DNA. The hybridization data showed that (1) all six clones were present in low copy number in the mouse genome; (2) all six showed an identical pattern of hybridization regardless of the sex of the mouse the DNA was prepared from, suggesting that none of the clones mapped to the Y chromosome; and (3) the probes hybridized to human DNA under conditions of moderate stringency suggesting some degree of evolutionary conservation (results not shown). One of these clones, called clone 46, was selected for extensive analysis.

### *Tissue-specific expression of clone 46*

Clone 46 was hybridized to a northern blot of RNA samples isolated from a wide range of somatic tissues including testis and female reproductive tract (Fig. 1C). The resulting hybridization was specific to testicular RNA even on prolonged exposure, thus confirming the strict tissue specificity of this gene.

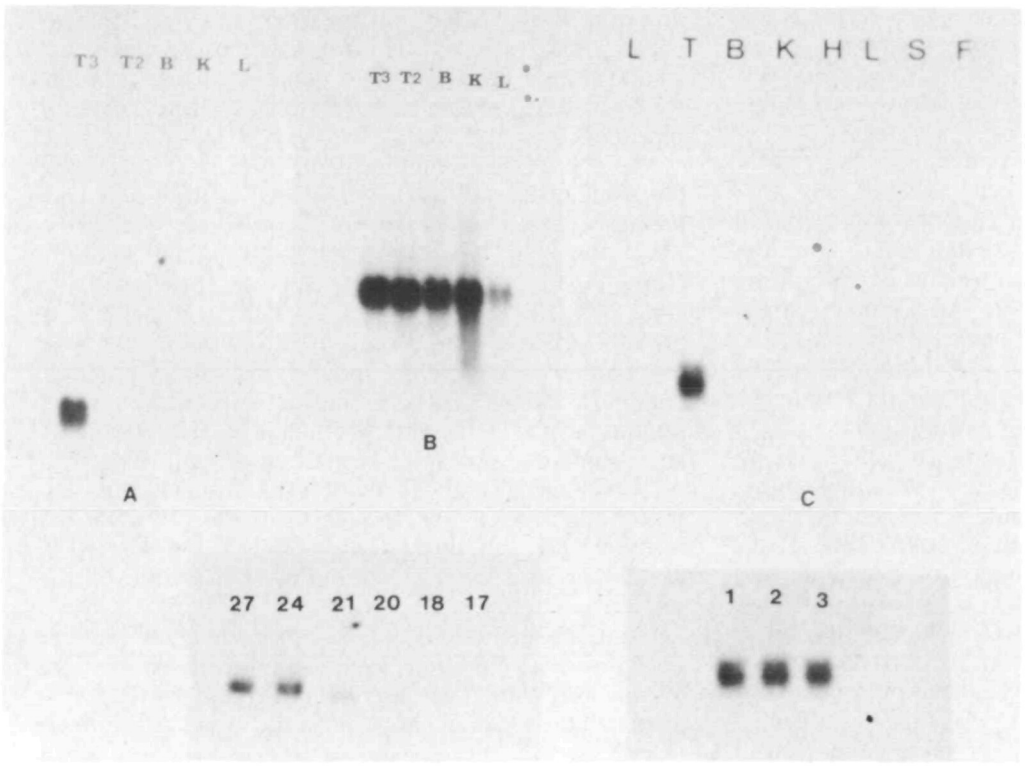



**Fig. 1.** Northern blot analysis of expression of clone 46 (A, C, D and E) and clone 1 (B). (A) The lanes indicated are T3 (3-week-old testis), T2 (2-week-old testis), B (brain), K (kidney) and L (liver). (B) The same samples as in A on a different blot probed with clone 1. (C) Clone 46 used to probe RNA from L (liver), T (adult testis), B (brain), K (kidney), H (heart), L (lung), S (spleen) and F (female reproductive tract). (D) Clone 46 used to probe RNA isolated from the testis of juvenile animals, the numbers indicate the age of the animal in days. (E) Clone 46 used to probe RNA made from the testis of (1) CBA/Ca mice, (2) $t^{w2}/t^{w1}$ *t*-mutants or (3) $T/t^{w1}$ *t*-mutants. Two transcripts, an upper (faint) of about 3 kb and a lower (intense) of 2 kb in size (judged in relation to the 18s and 28s rRNA species) are present in panels A,C–E. Clone 1 was isolated from the same differential screen as clone 46. It has a transcript of about 1 kb. All blots were hybridised at 42 °C for 18 h in 50 % formamide, 5×SSC with denatured sonicated salmon sperm DNA (50 $\mu g\,ml^{-1}$) as competitor. Washes were in 2×SSC at room temperature for 30 min, followed by 0.1×SSC, 0.1 % SDS at 55 °C for 30 min.

*Temporal and developmental expression of clone 46*

On a northern blot of RNA samples isolated from the testis of CBA/Ca mice ranging in age from 17 days up to 27 days clone 46 was shown to be expressed at just detectable levels at day 18 (Fig. 1D). This suggests that this clone corresponds to a gene first expressed in late pachytene spermatocytes/early secondary spermatocytes and that its expression is regulated in a developmental manner. The results showed the presence of two transcripts, approximately 2 and 3 kb in size, with the lower molecular weight transcript being more abundant and accumulating in the later stages of spermatogenesis (day 24 onwards). These transcripts could be detected after short autoradiographic exposure (30 min) indicating that they are highly abundant in the testis while their absence in testis RNA prepared from mice younger than 17 days (Fig. 1A) indicates that both transcripts are specific to the germ cells. Using clone 46 on northern blots of mouse embryo RNA no signal was detected between day 10 and the end of gestation (results not shown). This suggests either no expression of this gene during embryogenesis or expression in specific embryonic tissues at levels below detection using northern blotting of RNA isolated from whole embryos.

*Expression of 46 in* t*-mutant mice*

To determine whether there was any variation in the pattern of expression of clone 46 in the testes of *t*-mutant mice, RNA was prepared from the testes of mice of different haplotypes and probed on northern blots. Using homozygous and heterozygous animals bearing either complete or partial *t*-haplotypes, we were unable to detect any quantitative or qualitative differences between these animals and CBA/Ca (Fig. 1E).

*Sequencing of cDNA clones*

DNA sequencing revealed that clone 46 was a partial cDNA 1515 bp in length with a *Hin*dIII at its 5′ end, the result of digesting the cDNA with that enzyme during preparation of the library (Fig. 2). An 870 bp *Hin*dIII–*Acc*I fragment from the 5′ end of clone 46 was used to screen a $t^6/t^{w1}$ mouse testis cDNA library. A clone 2028 bp in length was isolated and shown to contain the entire coding sequence. This clone was called pβs13 and is the *t*-mutant equivalent of clone 46. pβs13 has an open reading frame starting at nucleotide 72 and ending at nucleotide 1770, encoding a polypeptide of 566 amino acids and relative molecular mass 61976 (Fig. 2). A comparison of the DNA sequence of clones 46 and pBs13 showed there are two silent base substitutions at nucleotides 510 and 789 and a G to C substitution at position 508 changing a glycine in pβs13 to an arginine in clone 46. The initiation of translation occurs at a sequence that conforms well with the established consensus (Kozak, 1984) and is completely identical to that of the preprovasopressin neurophycin II gene listed in the same paper. Analysis of the cDNA revealed that there was no upstream termination codon but genomic sequencing has revealed that there is a stop codon 75 nucleotides upstream from the start of translation, 3 nucleotides beyond the end of the cDNA (results not shown). The AAUAAA consensus polyadenylation signal (Proudfoot and Brownlee, 1976) is present 12 bp upstream of the start of the poly A tail.

A clustering of CpG dinucleotides is present at the 5′ end of pBs13 and the presence of single *Sac*II (nucleotide 141) and *Pvu*I (nucleotide 187) sites may indicate the presence of an HTF island at the 5′ end of this clone (Bird, 1986). However, HTF islands (*Hpa*II tiny fragment) are clusters of CpG dinucleotides in an unmethylated state and we currently do not know anything concerning the methylation patterns of pBs13. Four G/C boxes (GGGCGG or CCGCCC), which are usually associated with upstream promotor sequences, as well as 5′ coding regions in a number of genes including chicken β-actin and hamster HMG CoA reductase (Gardiner-Garden and Frommer, 1987), are also present at positions 39, 92, 220 and 329.

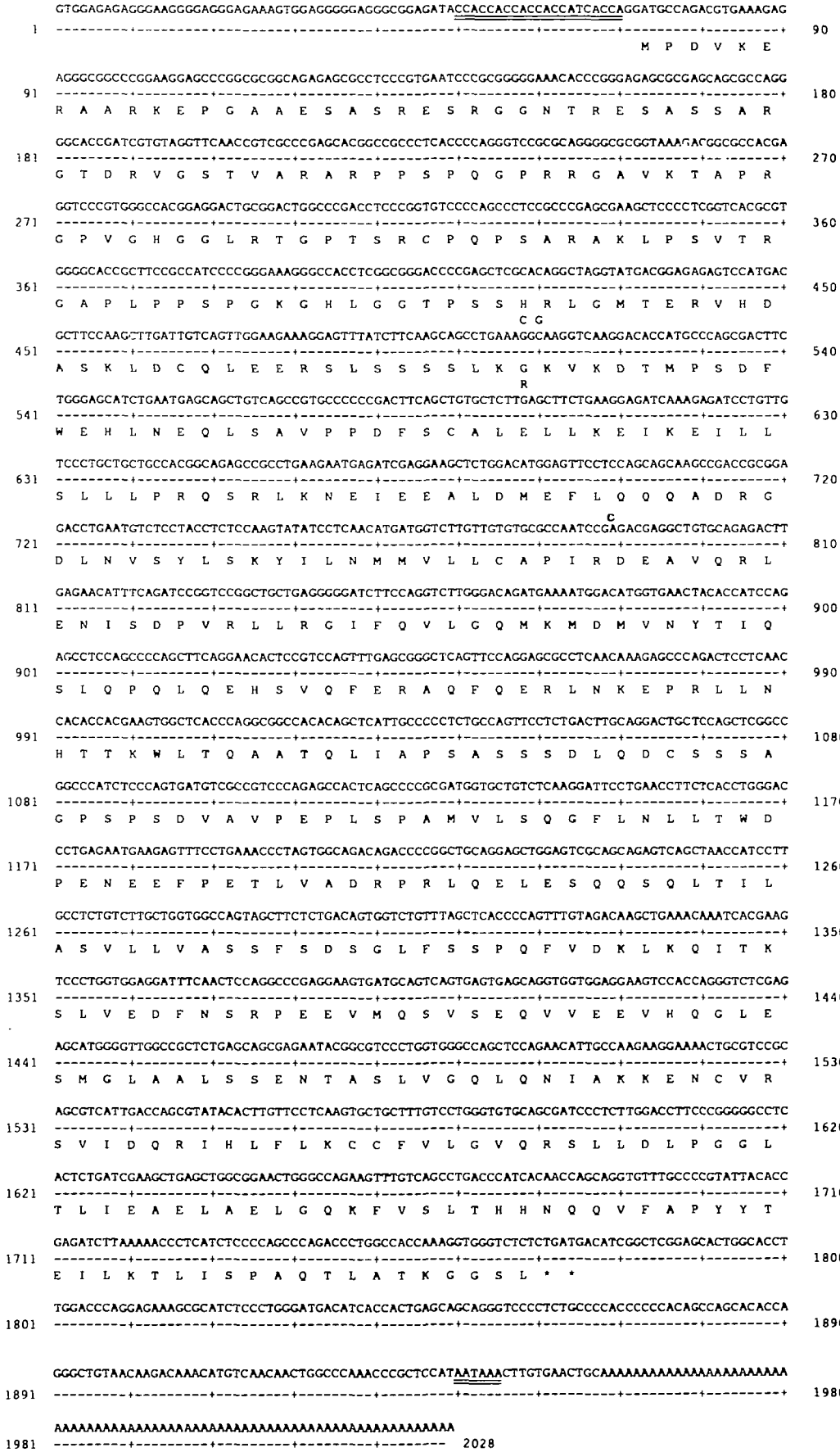
*Analysis of predicted protein structure*

Computer analysis of the protein products of clones pBs13 and 46 did not identify any potential hydrophobic membrane spanning domains. Secondary structure prediction analysis of the protein encoded by clone pBs13 using the combined algorithm of Eliopoulos *et al.* (1982) reveals extensive α-helical structure in this protein (Fig. 3). The prominent feature of the α-helices is the presence of heptad repeats of the consensus sequence (abcdefg)n with a high density of hydrophobic residues occupying positions a (72%) and d (65%) in this repeating unit. The presence of these apolar residues at positions a and d in the heptad repeat can lead to the formation of the α-helical structure through coiled-coil interactions with similar amino acids of other repeats (Crick, 1953).

The pBs13 amino acid sequence was used to search the protein data banks, as described in Materials and methods. Stretches of significant sequence similarity were observed between pBs13 and mouse laminin (1606–1764; pBs13:342–483) chicken apolipoprotein (125–220; pBs13:365–460) and with a hypothetical protein BGRF1 (102–253; pBs13:177–329) of unknown function encoded by the B95-8 Epstein–Barr virus (Baer *et al.* 1984). These similarities extended over the regions of pBs13 that are predicted to be alpha helix and contain imperfect heptad repeats. However, comparison of the pBs13 amino acid sequence with the proteins encoded by the testis-expressed cDNAs, including *Tcp-1* (Willison *et al.* 1986), *117c3* (Rappold *et al.* 1987) and *T66* (Schimenti *et al.* 1988), showed no sequence identity.

Using the PROSITE database, a number of consensus motifs with a possible significant role in the function of this protein have been identified in particular a protein kinase C phosphorylation site, a cell attachment

**Fig. 2.** Nucleotide sequence of pBs13. The position of nucleotides and amino acid differences between pBs13 (*t* form) and clone 46 (wild-type form) are indicated. The poly A addition motif and a CCA repeat are underlined.



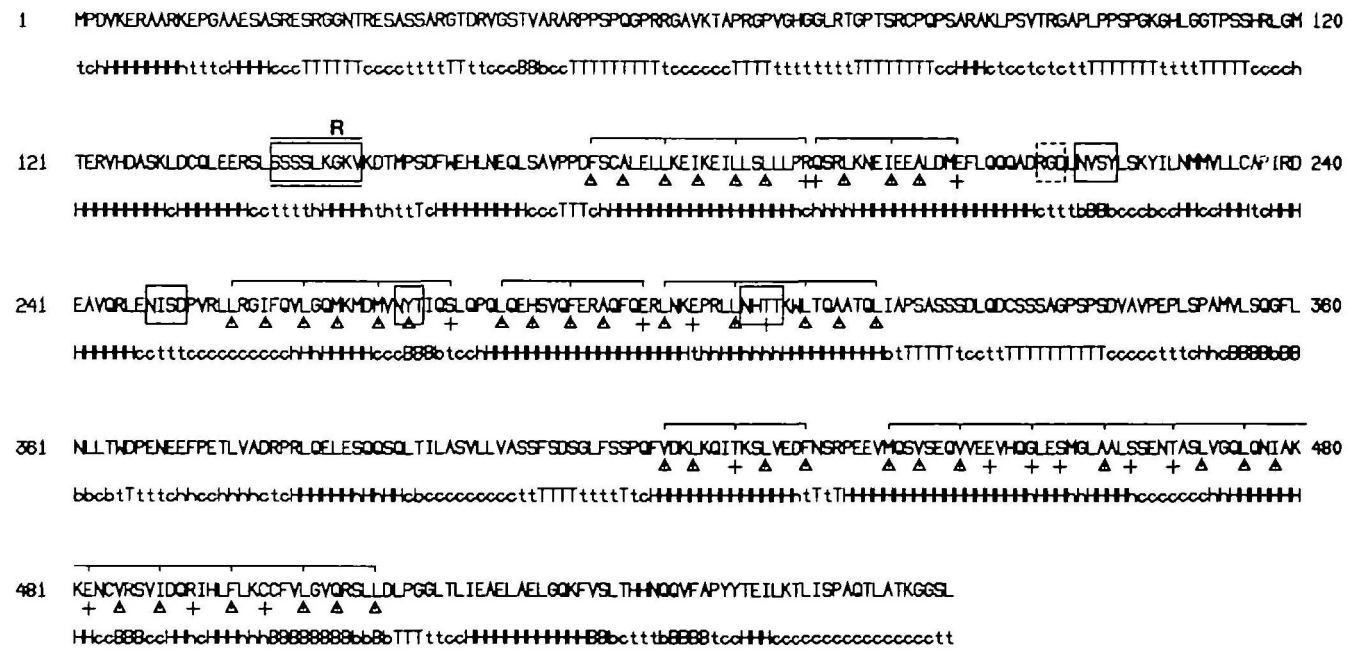


**Fig. 3.** The pBs13 predicted amino acid sequence is shown in single letter code. The combined secondary structure predictions from eight separate methods (Eliopoulos *et al.* 1982) are shown underneath the sequence and are presented by the letters H, B and T ($\alpha$ helix, $\beta$ sheet and turns predicted by >5 methods respectively) and h, b and t ($\alpha$ helix, $\beta$ sheet and turns predicted by <5 methods respectively) and c (ambiguous prediction). The heptad repeats (abcdefg)n (see text) are labelled by square brackets with positions a and d marked with triangles or crosses (triangles indicate hydrophobic residues), Potential N-glycosylation sites (□), a protein kinase C phosphorylation site ( ⊡ ), and a RGD cell attachment motif ( ⬚ ) are shown. The leucine heptad runs from amino acid 299 to amino acid 320. The amino acid substitution of arginine for glycine within the putative protein kinase C phosphorylation site is indicated (R).

RGD motif (but no export signal sequence), a leucine heptad repeat and four potential glycosylation sites (Fig. 3). The N-glycosylation site at position 248 has a proline at the C-terminal position, which may suppress modification at this site (Bause, 1983), although it remains unknown at present if any of the potential glycosylation and phosphorylation sites are utilized.

*Chromosomal mapping*

Southern analysis using clone 46 and the BXD recombinant inbred parental strains failed to reveal any

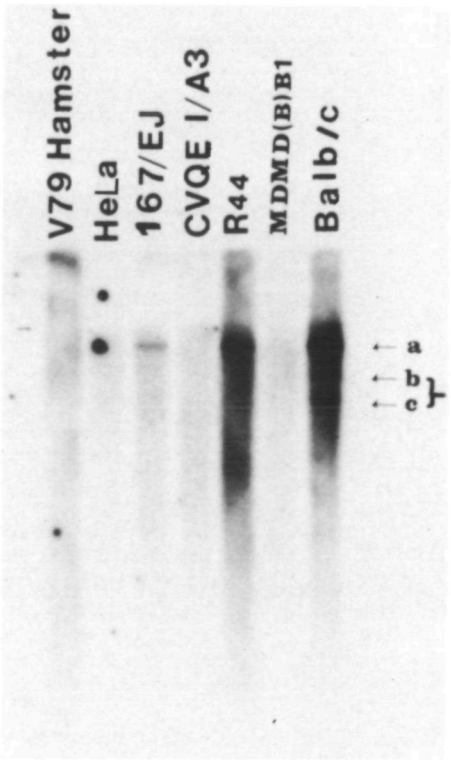
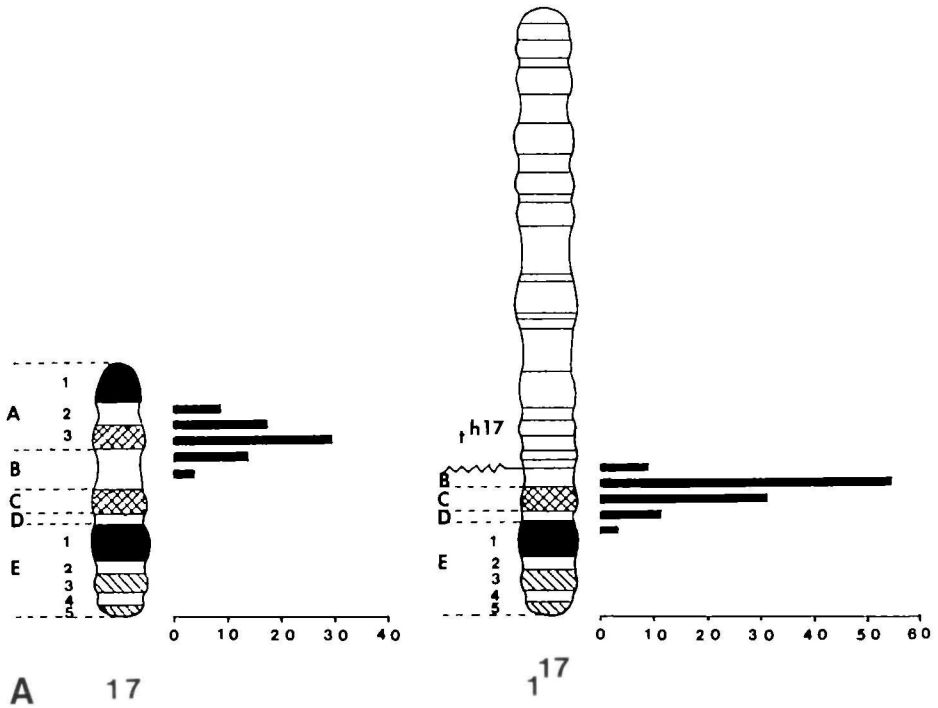



**Fig. 4.** Chromosomal assignment of clone 46. Clone 46 was used to probe a Southern blot of DNA prepared from a series of cell lines (see text for description) and restricted with *Eco*RI. a indicates an allele (15 kb) common to Balb/c genomic DNA and the R44 and 167/EJ cell lines both of which contain mouse chromosome 17. b and c are alleles (7 and 3 kb) present in Balb/c genomic DNA but which are not detected in the two cell lines. V79 Hamster and HeLa are controls for the chromosomal background on which the mouse chromosomes appear, and CVQE and MDMD(B)B1 contain proximal and distal portions of mouse chromosome 17 respectively on monkey or dog backgrounds. The CVQE and MDMD(B)B1 cell lines do not contain fragments covering the length of the t-complex. Hybridization was at 65°C in 0.5 M $NaPO_4$ buffer (pH 8.0) and 7% SDS for 20 h. Final wash was at 0.1×SSC, 0.1% SDS at 65°C. Autoradiography was on Fuji RX film at 70°C for 20 h with intensifying screens.

polymorphisms using 15 different restriction endonucleases (results not shown).

Somatic cell hybrids were used to determine the chromosomal localization of the gene corresponding to clone 46. Cell line R44 has mouse chromosomes 17 and 18 on a hamster background (Smiley *et al.* 1978) while cell line 167/EJ has mouse chromosomes 17 and 3 on a human background (P. Goodfellow, unpublished observations). On hybridization with clone 46, a 15 kb allele was detectable in the genomic DNA from both cell lines, identical to the allele detected in the BALB/c genomic DNA (Fig. 4). The common allele detected with these three genomic samples was not present in hamster DNA. Two other cell lines that had small parts of mouse chromosome 17 on dog or monkey backgrounds failed to show the common allele on hybridization with clone 46; cell line MDMB(B)B1 has a distal fragment of chromosome 17 on canine background (fragment distal to *Crya* locus spanning through the MHC to some point within *Tla-Qa*, Weis *et al.* 1986) while cell line CVQE has proximal parts of chromosome 17 on monkey background (D. Nelson, unpublished results). Two fainter restriction alleles (7 and 3 kb) present in the *Eco*RI-digested BALB/c DNA were absent from the R44 cell line indicating that these represent homologous loci residing on other mouse chromosome(s) but these have not yet been mapped. These data suggested strongly that clone 46 mapped to chromosome 17.

*In situ* hybridization experiments were performed in order to confirm the data from the somatic cell hybrids and establish where on chromosome 17 clone 46 mapped. Chromosome 17 was marked using the T(1;17)190Ca reciprocal translocation (Lyon *et al.* 1986), which also carries the distal partial *t*-haplotype $t^{h17}$ and which gives fairly recognisable long and short marker chromosomes. Metaphase spreads from a mouse heterozygous for the T190 Ca translocation,



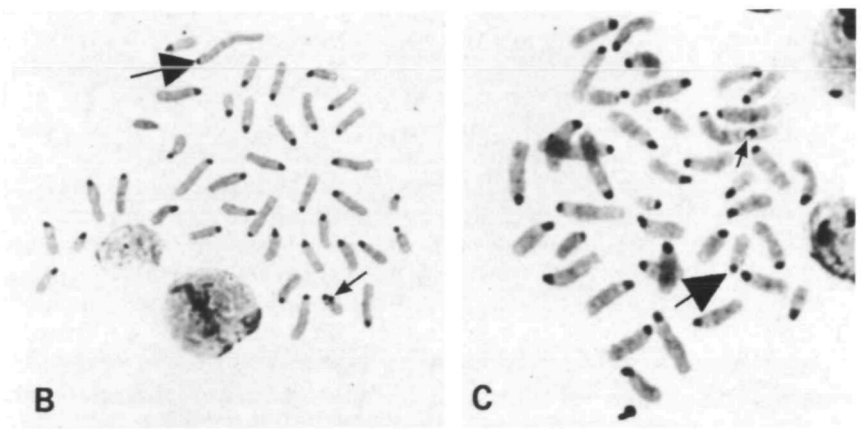
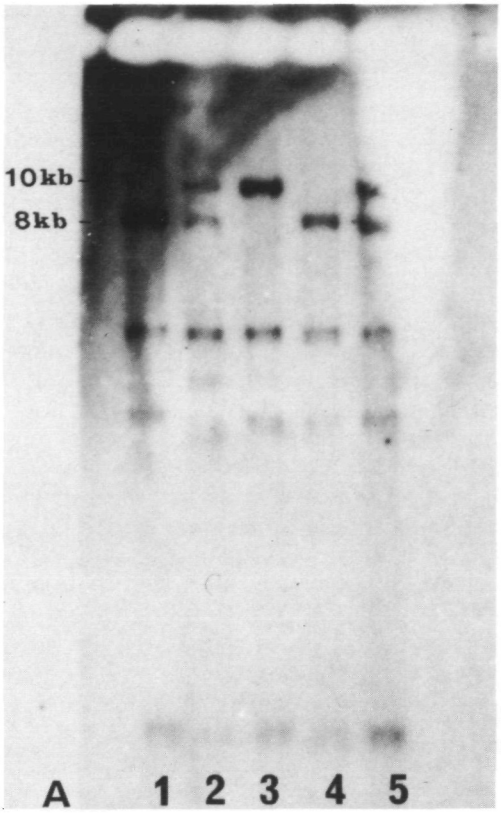




**Fig. 5.** (A) Diagram of grain distribution patterns after hybridization of clone 46 to, left, a normal chromosome and, right, to the long marker chromosome of T190. (B) Mitotic cell showing hybridization of clone 46 to band 17A3 (small arrowhead) on the normal chromosome 17. The long T190 translocation marker is indicated with a larger arrowhead. (C) Cell showing hybridization of clone 46 to the distal third of band 17B (small arrowhead) carried on the long T190 translocation marker chromosome. The normal chromosome 17 is indicated by the larger arrowhead.



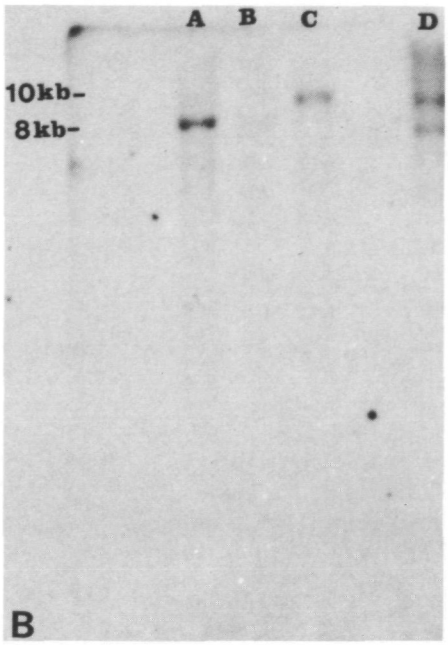




**Fig. 6.** (A) Mapping of pBs13 to the distal region of the mouse *t*-complex using partial haplotype analysis. pBs13 was used to probe a Southern blot of mouse genomic DNA digested with the restriction enzyme *Sst*1. (A) Lane 1: Wild-type (TO) DNA. Lane 2: $T/t^{w1}$ DNA. Lane 3: $t^{w1}/t^{w2}$ DNA. Lane 4: $t^{w18}/+$DNA. Lane 5: $t^{w18}/t^{6}$ DNA. The absence of the 10 kb *t*-allele in the $t^{w18}/+$DNA indicates that pBs13 either maps into the $t^{w18}$ deletion or lies distal to the deletion. The presence of both *t* and wild-type alleles in $t^{w18}/t^{6}$ DNA shows that pBs13 maps distal to the $t^{w18}$ deletion. (B) Lane A: Wild-type (TO) DNA. Lane B: $T/t^{w1}$ DNA. Lane C: $t^{w1}/t^{w2}$ DNA. Lane D: $t^{h20}/+$ DNA. The presence of both alleles in $t^{h20}/+$ DNA indicates that pBs13 lies proximal or distal to the $t^{h20}$ deletion. Blots were hybridized at 42°C for 24 h in 50% formamide, 5×SSC with denatured sonicated salmon sperm DNA (50 $\mu$g ml$^{-1}$) as competitor. Blots were washed twice in 2×SSC, 0.1% SDS at room temperature for 30 min, and then in 0.1×SSC, 0.1% SDS for 30 min at 65°C. Autoradiography was at −70°C for 72 h with intensifying screens.

have the distal *t*-haplotype inversion on the long marker chromsome while the homologous chromosome 17 is normal and wild type. The data from this experiment are shown in Fig. 5(A,B,C). 63 grains were detected in bands 17B/C of the translocated $1^{17}$ chromosome with 49 grains in region 17A3/B of the normal chromosome. 297 grains were scored with 38% in the relevant region. The difference in band position of the grains on the translocated and normal chromosomes is attributable to the presence of the *t*-complex distal inversion. Lyon *et al.* (1988) showed that this inversion extends from band A3 to band C. Thus since the peak grain distribution for clone 46 is in bands A3-B on a normal chromosome and bands B–C on a *t*-chromosome this clone maps within the distal inversion and near its proximal edge.

Attempts to map the gene more precisely were made with the pBs13 probe using an *Sst*I polymorphism detected between wild (TO) and *t*(*w1*/*w2*) mouse DNA. The proximal partial haplotype $t^{w18}$ carries a small deletion at the proximal edge of the distal inversion, with wild-type chromatin beyond (Bucan *et al.* 1989). The haplotype $t^{h20}$ also carries a small deletion, distal to and not overlapping the $t^{w18}$ deletion (Lyon *et al.* 1979), and has *t*-chromatin beyond. Southern analysis of *Sst*1 restricted genomic DNA from mice heterozygous for the $t^{w18}$ and $t^{h20}$ haplotypes reveals that the *t* specific 10 kb allele is absent in the $t^{w18}/+$ DNA but present in the $t^{h20}/+$ DNA (Fig. 6A,B). Animals of genotype $t^{w18/t6}$ showed both the *t* and wild-type allele. Thus, the gene must map either between the $t^{w18}$ and $t^{h20}$ deletions or distal to the $t^{h20}$ deletion.

*Hybrid selections* in vitro *translations*

In order to establish whether clone 46 encoded a protein that was a candidate for any of the polymorphic *t*-complex polypeptides (Silver *et al.* 1983) hybrid selection/*in vitro* translation was performed using DNA from pBs13 and testis RNA isolated from wild-type (C57BL/6) and homozygous *t* ($t^{w2}t^{w1}$) mice (Fig. 7A). The results showed that with both RNA samples a major band was detectable at about $63\times10^3$ $M_r$ with two less intense, but specific bands at about $74\times10^3$ and $52\times10^3$ (Fig. 7B). The estimated relative molecular mass of the major hybrid selected product (63 000) is in good agreement with the coding capacity of the pBs13 open reading frame (61 976). When these *in vitro* products were analyzed on two-dimensional gels the $63\times10^3$ $M_r$ polypeptide focused to a spot with an acidic pI of about 6.2 and the $74\times10^3$ $M_r$ protein band resolved into two spots of pI of about 7.0 (Fig. 7B). The $52\times10^3$ band did not focus under the conditions used. No polymorphisms were identified between *t* and wild mice on two-dimensional gels suggesting that these proteins do not correspond to any of the *t*-complex proteins described by Silver *et al.* (1983).

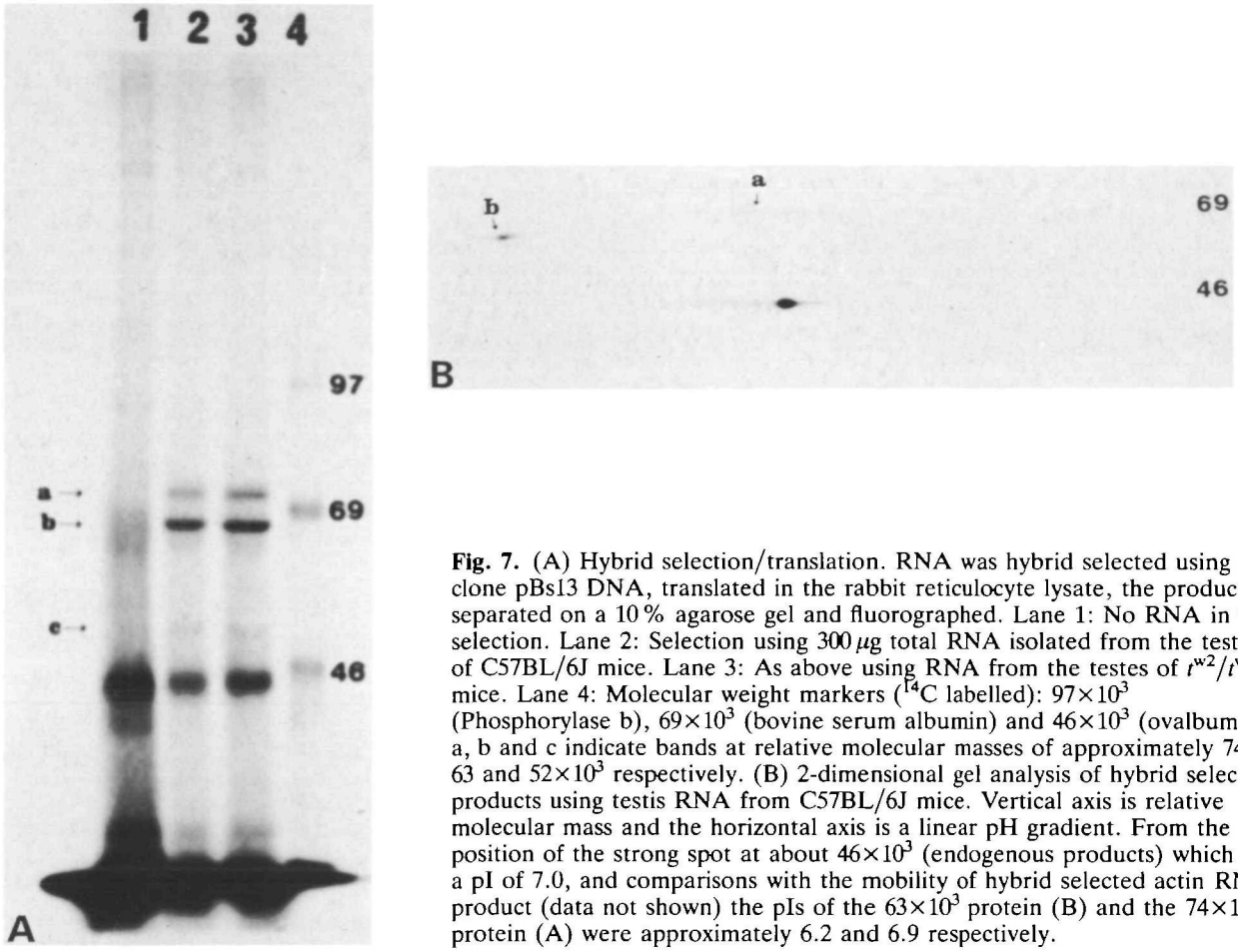



**Fig. 7.** (A) Hybrid selection/translation. RNA was hybrid selected using clone pBs13 DNA, translated in the rabbit reticulocyte lysate, the products separated on a 10% agarose gel and fluorographed. Lane 1: No RNA in the selection. Lane 2: Selection using 300 $\mu$g total RNA isolated from the testes of C57BL/6J mice. Lane 3: As above using RNA from the testes of $t^{w2}/t^{w1}$ mice. Lane 4: Molecular weight markers ($^{14}$C labelled): $97\times10^3$ (Phosphorylase b), $69\times10^3$ (bovine serum albumin) and $46\times10^3$ (ovalbumin). a, b and c indicate bands at relative molecular masses of approximately 74, 63 and $52\times10^3$ respectively. (B) 2-dimensional gel analysis of hybrid selection products using testis RNA from C57BL/6J mice. Vertical axis is relative molecular mass and the horizontal axis is a linear pH gradient. From the position of the strong spot at about $46\times10^3$ (endogenous products) which has a pI of 7.0, and comparisons with the mobility of hybrid selected actin RNA product (data not shown) the pIs of the $63\times10^3$ protein (B) and the $74\times10^3$ protein (A) were approximately 6.2 and 6.9 respectively.

## Discussion

In this paper, we describe the structure and expression of a novel gene mapping to the mouse *t*-complex. This gene has a strict testis-specific developmental pattern of expression similar to that described for other *t*-complex genes such as *T66* (Schimenti *et al.* 1988) and *117c3* (Rappold *et al.* 1987). Both these genes are only expressed in male germ cells and transcripts are first detected in pachytene spermatocytes as the cells progress through spermatogenesis. This pattern of expression has also been demonstrated for phosphoglycerate kinase 2 (Gold *et al.* 1983) and many as yet uncharacterised cDNA clones (Dudley *et al.* 1984; Thomas *et al.* 1989). The pachytene spermatocytes are known to be the most transcriptionally active of the germ cells (Bellvé, 1979) and the increasing number of cDNA clones which first detect transcripts at this stage may, in part, reflect this fact. For some of the clones mentioned above transcripts first detected at the pachytene stage are also detected in the secondary spermatocytes and in the haploid cells although in most cases there is no evidence that the genes continue to be transcribed in these later cells.

Clone pBs13 identifies three bands on a Southern blot of DNA isolated from BALB/c mice and digested with *Eco*RI. Data from the somatic cell hybrids reveal that only one of these hybridizing sequences maps to chromosome 17. The chromosomal location of the other two sequences remains unknown but it is significant that they hybridize only weakly to pBs13 suggesting that they have only limited sequence identity to the probe. In contrast, only one peak of hybridization, on chromosome 17, is seen on *in situ* hybridization to metaphase spreads. This supports our view that pBs13 is not derived from transcripts originating from the two weakly hybridizing alleles detected on Southern blots and has its origins on chromosome 17. Significantly, we detect two transcripts with pBs13 on northern blots of testes RNA and can identify three polypeptides after hybrid selection. In the absence of any evidence that pBs13 is the product of differential splicing we propose that the larger, and less abundant, transcript detected on northern blots may be the product of one of the alleles mapping outside chromosome 17. In a similar fashion, the $74\times10^3$ $M_r$ polypeptide detected on hybrid selection would be the result of translating this larger transcript. We have not detected a third transcript to account for the third polypeptide.

A comparison of the sequence of pBs13 (*t* clone) and clone 46 (wild-type clone) reveals that between nucleotides 458 (the start of clone 46) and the point at which the open reading frame is closed (nucleotide 1769) there are three nucleotide substitutions, one of

which results in a glycine in the *t* being replaced by an arginine in the wild-type protein. This substitution modifies a potential protein kinase C phosphorylation site, which if utilised could lead to altered function between the two proteins. Over a similar number of nucleotides a comparison of the *t* and wild-type forms of *Tcp-1* revealed 8 base changes leading to 6 amino acid substitutions (Willison *et al.* 1986). Within the coding sequence, then, it seems that clone pBs13 has diverged significantly less from the wild-type allele than *Tcp-1*, particularly at the amino acid level. The significance of this is unclear but it may reflect the fact that clone pBs13 and *Tcp-1* are located in the distal and proximal regions of the *t*-complex, respectively. There is an increasing body of evidence suggesting that the history of these two parts of the *t*-complex is different. There are many more partial *t*-haplotypes with breakpoints in the proximal than in the distal inversion. In addition, the strong linkage disequilibrium in the proximal inversion which has led to all partial haplotypes from this region having an identical allele for markers such as *Tcp-1* is not detected in the distal inversion. The evidence points to a mechanism involving segmental exchange in the distal inversion as opposed to single crossovers in the proximal (Erhart *et al.* 1988). An alternative explanation to account for the difference in sequence conservation between the *t* and wild-type forms of *Tcp-1* and clone 46 is that due to differences in function clone 46 protein is under greater evolutionary constraints than TCP-1. We are currently extending our analysis of clone 46 to establish how well the sequence is conserved between different *t*-haplotypes and different mouse strains.

The data presented here suggest that this gene may play an important role in sperm development. It is exclusively expressed in the male germ cells and is first detectable as the cells prepare to enter spermiogenesis. It has been mapped to the *t*-complex, a region of chromosome 17 known to harbour a number of genes expressed in the germ line. Furthermore from its position this gene becomes a candidate for distorter gene *Tcd-2* (Lyon, 1986). The presence of heptad repeats and an RGD sequence may indicate that this protein is transported to the cell surface but an understanding of its role in germ cell development will have to await further studies.

This work was carried out with the support of the Wellcome Trust and an SERC postgraduate studentship (N.M.). We acknowledge Keith Willison and Alan Ashworth for providing libraries and useful discussions, Lisa Stubbs for Southern filters and advice and Philip Cunningham for computer analysis.

## References

AKRIGG, D., BLEASBY, A. J., DIX, N. I. M., FINDLAY, J. B. C., NORTH, A. C. T., PARRY SMITH, D., WOOTTON, J. C., BLUNDELL, T. L., GARDNER, S. P., HAYES, F., ISLAM, S. A., STERNBERG, M. J. E., THORNTON, J. M. AND TICKLE, I. J. (1988). A protein sequence/structure database. *Nature* **335**, 745–746.

ARTZT, K., SHIN, H.-S. AND BENNETT, D. (1982). Gene mapping within the T/*t* complex of the mouse; anomalous position of the H-2 complex in t-haplotypes. *Cell* **28**, 471–476.

BAER, R., BANKSER, A. T., BIGGIN, M. D., DEININGER, P. L., FARRELL, P. J., GIBSON, T. J., HARTHILL, G.; HUDSON, G. S., SATCHWELL, S. C., SEGUIN, C., TUFFNELL, P. S. AND BARRELL, B. G. (1984) DNA sequence and expression of the B 95–8 Epstein-Barr virus. *Nature* **310**, 207–211.

BAUSE, E. (1983). Structural requirements of N-glycosylation of proteins. *Biochem. J.* **209**, 331–336.

BELLVÉ, A. R. (1979). In Finn C. A. (ed.) *Oxford Reviews of Reproductive Biology*, Vol. 1, pp. 159–261. London: Oxford University Press.

BENNETT, D. (1975). The T-locus of the mouse. *Cell* **6**, 441–454.

BIRD, A. P. (1986). CpG Rich Islands and the Function of DNA Methylation. *Nature* **321**, 209–213.

BUCAN, M., HERMANN, B. G., FRISCHAUF, A. M., BAUTCH, V. L., BODE, V., SILVER, L. M., MARTIN, G. R. AND LEHRACH, H. (1989). Deletion and duplication of DNA sequences is associated with the embryonic lethal phenotype of the *t* complementation group of the mouse *t*-complex. *Genes and Dev.* **1**, 376–385.

CHIRGWIN, J. M., PRZYBYLA, A. E., MACDONALD, R. J. AND RUTTER, W. J. (1979). Isolation of Biologically Active Ribonucleic Acid From Sources Rich in Ribonuclease. *Biochemistry* **24**, 5294–5298.

COLLINS, J. F., COULSON, A. F. W. AND LYALL, A. (1988). The significance of protein sequence similarities, *Computer Applications in the Biosciences* **4**, 67.

CRICK, F. H. C. (1953). The packing of helices; simple coiled-coils. *Acta Crystallogr.* **6**, 689–697.

DAYHOFF, M. (1978). *Atlas of Protein Sequence and Structure* **3** National Biomedical Research Foundation, USA.

DOOHER, G. B. AND BENNETT, D. (1977). Spermiogenesis and spermatazoa in sterile mice carrying different letal T/t locus haplotypes: A transmission and scanning electron microscopic study. *Biol. Reprod.* **17**, 269–288.

DUDLEY, K., POTTER, J., LYON, M. F. AND WILLISON, K. R. (1984). Analysis of male sterile mutations in the mouse using haploid stage expressed cDNA probes. *Nucl. Acids Res.* **12**, 4281–4293.

ELIOPOULOS, E. E., GEDDES, A. J., BRETL, M., PAPPIN, D. J. C. AND FINDLEY, J. B. C. (1982). A structural model of the chromophore binding domain of ovine rhodopsin. *Int. J Biol. Macromol.* **4**, 263–268.

ERHART, M. A., PHILLIPS, J. J. AND NADEAU, J. H. (1988). Contrasting patterns of evolution in the proximal and distal regions of the mouse t complex. In *Current Topics in Microbiology and Immunology* **137**, 71–75.

FEINBERG, A. P. AND VOGELSTEIN, B. (1984). A technique for radio labelling restriction endonuclease fragements to high specific activity. *Anal. Biochem.* **137**, 266–276.

FOX, H. S., MARTIN, G. R., LYON, M. F., HERMANN, B., FRISCHAUF, A-M., LEHRACH, H. AND SILVER, L. M. (1985). Molecular probes define different regions of the mouse t-complex. *Cell* **40**, 63–69.

GARDINER-GARDEN, M AND FOMMER, M. (1987). CpG Islands in Vertebrate Genomes. *J. molec Biol.* **196**, 261–282.

GOLD, B., FUJIMOTO, H., KRAMER, J. M., ERICKSON, R. P. AND HECHT, N. B. (1983). Haploid accumulation and translational control of phosphoglycerate Kinase-2 mRNA during Mouse Spermatogenesis. *Devl Biol.* **8**, 392–399.

HAMMER, M. F., SCHIMENTI, J. AND SILVER, L. M. (1989). Evolution of mouse chromosome 17 and the origin of inversions associated with *t*-haplotypes *Proc. natn. Acad. Sci. U.S.A.* **86**, 3261–3265.

HARTL, D. L. AND HAIRAIZUMI, Y. (1976). Segregation distortion. In *The Genetics and Biology of Drosophila* (ed. M. Ashburner and E Novitski). Vol. 16, pp. 615–666. New York: Academic Press.

HECHT, N. B. (1988). Regulation of gene expression during mammalian spermatogenesis. In *Experimental Approaches to Mammalian Embryonic Development* (ed. J. Rossant and R. A. Pedersen) pp. 151–194. Cambridge: Cambridge University Press.

HERMANN, B., BUCAN, M., MAINS, P. E., FRISCHAUF, A. M.,

Silver, L. M. and Lebrach, H. (1986). Genetic analysis of the proximal portion of the mouse t-complex: evidence for a second inversion within t-haplotypes. *Cell* **44**, 469–476.

Kozak, M. (1984). Compilation and analysis of sequences upstream from the translational start site of eukaryotic mRNAs. *Nucl. Acids Res.* **12**, 857–872.

Lader, E., Ha, H.-S., O'Neill, M., Arzt, K. and Bennett, D. (1989). *tctex*-1. a candidate gene family for a mouse *t*-complex sterility locus. *Cell* **58**, 969–979.

Lyon, M. F. (1984). Transmission ratio distortion in mouse *t*-haplotypes is due to multiple distorter genes acting on a responder locus. *Cell* **37**, 621–628.

Lyon, M. F. (1986). Male sterility of the mouse *t*-complex is due to homozygosity of the distorter genes. *Cell* **44**, 357–363.

Lyon, M. F., Jarvis, S. E., Sayers, I. and Johnson, D. R. (1979). Complementation reactions of a lethal mouse *t*-haplotype believed to include a deletion. *Genet. Res.* **33**, 153–161.

Lyon, M. F. and Meredith, R. (1964). The nature of *t*-alleles in the mouse: genetic analysis of a series of mutant derived from a lethal allele. *Heredity* **19**, 301–327.

Lyon, M. F., Zenthon, J., Evans, E. P., Burtenshaw, M. D., Dudley, K. and Willison, K. R. (1986). Location of the *t*-complex on mouse chromosome 17 by *in-situ* hybridisation. *Immunogenetics* **24**, 125–127

Lyon, M. F., Zenthon, J., Evans, E. P., Burtenshaw, M. D. and Willison, K. R. (1988). Extent of the mouse *t*-complex and its inversions shown by *in situ* hybridisation. *Immunogenetics* **27**, 375–382.

Maniatis, T., Fritsch, E. F. and Sambrook, J. (1982). *Molecular Cloning: A Laboratory Manual.* New York: Cold Spring Harbor

Maxam, A. M. and Gilbert, N. (1977). A new method for sequencing DNA. *Proc. natn. Acad. Sci. U.S.A.* **74**, 560–564.

Olds-Clarke, P. and Peitz, B. (1986). Fertility of sperm from *t*/+ mice; evidence that +− bearing sperm are dysfunctional. *Genet. Res. Camb.* **47**, 49–52.

Proudfoot, N. J. and Brownlee, G. G. (1976). 3′ non-coding regional sequences in eucaryotic mRNA. *Nature* **263**, 211–214.

Rappold, G. A., Stubbs, L., Labeit, S. M., Ckvenjakov, R. B. and Lehrach, H. (1987). Identification of a testis specific gene from the mouse *t*-complex next to at CpG rich island. *EMBO J.* **6**, 1975–1980.

Sanger, F., Nicklen, S. and Coulson, H. R. (1977). DNA sequencing with chain terminating inhibitors. *Proc. natn Acad. Sci. U.S.A.* **74**, 5463–5467.

Schimenti, J., Cebra-Thomas, J. A., Decker, C. L., Islam, S D., Pilder, S. H. and Silver, L. M. (1988). A candidate gene family for the mouse *t*-complex responder (Tcr) locus responsible for haploid effects on sperm function. *Cell* **55**, 71–78.

Seitz, A. W. and Bennett, D. (1985). Transmission ratio distortion of *t*-haplotypes is due to interactions between meiotic partners. *Nature* **313**, 143–144.

Shin, H.-S. (1988). T/t complex and mouse development. In *Human Immunogenetics* (ed. S. D. Litivin). pp. 443–471. New York: Marcel Dekker.

Silver, L. M. (1985). Mouse *t*-haplotypes. *A. Rev. Genet.* **19**, 179–208.

Silver, L. M. and Remis, D. (1987). Five of the genetically defined regions of mouse t-haplotypes are involved in transmision ratio distortion. *Genet. Res. Camb.* **49**, 51–56.

Silver, L. M., Uman, S., Danska, J. and Garrets, J. I. (1983). A diversified set of testicular cell proteins specified by genes within the *t*-complex. *Cell* **35**, 35–45.

Smiley, J. R., Steege, D. A., Juricek, D. K., Summers, W. P. and Ruddle, F. H. (1978). A herpes-simplex virus 1 integration site in the mouse genome defined by somatic cell genetic analysis. *Cell* **15**, 455–468.

Smith, T. F. and Waterman, M. S. (1981). Identification of common molecular sub-sequence. *J. molec. Biol.* **147**, 195–197.

Southern, E. M. (1975). Detection of specific sequences using DNA fragements separated by gel electrophoresis. *J. molec. Biol.* **98**, 503–511.

Taub, F. and Thompson, E. B. (1982). An improved method for preparing large arrays of bacterial colonies containing plasmids for hybridisation: *in situ* purification and stable binding of DNA on filter paper. *Anal. Biochem.* **126**, 222–230.

Thomas, K. H., Wilkie, T. M., Tomashefsky, P., Bellve, A. and Simon, M. I. (1989). Differential gene expression during mouse spermatogenesis. *Biology of Reproduction* **41**, 729–739.

Weis, J. H., Seidman, J. G., Housman, D. E. and Nelson, D. L. (1986). Eucaryotic chromosome transfer: production of a murine specific cosmid library from a neo linked fragement of murine chromosome 17. *Molec. cell. Biol.* **6**, 441–451.

Willison, K. and Ashworth, A. (1987). Mammalian spermatogenic gene expression. *TIG* **3**, 351–355.

Willison, K. R., Dudley, K. and Potter, J. (1986). Molecular cloning and sequence analysis of a haploid expressed gene encoding *t*-complex polypeptide. *Cell* **44**, 727–738.

(*Accepted 27 November 1990*)

*Note*: In order to facilitate the positioning of clone 46 on the mouse gene map, it has been given the name D17 KEN.1. The sequence of pBs13 has the accession number X52128 in the EMBL Data Library.