Volume 15 Number 7 1987 Nucleic Acids Research

Molecular cloning and sequence analysis of a mouse Y chromosome RNA transcript expressed in the testis

C.E.Bishop and D.Hatat

Unité d'Immunogénétique Humaine, INSERM U.276, Institut Pasteur, 75724 Paris, Cedex 15, France

Received December 29, 1986; Revised February 27, 1987; Accepted March 17, 1987

ABSTRACT

Using a Y specific probe (pY353/B) taken from a flow sorted mouse Y chromosome library we have identified a family of RNA transcripts encoded by the Y chromosome. These transcripts which are approximately 1.3 Kb in length are present in testis PolyA$^+$ RNA but can not be detected in either male liver, spleen, kidney, brain, heart or lung tissues. Isolation and sequence analysis of a corresponding cDNA shows it to contain a potential coding sequence of 696bp.These data show that the Y chromosome is transcriptionally active in the adult mouse testis.

INTRODUCTION

It is generally considered that dominant genes on the mammalian Y chromosome control primary sex determination by inducing the bipotential foetal gonad to differentiate as a testis. This can occur even when multiple copies of the X are present (1,2). Studies of Y chromosome deletions and rearrangements in both man and mouse suggest that the Y also encodes genes involved in several aspects of spermatogenesis and sperm function (3,4). At the molecular level very little is known about the way in which the Y controls these processes. To date no Y specific genes have been cloned although in the human an X-Y common gene encoding a cell surface antigen has been isolated (6). In the mouse, Y limited DNA probes obtained from a variety of sources have been used to great effect by localizing most of these functions to a small pericentric region of the Y. This region can be defined by simple quaternary repeated sequences originally isolated from a female heterogametic snake (7). When it is present in the genome of XXSxr (sex reversed) mice the individuals are phenotypically male although sterile (8).

The mouse Y also contains a large amount of murine retroviral related sequences estimated by Philips et al (9) to compose up to 3% of the chromosome. Using a new Y chromosome rearrangement, Eicher et al (4) have been able to show that the majority of these sequences are contained within the central region of the Y and further this region may be important in sperm motility.

Recently evidence has been presented for a functional steroid sulphatase gene on both the mouse X and Y chromosomes the alleles of which undergo obligatory crossing over during male meiosis (10). Although direct evidence is lacking this would imply that the genes would map to the telomeric pairing/exchange region of the sex chromosomes (6, 12, 13, 14).

In order to study the way in which the Y functions in molecular terms it is essential to identify and clone Y located genes. We and others have shown in both man and mouse that in the absence of more conventional markers random DNA sequences taken from Y chromosome enriched libraries can provide a means of directly probing this chromosome (15, 16, 17, 18, 19).

Using this approach we report here the identification of a family of testis specific RNA transcripts encoded by the mouse Y chromosome.

MATERIALS AND METHODS

DNA and RNA blots

10 µg samples extracted from a single liver or clonal cell line were digested with restriction endonuclease according to the manufacturers specifications, separated on O.75% agarose gels and transferred to zetopore filter membrane essentially according to Southern (2O). The blots were probed with DNA labelled with ^{32}P by nick-translation to an approx. specific activity of $1O^8$ cpm/µg. Hybridization was for 16 hrs at 42° in 5O% formamide and 5xSSC. Filters were then washed 3x3O min. in 2xSSC followed by 3x2O min. in O.1xSSC at 68°. Filters were exposed to Kodak X-AR5 film at -70° for 16-24 hours with intensifying screens. Total RNA was extracted from tissues using the LiCl/Urea method of Auffray and Rougeon (21). PolyA$^+$RNA was selected by several passages over a poly U sepharose column.

For Northern blot analysis 5 µg samples of polyA$^+$ RNA were denatured with 1OmM methyl mercury and separated on vertical 1.4% agarose gels containing 7.5mM methyl mercury. RNA was then transferred to nitrocellulose membrane. Hybridization to nick translated probes was as for DNA blots, washing was 3x3O min at 55° in O.1xSSC O.1% SDS.

cDNA cloning

Total polyA$^+$RNA extracted from the testis of 6 week old 129/sv male mice was used. First strand synthesis was performed using AMV reverse transcriptase and the second strand synthesised with RNA's H and DNA polymerase 1 according to Gubler and Hoffman (22). The cDNA was then methylated with EcoR1 methylase and EcoR1 linkers were directly ligated. After EcoR1 digestion the cDNA was selected on a Biogel ACA34 column. Approx. 5O ng was ligated to 1 µg of EcoR1 restricted NM1149 (12°,3hrs, total volume 3 µg) followed by *in vitro* packaging. Recombinant Cl$^-$ phage were selected on E.coli NM514 (*hfl* *hsdr*$^-$). 7O,OOO unamplified clones were screened *in situ* with ^{32}P labelled pY353/B. Filters were washed 3x3O min. in O.1xSSC, O.1%SDS at 68°. Eight positive clones were identified and plaque purified. This would suggest

that the abundance of the transcript in total testis mRNA is in order of 0.01% Insert size analysis showed all inserts to be between 1.1 and 1.3 Kb. The largest (1.3 Kb) was subcloned into pUC9 and designated pYMT2/B.

Nucleotide Sequence Analysis

Subclones of pYMT2/B were generated using a random sequencing strategy. 5 µg of purified self-ligated pYMT2/B insert was sonicated to an average size of 5OObp and ligated into the Sma1 site of M13 mp8 (23).Recombinant clones were then sequenced by the chain termination method (24).

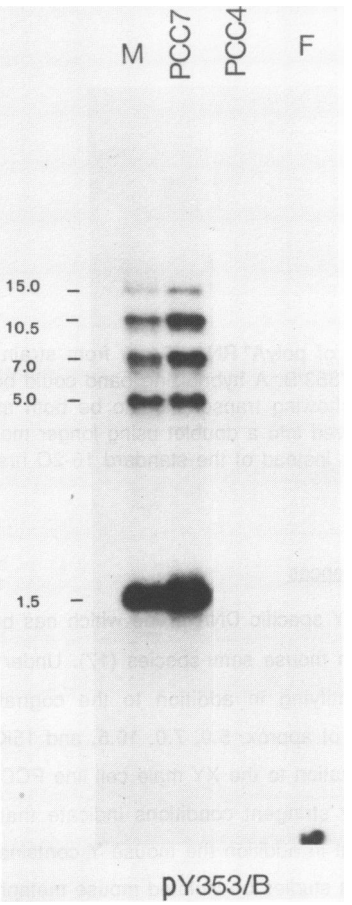

Fig.1 Southern blot analysis of EcoR1 digested DNA's from C57BL/6 male (M), C57BL/6 female (F), male (XY) embryonal carcinomal cell line PCC7, and the male (XO) cell line PCC4. Under stringent conditions pY353/B reacts strongly with the cognate 1.5 Kb band and also detects four homologous hybridizing bands of approx. 5.O, 7.O, 10.5 and 15.0 Kb. The probe does not react with female DNA or the male line PCC4 which has lost the Y in culture.

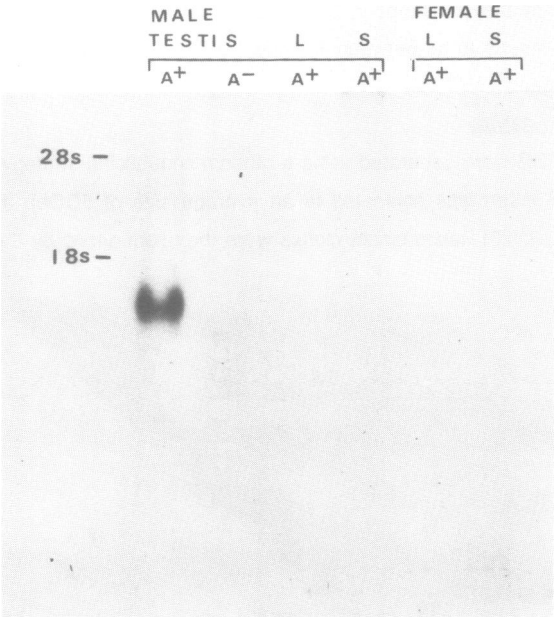

Fig.2 : Northern blot analysis of polyA$^+$RNA (5 µg) from strain 129/sv adult testis, liver (L) and spleen (S) using probe pY353/B. A hybridizing band could be detected only in the polyA$^+$ testis RNA (approx. 1.3 Kb) showing transcription to be both male and tissue specific. This band could frequently be resolved into a doublet using longer more resolutive gels and shorter exposure times (approx.4-5 hrs instead of the standard 16-2O hrs) as can be seen in Fig.5.

RESULTS

Detection of Transcribed Sequences

pY353/B is a 1.5Kb random Y specific DNA probe which has been used in our laboratory to discriminate between European mouse semi-species (17). Under stringent conditions it reacts strongly with male DNA identifying in addition to the cognate 1.5Kb EcoR1 band, four homologous hybridizing bands of approx. 5.0, 7.0, 10.5, and 15Kb (Fig.1). Its Y location can be demonstrated by its hybridization to the XY male cell line PCC7 but not to the X0 male line PCC4.. Titration studies under stringent conditions indicate that the cognate 1.5Kb band is present in about 3O copies, but in addition the mouse Y contains about 25O copies of related sequences. In situ hybridization studies on isolated mouse metaphase chromosomes show that the probe hybridizes along the entire length of the Y (Mattei M.G. personal communication).

When pY353/B was used to probe polyA$^+$RNA blots of male and female liver, male and female spleen and male testis from adult (6 week) 129/sv mice, a hybridizing band of approx. 1.3 Kb could be detected in the testis but not in the male or female liver or spleen (Fig.2).

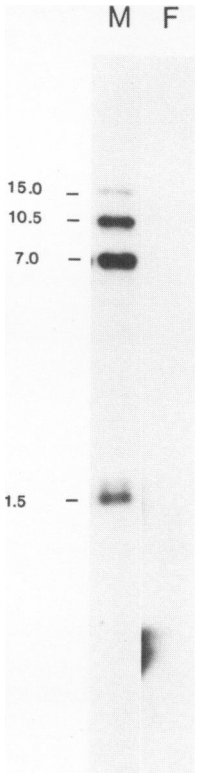

Fig.3 : Southern blot analysis of EcoR1 digested male (M) and female (F) C57/BL6 DNA using probe pYMT2/B (stringent conditions). The probe shows the same reaction pattern and Y specificity as the genomic sequence but no longer detects the 5.O Kb band.

Using longer more resolutive RNA blots and shorter exposure times this band could frequently be resolved into a doublet (see Fig 5). These data suggest that the probe detects several Y chromosome encoded transcripts which are expressed in a tissue specific fashion. All blots were routinely reprobed with a cloned mouse B_2 microglobulin probe to ensure the integrity of the transferred RNA's.

cDNA Analysis

When the cDNA pYMT2/B was used to probe EcoR1 digested male and female genomic blots under stringent conditions (Fig.3) it hybridized to the 1.5, 7.0, 10.5 and 15.0Kb bands recognised by the genomic probe pY353/B but not to the 5.0Kb band. In addition the intensity of hybridization to the 1.5Kb band was much weaker. The cDNA did not react with female DNA under these conditions but under non-stringent washing conditions (3xSSC, 68°) and prolonged exposure times (7-10 days) a band at 3.0Kb could be detected on female DNA. Preliminary mapping data using somatic cell hybrids has localized this band to either chromosome X or 16 (data not shown)

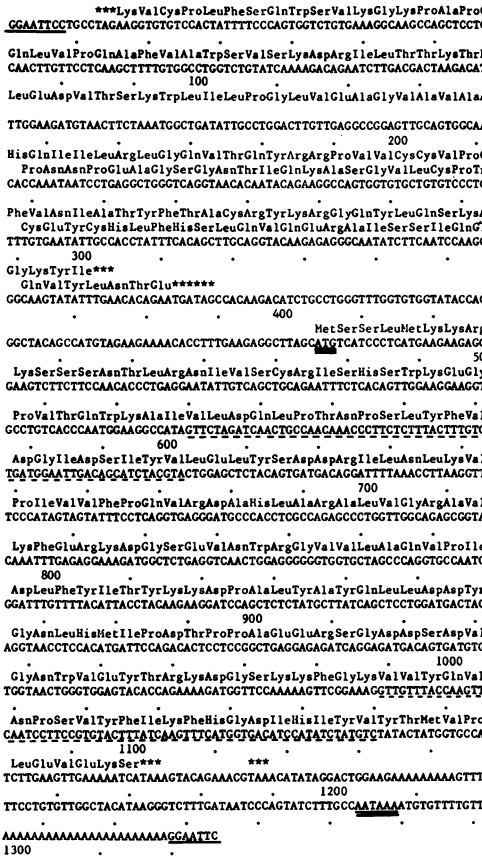

Fig. 4 : Complete nucleotide sequence of pYMT2/B. EcoR1 linkers used in cloning are underlined, consensus polyadenylation signal AATAAA at position 1274 is double underlined. The postulated protein coding sequence starts with ATG at position 476 (heavy underlined) and finishes at position 1171 after which it is followed by two stop codons (***). The corresponding amino-acid sequence is shown above. Amino-acids encoded by nucleotides 602-673 and 1061-1131 (broken underlined) are highly homologous showing 50% identity and 75% homology. The two open reading frames in the 5' end of the sequence are indicated by their corresponding amino-acid sequences given above. The reading frames of these two sequences and that of the proposed protein are all different.

Sequence analysis of pYMT2/B (Fig.4) showed it to be 1320bp long carrying a 26bp polyadenylated 3' tail (EcoR1 linkers used in cloning are underlined). A consensus polyadenylation signal sequence can be identified at position 1274 (double underlined). A significant potential coding region can be seen starting with an initiator ATG at position 476 (heavy underlined) and ending at position 1171 after which it is followed by two stop codons.

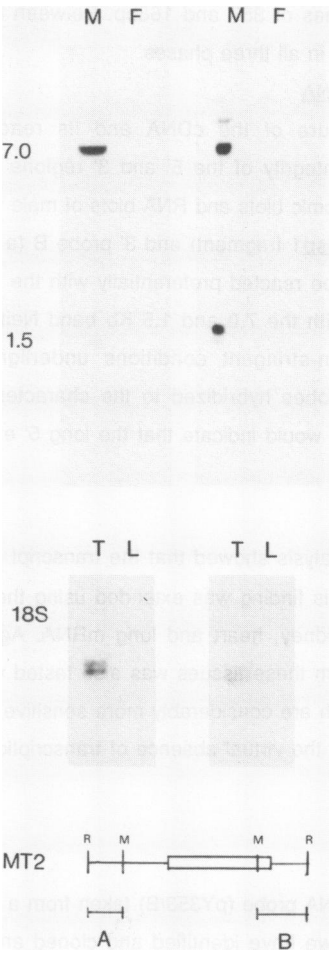

Fig.5 : (Above) Southern blot analysis (stringent conditions) of EcoR1 digested male (M) and female (F) DNA with 5' and 3' derived cDNA probes isolated as shown.
(Below) Northern blot analysis of male C57BL/6 testis (7) and liver (L) poltA$^+$ RNA with the same 5' and 3' probes.

This phase comprises 696bp of DNA which if translated would yield a protein of approx. 26Kd. Hydrophobicity analysis of the predicted amino acid sequence shows that the protein would lack a 5' hydrophobic leader, would be essentially hydrophilic, and in all probability would represent an internalized product not a secreted or cell surface protein.

An intriguing feature of this cDNA is its long 457bp sequence 5' to the proposed initiator ATG. This 5' sequence contains 5 ATG codons upstream of the proposed initiator ATG and two

overlapping open reading frames of 357 and 168bp. Between these two open reading frames there are multiple stop codons in all three phases.

Structural Analysis of the cDNA

Due to the unusual structure of the cDNA and its reaction with female DNA under non-stringent conditions the integrity of the 5' and 3' régions was investigated. Both EcoR1 digested male and female genomic blots and RNA blots of male testis and liver were probed with 5' probe A (a 194bp EcoR1-Msp1 fragment) and 3' probe B (a 353bp Msp1-EcoR1 fragment) as shown in Fig.5. The 5' probe reacted preferentially with the 7.0Kb genomic band and the 3' probe reacted preferentially with the 7.0 and 1.5 Kb band Neither probe hybridized to female DNA under stringent or non-stringent conditions underligning the Y specificity of the transcript. In addition both probes hybridized to the characteristic doublet RNA band in the testis but not in the liver. This would indicate that the long 5' end is not the result of a cloning artifact.

Tissue Specificity

Preliminary Northern blot analysis showed that the transcript could be detected in the testis but not the liver or spleen. This finding was extended using the potential coding portion of the cDNA to probe male brain, kidney, heart and lung mRNA. Again transcription could only be detected in the testis RNA from these tissues was also tested with an antisense RNA probe in RNAse protection assays which are considerably more sensitive. Probe protection could only be seen in testis RNA suggesting the virtual absence of transcription in the other tissues (data not shown).

DISCUSSION

Using a Y derived random DNA probe (pY353/B) taken from a highly enriched mouse Y chromosome genomic library we have identified and cloned an mRNA transcript expressed in the adult mouse testis. This transcript could not be detected in mRNA extracted from either male liver, spleen, brain, kidney, heart or lung. Showing it to be expressed in a tissue specific manner. These data provide evidence that the mouse Y chromosome is transcriptionally active in the male reproductive organ.

Probing the DNA of XXSxr (sex-reversed) male mice with either the genomic or cDNA clone gave negative results indicating that the transcript is not located within this critically important "Sxr" region of the Y chromosome (data not shown).

The function of the transcript is completely unknown. An analysis of the cDNA sequence itself shows it to carry a 26bp poladenylated tail preceded 20bp upstream by a characteristic consensus polyadenylation signal (AATAAA). Four 50bp regions of the sequence were chosen as

nucleotide search probes (one 5', two central and one 3') and used to search the Los Alamos nucleotide data Bank (the regions were positions 150-200, 667-717, 870-920, 1180-1230). No significant homologies were found at the 65% level. In addition the entire predicted amino-acid sequence (ie between nucleotides 476-1171) was split into 8x29 residue peptides which were then used to search the NFBP peptide data bank. Again no significant homologies were found to known peptides at the 25% level. These searches show that the RNA is not related to known retroviral sequence and represents a novel mammalian mRNA encoded by the Y chromosome. Within the potential coding region, the predicted amino-acid sequence coded by nucleotides 602-673 and 1061-1131 are highly related showing 50% identity and 75% homology. This could be due to an internal duplication event, the maintenance of the amino-acid homology serving some functional significance. A hydrophobicity analysis of the predicted protein sequence shows the amino terminus to be rather hydrophilic lacking any hydrophobic leader sequence indicating an internalised protein rather than a cell surface or secreted product. The sequence itself shows no evidence of the simple quaternary repeats GATA or their derivatives reported to identify male specific transcripts in the liver. An unusual feature of the cDNA is its very long sequence upstream of the proposed initiator ATG at position 476. It is highly unlikely that this represents a cloning artifact for the 5' EcoR1-Msp1 probe A remains male specific on genomic blots reacting preferentially with the 7Kb band. Further, on Northern blots it shows the characteristic testis doublet band and does not react with male liver RNA (Fig.5). Preliminary RNAse protection mapping shows that the entire 5' region from EcoR1 to BamHI (890bp) can be totally protected in either total testis RNA, total polyA$^+$RNA or, importantly cytoplasmic RNA. This argues against the possibility that the cDNA represents an unspliced or partially processed nuclear intermediate. These RNA studies also clearly indicate that multiple bands can be protected, suggesting that the transcript we have cloned forms part of a heterogeneous family of related transcripts which are expressed in a testis specific manner. Such an interpretation is consistent with the repeated nature of the probe itself and the doublet band seen on RNA blots. We are at present sequencing regions of the other isolated cDNA clones to define the regions of variability.

This family of transcripts may have originated from a sequence that has inserted in the Y and evolved by duplication events to it present repeated nature. The probable duplication detected in the cDNA, the heterogeneous nature of the transcripts and the detection of a weakly hybridizing band on the X or chromosome 16 (which could represent the ancestral sequence) would support this interpretation. Interestingly, the tissue specific expression suggests that they must be associated on the Y with testis specific promotor/enhancer elements.

Preliminary data, using antibodies raised against the proposed protein expressed as to lac Z

fusion protein in E.coli. have as yet failed to identify any protein product in vivo . It is possible that such antibodies do not recognize the protein in its native form. Alternatively, the transcript may be translated efficiently only at a specific stage of development. The presence of five upstream ATG triplets in the cDNA would certainly place severe limits on its translation and may explain why we have only obtained equivocal results with hybrid select-in vitro translation assays (unpublished observations). In this respect Kahane and Nathans have shown that the ornithine decarboxylase cDNA has a 737bp leader sequence containing 4 upstream ATG codons. They speculate that masking/unmasking of these start sites may regulate the expression of the enzyme.

At present we are trying to pinpoint in which cell type the transcript is expressed and at what time during embryogenesis.

ACKNOWLEDGEMENTS

We gratefully acknowledge that continued interest and support of Dr M. Fellous throughout this work. We would also like to thank Drs J. Weissenbach, I.Garner, G. Langsley, S.Wain-Hobson, P. Baldacci, P.Avner, O.Pujalon, C. Rijnders and C.Roberts for advice and stimulating discussion.

REFERENCES

1 - Jacobs P.A. and Strong J.A. (1959) - Nature 183,302-303.
2 - Russell L.B. and Chu E.H. (1961) - Proc. Natl. Acad. Sci (USA) 47, 571-575.
3 - Davis R.M. (1982) - J. Med. Genet. 18, 161-195.
4 Eicher E.M., Phillips S.J. and Washburn L.L. (1982) - in : Prospects for sexing mammalian sperm (Eds Amann R.P. and Seidel G) Colorado University Press Boulder.
5 - Burgoyne P.S., Levy E.R. and McLaren A.(1986) - Nature 320,170-172.
6 - Goodfellow P., Banting G., Sheer D., Ropers H.H., Caine A., Ferguson-Smith M.A. Povey S. and Voss R. (1983) Nature 302, 346-348.
7 - Singh L. and Jones K.W. (1982). Cell 28 205-216.
8 - Cattanach B.M., Pollard C.E. and Hawkes S.G. (1971) - Cytogenetics 10, 318-337.
9 - Phillips S.J., Nirkenmeir E.H., Callahan R. and Eicher E.M. (1982). Nature 297, 241-243.
10- Keitges E., Rivest M., Siniscalco M. and Gortler S.M. (1985) - Nature 315, 226-227.
11- Burgoyne P.S. (1982) - Human Genet. 6, 85-90
12- Evans E.P., Burtenshow M.D. and Cattenach (1982) - Nature 300, 443-445
13- Cooke H.J., Brown W.R. and Rappold G.A. (1985) - Nature 317, 687-691
14- Rouyer F., Simmler M.C., Johnson C., Vergnaud G., Cooke H.J. and Weissenbach J. (1986)-Nature 319, 291-295
15- Bishop C.E., Guellaen G., Geldwerth D., Voss R., Fellous M. and Weissenbach J. (1983)- Nature,303, 831-832
16- Bishop C.E., Guellaen G., Geldwerth D., Fellous M. and Weissenbach J.(1984)- J. Mol. Biol. 173, 403-417
17- Bishop C.E., Boursot P., Baron B., Bonhomme P. and Hatat D. (1985) - Nature 315, 70-72

18- Wolfe J., Erickson R.P., Rigby P.W.J. and Goodfellow P. (1984) - EMBO J. 3 1997-2003.
19 -Lamar E.E. and Palmer E. (1984) - Cell 37, 171-177.
20- Southern E. (1975). J. Mol. Biol. 98 503-517
21- Auffray C. and Rougeon F. (1980) - Eur. J. Biochem. 107, 303-314
22- Gubler U. and Hoffman B.J. (1983) - Gene 25, 263-269
23- Messing J. and Vieira J. (1982) - Gene 19, 269-276
24- Sanger F., Nicklen S. and Coulson A.R. (1977)-Proc. Natl. Acad. Sci (USA) 74, 5463-5467
25- Burnette W.N.(1981) - Analytic Biochem. 112 , 195-203