

Expression of *Sry*, the mouse sex determining gene

Adam Hacker¹, Blanche Capel^{1,*}, Peter Goodfellow² and Robin Lovell-Badge^{1,†}

¹Laboratory of Developmental Genetics, Medical Research Council, National Institute for Medical Research, London NW7 1AA, UK

²Department of Genetics, University of Cambridge, Cambridge CB2 3EH, UK

*Present address: Department of Cell Biology, Box 3709, Duke University Medical Center, Durham, North Carolina 27710, USA

†Author for correspondence

SUMMARY

In the mouse, *Sry* is expressed by germ cells in the adult testis and by somatic cells in the genital ridge. Transcripts in the former exist as circular RNA molecules of 1.23 kb, which are unlikely to be efficiently translated. We have used RNase protection to map the extent of the less abundant *Sry* transcript in the developing gonad. We demonstrate that it is a linear mRNA derived from a single exon. This begins in the unique region 5' of the protein coding region and extends several kilobases into the 3' arm of the large inverted repeat which bounds the *Sry* genomic locus. Knowledge of this transcript, which is very different from that of the human *SRY* gene, allows us to predict its protein product and reveals several features which may be involved in translational control. Our data is also consistent with there being two promoters for the *Sry* gene, a proximal one that gives functional transcripts in the genital ridge and a distal promoter used in germ cells in the adult

testis. As RNase protection is a quantitative technique, a detailed timecourse of *Sry* expression was carried out using accurately staged samples. *Sry* transcripts are first detectable just after 10.5 days post coitum, they reach a peak at 11.5 days and then decline sharply so that none are detected 24 hours later. This was compared with anti-Müllerian hormone gene expression, an early marker of Sertoli cells and the first known downstream gene of *Sry*. *Amh* expression begins 20 hours after the onset of *Sry* expression at a time when *Sry* transcripts are at their peak. While this result does not prove a direct interaction between the two genes, it defines the critical period during which *Sry* must act to initiate Sertoli cell differentiation.

Key words: sex determination, *Sry*, genital ridge, transcription, RNase protection, mouse

INTRODUCTION

Testis determination in mammals occurs in the gonadal anlage due to the action of the Y-linked gene *Sry*. *Sry* encodes a protein with an HMG box DNA binding domain and is believed to act within cells of the supporting cell lineage of the indifferent gonad to trigger a cascade of events resulting in the differentiation of these cells into Sertoli rather than follicle cells. The differentiating Sertoli cells then trigger other cell lineages within the gonad to follow the testicular pathway.

In terms of *SRY* function most of what is known concerns the HMG box. In vitro experiments have demonstrated that the HMG box has the ability to bind to specific target sequences, preferentially to AACAAAT, and to induce a bend of about 90° in the DNA (Nasrin et al., 1991; Ferrari et al., 1992; Harley et al., 1992; King and Weiss, 1993; Natesan and Gilman, 1993). In approximately 20% of humans with gonadal dysgenesis, mutations have been identified within the HMG box (for review see Hawkins, 1993; Poulat et al., 1994). Some of these mutations have been found to disrupt the interaction of *SRY* with DNA, affecting binding affinity or the angle of bend induced (Nasrin et al., 1991; Harley et al., 1992; Pontiggia et

al., 1994; Poulat et al., 1994). These findings are consistent with *SRY* having a role as the critical signal in the sex-determining pathway through an effect on transcription of one or more target genes, perhaps by organising local chromatin structure and facilitating other protein DNA interactions (Nasrin et al., 1991; Giese et al., 1992; Harley et al., 1992; Pontiggia et al., 1994). So far, very few mutations have been found in non-box regions of the *SRY* gene. Tajima et al. (1994) describe one mutation in the C-terminal domain and McElreavey et al. (1992) report a large 5' deletion which is presumed to affect regulation. However, of the other 80% of individuals with gonadal dysgenesis, some are likely to have mutations in as yet undefined regions of the *SRY* locus, while others may have mutations in genes elsewhere in the testis determining pathway.

To understand *SRY* in more detail it is necessary to identify other features of the locus and the protein it encodes. However, comparisons of the predicted *SRY* protein sequences from different species has shown that the only region conserved is the HMG box DNA binding domain. Regions N-terminal and C-terminal of this domain are highly divergent even amongst closely related species. Comparison of human and mouse

sequences is especially difficult since there is no region of extended homology apparent outside the HMG box and therefore no other reference points for alignment (Tucker and Lundrigan, 1993; Whitfield et al., 1993). This lack of homology makes it difficult to identify common regulatory regions that may control transcription or translation. It has therefore been necessary to characterize the genes independently.

There have been considerable efforts to identify the human *SRY* transcript (Vilain et al., 1992; Behlke et al., 1993; Clépet et al., 1993; Su and Lau, 1993). Each study is consistent with the human *SRY* transcript comprising a single exon with a single polyadenylation site. However, at the 5' end there is disagreement over the location and number of transcription initiation sites. These differences may reflect variations in the source of RNA used in experiments which include adult testis, a teratocarcinoma cell line NTERA-2 cl.D1 and a mouse fibroblast cell line transfected with a human *SRY*-containing cosmid. In addition, much of the data relied only on polymerase chain reaction (PCR) based techniques. These can be prone to artifacts, such as selection and amplification of specific products, which if they are rare or shortened transcripts will give misleading information. To date, no experiments have described the transcript present in the human developing gonad. Foetal gonad material is difficult to obtain in sufficient quantities and condition necessary for a detailed RNA analysis during the period when *SRY* is believed to be active, about 6 weeks post coitum just prior to testis cord formation.

In the mouse, *Sry* is expressed in the germ cells of the adult testis and the somatic cells of the genital ridge (Koopman et al., 1990). Northern analysis has demonstrated that the testis transcript is a relatively abundant 1.3 kb molecule, but such experiments have been too insensitive to detect *Sry* transcripts in the developing gonad. Characterization of the adult testis transcript using a variety of techniques demonstrated that it is a circular RNA molecule. We proposed (Capel et al., 1993) that such an unusual transcript might arise as a result of the structure of the mouse *Sry* genomic locus which consists of a large inverted repeat (absent in the human locus) extending for at least 15.5 kb 5' and 3' of a 2.8 kb unique region. The latter region contains a long open reading frame which could produce a protein of 395 amino acids including the HMG box DNA binding motif (Gubbay et al., 1990). However, transcription initiation in the 5' arm of the inverted repeat would produce large primary transcripts, able to form a stem loop structure from which the circle is spliced. These circles contain most of the ORF and are localised to the cytoplasm, but translation is unlikely to occur as we could not detect them loaded onto polysomes (Capel et al., 1993). Furthermore, it seems that the gene has no critical cell autonomous function in spermatogenesis, as XY germ cells deleted for *Sry*, within the founder male chimaera carrying the *Tdy*^{ml} mutation, were able to give rise to offspring (Lovell-Badge and Robertson, 1990). These data imply that there must be at least two transcripts, one expressed in the adult testis which is not translated, and one specific to the developing gonad where the gene functions. This would not be very surprising as many genes are expressed in different forms in the adult testis where they have no obvious function (e.g. Meijer et al., 1987; Fischman et al., 1990; Rubin and Nguyen-Huu, 1991; Sargent et al., 1991; Ruppert et al., 1992).

A number of different approaches have been used by us in an attempt to define the extent of the functional transcript expressed in the mouse genital ridge. The screening of 11.5 dpc genital ridge phage and plasmid cDNA libraries failed to identify any positive clones. This may have been partly due to low cloning efficiency and the instability of this particular insert DNA (as was seen for many genomic clones of *Sry*), but may also be due to the low level of *Sry* expression (Koopman et al., 1990; Capel et al., 1993). Attempts to over-express *Sry* from genomic DNA in tissue culture systems have either failed to produce any *Sry* transcripts or have produced them in the circular adult testis form (Capel et al., 1993). Reverse-transcription-PCR (RT-PCR) primer walking was also attempted along the *Sry* locus but, due to the presence of repeat sequences and microsatellites, results were highly variable and too unreliable to draw any conclusions. Similarly, because 5' and 3' rapid amplification of cDNA ends, RACE-PCR work efficiently only when the specific primers are close to the transcript ends we obtained many spurious results (unpublished observations). Primer extension at a number of sites 5' of the HMG box failed to yield any signal above background.

The low level of *Sry* expression and the spurious results obtained from other techniques has forced us to adopt the rigorous technique of RNase protection to analyze 11.5 dpc genital ridge RNA and map the extent of the *Sry* transcript. RACE-PCR could then be used to confirm the precise locations of transcription initiation and polyadenylation sites. The *Sry* transcript which we have mapped in the genital ridge is different from the circular transcript in the adult testis. The vast majority of transcripts show no splicing within the long ORF confirming assumptions made by us and others on the likely nature of the protein (Tucker and Lundrigan, 1993; Whitfield et al., 1993; Coward et al., 1994; Dubin and Ostrer, 1994). We are also able to identify several features of the *Sry* sequence that may be involved in transcriptional and translational regulation. Our results are consistent with the notion that a more distal promoter is used in the adult testis. Furthermore, RNase protection was used as a quantitative assay of *Sry* expression using samples dissected from accurately staged embryos, and it has been possible to compare the timecourse of *Sry* expression with that of the anti-Müllerian hormone gene, *Amh*, an early marker of Sertoli cells.

MATERIALS AND METHODS

Animals and tissues

Samples were dissected from Parkes outbred mice unless otherwise stated. Experiments involving the p741 transgene (Koopman et al., 1991) used line 32.10 which contains approx. 12 copies of the 14.6 kb mouse *Sry* genomic fragment. This line was maintained as a backcross to MF1 outbred mice where sex-reversal is not seen. 32.10 females were then crossed either to CBA males to look for the transgene expression in the genital ridge or to C57BL6/Y^{AKR} males to look for expression in adult testes. For all embryo dissections it was assumed that matings took place mid-way through the dark period; therefore midday on the day of appearance of the vaginal plug is approximately 0.5 dpc. For more accurate staging, e.g. Fig. 5, urogenital ridges were dissected from embryos after counting the number of tail somites posterior to the hind limb bud. Under this scheme, an embryo with 8 tail somites is approximately 10.5 dpc, 18 somites is 11.5 dpc and 30 somites is approximately 12.5 dpc.

Genotypic sex was ascertained by staining for sex chromatin in amniotic cell nuclei as described by Palmer and Burgoyne (1991). For the experiment involving transgenic mice, embryos were genotyped from crude tail lysate by multi-gene PCR. The primers used were specific for *Zfy-1* (5'-CCTAT TGCAT GGACA GCAGT CTTATG-3' and 5'-GACTA GACAT GTCTT AACAT CTGTCC-3') to identify Y chromosome carriers and for the p741 *Sry* transgene (5'-GAGGG CATGG TCAGT TGAAC-3' and 5'-CTCAG TGTGG AATTC ATCTG C-3'), the latter pair making use of the head-to-tail concatamer arrangement of the 14.6 kb sequences in this line. PCR conditions were as in Koopman et al. (1991), except that reactions were carried out in 25 μ l final volumes. Primers for myogenin were included in each reaction as a positive control for the quality of the lysate generated from each embryo.

RNA extraction

RNA from post partum tissue samples was prepared by the guanidinium thiocyanate-caesium chloride method (Chirgwin et al., 1979). For embryonic samples we used the method described by Sargent et al. (1986) except that the first phenol/chloroform extraction was carried out by mixing the homogenate with 0.5 volumes of water-saturated phenol, and then adding 0.5 volumes of chloroform before centrifugation at 4°C. RNA samples were stored at -70°C in isopropanol.

RNase protection probes

The probes used were as follows: probe 2, *NheI* (at base pair 6937 in the p741 *Sry* genomic sequence) - *XbaI* (bp 8298) genomic fragment subcloned into the *XbaI* site of pTZ18R (Pharmacia), the phagemid was linearised with *BspHI* (bp 7927) and transcribed with T7 RNA polymerase; probe 4, *BglIII* (bp 8211) - *PstI* (bp 8575) fragment subcloned into *PstI/BamHI* sites of pBluescriptKS (Stratagene), linearised with *NotI* in the polylinker and transcribed with T3 RNA polymerase; probe 8, *EcoRV* (bp 8714) - *HindIII* (bp 9666) fragment, subcloned into *EcoRV/HindIII* sites of pBluescriptSK (Stratagene), linearised with *SfaNI* (bp 9150) and transcribed with T7 RNA polymerase; probe 16, *AflIII* (bp 11317) - *AvaII* (bp 11694) fragment, subcloned blunt into *EcoRV* site of pBluescriptSK, linearised with *HindIII* in polylinker and transcribed with T3 RNA polymerase; probe 21, *MspI* (bp 10154) - *BglIII* (bp 13152) fragment, subcloned blunt into *EcoRV* site of pBluescriptSK, linearised with *HindIII* (bp 12791) and transcribed with T3 RNA polymerase; probe 19 (sense), *FokI* (bp12185) - *XbaI* (bp 12629), subcloned blunt into *EcoRV* site of pBluescriptSK, linearised with *HindIII* in the polylinker (so that a sense riboprobe is transcribed when T3 RNA polymerase is used); *Amh* probe, 315 bp *PstI SacI Amh* fragment (over the transcription start site) was sub-cloned into pBluescript KS, linearised with *HindIII* (in the polylinker) and transcribed with T7 RNA polymerase. Transcription was carried out in the presence of 800 Ci/mmol [³²P]UTP (Amersham) as described in Krieg and Melton (1987). Riboprobes were electrophoresed on a 6% polyacrylamide, 8 M urea gel, full length probe was excised and eluted from gel in a solution of 0.5 M ammonium acetate, 1 mM EDTA, 0.1% SDS; before being precipitated in ethanol (details of other probes shown in Fig. 1B are available on request).

RNase protection

Between 2 μ g and 5 μ g of total RNA was hybridized for 8-18 hours with 4 \times 10⁵ cpm of RNA probe in 20 μ l of 80% formamide, 40 mM Pipes (pH 6.7), 0.4 M NaCl, 1 mM EDTA and then digested with 400 μ l of 0.5 M NaCl, 10 mM Tris pH 7.5, 1 mM EDTA and 100 U RNase T1 and 2.5 μ g RNase A per ml at 37°C for 30 minutes. After proteinase K treatment and phenol-chloroform extraction, samples were ethanol precipitated and electrophoresed on 6% polyacrylamide, 8 M urea gels. Gels were exposed on PhosphorImager screens for a period of two to seven days for quantitative analysis (ImageQuant software) before being autoradiographed for seven days at -70°C with intensifying screens. For critical points in the protection walk experiments

were repeated at least ten times, and more than twice elsewhere. The complete timecourse experiments as shown in Fig. 5A,B were repeated twice with independent samples (and a number of partial experiments gave comparable results).

RACE-PCR

3' RACE-PCR

Experiments were undertaken over a number of sites 3' to the *Sry* ORF. RACE-PCR products were obtained at two sites. The major polyadenylation site (at bp 12970) was located as follows. Genital ridge RNA was reverse transcribed using the dT17-R1-R0 adaptor primer (Frohman et al., 1988) as described previously (Koopman et al., 1990). Primers R0 and Y54A (from bp 12768; 5'-CCCTG GGATA CTGTT CTTCT GG-3') were incubated for 7 minutes at 95°C, Taq polymerase was added and second strand cDNA synthesis carried out at 75°C for 5 minutes, 50°C for 2 minutes and 72°C for 15 seconds, followed by amplification for 30 cycles (91°C 45 seconds, 50°C 45 seconds, 72°C 1 minutes). Southern analysis of the PCR products showed a discrete band. PCR products were digested with *HindIII*, which cuts the *Sry* sequence immediately downstream of the primer Y54A, and *Clal* (which cuts in the adaptor primer) and cloned into *HindIII/Clal* pBluescript SK (Stratagene). One in forty clones was positive after screening by hybridization with *HindIII* (bp 12790)/*BglIII* (bp13149) probe. Two of the positive clones were sequenced and found to contain identical *Sry* inserts.

The minor 3' RACE-PCR product (at bp 11575/6) was obtained using a similar method as above but with two rounds of PCR. The first round of PCR was undertaken with primers R0 and Y53A (from bp 11395; 5'-CAGTT AAAAG ACTGT TCTCT AG-3'), and the second round with R1 and Y48A (from bp 11421; 5'-CAATG GAACA GTTAG CTGGC A-3'). PCR products were ligated into *HincII* digested pBluescript SK and 20 colonies were miniprep and digested, and two were sequenced.

5' RACE-PCR

Genital ridge RNA was analyzed as described in Capel et al. (1993). The first round of PCR was carried out using R0 and BY1A (from bp 8236; 5'-TGTTAG GGCTG AACAC TA-3') for 30 cycles, followed by a second round using R1 and BY1A. PCR products were digested with *BglIII* (this cuts in the *Sry* sequence upstream of the primer BY1A) and *Clal* (which cuts in the adaptor primer) and cloned into *BglIII/Clal* digested pBluescript SK. 36 clones were screened for hybridization with labelled oligonucleotide Y15A (from bp 8068; 5'-AAGTT TTGAC TTCTG TATCT GTCTG-3') which lies upstream of BY1A.

RESULTS

Analysis of the *Sry* transcript in the developing gonad

RNase protection experiments were initially carried out using probe 4 (Fig. 1A) which spans the HMG box region. A protection walk was then carried out moving in both directions using overlapping probes. A subset of the probes used is shown in Fig. 1A. In all experiments the RNAs used were from pooled 11.5 dpc genital ridges; adult testis (>6 weeks); SN18, a stably transfected cell line known to give the same *Sry* transcript as adult testis (Capel et al., 1993); and liver, as a negative control. Position within the *Sry* locus is according to the numbering of Genbank accession number, X67204.

5' analysis of *Sry*

A protected fragment of 97 nucleotides (nt) seen with probe 2 reveals the use of the splice acceptor site (see Capel et al.,

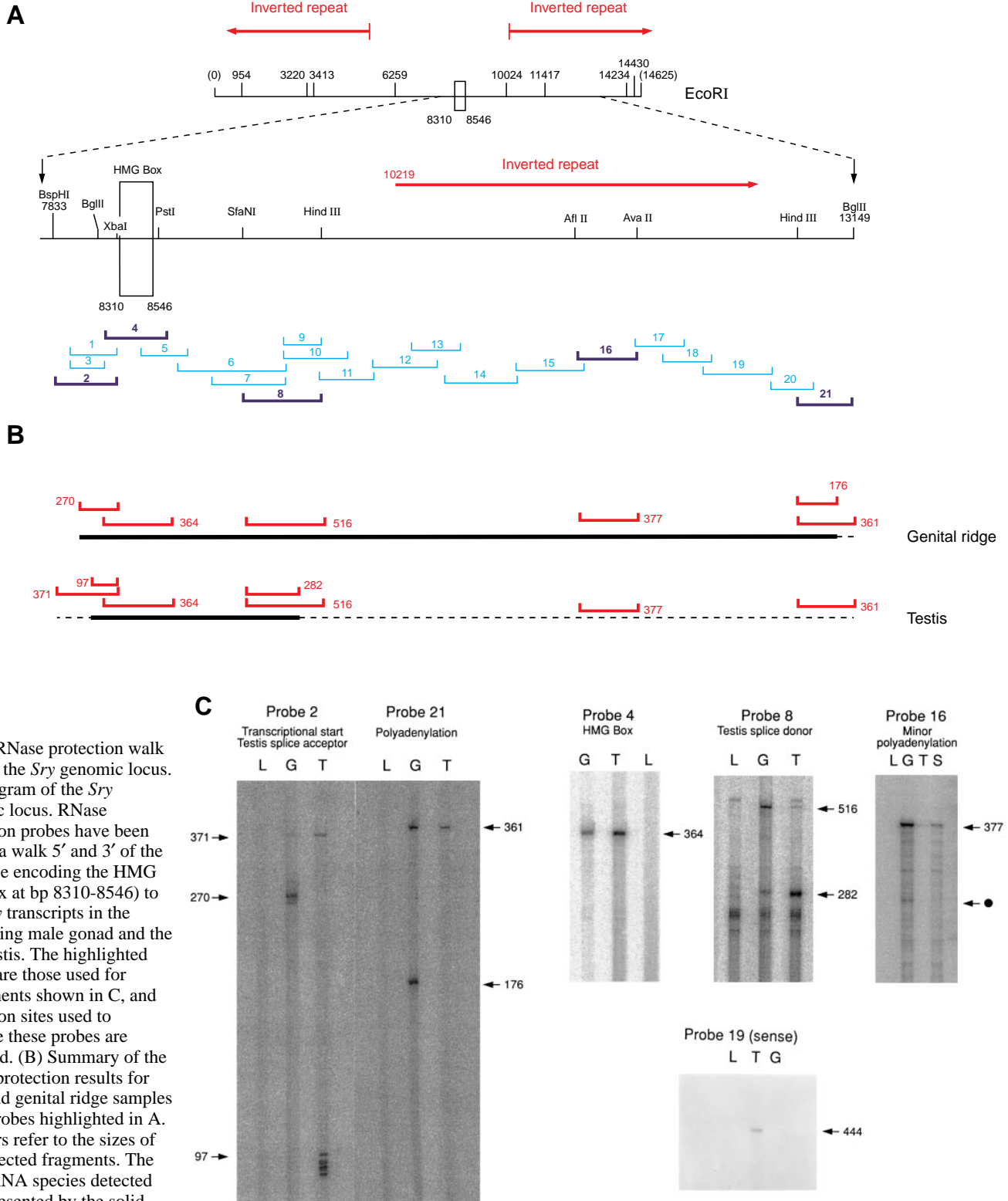


Fig. 1. RNase protection walk through the *Sry* genomic locus. (A) Diagram of the *Sry* genomic locus. RNase protection probes have been used in a walk 5' and 3' of the sequence encoding the HMG box (box at bp 8310-8546) to map *Sry* transcripts in the developing male gonad and the adult testis. The highlighted probes are those used for experiments shown in C, and restriction sites used to generate these probes are indicated. (B) Summary of the RNase protection results for testis and genital ridge samples using probes highlighted in A. Numbers refer to the sizes of the protected fragments. The major RNA species detected are represented by the solid line and minor species (<10%) by the dashed line. (C) RNase protection using probes highlighted in A. Probe 4 protects the region encoding the HMG box; probes 2 and 21 highlight the multiple transcriptional start sites and the polyadenylation site of the genital ridge transcript; probe 2 and probe 8 mark the splice acceptor and splice donor of the testis transcript; probe 16 gives a partially protected fragment (●) representing the minor polyadenylation site of the genital ridge transcript (as identified by RACE-PCR). The multiple bands seen with probe 2 using testis RNA are due to 'breathing' in the RNase digestion, other probes give a single band. Total RNA was hybridized; 5 µg of testis (T), liver (L); 2 µg SN18 (S) or 5 pairs of pooled unsexed 11.5 dpc genital ridges (G). In addition an RNase protection using probe 19 (sense) located within the 5' arm of the inverted repeat is shown. Total RNA was hybridized; 5 µg of testis (T), liver (L) and 5 pairs of 11.5 dpc XY genital ridges (G).

1993) at position bp 8201 in adult testis RNA (Fig. 1C). Using genital ridge RNA, no 97 nt fragment was observed with this probe, indicating that there is no splicing at this site. However, three major and one minor protected fragments of 259-272 nt were seen, which we believe correspond to multiple transcription start sites. Consistent results have been obtained by repeating this protection at least ten times. In addition, three different probes of varying sizes over this region have revealed the use of the same sites (data not shown), which map between bp 8027-8039. None of these sites are adjacent to an AG characteristic of splice acceptors, so transcription from these sites could be due to two upstream AT rich regions which are possibly degenerate TATA boxes (see Fig. 2A). There may also be some transcription initiation further 5', as shown by the faint band at 371 nt corresponding to complete protection of the *Sry* sequences in the probe. However, this is a very minor component of the total transcripts (less than 5%). The use of the sense probe 19, which due to the inverted repeat is identical to using an antisense probe over the region bp 5046-5490, demonstrates that no genital ridge transcripts extend into the 5' arm of the inverted repeat (Fig. 1C) unlike in the adult testis (see below).

In order to verify further the RNase protection data, 5' RACE-PCR was carried out using primers just 5' of the HMG box. A broad band was detected at the size expected from the RNase protection data. The band was excised, ligated into a *Bgl*III/*Cla*I digested vector and a number of clones were obtained and sequenced. The majority (7 out of 11 clones) demonstrate

transcription initiation between bp 8035 and 8039. This is in approximate agreement with the protection data. Although the protection identified two additional neighbouring sites it is possible that the PCR amplification does not favour these.

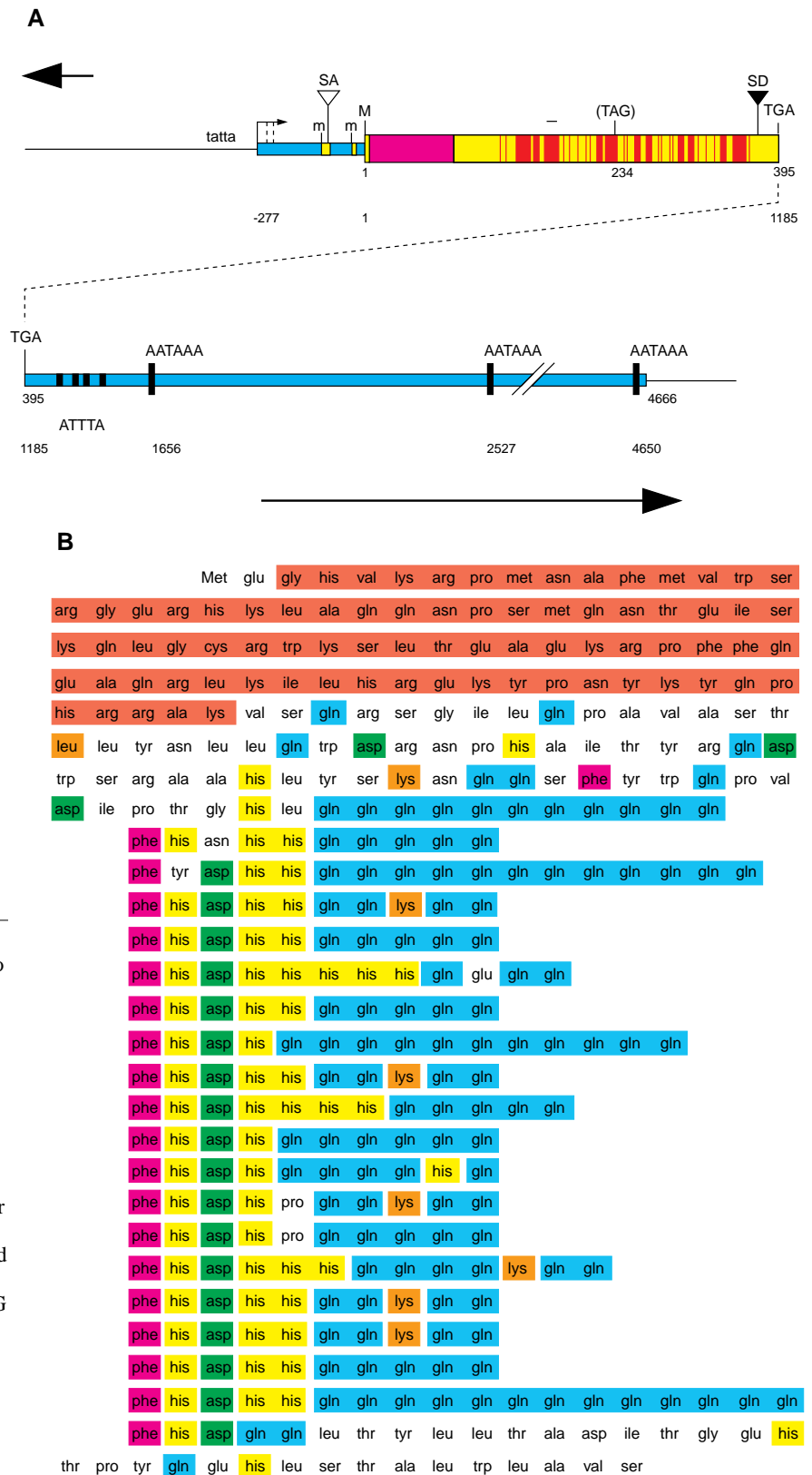


Fig. 2. (A) Features of the mouse *Sry* locus. The nucleotide 1 (and also amino acid 1) corresponds to bp 8303 within the 14.625 kb genomic clone. The long open reading frame of the gene is shown as a large box with the initiator methionine (M); red bands mark glutamine repeats; the HMG box is highlighted in pink; the 5' and 3' untranslated regions are shown as thin blue boxes. Features of interest are shown: the TATTA degenerate promoter element, three transcription initiation points, two short open reading frames (m); the SA and SD refer to the splice acceptor and splice donor sites used by the transcript expressed in the adult testis; transcript instability sequences (ATTTA) and polyadenylation signals (AATAAA). The inverted repeat is shown by arrows. Also marked is the TAG stop codon identified in *domesticus* type *Sry* sequences and the position of the variable number of glutamines (-). (B) The *Sry* open reading frame. The diagram shows the *Sry* open reading frame emphasizing the nature of the large glutamine/histidine repeat region at the carboxy terminal. The HMG box is highlighted in red.

Four shorter clones were obtained that begin further 3', two at bp 8107 and two at bp 8099. These may correspond to rare transcription initiation sites not detected by the protection assay or, more likely, may arise due to PCR slippage along the very T rich microsatellite. No clones were identified as extending beyond bp 8035, which, in combination with the protection data, implies that there is no additional exon spliced to this site.

3' analysis of *Sry*

3' of the HMG box, probes 5, 6 and 7 show complete protection of *Sry* sequences with both genital ridge and testis samples. As expected, protection of probe 8 with testis RNA gives a major 282 nt fragment (Fig. 1C) which is consistent with use of the splice donor at position bp 9432 (Capel et al., 1993). In the genital ridge a minor proportion (<10%) of the transcripts also appear to splice here. These do not seem to correspond to the circle as shown by RT-PCR (Capel et al., 1993 and data not shown). Attempts to use 3' RACE-PCR to identify sequences spliced to this donor site have consistently failed. However, the vast majority of transcripts extend further 3'. In fact, full protection of *Sry* sequences is observed with probes 9 to 20. Despite the presence of two AATAAA consensus polyadenylation sequences, which are spanned by probes 11 and 14, no major site of splicing or polyadenylation is apparent until probe 21. With this probe a 176 nt protected fragment can be observed. We believe this corresponds to the major polyadenylation site of the genital ridge transcript although approx. 25% of the transcripts extend further.

3' RACE-PCR confirmed the position of the major genital ridge polyadenylation site to be at bp 12969, 16 nt downstream from an AATAAA consensus polyadenylation signal. 3' RACE analysis of regions throughout the 3' end of the genomic locus identified another site at bp 11575/6, but this is clearly a minor polyadenylation site. When the probe covering this region (probe 16) was used in RNase protection it gave a protected fragment of expected size (● on Fig. 1C) at low levels (<10% of total signal). Furthermore, two rounds of PCR amplification were required to recover this RACE product compared with just one for the major polyadenylation site.

In the adult testis, the major RNA species detected is the exon comprising the circular transcript; however, a minor proportion (<5%) of unspliced transcripts were also found. These transcripts are not artefacts of the protection assay as they are resistant to digestion with DNase and the signal is completely absent in assays using adult liver RNA prepared in the same way. The *Sry* sequences in all probes used (e.g. 2, 8, 16, 21 and probe 19s in Fig. 1C) have been completely protected; demonstrating the presence of large transcripts extending from at least bp 5046 in the 5' arm of the inverted repeat to bp 13149 in the 3' arm. We believe that the circular transcripts are spliced from these large (>8103 bp) transcripts. In contrast to the cytoplasmic localisation of the circular transcript, these large possibly primary transcripts have been localised to the nucleus (Capel et al., 1993).

The data presented above are consistent with the majority of the *Sry* transcripts within the genital ridge comprising a single exon beginning 5' of the HMG box at approximately bp 8035 extending through the HMG box DNA binding motif, CAG repeat and into the 3' arm of the inverted repeat to bp 12969.

Analysis of the ORF

There are three ATGs 5' of the region encoding the HMG box (Fig. 2A). The first and second ATGs are followed by short ORFs of 9 and 3 amino acids respectively. The third ATG (at position 8295) is followed by a large ORF of 395 amino acids (Fig. 2B), encoding a protein containing the HMG box DNA binding domain. This binding domain is flanked by a short 2 amino acid N-terminal domain and a large 314 amino acid C-terminal domain which is characterized by a large repeat, rich in glutamines and histidines. In *Mus musculus domesticus Sry* alleles this glutamine repeat is truncated by the presence of a stop codon (Fig. 2A) thereby producing an ORF of 234 amino acids (see Discussion). Downstream of this region the next largest ORF within the *Sry* transcription unit could encode 141 amino acids. However, this ORF, which begins at bp 11475, contains a large microsatellite with a variable TC nucleotide repeat and has a poor translation initiation consensus sequence around the ATG.

Comparisons between mouse and human *SRY*

Regions of homology may denote functional protein domains or positions of common regulatory elements. However, comparison of the human adult testis transcript with the mouse genital ridge transcript reveals a surprising number of differences. Mouse *Sry* has a large 5' UTR of approx. 273 nt followed by an ORF of 395 amino acids and a large 3' UTR of 3481 nt. Human *SRY* has a shorter 5' UTR of approx. 77 nt, an ORF of only 204 amino acids and a short 137 nt 3' UTR. Within the transcription units the only region of homology is the HMG box domains (89% identity). However, two short regions of lower homology have been identified outside the transcripts, both lying 5' to the HMG box. One region of approx. 150 bp in length is located at the human and mouse transcription start sites and shows 62% homology; a second region lies approx. 1520 bp upstream of the human and approx. 1900 bp upstream of the mouse start sites and is 68% homologous over 95 bp (Fig. 3). Examination of the GCG Wisconsin data base has failed to identify any transcription factor binding sites at either location common to both human and mouse.

Sry transcripts in transgenic mice

We have examined expression of *Sry* in mice transgenic for the 14.6 kb genomic *Sry* clone, p741. This transgene is known to be capable of producing XX sex-reversal and expression of the transgene at 11.5 dpc had previously been shown by RT-PCR. However, it was not known whether the transgene contained all elements necessary to elicit the correct pattern of expression or the nature of the transcripts in genital ridge and adult testis.

Because XX *Sry* male mice lack germ cells in the adult testis we have examined expression of the transgene in XY transgenic mice. In order to distinguish transgene expression from endogenous *Sry* expression we have made use of the fact that p741 was derived from strain 129 mice and can be distinguished from the *Sry* locus in the AKR strain by virtue of a number of polymorphisms within the CAG repeat region 3' of the HMG box (our data not shown; Coward et al., 1994). To obtain transgenic animals of the correct genotype, we made use of one of our lines of mice, 32.10, which shows incomplete penetrance, producing XX *Sry* transgenic females or males, depending largely on genetic background (Vivian and R. L.-B., unpublished data). Female transgenics

can be crossed with XY^{AKR} males to produce XY^{AKR}*Sry* transgenic males which have normal testes. By using a probe over the polymorphic region (probe 8) it was possible to distinguish transgene expression from endogenous AKR *Sry* expression in an RNase protection assay. The transgene was expressed in the adult testis using the splice donor, thereby implying that the transcript exists in its circular form (Fig. 4A).

We have also confirmed by RNase protection that expression of *Sry* in the 11.5 dpc developing gonad of XX*Sry* transgenics gives rise to the linear form using the same transcription start sites but at approximately 40% of the level as their XY siblings (Fig. 4B). This low level of expression may be responsible for the incomplete penetrance of the sex-reversal in the 32.10 transgenic line, but given the profile of *Sry* expression as revealed below, many more samples and an accurate time-course need to be analysed.

Quantitative time course of *Sry* and *Amh* expression

Previous experiments examining the expression of *Sry* have used a non-quantitative method of RT-PCR which defined the window of expression loosely. We have used RNase protection assays as a quantitative method and have accurately staged samples so that a fine timecourse of *Sry* expression can be achieved. This timecourse has been carried out using probe 2 (Fig. 1) which can distinguish a spliced (97 nt protected fragment) testis type of transcript from a linear (approx. 270 nt protection) genital ridge transcript.

Sry expression is first observed in its linear form at 11/12 tail somite stage, expression peaks at 18 somites, turning off just after the 27 tail somite stage (these stages are equivalent to about 10.5 dpc, 11.5 dpc and just before 12.5 dpc respectively). *Sry* in the post partum testis gives rise to the circular transcript at 28 dpc, approximately coincident with the first wave of spermatogenesis and development of round spermatids (Fig. 5A).

Anti-Müllerian hormone has been proposed as a direct target for the action of SRY (see for example Haqq et al., 1993, 1994). An *Amh* timecourse was undertaken to define its temporal window of expression precisely in relation to the expression of *Sry*. A probe over the region of the *Amh* transcription start site was used for this analysis (Münsterberg and Lovell-Badge, 1991). We have identified at least three start sites, the major one lies approx. 10 nt upstream of the ATG (similar to that observed in human *AMH*). Full protection of *Amh* sequences in the probe corresponds to a mRNA being transcribed from a neighbouring ubiquitously expressed gene (Dresser et al. unpublished data). We have used this signal as a 'loading control' allowing us to normalise for the quantity of RNA used. Using the same samples as were used for *Sry*, expression of *Amh* is just detectable at the 18 tail somite stage and increases. After 7 dpp, transcript number gradually declines, reaching a low level at 43 dpp (Fig. 5B). Phosphoimager quantitation of the *Sry* timecourse and an *Amh* timecourse normalised with the loading control, has allowed expression to be represented graphically (Fig. 5C). *Amh* expression initiates when *Sry* expression is at its peak but continues to increase after *Sry* transcription is turned off.

DISCUSSION

Methodical use of RNase protection mapping in combination with 5' and 3' RACE-PCR has been used to accurately determine the transcription unit of the *Mus musculus musculus* *Sry* gene within the developing gonad. The majority of transcripts comprise a single exon of 4942 bp, which initiate 283 bp 5' of the HMG box and extend into the 3' arm of the inverted repeat. This confirms at least two predictions. First, there is no splicing within the single long ORF. Secondly, we show that the genital ridge transcript starts within the unique region of the *Sry* genomic locus, whereas the testis transcript initiates somewhere within the 5' arm of the inverted repeat. This suggests that there must be two promoters for mouse *Sry*, a distal one, specific to the germ cells of the adult testis, and a proximal one, specific to the somatic cells of the genital ridge. Large transcripts produced from the distal promoter would lead, via an unusual splicing event, to the formation of a circular transcript which is unlikely to be efficiently translated, whereas transcription initiating from the proximal promoter would lead to a normal linear transcript presumed to give rise to functional SRY protein.

The protein encoded by *Sry* can now be predicted with more certainty, based on this detailed analysis of the transcription unit. 5' of the HMG box within the transcription unit, there are three ATGs at bp 8193, 8242 and 8304. From these ATGs extend ORFs of 9, 3 and 395 amino acids respectively. It is the third of these ATGs that lies in an almost optimum locale, (as defined by mutational analysis, see Kozak, 1986b, 1987a, b; Grünert and Jackson, 1994) from which a large ORF, containing the HMG box, extends. However, the first ATG is flanked by a reasonable translational initiation sequence and must be considered as a site from which some translation might occur.

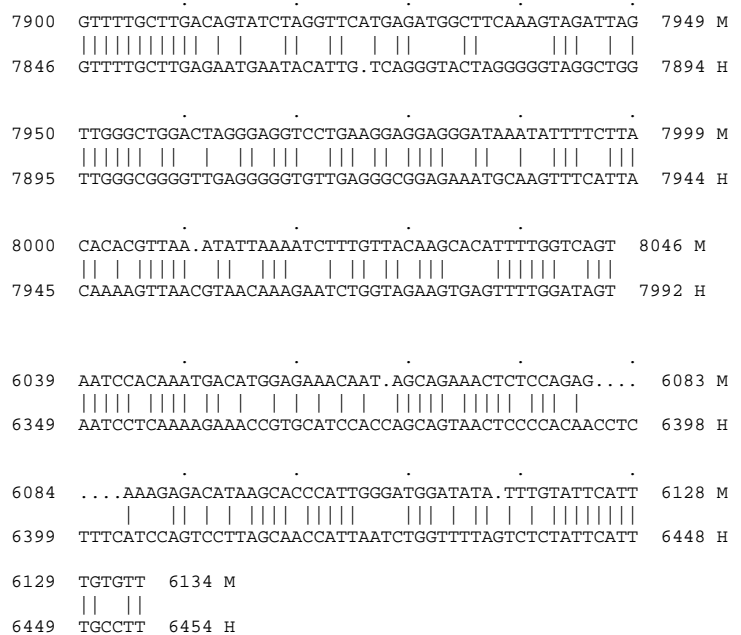


Fig. 3. Regions of homology between mouse and human *Sry*. Two areas of homology between mouse and human *Sry* 5' of the HMG box are shown, the first abuts the transcriptional start sites, the second lies 2-3 kb further upstream.

The short ORF associated with this ATG may, in the event of translation being initiated, allow scanning of the ribosomal 40S unit to resume until the next suitable ATG is reached (Kozak, 1983). A feature of the 5' UTR is the presence of a TCTG microsatellite. This might give rise to secondary structure which could reduce translational efficiency from any of the

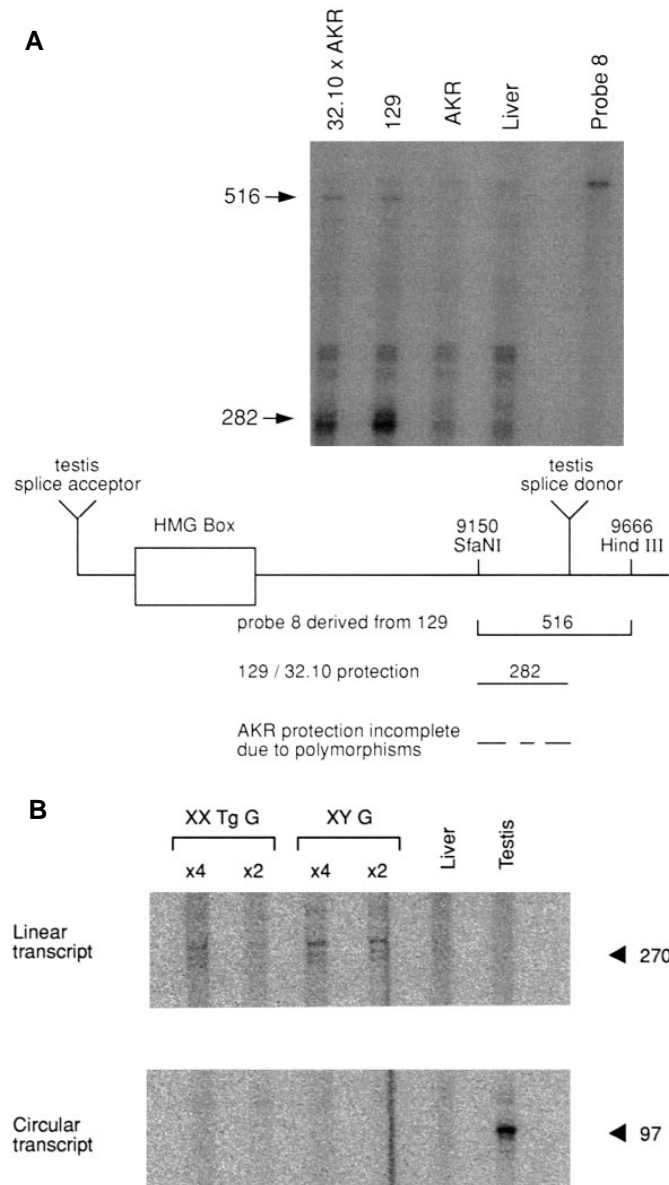


Fig. 4. Expression of *Sry* in mice transgenic for a 14.6 kb genomic fragment. (A) RNase protection assay on total adult testis RNA (5 µg) was carried out using antisense probe 8 (derived from a 129 *Sry* genomic clone). The AKR adult testis RNA shows no protected fragments. The 32.10 x AKR adult testis RNA shows a major protected 282 nt and a minor 516 nt fragment corresponding to the circular and unspliced transcripts. The expression is entirely due to the transgene. (B) RNase protection assay was carried out using antisense probe 2. XX transgenic (Tg) and their XY sibling 11.5 dpc gonads (G) were assayed (the number of genital ridges used in each lane are indicated). A protected fragment of approx. 270 nt demonstrates normal linear *Sry* transcription in the transgenics. The 97 nt fragment representative of the splice donor of the testis transcript is not seen in any of the gonad lanes.

ATGs (Pelletier and Sonenberg, 1985; Kozak, 1986a, 1990). The presence of multiple upstream ATGs and possible secondary structure might act to downregulate translation of an already rare *Sry* transcript. The 5' UTR of human *SRY* as defined in testis, NTERA-2 c1.D1 and transfected cell lines does not show any of these features (Vilain et al., 1992; Behlke et al., 1993; Clépet et al., 1993; Su and Lau, 1993). However, the single ATG lies in a non-optimal context which may also lead to inefficient translation.

The large ORF of mouse *SRY* (395 amino acids) is composed of the 79 amino acid HMG box flanked by a short two amino acid N-terminal domain and a 314 amino acid C-terminal domain characterized by a glutamine/histidine rich repeat region. The human *SRY* ORF is completely different. It is considerably smaller, consisting of only 204 amino acids with the HMG box flanked by a 57 amino acid N-terminal domain and a 68 amino acid C-terminal domain. Furthermore, the glutamine repeat, which appears as the major feature of the C-terminal domain of the mouse *SRY* protein, is completely absent in human *SRY*. Glutamine repeats, such as the one present in mouse *SRY*, have been identified in the context of genes that are associated with a number of diseases including spinal bulbar muscular atrophy, Huntington's disease and Spinocerebellar ataxia type 1 (for brief review see Green, 1993). A direct correlation between the number of glutamines and the onset of disease has been made. This can be either by expansion or by contraction of repeats (Huntington's Disease Collaborative Research Group, 1993; Mhatre et al., 1993; Orr et al., 1993; Schoenberg et al., 1994). The role of glutamine repeats in the function of proteins is yet to be fully understood. Glutamine repeats that occur in transcription factors may act as an activation domain (Gerber et al., 1994); removal or addition of extra glutamines modifies the properties of such a region (Mitchell and Tjian, 1989). As human *SRY* lacks a glutamine repeat its mode of action may be different from the mouse protein. Perhaps the human protein needs to interact with an additional protein that contributes an activation domain. However, the importance of the glutamine rich domain to *SRY* function in the mouse is difficult to understand. Sequence analysis of the ORF of *SRY* in a variety of mouse species (Coward et al., 1994, and our observations) has identified a mutation (marked by TAG at amino acid 234 in Fig. 2A) within the glutamine repeat of mice with a *Mus musculus domesticus* Y chromosome. This would lead to a truncated protein with loss of approximately half of the repeat region. Such *Sry* alleles clearly function, suggesting that this domain is not very important. It has been proposed that the presence of other polymorphisms within the locus may be responsible, in conjunction with this stop codon, for cases of sex-reversal when the Y chromosome of certain *M.m.domesticus* mice is placed on a C57BL6 background. For example, in *Sry* from the *Poschiavinus* Y chromosome one particular glutamine repeat region (marked by '-' in Fig. 2A) is contracted by a single amino acid, in the case of AKR an additional glutamine is present. The modification of the glutamine repeat may result in alteration of an activation domain thereby altering the interaction of *SRY* with target DNA or other proteins. However, recent data by P. Tucker (personal communications and Miller et al., 1994) has suggested polymorphisms are frequent within this region of the *SRY* ORF in wild mice. Sex-reversal seen with *Sry* alleles which have variations in the repeat sequence

may therefore simply reflect differences elsewhere, for example in regulatory regions.

Within the large 3' UTR (3481 nt) there are several features that may affect the stability of the RNA and efficiency of its translation. The presence of four ATTTA sequences, occurring between bp 9720 and 9790, may have a role in destabilising the RNA. Such sequences have been found in other genes e.g. *GM-CSF* and *c-fos* (Shaw and Kamen, 1986; Wilson and Treisman, 1988) where it is believed that an ARE-binding protein interacts with them to stimulate deadenylation (for review see Sachs, 1993; Decker and Parker, 1994). Polyadenylation itself is also likely to be carefully controlled. Within the long 3' UTR there are three AATAAA consensus polyadenylation signals but the first two are not used. The third, at which the majority of transcripts are polyadenylated, also has a (GT)_n(T)_n domain 17 nt downstream of the AATAAA thought to be important in poly(A) site selection (Levitt et al., 1989). Clearly there is selective use of a polyadenylation site. By using the third AATAAA signal a large 3' region extending into the inverted repeat is maintained within the transcript. We also identified an additional polyadenylation site at bp 11575/6. This is representative of a minor (<10%) proportion of the total transcripts and is preceded by an unusual AATATA signal. This site has also been identified independently by P. Koopman (personal communications). RNA specific oligonucleotides for RT-PCR can be designed based on either the major or minor polyadenylation sites. Unfortunately, the former is very AT rich and gives inconsistent results, however the latter is reliable (data not shown, and P. Koopman, personal communications). The human *SRY* transcript is different in this respect as well. All transcripts polyadenylate at the first AATAAA signal outside the ORF.

The presence of a wide variety of features contained within the *Sry* transcript, namely multiple ATGs, microsatellites in the 3' and 5' UTRs, ATTTA instability sequences and

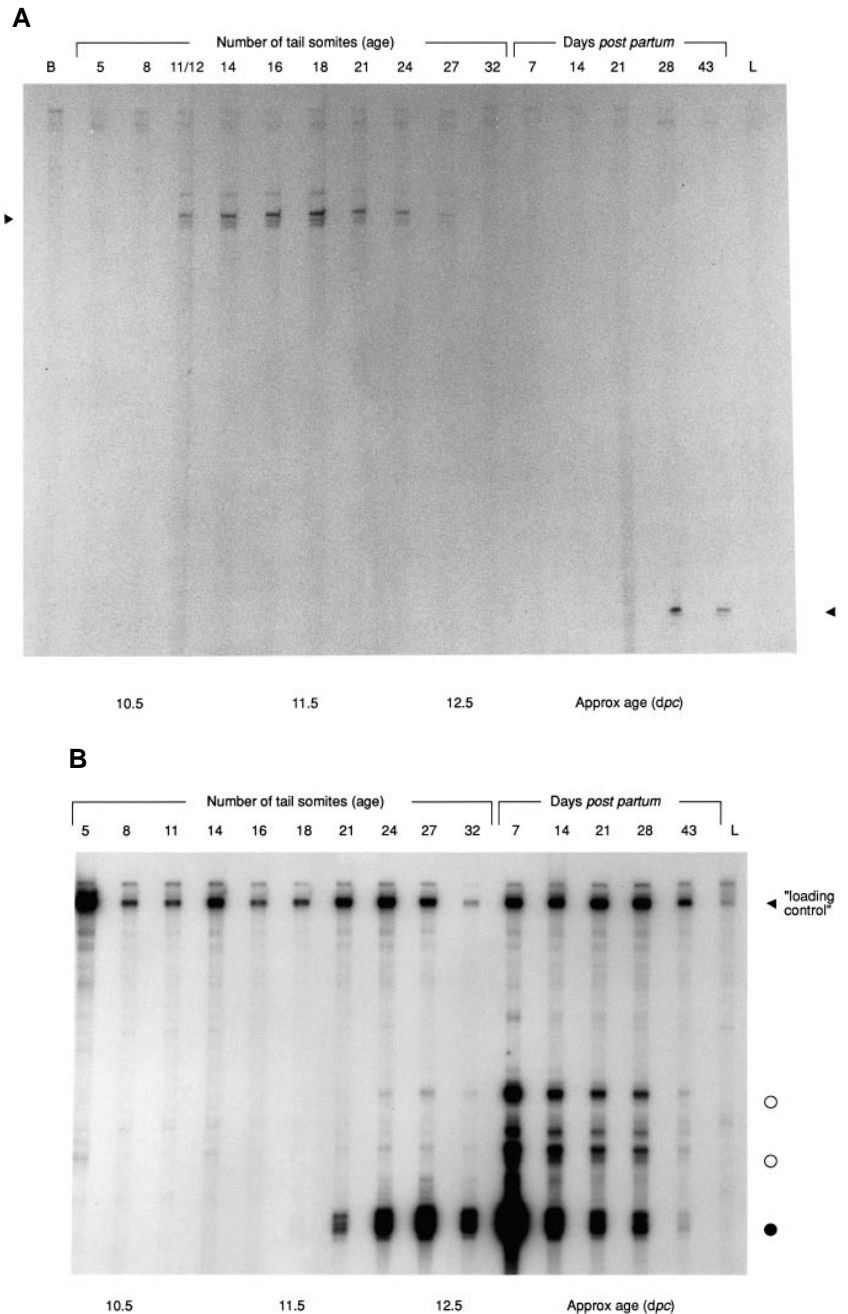
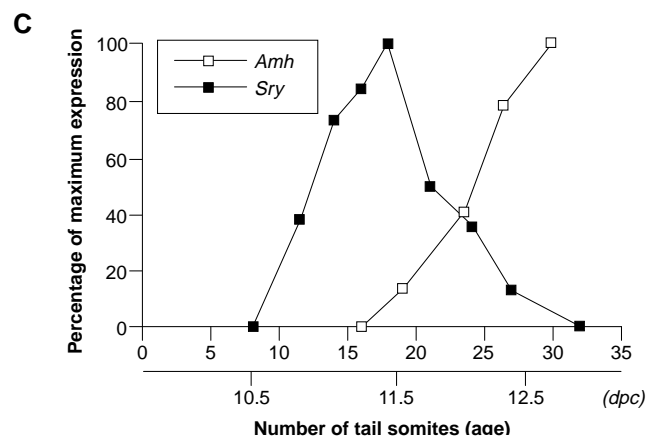


Fig. 5. Timecourse of expression of *Sry* and *Amh* in the developing gonad and testis. Total RNA was prepared using four pairs of sexed male genital ridges, 3/4 of the RNA was used in A, 1/4 used in B; 5 μ g of post partum testis and liver (L) were used in hybridizations. (A) *Sry* expression was analysed using antisense probe 2 to distinguish linear (large 270 nt protection) and circular (small 97 nt protection) transcripts. B=225 blastocysts. (B) Expression was analysed using an antisense probe over the 5' end of *Amh*. One major (●) and two minor (○) transcriptional start sites are observed. Expression of *Amh* at the 18 tail somite stage is more obvious with longer exposures. (C) RNase protection data for expression of *Sry* and *Amh* in males were quantitated using a Molecular Dynamics computer imaging system. The relative amounts of *Sry* and *Amh* transcripts were plotted taking expression of *Sry* at 18 tail somites and of *Amh* at 30 tail somites to be 100%.



specific choice of polyadenylation sites, all imply posttranscriptional regulation that may have a role influencing the rate and period of translation. This is in addition to the precise temporal and quantitative transcriptional regulation of the gene. In humans, there is no evidence for tight regulation of expression, in fact expression has been observed using RT-PCR in both adult and fetal tissues (Clépet et al., 1993). This could support the idea that the presence of an interacting protein is necessary to trigger the activity of human SRY in a tissue specific manner.

From 28 dpp *Sry* transcripts are seen in the testis in their circular form. This coincides approximately with the first occurrence of round spermatids (Nebel et al., 1961; Bellvé et al., 1977). Such data are consistent with expression of *Sry* in purified round spermatids in testis cell separations (Rossi et al., 1993) and the fact that *Sry* expression is absent in adult testes of XX Sxr males which are devoid of germ cells (Koopman et al., 1990). It seems likely that these transcripts are non-functional and may just reflect deregulated expression of genes on the Y chromosome (see Introduction). We have shown here that there are large perhaps primary transcripts in adult testis of more than 8 kb. Analysis of the 32.10 transgenic line on an AKR background suggests that a distal promoter element that gives rise to these transcripts lies in the 5' arm of the inverted repeat within the p741 14.6 kb clone. This is presumed to be different from that used in somatic cells of the genital ridge. The latter is presumed to be in the unique region flanking the observed initiation sites. The distal promoter gives a transcript containing both 5' and 3' arms of the inverted repeat which could hybridize to form a stem loop structure allowing splicing from splice donor and acceptor to give a circular exon. No circular transcripts have been identified in the genital ridge as the splice acceptor is not used (probe 2, Fig. 1), despite the transcript containing both splice acceptor and donor. The simple explanation is that the genital ridge RNA cannot form a stem loop as it lacks 5' inverted repeat sequences.

We have been unable to demonstrate any transcription termination or polyadenylation for the adult testis transcript. Presumably when the RNA is spliced to give the 1.23 kb circle the remaining 'intron', which will be mostly double stranded RNA, is rapidly degraded. Under these circumstances the presence or absence of a poly(A) tail would have no effect on the stability of the remaining transcript fragments. We have been unable to define the transcription initiation site used in adult testis, although it probably lies in the 5' arm of the inverted repeat within the 14.6 kb clone. Since the sequence surrounding this promoter is repeated at the 3' end of the *Sry* locus it might be expected that antisense transcripts would arise. We have not detected antisense transcripts over the region of the HMG box (data not shown). By sequencing further into the 3' repeat it may be possible to identify polymorphisms which serve to prevent antisense transcripts from occurring. Similarly, despite the knowledge that the promoter of the genital ridge transcript lies within the unique region, the whereabouts of any enhancer elements is unknown but their location within the inverted repeat would result in the presence of two copies, possibly acting synergistically. This could explain why the inverted repeat structure is maintained. Alternatively it could be a relic of multiple copies of *Sry*, as has been observed in other rodent species (Nagamine, 1994).

Recently, using RT-PCR analysis (Gaudette et al., 1993;

Zwingman et al., 1993; Boyer and Erickson, 1994), expression of *Sry* has been observed at a third site, namely the pre-implantation embryo. We have used RNase protection to analyse blastocysts for expression of *Sry* but have been unable to detect any transcripts (e.g. lane B Fig. 5A). Our RNase protection assays should have been sensitive enough to detect the 3.2×10^5 copies of *Sry* that Gaudette et al. (1993) suggest are present in a single blastocyst. There is no obvious role for *Sry* in pre-implantation development. Although it has been proposed that it could account for differences in embryonic growth rate of XY and XX embryos (e.g. Zwingman et al., 1993), this now seems unlikely as lack of *Sry* has been shown to have no effect (Burgoyne et al. unpublished data). Perhaps any transcription of *Sry* in the pre-implantation embryo is the same as in spermatids and reflects de-regulated expression.

Since we know the initiation sites for the mouse and human transcripts it is possible to align sequences 5'. This reveals two regions of limited homology. One abuts these initiation sites, but considerable differences are still apparent. In human *SRY* there is no obvious TATAA box element, whereas in the mouse this region is very AT rich, containing the sequence TTAAATATTAATAAAT, which may represent a slightly degenerate and expanded TATA box. The lack of any consensus TATA box may be responsible for the presence of multiple initiation sites for both human and mouse *SRY*. Additional elements often associated with promoter regions appear to be lacking, for example there are no obvious Sp1 binding sites. The other region of homology lies 1520 bp 5' of the mouse and 1900 bp from the human transcription initiation sites. However, a search of the transcription factor binding site database (GCG Wisconsin) identified no sites common to both sequences in this region.

The regulation of *Sry* must involve a combination of transcription factors, which define the level of expression, and the spatial and temporal pattern. Such factors must be present in both male and female embryos. It is important to define the profile of *Sry* expression precisely to correlate it with cellular and molecular events. The data presented here are in approximate agreement with those reported previously (Koopman et al., 1990; Rossi et al., 1993), however such data were obtained with pooled material from only crudely staged embryos. The use of carefully staged embryos allows independent comparisons with expression profiles of other genes. We find that *Sry* transcription begins after the genital ridge has developed as a prominent structure (approx. 10.5 dpc, 11/12 tail somites). This process is believed by many to involve an epithelial to mesenchymal transition. This must however, begin at approximately 10 dpc. From our data, it appears that *Sry* expression is not coincident with this, but occurs later, presumably within cells committed to the supporting cell lineage. *Sry* transcripts accumulate over the next 24 hours, reaching a peak at approx. 11.5 dpc (18 tail somites) and then decline sharply so that none are detected 22 hours later. Morphological organisation of cells in the genital ridge to form testis cords occurs over this period. It is assumed that *Sry* is responsible for triggering the differentiation of Sertoli cells, but it is clearly not required for the maintenance of the differentiated state. Conversely, if testicular differentiation fails, as with B6-Y^{DOM} sex-reversal, expression of *Sry* is prolonged (Lee and Taketo, 1994).

Any direct downstream target gene for SRY would be expected to begin its expression shortly after *Sry* expression is

initiated. One of the first products known to be made by Sertoli cells is AMH. Previous experiments had suggested that there was little or no overlap between *Sry* and *Amh* expression (Koopman et al., 1990; Münsterberg and Lovell-Badge, 1991) making it unlikely that there was a direct link between the two. However, we show here that *Amh* transcripts begin to accumulate once *Sry* transcription is at its peak. There is still a 20-hour delay after *Sry* transcription has started, which allows room for intervening step(s), but it is at least formally possible that there is a direct interaction between SRY and *Amh*. We and others have been unable to demonstrate significant binding of SRY to the *Amh* promoter region (Harley et al., 1994; Shen et al., 1994), although others do report binding in vitro (Haqq et al., 1993, 1994). The regulatory regions of *Amh* have not been well defined in vivo, although recent co-transfection experiments reported by Haqq et al. (1994) are consistent with SRY acting indirectly rather than directly on *Amh* expression. A recent candidate for a role in activating *Amh* is the steroidogenic factor SF1. Expression studies in the rat have demonstrated that there is a sexually dimorphic pattern of expression, presumably triggered by *Sry*, which may regulate *Amh* (Hatano et al., 1994; Shen et al., 1994). We are examining SF1 expression in the mouse, and if such a hypothesis is true, then just such a dimorphic expression pattern should be seen during the window of expression defined by *Sry* and *Amh*. In the context of the precise pattern of *Sry* expression which we have now established, it will be possible to understand whether a role in testis development is likely for other genes expressed in the developing gonad.

We are indebted to Amanda Swain, Silvia Nicolis, Nigel Vivian and Paul Burgoyne for their help and advice and to all members of the laboratories, past and present, who have given moral support and devoted time, effort and brain cells to the many problems encountered along the way.

REFERENCES

- Behlke, M. A., Bogan, J. S., Beer-Romero, P. and Page, D. C. (1993). Evidence that the SRY protein is encoded by a single exon on the human Y chromosome. *Genomics* **17**, 736-739.
- Bellvé, A. R., Cavicchia, J. C., Millette, C. F., O'Brien, D. A., Bhatnagar, Y. M. and Dym, M. (1977). Spermatogenic cells of the prepubertal mouse. *J. Cell. Biol.* **74**, 68-85.
- Boyer, T. R. and Erickson, R. P. (1994). Detection of circular and linear transcripts of *Sry* in pre-implantation mouse embryos: differences in requirement for reverse transcriptase. *Biochem. Biophys. Res. Commun.* **198**, 492-496.
- Capel, B., Swain, A., Nicolis, S., Hacker, A., Walter, M., Koopman, P., Goodfellow, P. and Lovell-Badge, R. (1993). Circular transcripts of the testis-determining gene *Sry* in adult mouse testis. *Cell* **73**, 1019-1030.
- Chirgwin, J., Przybyla, A., MacDonald, R. and Rutter, W. J. (1979). Isolation of biologically active ribonucleic acid from sources enriched in ribonuclease. *Biochemistry* **18**, 5294-5299.
- Clépet, C., Schafer, A. J., Sinclair, A. H., Palmer, M. S., Lovell-Badge, R. and Goodfellow, P. N. (1993). The human *SRY* transcript. *Hum. Molec. Genet.* **2**, 2007-2012.
- Coward, P., Nagai, K., Chen, D., Thomas, H. D., Nagamine, C. M. and Lau, C. Y.-F. (1994). Polymorphism of a CAG trinucleotide repeat within *Sry* correlates with B6.Y^{Dom} sex reversal. *Nature Genetics* **6**, 245-250.
- Decker, C. J. and Parker, R. (1994). Mechanisms of mRNA degradation in eukaryotes. *Trends Biochem. Sci.* **19**, 336-340.
- Dubin, R. A. and Ostrer, H. (1994). SRY is a transcriptional activator. *Mol. Endo.* **8**, 1182-1192.
- Ferrari, S., Harley, V. R., Pontiggia, A., Goodfellow, P. N., Lovell-Badge, R. and Bianchi, M. E. (1992). SRY, like HMG1, recognizes sharp angles in DNA. *EMBO J.* **11**, 4497-4506.
- Fischman, K., Edman, J. C., Shackelford, G. M., Turner, J. A., Rutter, W. J. and Nir, U. (1990). A murine *fer* testis-specific transcript (*ferT*) encodes a truncated Fer protein. *Mol. Cell. Biol.* **10**, 146-153.
- Frohman, M. A., Dush, M. and Martin, G. R. (1988). Rapid production of full-length cDNAs from rare transcripts: Amplification using a single gene-specific oligonucleotide primer. *Proc. Natn. Acad. Sci. USA* **85**, 8998-9002.
- Gaudette, M. F., Cao, Q. P. and Crain, W. R. (1993). Quantitative analysis of specific messenger RNAs in small numbers of preimplantation embryos. In *Guide to Techniques in Mouse Development* (ed. P. M. Wassarman and M. L. DePamphilis), pp. 328-344. London: Academic Press Inc.
- Gerber, H.-P., Seipel, K., Georgiev, O., Hifferer, M., Hug, M., Rusconi, S. and Schaffner, W. (1994). Transcriptional activation modulated by homopolymeric glutamine and proline stretches. *Science* **263**, 808-811.
- Giese, K., Cox, J. and Grosschedl, R. (1992). The HMG domain of lymphoid enhancer factor 1 bends DNA and facilitates assembly of functional nucleoprotein structures. *Cell* **69**, 1-20.
- Green, H. (1993). Human genetic diseases due to codon reiteration: relationship to an evolutionary mechanism. *Cell* **74**, 955-956.
- Grünert, S. and Jackson, R. J. (1994). The immediate downstream codon strongly influences the efficiency of utilization of eukaryotic translation initiation codons. *EMBO J.* **13**, 3618-3630.
- Gubbay, J., Collignon, J., Koopman, P., Capel, B., Economou, A., Münsterberg, A., Vivian, N., Goodfellow, P. and Lovell-Badge, R. (1990). A gene mapping to the sex-determining region of the mouse Y chromosome is a member of a novel family of embryonically expressed genes. *Nature* **346**, 245-250.
- Haqq, C. M., King, C. Y., Donahoe, P. K. and Weiss, M. A. (1993). SRY recognizes conserved DNA sites in sex-specific promoters. *Proc. Natn. Acad. Sci. USA* **90**, 1097-1101.
- Haqq, C. M., King, C.-Y., Ukiyama, E., Falsafi, S., Haqq, T. N., Donahoe, P. K. and Weiss, M. A. (1994). Molecular basis of mammalian sexual determination: activation of Müllerian inhibiting substance gene expression by SRY. *Science* **266**, 1494-1500.
- Harley, V. R., Jackson, D. I., Hextal, P. J., Hawkins, J. R., Berkovitz, G. D., Sockanathan, S., Lovell-Badge, R. and Goodfellow, P. N. (1992). DNA binding activity of recombinant SRY from normal males and XY females. *Science* **255**, 453-456.
- Harley, V. R., Lovell-Badge, R. and Goodfellow, P. N. (1994). Definition of a consensus DNA binding site for SRY. *Nucl. Acids Res.* **22**, 1500-1501.
- Hatano, O., Takayama, K., Imai, T., Waterman, M. R., Takakusu, A., Omura, T. and Morohashi, K.-i. (1994). Sex-dependent expression of a transcription factor, Ad4BP, regulating steroidogenic P-450 genes in the gonads during prenatal and postnatal rat development. *Development* **120**, 2787-2797.
- Hawkins, J. R. (1993). Mutational analysis of *SRY* in XY females. *Human Mutation* **2**, 347-350.
- Huntington's Disease Collaborative Research Group (1993). A novel gene containing a trinucleotide repeat that is expanded and unstable on Huntington's disease chromosomes. *Cell* **72**, 971-983.
- King, C.-Y. and Weiss, M. A. (1993). The SRY high-mobility-group box recognizes DNA by partial intercalation in the minor groove: A topological mechanism of sequence specificity. *Proc. Natn. Acad. Sci. USA* **90**, 11990-11994.
- Koopman, P., Münsterberg, A., Capel, B., Vivian, N. and Lovell-Badge, R. (1990). Expression of a candidate sex-determining gene during mouse testis differentiation. *Nature* **348**, 450-452.
- Koopman, P., Gubbay, J., Vivian, N., Goodfellow, P. and Lovell-Badge, R. (1991). Male development of chromosomally female mice transgenic for *Sry*. *Nature* **351**, 117-121.
- Kozak, M. (1983). Translation of insulin-related polypeptides from messenger RNAs with tandemly reiterated copies of the ribosome binding site. *Cell* **34**, 971-978.
- Kozak, M. (1986a). Influences of mRNA secondary structure on initiation by eukaryotic ribosomes. *Proc. natn. Acad. Sci. USA* **83**, 2850-2854.
- Kozak, M. (1986b). Point mutations define a sequence flanking AUG initiator codon that modulates translation by eukaryotic ribosomes. *Cell* **44**, 283-292.
- Kozak, M. (1987a). An analysis of 5'-noncoding sequences from 699 vertebrate messenger RNAs. *Nucl. Acids Res.* **15**, 8125-8148.
- Kozak, M. (1987b). At least six nucleotides preceding the AUG initiator codon enhance translation in mammalian cells. *J. Mol. Biol.* **196**, 947-950.
- Kozak, M. (1990). Downstream secondary structure facilitates recognition of

- initiator codons by eukaryotic ribosomes. *Proc. natn. Acad. Sci. USA* **87**, 8301-8305.
- Kreig, P. A. and Melton, D. A.** (1987). In vitro RNA synthesis with SP6 RNA polymerase. *Methods Enzymol.* **155**, 397-416.
- Lee, C.-H. and Taketo, T.** (1994). Normal onset, but prolonged expression, of *Sry* in the B6.Y^{DOM} sex-reversed mouse gonad. *Dev. Biol.* **165**, 442-452.
- Levitt, N., Briggs, D., Gil, A. and Proudfoot, N. J.** (1989). Definition of an efficient synthetic poly(A) site. *Genes Dev.* **3**, 1019-1025.
- Lovell-Badge, R. H. and Robertson, E.** (1990). XY female mice resulting from a heritable mutation in the murine primary testis determining gene, *Tdy*. *Development* **109**, 635-646.
- McElreavy, K., Vilain, E., Abbas, N., Costa, J.-M., Souleyreau, N., Kucheria, K., Boucekkine, C., Thibaud, E., Brauner, R., Flamant, F. and Fellous, M.** (1992). XY sex reversal associated with a deletion 5' to the *SRY* 'HMG box' in the testis-determining region. *Proc. natn. Acad. Sci. USA* **89**, 11016-11020.
- Meijer, D., Hermans, A., von Lindern, M., van Agthoven, t, de Klein, A., Mackenbach, P., Grootegoed, A., Talarico, D., Valle, G. D. and Grosveld, G.** (1987). Molecular characterization of the testis specific *c-abl* mRNA in mouse. *EMBO J.* **6**, 4041-4048.
- Mhatre, A. N., Trifiro, M. A., Kaufman, M., Kazemi-Esfarjani, P., Figlewicz, D., Rouleau, G. and Pinsky, L.** (1993). Reduced transcriptional regulatory competence of the androgen receptor in X-linked spinal and bulbar muscular atrophy. *Nature Genetics* **5**, 184-188.
- Miller, K. E., Lundrigan, B. L. and Tucker, P. K.** (1995). Length variation of CAG repeats in *Sry* across populations of *Mus domesticus*. *Mammalian genome* In Press.
- Mitchell, P. J. and Tjian, R.** (1989). Transcriptional regulation in mammalian cells by sequence-specific DNA binding proteins. *Science* **245**, 371-378.
- Münsterberg, A. and Lovell-Badge, R.** (1991). Expression of the mouse anti-Müllerian hormone gene suggests a role in both male and female sexual differentiation. *Development* **113**, 613-624.
- Nagamine, C. M.** (1994). The testis-determining gene, *Sry*, exists in multiple copies in Old World rodents. *Genet. Res.* **64**, 151-159.
- Nasrin, N., Buggs, C., Kong, X. F., Carnazza, J., Goebel, M. and Alexander-Bridges, M.** (1991). DNA-binding properties of the product of the testis-determining gene and a related protein. *Nature* **354**, 317-320.
- Natesan, S. and Gilman, M. Z.** (1993). DNA bending and orientation-dependent function of YY1 in the *c-fos* promoter. *Genes Dev.* **7**, 2497-2509.
- Nebel, B. R., Amarose, A. P. and Hackett, E. M.** (1961). Calendar of gametogenic development in the prepuberal male mouse. *Science* **134**, 832-833.
- Orr, H. T., Chung, M.-y, Banfi, S., Kwiatkowski Jr, T. J., Servadio, A., Beaudet, A. L., McCall, A., Duvick, L. A., Ranum, L. P. W. and Zoghbi, H. Y.** (1993). Expansion of an unstable trinucleotide CAG repeat in spinocerebellar ataxia type 1. *Nature Genetics* **4**, 221-226.
- Palmer, S. J. and Burgoyne, P. S.** (1991). The *Mus musculus domesticus Tdy* allele acts later than the *Mus musculus musculus Tdy* allele: a basis for XY sex reversal in C57Bl/6-Y^{Pos} mice. *Development* **113**, 709-714.
- Pelletier, J. and Sonenberg, N.** (1985). Insertion mutagenesis to increase secondary structure within the 5' noncoding region of a eukaryotic mRNA reduces translation efficiency. *Cell* **40**, 515-526.
- Pontiggia, A., Rimini, R., Harley, V. R., Goodfellow, P. N., Lovell-Badge, R. and Bianchi, M. E.** (1994). Sex-reversing mutations affect the architecture of SRY-DNA complexes. *EMBO J.* **13**, 5115-6124.
- Poulat, F., Soullier, S., Goz , C., Heitz, F., Calas, B. and Berta, P.** (1994). Description and functional implications of a novel mutation in the sex-determining gene *SRY*. *Human Mutation* **3**, 200-204.
- Rossi, P., Dolci, S., Albanesi, C., Grimaldi, P. and Geremia, R.** (1993). Direct evidence that the mouse sex-determining gene *Sry* is expressed in the somatic cells of male fetal gonads and in the germ cell line in the adult testis. *Mol. Reproduct. Dev.* **34**, 369-373.
- Rubin, M. R. and Nguyen-Huu, M. C.** (1991). Murine embryonic spinal cord and adult testis *Hox-1.4* cDNAs are identical 3' to the homeo box. *DNA Seq.* **1**, 329-334.
- Ruppert, S., Cole, T. J., Boshart, M., Schmid, E. and Schutz, G.** (1992). Multiple mRNA isoforms of the transcription activator protein CREB: generation by alternative splicing and specific expression in primary spermatocytes. *EMBO J.* **11**, 1503-1512.
- Sachs, A. B.** (1993). Messenger RNA degradation in eukaryotes. *Cell* **74**, 413-421.
- Sargent, M. N. C., O'Brien, J., Wolfe, J., Affara, N. A. and Ferguson-Smith, M. A.** (1991). Comparison of *ZFY* and *ZFX* gene structure and analysis of alternative 3' untranslated regions of *ZFY*. *Nucl. Acids Res.* **19**, 2579-2586.
- Sargent, T. D., Jamrich, M. and Dawid, I. B.** (1986). Cell interactions and the control of gene activity during early development of *Xenopus laevis*. *Dev. Biol.* **114**, 238-246.
- Schoenberg, M. P., Hakimi, J. M., Wang, S., Bova, G. S., Epstein, J. I., Fischbeck, K. H., Isaacs, W. B., Walsh, P. C. and Barrack, E. R.** (1994). Microsatellite mutation (CAG₂₄₋₂₈) in the androgen receptor gene in human prostate cancer. *Biochem. Biophys. Res. Commun.* **198**, 74-80.
- Shaw, G. and Kamen, R.** (1986). A conserved AU sequence from the 3' untranslated region of *GM-CSF* mRNA mediates selective mRNA degradation. *Cell* **46**, 659-667.
- Shen, W.-H., Moore, C. C. D., Ikeda, Y., Parker, K. L. and Ingraham, H. A.** (1994). Nuclear receptor steroidogenic factor 1 regulates the Müllerian inhibiting substance gene: a link to the sex determination pathway. *Cell* **77**, 651-661.
- Su, H. and Lau, Y.-F. C.** (1993). Identification of the transcription unit, structural organization, and promoter sequence of the human sex-determining region Y (*SRY*) gene, using a reverse genetic approach. *Am. J. Hum. Genet.* **52**, 24-38.
- Tajima, T., Nakae, J., Shinohara, N. and Fujieda, K.** (1994). A novel mutation localized in the 3' non-HMG box region of the *SRY* gene in 46,XY gonadal dysgenesis. *Hum. Mol. Genet.* **3**, 1187-1189.
- Tucker, P. K. and Lundrigan, B. L.** (1993). Rapid evolution of the sex determining locus in Old World mice and rats. *Nature* **364**, 715-717.
- Vilain, E., Fellous, M. and McElreavey, K.** (1992). Characterization and sequence of the 5' flanking region of the human testis-determining factor *SRY*. *Methods Molec. Cell. Biol.* **3**, 128-134.
- Whitfield, S., Lovell-Badge, R. and Goodfellow, P. N.** (1993). Rapid sequence evolution of the the sex determining gene *SRY*. *Nature* **364**, 713-715.
- Wilson, T. and Treisman, R.** (1988). Removal of poly(A) and consequent degradation of *c-fos* mRNA facilitated by 3' AU-rich sequences. *Nature* **336**, 396-399.
- Zwingman, T., Erickson, R. P., Boyer, T. and Ao, A.** (1993). Transcription of the sex-determining region genes *Sry* and *Zfy* in the mouse pre-implantation embryo. *Proc. Natn. Acad. Sci. USA* **90**, 814-817.

(Accepted 7 February 1995)