A gene from the human sex-determining region encodes a protein with homology to a conserved DNA-binding motif

Andrew H. Sinclair, Philippe Berta*, Mark S. Palmer, J. Ross Hawkins, Beatrice L. Griffiths, Matthijs J. Smith, Jamie W. Foster*, Anna-Maria Frischauf, Robin Lovell-Badge† & Peter N. Goodfellow

Human Molecular Genetics Laboratory, Imperial Cancer Research Fund, PO Box 123, Lincoln's Inn Fields, London, WC2A 3PX, UK
† Laboratory of Eukaryotic Molecular Genetics, National Institute for Medical Research, The Ridgeway, Mill Hill, London NW7 1AA, UK

A search of a 35-kilobase region of the human Y chromosome necessary for male sex determination has resulted in the identification of a new gene. This gene is conserved and Y-specific among a wide range of mammals, and encodes a testis-specific transcript. It shares homology with the mating-type protein, Mc, from the fission yeast *Schizosaccharomyces pombe* and a conserved DNA-binding motif present in the nuclear high-mobility-group proteins HMG1 and HMG2. This gene has been termed *SRY* (for sex-determining region Y) and proposed to be a candidate for the elusive testis-determining gene, *TDF*.

THE mammalian Y chromosome plays a crucial part in sex determination: an embryo that inherits a Y chromosome develops as a male whereas an embryo lacking a Y chromosome develops as a female[1]. The sex determining gene(s) on the Y chromosome induces testicular development, and subsequent male sexual differentiation is a consequence of the hormonal products of the testis[2]. The Y-encoded testis-determining gene has been named *TDF* (testis-determining factor) in humans and *Tdy* (testis-determining Y chromosome) in mouse. Although it is likely that many different genes are required for both male and female sex determination, understanding the mode of action of *TDF* may provide a general model for the genetic control of developmental decisions in mammals.

Attempts to identify and clone *TDF* have exploited detailed maps of the Y chromosome. Three types of map have been constructed: deletion maps from analysis of the genomes of sex-reversed XX males and XY females[3,4]; a meiotic map of the pseudoautosomal region which is shared by the X and Y chromosomes[5,6] and a long-range restriction map linking the first two maps[7]. Most XX males have inherited Y-derived sequences, including *TDF*, by terminal exchange between the X and Y chromosomes[8,9]. A map constructed on the basis of the Y fragments present in different XX males placed *TDF* in the distal part of the Y chromosome adjacent to the pseudoautosomal region[3]. The meiotic map of the pseudoautosomal region indicated that *MIC2* was the closest known pseudoautosomal locus to the sex-specific part of the Y chromosome[5]. A long-range restriction map beginning at *MIC2* and extending into the sex-specific region of the X and Y chromosomes identified a CpG-rich island on the Y chromosome which represented a possible candidate gene for *TDF*[7]. This gene was cloned after a chromosome walk initiated 130-kilobase (kb) proximal to the CpG-rich region, and subsequently named *ZFY* (ref. 10). The sequences present in a particular XX male (LGL203) and absent in an XY female (WHT1013) were used to define the position of *TDF* to within an interval of 140 kb and *ZFY* was isolated from this region. Other evidence consistent with identity between *ZFY* and *TDF* included the finding of ZFY-related sequences on the Y chromosome of all eutherian mammals tested; the presence of a *ZFY*-related gene, *Zfy-1*, in *Sxr'* (the smallest part of the mouse Y chromosome known to be sex-determining) and the structure of the *ZFY*-encoded protein, which has many features in common with transcription factors[10,11]. But there were several unexpected findings: first, *ZFX*—a homologue of *ZFY*—was found on the eutherian X chromosome and shown in humans to escape inactivation[11,12], and second, in metatherian mammals (marsupials), *ZFY*-related sequences were found not on the Y or X chromosome, but on the autosomes[13].

Two recent reports have further questioned the role of *ZFY* in male sex-determination. Koopman *et al.* studied the expression of *Zfy-1* and *Zfy-2*, the mouse homologues of *ZFY*, and found that their expression was linked to germ cells—a cell-type not required for normal male development. Furthermore, in W^e/W^e mutant mice, testicular development occurred in the absence of detectable *Zfy-1* and *Zfy-2* expression[14]. Palmer *et al.* described four sex-reversed XX individuals that had inherited Y-derived sequences not including *ZFY*. Assuming that these individuals are male because of their Y-derived sequences, this mapped *TDF* to the 60 kb proximal to the pseudoautosomal boundary. This result is inconsistent with the published breakpoint of the XY female (WHT1013) (ref. 10) and formally excludes *ZFY* as a candidate for *TDF* (ref. 15).

We have tested the hypothesis that the region immediately adjacent to the boundary contains *TDF* by searching for a gene in this region. We have found an open reading frame that is part of a conserved, Y-specific gene. This gene also shares homology with the *Mc* gene of *S. pombe* and with a conserved DNA-binding motif present in non-histone proteins related to HMG1 and HMG2.

Analysis of 35 kb of the Y chromosome

We have previously described a chromosome walk, comprising a series of overlapping λ and cosmid clones, from the pseudoautosomal region, across the pseudoautosomal boundary, to the sex-specific region of the Y-chromosome[16]. Probes from this walk were used by Palmer *et al.* on the genomes of XX males to define the region in which *TDF* must lie[15]. Since then, three additional probes have been used on the genomes of the same *ZFY*-negative XX males: pY4.1β, which was positive with the XX males, and pYH8, which was negative (Fig. 1). The third Y-specific probe, pYR0.4, is derived from sequences lying between pY4.1β and pYH8, and seems to define the break points in the XX males. The probe pYR0.4 detects an 8.5-kb *Hin*dIII fragment in normal males but only a 4-kb fragment in

* Permanent addresses: Centre de Recherches de Biochimie Macromoleculaire, CNRS LP8402/INSERM U.249, Route de Mende, 34033 Montpellier Cedex, France (P.B.) and Department of Genetics and Human Variation, La Trobe University, Bundoora, Victoria 3083, Australia (J.W.F.).

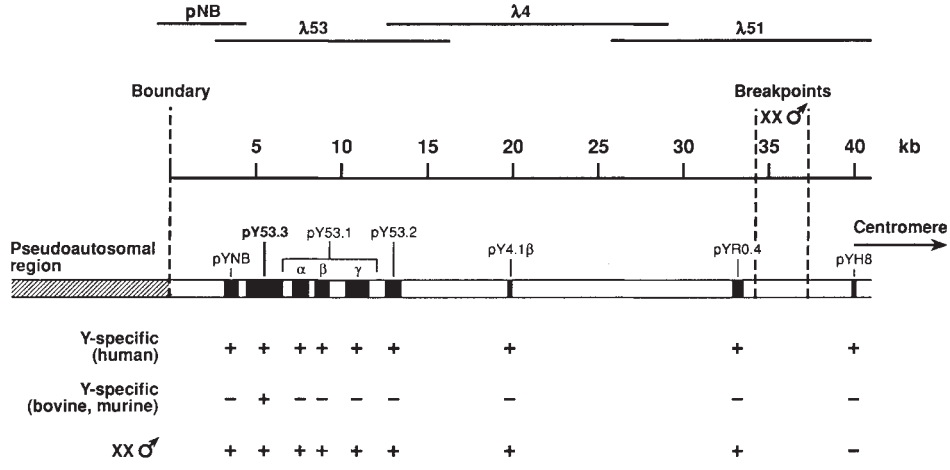

FIG. 1 Map of distal short arm of human Y chromosome: shaded region (left), pseudoautosomal region; broken line, boundary between pseudoautosomal and Y-specific regions. Top, the three overlapping lambda clones λ51, λ4, λ53 and the plasmid pNB obtained from walking along the Y chromosome. Breakpoints of the XX males, broken lines at 35 kb (see Fig. 2). Black boxes, probes detecting single-copy Y-specific human DNA fragments (+). When these probes were hybridized to male and female DNA from bovine and murine genomes, only pY53.3 detected Y-specific fragments (+). All of the probes except pYH8 hybridized to sequences in the XX males. pY53.3 was known as H2.1 and pY53.2 was called P0.9 (ref. 16).

two related individuals, an XX male and an hermaphrodite; a 6-kb fragment was detected in a third XX male (Fig. 2). The breakpoints in the XX males are clustered around a region that is roughly 35 kb proximal to the boundary (Fig. 1). This result implies that *TDF* is located in sex-specific sequences within 35 kb of the pseudoautosomal boundary. Further refinement of the positions of the breakpoints was not possible because of the highly repetitive nature of the sequences between pY4.1β, pYR0.4 and pYH8.

The following strategy was adopted to locate *TDF* within this 35-kb region of the Y chromosome. DNA from this region was subcloned into fragments of about 4 kb in size. Subsequently, each 4-kb fragment was cleaved with frequently cutting restriction enzymes such as *Rsa*I, to produce smaller fragments in the size range 500 base pairs (bp) to 1 kb. A total of 50 probes generated in this manner were examined. Each small fragment was radioactively labelled and used to probe Southern blots of DNA from human males and females, murine males and females, bovine males and females and human–hamster somatic cell hybrids containing the human X or the human Y chromosome. All probes were tested with and without prehybridization to total human DNA to suppress the contribution of repetitive sequences[17]. Despite this latter precaution, most probes tested failed to detect unique sequences in the genome of humans but reacted with repetitive elements distributed throughout the genome. These repetitive probes frequently also detected repeat sequences in the bovine genome.

Seven probes were found that detect single-copy Y-specific bands in *Eco*RI-digested human DNA (Fig. 1); pYNB (0.7 kb), pY53.3 (2.1 kb), pY53.1α (0.8 kb), pY53.1β (0.8 kb), pY53.1γ (1.3 kb), pY53.2 (0.9 kb) and pY4.1β (0.2 kb). But of the seven probes, only pY53.3 also reacted with Y-specific bands in the murine and bovine genomes. A 0.9-kb *Hin*cII subclone of pY53.3 hybridized most strongly to Y-specific fragments in human, murine and bovine genomic DNA. This subclone was used as a probe in subsequent experiments.

Conservation of pY53.3

A 'Noah's Ark' blot containing DNA from male and female pairs of eutherian mammals was hybridized with the 0.9-kb *Hin*cII subclone of pY53.3 (Fig. 3*a*).

The sequences detected by this probe are conserved and male-specific in a wide spectrum of eutherian mammals. At low stringency, additional fragments were found that are shared between males and females, suggesting the existence of pY53.3-related X-linked or autosomal sequences. But these fragments were not detected in a human–rodent somatic cell hybrid that retained the human X chromosome as the only human contribution (Fig. 3*b*).

Unique sequences that are conserved between the Y chromosomes of different mammalian species are very rare, the only other known example being *ZFY*[10]. This result is consistent with pY53.3 reacting with functional sequences on the eutherian Y chromosome.

Sequence analysis of Y-unique sequences

The sequence of pY53.3 was determined by primer-walking. A search of the EMBL DNA sequence database failed to find any sequence related to pY53.3. But inspection of the pY53.3 (2.1 kb) sequence revealed two long open reading frames corresponding

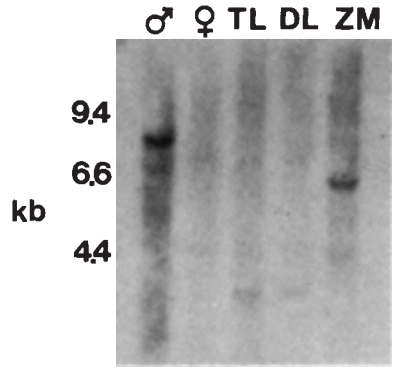

FIG. 2 Southern blot analysis of *Hin*dIII-digested genomic DNA from: ♂, 46XY cell line, PGF (ref. 28); ♀, 46XX cell line, WT49 (ref. 29); TL, familial cases (XX male); DL (hermaphrodite) and ZM (unrelated XX male) hybridized with pYR0.4. The probe detects an 8.5-kb *Hin*dIII fragment in normal males but only a 4-kb fragment in the two related individuals TL and DL, and a 6-kb fragment in ZM. This implies that breakpoints in these XX males and hermaphrodite are clustered around a region 35 kb from the boundary (see Fig. 1). DNA size markers (kb) to the left.
METHODS. Genomic DNA (10 μg) was digested with *Hin*dIII, separated on a 0.8% agarose gel, transferred to Hybond N^+ (Amersham) and fixed in 0.4 M NaOH (20 min). To suppress repeat sequences present in the probe pYR0.4 it was labelled with ^{32}P, denatured together with 500 μl (5 mg ml^{-1}) sonicated human placental DNA and prehybridized in 2 ml hybridization buffer at 65 °C for 2 h[17]. This probe mixture (2 ml) was added to the filter in a 5×SSPE buffer, 5×Denhardt's solution, 0.5% sodium dodecylsulphate, 200 μg ml^{-1} denatured salmon-sperm DNA and 10% sonicated denatured human placental DNA, and hybridized for 16 h at 65 °C. The filter was extensively washed with 0.2×SSC buffer, 0.2% SDS at 65 °C and autoradiographed for 6 days.

to 99 and 223 amino acids that overlap in different frames, reading 5′→3′ from the centromere towards the pseudoautosomal boundary (Fig. 4*a*). We screened the PIR protein database using a similarity search algorithm for sequences related to the proteins predicted to be encoded by these open reading frames[18,19]. The protein putatively encoded by the shorter open reading frame was unrelated to any sequence in the PIR protein database. But the protein putatively encoded by the longer open reading frame was found to have striking similarity both to a portion of the Mc protein encoded by the *mat3-M* of the fission yeast *S. pombe*[20] and to a conserved motif in several non-histone proteins related to HMG1 and HMG2[21,22]. The conserved motif covers a stretch of 80 amino acids and suggests a common structural role in these proteins, which could link them functionally (Fig. 4*b*).

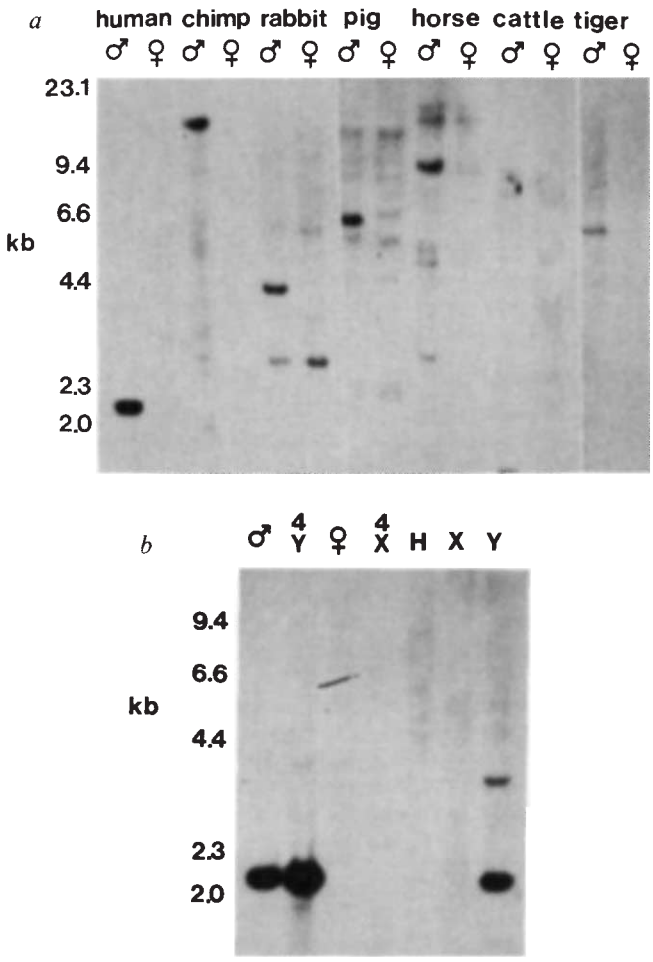

FIG. 3 *a*, 'Noah's Ark' blot of *Hin*dIII-digested genomic DNA from male and female pairs of eutherian mammals, hybridized with the 0.9-kb *Hin*cII fragment of pY53.3. The probe detects male-specific fragments in human (2.1 kb), chimpanzee (~18 kb), rabbit (4.2 kb), pig (6.6 kb), horse (~10 kb), cattle (1.6 kb) and tiger (6.3 kb). Rabbit ♀ lane is slightly overloaded. *b*, Southern blot of: ♂, 46XY cell line, PGF (ref. 28); 4Y, 49XYYYY cell line, Oxen (ref. 30); ♀, 46XX cell line, WT49 (ref. 29); 4X, 48XXXX cell line, GM1416B (Coriell Institute for Medical Research, Camden. New Jersey, USA); H, hamster parent cell line, W3GH (ref. 31); X, hamster–human hybrid cell line containing the human X chromosome, CL2D (ref. 32); Y, hamster–human hybrid cell line containing the human Y chromosome, 853 (ref. 33). The filter was hybridized with the (0.9 kb) *Hin*cII fragment of pY53.3, which detects a 2.1-kb *Hin*dIII fragment in ♂, 4Y and Y, the intensity of the bands reflects the number of copies of the Y chromosome present. DNA size markers (kb), to the left.

METHODS. Genomic DNA (10 μg) was digested with *Hin*dIII, separated on a 0.8% agarose gel, transferred to Hybond N^+ (Amersham) and fixed in 0.4 M NaOH. Probe pY53.3 was labelled with ^{32}P and added to the filter in 5 × SSPE buffer, 5 × Denhardt's solution, 0.5% SDS, 10% dextran sulphate, 200 μg ml^{-1} denatured salmon-sperm DNA, and hybridized for 16 h at 65 °C. The filter was washed in 1 × SSC buffer, 0.2% SDS at 65 °C and autoradiographed for 2 days.

At the 3′ end of the conserved motif in pY53.3, the sequence continues for 68 amino acids before reaching a stop codon. No potential splice donor site was found in the DNA sequence of this region but there is a putative polyadenylation signal 133 bp downstream of the stop codon. In the 5′ direction there is a potential splice acceptor signal in the pY53.3 DNA sequence at the point where homology between the conserved motif and pY53.3 ends. The open reading frame continues 5′ for a distance corresponding to another 75 amino acids and within this region two potential start codons are found in pY53.3.

To test the conservation and hence functional importance of the sequence motifs in pY53.3, the homologous Y-specific rabbit sequence was cloned and sequenced. Within the conserved motif the human and rabbit Y-specific sequences share 64 out of 80 amino acids (80%) with a further eight amino acids showing conservative changes (90% similarity overall) (Fig. 4*c*). Outside the motif the human and rabbit show only 54% identity. The high degree of homology within the motif strongly suggests pY53.3 contains the coding sequence of a gene.

The other Y-unique probes found in the original search were also sequenced. The pNB and pY53.2 probes did not reveal any open reading frames predicted to encode proteins related to sequences in the EMBL database or the PIR protein database. The probe pY4.1β is part of a larger 1.2-kb *Rsa*I fragment that contains an open reading frame predicted to encode a protein related to retroviral reverse transcriptase, commonly found in repetitive sequences. The probe pY53.1 encompasses a 5.6-kb region containing several open reading frames, but none of these was predicted to encode a protein related to sequences in the EMBL database or the PIR protein database. In total, 10.5 kb of the Y chromosome was sequenced in the search for potential coding sequence.

Tissue distribution and expression

A northern blot prepared with poly(A)$^+$ RNA from human tissues was hybridized with the 0.9-kb *Hin*cII fragment of probe pY53.3 (Fig. 5). The probe detects a fragment of about 1.1 kb in adult testis. Bands were not detected in ovary, lung (male) or kidney (male) cell lines, or in male and female lymphoblastoid cell lines. The filter was stripped and re-probed with β-actin to confirm the presence of poly(A)$^+$ RNA in the samples (Fig. 5). This result is consistent with a testis-specific transcript being encoded by the pY53.3 Y-specific sequence. This was confirmed using 3′ RACE (rapid amplification of complementary DNA ends) polymerase chain reaction (PCR)[23] from adult testis poly(A)$^+$ RNA, which showed a poly(A) tract 15 bases downstream from the potential polyadenylation signal, further indicating that this is the 3′ end of a Y-specific transcript (data not shown).

Discussion

We have extended the findings of Palmer *et al.*[15] and defined a 35-kb region of Y-specific sequence immediately adjacent to the pseudoautosomal boundary in which *TDF* must reside. An extensive Southern blot anlaysis of this 35-kb region of the Y chromosome revealed a Y-unique probe, pY53.3, which detects conserved Y-specific sequences in a wide range of eutherian mammals. In related studies, Gubbay *et al.*[24] have demonstrated that the equivalent murine sequence is present in *Sxr′* (ref. 25)—the smallest part of the mouse Y chromosome known to be male sex-determining—and is deleted from a mutant Y chromosome that has lost male sex-determining function[26]. The conservation of pY53.3-related sequences between the Y chromosomes of eutherian mammals suggests that these sequences have a functional role. The location of the pY53.3-related sequences on the Y chromosomes of man and mouse is consistent with a role in male sex determination.

The translated nucleotide sequence of pY53.3 (2.1 kb) contains two open reading frames. The longer open reading frame encodes a region of 120 amino acids shared with the homologous rabbit Y-specific sequence. Within this region there is a conserved motif of 80 amino acids which shows 80% identity, rising to 90% with conservative substitutions. Outside the conserved region the similarity between the residues drops to 54%.

The amino-acid sequence encoded by the longer open reading frame from pY53.3 showed striking similarity to 80 amino acids at the carboxy-terminal of the Mc protein encoded by the *mat3-M* locus of the fission yeast *S. pombe*[20]. This is the same stretch of 80 amino acids in pY53.3 that is conserved in the rabbit sequence. The mating-type locus, *mat-1*, in the fission yeast has two alternative alleles, *M* and *P*. These alleles are transposed during switch of mating type from either donor loci *mat3-M* or *mat2-P* to the *mat-1* locus. Both loci contain two transposable genes (*Mc* and *Mi*, and *Pc* and *Pi*); none of the four genes are related to each other in sequence. The precise function of the four genes is not known; *Mc* and *Pc* are, however, required for mating and all four genes are needed for meiotic competence[20]. By analogy to the budding yeast it has been suggested that genes of the *mat-1* locus function as transcription factors. This suggestion has been supported by the finding of a diverged homeobox domain in *Pi*[20].

The 80-residue conserved motif in pY53.3 also showed homology with a domain found in the nuclear non-histone proteins HMG1 and HMG2. High mobility group (HMG) proteins 1 and 2 are thought to play a part in chromosomal structure and gene activity, and some display enhanced DNA-binding to A+T-rich single-stranded sequences[27]. HMG1 and HMG2 are

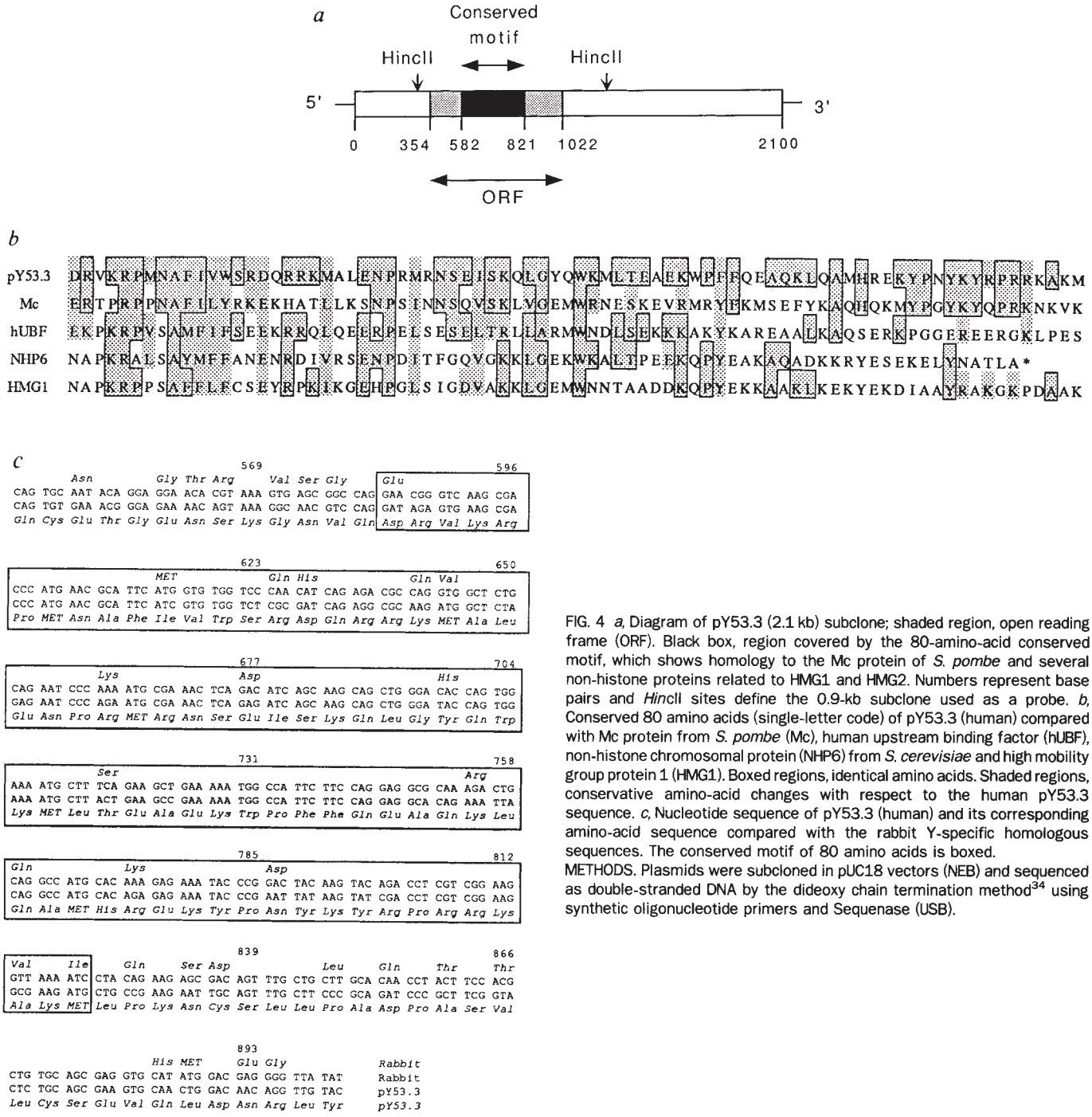

FIG. 4 *a*, Diagram of pY53.3 (2.1 kb) subclone; shaded region, open reading frame (ORF). Black box, region covered by the 80-amino-acid conserved motif, which shows homology to the Mc protein of *S. pombe* and several non-histone proteins related to HMG1 and HMG2. Numbers represent base pairs and *Hin*cII sites define the 0.9-kb subclone used as a probe. *b*, Conserved 80 amino acids (single-letter code) of pY53.3 (human) compared with Mc protein from *S. pombe* (Mc), human upstream binding factor (hUBF), non-histone chromosomal protein (NHP6) from *S. cerevisiae* and high mobility group protein 1 (HMG1). Boxed regions, identical amino acids. Shaded regions, conservative amino-acid changes with respect to the human pY53.3 sequence. *c*, Nucleotide sequence of pY53.3 (human) and its corresponding amino-acid sequence compared with the rabbit Y-specific homologous sequences. The conserved motif of 80 amino acids is boxed.
METHODS. Plasmids were subcloned in pUC18 vectors (NEB) and sequenced as double-stranded DNA by the dideoxy chain termination method[34] using synthetic oligonucleotide primers and Sequenase (USB).

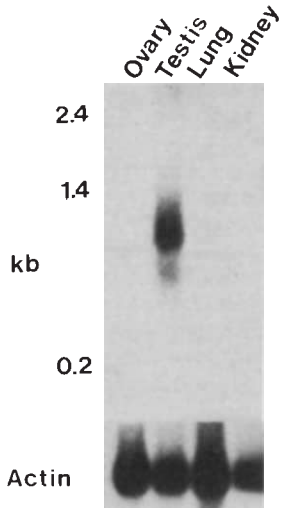

FIG. 5 Northern blot analysis of poly(A)$^+$ RNA from ovary, adult testis, lung (male) and kidney (male), hybridized with the 0.9-kb *Hin*cII fragment of pY53.3. The probe detects a transcript of ~1.1 kb in adult testis and in no other tissue tested. Below is the same filter stripped and re-probed with β-actin, confirming the presence of poly(A)$^+$ RNA in all the tissue samples. Poly(A)$^+$ RNA was also prepared from three lymphoblastoid cell lines 49XYYYY cell line (Oxen)[30], 46XY cell line (PGF)[28], 46XX cell line (WT49)[29], and probed as above, but no transcripts were detected (data not shown). METHODS. RNA was prepared from each tissue as described previously[35] and poly(A)$^+$ messenger RNA selected by poly(A) tract isolation system (Promega). Poly(A)$^+$ RNA (8 μg) was separated on a 1% agarose gel containing 2.2 M formaldehyde, transferred to Hybond N (Amersham), ultraviolet-fixed, and hybridized with ^{32}P-labelled 0.9-kb *Hin*cII fragment of pY53.3. Hybridization was at 65 °C in 3 × SSC buffer, 5 × Denhardt's solution, 200 μg ml^{-1} denatured salmon-sperm DNA, 6% polyethylene glycol and 0.1% SDS. The filter was washed in 1 × SSC buffer, 0.1% SDS at 65 °C and autoradiography was for 6 days at −70 °C. Conditions for re-probing the filter with β-actin were as above but the filter was washed in 0.1 × SSC buffer, 0.1% SDS at 65 °C and autoradiography was for 8 h.

not known to regulate specific gene sequences but are associated with regions of transcriptionally active chromatin. Within HMG1 and HMG2 there is a motif, the HMG box, which has been found in the non-histone chromosomal protein NHP6 of *Saccharomyces cerevisiae*[22], the yeast ARS-binding protein, ABF2, and the human nucleolar transcription factor hUBF (human upstream binding factor)[21]. The *hUBF* product is an RNA polymerase I transcription factor that interacts with sequence-specific DNA regions. This motif might represent a novel class of DNA-binding domains[21]. The conserved binding motif seems to be present in a large family of related sequences perhaps originating from an early HMG-like nonspecific DNA-binding structure.

It is tempting to speculate that Mc is also a transcription factor and that the conserved motif shared with the human pY53.3 sequence is a DNA-binding domain. The only structural evidence to support this conjecture is the high Arg–Lys content of both Mc and the conserved motif of pY53.3, which is 25% Arg–Lys.

We have termed the human Y-located gene defined by pY53.3 *SRY* (sex-determining region Y). The gene has been defined with respect to its location because only the homology to other genes suggests a DNA-binding function. But the putative nucleic-acid-binding motif within *SRY* and its testis-specific expression are consistent with *SRY* having a role in the developmental regulation of the testis.

The presence of a 3′ stop codon and a poly(A) tract in cDNA from 3′ RACE PCR implies that the open reading frame in pY53.3 corresponds to the last exon of *SRY*. At the 5′ end of the conserved motif sequence in *SRY* there is a potential splice acceptor site; but as homology between the rabbit and human genomic sequences continues past this site it may not represent an intron–exon boundary. This question will be resolved by the isolation and sequencing of transcripts from the *SRY* gene.

The 35 kb of Y-specific sequences immediately adjacent to the pseudoautosomal boundary are rich in repetitive sequences and this has hampered analysis. The open reading frame in pY53.3 is the only well-conserved sequence detected; but this approach could have failed to detect small exons, especially if they are close to repetitive elements or are only weakly conserved between species. Therefore, we cannot exclude the existence of genes other than *SRY* in the human sex-determining region.

We have described a novel, transcribed gene, *SRY*, present in the sex-determining region of man and mouse. Sequences homologous to *SRY* are located on the Y chromosome in all eutherian mammals tested. *SRY* encodes a protein with a potential DNA-binding domain, which is shared with the Mc protein of the mating-type locus of *S. pombe* and several non-histone proteins related to HMG1 and HMG2. *SRY* is currently the best candidate for *TDF*, the male sex-determining gene in humans. Proof of identity between *SRY* and *TDF* will require mutational analysis of XY females or the production of sex-reversed transgenic mice. □

Received 11 June; accepted 22 June 1990

1. Goodfellow, P. N. & Darling, S. M. *Development* **102,** 251–258 (1988).
2. Jost, A., Vigier, B., Prepin, J. & Perchellet, J. P. *Recent Prog. Horm. Res.* **29,** 1–41 (1973).
3. Vergnaud, G. *et al. Am. J. hum. Genet.* **38,** 109–124 (1986).
4. Guellaen, G. *et al. Nature* **307,** 172–173 (1984).
5. Goodfellow, P. J., Darling, S. M., Thomas, N. S. & Goodfellow, P. N. *Science* **234,** 740–743 (1986).
6. Weissenbach, J., Levilliers, J., Petit, C., Rouyer, F. & Simmler, M.-C. *Development* **101S,** 67–74 (1987).
7. Pritchard, C. A., Goodfellow, P. J. & Goodfellow, P. N. *Nature* **328,** 273–275 (1987).
8. Ferguson-Smith, M. A. *Lancet* **ii,** 475–476 (1966).
9. Petit, C. *et al. Cell* **49,** 595–602 (1987).
10. Page, D. C. *et al. Cell* **51,** 1091–1104 (1987).
11. Palmer, M. S., Berta, P., Sinclair, A. H., Pym, B. & Goodfellow, P. N. *Proc. natn. Acad. Sci. U.S.A.* **87,** 1681–1685 (1990).
12. Schneider-Gädicke, A., Beer-Romero, P., Brown, L.G., Nussbaum, R. & Page, D. C. *Cell* **57,** 1247–1258 (1989).
13. Sinclair, A. H. *et al. Nature* **336,** 780–782 (1988).
14. Koopman, P., Gubbay, J., Collignon, J. & Lovell-Badge, R. *Nature* **342,** 940–942 (1989).
15. Palmer, M. S. *et al. Nature* **342,** 937–939 (1989).
16. Ellis, N. A. *et al. Nature* **337,** 81–84 (1989).
17. Sealey, P. G., Whittaker, P. A. & Southern, E. M. *Nucleic Acids Res.* **13,** 1905–1922 (1985).
18. Smith, T. F. & Waterman, M. S. *J. molec. Biol.* **147,** 195–197 (1981).
19. Collins, J. F., Coulson, A. F. W. & Lyall, A. *CABIOS* **4,** 67–71 (1988).
20. Kelly, M., Burke, J., Smith, M., Klar, A. & Beach, D. *EMBO J.* **7,** 1537–1547 (1988).
21. Jantzen, H.-M., Adman, A., Bell, S. P. & Tjian, R. *Nature* **344,** 830–836 (1990).
22. Kolodrubetz, D. & Burgum, A. *J. biol. Chem.* **265,** 3234–3239 (1990).
23. Frohman, M. A., Dush, M. K. & Martin, G. R. *Proc. natn. Acad. Sci. U.S.A.* **85,** 8998–9002 (1988).
24. Gubbay, J. *et al. Nature* **346,** 245–250 (1990).
25. McLaren, A., Simpson, E., Tomonari, K., Chandler, P. & Hogg, H. *Nature* **312,** 552–555 (1984).
26. Lovell-Badge, R. & Robertson, E. *Development* **109,** 635–646 (1990).
27. Wright, J. M. & Dixon, G. H. *Biochemistry* **27,** 576–581 (1988).
28. Goodfellow, P. J. *et al. Ann. hum. Genet.* **53,** 15–22 (1989).
29. DeKretser, T. A., Crumpton, M. J., Bodmer, J. G. & Bodmer, W. F. *Eur. J. Immun.* **12,** 600–606 (1982).
30. Bishop, C. E. *et al. Nature* **303,** 831–832 (1983).
31. Westerveld, A., Visser, R. P. L. S., Meera Khan, P. & Bootsma, D. *Nature new Biol.* **234,** 20–24 (1971).
32. Goss, S. J. & Harris, H. *Nature* **255,** 680–683 (1975).
33. Burk, R. D., Ma, P. & Smith, K. D. *Molec. cell. Biol.* **5,** 576–581 (1985).
34. Sanger, F., Nicklen, S. & Coulson, A. R. *Proc. natn. Acad. Sci. U.S.A.* **74,** 5463–5467 (1977).
35. Chomczynski, P. & Sacchi, N. *Analyt. Biochem.* **162,** 156–159 (1987).

ACKNOWLEDGEMENTS. We thank M. Fellous for XX male DNA, I. Pickford for providing tissue samples, I. Goldsmith for preparing oligonucleotides and C. Middlemiss for typing this manuscript. A UK patent is pending on the *SRY* sequence.