Characterization of a murine gene expressed from the inactive X chromosome

Giuseppe Borsani, Rossana Tonlorenzi*, M. Christine Simmler†, Luisa Dandolo†, Danielle Arnaud†, Valeria Capra*, Markus Grompe, Antonio Pizzuti, Donna Muzny, Charles Lawrence‡, Huntington F. Willard§, Philip Avner† & Andrea Ballabio||

Institute for Molecular Genetics and ‡ Department of Cell Biology, Baylor College of Medicine, Houston, Texas 77030, USA
† Unite de Genetique Moleculaire Marine, Institut Pasteur, 25 rue du Docteur Roux, F-75724 Paris Cedex 15, France
§ Department of Genetics, Stanford University School of Medicine, Stanford, California 94305, USA

IN mammals, equal dosage of gene products encoded by the X chromosome in male and female cells is achieved by X inactivation. Although X-chromosome inactivation represents the most extensive example known of long range *cis* gene regulation, the mechanism by which thousands of genes on only one of a pair of identical chromosomes are turned off is poorly understood. We have recently identified a human gene (*XIST*) exclusively expressed from the inactive X chromosome[1]. Here we report the isolation and characterization of its murine homologue (*Xist*) which localizes to the mouse X inactivation centre region and is the first murine gene found to be expressed from the inactive X chromosome. Nucleotide sequence analysis indicates that *Xist* may be associated with a protein product. The similar map positions and expression patterns for *Xist* in mouse and man suggest that this gene may have a role in X inactivation.

There is genetic evidence that a single locus, the X inactivation centre (*XIC*) in man and the X controlling element (*Xce*) in mouse, is required in *cis* for inactivation to occur[2-5]. The chromosomal location of this locus in band Xq13 in man and bands D/E in the mouse X chromosome has recently been refined[6-9]. When the human *XIST* complementary DNA clone 14A (ref. 1) was hybridized under conditions of reduced stringency to DNAs of mice from an extensively typed interspecific backcross panel[10,11], restriction fragment length polymorphisms allowed localization of the corresponding locus to the central region of the mouse X chromosome in which *Xce* resides. As shown in Fig. 1*a* and *c*, the recombination breakpoints of the X chromosome in backcross animals 74, 114 and 194 establish the murine *Xist* locus as distal to *Ccg-1/Phka*, and those of animals 61 and 172 establish it as lying proximal to *Pgk-1*, consistent with the order of homologous loci in the human (*CCGl-XIST-PGKl*)[6]. Genetic analysis gives a recombination distance of 1.2+/−0.8 centimorgans between *Xist* and the *Ccg-1/Phka* loci (250 animals tested).

To isolate the murine *Xist* homologue, we hybridized the human *XIST* cDNA clone 14A to a cDNA library in λZAPII from thymus RNA of female mice (Stratagene). The longest of the isolated cDNA clones, MR20 (3.1 kilobases (kb)), gave results on the backcross panel indistinguishable from those found for the X-linked sequences detected with the human probe 14A (Fig. 1*b* and *c*) and consequently mapped it between the *Ccg-1/Phka* and *Pgk-1* loci. Independent confirmation of the mapping of the *Xist* locus to this region of the mouse X chromosome has been obtained from a series of somatic cell hybrids carrying various deletions of the mouse X chromosome[11] and from the use of a large panel of irradiation fusion somatic cell hybrids obtained after 50 krad irradiation[12] (data not shown). Combined analysis of the deletion and meiotic mapping data establishes the order: centromere-*Ccg-1*-*Phka*-*Xist*-DXPas19-*Pgk-1*-telomere.

Expression of the *Xist* gene has been studied by polymerase chain reaction amplification of reverse-transcribed RNA

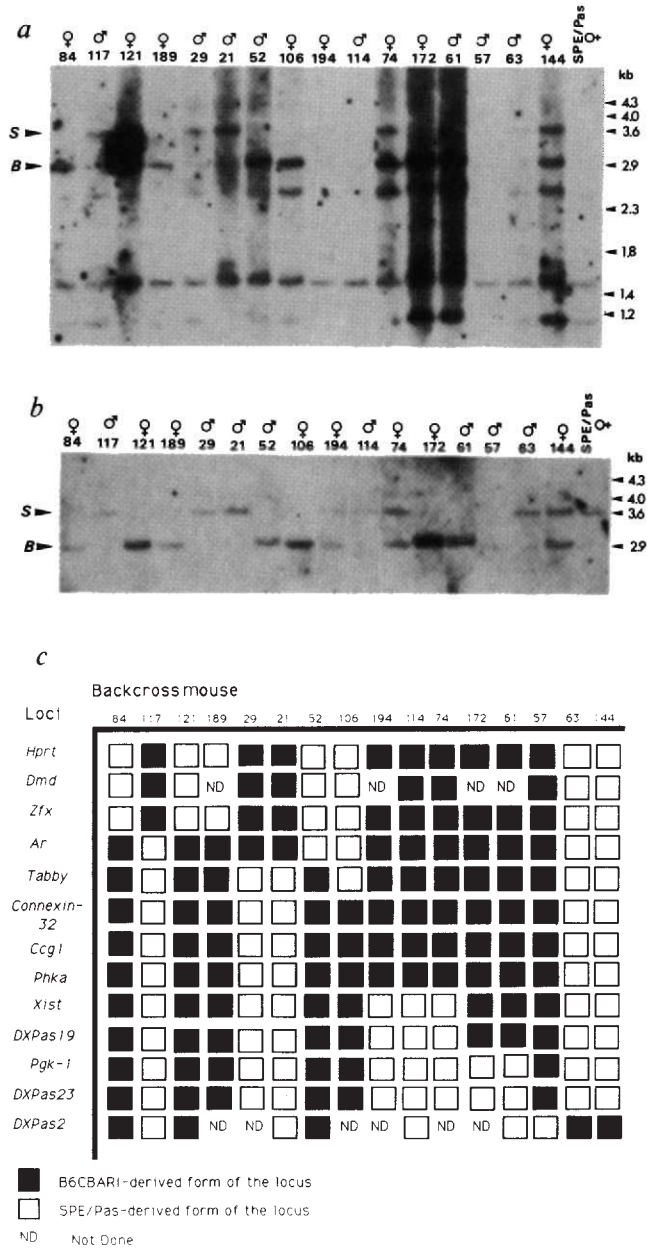


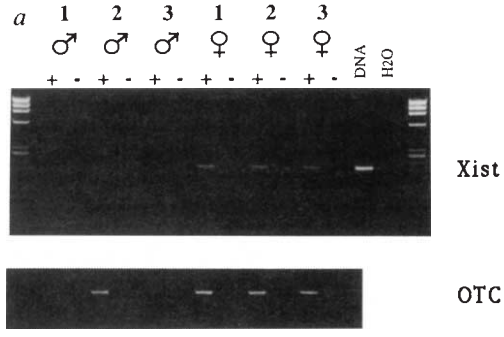
FIG. 1 Meiotic mapping of the *Xist* gene. *a*, DNAs from animals of an interspecific backcross, involving B6CBARI and inbred mouse strain SPE/Pas derived from *Mus spretus*, were hybridized to the human *XIST* cDNA 14A and the mouse *Xist* cDNA MR20. Under the conditions of reduced hybridization stringency used in the experiments with the human probe 14A, additional weak bands that do not apparently localize to the X chromosome are also observed. *b*, Hybridization pattern of the same panel with the mouse cDNA probe MR20. *c*, Pedigree analysis of individual backcross progeny from the *Mus spretus* backcross showing the position of the *Xist* locus. *S*, SPE/Pas-derived alleles; *B*, B6CBARI-derived alleles.
METHODS. For Southern blotting, the human probe 14A was hybridized in sodium phosphate-SDS buffer at 65 °C according to the method of Church and Gilbert[17], followed by washing with 5 × SSC at 50 °C. Mouse probe MR20 was hybridized under identical conditions but washed with 2 × SSC at 68 °C. All probes were labelled by random priming[18] and blots were exposed after hybridization in the presence of intensifying screens for various periods.

* On leave from G. Gaslini Institute, Genoa, Italy.
|| To whom correspondence should be addressed.

(RT-PCR). Figure 2*a* shows the RT-PCR amplification of liver RNA samples from three male and three female mice, using oligonucleotide primers derived from the mouse *Xist* cDNA sequence. Only samples corresponding to female RNAs gave any product, indicating that *Xist*, like its human homologue, is exclusively expressed from the inactive X chromosome. Levels of *Xist* RNA in XO and XY mouse liver were ~1,000-fold lower than those detected in the XX female samples containing an inactive X chromosome (Fig. 2*b*). The decreased transcription in our XO female samples, in which the X chromosome is of paternal origin, rules out the possibility that *Xist* transcriptional activity is controlled by paternal/maternal imprinting. Northern blot analysis of RNAs from liver samples of male and female mice gave a continuous smear only in the lanes corresponding to female samples (data not shown), as observed for the human *XIST* gene[1].

Figure 3*a* shows the PCR amplification of reverse-transcribed RNA, genomic DNA, cosmid clone MB4-14 (spanning the entire 3.1 kb of sequence contained in the cDNA clone MR20) and of clone MR20. The results demonstrate complete colinearity between the cDNA and the genomic DNA sequence, indicating that clone MR20 is a 3.1-kb unspliced transcript. Twenty-two additional *Xist* cDNA clones were identified using MR20 as a probe and characterized by restriction mapping, PCR amplification and hybridization of oligonucleotides designed from the sequence of clone MR20. The hybridization of nine oligonucleotides regularly spaced throughout the sequence of clone MR20 indicates the absence of alternative splicing, a conclusion based on the comparison of the length of each clone with its hybridization pattern (Fig. 3*b*). These findings differ from those for the human gene, in which an extremely high frequency of alternative splicing is observed in the equivalent region[1], and establish a clear difference in the splicing patterns of the human and mouse *Xist* genes.

Figure 4*a* shows the nucleotide sequence of 4.1 kb of the *Xist* gene (deposited in the GenBank/EMBL Data Library, accession number X59289). Sequence analysis revealed an open reading frame 894 base pairs (bp) long located at the 5′ end of the portion of the *Xist* gene characterized so far. The corresponding amino-acid sequence is rich in leucine and cysteine (25% and 12%, respectively). The putative protein is extremely hydrophobic, so possibly represents a membrane-bound domain. Analysis of the sequence revealed the presence of a 24-bp motif repeated eight times in the first 270 bp of the DNA sequence, reflected in a repeat of eight amino acids in the predicted protein (Fig. 4*a*). Screening of the GenBank sequence database using the *Xist* nucleotide sequence and of the Swiss-protein, PIR and GenPept databases using the predicted amino-acid sequence did not show significant similarity to any other sequence. We have also looked specifically for the repeat motif in the databases and have found no similar sequence. Although the long open reading frame strongly suggests that an *Xist* protein product exists, further studies are needed to provide a definitive answer on the protein-encoding capability of the *Xist* gene. The murine *Xist* sequence was compared with the longest unspliced stretch of cDNA sequence from the human *XIST* gene (J. Rupert, C. Brown and H. F. W., unpublished data). A homologous region of ~500 bp was found in a region corresponding to bases



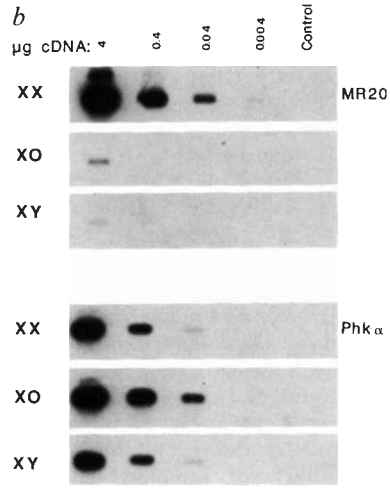
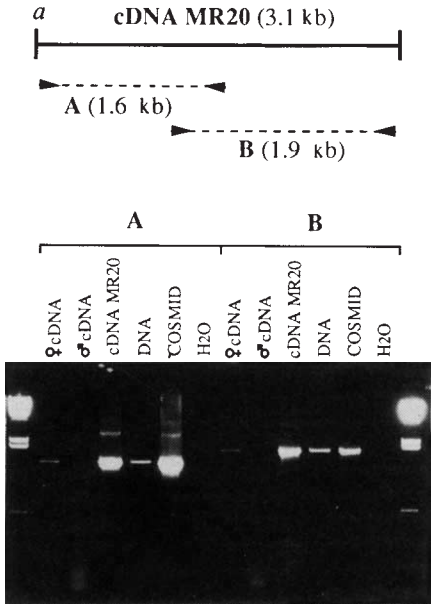


FIG. 2 Expression studies of the murine *Xist* gene. *a*, RT-PCR analysis of *Xist* expression in male and female mice. Reverse-transcribed liver RNA from six different mice (three males and three females) and mouse genomic DNA (lane 'DNA') were amplified using a set of primers from cDNA clone MR20 sequence. Plus and minus signs indicate the presence or absence respectively of reverse transcriptase in the reaction. *Phi*X-*Hae*III digested DNA has been used for size calibration (outer lanes). An amplification product of the expected size (203 bp) can only be detected in female RT^+ samples and in mouse genomic DNA with the PCR conditions used. A PCR with primers from the OTC cDNA sequence[19] was performed as a positive control for the quality of the cDNAs and is shown at the bottom. Absence of amplification in the RT samples rules out possible contamination of the samples with genomic DNA. 'H_2O' lane shows PCR without template DNA as a test for PCR contamination. *b*, Quantitative analysis of *Xist* expression in XX, XO and XY samples. Serial dilutions of cDNA obtained from liver RNA of XX, XO and XY mice were amplified using a set of primers from the 5′ end of the *Xist* gene. The PCR products (251 bp) were then transferred to a nylon membrane and hybridized with probe MR20. As an internal control we used a set of primers, amplifying a 240-bp fragment from the *Phka* gene, and a *Phka* cDNA probe[20].

METHODS. RT-PCR: total cellular RNA was prepared from liver of adult male and female mice by the guanidium thiocyanate method[21] and 5 μg were reverse-transcribed with murine Moloney leukaemia virus reverse transcriptase using random hexamer primers[22]. One-fiftieth of the cDNA was used for a PCR[23] amplification with 0.5 μmol primers for 40 cycles (94 °C for 1 min; 55 °C for 1 min; 72 °C for 1 min) with Cetus *Taq* polymerase and buffer. Primers used were: 594-1R (5′ GGGACCTAACTGTTGGCTTTATCAG 3′) and 594-1F (5′ GAAGTGAATTGAAGTTTTGGTCTAG 3′). Quantitative RT-PCR: 20 μg total liver RNA, isolated using the procedure described above, were treated with RQ1 DNase (Promega) before extraction with phenol and precipitation with ethanol. The RNA was then reverse-transcribed with 10 units AMV reverse transcriptase and 2 μg random hexamer primers[22]. Samples were treated with DNAse-free RNase (10 μg) and precipitated with ethanol. For each RNA sample, a control experiment was carried out in parallel omitting the reverse transcriptase. Samples containing from 4 μg to 0.004 μg of cDNA were amplified for 20 cycles using Cetus *Taq* polymerase and 1 μmol of the primers (94 °C for 1 min; 54 °C for 1 min; 72 °C for 1 min) in a total volume of 50 μl. *Xist* primers used: 588-2F (5′ ATCTAAGACAAAATACATCATTCCG 3′) and 588-2R (5′ CTTGGACTTAGCTCAGGTTTTGTGTC 3′). *Phka* primers[20]: A (5′ AATTCACTACTGCCCAGGGCTTCAAC 3′) and B (5′ GCTTCAGCTCAGCTGGGTTATAGTAT 3′). Thirty μl of each reaction product were electrophoresed on a 2% agarose gel, transferred to Hybond N+ (Amersham) and hybridized to random-primed MR20 *Xist* cDNA probe. The exposure time was 4 days, without intensifying screens.



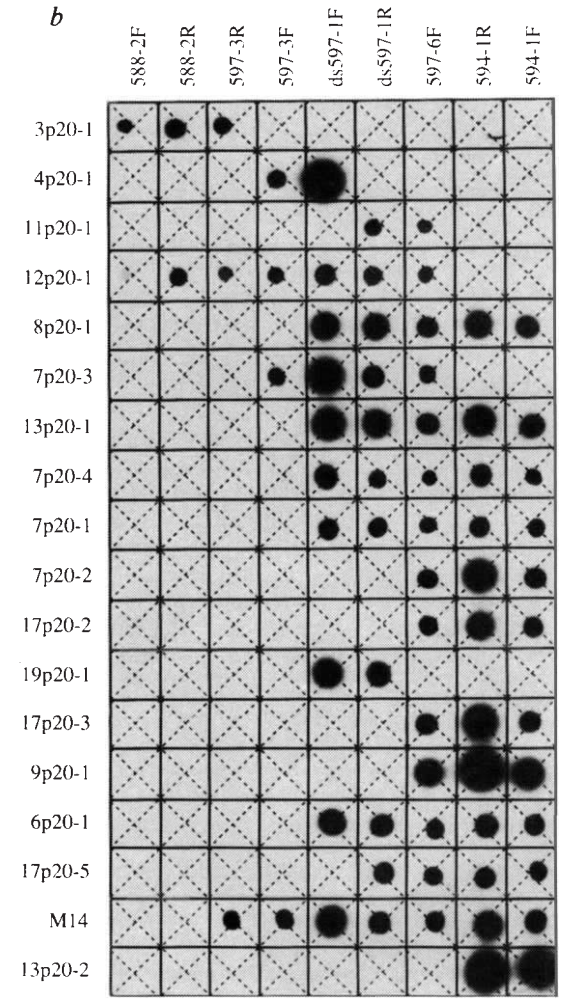


FIG. 3 *Xist* gene structure. *a*, PCR analysis of cDNA and genomic DNA (lane 'DNA') samples corresponding to the *Xist* gene. Two sets of oligonucleotide primers, A (588-2F + ds597-1R) and B (594-1F + ds597-1F), amplifying two overlapping segments from each end of clone MR20, were used. The position of each primer within clone MR20 sequence is shown in the scheme. A molecular size marker (λ DNA digested with *Hind*III; outside lanes) was loaded on the gel. Clone MB4-14 is one of seven positive clones obtained from the screening of a mouse genomic library in the cosmid vector pCOS2EMBL[24] using clone MR20 as a probe. 'H_2O' lane shows PCR without template DNA to test for PCR contamination. *b*, Characterization of *Xist* transcripts. Clone MR20 was used as a probe to rescreen the same cDNA library from which it was derived. Twenty-two more cDNA clones were identified and plaque-purified. Insert DNA from 18 cDNA clones was obtained by T7–T3 PCR amplification and spotted on a filter with a numbered grid. The name of each cDNA clone is indicated on the left. Vertical strips of the filter were cut and each strip was hybridized to an oligonucleotide located in a different position in the clone MR20 sequence. The nine oligonucleotide probes are indicated at the top in 5′–3′ order.

METHODS. 6×10^5 plaque-forming units of a λZAPII cDNA library were screened using the human *XIST* cDNA clone 14A (previously isolated by the immunoscreening of a λGT11 cDNA library using anti-STS antibodies[25]). Hybridization conditions were 65 °C in 1% SDS, 1M NaCl and 5% dextran sulphate. Washings were for 3 × 30 min at 65 °C in 3 × SSC, 0.05% SDS. The *Xist* cDNA clone MR20 was used as probe to rescreen the same library using identical conditions, except that the washings were at 65 °C in 0.2 × SSC. The latter conditions were also used for the screening of a mouse cosmid library (6×10^5 colony-forming units) in vector pCOS2EMBL with probe MR20. Two sets of oligonucleotide primers amplifying two overlapping segments from each end of clone MR20, were used for PCR analysis. Primers were 588-2F (5′ ATCTAAGACAAAATACATCATTCCG 3′) + ds597-1R (5′ CACTGACTTGGAAGTTACAGTAGGC 3′) and 594-1F (5′ GAAGTGAATTGAAGTTTTGGTCTAG 3′) + ds597-1F (5′ GGTTCATGGTCTTAGTTGCATATTCC 3′). PCR amplification was with 0.5 μmol primers for 40 cycles (94 °C for 1 min; 54 °C for 1 min; 72 °C for 6 min) with Cetus *Taq* polymerase and buffer. Oligonucleotide hybridization and washings: insert DNA from the cDNA clones was amplified using T3 and T7 primers (36 cycles of 94 °C for 1 min; 50 °C for 1 min; 72 °C for 6 min). 3 μl of a 1 in 10 dilution of the PCR products from each clone were spotted on a Genescreen Plus (NEN-Dupont) membrane. The membrane was then denatured in 0.4 M NaOH, neutralized in 0.2 M Tris, pH 7.5, 2 × SSC and baked at 80 °C. The nine oligonucleotides used in the analysis were: 588-2F, 588-2R, 597-3R (5′ GTAATTTTTCATTGTTGCCATTGCCC 3′), 597-3F (5′ CTGACATTGTTTTCCCCCTAACAAC 3′), ds597-1F, ds597-1R, 597-6F (5′ TTAACTGTACTCTTCCCATTATGGG 3′), 594-1R and 594-1F (see above). 100 ng of each oligonucleotide were labelled using 10 μCi of [γ-^{32}P]ATP and T4 polynucleotide kinase (Pharmacia). Hybridization was for 8 h at 43 °C in 1% SDS, 1M NaCl, 5% dextran sulphate; washing was at 42 °C in 2 × SSC, 0.1% SDS. Filters were exposed overnight with intensifying screens.

1,500–2,000 of the mouse sequence. This region contains several subregions with identity between mouse and human sequence of >80%. PCR primers designed from one of these subregions gave PCR products from feline and bovine DNA which were then subcloned and sequenced. Figure 4*b* shows the alignment between mouse, man, feline and bovine sequences. Homology with other species, including rabbit, chimpanzee, wallaby, chicken and carp, was also found by low-stringency hybridization using both the human and the mouse probes (data not shown).

There is evidence that genes which escape X inactivation in man, such as *STS*[13], *ZFX*[14] and *RPSX4*[15], are inactivated in the mouse (ref. 16, and A. Ashworth, personal communication), suggesting that there are important differences in X inactivation between man and mouse. *Xist* is the only gene known to be expressed from the murine inactive X chromosome and the only one among those also expressed from the human inactive X chromosome to show an identical expression pattern in the two species. *Xist* therefore shows novel features compared with other murine genes studied so far. Like its human homologue[6], the murine *Xist* gene localizes to the critical mapping interval to which the *Xce* locus required (in *cis*) for X inactivation has been assigned, and its expression is female-specific in chromosomally normal individuals. The conservation of these features between two species as distantly related as man and mouse, suggests a functional role for this gene in female mammals generally. Besides its putative role in X inactivation, the isolation of the *XIST* gene, first in man and now in the mouse, offers an ideal molecular tool for the study of X inactivation and a starting point for the characterization of the X inactivation centre region in the two species. □

a

```
1     TTTGTGTGTCTATTTCTTCCTTGCATTGTGTCTAATTCTTTGGTATATATATTTCTTCAT
1     PheValCysLeuPheLeuProCysIleValSerAsnSerLeuValTyrIlePheLeuHis

61    TGCTTTGTGTGTCTATGTCTCCTTGTGTTGTCTAATTCGTTGTTGCATCTATTTCTTCCT
21    CysPheValCysLeuCysLeuLeuValLeuSerAsnSerLeuHisLeuPheLeuPro

121   TGCTTTGTGTGTCTATTTCTTCCTTGCTTTGTGTGTCTATGTCTTCCTTGCTTTGTGTGT
41    CysPheValCysLeuPheLeuProCysPheValCysLeuCysLeuProCysPheValCys

181   CTATGTCTTCCTTGTTTTGTGTATCTACTTCTTCCTTGTGTGTCTAATTCTTTGTTACAT
61    LeuCysLeuProCysPheValTyrLeuLeuLeuProCysValSerAsnSerLeuLeuHis

241   CTATTTCTTCCTTCCTTTGCATGTCTCCTTCTTTCCTTTGTGTGTCTTTTCTGTCTGCAG
81    LeuPheLeuProSerPheAlaCysLeuLeuLeuSerPheValCysLeuPheCysLeuGln

301   TGTGTCTTACCTATTCCCATGTTTCTCCTGCATGTTCTTTCTTGCAGAGCTTTGAGCTTT
101   CysValLeuProIleProMetPheLeuLeuHisValLeuSerCysArgAlaLeuSerPhe

361   GTTTCACTTTCTCTGGTGCCTGTGTGGTCTGCTTTGTCTTCACTAGCTATGGCTCTCTGT
121   ValSerLeuSerLeuValProValTrpSerAlaLeuSerSerLeuAlaMetAlaLeuCys

421   TTTATCTATCTGGTTGCTATTTCTCTTAGCTTTTCTTTCACTCCTGCCTTTCGTGACTCC
141   PheIleTyrLeuValAlaIleSerLeuSerPheSerPheThrProAlaPheArgAspSer

481   CTTTGGGTCACATGTTGCATGCATCCCTCTCTTTTTCTTGTGCTCACCCCACTTGTTCTT
161   LeuTrpValThrCysCysMetHisProSerLeuPheLeuValLeuThrProLeuValLeu

541   TGTTCAAGTTCTCTTTGTCAGTCCATTTCAGTTTTCTTTCTGCTGCTTCTATCCTTAGTG
181   CysSerSerSerLeuCysGlnSerIleSerValPhePheLeuLeuLeuSerLeuVal

601   AATTCTTGTTTACATTTCTTCCCTGCCTTTCTTGGGCCACTTTCTCTGTTTTCTTTTGTA
201   AsnSerCysLeuHisPhePheProAlaPheLeuGlyProLeuSerLeuPheSerPheVal

661   TTTGTGTCTCTTTGCTATTGGTGGATTTCTTATCTCAGCATCATTCTGTTGCTTTGTGTT
221   PheValSerLeuCysTyrTrpTrpIleSerTyrLeuSerIleIleLeuLeuLeuCysVal

721   TGCTTGTGTTTCTATCTTCTACTTTCCTCCTTTCTGTTCACTTTGAGCATTTCATCTCTT
241   CysLeuCysPheTyrLeuLeuLeuSerSerPheLeuPheThrLeuSerIleSerSerLeu

781   TACAAGTCTGTGTCTCTCTTGTATTCTAAAGTAATCCTTTCTTGGATGTTTCTTTGTATG
261   TyrLysSerValSerLeuLeuTyrSerLysValIleLeuSerTrpMetPheLeuCysMet

841   TACATGTGCGTGTGTGCATGTGTGTTATGTGTGTCATGTGTGAGAGGAGCTTCATAGCCC
281   TyrMetCysValCysAlaCysValLeuCysValSerCysValArgGlyAlaSer

901   CTTCCCAATAGGTCCAGAATGTCACCCGTGGAGCCGTTCCTCACACCAGACTGCCCTGAG
961   AAATAATCTAAGACAAAATACATCATTCCGTCCGGTCAGGATTCAAGTGGCTCTGAAGTG
1021  AACGCCCAAGTAGAAGACAGAAGTTTTGCGACTTGAGATTTAAAAGGACCAAAATACACA
1081  GATGGCCCGTCTTGAGCTGGCTGGACAGAATGCTGACAACCCAAAGAAGAGGAACTGTTT
1141  CTACAGGACACCTGTGACTTCCAAGAGCGGGGAACTACGTATGTCATAAGACACAAAACC
1201  TGAGCTAAGTCCAAGCATAAGACCTAAGGACCCAATCCTATATGGACAGAATATTTAAGA
1261  GATAAAGGCCTATGGCCCAGAACTCTGGAAGGATATTTCTATCCTTCTATCCCCAAGACC
1321  AAGAAGGGAAATTCGAAGATGAGACCTGCCCCCCAACCCCAGCATCCCTTTCCATTTCTT
1381  ATATTTCTATTTAAGCTGTCTTCACTTGAGATGTAATTTTTCATTGTTGCCATTGCCCAT
1441  AAAGGAATACGTTTTTAGCTGGATAGTATTGTGCAAGGGTCTGTTTTAAACTGGGTCTTA
1501  GCCATTTGTTAAATTGTTGATGTTTTACAACTTCCATTTCTCTTCACATCTGCTCCACTT
1561  GAGACGGAACTAAATCCAGCCAGTGTATATAGCCTGACTATTGAAACTTCCCTAGGAATA
1621  AGCATGCATACAGATATGCATACTGCCATCCTCCCTACCTCAGAAGCCCTAGGCTGACAA
1681  GAAAAGGAAAGCATCAGGTTGTTAGGGGGAAAACAATGTCAGGCTATCTAGAGAAAATAT
1741  AAAGAGTTGTTCCAGACCAATGAGAAGAATTAGACAAGCAATATGCAGATGTGCCAACCC
1801  TCTGAGAAGCACCAGCCAGTGTCACCTTCTTTCTTTGGGCTTAGGTGAGCAGGGTATGGT
1861  TTTCTAATAATGGTTTGGGGACAAAATGAGGTCTGAACTCCCTGCTCATAGTAGTGGCCG
1921  AGTAATTTGGTGCATTTCACCAAAGGAACTCCTGGGTCTAATACCTACCTTTAAAATTAA
1981  TGATGAGAGACTCTAAGGACTACTTAACGGGCTTAATCTTTTTCGTGCCTTCCTCTTCCT
2041  CTGTAAGAGGGAAGTTAAATGACACAGGATGAAAAAGTAACATGCTCATAGCACATTGGC
2101  AATTATACATGGTTATTATCTGAAAGTGTAGAGCTTTTCCTATAAGGCATCAGACTAAGT
2161  ACCTGAAGCTTTGTGGGTTCATGGTCTTAGTTGCATATTCCTTAGTTGCAAATCCTTTTC
2221  AAAAGGTAAGAAAAAGGCACACTGGTCTATTGCCTGTACTTGATCAAGCCCTGATATGAA
2281  TGCCAGGGAATGTCTGAGTAACATTAATTTCCTTCCCTGCATATTTTTTGTGCTGAATAC
2341  TAAGGCTGTGATGCTTCACTGTGGTCACCCCCAGGTAACAAGATATTACCAGGTAACCAG
2401  GAAACGTATGAATACGTAAACCATGAAGCCTACTGTAACTTCCAAGTCAGTGCTGAGTAT
2461  GTATTACATAGTAGCTAAAGTCTACGCCTCTGTGTGCTATAGGCACAAAGATTGCTCTAG
2521  GAATAACATGCTTTGTAAAAACAAATATATGAACATAACGGGGCTTGAATGAATAACAGT
2581  CCATATACTTAAGGCCAGTGTGTTTCTTCTGCTTTGGTGAGGCTCAGTAAGTTATATTAT
2641  ACCAGGTAGCAGAAGAGAAAACACATGGAAACTGATTTTAAACTACAAACTAGGTCACTA
2701  ATGCAGGTGATTGATTACCCTATTCTGATCACCTTCTAATTTCTGAATACCCATGTTCAG
2761  CACTGGGAATAACAAAGGGGGACATTACCACAGAACTAGAATTTACAAAAGAATGCATTA
2821  AATAAAGCATTATACAGCTATCAATTGTTCCATGTGTGCAAATGAATGACTACTAACTAC
2881  CTCTGATGTATCCGATATTGTTTTGGGTACATGAAATATTCATGAGTAACTGCCATGAAA
2941  TAAGAATGTTTGCATTCCATACTATTCATAAGGAATGAGCCAATGCTTAATTTAATCAGT
3001  CAAAACTTGAGTGATAAGGGCATGTTAATACAAGAACATTTGCCCAGGTCACATTATGGT
3061  TGTGGGTACTTTCTTAACTATAAAGCAGTTCAGTAGTATAAGACAAGACAAATTTTCTAT
3121  AGAAATAAAGCTGCCTATAAAATAGGCATAGTCTCTACAAAATTTTCATTGTACTTTTTA
3181  GCCCATAATGGGAAGAGTACAGTTAACAAGCTGGGTGTGGTAGCATGTGCTCTGAGCTGA
3241  AGCAACAGGACCACTTGAGCCCAGAAATTGGAGGCTAGCCTGGGAAGACCATAAGGTCAA
3301  TCTCAAACCTGGAGGCTAAATATTGTCTCCATGTGTATATTCTCTTTCATGGGTACTGG
3361  AGAGATACACAGACGTACATTTCAGTGTGTCCACACTTGAGAATAATATGTACGTTGGCA
3421  TTTTATGAACTCGGAGGTACCATATAAATGTAACAATTCATTTTCTTACTTGGTATCAAT
3481  TTCCAGGCTTTAAAATTCTGCCACATTTATTATACTGTGAAAATAAAGTAAATAAGTAA
3541  CTGTGAACCACTGAATATATGAAGCATTCAATACTTGATGAGTACATACTGAATGGCAGT
3601  CATTTATTACAAAACAGTGCCCTTGCTAGGCACTGGGATGCAAAGAGCATTCTCATTGTC
3661  CTGTGTATCTAAAGAAATTATGCATGAGATTAATTTATAATTTGTAAACTGCCATATATA
3721  TGTGTATATATGCAATATTTGCCTGGTGTGCAATGACTTTGCTTTTATCCCAGGCATGCA
3781  CAACAGATCTGTGTGGAGCTTTGTGAAGTCTACAGTTCTATAAAGCCGGGACCTAACTGT
3841  TGGCTTTATCAGTGAACAGTGATTACTTTCTAAGTTTCATAATGGCTGAAACTTAATCAT
3901  AATGCTTATCACCTAACACCACCTAATAATAATTTTACCATGCTATGTGTTGAGCGAACA
3961  CATAGATTGCTTTCTAGCATTATGTAGCACTTATAGGAGTGAAATCTAGACCAAAACTTC
4021  AATTCACTTCAATGAGGAAATGAAAAAAAAAAAAAAA
```

b

```
mouxist : CGGAACT..AAATCCAGCCAGTGTATATAGCCTGACTATTGA.AACTTCCCT.AGGAATA
humxist : GA--CACca--------T----..--C--AT---G--T---Tt---------c----GC-
catxist : T---CA-....--.....................................-GGT
cowxist : A---CAC....--.....................................-GCT

mouxist : AGCATGCATACAGATATGCATACTGCCATCCTCCCTA..CCTCAGAAGCCCTAGGCTGAC
humxist : GA---T----T--G-G...------TAT-T-AGT-CTtt-T-TT--..---C--AAGCC-
catxist : -A---T.............G--T-C-AG--------Ttt---T----A-T----AA---T
cowxist : GA---T.............---T-T-AG--C-T--Tt.----------T---------T

mouxist : AAGAAAAGGAAAGCATCAGGTTG....TTAGGGGGAAA.....ACAATGTCAGGCTATCT
humxist : T---CTGA---GAT-AA-T-G-Caggt-GTT-------aa...-A-G--C------C---
catxist : G-------------------C--g.....T-A---GTGtggagGA-GA-AT---------
cowxist : G---G--------T--------CagttA-T-A---G--ag...-A-G--C--A-------

mouxist : AGAGAAAA.TATAAAGAGTTGTTCCAGA
humxist : --------a-G-G-----A--C-----G
catxist : --------.-G-G-----AC-------G
cowxist : --------.-G-G-----A--C----C-
```

FIG. 4 *Xist* sequence analysis and comparison with other species. *a*, Nucleotide sequence of *Xist* cDNA. The sequence has been derived mainly from cDNA clone MR20 (bases 975–4,057) and partly from a *Hin*dIII subclone of cosmid MB4-14 (1–974). RT–PCR of female RNA confirmed that the portion of the sequence deriving from genomic DNA is also transcribed (data not shown). The predicted amino-acid sequence from the 894-bp open reading frame located at the 5′ end of the sequence is also shown. Arrows indicate different units of the 24-bp repeat. *b*, Comparison between mouse, human, feline and bovine sequences. Oligonucleotide primers were designed from a region of high homology between the human and mouse sequences. These primers amplify a product in both human (257 bp) and mouse DNA (250 bp) and were used to amplify genomic DNA from cat and cow. A PCR product of 200 bp and of 203 bp was identified in cat and cow, respectively, and was subcloned and sequenced. The nucleotide sequences of this region in the four species are compared. Upper-case letters, aligned non-identical bases; lower-case letters, unaligned bases; dashed lines, aligned identical bases; dotted lines, gap.

METHODS. λ cDNA clones were subcloned in plasmids using the λZAPII plasmid rescue procedure according to the manufacturer's specifications (Stratagene). All nucleotide sequences were determined automatically using an Applied Biosystem ABI 370A fluorescent sequencer, as described by Gibbs *et al.*[26], and manually using a Sequenase sequencing kit (USB). Assembly of DNA sequences and search for open reading frames used the computer program DNA Strider[27]. The sequence databases were searched using a previously described program[28]. Multiple sequence alignment was according to ref. 29. Primers used for PCR were derived from the region of high homology and were TOP1 (5′ caggatccTTCACATCTTCTCCACTTGAGA 3′) and TOP2 (5′ caggatcCTTGTCTAATTCTTCTCATTGG 3′), the lower-case letters indicating a *Bam*HI cloning tail. Genomic DNA (500 μg) from each species was amplified. PCR conditions were eight cycles at 94 °C for 1 min; at 50 °C for 1 min; at 72 °C for 1 min, followed by 27 cycles at 94 °C for 1 min; 53 °C for 1 min; 72 °C for 1 min. PCR products were digested with *Bam*HI followed by phenol–chloroform extraction and precipitation with ethanol. The fragments were subcloned in M13mp18 vector predigested with *Bam*Hi, and subclones sequenced as described above.

Received 21 February; accepted 9 April 1991.

1. Brown, C. J. *et al. Nature* **349,** 38–44 (1991).
2. Gartler, S. M. & Riggs, A. D. *A. Rev. Genet.* **17,** 155–190 (1983).
3. Grant, S. G. & Chapman, V. M. *A. Rev. Genet.* **22,** 199–233 (1988).
4. Lyon, M. F. *Nature, new Biol.* **232,** 229–232 (1971).
5. Cattanach, B. M. *A. Rev. Genet.* **9,** 1–18 (1975).
6. Brown, C. J. *et al. Nature* **349,** 82–84 (1991).
7. Mitchell, M. D. *et al. Genetics* **121,** 803–809 (1989).
8. Keer, J. T. *et al. Genomics* **7,** 566–572 (1990).
9. Cattanach, B. M. *et al. Cytogenet. Cell Genet.* (in the press).
10. Amar, L. C. *et al. Genomics* **2,** 220–230 (1988).
11. Avner, P. R. *et al. Proc. natn. Acad. Sci. U.S.A.* **84,** 5330–5334 (1987).
12. Sefton, L., Arnaud, D., Goodfellow, P., Simmler, M. C. & Avner, P. *Mammalian Genome* (in the press).
13. Mohandas, T. *et al. Proc. natn. Acad. Sci. U.S.A.* **84,** 4954–4958 (1987).
14. Schneider-Gadicke, A., Beer-Romero, P., Brown, L. G., Nussbaum, R. & Page, D. C. *Cell* **57,** 1247–1258 (1989).
15. Fisher, E. M. C. *et al. Cell* **63,** 1205–1218 (1990).
16. Jones, J., Peters, J., Rasberry, C. & Cattanach, B. M. *Genet. Res.* **53,** 193–199 (1989).
17. Church, G. & Gilbert, W. *Proc. natn. Acad. Sci. U.S.A.* **81,** 1991–1995 (1984).
18. Feinberg, A. P. & Volgestein, B. *Analyt. Biochem.* **132,** 6–13 (1983).
19. Scherer, S. E., Veres, G. & Caskey, C. T. *Nucleic Acids Res.* **16,** 1593–1601 (1988).
20. Bender, P. K. & Lalley, P. A. *Proc. natn. Acad. Sci. U.S.A.* **86,** 9996–10000 (1989).
21. Chirgwin, J. M., Przybyla, A. E., MacDonald, R. J. & Rutter, W. J. *Biochemistry* **18,** 5294–5299 (1979).
22. Kawasaki, E. S. in *PCR Protocols. A Guide to Methods and Applications* (eds Innis, A., Golfond, D. H., Sninsley, J. J. & White, T. J.) (Academic, New York, 1990).
23. Saiki, R. K. *et al. Science* **239,** 487–491 (1988).
24. Ehrich, E., Craig, A., Poustka, A., Frischauf, A. M. & Lehrach, H. *Gene* **57,** 229–237 (1987).
25. Ballabio, A. *et al. Proc. natn. Acad. Sci. U.S.A.* **84,** 4519–4523 (1987).

26. Gibbs, R. A., Nguyen, P. N., McBride, L. J., Koepf, S. M. & Caskey, C. T. *Proc. natn. Acad. Sci. U.S.A.* **86**, 1919–1923 (1989).
27. Marck, C. *Nucleic Acids Res.* **16**, 1829–1836 (1988).
28. Altschul, S. F., Gish, W., Miller, W., Myers, E. W. & Lipman, D. J. *J. molec. Biol.* **215**, 403–410 (1990).
29. Feng, D. F. & Doolittle, R. F. *J. molec. Evol.* **25**, 351–360 (1987).

ACKNOWLEDGEMENTS. We thank D. Toniolo for the mouse cosmid library, C. Brown and J. Rupert for unpublished sequence data and D. Ledbetter and C. Chinault for critical reading of the manuscript. Sequence comparison was performed using the facilities of the Molecular Biology Information Resource, Baylor College of Medicine. The nucleotide sequence data reported will appear in the EMBL, GenBank and DDBJ Nucleotide Sequence Databases under the accession number X59289 MURINE X IST CDNA.

Conservation of position and exclusive expression of mouse *Xist* from the inactive X chromosome

Neil Brockdorff, Alan Ashworth*, Graham F. Kay, Penny Cooper, Sandy Smith, Veronica M. McCabe, Dominic P. Norris, Graeme D. Penny, Dipika Patel & Sohaila Rastan†

Section of Comparative Biology, MRC Clinical Research Centre, Harrow HA1 3UJ, UK
* Chester Beatty Laboratories, The Institute of Cancer Research, London SW3 6JB, UK
† To whom correspondence should be addressed

X-CHROMOSOME inactivation in mammals is a regulatory phenomenon whereby one of the two X chromosomes in female cells is genetically inactivated, resulting in dosage compensation for X-linked genes between males and females[1]. In both man and mouse, X-chromosome inactivation is thought to proceed from a single *cis*-acting switch region or inactivation centre (XIC/Xic)[2–5]. In the human, XIC has been mapped to band Xq13 (ref. 6) and in the mouse to band XD (ref. 7), and comparative mapping has shown that the XIC regions in the two species are syntenic[8]. The recently described human *XIST* gene maps to the XIC region[6] and seems to be expressed only from the inactive X chromosome[9]. We report here that the mouse *Xist* gene maps to the Xic region of the mouse X chromosome and, using an interspecific *Mus spretus*/*Mus musculus domesticus* F_1 hybrid mouse carrying the T(X; 16)16H translocation, show that *Xist* is exclusively expressed from the inactive X chromosome. Conservation between man and mouse of chromosomal position and unique expression exclusively from the inactive X chromosome lends support to the hypothesis that *XIST* and its mouse homologue are involved in X-chromosome inactivation.

We have used a 1.3- kilobase (kb) human probe, generated by the polymerase chain reaction from the published human *XIST* sequence, to screen an oligo(dT)-primed complementary

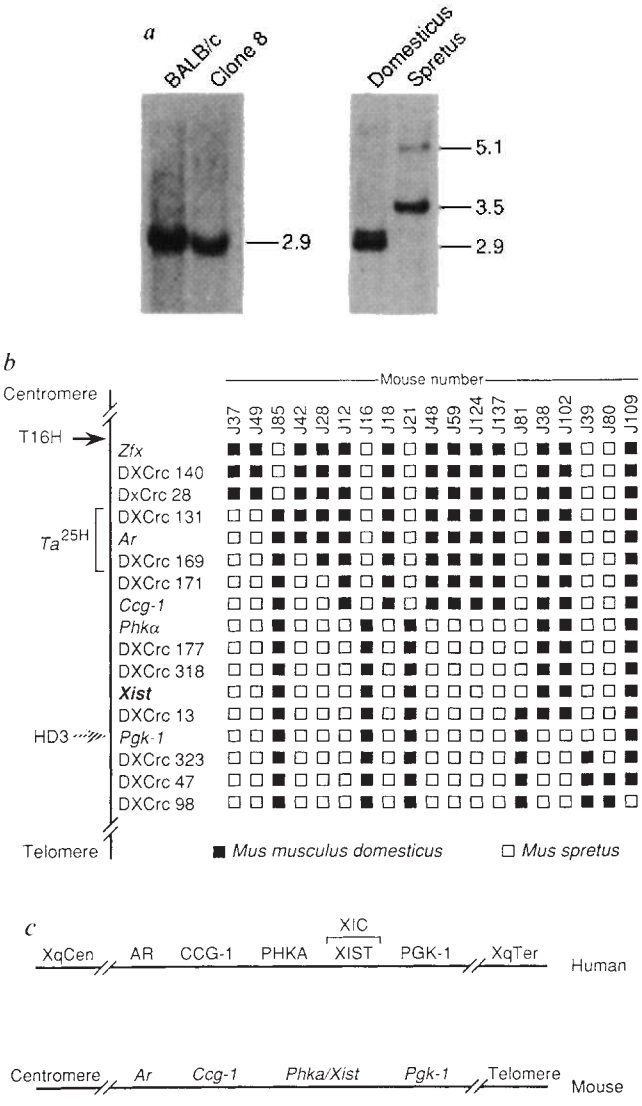


FIG. 1 Genetic mapping of the mouse *Xist* gene. *a*, Hybridization of the 2.7-kb insert from the *Xist* cDNA to a Southern blot of *Taq*I-restricted DNA from a BALB/c mouse and from the clone 8 hybrid cell line[17] carrying only the X chromosome of mouse together with human HeLa cell chromosomes. The *Xist* gene maps to the mouse X chromosome (panel 1). A single band corresponding to 2.9 kb was detected in both BALB/c and clone 8 DNA. For interspecific backcross pedigree analysis, a *Taq*I restriction fragment-length variant between *Mus musculus domesticus* (2.9 kb) and *Mus spretus* (5.1 and 3.5 kb) was detected (panel 2). An additional 3.1-kb weak band, consegregating with the 2.9-kb band, was seen in the *domesticus* mice used for the backcross. This band was not seen in BALB/c or clone 8 DNA, and is thought to represent a polymorphism in the genetic background (strains 101 and C3H) of the *domesticus* parent used for the backcross. Molecular sizes are indicated in kbp. *b*, The *Taq*I restriction-fragment length variant between *domesticus* and *spretus* was used to map *Xist* to the Xic region of the mouse X chromosome. This region is delineated by the breakpoint in T16H (proximal limit) and the deletion breakpoint in embryonic stem cell line HD3 (the distal limit)[7]. A small region within these limits, deleted in the Ta^{25H} mutation, is excluded as a candidate region for the Xic (refs 18, 19). A panel of 17 interspecific backcross mice with recombinant break points within Xic were used to map the *Xist* gene with respect to several other molecular markers in this region[10]. This analysis locates the *Xist* gene between *Ccg*-1 and the DXCrc13 loci, and shows that it consegregates with *Phkα*, DXCrc177 and DXCrc318. The haplotypes of the recombinant X chromosome of the 17 backcross progeny are shown for each of the probes used. *c*, A comparative map of the X-inactivation centre region of the human and mouse X chromosome illustrates that the genetic map position of *Xist* with respect to flanking markers is identical in the two species.

METHODS. The human *XIST* probe used to screen the mouse cDNA library was a 1.3-kb fragment generated from the published human *XIST* sequence[9] by PCR from HL60 cell line cDNA. The primers used were AAGGTGGAAGGCTCATAGG and CTGCATGATTGCCAATACAC, corresponding to nucleotides 121–140 and 1,462–1,443 of the human sequence[9]. The cDNA library, an oligo (dT)-primed library from 17.5-day-old mouse embryos (Clontech), was screened at low stringency (5 × SSC, 10% dextran sulphate, 1% SDS and 100 μg ml^{-1} salmon sperm DNA, at 50 °C overnight; wash conditions: 2 × SSC for 2 × 15 min at room temperature followed by 2 × SSC, 1% SDS for 2 × 30 min at 50 °C). A single positive hybridizing clone with a 2.7-kb insert was obtained. Partial sequence analysis showed that it coded for the murine homologue of *XIST*. This clone overlaps with part of the published human sequence and shows about 75% sequence homology; multiple termination codons were present in all reading frames (data not shown). Southern hybridizations were carried out under standard conditions[17]. The production of the interspecific backcross progeny and the detailed molecular mapping of the 17 backcross progeny used in the mapping panel is described elsewhere (ref. 10, and G.F.K., R. V. Thakker and S.R., manuscript submitted).