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 We thank D. Toniolo for the mouse cosmid library, C. Brown and J. Rupert for unpublished sequence data and D. Ledbetter and C. Chinault for critical reading of the manuscript.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

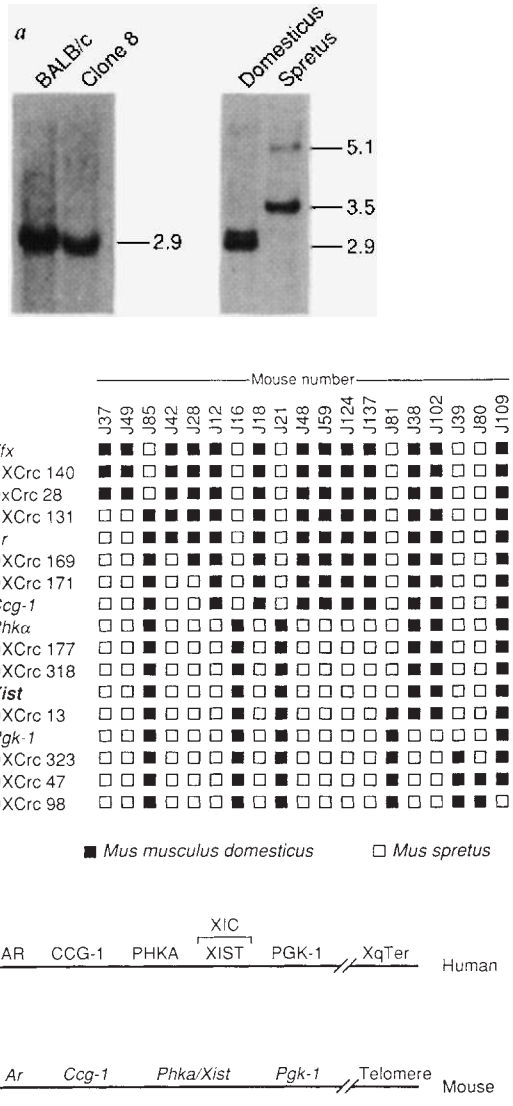

FIG. 1 Genetic mapping of the mouse *Xist* gene. *a*, Hybridization of the 2.7-kb insert from the *Xist* cDNA to a Southern blot of *Taq*I-restricted DNA from a BALB/c mouse and from the clone 8 hybrid cell line[17] carrying only the X chromosome of mouse together with human HeLa cell chromosomes. The *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).

DNA library from 17.5-day-old mouse embryos. A single positive clone containing a 2.7-kb insert was isolated, which corresponded to part of the mouse *Xist* gene. We demonstrated that the *Xist* cDNA maps to the mouse X chromosome by analysing a human–mouse somatic cell hybrid, clone 8, which carried only the X chromosome of mouse, together with HeLa-cell chromosomes (Fig. 1*a*, panel 1). To define the position of *Xist* on the mouse X chromosome, *Mus musculus domesticus*/*Mus spretus* interspecific backcross pedigree analysis was used. A *Taq*I restriction fragment-length variant between *domesticus* and *spretus* was found for the *Xist* probe (Fig. 1*a*, panel 2). The segregation of this variant was then analysed through a panel of 17 interspecific backcross progeny. These progeny had been previously characterized with many molecular and genetic markers in the Xic region of the mouse X chromosome[10]. The progeny were selected on the basis of their having recombination breakpoints between *Zfx* (the proximal limit of the Xic region) and DXCrc98 (the distal limit of the Xic region)[10] (Fig. 1*b*). The *Xist* gene maps in the position shown in Fig. 1*b* between *Ccg*-1 and DXCrc13, and cosegregates with *Phk* α, DXCrc177 and DXCrc318. Comparison of maps of the XIC/Xic region on the human and mouse X chromosome shows that the position of *Xist* with respect to flanking markers is identical in the two species (Fig. 1*c*).

We examined the expression of the *Xist* gene in the mouse by northern blotting (Fig. 2). *Xist* messenger RNA was detected in females only, suggesting that expression is confined to the inactive X chromosome, as seems to be the case for the human *XIST* gene[9]. The strong heterogeneous signal in female RNA indicates a series of many different transcripts, showing size and/or conformation heterogeneity, as for the human gene. But large discrete bands at positions corresponding to 14.3 kb and 11.8 kb could be clearly seen (Fig. 2*a*), indicating that there are two predominant transcription products of *Xist*. *Xist* expression was examined in both *Mus musculus domesticus* (C57B1/10 strain) mice and *Mus spretus* mice (Fig. 2*a*). The level of expression was markedly lower in the *Mus spretus* female than in the C57B1/10 female (Fig. 2*a*), and reprobing with actin cDNA (Fig. 2*b*) confirmed that equal quantities of intact RNA had been loaded in each lane. The *Xce* locus, which maps to the Xic region[11,12], has been proposed as a candidate for Xic in the mouse because different alleles affect the randomness of X-inactivation[13,14]. That *Xist* expression in *Mus spretus* is less than that in C57B1/10 mice is intriguing, because *Mus spretus* seems to have a strong *Xce* allele; that is, the *Mus spretus* X chromosome is more likely to remain active in F_1 interspecific hybrids between *Mus spretus* (spe) and *Mus musculus domesticus* (dom)[15].

To prove that expression of *Xist* was from the inactive X chromosome, rather than from the active X chromosome in response to the presence of the inactive X chromosome, we examined *Xist* expression using an F_1 interspecific hybrid mouse between *Mus spretus* (spe) and *Mus musculus domesticus* (dom) carrying the T(X; 16)16H translocation (T16) (ref. 15). These animals provide an *in vivo* system to assess whether *Xist* is expressed from the active ($T16^{dom}$) or inactive (X^{spe}) chromosome. This translocation causes nonrandom inactivation of the X^{spe} chromosome in $T16^{dom}/X^{spe}$ mice[16]: only the $T16^{dom}$ alleles of genes known to be X-inactivated are expressed[15]. These animals provide an *in vivo* system to assess whether *Xist* is expressed from the active ($T16^{dom}$) or inactive (X^{spe}) chromosome.

Control interspecific X^{dom}/X^{spe} females, in which either X chromosome may be inactivated, express both $Xist^{dom}$ and

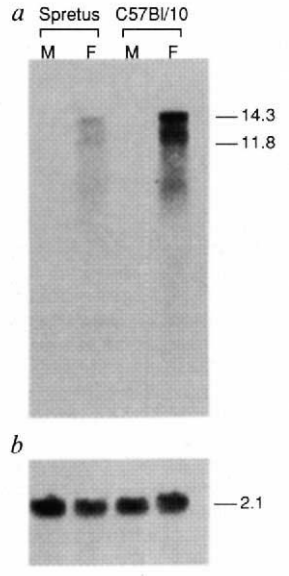

FIG. 2 Expression of the *Xist* gene in male and female mice. *a*, Northern blot of total cellular RNA from kidney of male (M) and female (F) C57B1/10 (*Mus musculus domesticus*) and *Mus spretus* mice hybridized with the 2.7-kb mouse *Xist* cDNA. The probe hybridizes only to the female RNA samples, suggesting that the transcript is produced from the inactive X chromosome. Major bands are detected at positions corresponding to 14.3 and 11.8 kb. *b*, The same northern blot filter stripped and reprobed with mouse actin cDNA, demonstrating that RNA is present and intact in all lanes. Molecular sizes are indicated in kb.

METHODS. Total cellular RNA was prepared using the guanadinium thiocyanate method[20]. Northern blots were carried out using the glyoxal method[21]. Total RNA (20 μg) was electrophoresed on a 0.8% agarose gel and then transferred to nylon membranes. The blots were hybridized with ^{32}P-labelled *Xist* cDNA insert in 50% formamide, 5 × SSC, 10% dextran sulphate, 1 × Denhardt's solution, 1% SDS and 100 μg ml^{-1} denatured salmon sperm DNA at 42 °C overnight. Wash conditions, 2 × SSC for 2 × 15 min at room temperature followed by 2 × SSC, 1% SDS for 2 × 20 min at 65 °C. The actin probe was a 1.35-kb mouse cDNA[22].

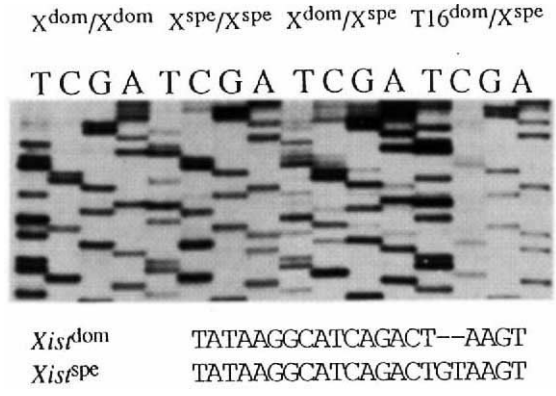

FIG. 3 Xist is only expressed from the inactive X chromosome in mice. RNA prepared from the livers of X^{dom}/X^{dom}, X^{spe}/X^{spe}, X^{dom}/X^{spe} and $T16^{dom}/X^{spe}$ mice was reverse-transcribed into cDNA. *Xist* cDNA was amplified using gene-specific primers and the PCR products directly sequenced. $Xist^{spe}$ differs from $Xist^{dom}$ by the presence of 2 bp at the position indicated. In the control interspecific F_1 animal (X^{dom}/X^{spe}) both alleles are expressed. Interspecific F_1 $T16^{dom}/X^{spe}$ animals, however, express only the $Xist^{spe}$ allele confirming that *Xist* is expressed exclusively from the inactive X chromosome.

METHODS. Generation of $T16^{dom}/X^{spe}$ and X^{dom}/X^{spe} animals and analysis of gene expression were performed as described[15]. *Xist* cDNA was amplified with primers (TAAGGACTACTTAACGGGCT and TCACATCTGCTCCACTTGAG) designed on the basis of the partial sequence of the murine cDNA clone (not shown). The product of the PCR is 300 bp. Control PCR reactions on RNAs incubated without reverse transcriptase demonstrated that amplification was not due to contaminating genomic DNA. Sequence variants between *Mus musculus domesticus* and *Mus spretus* were identified by DNA sequencing. The sequence shown is of *Xist* mRNA and is the opposite strand to that shown on the sequencing gel.

Xist^{spe} transcripts (Fig. 3). The relative expression of *Xist*^{dom} is higher than *Xist*^{spe}. This is evident in the C track which allows a within-track comparison, where the CC^{dom} doublet is significantly stronger than the CC^{spe} doublet. This result supports our northern-blot analysis which showed an inverse correlation between the level of *Xist* expression and the strength of the *Xce* allele (Fig. 2). By contrast, T16^{dom}/X^{spe} mice, in which the X^{spe} is always the inactive X chromosome, express only the *Xist*^{spe} allele (Fig. 3). Thus, expression of *Xist* is exclusively from the inactive X chromosome *in vivo*, and is not from the active X chromosome in response to a signal from the inactive X chromosome(s).

The conservation of map position between human *XIST* and mouse *Xist*, together with their exclusive expression from the inactive X chromosome, strongly supports the hypothesis that they are involved in X-chromosome inactivation. □

Received 13 February; accepted 23 April 1991.

1. Lyon, M. F. *Nature* **190,** 372–373 (1961).
2. Russell, L. B. *Science* **140,** 976–978 (1963).
3. Cattanach, B. M. *A. Rev. Genet.* **9,** 1–18 (1975).
4. Mattei, M. G., Mattei, J. F., Vidal, I & Giraud, F. *Hum. Genet.* **56,** 401–408 (1981).
5. Therman, G., Sarto, G. E. & Patau, K. *Chromosoma* **44,** 361–366 (1974).
6. Brown, C. J. *et al. Nature* **349,** 82–84 (1991).
7. Rastan, S. & Robertson, E. J. *J. Embryol. exp. Morph.* **90,** 379–388 (1985).
8. Keer, J. T. *et al. Genomics* **7,** 566–572 (1990).
9. Brown, C. J. *et al. Nature* **349,** 38–44 (1991).
10. Brockdorff, N. *et al. Genomics* **10,** 17–22 (1991).
11. Cattanach, B. M., Rasberry, C. & Andrews, S. J. *Mouse News Letter* **83,** 165 (1989).
12. Rastan, S. & Brown, S. D. M. *Genet. Res.* **56,** 99–106 (1990).
13. Cattanach, B. M. & Papworth, D. *Genet. Res.* **38,** 57–70 (1981).
14. Johnston, P. G. & Cattanach, B. M. *Genet. Res.* **37,** 151–160 (1981).
15. Ashworth, A., Rastan, S., Lovell-Badge, R. & Kay, G. *Nature* (in the press).
16. Lyon, M. F., Searle, A. G., Ford, C. E. & Ohno, S. *Cytogenetics* **3,** 306–323 (1964).
17. Brockdorff, N., Montague, M., Smith, S. & Rastan, S. *Genomics* **7,** 573–578 (1990).
18. Cattanach, B. M. *et al. Cytogenet. Cell Genet.* (in the press).
19. Brockdorff, N. Kay, G., Cattanach, B. M. & Rastan, S. *Mammalian Genome* (in the press).
20. Chirgwin, J. M. *et al. Biochemistry* **18,** 5294–5299 (1979).
21. Sambrook, J., Fritsch, E. F. & Maniatis, T. *Molecular Cloning: A Laboratory Manual* (Cold Spring Harbor Laboratory Press, New York 1989).
22. Minty, A. J. *et al. J. biol. Chem.* **256,** 1008–1014 (1981).

ACKNOWLEDGEMENTS. We thank J. McVey for the actin cDNA probe, D. Shepherd and other animal husbandry staff at the MRC Clinical Research Centre, the medical illustration department at the centre, and S. Jenks for typing the manuscript. This work was supported by the MRC and the Cancer Research Campaign.

Tetramerization of an RNA oligonucleotide containing a GGGG sequence

Joseph Kim, Chaejoon Cheong & Peter B. Moore

Department of Chemistry, Yale University, 225 Prospect Street, New Haven, Connecticut 06511, USA

POLY rG can form four-stranded helices[1]. The Hoogsteen-paired quartets of G residues on which such structures depend are so stable that they will form in 5′-GMP solutions, provided that Na^+ or K^+ are present (see for example, refs 2–4). Telomeric DNA sequences, which are G-rich, adopt four-stranded antiparallel G-quartet conformations *in vitro*[5,6], and parallel tetramerization of G-rich sequences may be involved in meiosis[7,8]. Here we show that RNAs containing short runs of Gs can also tetramerize. A 19-base oligonucleotide derived from the 5S RNA of *Escherichia coli* (strand III), 5′GCCGAUGGUAGUGUGGGGU3′, forms a K^+-stabilized tetrameric aggregate that depends on the G residues at its 3′ end. This complex is so stable that it would be surprising if similar structures do not occur in nature.

Strand III, which is prepared from the 5S RNA of *E. coli* nucleolytically[9], forms an aggregated species called III*, which migrates much more slowly than strand III on polyacrylamide gels. This III* species is very stable; unlike ordinary double helices it persists in low-salt buffers containing 8 M urea. Recent experiments demonstrated that III* is preferentially stabilized by K^+, suggesting that it might be a four-stranded structure, like the several G-rich DNA structures reported recently, which are also K^+-stabilized[5–8].

If III* is a four-stranded G structure, the residues most likely to be responsible for aggregation are the four G residues near the 3′ end of strand III, and strand III should be able to aggregate with any RNA that includes a run of four (or more) G residues, as has been demonstrated for some DNA tetramers[8]. An RNA hexamer was synthesized to test this prediction (5′UGGGGU3′), and equimolar mixtures of hexamer and strand III were analysed by PAGE.

If heterotetramers form in these mixtures, as expected, this experiment could have two different outcomes, depending on whether the aggregates formed are parallel stranded[7,8,10] or antiparallel stranded[5,6]. Parallel aggregation will generate five tetrameric species distinguishable on gels (Fig. 1), but antiparallel tetramerization should give rise to six. The 'extra' antiparallel species arises because the critical G-run in strand III is near one end of the molecule. Consequently there are two possible antiparallel III_2–H_2 tetramers, one with its strand III components parallel and the other with them antiparallel, and they should be distinguishable electrophoretically because they differ in hydrodynamic radius.

Only five species can be detected in mixtures of strand III and hexamer in addition to their single-stranded forms: H_4, III–H_3, III_2–H_2, III_3–H and III_4 (Fig. 2). We conclude therefore that III* is tetrameric, that the interactions stabilizing it involve residues at its 3′ end, and that III* is parallel-stranded. Hexamer tetramizes with itself, as expected.

If III* is stabilized primarily by interactions involving the four G residues at its 3′ end, it should be possible to convert the 5′ portions of each oligonucleotide in III* into double helix by hybridization with an RNA of appropriate sequence without

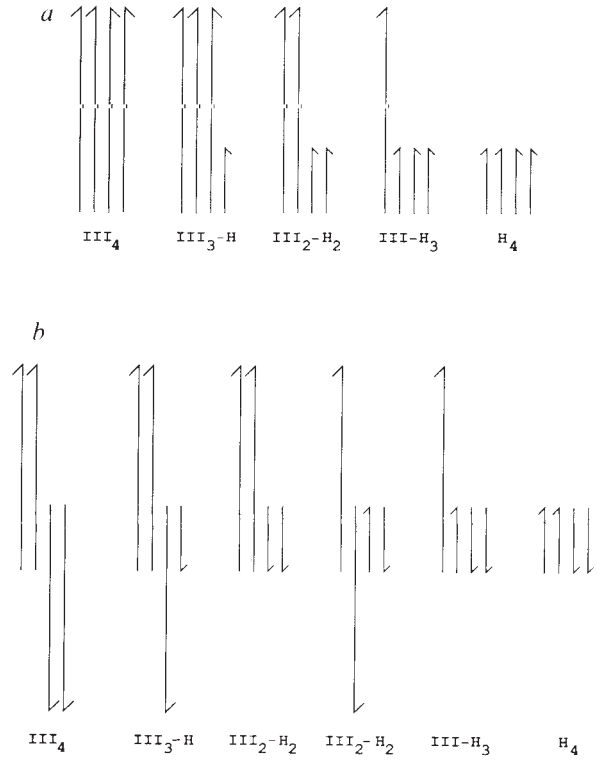

FIG. 1 The dependence of the number of tetrameric aggregates expected on relative strand orientation. The tetramers that can form in mixtures of hexamer and strand III depend on whether the structures in question are parallel stranded (*a*) or antiparallel (*b*), as shown.