AD_____

Award Number: DAMD17-99-1-9032

TITLE: An Experimental System to Evaluate LOH in Prostate Cancer

PRINCIPAL INVESTIGATOR: William M. Strauss, Ph.D.

CONTRACTING ORGANIZATION: Beth Israel Deaconess Medical Center
Boston, Massachusetts 02215

REPORT DATE: July 2001

TYPE OF REPORT: Annual

PREPARED FOR: U.S. Army Medical Research and Materiel Command
Fort Detrick, Maryland 21702-5012

# REPORT DOCUMENTATION PAGE

*Form Approved*
*OMB No. 074-0188*



| 1. AGENCY USE ONLY (Leave blank) | 2. REPORT DATE | 3. REPORT TYPE AND DATES COVERED |
|---|---|---|
| | July 2001 | Annual (1 Jan 00 - 30 Jun 01) |

**4. TITLE AND SUBTITLE**
An Experimental System to Evaluate LOH in Prostate Cancer

**5. FUNDING NUMBERS**
DAMD17-99-1-9032

**6. AUTHOR(S)**
William M. Strauss, Ph.D.

**7. PERFORMING ORGANIZATION NAME(S) AND ADDRESS(ES)**
Beth Israel Deaconess Medical Center
Boston, Massachusetts 02215

E-Mail: wstrauss@hihg.med.harvard.edu

**8. PERFORMING ORGANIZATION REPORT NUMBER**

**9. SPONSORING / MONITORING AGENCY NAME(S) AND ADDRESS(ES)**
U.S. Army Medical Research and Materiel Command
Fort Detrick, Maryland 21702-5012

**10. SPONSORING / MONITORING AGENCY REPORT NUMBER**

**11. SUPPLEMENTARY NOTES**

**12a. DISTRIBUTION / AVAILABILITY STATEMENT**
Approved for Public Release; Distribution Unlimited

**12b. DISTRIBUTION CODE**

**13. ABSTRACT** *(Maximum 200 Words)*

Our goal is to help define the human genes that are associated with prostate cancer. One of the steps in development of prostate cancer may be the loss of important genes called tumor suppressor genes. These genes suppress the formation of tumors. When they are not functional, tumor forming cells are no longer inhibited and cancers may develop. Identification of these genes is a complex process that relies on sorting through may different patient's tumor samples and looking for changes in number of genes. When a researcher detects a gene loss there is potential identification of a tumor suppressor gene. One problem with this approach is that there is no way to test these genes fro actual tumor suppressor activity. The purpose of this grant is to develop the technology of directly evaluating portions of chromosomes and genes for their role as tumor suppressors. Ultimately, and clearly beyond the scope of this granting period, we hope this will lead to new treatments for prostate cancer.

| 14. SUBJECT TERMS | 15. NUMBER OF PAGES |
|---|---|
| Loss of Heterozygosity, Mouse Transgenic Model, Xist, Disease Tumor suppressor gene, prostate cancer, inducible gene expression | 26 |
| | 16. PRICE CODE |

| 17. SECURITY CLASSIFICATION OF REPORT | 18. SECURITY CLASSIFICATION OF THIS PAGE | 19. SECURITY CLASSIFICATION OF ABSTRACT | 20. LIMITATION OF ABSTRACT |
|---|---|---|---|
| Unclassified | Unclassified | Unclassified | Unlimited |

## Table of Contents

# Annual Report for the Grant: DAMD17-99-1-9032

## Introduction

The scope of our proposal is both bold and innovative. We are creating a strain of mice which can be manipulated to create functional Loss of Heterozygosity (LOH) on murine chromosome 8. Functional LOH will be created inducibly, by expression of murine Xist under control of the tetracycline repressor/operator.

## Body

The progress to our goals is primarily dependent upon the creation of mouse lines containing Xist transgenes integrated into mouse chromosome 8. Secondarily this transgene must be expressed and functionally capable of inducing chromosomal silencing. It is this silencing that will result in functional LOH. Thus the progressive set of experiments outlined in the "statement of work" include introduction of transgene into J1 ES cells, selection with hygromycin B, mapping of the transgene by DNA-FISH, functional characterization of the cell line and subsequent injection of cells into blastocyst for chimera production.

We constructed a cell line which is transgenic on murine chromosome 8 (see figure 4, as we described in our phase II application). During the expansion of this cell line in preparation for chimera production it was noticed that the cell line had undergone a chromosomal alteration. In particular, chromosome 8 had undergone a Robertsonian translocation. This mutation is a technical set back and completely precludes the use of this cell line for our experiments. As a result, our originally proposed schedule is impossible. We have re-started our efforts to derive ES cell lines with chromosome 8 transgenes. To this end we have undertaken a series of three (3) ES cell transfections, from these transfections 506 primary colonies have been picked, archived, and expanded for initial analysis by RNA-FISH to determine if the transgene was capable of expression. Of the 506 clones only 94 clones appear capable of expression of Xist. (19%). Currently these clones are being screened by DNA-FISH to determine which clones have an Xist transgene integration into murine chromosome 8.

Clearly the entire success of this project is dependent upon a full appreciation of Xist function. Consequently, in addition, to the efforts to derive the ES cell line with the desired chromosome 8 transgene, we have continued our efforts to define the structure and function of the Xist transgene so pivotal to the success of this project. In this regard, first we are continuing our structural analysis of the Xist gene, especially to verify the published structure. Second we are determining the structural and developmental requirements for Xist to cause functional LOH in the ES cellular environment.

To these ends our efforts were specifically directed to (1) validation of the new 3'end that we discovered (PNAS 1999), (2) Refining the shuttle system for construction and expression of Xist transgenes, (3) Defining the requirement for Xist to bind to chromosomes as a prerequisite for function and (4) Development of a non-disruptive assay to measure LOH in situ. As all of these projects have resulted in publications, they

will be discussed under Key research accomplishments and the off-prints attached (see appendix).

## Key Research Accomplishments

Our accomplishments fall into two general groups, unpublished and published. The unpublished results center around the derivation of the ES cell line with the chromosome 8 transgene. These specific results: three (3) ES cell transfections have been carried out. From these transfections 506 primary colonies have been picked, archived, and expanded for initial analysis by RNA-FISH to determine if the transgene was capable of expression. Of the 506 clones only 94 clones appear capable of expression of Xist. (19%). Currently these clones are being screened by DNA-FISH to determine which clones have an Xist transgene integration into murine chromosome 8.

The published accomplishements fall into 4 sets of publications. 1) Memili et al 2001(Gene) details the confirmation of the structure for the Xist 3'end that we published in 1999 (Hong et al Proc Natl Acad Sci USA). The confirmation was based upon RNAse protection, Northern blot, and database analysis. 2) Hong et al 2001(Anal Biochemistry) details the creation of an efficient shuttle system for construction of inducible Xist transgenes as outlined in our grant. This system should be of general use to researchers. 3) Beletskii et al 2001 )Proc Natl Acad Sci USA) details the first demonstration that the Xist RNA has functional domains and that Xist RNA must bind in order to function in chromosome silencing (functional LOH). This major accomplishment was the result of an innovative new technique which we call P-IMP. 4) Kim et al 2001 (BioTechniques) describes a efficient, non-invasive general method to evaluate chromosomal LOH in the cells of tissues in situ. Later in the granting period this technology will be very useful.

## Reportable Outcomes

Memili, E. Y-K Hong, D. K. Kim, S.D. Ontiveros, W.M. Strauss. Murine *Xist* RNA isoforms are different at their 3'ends: a role for differential polyadenylation.. *Gene* **266**:131-137(2001).

Hong. Y-K,. D-H Kim, A. Beletskii, C. Lee, E. Memili, and W. M. Strauss. Development of two bacterial artificial chromosome shuttle vectors for a recombination-based cloning and regulated expression of large genes in mammalian cells. *Analytical Biochemistry* **291**:142-148(2001).

Beletskii, A. Hong,Y-K, Pehrson, J. , Egholm , M. & Strauss, W.M. PNA Interference Mapping Demonstrates Functional Domains in the Non-coding RNA *Xist*. *Proc. Natl. Acad. Sci. U.S.A.* **98**:9215-9220(2001)

Kim, D-H, Hong, Y-K, Egholm, M., & Strauss, W.M. A non-disruptive PNA-FISH protocol for formalin-fixed and paraffin-embedded tissue sections. *BioTechniques* In Press.

## Conclusions

There have been some major accomplishments during this reporting period as well as some set backs. Given the set backs we have been forced to recalculate our time table.

Nevertheless the accomplishements have been ample and in particular the paper by Beletskii et al 2001 PNAS USA is a major contribution to the field of research and we are very proud of it.

Our adjusted time table would estimate that by the end of this year or so we will have defined ES cells capable of being used in chimera production. At this time we can start the establishment of chimeric mice and breeding stable mouse lines containing chromosome 8 transgenes.

## References

Hong, Y-K, S.D.Ontiveros, C.Chen, W.M. Strauss. A New Structure for the Murine *Xist* Gene and its Relationship to Chromosome Choice/Counting during X-chromosome Inactivation. *Proc. Natl. Acad, Sci. U.S.A.* **96**(12): 6829-6834(1999).

Hong, Y-K, S.D. Ontiveros, W.M. Strauss. A revision of the human *XIST* gene organization and structural comparison to mouse *Xist*. . *Mammalian Genome.* **11**:220-224(2000).

Chen, C., T.Wei, M. Egholm, W.M. Strauss. Unique Chromosome Identification and Sequence-Specific Structural Analysis Using Short PNA Oligomers. *Mammalian Genome.* **11**:384-391(2000).

Memili, E. Y-K Hong, D. K. Kim, S.D. Ontiveros, W.M. Strauss. Murine *Xist* RNA isoforms are different at their 3'ends: a role for differential polyadenylation.. *Gene* **266**:131-137(2001).

Hong. Y-K,. D-H Kim, A. Beletskii, C. Lee, E. Memili, and W. M. Strauss. Development of two bacterial artificial chromosome shuttle vectors for a recombination-based cloning and regulated expression of large genes in mammalian cells. *Analytical Biochemistry* **291**:142-148(2001).

Beletskii, A. Hong,Y-K, Pehrson, J. , Egholm , M. & Strauss, W.M. PNA Interference Mapping Demonstrates Functional Domains in the Non-coding RNA *Xist*. *Proc. Natl. Acad. Sci. U.S.A.* **98**:9215-9220(2001).

Kim, D-H, Hong, Y-K, Egholm, M., & Strauss, W.M. A non-disruptive PNA-FISH protocol for formalin-fixed and paraffin-embedded tissue sections. *BioTechniques* In Press.

## Appendices

See attachements.

ELSEVIER

Gene 266 (2001) 131–137

GENE
AN INTERNATIONAL JOURNAL ON GENES AND GENOMES

www.elsevier.com/locate/gene

# Murine *Xist* RNA isoforms are different at their 3′ ends: a role for differential polyadenylation

Erdogan Memili, Young-Kwon Hong, Duk-Hwan Kim, Sara D. Ontiveros, William M. Strauss*

*Harvard Institute of Human Genetics, Harvard Medical School and Beth Israel Deaconess Medical Center, Room 441A, 4 Blackfan Circle, Boston, MA 02115, USA*

Received 25 September 2000; received in revised form 23 December 2000; accepted 15 January 2001
Received by M. D'Urso

**Abstract**

Murine *Xist* is an essential transcript for X chromosome inactivation (X inactivation). According to recently revised structure, *Xist* is at least 17.8 kb long. It consists of seven exons and there are two major transcripts in female somatic cells. In this study we further defined the molecular structures of the two isoforms, namely short (S) and long (L) forms by northern blot and RNAse protection assay (RPA). The following lines of evidences suggest that mouse *Xist* depends on differential polyadenylation, not alternative splicing, to generate the two RNA isoforms: (1) only one band was detectable with the northern probes spanning the 3′ end of *Xist*. (2) RPA showed the 3′ termini of both *S* and *L* forms, and there are putative polyadenylation signals and hairpin structures close to these ends. (3) Analyses by splice site prediction program did not show any evidence of splice motifs in the sequence of *L* form. (4) Alignments between Xist 3′ end (ESTs) and genomic sequence support the absence of splicing event in the region. The newly revised structure of *Xist* isoforms may have different stability and roles in the process of X inactivation. 

*Keywords*: X inactivation; *Xist*; Structure; Polyadenylation; RNA stability

## 1. Introduction

One of the two X chromosomes in female mammals is inactivated in order to maintain equal levels of expression of X-linked genes between males and females (Lyon, 1961). Prior to implantation of the embryo in the uterus, female embryos express genes from both X-chromosomes. After implantation, X-linked genes are expressed only from one X chromosome. In the embryo proper, either the maternal or paternal X chromosome is randomly chosen for inactivation. X chromosome inactivation, mediated by a locus called the *Xic* (X chromosome inactivation center), first occurs in the extraembryonic trophectoderm and primitive endoderm, and then proceeds to the inner cell mass of the blastocyst. The process of X chromosome inactivation involves: counting of number of the X chromosomes as compared to the number of autosomes, choosing a single X chromosome to remain active per autosome complement, initiation, spreading and maintaining the inactivity. Characteristics of the inactive X chromosome include expression of X-inactive specific transcript (*Xist*), transcriptional silencing of X-linked genes, late replication of the inactive X chromosome, hypoacetylation of histone H4, and hypermethylation of CpG islands in X-linked gene regions (Heard et al., 1997).

A number of studies demonstrate that *Xist* is essential for X chromosome inactivation (Penny et al., 1996; Brockdorff and Duthie, 1998). Upregulation of *Xist* occurs through post-transcriptional mechanisms. *Xist* accumulates in the nucleus, coats X chromosome and thereby involving in transcriptional silencing of the genes in the chromosome. Molecular details of the mechanism by which the *Xist* coating causes X chromosome inactivation are not known. Another gene, *Tsix*, has been recently shown to be transcribed from the antisense strand of the *Xist* gene and it is proposed that it regulates *Xist* expression in *cis*. In female cells, *Tsix* is expressed biallelicly before X inactivation and is not expressed after differentiation (Lee et al., 1999). Targeted deletion of *Tsix* results in skewed X inactivation, primarily non-random inactivation of the targeted X chromosome, in heterozygous females (Lee and Lu, 1999).

Initially, murine *Xist* was shown to be 14.7 kb RNA

Abbreviations: bp, base pair(s); CDNA, DNA complementary to RNA; EST, expressed sequence tag(s); kb, kilobase(s); nt, nucleotide(s); RPA, RNAse protection assay; RT-PCR, reverse transcription polymerase chain reaction; Xist, X inactive specific transcript

* Corresponding author. Tel.: +1-617-432-8475; fax: +1-617-432-8476.
*E-mail address:* wstrauss@hihg.med.harvard.edu (W.M. Strauss).

0378-1119/01/$ - see front matter 
PII: S0378-1119(01)00353-5

consisting of six transcribed exons and a rarely transcribed exon that is homologous to human *XIST* (Brockdorff et al., 1992; Simmler et al., 1996). It was also shown that stability of *Xist* is determined by differential usage of promoters (Johnston et al., 1998). According to the latter study, unstable *Xist* is transcribed from P0 promoter in undifferentiated cells. In differentiated XX female cells, a promoter switch occurs and stable *Xist* transcripts are produced from the P1 (~6 kb downstream of P0) and P2 (1.5 kb downstream of P1) promoters. In the soma 70–90% of stable *Xist* transcripts are produced from P2 promoter. Warsawsky et al. (1999) showed that there was no *Xist* transcription from the P0 promoter, and *Xist* transcript from P1 was unstable and did not accumulate in undifferentiated ES cells. A recent study examining murine *Xist* structure more closely demonstrated that murine *Xist* is at least 17.8 kb with an additional 3.1 kb of information at the 3′ end, and has seven exons, not six (Hong et al., 1999). The same study also showed that there were two major species of *Xist* detectable by northern analyses.

In this study, we attempted to do a precise analysis of *Xist* structure in differentiated cells. We further analyze the two isoforms of *Xist* reported by Hong et al. (1999) and show that these isoforms are different at their 3′ termini. The data are consistent with the interpretation that mouse *Xist* RNA isoforms are the product of differential polyadenylation, not splicing. Finally we report that most of the two detectable isoforms are transcribed exclusively from the P2 promoter.

## 2. Materials and methods

### 2.1. PCR amplification and cloning

Probes used for Norhtern blot include Mx8, pWS1231 and pWS850 (Fig. 1A). Probe Mx8 spans 5 kb of sequence upstream of the P1 promoter (Johnston et al., 1998). Probe pWS1231 is 3.3 kb long and spans 2 kb upstream and 1.1 kb downstream of the P1 promoter. Probe pWS850 is approximately 6 kb long and spans exon 6 and the majority of exon 7. Total size of probes pWS855, 859 and 860 is 853 bp long and span *Xist L* form. Probes used for RPA were subcloned from previously described vectors (Hong et al., 1999) into pBluscript II SK+ (Stratagene) resulting in pWS878, pWS970 and pWS971 for probes 1, 2 and 3, respectively. Probe 1 spans the end of *Xist* previously proposed by Brockdorff et al. (1992), probes 2 and 3 span putative polyadenylation sites in the *Xist* sequence (Fig. 1B).

### 2.2. RNA blotting

Total RNA was isolated from kidneys of male and female mice by using the guanidium thiocyanate method (Chirgwin et al., 1979). Total RNA (25 μg) was electrophoresed on 0.8% agarose gel containing 2.2 M formaldehyde for 20 h at 100 V/35 cm and then transferred to positively charged nylon membrane, yielding duplicate matched lanes. Blots were then hybridized with $^{32}$P-labeled DNA fragments from Mx8, pWS1231 digested with BssHII, pWS850 digested with SacII/SacI, or mixtures of pWS855, 859 and 860 (see Fig. 1A). Results of northern blot experiments are shown in Fig. 2 (upper panels). Lower panels in Fig. 2 show the Ethidium bromide staining of the 28 S RNA of the gels used for then experiments to demonstrate the presence of RNA in each lane (Fig. 2).

### 2.3. RNAse protection assay (RPA)

RPA probes and the regions they span in the murine *Xist* are shown in Fig. 1B. Ten micrograms total RNA was used for RPA. Probes pWS878, 970 and 971 were generated from *Xist* cDNA spanning exon 7. Total size of probe pWS878 is 271 bp of which 191 is *Xist* sequences spanning originally proposed end of *Xist* (Brockdorff et al., 1992) and the rest of the sequences are vector sequences from pBSIISK+. Probe pWS878 produces a 191 nt protected product which is the full length *Xist* sequences in the vector. Total size of probe pWS970 is 645 bp of which 557 bp is from *Xist* sequences spanning *L* form and part of the *S* form, and the rest of the sequences are from vector pBSIISK+. PWS970 produces 557 and 354 nt protected products for *L* or *S* forms respectively. Total size of probe pWS971 is 537 bp of which 437 bp is from *Xist* sequences spanning part of the *L* form and produces a 328 nt protected product corresponding the end of *L* form. $^{32}$P-radiolabeled UTP (NEN) was incorporated into probes by in vitro transcription using in vitro transcription kit (Promega). RPA was performed by using the RPA II kit (Ambion). Protected fragments were separated on 6% 8 M urea denaturing acrylamide gels and visualized using X-ray film (Fig. 2A–D). M13 mp18 ssDNA was sequenced using Sequenase Version 2 (USB) to use as molecular weight markers.

## 3. Results

### 3.1. Two isoforms of Xist are detectable in the mouse somatic cells and they are transcribed from P2 promoter

To determine the structure of *Xist* transcript, we performed northern blot experiments. Total RNA was extracted from murine male and female kidney, electrophoresed on denaturing agarose gels and transferred to nylon membranes. The blots were hybridized to the probes Mx8, pWS1231, pWS850 and mixture of probes pWS855, 859 and 860 (Fig. 1A). Mx8 probe, which spans 5 kb upstream of the P1 promoter, did not give any detectable signal, indicating no detectable transcript from upstream of the P1 promoter (Fig. 2A). Furthermore, no signal was detected with probe pWS1231, which spans a 1.1 kb sequence between the P1 and P2 promoter (Fig. 2B). However, probe pWS850, spanning exon 6 and 7, gave two different bands (Fig. 2C). This indicates that *Xist* has two distinct isoforms (long and short), likely produced from the P2

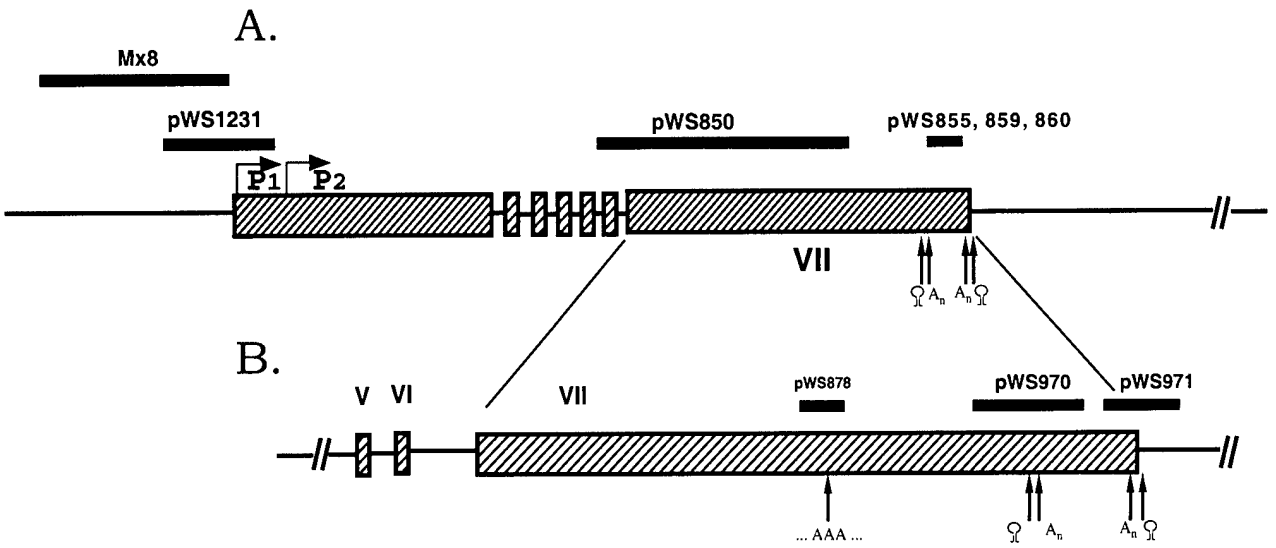


Fig. 1. *Xist* 3′ ends and probes for Northern blot (A) and RPA (B). P1 and P2 indicate transcription start sites from the P1 and P2 promoters respectively. Probes used for northern blot are shown at the top. Probe Mx8 spans upstream of the P1 promoter, probe pWS1231 spans 2 kb upstream and 1.1 kb downstream of the P1 promoter, probe pWS850 spans exon 6 and majority of exon 7 while probes pWS855, 859 and 860 span 853 nt of exon 7 at the 3′ terminus. Probes used for RPA are shown at the bottom row. Probe pWS878 spans previously proposed end of *Xist*, probe pWS970 and 971 span possible ends of *Xist S* and *L* forms due to differential polyadenylation. Probe pWS971 also spans the proposed rarely transcribed exon 8. Stretches of As under the pWS878 are shown as ... AAA ...which was the previously proposed end of *Xist*. Hairpin structures and putative polyadenylation signals are shown under the probes pWS970 and 971.

promoter. Mixture of probes pWS855, 859 and 860 spanning 853 bp of exon 7 hybridized only to the upper isoform detected with probe pWS850. This is consistent with the results from P3 of the RPA experiments (see Section 3.2), indicating that one of the isoforms ends before the region spanned by probes pWS855, 859 and 860 (Fig. 2D).

### *3.2. The two isoforms of Xist differ in size at their 3′ ends due to differential polyadenylation*

We further investigated the previously revised isoforms of murine *Xist* transcripts and their 3′ ends by RNAse protection assay (RPA). Earlier studies in our laboratory (Hong et al., 1999) indicated that *Xist* has 3.1 kb additional information at its 3′ end from what had been previously proposed (Brockdorff et al., 1992). Furthermore we reported that the murine *Xist* gene has seven exons, and two distinct isoforms, namely, short (S) and long (L) forms (Fig. 1A). In an attempt to determine the end of each isoform, and to disprove the previously reported end of *Xist* (Brockdorff et al., 1992), we performed RPA. Three probes were designed in such a way that the protected fragments would show the originally proposed 3′ end, S and L ends of *Xist* (Fig. 1). Total size of the first probe (pWS878) is 271 bp and it includes 191 bp of *Xist* sequences spanning the originally proposed end of *Xist*, and 80 bp of vector sequences from pBSIISK + . Second probe (pWS970) is 645 bp long and it includes 557 bp of *Xist* sequences spanning a possible polyadenylation site for the *S* form, and 88 bp of vector sequences from pBSIISK+. The third probe (pWS971) is 537 bp long and it includes 437 bp of *Xist* sequences spanning possible polyadenylation site for the *L* form. *Xist* RNA was detected with pWS878 as a protected fragment, the full length *Xist* sequences in the pWS878, 191 nt. The two bands in lane 4 of Fig. 3 are 3–4 nucleotides apart from each other, which is most likely due to differential restriction digest of the major protected fragment by the probe pWS878. There was no band detected at 70 nt long, as this would be the predicted fragment for the originally proposed end (Fig. 3A, lane 4). The full length, undigested, probe (271 bp) is seen in lane 1. There were no bands in the control lanes 2 and 3, yeast t-RNA and mouse male RNA (Fig. 3A). This confirms the results reported by Hong et al. (1999) and indicates that *Xist* does not end at the previously proposed site (Brockdorff et al., 1992). RPA with probe pWS970 showed two major protected fragments of 557 (full length *Xist* sequences in the pWS970) and 354 nt. The larger band represents the *Xist L* form while the smaller band of 354 represents the end of the *S* form (Fig. 3B). The end of *Xist L* form is demonstrated by detection of a 328 nt protected fragment from probe pWS971 (Fig. 3C). Full length probe containing *Xist* and the vector sequences is shown in lane 1. Results with the probe pWS971 do not support the suggestion that there is an additional exon 8, proposed to be homologous to human *Xist* (Fig. 3C) (Brockdorff et al., 1992).

The data obtained from northern and RPA experiments show that total lengths of *Xist* produced from P2 promoter are 15527 bp (*S* form) and 16370 bp (*L* form). The end of *L* form is consistent with the results obtained by Hong et al. (1999) in which attempts to amplify cDNA starting immediately 3′ of *L* form did not give any detectable PCR product. Previously shown P1 driven *Xist* transcripts

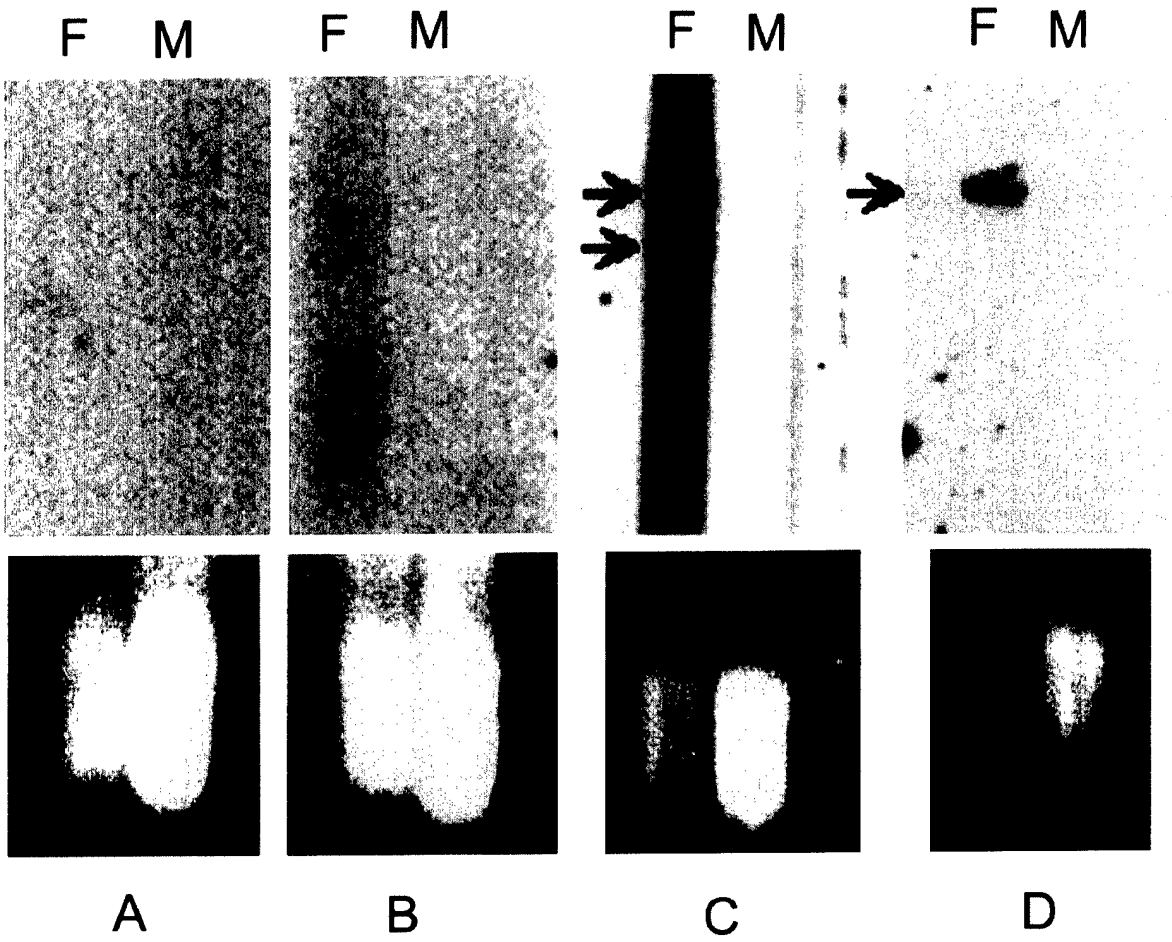


Fig. 2. Northern blot analysis of *Xist* 3′ ends (upper panels). Total RNA (25 μg) was isolated from male and female mouse kidney, fractionated on a 0.8% denaturing agarose gel and transferred to a nylon membrane. Resulting blots were hybridized with Mx8 (A), pWS1231 (B), pWS850 (C) or mixtures of pWS855, 859 and 860 (D). Hybridization with probes Mx8 and pWS1231 did not give any signal. Probe pWS850 hybridized to long and short isoforms as indicated by arrows while probes pWS855, 859 and 860 hybridized only to the larger isoform (long form) detected by pWS850. Lower panels show the ethidium bromide staining of 28S ribosomal RNAs in the RNA gels indicating the presence of RNA in each well. M, RNA from male mouse; F, RNA from female mouse.

(Brockdorff et al., 1992; Johnston et al., 1998) would be 1503 bp longer than these isoforms.

## 4. Discussion

Even though *Xist*, an essential gene for X inactivation, has been studied by a number of investigators, information about the structure of *Xist* RNA has been elusive. It had been reported that murine *Xist* is 14.7 kb and consists of six exons, and a rarely transcribed exon that is homologous to human *XIST* (Brockdorff et al., 1992; Simmler et al., 1996). However, our recent studies showed that *Xist* is at least 17.8 kb long, consists of seven exons and has two isoforms (Hong et al., 1999). Prior to differentiation, an unstable *Xist* had been reported to be transcribed from a promoter (P0) (Johnston et al., 1998). The same investigators suggested that, during cell differentiation, a promoter switch occurs from P0 to P1 (6.5 kb downstream of P0). Another promoter, P2, starts 1.5 kb downstream of the P1 promoter. However, Warsawsky et al. (1999) recently showed that there was no P0 promoter, and that unstable *Xist* is transcribed from the P1 promoter in undifferentiated ES cells. Also Johnston et al. (1998) proposed that stable *Xist* is initiated from a novel promoter located 1.5 kb downstream of the previously reported P1 promoter. We investigated any message transcribed upstream of the P1 promoter and detected no transcript (Fig. 2A). Our next attempt was to detect *Xist* produced from the P1 promoter by using a northern probe spanning 2 kb upstream and 1.1 kb downstream of the P1 promoter. This probe would hybridize to any transcript synthesized from the P1 promoter. Our results showed that there was not any detectable *Xist* synthesis from P1 promoter in differentiated cells (Fig. 2B). This result suggests that northern blot is not sensitive enough to detect low level of transcripts from P1 as it was previously shown

by nuclease protection assay that 10–20% of total *Xist* transcripts in the somatic cells is produced from the P1 promoter, 70–90% is transcribed from the P2 promoter (Johnston et al., 1998). We next used another probe, pWS850, spanning exon 6 and the majority of exon 7. This probe hybridized to two isoforms of *Xist* (Fig. 2C) indicating that both of these isoforms are synthesized from a promoter other than P1. They are likely to be produced from the previously proposed P2 promoter (Johnston et al., 1998). Both membranes that were hybridized with Mx8 and pWS1231 probes were rehybridized with pWS850 and signals detected showing the presence of *Xist* RNA in the membranes. We

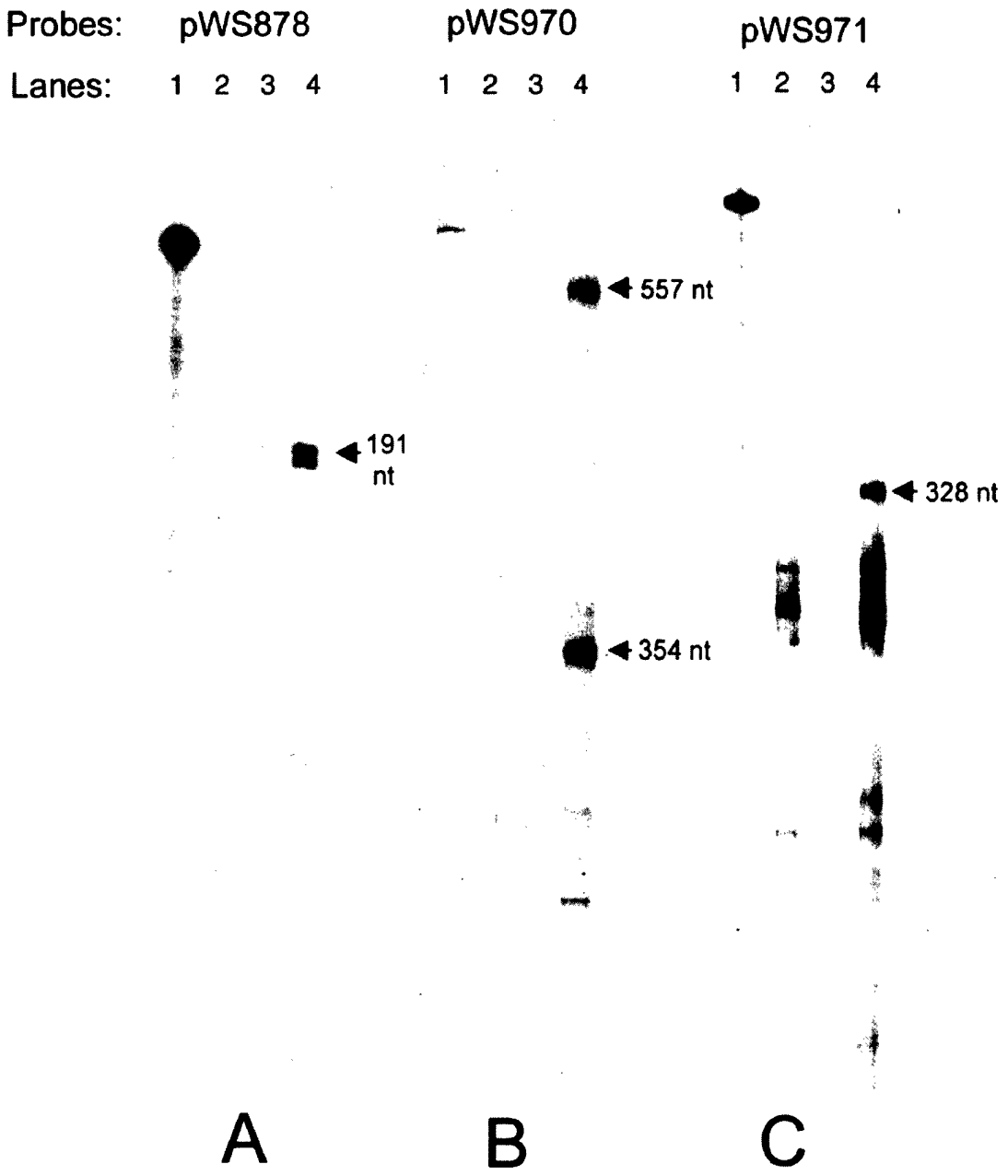


Fig. 3. RPA analysis of *Xist* structure. Probes 1–3 were designed as mentioned in Fig. 1. Mouse somatic cell RNA (10 μg total RNA) was isolated from male and female kidney, fractionated on a 6% denaturing acrylamide gel and RPA analyses was performed as described in Section 2. (A) Lane 1, full length Probe pWS878 (272 bp) of which 191 bp is *Xist* sequences and the rest is vector sequences from pBSIISK+. Lane 2, yeast t-RNA. Lane 3, male RNA. Lane 4, female RNA. Probe 1 protects full-length *Xist* sequences of 191 nt as a doublet, 3–5 nt apart from each other in lane 4. (B) Lane 1, full length Probe pWS970 (647 bp) of which 557 bp is *Xist* sequences and the rest is vector sequences from pBSIISK+. Lane 2, yeast t-RNA. Lane 3, male RNA. Lane 4, female RNA. Probe pWS970 protects two fragments of 557 nt (full length *Xist* sequences) and 354 nt in Lane 4. (C) Lane 1, full length Probe pWS971 (537 bp) of which 437 bp is *Xist* sequences and the rest is vector sequences from pBSIISK+. Lane 2, yeast t-RNA. Lane 3, male RNA. Lane 4, female RNA. Probe 3 protects a 328 nt fragment in lane 4.

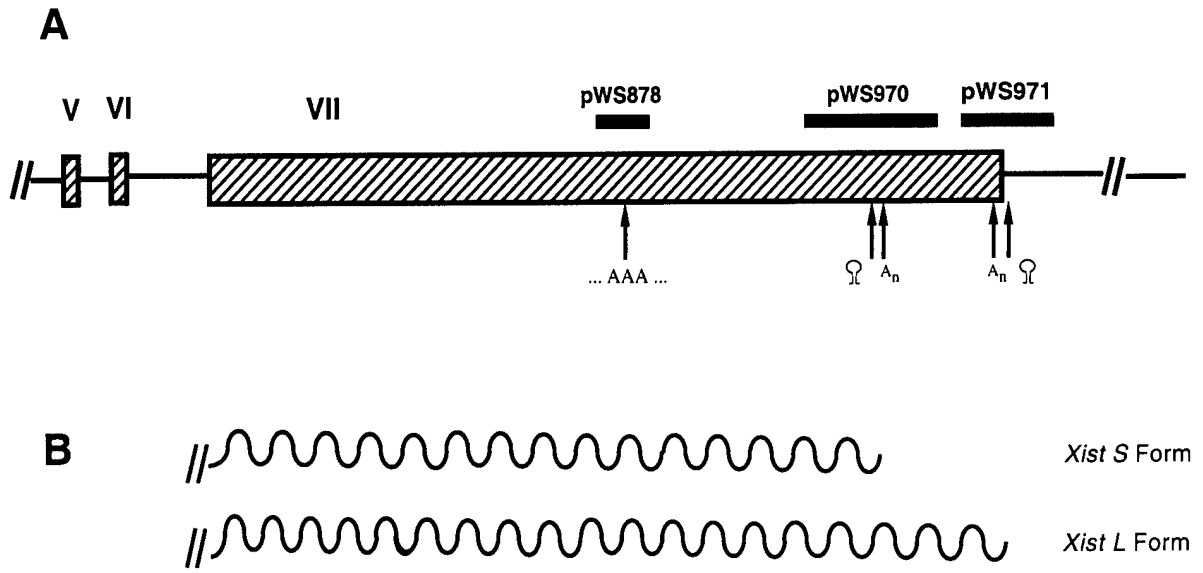


Fig. 4. *Xist* exons V–VII and the RPA probes over exon VII. (A) There are putative polyadenylation signals and hairpin structures close to 3′ termini of both *S* and *L* forms. The two isoforms of *Xist* transcripts; *S* and *L* forms. (B) The *S* and *L* forms are 17030 and 17873 nt long respectively. The 843 nt extra sequence information in the *L* form has significant homology to the 3′ intron of human *XIST*. Stretches of As under the pWS878 are shown as ... AAA ...which was the previously proposed end of *Xist*. Hairpin structures and putative polyadenylation signals are shown under the probes pWS970 and 971.

wanted to determine whether these two isoforms are different at their 3′ ends by using a mixture of probes (pWS855, 859 and 860) spanning part of exon 7 at its 3′ end. Surprisingly, this probe hybridized only to the longer isoform detected with probe pWS850, suggesting that the isoforms are different at their 3′ ends.

Next, RPA experiments were performed in order to precisely determine the 3′ ends of the *Xist* isoforms and also whether the difference in the 3′ end is a result of polyadenylation or splicing. When we performed RPA with probe pWS878 spanning the previously proposed site for the end of *Xist* (Brockdorff et al., 1992), only full length protected fragment was detected, indicating that *Xist* does not end there (Fig. 3A). The *Xist* 3′ end was previously defined by analysis of clones containing mouse cDNA as described by Brockdorff et al., 1992: 'Previously we described a 2.7 kb mouse Xist cDNA clone, mXist1, that was isolated from oligo(dT)-primed library (Brockdorff et al., 1991)'. The presence of long stretches of As may have been thought to be the end of

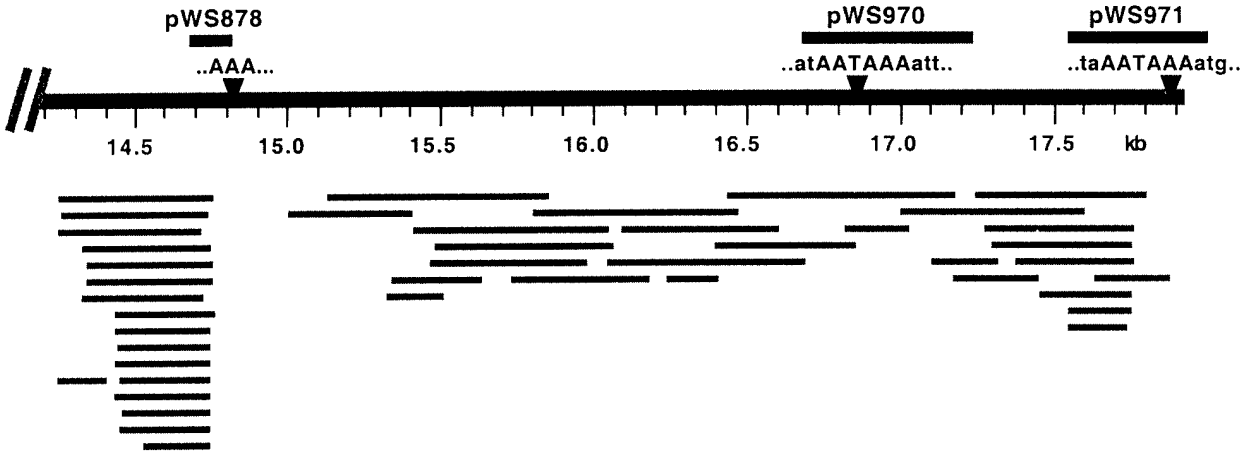


Fig. 5. Alignments of *Xist* 3′ end (ESTs) and genomic sequence. Probes used in the RPA experiments are shown on top. Partial sequences spanned by the RPA probes and stretches of As are shown under pWS878 and putative polyadenylation signals are shown under the probes pWS970 and 971. Alignments of ESTs are demonstrated under the 3′ sequences of *Xist*. Overlapping ESTs cover majority of the 3.1 kb sequences at the 3′ terminus of the *Xist L* form indicating that there is no alternative splicing within the region. A region starting from the previously proposed end of *Xist* does not have any available any matching ESTs, however, this region is spanned by the RPA probe pWS878 which was shown to be protected fully in the RPA experiments indicating that *Xist* does not end there.

the transcript (Brockdorff et al., 1992). Probe pWS970 protected two fragments of 354 and 557 nucleotides, showing the two isoforms and the location of the 3′ end of one isoform (Fig. 3B). We named these isoforms the long (L) and short (S) forms. We were able to locate the 3′ end of the *L* form from the data obtained by probe 3 (Fig. 3C). The *Xist L* form is approximately 843 nt longer than the S form (Figs. 1; 3B,C; 4A,B). These two isoforms and the extra information in the *L* form (Fig. 4) may play important roles in increased stability of *Xist* and the process of X inactivation (Panning et al., 1997). It is interesting to note that according to recently revised structure of human *XIST*, there is a high sequence similarity between the 3′ regions of human *XIST* and mouse *Xist* (Hong et al., 2000). However, human *XIST* has an intron in this region and this intron spans the majority of the high sequence similarity between mouse and human. This finding and the data presented here indicate that mouse *Xist* is differentially polyadenylated while human *XIST* is alternatively spliced. Evidence for differential polyadenylation of *Xist* includes: (1) we have performed EST database search and alignment with the 3′ of *Xist* sequences, starting from the previously proposed end of *Xist*. Our results show all of the available EST's in GenBank covering the 3′ region, further the ESTs are ungapped and overlapping each other, i.e. there is no splicing at the region (Fig. 5). There is only one region (immediately downstream of the previously proposed end of *Xist*) for which there is no ESTs yet. However, the RPA probe pWS878 spanning this region gave only full length protected probe, i.e. there is no end or no splicing there (Fig. 3). (2) Previously RT-PCR was performed in the 3′ region of *Xist* and only one PCR product of 3.1 kb was observed (Hong et al., 1999 and data not shown). (3) No splice sites were detected with analyses of the 3′ termini of *S* and *L* forms with splice site prediction program (http://www.fruitfly.org/seq_tools/splice.html). (4) Only one band corresponding to the *L* form with the mixtures of probes pWS855, 859 and 860 was detected by northern blot experiments (Fig. 2D).

In summary, we precisely define the structure of two *Xist* transcript isoforms, the short (*S*) and long (*L*) forms. The majority of these isoforms are synthesized from the P2 promoter and they differ in their 3′ termini as a result of differential polyadenylation. The two isoforms may have different stability and roles in the process of X inactivation. Elucidation of the correct structure of *Xist* and its isoforms will be helpful in experiments aimed at a better understanding *Xist* of function in the process of X inactivation.

## Acknowledgements

This work was supported by an NIH grant CA81732, a US Army Prostate Cancer Research Award (DAMD 17-99-1-9032) and the Harvard Nathan Shock Center for Biology of Aging Core C (5 P30-AG13314), all awarded to W.M. Strauss.

## References

Brockdorff, N., Duthie, S.M., 1998. X chromosome inactivation and the *Xist* gene. Cell Mol. Life Sci. 54 (1), 104–112.

Brockdorff, N., Ashworth, A., Kay, G.F., Cooper, P., Smith, S., McCabe, V.M., Norris, D.P., Penny, G.D., Patel, D., Rastan, S., 1991. Conservation of position and exclusive expression of mouse *Xist* from the inactive X chromosome. Nature 351, 329–331.

Brockdorff, N., Ashworth, A., Kay, G.F., McCabe, V.M., Norris, D.P., Cooper, P.J., Swisft, S., Rastan, S., 1992. The product of mouse Xist gene is a 15 kb inactive X-specific transcript containing no conserved ORF and located in the nucleus. Cell 71, 515–526.

Chirgwin, J.M., Przybyla, A.E., MacDonald, R.J., Rutter, W.J., 1979. Isolation of biologically active ribonucleic acid from sources enriched in ribonuclease. Biochemistry 18, 5294–5299.

Heard, E., Clerc, P., Avner, P., 1997. X-chromosome inactivation in mammals. Ann. Rev. Genet. 31, 571–610.

Hong, Y.K., Ontiveros, S.D., Chen, C., Strauss, W.M., 1999. A new structure for murine *Xist* gene and its relationship to chromosome choice/counting during X-chromosome inactivation. Proc. Natl. Acad. Sci. USA 96, 6829–6834.

Hong, Y.K., Ontiveros, S.D., Strauss, W.M., 2000. A revision of the human *XIST* gene organization and structural comparison with mouse *Xist*. Mamm. Genome. 11 (3), 220–224.

Johnston, C.M., Nesterova, T.B., Formstone, E.J., Newall, A.E.T., Duthie, S.M., Sheardown, S.A., Brockdorff, N., 1998. Developmentally regulated *Xist* promoter switch mediates initiation of X inactivation. Cell 94, 809–817.

Lee, J.T., Lu, N., 1999. Targeted mutagenesis of *Tsix* leads to non-random X inactivation. Cell 99, 47–57.

Lee, J.T., Davidow, S.L., Washawsky, D., 1999. *Tsix*, a gene antisense to *Xist* at the X-inactivation center. Nat. Genet. 21, 400–404.

Lyon, M.F., 1961. Gene activation in the X-chromosome of the mouse (*Mus musculus L.*). Nature (London) 190, 372–373.

Panning, B., Dausman, J., Jaenisch, R., 1997. X chromosome inactivation is mediated by *Xist* RNA stabilization. Cell 90, 907–916.

Penny, G.D., Kay, G.F., Sheardown, S.A., Rastan, S., Brockdorff, N., 1996. Requirement for *Xist* in X chromosome inactivation. Nature (London) 379, 131–137.

Simmler, M.C., et al., 1996. Localization and expression analysis of a novel conserved brain expressed transcript, Brx/BRX, lying within the Xic/XIC candidate region. Hum. Mol. Genet. 5, 1713–1726.

Warsawsky, D., Stavropoulos, N., Lee, J.T., 1999. Further examination of the *Xist* promoter switch hypothesis in X inactivation: evidence against existence and function of a P0 promoter. Proc. Natl. Acad. Sci. 96, 14424–14429.

Analytical Biochemistry **291,** 142–148 (2001)
doi:10.1006/abio.2001.5010, available online at http://www.idealibrary.com on IDEAL®

# Development of Two Bacterial Artificial Chromosome Shuttle Vectors for a Recombination-Based Cloning and Regulated Expression of Large Genes in Mammalian Cells

Young-Kwon Hong,* Duk-Hwan Kim,* Anton Beletskii,* Charles Lee,† Erdogan Memili,* and William M. Strauss*,[1]

**Harvard Institute of Human Genetics, Harvard Medical School, Beth Israel Deaconess Medical Center, 4 Blackfan Circle, Boston, Massachusetts 02115; and †Department of Obstetrics and Gynecology, Brigham & Women's Hospital, and Harvard Medical School, 75 Francis Street, Boston, Massachusetts 02115*

Received November 10, 2000; published online February 26, 2001

**Most conditional expression vectors designed for mammalian cells have been valuable systems for studying genes of interest by regulating their expressions. The available vectors, however, are reliable for the short-length cDNA clones and not optimal for relatively long fragments of genomic DNA or long cDNAs. Here, we report the construction of two bacterial artificial chromosome (BAC) vectors, capable of harboring large inserts and shuttling among *Escherichia coli*, yeast, and mammalian cells. These two vectors, pEYMT and pEYMI, contain conditional expression systems which are designed to be regulated by tetracycline and mouse interferons, respectively. To test the properties of the vectors, we cloned in both vectors the green fluorescence protein (GFP) through an *in vitro* ligation reaction and the 17.8-kb-long X-inactive-specific transcript (*Xist*) cDNA through homologous recombination in yeast. Subsequently, we characterized their regulated expression properties using real-time quantitative RT-PCR (TaqMan) and RNA-fluorescent *in situ* hybridization (FISH). We demonstrate that these two BAC vectors are good systems for recombination-based cloning and regulated expression of large genes in mammalian cells.** © 2001 Academic Press

*Key Words:* **bacterial artificial chromosome (BAC); tetracycline-inducible promoter; interferon-inducible promoter; yeast artificial chromosome (YAC).**

With the completion of the Human Genome Project, there is an increased need for special vector systems that allow cloning of large fragments of DNA and subsequent regulated expressions of the fragments in target cells. Yeast artificial chromosomes (YACs)[2] have been a valuable tool for cloning and manipulating large pieces of genomic DNA (1–5). Furthermore, highly efficient homologous recombination in the yeast *Saccharomyces cerevisiae* permits direct subclonings of specific regions from a YAC into yeast-bacteria shuttle plasmids for further investigations. For example, previously reported bacterial artificial chromosome (BAC) vectors, pClasper and pSURF-2, have been excellent recombination-based cloning vectors, capable of harboring DNA fragments much larger than most bacteria-yeast shuttle plasmids can accommodate (6, 7). However, when it is necessary to study a large gene by conditionally regulating its expression in mammalian cell, the vectors are not optimal because they were initially designed as cloning vectors (6, 7).

Conditional expression systems have been widely used as a molecular switch from bacteria to mammalian cells. For mammalian cells, the tetracycline (Tc)- and the interferon-inducible systems have been previously described (8–12). The tetracycline-regulated promoter contains tetracycline-responsive elements (TRE) and regulates the transcription of a downstream gene when the tetracycline-controlled transactivator (tTA) binds to the promoter. The transactivator is a fusion protein, composed of the Tet repressor from *Esche-*

[1] To whom reprint requests should be addressed. E-mail: wstrauss@hihg.med.harvard.edu.

[2] Abbreviations used: BAC, bacterial artificial chromosome; YAC, yeast artificial chromosome; Tc, tetracycline; TRE, tetracycline-responsive elements; tTA, tetracycline-controlled transactivator; ES, embryonic stem; GFP, green fluorescence protein; *Xist,* X chromosome-inactive-specific transcript; PCR, polymerase chain reaction; FISH, fluorescent *in situ* hybridization.



0003-2697/01 $35.00

*richia coli* and the activating domain from the VP16 protein (pTet-Off) (11). In the presence of the tetracycline, the activator protein is prevented from binding to the Tet operator and therefore expression of the cloned gene does not occur. With removal of tetracycline, the transactivator binds to the TRE sequences and stimulates transcription of the cloned gene. Alternatively, a mutated version of the transactivator (rtTA) has the reversed binding property (pTet-On) (11). In this case, the rtTA protein binds to the TRE promoter only when tetracycline is present, functioning as an inducer. On the other hand, the interferon-inducible promoter system was initially reported as a controlled expression system for mouse embryonic stem (ES) cells as well as somatic cells (12). Transcription from the interferon promoter system can be properly regulated and this dosage-dependent induction was observed from both transiently and stably transfected ES cells (12).

In order to develop vector systems that allow cloning large genes from YACs and regulating their expressions in mammalian cells, we constructed two BAC vectors, pEYMT and pEYMI. These vectors, capable of shuttling among bacteria, yeast and mammalian cells, contain a conditional expression system that can be regulated by either tetracycline (for pEYMT) or mouse interferons (for pEYMI). They are identical in sequences except the promoter regions. We have characterized the properties of these two vectors by cloning and regulating expressions of two genes in both vectors, the green fluorescence protein (GFP) gene by *in vitro* ligation reactions and the unusually long (17.8-kb cDNA) murine X chromosome-inactive-specific transcript (*Xist*) gene through yeast homologous recombination. Results presented in this paper demonstrate the vectors excellent systems both for cloning large genes and for subsequent conditional expression of the genes in mammalian cells.

## MATERIALS AND METHODS

### *General Methods, Generating Stable Cell Lines, and Induction Conditions*

Standard molecular biology methods were employed (13). DH10B was used as an *E. coli* host strain. Final concentrations of 50 μg/ml of ampicillin and 12.5 μg/ml of chloramphenicol were used for bacterial transformations and selections. yWS116 (the same as Y116) was previously described (1). Yeast transformation was performed by the lithium acetate method and standard yeast protocols were exercised (14). pEYMT-GFP and pEYMT-*Xist* were tested in a modified NIH3T3 cell line that expresses the tetracycline-controlled transactivator (Tet-Off) (15). Standard NIH3T3 cell line was used to test the pEYMI-GFP. Mouse J1 embryonic stem cell line was cultured as described (1) and used to test pEYMI-*Xist*. Transfections of both mouse fibroblasts and embryonic stem cells were done using DOTAP (Boehringer Mannheim) and selected for in media containing 400 μg/ml hygromycin. Expression from the pEYMT-derived vectors was induced upon removal of doxycycline (10 μg/ml) for 36 to 120 h. Expressions of the cloned genes in the pEYMI vector were induced by addition of 1000 IU/ml of mouse interferons (α and β mixture from Sigma) for 3 to 5 days.

### *Construction of Vectors*

pEYMT and pEYMI were constructed as follows. The *Eco*RV fragment containing the hygromycin-resistance gene from pPGK-hyg (16) was inserted into the blunted *Afl*III site of pTRE (Clontech) to make pTH1. *Pac*I linker (TTAATTAA) and *Not*I linker (AGCGGCCGCT) were sequentially inserted into the blunted *Cla*I and blunted *Xba*I site of pTH1, respectively, to make pWS1280. Two complementary DNA oligonucleotides (GGCCGGCCGTTTAAACGTCGACGCTAGC and GGCCGCTAGCGTCGACGTTTAAACGGCCGGCCGC), representing the multiple cloning sites, were annealed and inserted into the *Sac*II/*Not*I sites of pWS1280. Finally, the *Xho*I/*Pac*I fragment from the resulting vector was transferred into the *Sal*I/*Pac*I sites of pClasper, which generated pEYMT. The promoter of the human interferon-inducible gene 6-16 was PCR-amplified from female placenta genomic DNA using two DNA oligonucleotides (AAAGAGCTCGGCGCGCCATGATGGCCACATGAACACA and AAACCGCGGCTGCTGATAGATGGGCACAG). The resulting PCR fragment containing the interferon-inducible promoter was digested with *Asc*I and *Sac*II and used to replace the *Asc*I/*Sac*II fragment containing the tetracycline-inducible promoter flanked by the *Asc*I/*Sac*II sites, which generated the pEYMI vector.

pEYMT-GFP and pEYMI-GFP were generated by cloning the *Xho*I/*Not*I fragment containing eGFP from pEGFP-N1 (Clontech) at the *Sal*I and *Not*I sites of pEYMT and pEYMI, respectively. The cloning of *Xist* cDNA is described in detail elsewhere (Hong *et al.*, in preparation). Briefly, the first 1-kb fragment of Xist, which serves the 5′ homologous sequence for the homologous recombination, was cloned in pBluescript IISK+ from a lambda clone containing the *Xist* gene (Strauss *et al.*, unpublished work). DNA fragments, covering from the end of the first exon to the end of gene, were PCR-amplified from a female mouse cDNA library from several PCRs. The resulting PCR fragments were then assembled adjacent to the 5′ homologous region in pBluescript IISK+. The entire fragment was transferred to either pEYMT or pEYMI. The resulting prerecombination vectors were linerized using a *Sal*I site and then introduced into the yeast strain, yWS116, which harbors a YAC containing the *Xist* gene (1). The successful gap-filled full-length *Xist*

clones generated through the homologous recombination were identified and characterized.

### *Real-Time Quantitative RT-PCR and RNA-FISH*

Total RNA was obtained by using Tri-Reagent (Sigma) according to the manufacturer's instructions. Total RNA was treated with 1 unit RQ-DNAse (Promega) per 10 $\mu$g of RNA. TaqMan real-time quantitative PCR analysis was performed using the EZ-RT PCR Core reagent from PE Applied Biosystems. Four hundred nanograms of total RNA was used for analysis using WS483 (AACAGTTAGGTCCCGGCTTT) and WS831 (CTTTGCTTTTATCCCAGGCA) as the primers and WS869 (6FAM-TCTGTGTGGAGCTTTGTGAAG-TAMRA) as the dual-labeled fluorescent probe. A standard curve was obtained by using serial dilutions of the known concentrations of pWS889 (17). Real-Time reactions and quantification were done using the ABI 7700 sequence analyzer. Fluorescence *in situ* hybridization was performed as previously described (17).

## RESULTS

### *Features of Two BAC Vectors*

We have constructed two vectors which were designed to directly clone specific regions of interest from YACs via homologous recombination and then to regulate their expressions using two independent inducible promoters in mammalian cells. Figure 1 represents the schematic diagram of the pEYMT and pEYMI vectors. A mini-F factor replication origin and the chloramphenicol-resistance gene allow an *E. coli* host to maintain the vector in low copies in the presence of chloramphenicol. CEN6/ARS4 and the LEU2 gene are required for propagation in yeast cells and the hygromycin-resistance gene as a marker for mammalian cells (16). For regulated expression in mammalian cells, pEYMT and pEYMI harbor the tetracycline-inducible promoter (11) and the promoter of the human interferon-inducible gene 6-16 (12), respectively. The promoters are followed by multiple cloning sites and SV40 poly(A) signal sequence. The multiple cloning sites of the vectors contain both 6- and 8-base-recognition enzyme sites including *Fse*I, *Pme*I, *Sal*I, *Nhe*I, and *Not*I as the unique sites (*Bam*HI is unique only for pEYMT) (Fig. 1). Furthermore, recognition sites for two homing endonucleases, I-*Sce*I and I-*Ppo*I, are flanking the cassette of the inducible promoters, multiple cloning sites, SV40 poly(A) signal sequence, and hygromycin-resistance gene. I-*Sce*I and I-*Ppo*I recognize asymmetric 18- and 15-base sequences, respectively. Due to the rare occurrence of these recognition sequences, there is extremely low chance for the cloned gene to have their recognition sites. Therefore, after the gene of interest is cloned into the multiple cloning sites, the cassette containing the promoter, the cloned gene, poly(A) signal, and the hygromycin-resistant gene can be readily transferred into other vectors for other purposes.

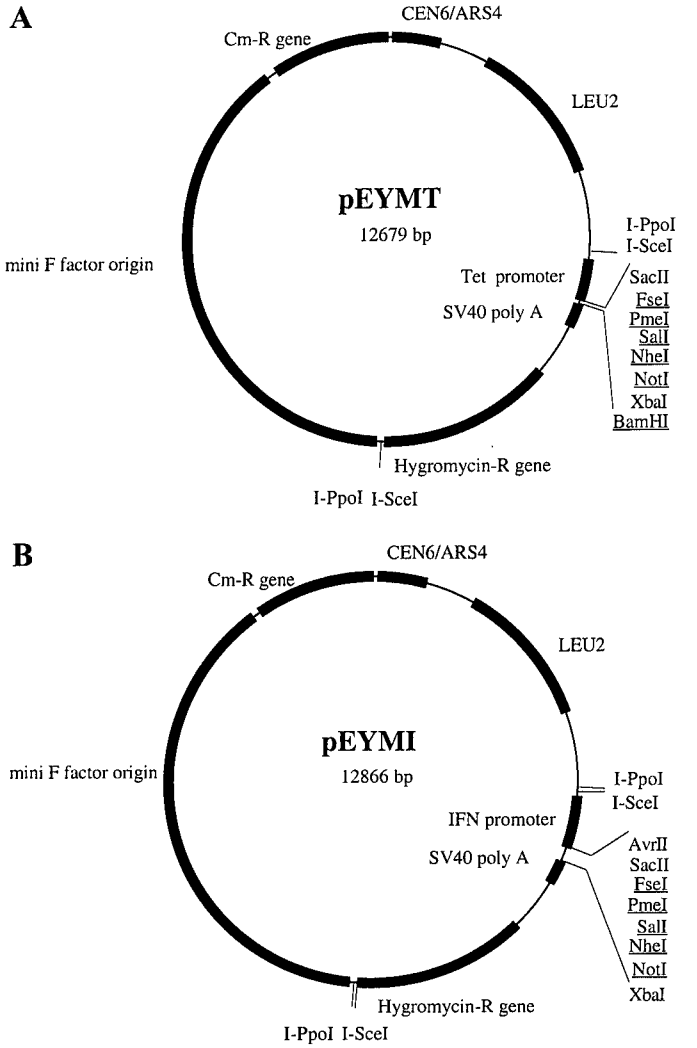


**FIG. 1.** Schematic representation of the pEYMT (A) and pEYMI (B) vectors. Unique restriction enzyme sites are underlined. Cm-R gene, chloramphenicol-resistance gene; hygromycin-R gene, hygromycin-resistance gene; Tet promoter, tetracycline-inducible promoter; IFN promoter, human interferon-inducible gene 6-16.

### *Functional Analysis of the Vectors Using the GFP Gene as a Reporter*

In order to analyze the vectors, we cloned the enhanced green fluorescent protein gene in both vectors and generated stable cell lines containing either construct. Figure 2 shows tight regulation in the expressions of the GFP reporter gene from these vectors. The pEYMT-GFP vector was transfected into a mouse embryonic fibroblast cell line which expresses the tetra-

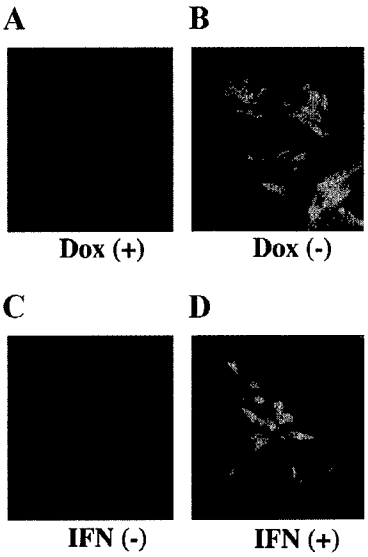


**FIG. 2.** Regulated expression of GFP from pEYMT-GFP (A and B) and pEYMI-GFP (C and D). The transfectants containing pEYMT-GFP were cultured in two separate dishes in the presence (A) or absence (B) of doxycycline (10 $\mu$g/ml) for 2 days before images were captured. Similarly, cells containing pEYMI-GFP were cultured in two separate dishes in the absence (C) or presence (D) of mouse interferon mixture of $\alpha$ and $\beta$ (1000 IU/ml) for 3 days before images were captured.

cycline-controlled transactivator (15) and stable cell lines were obtained after selection in a medium containing hygromycin and doxycycline. Stable transfectant cell cultures were grown in the presence (Fig. 2A) or absence (Fig. 2B) of doxycycline for 36 h before image capturing. Because the host cells represent the pTet-Off system, removal of doxycycline activates the transcription of the clone gene from the promoter (11). Similarly, the pEYMI-GFP vector was transfected into the NIH3T3 cell line and resulting transfectants were grown in the absence (Fig. 2C) or presence (Fig. 2D) of 1000 IU/ml of mouse interferon-$\alpha$ and -$\beta$ mixture. After 3 days, the cells were examined under microscope and the images were captured. These data show that the expressions of the GFP from the two vector systems are tightly regulated by the presence and absence of doxycycline or interferons, consistent with previously reported data (11, 12).

### *Characterization of pEYMT and pEYMI by Cloning and Expressing the Xist Gene*

Since pEYMT and pEYMI are derived from pClasper, a bacteria-yeast shuttle BAC vectors, they are expected to be capable of cloning large genes directly from YACs by homologous recombination (6). In order to confirm this important feature of the vectors, we cloned the mouse X chromosome-inactive-specific transcript gene. The *Xist* gene is extraordinarily long with the lengths of its cDNA and genomic DNA, 17.8 and 22.8 kb, respectively, and contains regions of highly repeated sequences (17, 18). These features of *Xist* make it very difficult to clone the gene in a routine high-copy vectors by *in vitro* ligation reactions or PCR amplifications. In fact, we observed that a fragment of *Xist,* cloned in the pBluescript vector, tended to be easily rearranged in an *E. coli* host (data not shown). In addition, PCR amplification reactions of the first exon of *Xist* (about 9 kb), which contains most highly repeated sequences, yielded DNA products with multiple incorrect sizes during our attempt to clone the gene. Therefore, successful cloning of the *Xist* gene directly from a YAC through homologous recombination would verify the function of the vectors. Another striking feature of *Xist* is that the *Xist* transcript accumulates on chromosomes where the gene is expressed (17, 18). This feature serves as a good assay to monitor the regulated expressions by RNA-FISH (1, 18). Figure 3 represents the scheme of recombination-based cloning of the *Xist* gene using pEYMT and pEYMI. Relying on this strategy, we cloned both the cDNA and the full length of *Xist* in both vectors from the previously reported YAC that contains the *Xist* gene (1). Subsequently, we isolated the recombinants of the cDNA and the genomic DNA of *Xist* and verified by a series of sequencing as well as Southern blot analyses (data not shown). Only the cDNA constructs of the gene (pEYMT-*Xist* and pEYMI-*Xist*) were used for further analysis.

Then, we generated stable fibroblast cell lines containing pEYMT-*Xist* (The *Xist* cDNA cloned in pEYMT) by transfecting the vector into the previously described NIH3T3 line that expresses the tTA transactivator (pTet-Off) (15). Since the tTA transactivator activates transcription of the cloned gene only in the absence of the effector, tetracycline (or doxycycline), the induction of expression of the *Xist* gene would only occur upon removal of the effector from medium. Figure 4 shows the regulated expression of the *Xist* gene by RNA-FISH. Upon the removal of doxycycline, the expression of the *Xist* gene was induced and accumulation of the *Xist* transcript was detected by RNA-FISH (Figs. 4A and B). Subsequently, the induction level was quantified using real-time quantitative RT-PCR (TaqMan RT-PCR). We used primers against the 3′ terminal region of *Xist* as well as a dual-labeled fluorescent probe that binds between the two primers. The quantitative RT-PCR shows that the induction of expression can result in roughly 80-fold expression over basal levels seen before induction (Fig. 4C). The control cell line does not express the endogenous *Xist* gene, since it is expressed only in female cells after differentiation is completed (1, 18).

We also generated stable cell lines containing pEYMI-*Xist* (the *Xist* cDNA cloned in pEYMI) by trans-

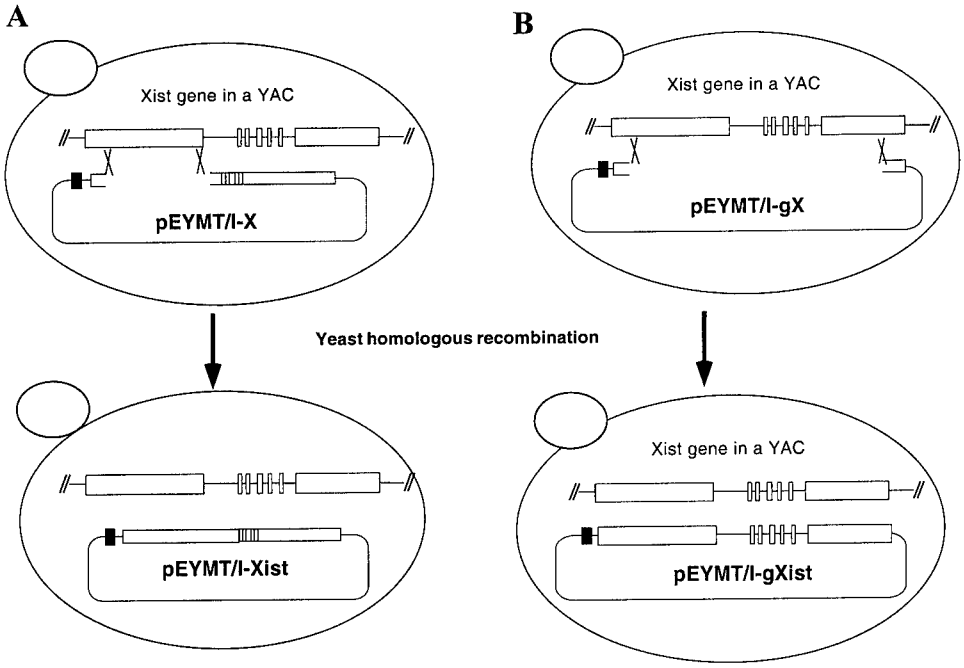


**FIG. 3.** Schematic representation cloning of the *Xist* cDNA (A) or the *Xist* gene (B) from the *Xist*-containing YAC into the pEYMT/I vectors. The prerecombination vector containing the 5′ and 3′ homologous sequences of the *Xist* gene for the cDNA (pEYMT/I-X) or the gene (pEYMT/I-gX) is linearized by using a restriction enzyme site located between the 5′ and the 3′ homologous sequences. Then, the linearized vectors were introduced into the yeast yWM116 strain and the transformants were selected for on leucine-deficient media. The successful recombinant clones (pEYMT/I-*Xist* or pEYMT/I-g*Xist*) through the "gap-filling" are isolated and only the cDNA constructs were used for further analysis.

fecting the vector into the mouse embryonic stem cells. Figure 5 shows the controlled expression of the *Xist* cDNA from pEYMI-*Xist* by RNA-FISH. Addition of mouse interferon-$\alpha$ and -$\beta$ mixture (1000 IU/ml) to the media stimulated the transcriptional induction of the *Xist* transcripts from the ectopic copy of *Xist* after 3 days, whereas the ES cell-specific expressions of the endogenous *Xist* gene did not change, regardless of presence or absence of the inducer, serving as an internal control.

## DISCUSSION

We have constructed two bacteria-yeast-mammalian shuttle vectors with two different regulated promoter systems. These vectors are derived from the previously described vector, pClasper (6). Both pClasper and our vectors share important features, namely bacteria-yeast shuttling, recombination-based cloning of specific regions from YAC, and stable maintenance of the large inserts in bacteria. In addition to these functions, pEYMT and pEYMI are designed for regulating the expression of a cloned gene in mammalian cells, whereas pClasper serves only for a cloning vector.

In order to test their functions, we cloned the GFP gene by conventional *in vitro* ligation and the *Xist* gene by recombination in both pEYMT and pEYMI vectors. We found that the two inducible promoters of the vectors were fully functional and responded well to the induction conditions. Although we obtained several cell lines with a tight regulation of expression, a number of stably transfected cell lines displayed less refined regulation either having a relatively high basal level of expression or low/absent induction. Fifty to 60 transfected clones were screened to recover 12 clones which exhibited regulated GFP expression; alternatively we had to screen more than a hundred transfected clones to screen to recover 2 clones which exhibited regulated *Xist* expression. The difference in efficiency of useful clone recovery may be attributed to the identity of the transgene, the structural characteristics of the transgene, the copy number, or the positional effects of the integrated transgene, rather than from intrinsic problems with the vectors.

The pEYMT and pEYMI vectors have advantages and disadvantages over each other. While pEYMI works by itself, pEYMT requires another vector which encodes a transactivator. Therefore, when using pEYMT, a stable cell line which expresses the transactivator should first be generated by transfecting either pTet-On or pTet-Off depending on the goals of the given experiment. Recently, a relatively easy way to

screen for cells expressing the transactivator has been developed using the GFP-transactivator fusion protein (19). Compared to the amount of time and effort required to establish pEYMT-based systems, it can be more rapid to set up the pEYMI-based system as it does not require additional vector systems. Nevertheless, exposure of test cells to inducing doses of mouse interferon may directly or indirectly cause an undesirable interference of cellular process. Although we have not observed any confounding problems, transcriptional induction of other interferon-responsive genes may result in alteration of cell growth and differentiation.

The pEYMT and pEYMI vectors contain the mini-F factor origin and therefore are maintained at low copy number in bacteria. Yet, the alkaline mini-prep of 1.5-ml overnight culture yields a good amount of DNA to do routine laboratory work, including enzymatic digestions. In order to obtain a larger quantity of DNA, one may resort to the use of a column-based purification protocol (20) or the cesium chloride ultracentrifugation method (13) using a starting culture volume of 1 liter. Our experience shows that exceptional precautions are not necessary for handling of the purified vectors and constructs.

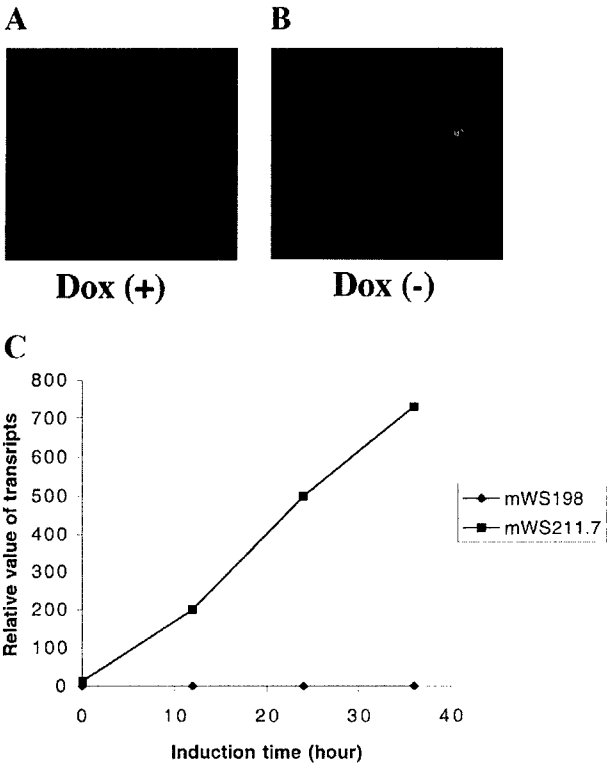


**FIG. 4.** Regulated expression of *Xist* from pEYMT-*Xist* in the mouse fibroblast cell lines expressing the tetracycline-controlled transactivator. (A) RNA-fluorescent *in situ* hybridization (FISH) was performed on the transfected cells grown in the presence (A) and absence (B) of doxycycline (10 $\mu$g/ml) for 36 h. (C) Transfected cell line (mWS211.7) that contains pEYMT-*Xist* and untransfected control cell line (mWS198) were culture in media containing doxycycline and the expression of *Xist* was induced by removal of doxycycline for 12, 24, and 36 h. The induction was quantified using the real-time quantitative RT-PCR.

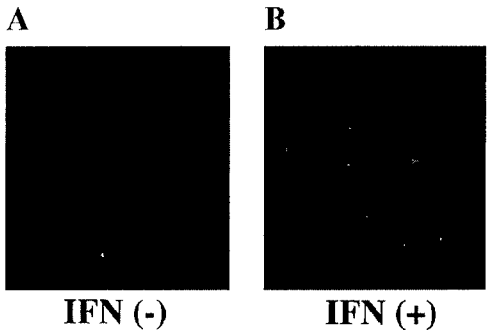


**FIG. 5.** Regulated expression of *Xist* from pEYMI-*Xist* in the mouse male embryonic stem cells. RNA-FISH was performed on the transfected cells grown in the absence (A) and presence (B) of mouse interferon-$\alpha$ and -$\beta$ mix (1000 IU/ml) for 3 days. The expressions of the endogenous *Xist* gene are detectable under both conditions and serve as an internal control.

## ACKNOWLEDGMENTS

The authors thank Drs. S. Bradshaw and P. Rathjen for their generous gifts. The authors acknowledge the support of U.S. Army Prostate Cancer Research Award DAMD17-99-1-9032, the Harvard Nathan Shock Program Rant Core C P30 AG13314, and NIH/NCI R21CA81732, awarded to W.M.S.

## REFERENCES

1. Lee, J. T., Strauss, W. M., Dausman, J. A., and Jaenisch, R. (1996) *Cell* **86,** 83–94.
2. Strauss, W. M. (1996) *Methods Mol. Biol.* **54,** 307–327.
3. Strauss, W. M., Dausman, J., Beard, C., Johnson, C., Lawrence, J. B., and Jaenisch, R. (1993) *Science* **259,** 1904–1907.
4. Smith, D. R., Smyth, A. P., Strauss, W. M., and Moir, D. T. (1993) *Mamm. Genome* **4,** 141–147.
5. Strauss, W. M., and Jaenisch, R. (1992) *EMBO J.* **11,** 417–422.
6. Bradshaw, S., Bollekens, J. A., and Ruddle, F. H. (1995) *Nucleic Acids Res.* **23,** 4850–4856.
7. Boyd, A. C., Davidson, H., Stevenson, B., McLachlan, G., Davidson-Smith, H., and Porteous, D. J. (1999) *Biotechniques* **27,** 164–170, 172, 175.
8. Bujard, H. (1999) *J. Gene Med.* **1,** 372–374.
9. Baron, U., Gossen, M., and Bujard, H. (1997) *Nucleic Acids Res.* **25,** 2723–2729.
10. Freundlieb, S., Baron, U., Bonin, A. L., Gossen, M., and Bujard, H. (1997) *Methods Enzymol.* **283,** 159–173.
11. Gossen, M., Freundlieb, S., Bender, G., Muller, G., Hillen, W., and Bujard, H. (1995) *Science* **268,** 1766–1769.
12. Whyatt, L. M., Duwel, A., Smith, A. G., and Rathjen, P. D. (1993) *Mol. Cell. Biol.* **13,** 7971–7976.

13. Sambrook, J., Fritsch, E. F., and Maniatis, T. (Eds.) (1989) Molecular Cloning: A Laboratory Manual, Cold Spring Harbor Laboratory Press, Cold Spring Harbor, New York.
14. Guthrie, C., and Fink, G. R. (1991) Guide to Yeast Genetics and Molecular Biology, Academic Press, San Diego.
15. Wu, Z., Xie, Y., Bucher, N. L., and Farmer, S. R. (1995) *Gene Dev.* **9,** 2350–2363.
16. te Riele, H., Maandag, E. R., Clarke, A., Hooper, M., and Berns, A. (1990) *Nature* **348,** 649–651.
17. Hong, Y. K., Ontiveros, S. D., Chen, C., and Strauss, W. M. (1999) *Proc. Natl. Acad. Sci. USA* **96,** 6829–6834.
18. Brockdorff, N., Ashworth, A., Kay, G. F., McCabe, V. M., Norris, D. P., Cooper, P. J., Swift, S., and Rastan, S. (1992) *Cell* **71,** 515–526.
19. Callus, B. A., and Mathey-Prevot, B. (1999) *Biochem. Biophys. Res. Commun.* **257,** 874–878.
20. Kirschner, L. S., and Stratakis, C. A. (1999) *Biotechniques* **27,** 72–74.

# PNA interference mapping demonstrates functional domains in the noncoding RNA *Xist*

**Anton Beletskii*†, Young-Kwon Hong*†, John Pehrson‡, Michael Egholm§, and William M. Strauss†¶**

†Harvard Institute of Human Genetics, Beth Israel Deaconess Medical Center, Harvard Medical School, 4 Blackfan Circle, Boston, MA 02115; ‡University of Pennsylvania, School of Veterinary Medicine, 3800 Spruce Street, Philadelphia, PA 19104-6046; and §Applied Biosystems, 500 Old Connecticut Path, Framingham, MA 01701

Edited by Stanley M. Gartler, University of Washington, Seattle, WA, and approved May 23, 2001 (received for review April 8, 2001)

**The noncoding RNA *Xist* has been shown to be essential for X-chromosome inactivation and to coat the inactive X-chromosome (Xi). Thus, an important question in understanding the formation of Xi is whether the binding reaction of *Xist* is necessary for X-chromosome inactivation. In this article, we demonstrate the failure of X-chromosome silencing if the association of *Xist* with the X-chromosome is inhibited. The chromatin-binding region was functionally mapped and evaluated by using an approach for studying noncoding RNA function in living cells that we call peptide nucleic acid (PNA) interference mapping. In the reported experiments, a single 19-bp antisense cell-permeating PNA targeted against a particular region of *Xist* RNA caused the disruption of the Xi. The association of the Xi with macro-histone H2A is also disturbed by PNA interference mapping.**

X-chromosome inactivation is an early developmental process occurring in female mammals to compensate for differences between male and female mammals in dosage of genes residing on the X-chromosome (1, 2, 4). In mammals, dosage compensation is achieved by the transcriptional silencing of genes on one of the two X-chromosomes in females (5). The inactivated X-chromosome (Xi) can be microscopically observed during interphase as a condensed body at the nuclear periphery (6). Moreover, the Xi has been shown to form a nuclear structure termed the macrochromatin body (MCB) known to be enriched for the variant histone, macrohistone H2A (7–10).

On the basis of the study of chromosomal translocations, an interval called the *X inactivation center* (XIC) of the X chromosome has been identified to control the process of inactivation (11, 12). The XIC has been shown to be a complex transcription unit consisting of at least two genes, *Xist* and *Tsix*, which are transcribed off the opposite strands of the same DNA duplex (3, 13–18). Although the function of *Xist* is not known, deletion of the gene leads to failure of X-inactivation, and female knockout mice die near gastrulation (19, 20). In embryonic stem (ES) cells, the *Xist* transcript supplied in an inducible manner by a transgene has been shown to be necessary and sufficient to cause silencing of genes in cis to the transgene (21).

The gene *Xist* is expressed exclusively from the Xi and shows several interesting features. First, both human and mouse *XIST/Xist* cDNAs are unusually long, 19.3 and 17.8 kb, respectively (22, 23). Second, the transcript does not seem to encode a protein (3, 16). Third, the *Xist* RNA physically associates with, or "coats," the Xi (3). The silencing was always associated with coating of the chromosome. Two basic questions about *Xist* are: (*i*) How is the coating/binding process related to the structure of the *Xist* RNA; and (*ii*) is the binding reaction of *Xist* to Xi necessary for *Xist* function?

We have developed a technology that we term Peptide Nucleic Acid (PNA)-Interference Mapping (P-IMP) to define in living cells how specific regions of the *Xist* transcript contribute to X-inactivation. These experiments take advantage of the unique properties of PNAs. PNAs are nucleic acid mimics, which contain a pseudopeptide backbone, composed of charge neutral and achiral *N*-(2-aminoethyl) glycine units to which the nucleobases are attached via a methylene carbonyl linker (24–26). PNAs hybridize with high affinity to complementary RNA sequences forming PNA–RNA complexes via Watson–Crick or Hoogsteen binding (26). PNAs are not readily degraded and do not participate in or activate repair or degradative pathways for DNA or RNA, ensuring a very selective range of activity. In addition to the high thermal stability of complexes, PNA–RNA binding is highly sensitive to mismatches (27–30). The PNAs used for P-IMP were conjugated to transportan via a cleavable disulfide linkage. Transportan is a 27-aa chimeric peptide consisting of the N-terminal fragment of the neuropeptide galanin and the membrane-interacting wasp venom peptide mastoparan. The conjugation of PNA to transportan results in rapid energy-independent nonsaturable transport of the PNA–transportan conjugate across the plasma membrane (31–33).

Using P-IMP, we found that a group of PNAs complementary to a distinct repeat region in the first exon of *Xist* completely abolished binding of *Xist* to the X-chromosome and, in so doing, prevented formation of Xi.

## Experimental Procedures

RNA folding experiments were performed by using MFOLD software, available at http://mfold2.wustl.edu/~mfold/rna/form1.cgi. Many probable structures were generated by submission of sequence derived from the C region.

For PNA conjugate treatment, we plated $10^4$ tetraploid fibroblast cells per well in eight-well Lab-Tek slides (Nalge) 12 h before treatment to allow for cell attachment. PNA conjugates were dissolved in water at 50 $\mu$M concentration. The PNA stock was diluted with DMEM with 10% vol/vol Cosmic Calf Serum (HyClone) to a concentration 1 $\mu$M. PNA conjugate and medium mixture (0.2 ml) was then added to the cells. Afterward, PNA conjugate treatment slides were cooled on ice and processed for RNA fluorescence *in situ* hybridization (FISH).

Diploid female ES cell (mWS244.6) culture and differentiation were accomplished by standard method (21). Cells ($5 \times 10^4$ per well) were treated for 6 days with 1 $\mu$M PNA in the presence of $10^{-7}$ M retinoic acid. To maintain high PNA concentration, medium was changed every 12 h. PNAs used are indicated below and in the text.

To perform the TaqMan assay, total RNA was isolated by using Tri-Reagent (Sigma) according to the manufacturer's instructions. Total RNA was treated with 1 unit of RQ-DNase (Promega) per 10 $\mu$g of RNA. TaqMan quantitative PCR analysis was performed by using the EZ-RT-PCR Core Reagent from Applied Biosystems. Four hundred nanograms of total

This paper was submitted directly (Track II) to the PNAS office.

Abbreviations: MCB, macrochromatin body; PNA, peptide nucleic acid; P-IMP, PNA Interference Mapping; Xi, the inactive X-chromosome; ES, embryonic stem; FISH, fluorescence *in situ* hybridization; RT-PCR, reverse transcription–PCR; MCB, macrochromatin body; Xa, the active x-chromosome; BAC, bacterial artificial chromosome; Ch, chromosome.

*A.B. and Y.-K.H. contributed equally to this work.

¶To whom reprint requests should be addressed. E-mail: wstrauss@hihg.med.harvard.edu.

The publication costs of this article were defrayed in part by page charge payment. This article must therefore be hereby marked "*advertisement*" in accordance with 18 U.S.C. §1734 solely to indicate this fact..

GENETICS

RNA was used for analysis. A standard curve was obtained by using serial dilutions of the known concentrations of pWS889, the plasmid containing the 3′ end region of *Xist*. Reverse-transcription reactions and quantification were done by using the ABI 7700 Sequence Analyzer (Applied Biosystems). For *Xist*, two different assays were performed at the 5′ and 3′ regions of the transcript; only data for the 3′ assay are presented (see Fig. 3, which is published as supplemental data on the PNAS web site, www.pnas.org), as both of the assays yielded substantially identical results.

*Tsix*

pWS1049: (GCCAAGGTGTAAGTAGACTAGCCACT) F. primer

pWS1048: (CGTGGCGGTGCAAACTAAA) R. primer

pWS1304: (6FAM-CTCAGCCCGTTCCATTCCTTTG-TATTGTT-TAMRA) TaqMan probe

mouse *Idh1*

pWS1394: (ACCGCATGTACCAGAAAGGG) F. primer

pWS1395: (CTCGGGACCAGGCAAAAAT) R. primer

pWS1396: (6FAM-AGAGACGTCCACCAACCCCATT-GCTT-TAMRA) TaqMan probe

mouse *Dnmt3b*

pWS1397: (CAGGTCTCGGAGACGTCGAG) F. primer

pWS1398: (CTTCCATGAAGTCGACGCTG) R. primer

pWS1399: (6FAM-TCGTCTTCAGCAAGCACGCCATG-TAMRA) TaqMan probe

mouse *Hmg2*

pWS1400: (GGGCAAAATGTCCTCGTACG) F. primer

pWS1401: (CGAGTCGGGATGCTTCTTCT) R. primer

pWS 1402: (6FAM-CAGACCTGCCGCGAGGAGCAC-TAMRA) TaqMan probe

5′ *Xist* assay

pWS1048: (CGTGGCGGTGCAAACTAAA) F. primer

pWS1049: (GCCAAGGTGTAAGTAGACTAGCCACT) R. primer

pWS1304: (6FAM-CTCAGCCCGTTCCATTCCTTTG-TATTGTT-TAMRA) TaqMan probe

3′ *Xist* assay

pWS483: (AACAGTTAGGTCCCGGCTTT) F.primer

pWS831: (CTTTGCTTTTATCCCAGGCA) R. primers

pWS869: (6FAM-TCTGTGTGGAGCTTTGTGAAG-TAMRA) TaqMan probe

RNA-FISH was done according to refs. 22 and 23. RNA-FISH probes: for *Xist*, full length cDNA = pWS1081, and for $\beta$-actin, a murine BAC from Genome Systems (St. Louis) (assignment no. 324C19). Histone macroH2A1 immunofluorescent labeling was done according to Costanzi and Pehrson (9).

For actinomycin D treatment, we plated $10^4$ cells per well in eight-well Lab-Tek slides (Nalge) 12 h before treatment to allow for cell attachment. Actinomycin D stock was diluted with DMEM with 10% vol/vol Cosmic Calf Serum (HyClone) to a final concentration of either 2.5 or 5 $\mu$g/ml. Two hundred microliters of actinomycin D solution was applied to each well. After treatment, slides were cooled on ice and processed for RNA-FISH.

## Results

**Experimental System.** We decided to test whether the structure of Xi could be disrupted in living cells by the administration of sequence-specific PNAs against the *Xist* transcript. PNA–transportan conjugates were selected on the basis of a careful survey of the sequence composition of the *Xist* transcript. Analysis of the cDNA encoding *Xist* shows four repeated regions [Fig. 1 (3)]. This analysis reveals the striking repetitive structure of four distinct regions, termed A, B, C, and D. Further analysis of the C region revealed that it consists of 14 direct repeats of ≈110–120 bases. These repeats contain several conserved regions, in three locations (termed I, II, and III), in which the motif UCAY was observed. The sequence GAGUCAU observed in region II was very conserved from repeat to repeat, with the UCAU being invariant. Similarly, the sequence GAAUUUCACUU in region III was almost invariant throughout the C region.

**Table 1. PNA used in this study**

| 5′-sequence | Name | Orientation |
|---|---|---|
| Experimental PNA sequences directed against the C-region | | |
| aaattctatgactctggaa | #pWS1246 | Antisense |
| aaattccatgactctgtaa | #pWS1248 | Antisense |
| aaattccatgactctagaa | #pWS1250 | Antisense |
| Experimental PNA sequences directed against the B-region | | |
| ggggcaggggctggggca | #pWS1458 | Antisense |
| Experimental PNA sequences directed against the D-region | | |
| ggagaaatagacacacaaa | #pWS1380 | Antisense |
| Nonspecific control | | |
| cggactaagtccattgc | #pWS1252 | Scramble sequence |
| PNA control: Sense sequences directed against the C-region | | |
| ttccagagtcatagaattt | #pWS1247 | Sense |
| ttacagagtcatggaattt | #pWS1249 | Sense |
| PNA control: 3-bp mismatch on C region | | |
| aaattccacagctctggaa | #pWS1290 | Antisense |
| PNA control: Directed to the 3′ end (used as a cocktail) | | |
| ctgattgttgtcaattttattattc | #pWS1295 | Antisense |
| tttatgcaagaatgttaaac | #pWS1296 | Antisense |
| tcaatttggtctttcgttcctccag | #pWS1284 | Antisense |

These motifs have been observed in other systems where they are involved in RNA–protein interactions (34–36). Several lines of evidence, including x-ray crystallographic analysis, show that the NOVA2/NOVA1-binding site is UCAY, where Y stands for a pyrimidine. Furthermore, the sequence GAGUCAU was shown to be an optimal binding site for the NOVA2 protein by *in vitro* selection (SELEX) experiments (34–36).

We performed RNA folding experiments with sequences from the C-repeat region, by using the MFOLD software (see *Experimental Procedures*). The modeling experiments revealed that sequences from the C region consistently formed ordered structures, which we call RNA fingers. These RNA fingers have consensus hairpin structures seen in other systems (35).

These observations provided the rational basis for our design of targeting PNA conjugates (Table 1). Antisense PNAs specific for the C region were designed to span region II containing one of the UCAY sites in the RNA fingers. PNA conjugates were also directed at the B and D regions. Also used were the sense version of the *Xist* C-region-binding antisense PNA, an antisense *Xist*-binding PNA with three sequence mismatches, a set of three PNA conjugates (used as a mixture, bind at the *Xist* 3′ end) (22, 37), and a scrambled sequence control.

To determine the effect of PNA on *Xist* binding to Xi, we treated tetraploid male and female murine C57BL/6 dermal fibroblasts with PNA conjugate under optimal conditions (see *Experimental Procedures*). We chose tetraploid cells because a measure of the effect of PNA administration would be more credible if it altered both copies of Xi. The cells were then fixed and hybridized to a full length *Xist* DNA probe to visualize the location of *Xist* RNA by RNA-FISH (see *Experimental Procedures*). The cells were then analyzed under an epifluorescence microscope for the association of *Xist* with Xi.

**PNA Conjugates Do Not Alter Steady-State Levels of *Xist* RNA.** The first concern in this project was whether antisense PNA-conjugate administration would have an effect on the steady-state levels of RNA. Thus, we first designed experiments to determine the effect of C-region antisense PNA on *Xist* steady-

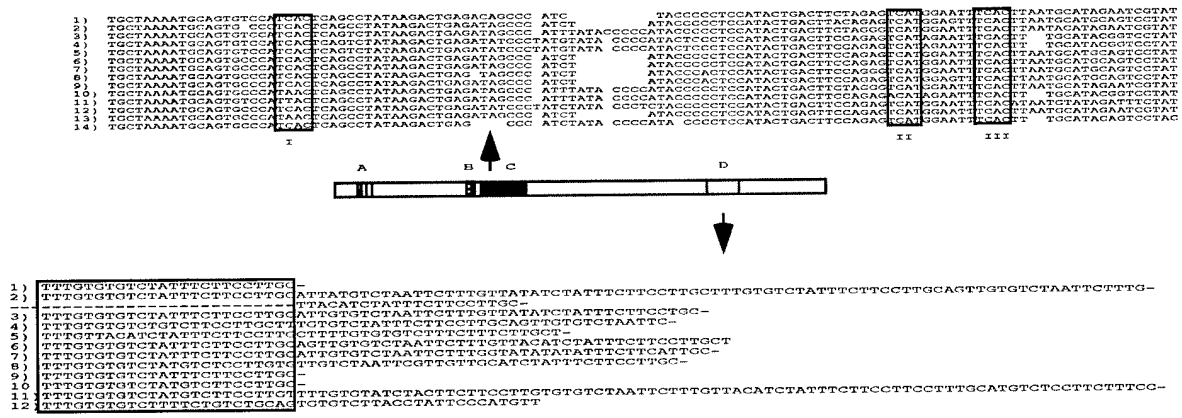

**Fig. 1.** Description of murine *Xist* repetitive regions used in this study. Schematic line drawing of *Xist*. Repetitive sequence regions are colored and labeled with letters A, B, C, and D (3). The majority of *Xist* transcripts (>90%) do not contain the A region, and thus it was not investigated. Regions C and D are approximately the same size, and antisense oligomers for the C region are directed at the region labeled II and for the D region against the boxed motif.

state RNA levels. Total RNA was isolated from fibroblasts treated with PNA-conjugate for a variety of times. Quantitative reverse transcription–PCR (RT-PCR) was performed by using the TaqMan system (see *Experimental Procedures* and Fig. 3*A*). No difference in *Xist* RNA steady-states levels was observed after PNA treatment.

**Specific PNA Conjugates Cause Loss of the *Xist* Body.** As steady-state levels of *Xist* did not change after PNA treatment, we wished to determine whether other aspects of *Xist* metabolism might have been altered by treatment. Therefore, RNA-FISH experiments on murine fibroblast cells were performed to identify the presence of the macromolecular complex called the *Xist* body, which has been shown to be congruent with Xi. PNA conjugates directed against the B, C, and D regions were compared. The results of these experiments were tabulated to yield the percentage of cells with *Xist* bodies after different PNA-conjugate treatments (Table 2). Representative microscopic fields are presented in Fig. 2. Cells were treated with the PNA conjugates, as indicated in Fig. 2, Table 2, and notes therein. The results of this experiment showed a clear effect of antisense PNA-conjugates directed to the C region of the *Xist* RNA. PNAs directed against the C region caused the loss of localized *Xist* RNA and the disappearance of the *Xist* body. Antisense PNAs against the B and D regions showed no effect. PNA conjugates, that also had no effect on *Xist* binding, were: C region sense PNA conjugates, scramble sequence PNA conjugate, mismatch C-region antisense PNA conjugate, and a mixture of antisense PNA conjugates directed to the 3′end of *Xist*.

To evaluate the mechanism of action of PNA on *Xist* body loss, we decided to compare the effects of PNA treatment to the effects of the RNA polymerase inhibitor actinomycin D. To evaluate the presence of *Xist* body loss after each treatment, RNA-FISH was performed. The time course of antisense C-region PNA conjugate activity was compared with matched cell cultures treated with actinomycin D. Results were determined by counting percentage of cells with *Xist* bodies. The results of this experiment (see Fig. 3*B*) show that PNA conjugates exhibit a pronounced effect after 18 h of action, with complete activity after 24 h. In contrast, actinomycin D showed complete activity (loss of *Xist* body) in 6 h at 5 $\mu$g/ml.

**Specific PNA Conjugates Cause Loss of the Histone mH2A Body [Macrochromatin Body (MCB)].** Previously, it has been shown that histone mH2A forms associates with the Xi (9). To establish whether the MCB would disappear after PNA treatment, experiments using the antisense C-region PNA conjugate were performed. PNA conjugate activity was evaluated by Immuno-FISH by using antibodies against the variant histone macrohistone H2A (9). The percentage of cells showing the presence of the macrohistone body was determined as a function of hours of PNA conjugate treatment. The results were plotted with RNA-

**Table 2. Numerical results of RNA-FISH experiments depicted in Fig. 2**

| PNA name | PNA description | No. cells with *Xist* body | No. cells without *Xist* body | Total no. cells | Percent cells with *Xist* body |
|---|---|---|---|---|---|
| pWS1246 | Xist antisense "C" region | 12 | 540 | 552 | 2.1 |
| pWS1248 | Xist antisense "C" region | 44 | 526 | 570 | 7.7 |
| pWS1250 | Xist antisense "C" region | 0 | 505 | 505 | 0 |
| pWS1290 | Xist antisense "C" region with 3bp-mismatch | 468 | 72 | 546 | 86.6 |
| pWS1295/1296/1284 | Xist antisense to 3′ end | 476 | 53 | 529 | 89.9 |
| pWS1249 | Xist sense "C" region | 345 | 167 | 512 | 67.3 |
| pWS1247 | Xist sense "C" region | 427 | 179 | 606 | 70.4 |
| pWS1458 | Xist antisense "B" region | 479 | 59 | 538 | 89.0 |
| pWS1380 | Xist antisense "D" region | 451 | 53 | 504 | 89.5 |
| pWS1252 | Scramble sequence | 429 | 106 | 535 | 80.1 |
| Female cells no PNA | N/A | 501 | 50 | 551 | 90.9 |
| Male cells no PNA | N/A | 0 | 510 | 510 | 0 |

The images presented in Fig. 2 are representative of the response of cells in the experiment. To quantitate the response of cells to PNA-transportan treatment, we counted the cells with *Xist* bodies. The total number of cells counted and number of cells with and without *Xist* bodies, as well as the percentages, are presented. Treatment time was 36 h.

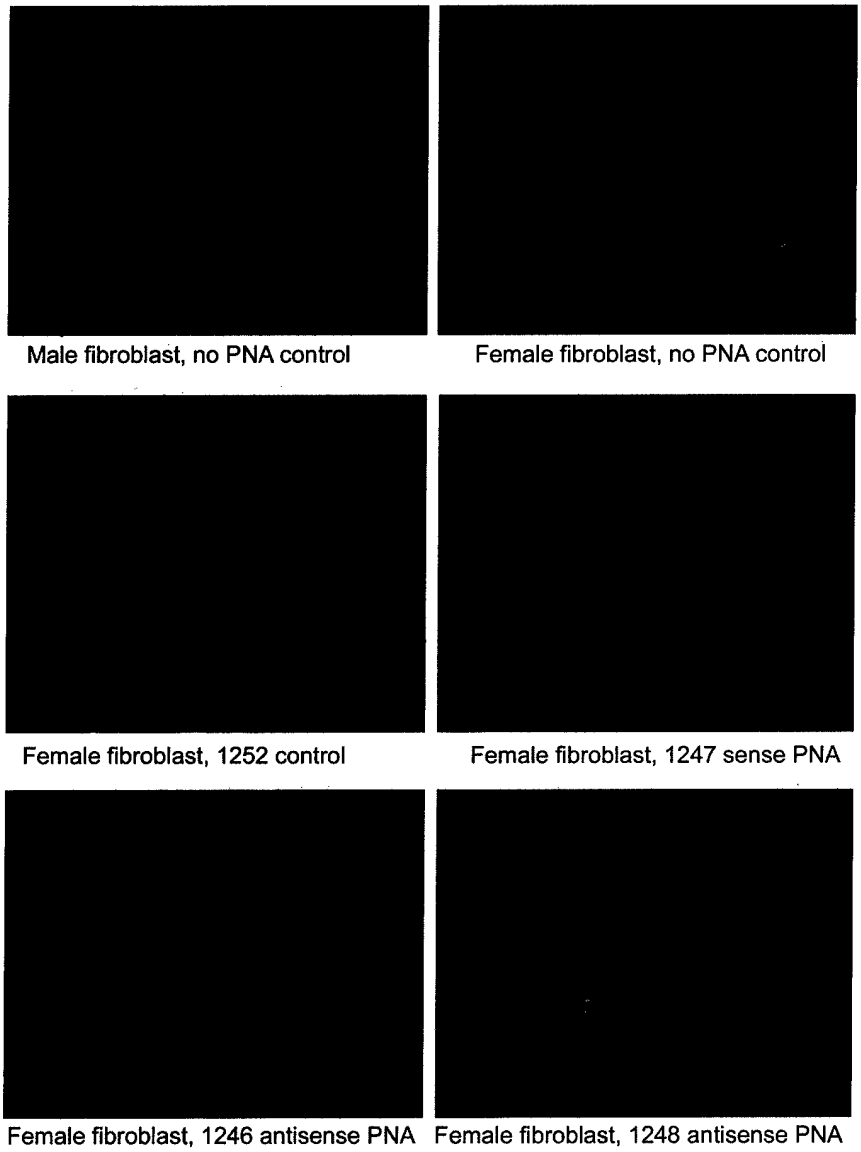


**Fig. 2.** RNA-FISH of fibroblast cells from various PNA treatments. Female and male fibroblastic cells were treated for 36 h as indicated. After treatment, cells were fixed and processed for RNA-FISH .

FISH time-course results. In these experiments, the MCB was seen to disappear after C-region antisense PNA treatment (see Fig. 3*B*).

***Xist* Must Bind to the X-Chromosome to Function.** To show that *Xist* binding is essential for *Xist* function, diploid female ES cells were differentiated *in vitro*; during the course of differentiation PNA conjugates were added to the media. After the experimental course, samples were processed for analysis.

The effect of PNA administration on the steady-state levels of *Xist* and *Tsix* was measured before or during differentiation. The results of PNA (pWS1248) administration were evaluated by quantitative RT-PCR (TaqMan) for *Xist* and *Tsix*. *Xist* was up-regulated normally in the presence of PNA (see Fig. 3*C*). *Tsix* was down-regulated as previously reported (14, 15) (see Fig. 3*C*). Similarly, PNA administration does not alter expression for genes at autosomal locations. Genes at autosomal locations were evaluated by quantitative RT-PCR. The loci evaluated were *Idh1*:(Chromosome 1), *Dnmt3b*:(Chromosome 2), and *Hmg2*:(Chromosome 8) (see Fig. 3*C*).

PNA-conjugate treatment does not alter the capacity of female ES cells to differentiate normally, as manifested by the appearance of the *Xist* body. Table 3 shows the effect of either

**Table 3. Effect of control PNAs on *Xist* body formation in female ES cells**

| PNA | No signal | *Xist* double dot | *Xist* body | Number nuclei | Percent *Xist* bodies |
|---|---|---|---|---|---|
| (−) | 17 | 196 | 318 | 531 | 60 |
| 1290 | 32 | 207 | 359 | 598 | 60 |
| 1252 | 20 | 208 | 307 | 535 | 57 |

We wished to test whether PNA administration by itself could alter the capacity of ES cells to differentiate normally and form *Xist* bodies. We evaluated scrambled sequence (pWS1252) or mismatched sequence (pWS1290) PNAs for their capacity to prevent *Xist* body formation during differentiation of female ES cells (see *Experimental Procedures*). Female ES cells were differentiated and treated with PNA–transportan conjugates for 6 days. The cells were then fixed and assayed for the number of *Xist* bodies by RNA-FISH.

**Table 4. Effects of PNA administration on X-linked gene expression as measured by RNA-FISH experiments**

| | *Mecp2* single dot | *Mecp2* double dot | Number nuclei | Percent single dot | Percent double dot | (+) *Xist* bodies | (−) *Xist* bodies | Number nuclei | Percent *Xist* bodies |
|---|---|---|---|---|---|---|---|---|---|
| (−)PNA | 288 | 264 | 552 | 52 | 48 | 332 | 220 | 552 | 60 |
| (+)PNA | 65 | 415 | 480 | 13 | 87 | 50 | 430 | 480 | 10 |

To quantitate the response of cells to PNA–transportan (pWS1248) treatment, we counted the cells with *Xist* bodies, one or two sites of expression for either *Pgk1* or *Mecp2*, as measured by RNA–FISH. The total number of cells counted, number of cells with and without *Xist* bodies, and the number of cells with one or two spots of expression for *Mecp2* are presented.

scramble PNA (pWS1252) or mismatch PNA (pWS1290) on the formation of the *Xist* body. Treated cultures showed the same number of *Xist* bodies as untreated cultures. Further, in Table 3, 60% is a measure of the percentage of female ES cells that have differentiated during the experimental time course of 6 days.

We evaluated the specific effect of *Xist* antisense PNA administration on coating/binding of *Xist* in ES cell culture. Treated and untreated female ES cells at day 6 of differentiation were examined by RNA-FISH. PNA-untreated and differentiated ES cells show the expected *Xist* body and a single site of expression for *Pgk1* and Mecp2 loci. In contrast, PNA-treated and differentiated ES cells do not show a *Xist* body and exhibit two sites of expression for *Pgk1* and loci. When transcription at the murine β-actin locus on Ch 4 was determined by RNA-FISH, no difference between untreated ES cells and ES cells treated with PNA pWS1248 was observed. The RNA-FISH data are tabulated in Tables 3–6. In Tables 3–6, over 500 interphase nuclei from treated and untreated cultures were examined after differentiation and scored for the presence of *Xist* body and the suppression of transcription from either *Pgk1*, *Mecp2*, or β-actin loci.

## Discussion

Here we report the first functional demonstration that the large nontranslated RNA, *Xist*, can be organized into functional domains. These studies also confirm that the function of *Xist* is mediated through its binding to X-chromosome. It is also apparent from the kinetic studies in this report that macro histone H2A and *Xist* are intimately associated in a macromolecular complex on Xi.

The process of *Xist*-mediated gene repression observed in X-chromosome inactivation is thought to be a special case of a general developmental pathway for chromatin reorganization directed at transcriptional repression (38, 39). As the nontranslated RNA *Xist* has been shown to be necessary and sufficient to initiate the silencing pathway (21), it is of great interest to discern how the nontranslated RNA *Xist* interacts with the X-chromosome to initiate silencing. One of the most striking qualities of the *Xist* transcript is its localization to the Xi. To make the connection between the structure of *Xist*, its binding to the X-chromosome, and silencing of gene expression, a functional analysis of the *Xist* transcript is vital. Here we report evidence for the existence of a functional domain in *Xist* responsible for *Xist* coating/binding to Xi. We argue that this functional domain appears to be encoded in the repetitive sequence in the large first exon (C region, Fig. 1).

The data presented here definitively demonstrate that C-region antisense PNA-conjugates can alter *Xist* coating/binding. PNA conjugates directed to other regions of *Xist* do not alter its localization. Perfectly matched PNAs against the C region can bind a maximum of seven times; if one or two mismatches are allowed, they can bind a maximum of 14 times. D-region PNA can bind to six perfectly matched target sites, and if mismatches are permitted to at least eight sites. Thus, the amounts of PNA bound to the C and D regions are similar. Nonetheless, antisense PNAs directed against the *Xist* B and D regions, and 3′ end have no effect on *Xist* coating. In addition, the pWS1290 PNA conjugate, which represents a 3-bp mismatch of the active antisense PNA (pWS1248) PNA conjugate, has no effect on coating/binding.

Time-course results show striking differences between the kinetics of PNA action and actinomycin D inhibition on *Xist* body loss. In our experiments, complete loss of *Xist* bodies was observed 6 h after actinomycin D treatment, whereas for PNA treatment, this period was 18–24 h. In the context of no change in *Xist* RNA expression (see Fig. 3), PNA treatment is governed by a slow kinetic path, which remains to be defined. We postulate potential steric factors that might contribute to the inaccessibility of *Xist* to PNA action.

Under conditions where the intracellular levels of steady-state *Xist* RNA do not change (see Fig. 3), the slow step in *Xist* body loss is comparable to the kinetics of MCB loss. The variant histone mH2A has been shown to be associated with the Xi during development (8, 9). Previous attempts at describing the association between the *Xist* body and the MCB required the recombinational loss of the *Xist* gene and subsequent loss of *Xist* expression to measure the association between *Xist* body and the MCB (7). The current kinetic study does not suffer this limitation, as PNA uptake is essentially instantaneous, with high concentrations of PNA available in the nucleus in minutes; further, the rate of change for *Xist* and the MCB can be monitored from the inception of PNA administration. The data presented here would suggest that the binding of *Xist* and the MCB to the Xi are equally destabilized by antisense PNA against the C region. This simultaneous destabilization suggests concerted breakdown of a macromolecular structure by PNA action. The kinetics of dissociation for *Xist* and macrohistone H2A imply that the MCB and *Xist* are in intimate association in the Xi. The loss of the MCB relative to *Xist* is the predicted result if one were to suppose a hierarchical structure where *Xist* represents a foundation on which proteins like mH2A might associate with Xi.

**Table 5. Effects of PNA administration on X-linked gene expression as measured by RNA-FISH experiments**

| | *Pgk1* single dot | *Pgk1* double dot | Number nuclei | Percent single dot | Percent double dot | (+) *Xist* bodies | (−) *Xist* bodies | Number nuclei | Percent *Xist* bodies |
|---|---|---|---|---|---|---|---|---|---|
| (−)PNA | 241 | 277 | 518 | 46 | 53 | 295 | 223 | 518 | 57 |
| (+)PNA | 122 | 396 | 518 | 23 | 76 | 64 | 454 | 518 | 12 |

To quantitate the response of cells to PNA–transportan (pWS1248) treatment, we counted the cells with *Xist* bodies, one or two sites of expression for either *Pgk1* or *Mecp2*, as measured by RNA–FISH. The total number of cells counted, number of cells with and without *Xist* bodies, and the number of cells with one or two spots of expression for *Pgk1* are presented.

**Table 6. Effect of PNA administration on autosomal expression measured by RNA-FISH**

| | β-Actin single dot | β-Actin double dot | Number nuclei | Percent single dot | Percent double dot |
|---|---|---|---|---|---|
| (−)PNA | 16 | 530 | 546 | 3 | 97 |
| (+)PNA | 20 | 550 | 570 | 3 | 97 |

PNA conjugate-treated cells were evaluated for alterations in β-actin expression. Actin mRNA expression was evaluated for single or double sites of expression and then tabulated.

Two lines of evidence support the conclusion that PNA activity is caused by the interference with a particular RNA structure. The theoretical folding of the *Xist* RNA by MFOLD algorithm predicts highly stable hairpins. Thus PNA activity against these structures is based on the specificity, not the accessibility, of PNA to *Xist* RNA. Secondly, unique kinetics of MCB disruption by PNA interference is different from actinomycin D time-course, implying that dissociation of *Xist* and the MCB is coupled with, and not attributable to, an alteration in *Xist* metabolism per se.

Coating/binding of *Xist* is necessary for X-inactivation. During female development, a number of changes occur within the XIC. On Xi, *Xist* expression is up-regulated by an unknown mechanism, and *Tsix* expression is silenced. On Xa, expression of both *Xist* and *Tsix* is suppressed. Formally, it has not been proven which of these steps is required for the biochemical process of silencing. Using antisense PNA conjugates that can bind only to the *Xist* transcript, we demonstrate that, despite the ordered progression of developmental stages in differentiating female ES cells, if up-regulated *Xist* cannot bind to the Xi, there is no chromosomal silencing. Thus, the described experiments distinguish between the expression of *Xist* and the correct localization of the transcript with respect to function.

One of the major limitations for analysis of the nuclear compartment and *Xist* function within it has been the lack of a functional technology that would specifically connect DNA sequence-based knowledge to the biochemistry of the nucleus. One of the goals of the present work was to make the connection between *Xist* sequence information and its function. The technology described here, P-IMP, is a functionally based technology to probe the RNA–RNA and RNA–protein interactions that occur in the living cell. P-IMP experiments can be envisaged for a variety of processes that involve nontranslated RNAs, including splicing, telomere formation and maintenance, gene imprinting, and chromatin-mediated gene silencing. It will be exciting to envisage a high-throughput functional genomic analysis of the nuclear compartment by P-IMP.

We thank Ralph A. Casale and Eric G. Anderson for their dedication and skill in the preparation of the PNA conjugates. We acknowledge the support of U.S. Army Prostate Cancer Research Award DAMD17–99-1–9032 and National Institutes of Health Grants R21 CA81732 and RO1 GM61079 (to W.M.S.).

1. Brockdorff, N. (1998) *Curr. Opin. Genet. Dev.* **8,** 328–333.
2. Brockdorff, N. & Duthie, S. M. (1998) *Cell Mol. Life Sci.* **54,** 104–112.
3. Brown, C. J., Hendrich, B. D., Rupert, J. L., Lafreneire, R. G., Xing, Y., Lawrence, J. & Willard, H. F. (1992) *Cell* **71,** 527–542.
4. Lyon, M. F. (1961) *Nature (London)* **190,** 372–373.
5. Gartler, S. M., Chen, S. H., Fialkow, P. J., Giblett, E. R. & Singh, S. (1972) *Nat. New Biol.* **236,** 149–150.
6. Barr, M. L. & Bertram, E. G. (1949) *Nature (London)* **163,** 676–677.
7. Csankovszki, G., Panning, B., Bates, B., Pehrson, J. R. & Jaenisch, R. (1999) *Nat. Genet.* **22,** 323–324.
8. Mermoud, J. E., Costanzi, C., Pehrson, J. R. & Brockdorff, N. (1999) *J. Cell Biol.* **147,** 1399–1408.
9. Costanzi, C. & Pehrson, J. R. (1998) *Nature (London)* **393,** 599–601.
10. Rasmussen, T. P., Mastrangelo, M. A., Eden, A., Pehrson, J. R. & Jaenisch, R. (2000) *J. Cell Biol.* **150,** 1189–1198.
11. Russell, L. B. & Mongomery, C. S. (1965) *Genetics* **52,** 470–471.
12. Russell, L. B. (1963) *Science* **140,** 976–978.
13. Lee, J. T., Strauss, W. M., Dausman, J. A. & Jaenisch, R. (1996) *Cell* **86,** 83–94.
14. Lee, J. T., Davidow, L. S. & Warshawsky, D. (1999) *Nat. Genet.* **21,** 400–404.
15. Lee, J. T. & Lu, N. (1999) *Cell* **99,** 47–57.
16. Brockdorff, N., Ashworth, A., Kay, G. F., McCabe, V. M., Norris, D. P., Cooper, P. J., Swift, S. & Rastan, S. (1992) *Cell* **71,** 515–526.
17. Borsani, G., Tonlorenzi, R., Simmler, M. C., Dandolo, L., Arnaud, D., Capra, V., Grompe, M., Pizzuti, A., Muzny, D. & Lawrence, C. (1991) *Nature (London)* **351,** 325–329.
18. Brown, C. J., Ballabio, A., Rupert, J. L., Lafreniere, R. G., Grompe, M., Tonorenzi, R. & Willard, W. F. (1991) *Nature (London)* **349,** 38–44.
19. Marahrens, Y., Panning, B., Dausman, J., Strauss, W. & Jaenisch, R. (1997) *Genes Dev.* **11,** 156–166.
20. Penny, G. D., Kay, G. F., Sheardown, S. A., Rastan, S. & Brockdorff, N. (1996) *Nature (London)* **379,** 131–137.
21. Wutz, A. & Jaenisch, R. (2000) *Mol. Cell.* **5,** 695–705.
22. Hong, Y. K., Ontiveros, S. D., Chen, C. & Strauss, W. M. (1999) *Proc. Natl. Acad. Sci. USA* **96,** 6829–6834.
23. Hong, Y. K., Ontiveros, S. D. & Strauss, W. M. (2000) *Mamm. Genome* **11,** 220–224.
24. Nielsen, P. E., Egholm, M., Berg, R. H. & Buchardt, O. (1991) *Science* **254,** 1497–1500.
25. Egholm, M., Buchardt.O., Nielsen, P. E. & Berg, R. H. (1992) *J. Am. Chem. Soc.* **114,** 1895–1897.
26. Egholm, M., Buchardt, O., Christensen, L., Behrens, C., Freier, S. M., Driver, D. A., Gerg, R. H., Kim, S. K., Norden, B. & Nielsen, P. E. (1993) *Nature (London)* **365,** 566–568.
27. Verheijen, J. C., Van der Marel, G. A., Van Boom, J. H., Bayly, S. F., Player, M. R. & Torrence, P. F. (1999) *Bioorg. Med. Chem.* **7,** 449–455.
28. Bonham, M. A., Trown, S., Boyd, A. L., Brown, P. H., Bruckenstein, D. A., Hanvey, J. C., Thomson, S. A., Pipe, A., Hassman, F., Bisi, J. E., *et al.* (1995) *Nucleic Acids Res.* **23,** 1197–1203.
29. Brewer, G. & Ross, J. (1989) *Mol. Cell. Biol.* **9,** 1996–2006.
30. Uhlmann, E. (1998) *Biol. Chem.* **379,** 1045–1052.
31. Pooga, M., Lindgren, M., Hallbrink, M., Brakenhielm, E. & Langel, U. (1998) *Ann. N.Y. Acad. Sci.* **863,** 450–453.
32. Pooga, M., Soomets, U., Hallbrink, M., Valkna, A., Saar, K., Rezaei, K., Kahl, U., Hao, J. X., Xu, X. J., Wiesenfeld-Hallin, Z., *et al.* (1998) *Nat. Biotechnol.* **16,** 857–861.
33. Pooga, M., Hallbrink, M., Zorko, M. & Langel, U. (1998) *FASEB J.* **12,** 67–77.
34. Jensen, K. B., Musunuru, K., Lewis, H. A., Burley, S. K. & Darnell, R. B. (2000) *Proc. Natl. Acad. Sci. USA* **97,** 5740–5745. (First Published May 16, 2000; 10.1073/pnas.090553997)
35. Lewis, H. A., Chen, H., Edo, C., Buckanovich, R. J., Yang, Y. Y., Musunuru, K., Zhong, R., Darnell, R. B. & Burley, S. K. (1999) *Struct. Folding Des.* **7,** 191–203.
36. Lewis, H. A., Musunuru, K., Jensen, K. B., Edo, C., Chen, H., Darnell, R. B. & Burley, S. K. (2000) *Cell* **100,** 323–332.
37. Memili, E., Hong, Y.-K., Kim, D. K., Ontiveros, S. D. & Strauss, W. M. (2001) *Gene* **266,** 131–137.
38. Wolffe, A. P. & Matzke, M. A. (1999) *Science* **286,** 481–486.
39. Bird, A. P. & Wolffe, A. P. (1999) *Cell* **99,** 451–454.