Mys retrotransposons in Peromyscus leucopus and transgenic Mus musculus

Daniel S.Pine*, Eric C.Bourekas and S.Steven Potter⁺

Children's Hospital Research Foundation, Elland and Bethesda Aves, Cincinnati, OH 45229, USA

Received December 4, 1987; Revised and Accepted February 2, 1988

ABSTRACT

The mys family of retrotransposons exhibits an interesting phylogenetic distribution with 500-1000 copies per haploid genome in the white-footed mouse Peromyscus leucopus and no copies detectable in the house mouse Mus musculus, even though most other repeated sequences are shared by these two species. Comparison of the DNA sequences from the 3' ends of five mys elements show that insertion occurs just upstream of a well-conserved 11 bp target sequence. Transcription patterns of the elements in brain, liver, heart, kidneys and gonads in Peromyscus leucopus are analyzed. Transcripts are found in all tissues examined, but they are remarkably heterogeneous in size. When four cloned elements are introduced into transgenic Mus musculus, however, discrete patterns of expression are revealed. Furthermore, a study of the structure of the concatemers of mys elements in the transgenic mice demonstrates that recircularization of injected linear molecules is an important event in concatemer generation.

INTRODUCTION

Using a phylogenetic DNA screening procedure we have found a family of retrotransposons with 500-1000 elements per haploid genome in the white-footed mouse Peromyscus leucopus, but none detectable in the genome of the house mouse Mus musculus, even though most other families of repetitive elements are shared by these two species (1). These elements, called mys from the Greek word for mouse, exhibit many of the structural features of retroviruses. A sequenced element of 2843 bp carries long terminal repeats (LTRs) of 343 bp, a polypurine tract at one inner LTR junction and a lysine tRNA binding sequence at the other, open reading frame encoding an amino acid sequence with homology to reverse transcriptase, and generated a six bp target site duplication (1).

These data suggest that the mys family might be derived from a retrovirus in relatively recent evolutionary time, since the divergence of the white-footed mouse and the house mouse. Nevertheless the elements are quite distinct from retroviruses because of their small size, limited coding potential, and unusual structural organization (1). If one assumes that the family began from a single provirus then it is clear that there has been a dramatic expansion of the number of mys elements in the white-footed mouse genome. Such an expansion obviously cannot continue indefinitely and a question of key interest is how the growth of such a family is halted. In Drosophila it is known that some elements, such as P elements can rapidly increase in number after initial introduction, but they then soon become relatively inactive (2). Other "DNA intermediate" elements in Drosophila,

such as the Foldback elements, show steady state levels of activity that do not change over decades (3). And although there is evidence that copy number of retrotransposons in Drosophila is controlled (4), the mechanism is not understood. In this report we investigate possible regulatory mechanisms that might be involved in controlling the copy number of mys elements.

Retrotransposons must be transcribed in order to generate new copies of elements, and one possible control mechanism might be to severely limit their transcriptional activity. We have, therefore, analyzed the transcription patterns of mys elements in brain, liver, heart, kidney and gonads in order to gain some measure of their current ability to function. We have also directly tested four mys elements for transpositional activity, and further studied their transcription, by introducing them into transgenic M. musculus.

Another possible copy control mechanism could be based on a strong target site preference during transposition of the elements. Once all available target sites are saturated, or occupied by elements, then further expansion of the family would cease. We have previously noted that the mys elements are indeed found inserted into similar sequences and in this report these observations are extended with the comparison of five target sites, revealing a well-conserved consensus sequence.

Finally, in the process of analyzing transgene concatemers certain structures were observed that partially elucidate the mechanism of concatemer formation.

MATERIALS AND METHODS

DNA Purification and Restriction Enzymes

Mys elements flanked by P. leucopus genomic DNA were cloned into the Bam HI site of pBR322. DNA was prepared by the alkaline lysis procedure (5) or by the rapid boil method (6) supplemented with the previously described modifications (7). All restriction enzymes were from New England Biolabs and digestion conditions were as recommended by the manufacturer. A five to ten fold overdigestion was typically performed.

DNA Sequencing

Sequencing was performed according to the procedure of Maxam and Gilbert (8) or by the dideoxy chain termination method with the modifications of Duncan (9).

Transgenic Mice

Mys clones 1, 2, 4 and 7 (1) were linearized at the Sal I site in pBR322. The four plasmids were then pooled in approximately equal ratios and twice purified in CsCl equilibrium gradients without ethidium bromide. Var Lac Oid ultrapure CsCl was used. Banded DNA was retrieved by collecting gradient fractions and electrophoresing 5 μ l aliquots directly on a 0.8% agarose gel to determine which contain the DNA. These fractions were then dialyzed for 48 hours against four, one liter volumes of injection buffer (5 mM NaCl, 5 mM Tris-HCl pH 7.4, 0.5 mM EDTA) and assayed by spectrophotometry. The DNA solution was diluted to 13.78 μ g/ml, giving 2,000 plasmids per 2 μ l injection volume, and microinjected into the male pronuclei of C57BL/6J X C3H F₂ zygotes. Transgenic mice were identified by assaying for the presence of mys element sequences in tail DNA, purified according to the procedure of D. Hanahan (10).

DNA Electrophoresis, Blotting, and Hybridization

Following digestion, 10 μ g of genomic DNA was loaded per well and electrophoresed through 0.75% agarose gels. Nucleic acid concentrations were determined by spectrophotometry or by a fluorimetric method (11) using

the DNA-specific fluorescent dye bisbenzimidazole (Hoechst 33258). Copy number control lanes were prepared by mixing the appropriate amount of positive marker DNA (*mys* 1, *mys* 2, or the four *mys* elements combined) with 10 µg of control (negative) house mouse DNA, digesting with the desired enzyme(s), and electrophoresing in parallel with the 10 µg genomic samples. In order to standardize the electrophoretic mobility and transfer efficiency of different sample lanes, genomic digests of less than 10 µg were also supplemented with enough control DNA (prior to digestion) to make a total of 10 µg/well. Following photography, gels were alkaline blotted onto GeneScreen Plus by the procedure of Chomczynski and Qasba (12) and hybridized at 65°C with 10% dextran sulfate. Filters were washed according to the protocol supplied by New England Nuclear and autoradiographs were exposed at -70°C with a Dupont Lightning plus intensifying screen.

RNA Isolation, Electrophoresis, Blotting, and Hybridization

Tissues from *P. leucopus* and three *mys* transgenic *M. musculus* lines were removed, frozen in liquid nitrogen, ground to a fine powder using a mortar and pestle, then suspended in extraction buffer (4M guanidine thiocyanate, 0.2% sarkosyl, 25 mM sodium citrate, pH 7.0, 0.2 M β-mercaptoethanol) and Dounce homogenized. RNA was then pelleted through 5.7 M CsCl + 0.1 M EDTA using a SW28 rotor (70,000 g, 20°C, 18 hr), dissolved in TE (10 mM Tris pH 7.4, 1 mM EDTA), EtOH precipitated and redissolved in TE. Portions of the total RNA were loaded onto an oligo(dT)cellulose column

Figure 1. Genomic DNA sequences flanking the 3' end of five *mys* elements and the consensus target site deduced. Thin underlines mark the position of the ATCC quadruplet. The thick underlining indicates seven additional target base pairs. All eleven conserved bases are printed in bold type and their consensus sequence is shown beneath the *mys* element sequence data. Exceptions to the consensus in *mys* 4 and *mys* 6 are indicated by a thick box below the underlining. An AATAAA stretch is also present in the flanking regions of *mys* 1 and *mys* 2 (positions 24-30) but could not be found within 80 bases of the 3' ends of the other three elements.

Figure 2. Northern blots showing *mys* transcription in male *Peromyscus leucopus* mice. RNA was electrophoresed through a 1% agarose gel containing formaldehyde, blotted to GeneScreenPlus and simultaneously hybridized to ^{32}P -labeled 2.5 kb *Eco* RV *mys* 1 probe and ^{32}P -labeled 6.1 kb mouse β -globin (digested with *Bam* HI) probe as described in Materials and Methods. The 0.7 kb band seen is the full length β -globin transcript. The markers indicated are ribosomal RNA, 28S and 18S from calf liver and 23S and 16S from *E. coli*. The markers were electrophoresed on the same gels but then cut-off, stained separately with ethidium bromide, and photographed. Lanes A-E contain 10 μg total RNA: lane A, brain; lane B, testes; lane C, kidney; lane D, liver; lane E, heart. Lane F, 3 μg brain poly A⁺ RNA; lane G, 3 μg liver poly A⁺. Lanes H-J, 10 μg poly A⁻ RNA: lane H, brain; lane I, testes; lane J, liver. Lane K, 12 μg control poly A⁺ liver RNA from *Mus domesticus*. Lane L, 4 μg liver poly A⁺ RNA; lane M, 4 μg testes poly A⁺ RNA. Lanes L and M are from different gels than the rest of the lanes. The arrow indicates the 3.1 kb transcript evident in lanes F, G and L.

(Type 3-Collaborative Research) to isolate poly A⁺ RNA. Once loaded with RNA, columns were washed with TES (TE + 0.5 M NaCl) and poly A⁺ RNA was eluted with TE and concentrated by ethanol precipitation. RNA samples were electrophoresed through 1% agarose gels containing formaldehyde (13), and then transferred onto GeneScreen Plus using 10X SSC as transfer solution. Filters were hybridized with 10% dextran sulfate at 65°C and washed according to the New England Nuclear protocol and autoradiographs exposed at -70°C with an intensifying screen.

Figure 3. Restriction enzyme maps of the four *mys* element clones inserted into transgenic mice, and hybridization probes used in Southern and Northern blot analyses. The thin line indicates the *pBR322* plasmid vector. The bars represent *P. leucopus* genomic DNA ligated into the *Bam* *HI* site. All four clones were linearized (as shown) with *Sal* *I* prior to microinjection. The *mys* elements are depicted with solid black arrows as the LTRs and hatched regions as the internal portion. The orientation of each element within the clone is signalled by the direction of arrows and hatch marks. The 3' end referred to in discussions of the target sequence (Fig. 1) is according to the orientation found in *mys* 1 and *mys* 4. H, *Hind* *III*; P, *Pvu* *II*; R, *Eco* *RI*; T, *Pst* *I*; V, *Eco* *RV*; X, *Xho* *I*. Hybridization probes are indicated by thin open bars below the clones. The *mys* 1 probes are as follows: 1) 1.2 kb *Eco* *RI*, *Hind* *III* fragment (subcloned into M13pml8 and M13pml9); 2) 2.5 kb *Eco* *RV* fragment; and 3) 0.7 kb *Pvu* *II* single-copy flanking fragment. The probe from *mys* 4 is a 0.8 kb *Hind* *III*, *Eco* *RV* flanking fragment. The *mys* 2 clone contains two additional *Pvu* *II* sites (not shown), which split its right most *Pvu* *II* fragment illustrated into pieces of approximately 1150, 195, and 65 bp.

Labelling DNA Probes

Probe fragments were labelled with ^{32}P dATP (New England Nuclear) by the primer extension method (14) using random hexanucleotide primers (Pharmacia 27-2166-01). We routinely achieved specific activities of 1×10^9 CPM/ μg with this technique. Alternatively, single stranded hybridization probes were made from M13 subcloned fragments using a hybridization probe primer (Pharmacia 27-1544-01) according to the procedure of Hu and Messing (15).

RESULTS

A Consensus Target Sequence

In order to determine if *mys* elements exhibit a target site preference we have analyzed the flanking DNA of five *mys* elements cloned from the white-footed mouse, *P. leucopus*, genome. Figure 1 shows the genomic

sequences flanking the 3' ends of these elements with conserved base pairs in bold type. Two sequences are found to be common to these target sites. There is an "ATCC" quadruplet near the last base of the target site duplication followed at a variable distance of 3-13 bases by the seven base sequence "T(T/G)AAGTT". While the precise spacing of the eleven conserved bases is somewhat flexible, the degree of homology is remarkable, with the five target sites having an average of 89% faithfulness to the consensus sequence (only six mismatches out of 55 bp). Although this degree of target specificity is unusual, it is nevertheless clear that unless there are other constraints on integration that are not revealed by our linear DNA analysis then there are many more target sites than mys elements present in the genome; the availability of target sites is likely not yet, therefore, a limiting factor in controlling mys element movement and maintaining copy number at present levels.

Mys Transcripts in Peromyscus leucopus

To study the transcription patterns of mys elements, RNA from various P. leucopus tissues was isolated. Following electrophoresis and blotting to GeneScreen Plus, filters were hybridized to a radiolabelled 2.5 Kb Eco RV segment from mys 1. As shown in Figure 2, hybridizing transcripts ranging in size from 0.5 kb to over 15 kb were detected in all tissues examined. Control hybridization of the same blots to a mouse B-globin probe revealed discrete bands of the predicted size for globin mRNA suggesting that RNA degradation had not caused the heterogeneous mys transcription pattern observed. It should be noted, however, that the various tissues tested carry different amounts of blood, accounting for the considerable differences in intensities of the control globin bands. In other controls pretreatment of samples with RNase eliminated all hybridization, while treatment with DNase did not change the patterns (data not shown). There was also no hybridization of the mys probe to control RNA from M. musculus mice which have no mys elements in their genome.

Quantitative differences in mys transcription levels were found amongst the sampled tissues. For poly A⁺ RNA, more mys transcripts were found in the brain and liver than in testes, kidney or heart, while for poly A⁻ RNA, the brain was dramatically richer in mys transcripts than the other tissues (Fig. 2). In addition, distinct hybridization bands were sometimes visible within the heterogeneous patterns obtained. In particular, the brain and liver poly A⁺ RNAs are rich in 3.1 kb mys homologous transcripts (lanes F, G, and L). This is of special interest because 3.1 kb corresponds roughly in size to a polyadenylated full-length mys transcript, initiated within one LTR and terminated in the other, as might be expected for a retrotransposon.

Mys Transcripts in Transgenic Mice

The pattern of heterogeneous transcripts observed in P. leucopus, a species with at least 1000 mys elements scattered throughout its diploid genome, could be the result of aberrantly structured elements (1) and/or mys sequences inserted into other transcriptional units. In order to investigate the in vivo transcription of a few elements within a small number of genomic contexts transgenic mice were generated by injecting high concentrations of the four cloned mys elements shown in Figure 3 into M. musculus zygotes.

Southern blot experiments using digested tail DNA from the three resulting transgenic mice (9-5, 9-6 and 17) confirmed that each mouse carried a mix of all four injected elements. The enzyme Pvu II cleaves once within each mys element, giving two fragments the sizes of which depend on the placement of flanking Pvu II sites. Since these flanking sequences were part of the injected clones and had been previously mapped, the sizes of the two unique Pvu II segments produced by each element were known. In Figure 4

Figure 4. Southern blots identifying all four injected *mys* elements within each transgenic mouse. *Pvu* II digested control and transgenic tail DNA was electrophoresed, immobilized on a filter, and hybridized to the 2.5 kb *Eco* RV probe from *mys* 1. The 5X and 1X copy-marker lanes (made by mixing *Sal* I linearized *mys* 1, 2, 4, and 7 clones in equal numbers) represent that number of copies of each *mys* clone per diploid genome. Eight characteristic *Pvu* II fragments are easily detectable in the 5X lane, however, the 3.0 and 3.1 kb bands are not separated. *Mys* 1 produces the 0.7 and 3.1 kb fragments; *mys* 2, the 1.1 and 8.8 kb; *mys* 4, the 3.0 and 4.9 kb; and *mys* 7, the 1.3 and 2.3 kb. The 17, 9-6, and 9-5 lanes each contain 10 μ g of digested DNA from one of the three resulting transgenic mice. Data here and elsewhere (not shown) indicates that mouse 9-6 has acquired approximately 50 *mys* elements total, with about 10-15 copies of each of the four different clones. Mice 17 and 9-5 were mosaic and copy number determinations were therefore done with their non-mosaic progeny. High molecular weight fragments in lane 9-5 have migrated more slowly than their counterparts in other lanes due to an artifact that has not been seen in other gels.

the presence of these eight characteristic bands is revealed by hybridization to the 2.5 kb *Eco* RV probe from *mys* 1 (Figure 3), thus proving that each mouse contains copies of all four injected elements.

Transcription patterns of *mys* sequences in the two lines of transgenic *M. musculus* mice analyzed were found to differ dramatically from those observed in *P. leucopus*. First, *mys* transcription in transgenic mice was only found in testes and heart, and not in brain, liver, kidney or ovaries (Figure 5). Second, the intense heterogeneous patterns are reduced in

Figure 5. Northern blot showing *mys* transcription in transgenic mice. RNA was electrophoresed, blotted, and hybridized as previously described in figure 2. The markers are the same as in figure 2 and the 0.7 kb band is again the full-length B-globin transcript. Lanes A-D contain 12 μ g of total RNA from the 'four band' transgenic line of mice: lane A, brain; lane B, testes; lane C, kidney; lane D, liver. Lane E, 9 μ g heart total RNA, from 'four band' mice. Lane F, 12 μ g ovarian total RNA, mix of '3 and 4 band' transgenic mice. Lanes G-K, 12 μ g total RNA, from 'thirteen band' mice: lane G, brain; lane H, testes; lane I, kidney; lane J, liver; lane K, heart. Lanes L-N, poly A⁺ RNA, 'four band': lane L, 1 μ g brain; lane M, 5 μ g kidney; lane N, 5 μ g liver. Lanes O-R, poly A⁺ RNA, 'thirteen band': lane O, 0.6 μ g brain; lane P, 5 μ g testes; lane Q, 5 μ g kidney; lane R, 5 μ g liver. Lane S, 12 μ g testes poly A⁻ RNA, 'thirteen band'. All the RNA, except ovarian, was isolated from male mice. The three arrows indicate the *mys* transcripts (7.2 kb, 4.8 kb, 3.1 kb) found in testes RNA.

transgenic mice, with small numbers of discrete bands now predominant. In both transgenic lines analyzed there are three testes-specific transcripts (3.1, 4.8, and 7.1 kb) seen in total RNA (lanes B and H). The 3.1 kb transcript is the only one that is polyadenylated (lane P) and it also corresponds in size to the abundant transcript observed in *P. leucopus* (Figure 2).

No Transposed Elements Are Found in Transgenic Mice

We investigated the possibility that one or more of the four elements inserted into the transgenic mice would be active and capable of transposing to a new genomic site. Such transposition events by members of the IAP family, which resemble *mys* and other copia-like elements in many respects, have been observed in cell lines from *M. musculus* (16-18). The chance of our observing such an event in transgenic mice was increased by the large numbers of elements injected (2000-3000 per egg) and by the *in vivo* recombination that may have created functional hybrid elements from inert parents.

If one assumes that the transposed element would integrate randomly at any genomic site containing the target sequence, then in most cases it would insert into a separate chromosome from that carrying the parent concatemer of *mys* elements. This chromosome would then segregate independently during meiosis, often generating progeny mice carrying only the single transposed *mys* element. In screening 30 first generation, 40 second generation mice and 70 third generation mice, however, none were found to carry single copy *mys* elements that might signify a transposition event (Figure 6 and data not shown).

Structure of Concatemer DNA

The foreign DNA integrated into the genome of the transgenic mouse is often found as a head to tail concatemer, but sometimes the structure has been found to be more complex, with interrupting islands of genomic DNA (19). The family tree in Figure 6 shows Southern blot patterns of the original transgenic mice, the second generation mice inheriting *mys* sequences, and representative third generation mice. Genomic DNAs were digested with *Eco* RV which cleaves within each element LTR, normally releasing segments of uniform 2.5 kb size that hybridize to the internal *Eco* RI- *Hind* III *mys* probe (described in Figure 3).

The numerous additional fainter bands observed in Figure 6 likely represent junction sequences where the elements are interrupted by genomic DNA. While a concatemer will normally have two terminal junctions at either end, the large number of faint bands shown in Figure 6 presumably reflects the presence of many internal "islands of genomic DNA" interrupting the concatemer *mys* elements and thereby generating numerous junctions. The patterns of faint bands observed serve to distinguish the various *mys* transgenic lines of mice, and they are named accordingly. For examples, the "four band" line pattern is shown in Figure 6 lane O and the "thirteen band" line pattern is seen in Figure 6 lane H.

These complex concatemer structures were observed to be somewhat unstable. As shown in Figure 6, progeny mice with fewer of the faint "satellite" bands were sometimes observed, (for examples, compare lane H with I, and K with O) suggesting that portions of the concatemer can be lost by recombination. The large over digestions used, the fact that some of the junction segments observed are smaller than 2.5 kb, and the reproducibility of the patterns, indicate that the faint bands are not the result of partial digestion.

Recircularization precedes concatemer formation. As a byproduct of this work it was possible to examine the structures of the concatemers formed from the mixture of the four injected *mys* clones, and thereby to learn something of the process of concatemer formation itself. Several other studies have demonstrated the efficient occurrence of end-to-end ligation and homologous recombination in DNA transfected (20-23) or microinjected (20,24) into tissue culture cells. The extent of homology required for recombination (25), the ability of cells to ligate similar and dissimilar ends equally well (26,27), and the mutations that sometimes occur

in the process (26,28) have also been studied. Furthermore, Palmiter et al. (29) have clearly demonstrated that homologous recombination of injected DNA can also occur in mouse zygotes.

Southern blot analysis of the mys concatemer structures indicates that most injected linear molecules re-circularized, by intra-molecular end-to-end ligation, as the first step in concatemer formation. That is, the ends of a particular injected molecule are found within the concatemer joined to each other and not randomly ligated to any other available end among the injected linears. Each Sal I cut end of the four injected plasmids is in principle capable of ligating to any of the other seven types of ends or to another copy of itself. Analysis of the arrangement of restriction sites within the concatemer, however, indicates that almost all ligation events involve the two ends of a single injected molecule. Figure 7 shows the Southern blot patterns obtained following digestion of genomic DNA from the 4-band and 13-band third generation transgenic mice with Eco RI or Pst I, and hybridization to a single copy Pvu II segment probe flanking the 3' side of mys 1 (Figure 3). The size of the predominant band of hybridization observed with each enzyme is the same as that seen with the control digested mys 1 supercoiled plasmid DNA. The same result was obtained in a similar experiment using the mys 4 Hind III- Eco RV flanking probe illustrated in Figure 3 (data not shown).

Recircularization is also indicated by the data of Figure 4. The 1.3 Kb segments that hybridize in the control lanes, which have Sal I linearized plasmid DNAs (as injected into zygotes), are absent in the lanes with digested transgenic mouse DNA. Recircularization of the mys 7 DNA converted this 1.3 Kb end segment into a 2.7 Kb Pvu II segment, which is present in transgenic DNA lanes and absent in controls.

DISCUSSION

The mys transposable elements of mice are of particular interest because of their phylogenetic distribution. They are found in the white-footed mouse, P. leucopus, but absent in the house mouse, M. musculus, even though most repetitive sequences are shared by these two species (1). This suggests that perhaps these elements have been acquired relatively recently

Figure 6. Pedigree and inheritance patterns of transgenic mice from lines 9-5 and 9-6. 10 µg of Eco RV digested tail DNA was electrophoresed in each lane, Southern blotted, and then hybridized to the single-stranded 1.2 kb Eco RI-Hind III probe from mys 1 (subcloned into M13). Thick arrows indicate the expected 2.5 kb fragment liberated from within each intact mys element. Thin arrows give the sizes for three of the many single-copy 'satellite' fragments observed in the progeny of mouse 9-5, but not in the 9-6 line. Thin lines between second generation mice lanes A, B, and C, and third generation mice lanes H and I connect fragments of identical size. Mouse B contains both the 13-band concatemer (seen in mouse A) and a 3-band form (related to the 4-band concatemer in mouse C). The different fragment patterns can be clearly distinguished in various third generation mice. Lanes H and M have the 13-band concatemer, lane K the 3-band, lane O the 4-band, lane L the 13-band plus 3-band, and lane I contains just two satellite fragments either deleted, or recombined and segregated away from the 13-band concatemer. The quantities in parentheses below each lane are the number of mice which inherited that particular pattern of fragments. Mouse 9-5 is mosaic, with only a small fraction of cells carrying the transgene concatemer, resulting in a fainter hybridization signal than for progeny transgenics.

Figure 7. Southern blots demonstrating recircularization of linear *mys 1* clones during concatemer formation in microinjected zygotes. The filter was hybridized to the single-copy 0.7 kb *Pvu II* probe from *mys 1*. Lanes A and C are 10 μ g aliquots of tail DNA from two different third generation (4-band) transgenic mice digested with *Eco RI* or *Pst I*, as indicated. Lanes B₁ and B₂ are DNA from two 13-band transgenic mice that were similarly digested. Lane D is 10 μ g of *M. musculus* (control) DNA digested as shown, revealing the endogenous diploid sequence recognized by the *P. leucopus* probe (despite the evolutionary distance separating these two species). Lanes E-H are the same as D but with increasing amounts of digested supercoiled *mys 1* plasmid DNA added. Lane E contains the equivalent of one *mys* element per cell. Lanes F, G, and H contain four, ten, and sixty copy equivalents per cell. Large arrows mark the 5.7 kb and 4.0 kb fragments produced by *Eco RI* and *Pst I*, respectively. These predominant hybridization bands in the transgenic mice lanes correspond exactly to the fragments produced by similarly digested *mys 1* plasmids, indicating that most of the injected linear sequences had recircularized prior to their incorporation into a concatemer. The 4.8 kb fragment in the B₂ lane digested with *Eco RI*, is the correct size for a *mys 1*, *mys 2* intermolecular ligation, suggesting that these events can occur although far less frequently than recircularizations. The corresponding fragment in the *Pst I* digested B₁ and B₂ lanes should be 4.0 kb (see Fig. 3) and is therefore masked by the intense recircularization fragment. Most of the other single-copy fragments in the B₁ and B₂ lanes (some of which have their sizes labelled) are satellite or junction fragments similar to those seen in Fig. 6. The few exceptions are faint fragments seen in both the B and H lanes which are due to contamination of the gel-purified probe with other fragments from the *mys 1* clone. This Southern blot also reveals that approximately 10 of the 50 elements in the 4-band mice (lanes A and C) are *mys 1*, and that the 13-band mice (lanes B₁ and B₂) have around 60 *mys 1* sequences in their 250 copy concatemer.

in evolutionary time, since the divergence of these two species. It also raises several questions. Are the mys elements still active? Are they transcribed? Can they still transpose? Are they still expanding in number within the genome or are they now "under control"? To begin to address these questions we have studied the transcription patterns of mys elements in both P. leucopus and in transgenic M. musculus. Moreover the elements have been tested for transpositional activity in transgenic mice and possible copy control mechanisms have been investigated.

The detection of mys hybridizing transcripts in all organs tested in P. leucopus suggested that the elements are indeed transcriptionally active. Nevertheless the heterogeneity of the transcript sizes, although quite common among other retrotransposon families (30), was somewhat unsatisfying and left open the possibility that the detected transcripts were all merely read-through products driven by promoters upstream of the mys elements.

The transgenic M. musculus, carrying only four different mys elements within a limited number of genomic contexts displayed a much simplified transcription pattern. In these mice discrete transcript sizes were observed and in particular a testes 3.1 kb polyadenylated RNA predominated. Within experimental error this transcript is equal in size to an abundant species observed within the heterogeneous transcription pattern of P. leucopus. It also approximately matches the predicted size of a polyadenylated full-length mys transcript. These results suggest that at least some mys elements, in some tissues, are capable of retrovirus-like transcription. It is also interesting to note that transcription of functional mys elements in germ cells (testes) is all that would be needed for a retrotransposon to spread throughout a population and/or to increase in copy number within the genome.

The absence of detectable mys transposition in transgenic mice was not surprising. Although retrotransposons have been shown to move rapidly in tissue culture cells (17,31), detectable transposition in higher organisms is rare. Nevertheless, certain models for copy number control, such as a limited availability of target sites, or a diffusible repressor, predict a burst of transposition when active elements are first introduced into a new, unrestricted genome (32). The observed absence of any mys transposition events suggests that either these control mechanisms are not at work, or the elements injected (and their recombinant products) are simply incapable of movement at a detectable rate.

The mys elements were, however, found to exhibit an unusual degree of preference for a particular target sequence. Comparison of DNAs flanking five mys elements in P. leucopus revealed an 11 bp conserved sequence 3' of the elements. This is in marked contrast to most transposable elements which show little if any detectable target sequence similarities, although the P elements of Drosophila (33) and the alu elements in primates (34), as well as the retrotransposons IAP (16) and 297 (33) do show some preferences. While the mys element specificity is interesting, it remains likely that there are far more target sites in the P. leucopus genome than there are mys elements. A lack of insertion targets is not yet, therefore, limiting the expansion of the mys family within the genome.

In the process of performing these experiments several observations were made concerning the formation, structure and instability of foreign DNA concatemers in transgenic mice. Recircularization of injected linear molecules was found to frequently precede concatemer formation. Furthermore, when four distinct yet homologous linear molecules are injected into zygotes, all four can be incorporated into a single concatemer, in approximately equal ratios. And the large concatemers studied in this report, with multiple interrupting islands of genomic DNA, are rather

unstable, allowing several independent partial deletion events to be observed.

ACKNOWLEDGEMENTS

We thank Frank Costantini for helping us learn how to make transgenic mice. James Deddams provided valuable expertise in helping us with probability calculations. Carol Lynch helped us establish a mouse colony.

This work was supported by Public Health Service grant CA 0140 from the National Cancer Institute and GM 36266 from NIH.

⁺Address for correspondence and reprint requests: Division of Basic Science Research, Children's Hospital Research Foundation, Elland and Bethesda Aves, Cincinnati, OH 45229, USA

*Current address: University of California, San Francisco; School of Medicine; Office of Student Affairs, S-245; San Francisco, CA 94143-0454, USA

REFERENCES

1. Wichman, H.A., Potter, S.S. and Pine, D.S. (1985) *Nature* **317**, 77-81.
2. O'Hare, K. (1985) *Trends in Genetics* **1**, 250-254.
3. Collins, M., and Rubin, G.M. (1983) *Nature* **303**, 259-260.
4. Young, M.W. (1979) *Proc. Natl. Acad. Sci. USA* **76**, 6274-6278.
5. Maniatis, T., Fritsch, E.F. and Sambrook, J. (1982.) *Molecular Cloning: A Laboratory Manual*, Cold Spring Harbor Laboratory, Cold Spring Harbor, New York.
6. Holmes, D.S. and Quigley, M. (1981) *Anal. Biochem.* **114**, 193-197.
7. Potter, S.S. (1982) *Mol. Gen. Genet.* **188**, 107-110.
8. Maxam, A.M., and Gilbert, W. (1980) *Methods Enzymol.* **65**, 499-560.
9. Duncan, C.H. (1985) *New England Nuclear Product News* **4** (#3), 6-7.
10. Hannahan, D. (1986) In Hogan, B., Costantini, F. and Lacy, E. (eds), *Manipulating the Mouse Embryo: A Laboratory Manual*, Cold Spring Harbor Laboratory, New York, pp. 174-175.
11. Labarca, C., and Paigen, K. (1980) *Anal. Biochem.* **102**, 344-352.
12. Chomczynski, P., and Qasba, P.K. (1984) *Biochem. Biophys. Res. Commun.* **122**, 340-344.
13. Lehrach, H., Diamond, D., Wozney, J.M. and Boedtker, H. (1977) *Biochemistry* **16**, 4743.
14. Feinberg, A.P., and Vogelstein, B. (1983) *Anal. Biochem.* **132**, 6-13.
15. Hu, N.-T., and Messing, J. (1982) *Gene* **17**, 271-227.
16. Hawley, R.G., Shulman, M.J. and Hozumi, N. (1984) *Mol. Cell. Biol.* **4**, 2565-2572.
17. Kuff, E.L., Feenstra, A., Lueders, K., Rechavi, G., Givol, D. and Canaani, E. (1983) *Nature* **302**, 547-548.
18. Kuff, E.L., Feenstra, A., Lueders, K., Smith, L., Hawley, R., Hozumi, N. and Schulman, M. (1983) *Proc. Natl. Acad. Sci. USA* **80**, 1992-1996.
19. Covarrubias, L., Nishida, Y. and Mintz, B. (1986) *Proc. Natl. Acad. Sci.* **83**, 6020-6024.
20. Kopchick, J.J. and Stacey, D.W. (1984) *Mol. Cell. Biol.* **4**, 240-246.
21. Kucherlapati, R.S., Eves, E.M., Song, K.-Y., Morse, B.S., and Smithies, O. (1984) *Proc. Natl. Acad. Sci. USA* **81**, 3153-3157.
22. Miller, C.K. and Temin, H. (1983) *Science* **220**, 606-609.
23. Roth, D.B. and Wilson, J.H. (1985) *Proc. Natl. Acad. Sci.* **82**, 3355-3359.

24. Folger, K.R., Wong, E.A., Wahl, G. and Capecchi, M.R. (1982) *Mol. Cell. Biol.* **2**, 1372-1387.
25. Rubnitz, J., and Subramani, S. (1984) *Mol. Cell. Biol.* **4**, 2253-2258.
26. Wake, C.T., Gudewicz, T., Porter, T., White, A. and Wilson, J.H. (1984) *Mol. Cell. Biol.* **4**, 387-398.
27. Wilson, J.H., Berget, P.B. and Pipas, J.M. (1982) *Mol. Cell. Biol.* **2**, 1258-1269.
28. Lebkowski, J.S., Dubridge, R.B., Antell, E.A., Greisen, K.S. and Calos, M.P. (1984) *Mol. Cell. Biol.* **4**, 1951-1960.
29. Palmiter, R.D., Hammer, R.E. and Brinster, R.L. (1985) In Constantini, F. and Jaenisch, R. (eds), *Genetic Manipulation of the Early Mammalian Embryo*, Banbury Report #20, Cold Spring Harbor Laboratory, Cold Spring Harbor, New York, pp. 123-132.
30. Parkhurst, S.M. and Corces, V.G. (1987) *EMBO J.* **6**, 419-424.
31. Potter, S.S., Brorein, W.J., Dunsmuir, P. and Rubin, G.M. (1979) *Cell* **17**, 415-427.
32. Rubin, G., Kidwell, M.G. and Bingham, P.A. (1982) *Cell* **29**, 987-994.
33. Rubin, G.M. (1983) In Shapiro, J.A. (ed), *Mobile Genetic Elements*, Academic Press, New York, pp. 329-361.
34. Daniels, G.R., and Deininger, P.L. (1985) *Nucleic Acids Research* **24**, 8939-8954.