2976–2981 Nucleic Acids Research, 1994, Vol. 22, No. 15

Tandemly repeated transgenes of the human minisatellite MS32 (D1S8), with novel mouse gamma satellite integration

Maxine J.Allen*, Alec J.Jeffreys, M.Azim Surani[1], Sheila Barton[1], Michael L.Norris[2] and Andrew Collick

Department of Genetics, University of Leicester, Leicester LE1 7RH, [1]Wellcome CRC Institute of Developmental Biology, Cambridge CB2 1QR and [2]AFRC Institute, Babraham, Cambridge CB2 4AT, UK

Received May 16, 1994; Revised and Accepted June 22, 1994 EMBL accession nos X78690–X78702 (incl.)

ABSTRACT

The human hypervariable minisatellite MS32 has a well characterised internal repeat unit array and high mutation rates have been observed at this locus. Analysis of MS32 mutants has shown that male germline mutations are polarised to one end of the array and frequently involve complex gene conversion-like events, suggesting that tandem repeat instability may be modulated by *cis*-acting sequences flanking the locus. In order to investigate the processes affecting MS32 mutation rate and mechanism, we have created transgenic mice harbouring an MS32 allele. Here we describe the organisation of eight transgenic insertions. Analysis of these transgenic loci by MVR-PCR and structural analysis of the junctions between mouse flanking DNA and the transgenic loci has shed light on mechanisms of integration and rearrangement of the tandem repeated transgenes. Sequence analysis of the mouse DNA flanking these transgenes has shown that 5 of the 8 insertions have integrated into mouse gamma satellite repeated sequence. This suggets a non-random integration of the MS32 transgene construct into the mouse genome.

INTRODUCTION

Tandem repeated sequences are common in higher eukaryotic genomes. These tandem arrays are often dispersed throughout the genome and cover a wide range of size classes from microsatellite loci less than a hundred base pairs long to megabases of satellite DNA. Minisatellite repeat arrays fall into an intermediate size group with a total length typically between 0.5–30 kb. Many minisatellite loci identified to date in humans show a GC-rich sequence with some similarity to the *chi* recombination sequence of *Escherichia coli*. This information combined with their clustering at the ends of chromosomes (1) and the existence of minisatellite-like sequences at or near chiasmata (2) suggests that these loci may be involved in meiotic recombination processes. Some minisatellites are very polymorphic, with variation both in allelic repeat unit number and in internal sequence variation of the repeated sequence blocks. Their high variability makes these loci important as markers in linkage analysis and for individual identification techniques (3).

High mutation rates altering allelic repeat unit number at minisatellite loci are responsible for the high levels of allele length polymorphism observed. *De novo* mutation can be detected both in pedigrees (4) and directly in germline (sperm) DNA by single molecule PCR amplification approaches (5, 6). The mutation processes can be further investigated by mapping the interspersion pattern of minisatellite variant repeat (MVR) units along an allele using MVR-PCR (7). Comparison of internal allele structures before and after mutation allows the structural basis of mutation to be defined (7). The best studied human minisatellite is MS32 (locus D1S8) which is comprised of a number of different sequence variants of a 29 bp repeat unit. It has an allele length heterozygosity of 97.5% (8) and a germline mutation rate of 1% per gamete (4). Mutation analysis at this locus has shown that mutation events occuring in the male germline are more common than, and structurally distinct from, those seen in the soma (6). Most sperm mutation events are polar, involving a gain of repeat units at one end of the tandem repeat array. These mutations frequently involve complex gene conversion-like events involving the transfer of repeat units between alleles without the exchange of flanking DNA markers. It is likely, though not proven, that these mutations occur during meiosis. Polar gene conversion appears to be a common phenomenon at human minisatellites (6, 9, 10) and suggests that tandem repeat instability may be modulated by DNA sequences flanking the minisatellite. The simplest model (6) involves a mutation initiator element in the flanking DNA responsible for activating an allele for mutation/conversion, for example by directing a double-strand break into the beginning of the tandem repeat array. The biological significance of such a hypothetical mutation regulator

*To whom correspondence should be addressed

is unknown, though possibilities include the promotion of homologue recognition at synapsis and the initiation of meiotic recombination.

As a first stage in dissecting the complex mutation pathways at minisatellite MS32, we have created transgenic (or more correctly transminisatellitic) mice harbouring MS32, by microinjection of an MS32 allele into mouse 1-cell embryos. The aims were two-fold: first, to study the insertional behaviour of a repeated DNA transgene, and second, to ascertain if the mutation processes at any of the transgenic loci generated reflect those of the endogenous human locus. In this paper we describe the structures of eight transminisatellitic integration events and analyse the complex rearrangements which can occur during integration. Importantly, we show that five out of eight transgenes have integrated into mouse gamma satellite DNA sequences, providing further information relating to the processes by which microinjected transgenes integrate into the mouse genome.

MATERIALS AND METHODS

Construction of transgenic DNA

An MS32 allele was amplified by PCR from 150 ng genomic DNA using primers E1-A and E2-D (see Fig 1) for 15 cycles of 96°C 1.3 min, 66°C 1 min, 70°C 10 min. Primer sequences were E1-A 5′ CCCAAGCTTGGCCTAATGGCcggtgaattccac-agacact 3′ and E2-D 5′ TGCGCGGAATTCTGCAGTCGactc-gcagatggagcaatgg 3′. These primers consist of a 20 nucleotide synthetic sequence (upper case) followed by 20 nucleotides of sequence complementary to MS32 flanking DNA (lower case). PCR conditions were as described previously (7). A 1/50 dilution of the product was re-amplified with primers E1 and E2 for 16 cycles under the same conditions until the product was detectable by agarose gel electrophoresis and ethidium bromide staining. 500 ng PCR product was electroeluted onto dialysis membrane. The DNA was purified on an NACS column (11).

Microinjection

Approximately 200 copies of the MS32 DNA in linear construct form were microinjected into 1-cell embryos of C57BL/6J/CBA/ca F_2 genotype. Microinjection was by standard techniques. Embryos surviving to the 2-cell stage were transferred into recipient female (CFLP) mice. All transgenic production techniques were as described previously (11).

Tail DNA analysis

Tail DNA was prepared by conventional methods of proteinase K digestion and phenol/chloroform extraction (12). Transgene-positive mice were identified by Southern blot analysis of DNA digested with *Pst*I and hybridised with an MS32 repeat unit probe. Southern blot and hybridisation conditions were as described previously (7).

MVR-PCR mapping

Four-state MVR-PCR reactions (13), which analyse two variant sites in MS32 repeats to enable four repeat unit types to be distinguished, were performed on 100 ng aliquots of mouse tail DNA.

Isolation of the transgene flanking sequences by vectorette PCR

Mouse DNA sequences flanking the transgene integration sites were recovered from total genomic DNA using vectorette PCR (14, 15, M.J.Allen, A.Collick and A.J.Jeffreys, manuscript in preparation). PCR conditions were as described previously (7) with a 'hot start' (80°C hold before addition of the primers to the reaction) for the initial amplification from the vectorette linker library. The sequence of vectorette PCR products too long to be fully sequenced by PCR sequencing methods was obtained by generating sub-vectorette products. This approach utilised restriction sites within the length of the DNA fragment as new targets for different vectorette linkers from those used to produce the initial libraries. The series of shorter DNA fragments produced could be sequenced completely.

Vectorette and sub-vectorette DNA fragments containing the transgene flanking DNA were amplified to levels detectable by ethidium bromide staining using nested primers where possible. After agarose gel electrophoresis these DNA fragments were purified by electroelution followed by ethanol precipitation. DNA was dissolved in water and used in double-stranded sequencing reactions. Sequencing was by the method of Winship (16) or by Taq cycle sequencing (17). Some fagments proved refractory to regular double-stranded sequencing methods and were subcloned into the plasmid vector pBluescript SK^+ or KS^- (Stratagene, USA) using standard recombinant DNA techniques. Single-stranded template was produced from the pBluescript clones and sequencing performed by the Sequenase protocol (Pharmacia). These sequence data have EMBL database accession numbers X78690–X78702.

RESULTS

Transminisatellite founder mice and transmission to F_1 mice

An MS32 allele containing 71 repeat units was selected for transgenesis based upon an informative MVR code suitable for detecting subsequent rearrangements. The allele was PCR amplified from human genomic DNA using primers flanking the minisatellite and carrying 5′ extensions (E1 and E2) of synthetic oligonucleotides containing diagnostic restriction sites for subsequent transgene analysis. PCR products containing the repeat array, plus 212 bp of DNA flanking the 5′ mutation hotspot of MS32 and 214 bp of 3′ flanking DNA (Fig.1a), were microinjected into 554 1-cell embryos. 375 embryos surviving to the 2-cell stage were transferred into 11 recipient female mice. Subsequently 13 progeny survived to weaning, when tail DNA was prepared. Tail DNA from these mice was restricted with *Pst*I, an enzyme that cleaves in the E2 linker at the 5′ end of the construct. Southern blot analysis of these digests probed with a fragment of the MS32 locus containing only repeat unit sequence revealed four of the mice to be carrying transminisatellite DNA by the presence of one or more bands hybridising to this probe. Three of the four mice gave multiple hybridising DNA fragments indicating that they were either multi-copy insertions or the result of more than one separate insertion into the same embryo. Mouse 102 only showed one hybridising band, suggesting at first that it contained only one single-copy insertion.

By breeding these mice with non-transgenic C57BL/6J/CBA/ca F_2 mice, we were able to show transmission of the transminisatellites and to establish lines containing a single transminisatellite locus (data not shown). From founder male 110 three separate loci segregated to F_1 mice: 110A (multi-copy/tandem insertion), 110C and 110D (single-copy insertion). Founder male 102 gave two single-copy inserts, 102A and 102B,

whilst founder male 109 gave progeny with only one multi-copy insertion. We have no evidence of linkage between transgenes 110A, C and D. From 36 progeny of the founder mouse 110, A/C, θ = 0.4, z = 0.48, A/D, θ = 0.43, z = 0.23, C/D, θ = 0.43, z = 0.22 or between 102A and B, θ = 0.41, z =0.2, for 27 progeny. Founder 105, a female, died before breeding, preventing the demonstration of independent segregation of the two transminisatellites carried by this animal: 105A (multi-copy) and 105B (single-copy). A total of eight insertions of the MS32 locus were thus created in four of the thirteen progeny. This is a good recovery of integration events when compared to other microinjection experiments (M.L.Norris, personal communication). This may be a reflection of reduced deleterious effects of a non-coding integrant on transgenic progeny, or due to a relatively extensive analysis of the founder mice and their progeny.

Physical mapping of transminisatellites

The restriction maps of the eight transminisatellites are shown in Fig. 1b and c. Initially no MS32 *Pst*I fragments were detected for the single-copy transgenes 102B and 110D, suggesting a lack of neighbouring *Pst*I sites. Further analysis failed to detect the expected frequency of sites for most other restriction enzymes tested, both around these two transgenes, and around the single-copy transgene 102A and the multi-copy transgenes 105A and 109. The latter two transgenes can however be detected in a *Pst*I digest by the release of internal transgene blocks from the multicopy integrant. All five of these transgenes did show sites for the enzymes *Ava*II and *Mnl*I close to the flanking ends of each transgene. Furthermore, partial digestion of *Pst*I digests of 102A with *Mnl*I showed the presence of a restriction site every ~230 bp over most of the 3′ flanking DNA. *Ava*II gave a similar pattern though it did not cleave every ~230 bp. These data suggested that these five transgenes had integrated into mouse gamma satellite DNA which consists of tandem arrays of a 234 bp repeating unit, most of which contain one site for *Mnl*I and often for *Ava*II (18). In contrast, the single-copy transgenes 105B and 110C and the multi-copy transgene 110A showed a normal frequency of restriction sites in mouse flanking DNA. The single-copy transgene 102A shows loss of the 3′ *Eco*RI and *Hin*dIII sites diagnostic of the synthetic sequence E2, suggesting that the integrant was truncated at the 3′ end. Similar analysis showed that 102B was 5′ truncated. 105B and 110D appeared intact by analysis of E1 and E2 diagnostic sites, but were shorter than predicted from the original construct length, suggesting loss of MS32 repeat units. 110C consisted of two adjacent MS32 repeat unit blocks separated by *Hin*dIII, *Pst*I and *Eco*RI sites, consistent with a rearranged integrant.

The multi-copy transgenes 109 and 110A contain 6 and 11 blocks of repeats respectively, some of which show size changes relative to the original construct. A detailed analysis of the structure of these complex transgenes will be presented elsewhere.

MVR mapping

The internal structures of the minisatellite arrays of the five single-copy transgenes were analysed using four-state MVR-PCR (13) which assays two polymorphic sites within MS32 repeat units to enable the interspersion patterns of the resulting four types of repeat unit (**E**, **e**,**Y**, **y**) to be determined. MVR-PCR was performed using the flanking primer MS32-O (Fig. 1a) 5′ to the array. With those 5′ truncated transgenes which lacked MS32-O, reverse four-state MVR-PCR was performed using the primer

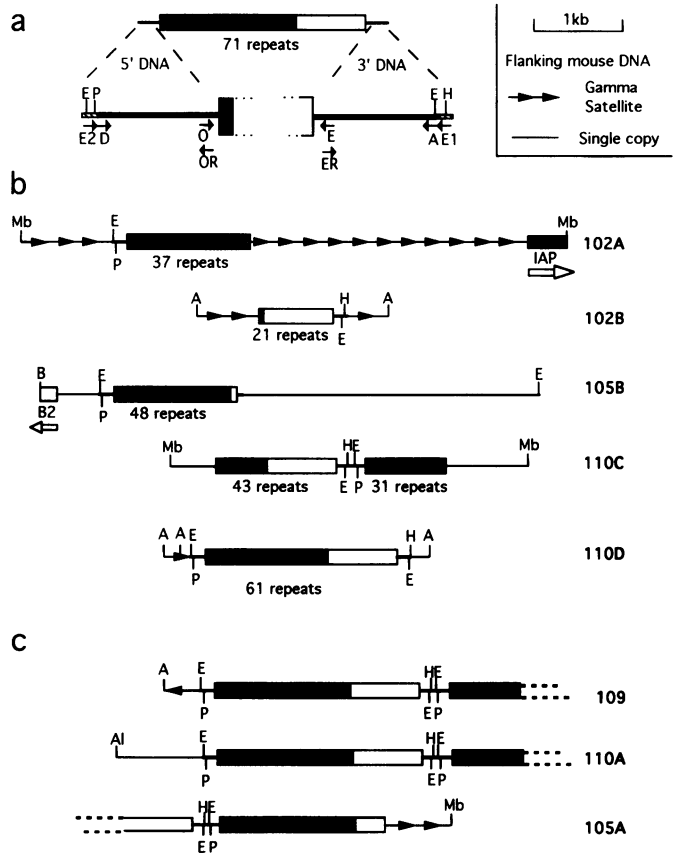

Figure 1. Structure of the MS32 minisatellite integrants in transgenic mice. (**a**) The structure of the original construct for microinjection. 212 bp of 5′ and 214 bp of 3′ MS32 flanking DNA are included in the construct. Oligonucleotide primer sites for the MS32 primers 32A, O, OR, E, ER, and D are depicted by arrows. The sequences of these primer sites are detailed elsewhere (7). The initial amplification of the construct used primers A and D with attached synthetic linkers E1 and E2. Subsequently the E1 and E2 priming sites can be used separately from A and D. The terminal primers E1 and E2 have the following sequences, E1 = cccaagcttggcctaatggc, E2 = tgcgcggaattctgcagtcg. There are 71, 29 bp repeat units, and the total length of the construct is 2525 bp. Restriction sites found in the MS32 flanking DNA and synthetic linkers are shown, *Pst*I (P), *Hin*dIII (H), *Eco*RI (E). (**b**) Structures of the five single copy transgenes. The blocks of shading represent the regions of repeat units of the original construct maintained in each insertion. Restriction sites in the flanking DNA used to make the initial vectorette products are shown 5′ and 3′ of the insert. Al = *Alu*I, A = *Ava*II, B = *Bam*HI, Mb = *Mbo*I, S = *Ssp*I. The initial vectorette products were made with blunt (e.g. *Alu*I) or sticky ended (e.g. *Mbo*I) vectorette linkers or vectorette linkers with a degenerate *Ava*II overhang. The flanking mouse DNA is either single-copy (line) or mouse gamma satellite (arrows). Unshaded arrows represent the orientation of the B2 and IAP element homologies. (**c**) A partial structure of the multi-copy transgenes is shown. The full structure of these complex insertions will be discussed elsewhere.

MS32-E in the 3′ flanking human DNA. The most 5′ minisatellite array of the multi-copy transgene 110A was also MVR mapped. MVR codes of the progenitor allele and each mapped transgene are shown in Fig. 2.

The first minisatellite array in the multicopy transgene 110A carries an MVR code identical to the input MS32 allele. In contrast all five single-copy transgenes show different MVR codes, though all clearly align with the progenitor allele. In some cases null or o-type repeat units were seen. These are repeat units that do not amplify with the **E**-, **e**-, **Y**-, or **y**-type repeat specific

primers (19). There are no null repeats in the original allele and their appearance in transgenes 110C, 105B and 110D may be the result of Taq polymerase errors during PCR amplification of the original allele prior to injection (20). Transgene 110D contains full 5′ and 3′ flanking human DNA, but has lost 10 repeat units, 9 at the beginning of the allele and one in the centre of the array. Likewise 105B has lost 23 repeats from the 3′ end of the array and has undergone a complex rearrangement over repeat units 20−24 where a segment of code reading **yEyeE** in the original allele has changed to **eeyye** in the transgene. As predicted by restriction mapping, transgene 102A has lost the 3′ flanking human DNA together with the terminal 34 repeat units of the array. Transgene 102B is the apparent complement of 102A with no 5′ flanking human DNA and the loss of the first 50 repeat units. Transgene 110C results from a circular permutation of the input allele and consists of the last 43 repeats followed by 3′ flanking human DNA then 5′ flanking DNA followed by the first 31 repeats of the original array (Fig.1b). The two MVR maps combine to give a full original construct plus a duplication of three repeat units (**Yey**, positions 28−31) at the point of integration of the MS32 allele.

Flanking DNA sequence analysis

Vectorette PCR (14, 15) was used to recover mouse DNA flanking the 5′ and 3′ ends of each transgene. Vectorette libraries were constructed from transgenic mouse DNA digested with *Mbo*I, *Bam*HI, *Bgl*II or *Ava*II. 5′ junctions were isolated using the vectorette primer and primer MS32-OR (Fig. 1a.); for 5′ truncated transgenes the 5′ junction was amplified using primer MS32-E to yield products extending from the 5′ flanking mouse DNA across the MS32 repeat array and into 3′ flanking human DNA. 3′ junctions were likewise recovered using primer MS32-ER or, if necessary, MS32-O. All junctions were successfully recovered except for the 3′ end of the multicopy transgenes 110A and 109 and the 5′ end of the multicopy transgene 105A. Junction fragments were sequenced, if necessary after reducing the size of the original PCR product by sub-vectoretting (see Materials and Methods), and all sequences scanned against the EMBL database. The sequence organisation around the transgenes is summarised in Fig. 1b and c. Where 5′ and 3′ human flanking sequences are shown they are complete, with the exception of transgene 109 which has lost 33 bp of 5′ flanking DNA at the mouse human junction and 105B which has only the first 27 bp of 3′ flanking human DNA sequence. Sequencing confirmed that transgenes 102A, 102B, 105A, 109 and 110D had integrated into mouse gamma satellite DNA. The 3′ and 5′ truncation of 102A and 102B respectively was confirmed; in both cases the junction consisted of MS32 repeats joined directly to gamma satellite DNA. The gamma satellite DNA 3′ to 102A switches after 3.45 kb to a region showing 92% sequence identity to an LTR sequence from a mouse IAP element (21). Restriction sites in this LTR are present in 102A genomic DNA, indicating that this junction fragment is authentic and not a vectorette PCR artefact. Mouse DNA sequences around 105B, 110C and 110A show no similarity with any known mouse sequence, except for an 80.5% sequence identity with the 3′ end of a mouse B2 element (22), upstream of 105B. Sequence analysis of 110C also confirmed the circularly permuted structure of this transgene, which consists of MS32 repeats followed by a complete 3′ flanking human DNA region joined to a complete 5′ flanking human region followed by MS32 repeats.

Short regions of homology at transgene/gamma satellite junctions

Figure 3a shows the sequence of the 3′ flanking mouse gamma satellite DNA adjacent to two of the transminisatellite loci, 102A and 105A. Both gamma satellite integrations show 4−5 nucleotides of sequence shared between the minisatellite sequence and the gamma satellite sequence at the integration site. Small regions of homology have been observed in previous studies of non-homologous recombination involving transgenes (23, 24). The position of insertion into the 234 bp gamma satellite is different in both cases and in each of the other sequenced gamma satellite/transgene junctions (data not shown). The orientation of the gamma satellite with respect to the 5′ end of the transminisatellite is the same for 4 out of 5 loci, with the exception of the 5′ sequence of the 109 locus.

Deletions of endogenous DNA at the integration site

Deletions of mouse DNA at sites of integration have been observed at other transgenes (25, 26) and we see evidence for this type of structural change around the 110D, 105B and 102A transminisatellite junctions. The 5′ mouse flanking DNA of the 110D transminisatellite shows homology with gamma satellite DNA, but the 3′ flanking DNA does not. This may be the result of a deletion of the region joining the gamma satellite to flanking non-repetitive DNA. A deletion of mouse DNA at the junction site has also been observed with the transminisatellite 105B, where PCR primers in the mouse DNA are able to amplify across the transminisatellite efficiently, but cannot produce a product in mice negative for insertion (data not shown). This would suggest a deletion of mouse DNA too large to be amplified by PCR (8 kb). With the insertions flanked both 5′ and 3′ by mouse gamma satellite DNA there is also evidence of rearrangement during insertion since the gamma satellite repeat structure does not continue in phase after the transgene integration, as shown in Fig. 3b. The mechanism of transgene insertion for 102A into gamma satellite may involve the loss of a number of full gamma satellite 234 bp repeats plus 205 bp of one repeat leaving a 29 bp portion, or it may involve a duplication of 29 bp flanking the site of insertion.

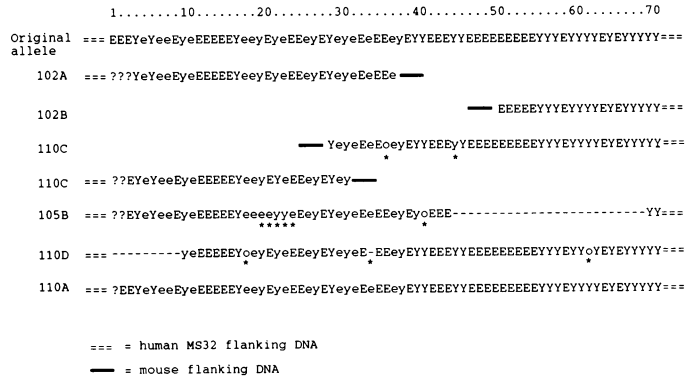

Figure 2. MVR-PCR codes of the MS32 repeat arrays in trangenic mice. Four-state MVR codes (repeat types **E**, **e**, **Y**, **y**) are shown for the progenitor allele, for each single-copy transgene and for the first repeat block in the multi-copy transgene 110A. o = 'null' repeat unit unamplifiable with **E**-, **e**-, **Y**-, or **y**-specific MVR primers; ? = repeat type not known. Differences between the transgene and progenitor are marked *. Deleted repeats are indicated by dashes (−).

DISCUSSION

This work demonstrates for the first time that it is possible to create, by microinjection, mice transgenic for human minisatellites. Single-copy integrants were obtained together with novel multi-copy inserts containing tandem repeats of the MS32 array plus flanking DNA. However, MVR mapping and sequence analysis showed that all five single-copy integrants had undergone rearrangement prior to, during, or post-insertion. It is possible that some of the more subtle rearrangements, for example minor changes in the MVR maps of the transgenes, may have resulted from changes in the construct during the initial PCR amplification prior to microinjection; however, larger structural rearrangements have most likely arisen during or after insertion. Pedigree analysis has shown that the single-copy transgenes 102A, 102B, 110C and 110D are inherited stably, without change in the repeat copy number (e.g. 174 descendants of founder 110 which were positive for the 110D transgene showed no evidence for mutation; data not shown). Abnormal structures in the single-copy transgenes are therefore most likely to have arisen either pre-insertion, where the linear constructs are free to recombine with each other and exchange sequence information, or during the insertion process.

Various types of insertional rearrangement are seen in the single-copy transgenes. For example, 105B has lost all of the 3′ human flanking DNA except for the proximal 27 bp (not shown), together with 26 repeat units from within the MS32 array. In addition, repeats 20–24 have switched MVR code which may be analogous to the products of complex gene conversion events seen in the male germline at the endogenous MS32 locus (6) and may have arisen by mismatch repair at a heteroduplex formed during recombination of two progenitor molecules.

Another type of rearrangement is seen with transgene 110C which consists of a circular permutation of a single original construct containing two blocks of MS32 repeats separated by a tail-to-head fusion of human flanking sequences. This structure suggests that transgene insertion most likely occurred via a circular recombination intermediate containing a single construct (Fig. 4) to disrupt the MS32 repeat block into two halves plus a three repeat duplication at the site of insertion into mouse DNA. It is possible that a linear tandem array of transgene constructs was involved prior to integration, but the fact that the transgene consists of only one full original construct (plus a three repeat duplication) suggests it is more likely that a single circularised intermediate was involved. Previously it has been suggested that circular DNA molecules may be recombinationally active in transgene integration (24).

A further curious rearrangement is seen with transgenes 102A and 102B, which have arisen in the same founder mouse and contain the 5′ and 3′ halves of the original construct. Both loci are flanked by mouse gamma satellite DNA but show no detectable linkage. It is possible that these two transgenes have resulted from a single original full length insert which has been disrupted by a rearrangement such as a translocation together with the loss of the central 13 repeat units. The absense of any trace of a full length construct in the founder mouse or in progeny DNA as shown by PCR (data not shown) suggests that this disruption has occured during integration.

Microinjection of DNA into the mouse pronucleus is a commonly used technique for introducing genetic material into mice, to provide an important system for studying gene expression, cancer and developmental biology (27). Typically several hundred copies of the construct are injected into the male pronucleus and the integrations can result in either single or multi-copy transgenes. Surprisingly, 5 out of 8 of the transminisatellites have integrated into mouse gamma satellite DNA. Since this satellite comprises only 10% of the mouse genome (28), this suggests a non-random integration of these constructs (P = 0.0046). There are a number of possible explanations for this pattern of integration of the MS32 construct into the mouse genome.

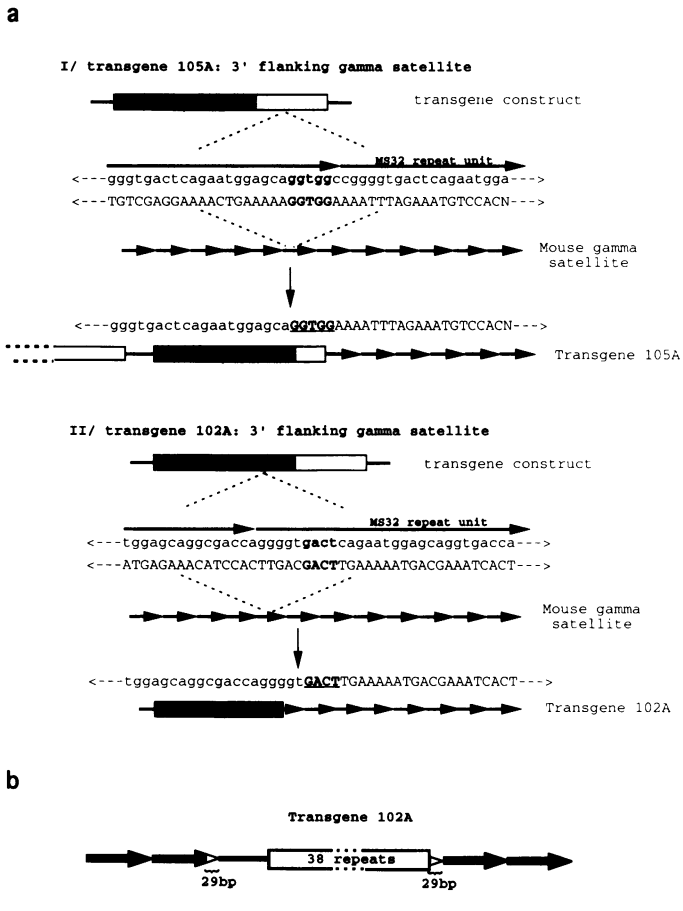

Figure 3. DNA sequences at the junctions between mouse DNA and MS32 transgenes. (**a**) I, II, examples of homologies between MS32 repeat sequence (lower case) and gamma satellite sequence (upper case) at transgene junctions involving fusion of minisatellite and satellite repeats. (**b**) Disruption of gamma satellite repeat phase by transgene insertion. Gamma satellite repeats are indicated by arrows.

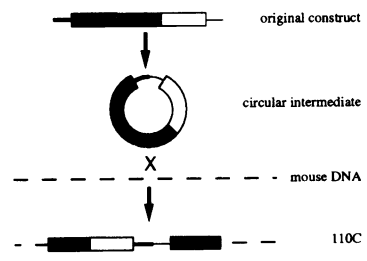

Figure 4. A putative insertion mechanism for the 110C transminisatellite.

In two cases, 4–5 bp regions of homology between the transgene and mouse gamma satellite are seen at the junction breakpoint; since regions of homology as small as 1 bp have previously been suggested as having an effect on recombination leading to integration (23, 24), it is possible that these two integrations reflect sequence-specific targetting. However, this targetting is not site-specific since all five integrations into gamma satellite are at different positions in the 234 bp repeat sequence. An alternative explanation may be that gamma satellite is more likely to receive transgenes than other DNA regions. Perhaps the condensed chromatin structure of the gamma satellite DNA (28) is more likely than other regions of the genome to break under the forces of microinjection. If so, repair processes may be more likely to integrate a transgene into the gamma satellite DNA. A third possibility is that integration into single-copy DNA can result in dominant lethality, but that integration into gamma satellite causes no such deleterious effects.

It is not clear whether preferential integration into gamma satellite is a general phenomenom for all transgenes, since few studies have systematically addressed the nature of mouse DNA flanking sites of integration. Transfection of DNA into mammalian cell lines has yielded integrants flanked by centromeric DNA (29, 30), with centromeric integration occuring at a frequency of 1/4 and 1/3 in these two studies. In some cases integrant expression was modulated by the surrounding centromeric DNA (31). If transgene expression is repressed by centromeric gamma satellite DNA, then screening mice for functional transgenes may overlook many mice containing a normal but inactive transgene. Even if transgene detection is by Southern blot analysis of genomic DNA rather than functional assays, the relative lack of restriction sites in gamma satellite DNA may make some transgenes undetectable, as was initially the case with transminisatellite 102B and 110D. It may therefore be more appropriate to use enzymes such as *Ava*II, which cleave frequently in mouse gamma satellite, to identify satellite integrants.

Mouse lines with transgenes integrated into satellite DNA provide a novel resource for studying the biology of centromeres, though we should stress that the exact genomic locations of the five transminisatellitic integrations flanked by gamma satellite mouse sequence are unknown, and it remains possible that integrations have occurred into ectopic blocks of satellite sequence, rather than into centromeric gamma satellite.

The analysis of somatic and germline instability at these transminisatellitic loci will be presented elsewhere.

ACKNOWLEDGEMENTS

We thank Robert Kelly, John Armour, Nicola Royle, Paul Overbeek, Richard Woychik and Richard Palmiter for helpful discussions. The work of A.J.J. was supported in part by an International Research Scholars Award from the Howard Hughes Medical Institute, and in part by grants from the Medical Research Council and the Royal Society. The work of A.S. was supported by grant number 036481 from the Wellcome Trust.

REFERENCES

1. Royle,N.J., Clarkeson,R.E., Wong,Z., and Jeffreys,A.J. (1987) *Genomics*, **3**, 352–360.
2. Chandley,A.C. and Mitchell,A.R. (1988) *Cytogenetics Cell Genetics*, **48**, 152–155.
3. Jeffreys,A.J., Wilson,V., and Thein,S.L. (1985) *Nature*, **314**, 67–73.
4. Jeffreys,A.J., Royle,N.J., Wilson,V., and Wong, Z. (1988) *Nature*, **332**, 278–281.
5. Jeffreys,A.J., Neumann,R., and Wilson, V. (1990) *Cell*, **60**, 473–485.
6. Jeffreys,A.J., Tamaki,K., MaCleod,A., Monkton,D.M., Neil,D.L., and Armour, J.A.L. (1994) *Nature Genet.*, **6** (2), 136–145.
7. Jeffreys,A.J., MaCleod,A., Tamaki,K., Neil,D.L., and Monkton,D.G. (1991) *Nature*, **354**, 204–209.
8. Wong,Z., Wilson,V., Patel,I., Povey,S., and Jeffreys,A.J. (1987) *Ann. Hum. Genet.*, **51**, 269–288.
9. Neil,D.L. and Jeffreys,A.J. (1993) *Human Molecular Genetics*, **2** (8), 1129–1135.
10. Armour,J.A.L., Harris,P.C. and Jeffreys,A.J. (1993) *Human Molecular Genetics*, **2** (8), 1137–1145.
11. Allen,N.D., Barton,S.C., M. Azim Surani, and Reik, W. (1987) In Monk,M. (ed.), Production of transgenic mice, in Mammalian development; a practical approach. IRL Press, Oxford, pp. 217–233.
12. Kelly,R., Bulfield,G., Collick,A., Gibbs,M., and Jeffreys,A.J. (1989) *Genomics*, **5**, 844–856.
13. Tamaki,K., Monkton,D.G., MaCleod,A., Allen,M.J., and Jeffreys,A.J. (1993) *Human Molecular Genetics*, **2** (10), 1629–1632.
14. Riley,J., Butler,R., Ogilvie,D., Finniear,R., Jenner,D., Powell,S., Anand,R., Smith,J.C. and Markham, A.F. (1990) *Nucleic Acids Res.*, **18**, 2887–2890.
15. Arnold,C. and Hodgson,J. (1991) *PCR Methods Applications*, **1**, 39–42.
16. Winship,P. (1989) *Nucleic Acids Res.*, **17** (3), 1266.
17. Murray,V. (1989) *Nucleic Acids Res.*, **17** (21), 8889.
18. Vissel,B. and Choo,K.H. (1989) *Genomics*, **5**, 407–414.
19. Tamaki,K., Monkton,D.M., MaCleod,A., Neil,D.L., Allen,M.J. and Jeffreys,A.J. (1992) *Human Molecular Genetics*, **1** (6), 401–406.
20. Dunning,A.H., Talmud,P.,and Humphries,S.E. (1988) *Nucleic Acids Res.*, **16** (21), 10393.
21. Mietz,J.A., Grossman,Z., Lueders,K.K. and Kuff,E.L. (1987) *J. Virol.*, **61**, 3020–3029.
22. Krayev,A.S., Markusheva,T.V., Kramerov,D.A., Ryskov,A.P., Skryabin,K.G., Bayev,A.A. and Georgiev, G.P. (1982) *Nucleic Acids Res.*, **10**, 7461–7475.
23. Rohan,R.M., King,D., and Freis,W.I. (1990) *Nucleic Acids Rese.*, **18** (20), 6089–6095.
24. Hamada,T., Sasaki,H., Seki,R. and Sasaki, Y. (1993) *Gene*, **128**, 197–202.
25. Woychik,R.P., Stewart,T.A., Davis,L.G., D'Eustachio,P. and Leder,P. (1985) *Nature*, **318**, 36–40.
26. Kato,S., Anderson,R.A., and Camerini-Otero, R.D. (1986) *Molecular Cellular Biology*, **6** (5), 1787–1795.
27. Palmiter,R.D. and Brinster,R.L. (1986) *Annual Review Genetics*, **20**, 465–499.
28. Hastie,N.D. (1989) In Lyon,M.F. and Searle,A.G.(ed.) Genetic variants and strains of the laboratory mouse. Oxford University Press, pp.
29. Lo,C.W. (1983) *Molecular Cellular Biology*, **3** (10), 1803–1814.
30. Wahl,G.M., Robert de Saint Vincent,B. and Rose, M.L. (1984) *Nature*, **307**, 510–516.
31. Butner,K. and Lo,C.W. (1986) *Molecular Cellular Biology*, **6** (12), 4440–4449.