Direct sequencing of PCR-amplified junction fragments from tandemly repeated transgenes

Richard M.Rohan+, Donna King*§ and William I.Frels
US Department of Agriculture, Agricultural Research Service, Reproduction Laboratory, Beltsville Agricultural Research Center, Beltsville, MD 20705, USA

Received June 4, 1990; Revised and Accepted September 4, 1990

ABSTRACT

When microinjected foreign genes integrate into the genomes of mice, multiple copies are frequently found clustered together at one location. How they concatamerize—by the integration of large linearized concatamers that are formed by simple end-to-end linkage, by circularization of individual DNA fragments and recombination, or by some other means—is not understood. In the transgenic animals studied thus far by ourselves and others, integration frequency and transgene copy number do not seem to be significantly influenced by the complementarity of the ends of the DNA fragments that have been microinjected. We have utilized PCR amplification and DNA sequence analysis to study selected transgene junctions at the nucleotide level. In two transgenic mice carrying the synthetic RSVcat gene (injected with noncomplementary overhangs on the fragment ends), ends were 'nibbled' from 1 to 62 bases before being joined to an adjacent gene copy. Repeated dinucleotides, providing the most minimal of homologies, are present in half of the characterized junctions. Determination of the relative copy number of the junctions in each mouse supports the idea that transgene complexes can undergo additional rearrangements after the initial formation event.

INTRODUCTION

Microinjection of purified gene fragments has become a common laboratory technique for producing transgenic animals (1–3). In mice, the foreign DNA is inserted into the genome in approximately 10 to 20 percent of the animals born (4). While the usefulness of the technique in studying gene expression has been thoroughly demonstrated, very little is actually known about how the foreign DNA becomes associated with the animal's chromosomes. Typically several hundred gene copies are microinjected into one of the pronuclei of a recently fertilized ovum. Sometimes a single copy of the gene becomes integrated into the genome. Often, a concatamer of genes is detected in the transgenic animal; two to over one-hundred foreign gene copies are linked together at the same locus. Occasionally integrations will occur at more than one locus. Rearrangements of host genomic sequences including duplications (5), deletions (6–8), and translocations (9) have been associated with the integration of transgenes.

Models for the concatamerization of identical linear DNA molecules require circularly permuted intermediates (2,4,10). DNA molecules with compatible endonuclease-restricted ends can be joined by simple ligation of the ends. Joining of microinjected linear DNA at compatible ends without DNA sequence alteration occurs in cultured cells (11,12). Linear DNA with non-compatible ends has also been demonstrated to be efficiently circularized in transfected cell cultures (13). The direct ligation of compatible ends has occurred in microinjected embryos as evidenced by the frequent regeneration of restriction sites between the tandemly arrayed transgenes (8, 14–16).

In an effort to understand how gene concatamers are formed in transgenic mice, the DNA sequences of a group of RSVcat transgene junctions were obtained. To eliminate the most trivial cases of end-to-end ligation, the transgenic mice were created by microinjection of a gene fragment which had nonhomologous ends. Polymerase chain reaction (PCR; 17,18) performed with end-labelled oligonucleotides was then used to synthesize copies of each transgene junction for direct sequence analysis.

MATERIALS AND METHODS

Animals

Mouse ova were collected for microinjection 8 to 12 hours after the presumed time of mating (1–3). The RSVcat gene (19), consisting of a Rous Sarcoma Virus promoter fused to the bacterial chloramphenicol acetyltransferase coding region, was microinjected as a 2.7kb fragment at a concentration of 5–10 ng/μl in 10 mM tris, pH 7.4, 0.1 mM EDTA. The gene was purified away from the vector sequences by digestion with PstI and NdeI, generating a 5′ overhang of 2 bases at the 5′ end and a 3′ overhang of 4 bases at the 3′ end. Approximately 500 copies of the gene were injected into each ovum.

* To whom correspondence should be addressed

Present addresses: +Department of Obstetrics and Gynecology, Division of Reproductive Endocrinology, School of Medicine, University of Maryland at Baltimore, 655 W. Baltimore St., Baltimore, MD 21201 and §Department of Pharmacology and Molecular Biology, The Chicago Medical School, 3333 Green Bay Rd., North Chicago, IL 60064-3095, USA

Transgenic mice were identified by dot blot analysis of DNA prepared from tail tissue, and further characterized for gene copy number and integrity by Southern blotting (20).

Polymerase chain reactions

Figure 1 depicts the relative location of the oligonucleotide primers on the injected RSVcat fragment and their sequences. A typical polymerase chain reaction contained 0.016 OD_{260} units of each primer and 200 ng of total genomic DNA from tail tissue. Buffer, deoxynucleotide triphosphates and *Thermus aquaticus* DNA polymerase were from Gene-Amp kits (Perkin-Elmer Cetus, Norwalk CT) and were used according to manufacturer's protocols. Temperature cycles were carried out in a DNA Thermal Cycler (Perkin-Elmer Cetus) using step-cycle settings of 1.5 min at 94°C (denaturation), 1.5 min at 60°C (hybridization), and 3 min at 72°C (elongation). For amplification of longer target sequences (198/200, 274/200, and 273/199 oligonucleotide primer pairs), the elongation step was increased to 6 min.

For DNA sequencing, one member of each pair of oligonucleotides (0.016 OD_{260} units) was end-labelled in 50 mM Tris, pH 7.5, 10 mM $MgCl_2$, 5 mM dithiothreitol, 1 mM spermidine and 35–70 μCi of gamma ^{32}P-ATP (>3000 Ci/mmole, New England Nuclear, Boston MA) with 10 units of T4 polynucleotide kinase (New England BioLabs, Beverly MA). Reactions were incubated at 37°C for 30 min and then 65°C for 15 min to denature the kinase. One-tenth volume of the kinase reaction was added directly to the polymerase chain reaction with one-tenth of the typical amount of unlabelled opposing primer.

Quantitation of transgene junctions

Oligonucleotide primers 199 and 200 (0.016 OD_{260} units each) were labelled with gamma ^{32}P-ATP by polynucleotide kinase. After heat denaturation, the kinase reaction was mixed with 2 μg of total genomic DNA from transgenic mouse M2334 or M2315. The primer/DNA mixture was aliquotted to nine parallel PCR reactions. Individual reactions were amplified for 20, 21, 22, 23, 24, 25, 26, 28 or 30 cycles. Equal aliquots of each were electrophoresed on a 6% polyacrylamide gel. Bands were localized by autoradiography, excised, and quantitated in a scintillation counter. The amount of radioactively labelled primer in each band at cycles 20 to 25 was fit, by linear regression, to the equation: $\log A_N = \log A_0 + N\,[\log(R+1)]$ where N=number of cycles, A_N=the amount of radioactivity (CPM) incorporated into the band after N cycles, A_0=the initial amount of target sequence, and R=the efficiency of the amplification reaction (21). Calculation of the y-intercept gives the value of Log A_0, and calculation of the slope gives the value of log (R+1). A_0 is a relative quantity that is useful for comparison within a particular assay. The value of R should approach 100%.

DNA sequencing

After PCR amplification the unincorporated nucleotide and primers were removed with Centricon-30 columns (Amicon, Danvers MA). DNA was electrophoresed on 6% polyacrylamide gels. Bands were visualized by autoradiography, excised, and eluted by the crush-soak method (22). In some cases purified native fragments were purified further on strand separation gels. Eluted DNA fragments were desalted with NENSORB columns (New England Nuclear). DNA sequencing reactions were carried out as described (23).

RESULTS

The DNA from eight transgenic mice, created by microinjection of the RSVcat fragment depicted in figure 1, was tested by the polymerase chain reaction using various combinations of primers. The 199/200 pair of primers was designed to amplify across transgene junctions. Other pairs of primers were designed to amplify internal regions of the RSVcat fragment. The results are summarized in table 1. Four of the mice (M1863, M1878, M1882, and M1936) had integrated RSVcat sequences as assayed by dot blot hybridization but gave no positive signal with any of the pairs of primers which were tested. Data from Southern blot hybridizations indicated that these mice had partial or rearranged RSVcat sequences at less than one copy per cell (data not shown). Mice M1927 and M2313 each integrated a partial copy of the middle portion of the RSVcat gene. Although these six mice did not appear to have junction sequences, the presence of inverted junctions cannot be detected by PCR (due to kinetically favored fold-back of individual template strands on themselves) and should not be ruled out.

DNA from two of the founders, M2315 and M2334, each produced multiple bands on a gel when amplified with the 199–200 oligo pair (figure 2). The number of transgene copies estimated by dot blot hybridization was 2–5 for M2315 and 25 for M2334. Southern blot hybridization indicated that the transgenes in both of these mice were integrated as tandem (direct) repeats.

The products of PCR amplifications with M2334 DNA were purified from nondenaturing gels. Denaturation and

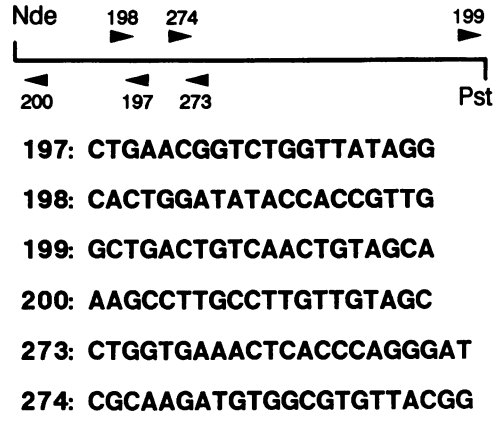


Figure 1. The 2.75 kb NdeI- PstI fragment of pRSVcat (19) is shown. The relative location of the oligonucleotide primers used in this study are denoted by arrows. The sequence of each of the primers is shown below.

Table 1. Amplification of RSVcat sequences with different oligonucleotide primer pairs in eight transgenic mice. + presence of visible products in ethidium stained gels, – no visible products, ND not done.

	197–198	273–274	199–200	198–200	274–200	273–199
M1863	–	–	–	–	–	–
M1878	–	–	–	–	–	–
M1882	–	–	–	–	–	–
M1927	+	+	–	–	–	–
M1936	–	–	–	–	–	–
M2313	+	+	–	–	–	–
M2315	+	ND	+	+	+	+
M2334	+	ND	+	+	+	+

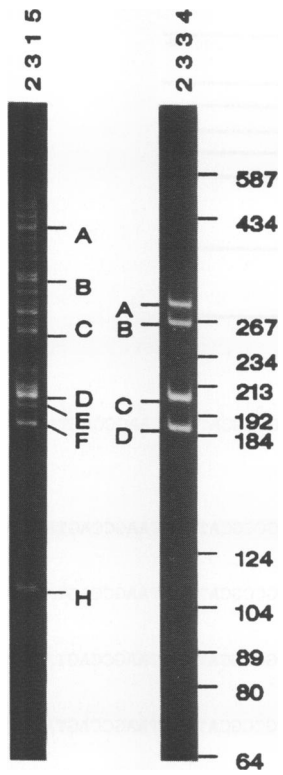


Figure 2. Products of PCR amplification with genomic DNA from transgenics M2315 and M2334. Thirty PCR cycles were performed and one quarter of the reaction volume was electrophoresed on a 6% polyacrylamide gel, which was then stained with ethidium bromide. The positions of DNA size markers are given on the right. The location of DNA fragments discussed in the text are identified by the letters A through H.

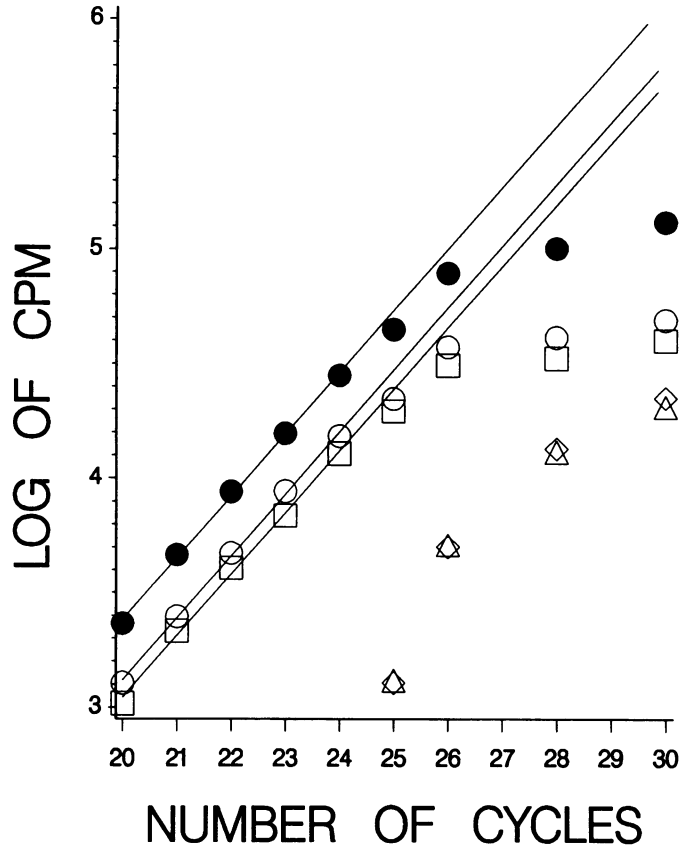


Figure 3. Plot of the logarithm of incorporated ^{32}P in each band versus the number of PCR cycles for analysis of mouse M2334 transgene junctions. Band A (triangle); Band B (diamond); Band C (circle); Band D (square); and total counts per minute in all bands (dot).

electrophoresis of these purified fragments on strand separating gels showed that bands A and B are heteroduplexes composed of single strands from bands C and D.

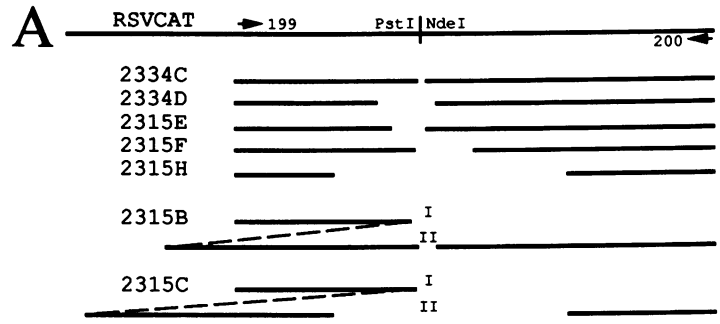
The relative amounts of the PCR amplification targets were determined as described in Materials and Methods and shown in figure 3. Parallel reactions were stopped after various numbers of cycles and the amount of product was quantitated. Fragments C and D from M2334 accumulate logarithmically from cycles 20 to 25. Bands A and B only become evident after 25 cycles, which corresponds to the cycle when the increase in bands C and D is no longer logarithmic. The efficiency of production of both DNA fragments, under these conditions, was 86%. The A_0 ratio for bands C and D is 1.1, indicating that essentially equal amounts of the two different junctions were incorporated in the genome. This mouse was determined to be sterile, since she did not produce any offspring despite repeated breeding attempts.

Similar quantitative analysis of the products of PCR reactions with M2315 DNA showed that bands A through H (figure 2) increased logarithmically from cycles 20 to 25. Band G, which migrated midway between F and H, was visible on autoradiographs but did not accumulate to sufficient quantities to be visible in ethidium stained gels. The other bands seen in figure 2 only became evident after 25 cycles. Some of the products of amplification with M2315 DNA and the 199/200 oligo pair were found to be heteroduplexes of the primary products. Additionally, the structures of some of the primary products (see discussion) allow those templates to produce concatamers of ever-increasing length at later cycles of amplification. The initial amount of template and the efficiency of amplification (A_0 and R) calculated for each of the primary products are given in table 2. The initial amount of the various transgene junctions in the DNA from M2315 varied by approximately 30-fold. She failed to pass any RSVcat sequences to her 21 offspring.

Table 2. Initial amounts of target sequences (A_0) and efficiency of amplification (R) for PCR products with the 199/200 oligonucleotide primer pair and transgenic M2315 genomic DNA. The experiment was carried out as described in figure 3. Sequence data indicates that fragment D represents a mixture of nearly full−length junctions.

Fragment	A_0	R
A	8.3×10^{-5}	0.95
B	8.2×10^{-5}	0.99
C	7.5×10^{-5}	1.00
D	2.3×10^{-3}	0.89
E	9.7×10^{-4}	0.81
F	5.0×10^{-4}	0.92
G	2.2×10^{-4}	0.74
H	8.6×10^{-5}	0.94

The DNA sequence of many of the transgene junctions from M2315 and M2334 was determined directly by purifying the PCR products from reactions in which only one strand was labelled by incorporation of a radioactive primer. PCR products were sequenced in both directions by labelling one primer or the other



B

RSVCAT PstI NdeI
TTGAGCAGGATATTTGGTCCTGTAGTTTGCTAACACACCCTGCA/TATGGTGCACTCTCAGTACAATCTGCTCTGATGCCGCATAGTTAAGCCAGTATCTGCTCCCTGCTTGTGT

2334C
TTGAGCAGGATATTTGGTCCTGTAGTTTGCTAACACACCCTGC TGGTGCACTCTCAGTACAATCTGCTCTGATGCCGCATAGTTAAGCCAGTATCTGCTCCCTGCTTGTGT

2334D
TTGAGCAGGATATTTGGTCCTGTA() GTGCACTCTCAGTACAATCTGCTCTGATGCCGCATAGTTAAGCCAGTATCTGCTCCCTGCTTGTGT

2315B-II
TTGAGCAGGATATTTGGTCCTGTAGTTTGCTAACACACCCTGC GCACTCTCAGTACAATCTGCTCTGATGCCGCATAGTTAAGCCAGTATCTGCTCCCTGCTTGTGT

2315E
TTGAGCAGGATATTTGGTCCTGTAGTTTGC() TATGGTGCACTCTCAGTACAATCTGCTCTGATGCCGCATAGTTAAGCCAGTATCTGCTCCCTGCTTGTGT

2315F
TTGAGCAGGATATTTGGTCCTGTAGTTTGCTAACACACCCTT* TGCTCTGATGCCGCATAGTTAAGCCAGTATCTGCTCCCTGCTTGTGT

2315H (2315C-II)
TTG() GCTTGTGT

RSVCAT PstI
TTGAGCAGGATATTTGGTCCTGTAGTTTGCTAACACACCCTGCA/TTGGATCAGGAACCCAGCACTCCACTGGATAAGCATTATCCTTATCCAAAACAGCCTTGTGGTCAGTGTT

2315B-I
TTGAGCAGGATATTTGGTCCTGTAGTTTGCTAACACAC() CCAAAACAGCCTTGTGGTCAGTGTT

2315C-I
TTGAGCAGGATATTTGGTCCTGTAGTTTGCTAACACACCCTGA* AACCCAGCACTCCACTGGATAAGCATTATCCTTATCCAAAACAGCCTTGTGGTCAGTGTT

Figure 4. The structures of the seven PCR products which were sequenced are depicted in A. The position of the PstI and NdeI restriction endonuclease sites and the 199 and 200 oligonucleotide primer binding sites are shown on the top line. The solid lines denote the DNA segments that are present in each of the junction fragments. Fragments 2315B and 2315C each contain two junctions (I and II). In the first junction of each fragment, sequences near the PstI site are joined to sequences 5′ of the 199 oligonucleotide site (dashed lines). The sequences of the individual junctions are given in B. The top line shows the sequence of the corresponding segments of pRSVcat from the GenBank database. The 3′ overhang from the PstI site and the 5′ overhang from the NdeI site are indicated by boxes underneath the sequence. Rat liver topoisomerase I cleavage sites determined in vitro (38) are shown by . (... strong cleavage; .. intermediate cleavage; . weak cleavage). Homologies to the chi consensus sequence are underlined in the RSVcat sequences. Dinucleotide repeats at the junctions are denoted by () and nucleotides which differ from the RSVcat sequence are marked with an *. In the second part of B, the RSVcat sequence is given to the PstI site and then continues from a position 75 nucleotides 5′ of the oligonucleotide 199 binding site. The sequence of the junction in the 2315H fragment was identical to the sequence of the 2315C-II junction.

and performing parallel reactions. An unambiguous sequence across the deletion junction could not be attained for bands A, D, and G from M2315. The sequences of bands A and D from M2315 were readable in both directions nearly to the ends of the injected fragment before becoming unreadable. Preparations of band A were presumably contaminated with artifactual bands which were produced in the later cycles. Band D apparently contains a mixture of fragments which differ in length by only a few nucleotides. There was an insufficient amount of radioactivity in band G for sequencing.

The sequences of seven of the transgene junction fragments, containing eight novel deletion endpoints, are shown in figure 4. These sequences are compared to the sequence of the RSVcat plasmid compiled from the GenBank database. Four of the eight deletion endpoints occurred in a dinucleotide which was present at both ends of the linear fragment. Two of the eight deletion endpoints contained a nucleotide at the junction which differed from the consensus sequence. No other base changes were found in these seven fragments. Two of the seven fragments contained two junctions. As depicted in part A of figure 4, these fragments (2315B and 2315C) start at the 199 oligo and extend nearly to the PstI site at the end of the injected fragment. They are then joined to plasmid sequences upstream of the oligo 199 binding site. The sequences continue through a second oligo 199 binding site and are joined to the opposite end of the next gene fragment at different nucleotides.

The junction in fragment 2315H is identical to the second junction (II) in fragment 2315C. The latter fragment undoubtedly

arose by amplification from the internal 199 oligonucleotide binding site. Internal amplification of the second junction of fragment 2315B is expected to produce a fragment of the same size as 2315D. This sub-fragment of 2315B is presumably one of the mixture of fragments in the 2315D preparation.

In addition to the junctions that were sequenced, amplification of M2315 DNA with the 274/200 oligonucleotide pair produced a fragment of approximately 480 bp. The restriction map of this fragment indicated that it arose from a junction in which the NdeI end had been joined to RSVcat sequences approximately 1430 bp upstream of the PstI end.

DISCUSSION

Seven different junction fragments between integrated RSVcat genes in two transgenic mice were characterized at the level of DNA sequence. Several interesting features emerged. In no case was the sequence of the original microinjected fragment completely conserved; deletions of 1 to 62 bases from each end occurred. The distribution of deletion endpoints appears to be random. A single base mutation (transversion) of the last base at the 3′ end of the fragment occurred in two of the junctions. In both instances the mutation is found in what was originally the overhanging 3′ portion of the PstI sticky end. The exact location of four of the junctions could not be determined since joining occurred in a dinucleotide that could have come from either of the two ends of the linear DNA fragment. As discussed below, these dinucleotide repeats may have provided limited homology to facilitate ligation of the altered ends.

The transgene junctions characterized here serve as examples of some of the complexities involved in the incorporation of foreign microinjected DNA into the genome. Genomic DNA from mouse M2334 contained two different transgene junctions with equal copy numbers. This mouse had approximately 25 copies of the transgene arrayed in tandemly repeated units. At the nucleotide level, junction formation appears to be a random process. The simplest explanation for the low number of junctions in this mouse is that an initial three-gene concatamer, containing one copy of the 2334C junction and one copy of the 2334D junction, was amplified by DNA replication. Expansion of transgene complexes by unscheduled localized DNA replication has been postulated by others (24).

Mouse M2315 integrated the RSVcat DNA in a dramatically different fashion. This mouse was highly mosaic, and analysis of the accumulation rates of PCR products indicates a 30-fold range in the abundance of junction templates. In other words, these junctions were fixed into the genome at different times in the early developmental period of this animal. We envision two different scenarios to explain this difference in copy number. A large concatamer may have initially integrated at one unstable site and portions of the complex were subsequently deleted in a subset of cells in the early zygote. A transgenic mouse with a germline that was mosaic for a deleted form of the transgene complex found in somatic tissue has recently been described (25). In the case of transgenic mouse M2315, the junctions in the 2315B and C fragments would then be the products of separate and imprecise deletion events that removed most of a transgene segment.

Alternatively, the junctions in the 2315B and C fragments may have been created extrachromosomally and recombined with the original integrant at a later developmental time. These junctions would then be the products of end-joining reactions between the ends and double-stranded breaks within the microinjected DNA. Transfected DNA undergoes fragmentation (26,27) and it is not unreasonable to assume that microinjected DNA suffers the same type of damage.

PCR artifacts

The polymerase chain reaction technique made direct sequencing of the junctions possible, but it also produced some artifactual bands. The primary artifacts were heteroduplexed copies of different junctions. These fragments contain loops of single-stranded DNA which cause them to migrate anomalously on polyacrylamide gels (28–30). This was confirmed by sequence analysis. As the amplification products accumulate during the cyclic reactions, they eventually become able to compete with the oligonucleotide primers for binding to template DNA. The correspondence between the appearance of the heteroduplexes and the change from logarithmic to linear amplification suggests that this competition eventually limits further amplification.

We have also found that amplification of segments of DNA containing direct repeats (e.g., fragments B and C from M2315) can give rise to PCR products of ever-increasing length. The products of amplification can hybridize out of register such that the oligo 199 binding site at the 3′ end of one strand base pairs with the internal oligo 199 sequence on the other strand. This produces a hybrid with a large 5′ overhang containing the region between the two oligo 199 sequences. The action of DNA polymerase on the 3′ end would add another copy of the segment between the two oligo 199 sequences. Thus, the strand that was originally synthesized from the oligo 200 primer would be lengthened. The possibility that amplification products act as primers in subsequent cycles to create artifacts has been noted by others (31–33). Performing fewer rounds of PCR to insure that samples are taken during the logarithmic phase of amplification reduces the incidence of both of these types of artifacts.

Mechanisms of concatamer formation

The copy-choice (or strand slippage) model of DNA replication has been evoked as a model for nonhomologous recombination (34; 35, and references therein). Junctions which arise from this model usually display short homologies of more than four base pairs. This mechanism could explain two of the junctions between transgene segments described by Wilkie and Palmiter (5). Our junctions, however, are unlikely to have arisen by this mechanism because the maximum homology at any of the junctions is only two base pairs.

Topoisomerase I can mediate illegitimate recombination in vitro (36). Junctions that are formed by excision and circularization of integrated SV40 DNA are associated with topoisomerase I cleavage sites (37). Cleavage sites for topoisomerase I are difficult to predict, however, because the enzyme recognizes a degenerate sequence in vitro. None of the transgene junctions in this study correspond to the consensus sequence (A/T,G/C,T/A,T) for the strongest cleavage sites. Furthermore, the region of SV40 DNA contained in the RSVcat construct around the Pst I site has been tested for cleavage by wheat germ and rat liver topoisomerase (38). There is no strong association between the junction endpoints and cleavage sites in this region.

An interesting observation is that two sequences with strong homology to chi sites occur within 62 nucleotides of all of the sequenced junctions. Chi sites act as recombinational hotspots in bacteriophage lambda (39). These sites have also been noted

near other transgene junctions and have been suggested as helping to mediate transgene concatamerization and/or integration (5). Computer-assisted scanning of the microinjected RSVcat fragment revealed that it contains 15 sites with at least 75% (6/8 bases) homology to the chi consensus sequence (5'GCTGGTGG3'; ref. 40). The best match occurs between the oligo binding sites 197 and 274 (refer to figure 1). Amplification reactions using the 198/200 and 273/199 oligo pairs did not detect any junctions involving the ends of the injected fragment and this internal site.

The junctions between these transgene segments closely resemble those seen when transfected linear DNA is circularized in tissue culture cells (13). In this report, it was found that DNA fragments with a 3' overhang on one end and a 5' overhang on the other end were circularized by two different mechanisms. Approximately half the junctions were formed by homology-independent abutting of overhanging ends. The remainder were termed homology- dependent because they contained one to six base pairs of homology at the junction. In our study of transgene junctions formed in microinjected zygotes, four out of eight junctions are of the homology-dependent type. They contain two base pair homologies at the crossover site. The remainder of the transgene junctions are of the homology-independent type. It is of interest to note that all of these latter type junctions retain some portion of the PstI overhanging end. Homology-independent end-joinings in tissue culture cells nearly always retain the full overhangs of both ends (13). Also, the terminal nucleotide in two of the homology-independent transgene junctions was mutated. Mutated nucleotides were not observed in tissue culture.

The simplest explanation of the end-joining mechanism is a combination of exonuclease and ligation activities. Homology-independent joining would arise by ligation of DNA molecules with ends which are blunt or contain different types of overhanging ends. The exonuclease(s) may occasionally produce DNA molecules with compatible overhanging ends which.could be joined by sticky-end ligation (homology-dependent joining). The dinucleotide repeats at some of the junctions are indicative of this joining mechanism.

CONCLUSION

The mosaic nature of transgene complexes that have undergone deletions or additions of transgene segments at later stages of zygote development may go undetected by Southern blot analysis unless the mosaicism is present in the germline. Quantitation of different junction fragments by the more sensitive PCR amplification technique would be useful for detecting mosaic transgene structures in somatic tissues. Complications may be generated, however, due to the presence of cells of separate developmental origins in most tissues.

The integration of foreign microinjected DNA into the genomes of mouse embryos is a complex multistep process; evidence presented here indicates that alternate pathways to the end product are possible. It is unknown if the types of junctions we have described are general structures or are peculiar to this set of microinjections. The junctions in one mouse described by Wilkie and Palmiter (5) are distinctly different from the junctions characterized in our study. The ability to directly sequence junctions via PCR will allow the characterization of more junctions produced by other types of microinjected fragments.

ACKNOWLEDGEMENTS

The authors thank John Centanni for expert technical assistance.

REFERENCES

1. Gordon, J.W., Scangos, G.A., Plotkin, D.J., Barbosa, J.A. and Ruddle, F.H. (1980) Genetic transformation of mouse embryos by microinjection of purified DNA. Proc. Natl. Acad. Sci., USA, **77**, 7380–7384.
2. Brinster, R.L., Chen, H.Y., Trumbauer, M., Senear, A.W., Warren, R. and Palmiter, R.D. (1981) Somatic expression of herpes thymidine kinase in mice following injection of a fusion gene into eggs. Cell, **27**, 223–231.
3. Hogan, B., Costantini, F. and Lacy, E. (1986) Manipulating the Mouse Embryo—A Laboratory Manual. Cold Spring Harbor: Cold Spring Harbor Laboratory.
4. Brinster, R.L., Chen, H.Y., Trumbauer, M.Y., Yagle, M.K. and Palmiter, R.D. (1985) Factors affecting the efficiency of introducing foreign DNA into mice by microinjecting eggs. Proc. Natl. Acad. Sci., USA, **82**, 4438–4442.
5. Wilkie, T.M. and Palmiter, R.D. (1987) Analysis of the integrant in MyK-103 transgenic mice in which males fail to transmit the integrant. Mol. Cell. Biol. **7**, 1646–1655.
6. Covarrubias, L., Nishida, Y. and Mintz, B. (1986) Early postimplantation embryo lethality due to DNA rearrangements in a transgenic mouse strain. Proc. Natl. Acad. Sci., USA, **83**, 6020–6024.
7. Covarrubias, L., Nishida, Y., Terao, M., D'Eustachio, P. and Mintz, B. (1987) Cellular DNA rearrangements and early developmental arrest caused by DNA insertion in transgenic mouse embryos. Mol. Cell. Biol., **7**, 2243–2247.
8. Ninomiya, T., Hoshi, M. and Yuki, A. (1988) Structures of integrated DNA containing human adenovirus E1A genes in transgenic mice. Agric. Biol. Chem., **52**, 2537–2546.
9. Mahon, K.A., Overbeek, P.A. and Westphal, H. (1988) Prenatal lethality in a transgenic mouse line is the result of a chromosomal translocation. Proc. Natl. Acad. Sci., USA, **85**, 1165–1168.
10. Bishop, J.O. and Smith, P. (1989) Mechanism of chromosomal integration of microinjected DNA. Mol. Biol. Med., **6**, 283–298.
11. Folger, K.R., Wong, E.A., Wahl, G. and Capecchi, M.R. (1982) Patterns of integration of DNA microinjected into cultured mammalian cells: evidence for homologous recombination between injected plasmid DNA molecules. Mol. Cell. Biol., **2**,1372–1387.
12. Kopchick, J.J. and Stacey, D.W. (1984) Differences in intracellular DNA ligation after microinjection and transfection. Mol. Cell. Biol., **4**, 240–246.
13. Roth, D.B. and Wilson, J.H. (1986) Nonhomologous recombination in mammalian cells: role for short sequence homologies in the joining region. Mol. Cell. Biol., **6**, 4295-4304.
14. Palmiter, R.D., Brinster, R.L., Hammer, R.E., Trumbauer, M.E., Rosenfeld, M.G., Birnberg, N.C. and Evans, R.M. (1982) Dramatic growth of mice that develop from eggs microinjected with metallothionein-growth hormone fusion genes. Nature, **300**, 611–615.
15. Khillan, J.S., Overbeek, P.A. and Westphal, H. (1985) Drosophila P element integration in the mouse. Dev. Biol., **109**, 247–250.
16. Yamamura, K., Miki, T., Suzuki, N., Ebihara, T., Kawai, K., Kumahara, Y. and Honjo, T. (1985) Introduction of mouse CE genes into COS7 cells and fertilized mouse eggs. J. Biochem., **97**, 333–339.
17. Saiki, R.K., Scharf, S., Faloona, F., Mullis, K.B., Horn, G.T., Erlich, H.A. and Arnheim, N.(1985) Enzymatic amplification of beta-globin genomic sequences and restriction site analysis for diagnosis of sickle cell anemia. Science, **230**, 1350–1354.
18. Saiki, R.K., Gelfand, D.H., Stoffel, S., Scharf, S.J., Higuchi, R., Horn, G.T., Mullis, K.B.and Erlich, H.A. (1988) Primer-directed enzymatic amplification of DNA with a thermostable DNA polymerase. Science, **239**, 487–491.
19. Gorman, C.M., Merlino, G.T., Willingham, M.C., Pastan, I. and Howard, B.H. (1982) The Rous sarcoma virus long terminal repeat is a strong promoter when introduced into a variety of eukaryotic cells by DNA-mediated transfection. Proc. Natl. Acad. Sci., USA, **79**, 6777–6781.
20. Southern, E. (1975) Detection of specific sequences among DNA fragments separated by gel electrophoresis. J. Mol. Biol., **98**, 503–517.
21. Chelly, J., Kaplan, J.-C., Maire, P., Gautron, S. and Kahn, A. (1988) Transcription of the dystrophin gene in human muscle and non-muscle tissues. Nature, **333**, 858–860.

22. Smith, H.O. (1980) Recovery of DNA from gels. Methods in Enzymology, **65**, 371–380.
23. Maxam, A.M. and Gilbert, W. (1980) Sequencing end-labeled DNA with base-specific chemical cleavages. Methods in Enzymology, **65**, 499–560.
24. Wagner, E.F., Covarrubias, L., Stewart, T.A. and Mintz, B. (1983) Prenatal lethalities in mice homozygous for human growth hormone gene sequences integrated in the germ line. Cell, **35**, 647–655.
25. Xiang, X., Benson, K.F. and Chada, K. (1990) Mini-mouse: disruption of the pygmy locus in a transgenic insertional mutant. Science, **247**, 967–969.
26. Lebkowski, J.S., DuBridge, R.B., Antell, E.A., Greisen, K.S. and Calos, M.P. (1984) Transfected DNA is mutated in monkey, mouse and human cells. Mol. Cell. Biol., **4**, 1951–1960.
27. Wake, C.L., Gudewicz, T., Porter, T., White, A. and Wilson, J.H. (1984) How damaged is the biologically active subpopulation of transfected DNA? Mol. Cell. Biol., **4**, 387-398.
28. Bhattacharya, A. and Lilley, D.M.J. (1989) The contrasting structures of mismatched DNA sequences containing looped-out (bulges) and multiple mismatches (bubbles). Nucleic Acids Res., **17**, 6821.
29. Rice, J.A. and Crothers, D.M. (1989) DNA bending by the bulge defect. Biochemistry, **28**, 4512–4516.
30. Hsieh, C.-H. and Griffith, J.D. (1989) Deletions of bases in one strand of duplex DNA, in contrast to single-base mismatches, produce highly kinked molecules: Possible relevance to the folding of single-stranded nucleic acids. Proc. Natl. Acad. Sci., USA, **86**, 4833–4837.
31. Jeffreys, A.J., Wilson, V., Neumann, R. and Keyte, J. (1988) Amplification of human minisatellites by the polymerase chain reaction: towards DNA fingerprinting of single cells. Nucleic Acids Res., **16**, 10953–10971.
32. Shuldiner, A.R., Nirula, A. and Roth, J. (1989) Hybrid DNA artifact from PCR of closely related target sequences. Nucleic Acids Res., **17**, 4409.
33. Meyerhans, A., Vartanian, J.P. and Wain-Hobson, S. (1990) DNA recombination during PCR. Nucleic Acids Res., **18**, 1687–1691.
34. Canning, S. and Dryja, T.P. (1989) Short, direct repeats at the breakpoints of deletions of the retinoblastoma gene. Proc. Natl. Acad. Sci., USA, **86**, 5044–5048.
35. Mita, S., Rizzuto, R., Moraes, C.T., Shanske, S., Arnaudo, E., Fabrizi, G.M., Koga, Y., DiMauro, S. and Schon, E.A. (1990) Recombination via flanking direct repeats is a major cause of large-scale deletions of human mitochondrial DNA. Nucleic Acids Res., **18**, 561–567.
36. Halligan, B.D., Davis, J.C., Edwards, K.A. and Liu, L. (1982) Intra- and intermolecular strand transfer by HeLa DNA topoisomerase I. J. Biol. Chem., **251**, 3995–4000.
37. Bullock, P., Champoux, J.J. and Botchan, M. (1985) Association of crossover points with topoisomerase I cleavage sites: a model for nonhomologous recombination. Science, **230**, 954–958.
38. Been, M.D., Burgess, R.R. and Champoux, J.J. (1984) Nucleotide sequence preference at rat liver and wheat germ type 1 DNA topoisomerase breakage sites in duplex SV40 DNA. Nucleic Acids Res., **12**, 3097–3114.
39. Stahl, F.W. (1979) Special sites in generalized recombination. Ann. Rev. Genet., **13**, 7–24.
40. Smith, G.R., Kunes, S.M., Schultz, D.W., Taylor, A. and Triman, K.L. (1981) Structure of chi hotspots of generalized recombination. Cell, **24**, 429–436.