MOLECULAR AND CELLULAR BIOLOGY, Jan. 1996, p. 10–18 Vol. 16, No. 1
0270-7306/96/$04.00+0

High-Frequency Illegitimate Integration of Transfected DNA at Preintegrated Target Sites in a Mammalian Genome

RAYMOND V. MERRIHEW,[1] K. MARBURGER,[1] SANDRA L. PENNINGTON,[2] DAVID B. ROTH,[3] AND JOHN H. WILSON[1]*

The Verna and Marrs McLean Department of Biochemistry[1] and Department of Microbiology and Immunology,[3] Baylor College of Medicine, Houston, Texas 77030, and Division of Basic Sciences, Fred Hutchinson Cancer Research Center, Seattle, Washington 98104[2]

Received 14 April 1995/Returned for modification 9 June 1995/Accepted 16 October 1995

To examine the mechanisms of recombination governing the illegitimate integration of transfected DNA into a mammalian genome, we developed a cell system that selects for integration events in defined genomic regions. Cell lines with chromosomal copies of the 3′ portion of the adenine phosphoribosyltransferase (APRT) gene (targets) were established. The 5′ portion of the APRT gene, which has no homology to the integrated 3′ portion, was then electroporated into the target cell lines, and selection for APRT gene function was applied. The reconstruction of the APRT gene was detected at frequencies ranging from less than 10^{-7} to 10^{-6} per electroporated cell. Twenty-seven junction sequences between the integrated 5′ APRT and its chromosomal target were analyzed. They were found to be randomly distributed in a 2-kb region immediately in front of the 3′ portion of the APRT gene. The junctions fell into two main classes: those with short homologies (microhomologies) and those with inserted DNA of uncertain origin. Three long inserts were shown to preexist elsewhere in the genome. Reconstructed cell lines were analyzed for rearrangements at the target site by Southern blotting; a variety of simple and complex rearrangements were detected. Similar analysis of individual clones of the parental cell lines revealed analogous types of rearrangement, indicating that the target sites are unstable. Given the high frequency of integration events at these sites, we speculate that transfected DNA may preferentially integrate at unstable mammalian loci. The results are discussed in relation to possible mechanisms of DNA integration.

Mammalian genomes undergo many dynamic events and modifications. Chromosomal alterations under genetic control in normal cells include immune system rearrangements (31) and transposition events (75). In addition, genomic instability is a hallmark of tumor cell progression (24). Chromosomal deletions, translocations, and gene amplification are rearrangements commonly resulting from illegitimate (or nonhomologous) recombination in cancer cells. Although the processes that govern these rearrangements remain largely undefined, analysis of the DNA sequences at illegitimate recombination junctions has provided important information concerning the mechanisms involved (35).

The dynamic properties of mammalian chromosomes are evident in the illegitimate (nonhomologous) integration of exogenously introduced DNA. Mammalian cells integrate foreign DNA widely throughout the genome in a process often referred to as random integration (27, 58). The integration of DNA into the mammalian genome is efficient, with up to 20% of cells integrating microinjected DNA (10). Furthermore, integration events are usually associated with major chromosomal rearrangements that remain largely undefined (62). Fully characterized rearrangements include a 22-kb deletion (33) and a 5-kb duplication of the chromosomal sequence flanking the integrated DNA (76). Additional analysis of random integrants may provide insights into the processes involved in genomic instability.

Although free DNA ends stimulate random integration (18, 50), little is known about how the ends are joined to the chromosomal DNA at the sequence level. The traditional analysis of random integrants has proven tedious. Because each integration site is different, one must first isolate the chromosomal sequences that flank the integrated DNA and then use them to isolate the undisrupted DNA in the parental cells so that the chromosomal sequences on both sides of the junction can be determined. Such characterizations are further complicated by the major chromosomal rearrangements that usually accompany the integration events.

In this communication we describe a mammalian cell system that works around these problems by selecting for events in a small region of the genome (the target site). In order to survive selection, a cell must have an integration event that reconstructs a functional adenine phosphoribosyltransferase (APRT) gene. A collection of the resulting reconstructed $APRT^+$ cell lines have been examined. To gain insights into the mechanism of integration, we have determined sequences for junctions between the input DNA and the chromosomal DNA. In addition, we have examined the stability of the target sites prior to integration and the rearrangements associated with the integration events.

MATERIALS AND METHODS

Cell culture and subclone isolation. All cell lines used in these experiments are derivatives of Chinese hamster ovary line AT3-2, which is hemizygous for APRT (1, 3). The U1S36-TG^R cell line (55) is a hypoxanthine phosphoribosyltransferase-deficient derivative of U1S36 (42), a DNA-repair-deficient (UV-hypersensitive) line with a 4.1-kb deletion which covers the entire body of the hemizygous APRT locus. Cell lines DR15, DR11, DR33, RKM3, and RKM4 (Table 1) were generated by electroporating U1S36-TG^R with pDR27 (uncut for DR15; *Eco*RI-linearized for DR11, DR33, RKM3, and RKM4) and then selecting for gpt^+ colonies in HAT medium (100 μM hypoxanthine, 0.4 μM aminopterin, 16 μM thymidine; Sigma). Independent colonies were isolated and grown to confluency in two 100-mm-diameter plates for genomic DNA extraction and Southern blotting; cells at this point (about 25 total generations) were considered in

* Corresponding author. Mailing address: Department of Biochemistry, Baylor College of Medicine, Houston, TX 77030. Phone: (713) 798-5760. Fax: (713) 795-5487.

TABLE 1. Illegitimate reconstruction of APRT

Random integrant or native locus	No. of:				Reconstruction frequency[b]
	Genomic sites	Target copies	Electroporations[a]	APRT$^+$ colonies	
Random integrants					
DR15	1	3–4	10	95	4.7×10^{-7}
DR11	2	ND	1	2	1.0×10^{-7}
DR33	4	ND	1	0	$<5.0 \times 10^{-8}$
RKM3	1	1	4	6	7.5×10^{-8}
RKM4	1	1	1	0	$<5.0 \times 10^{-8}$
Native loci					
T2S24	1	1	1	1[c]	$<5.0 \times 10^{-8}$
DEPO1	1	1	3	0	$<1.7 \times 10^{-8}$
DEPO2	1	1	3	0	$<1.7 \times 10^{-8}$

[a] For each electroporation, 60 μg of pDR28 was digested with *Eco*RI, ethanol precipitated, and mixed with 2×10^7 cells in electroporation buffer. Treated cells were selected the next day for APRT function in ALASA.

[b] Frequencies of illegitimate reconstruction of the APRT gene were calculated as total APRT$^+$ colonies/total number of electroporated cells. Typical survival after electroporation under our conditions ranges from 60 to 80%.

[c] Analysis of this single colony showed that it resulted from homologous recombination between the transfected 5′ APRT and the native APRT locus.

early passage. Cell line T2S-24 TG^r-1 is a 6-thioguanine-resistant derivative of T2S-24 (42). Cell lines DEPO1 and DEPO2 were generated by a multistep targeting procedure that utilizes FLP/FRT site-specific recombination (34, 48).

Subclones of several cell lines were obtained. Cell lines DR15 and RKM3 were maintained in HAT medium, and APRT$^+$ derivatives of DR15 (DR15-1, DR15-19, and DR15-43) were maintained in ALASA medium (50 μM azaserine, 25 μM alanosine, 100 μM adenine; Sigma, National Cancer Institute, and Sigma). Subclones were isolated by dilution into 96-well dishes. Independent colonies were grown in HAT or ALASA for genomic DNA extraction and Southern blotting.

All cell cultures were maintained in Dulbecco's modified Eagle's medium supplemented with 10% fetal bovine serum (JRH Biosciences), penicillin, streptomycin, and minimum nonessential amino acids (Sigma). Cultures were incubated at 37°C in a humidified incubator with 5% CO_2.

Plasmid construction. The pDR27 plasmid was derived from pAG-1 (55) by deleting 1.3 kb of the 5′ APRT sequence between *Eco*RI sites, leaving APRT exons 3 through 5, a *gpt* cassette, and a pBR322 backbone. This 3′ APRT sequence includes the downstream portion (400 nucleotides) of intron 2. Plasmid pDR28 (Fig. 1) was generated by subcloning the 1.3-kb 5′ APRT *Eco*RI fragment from pAG-1 into pUC7. Plasmid pRM9 was generated by subcloning the 5′ portion of the APRT sequence into pACYC184 (New England Biolabs). The 5′ APRT sequence includes the upstream portion (600 nucleotides) of intron 2. Plasmid pGAL contains the full-length APRT sequence and is described elsewhere (29). Plasmid pMC1gptPolA (kindly provided by Geoff Sargent) is a derivative of pMC1neoPolyA (73) in which the *neo* gene has been replaced by the bacterial *gpt* cassette from pSV2gpt (40).

Electroporations. Conditions for electroporations were 1,000 V and 25 μF with HEPES (*N*-2-hydroxyethylpiperazine-*N*′-2-ethanesulfonic acid)-buffered glucose (12). For each experimental electroporation, 60 μg of pDR28 was digested with *Eco*RI, ethanol precipitated, and resuspended in electroporation buffer with 2×10^7 cells. Following electroporation, cells were plated at 5×10^5 or 1×10^6 cells per 100-mm-diameter dish and selected the following day for APRT expression in ALASA medium. After 2 weeks, ALASA-resistant colonies were picked and propagated. Transfection efficiencies were measured by transfecting a different selectable plasmid in parallel electroporations with the same population of cells used in the experiments. Transfection efficiencies were determined for U1S36-TG^R, DR15, RKM3, RKM4, and T2S-24 TG^r-1 cells by electroporating 5×10^6 cells with 5 μg of linearized pAG-1 or pGAL. A transfection efficiency was determined for DEPO1 cells by including 5 μg of linearized pMC1gptPolA (together with pDR28) in the experimental electroporation. Cells were plated in 100-mm-diameter dishes and selected the following day in ALASA (for pAG-1 and pGAL) or HAT (for pMC1gptPolA). After 2 weeks, colonies were stained and counted. Transfection efficiencies in these cell lines ranged from 1×10^{-3} to 2×10^{-3} in different experiments. A transfection efficiency of 2.8×10^{-3} for DR15 cells, described below (see Discussion), was obtained with conditions similar to those for experimental electroporations: 2×10^7 cells were transfected with 60 μg of pAG-1 and plated for selection in ALASA.

PCR. Various combinations of primer pairs specific to the 5′ APRT sequence and to the target site were used in PCRs to amplify across junctions in genomic DNA extracted from APRT$^+$-reconstructed cell lines. Primers were designed on the basis of maps of APRT (46) and pBR322 (71). Forward primers specific to 5′ APRT, with map positions shown in parentheses, were as follows: P1 (684 to 707), P2 (795 to 818), P3 (923 to 946), P4 (1078 to 1101), and P5 (1224 to 1247). Primers specific to target sequences included two reverse primers specific to pBR322, P6 (4151 to 4128) and P7 (4300 to 4277), and five reverse primers specific to 3′ APRT, P8 (1423 to 1400), P9 (1479 to 1456), P10 (1545 to 1522), P11 (1662 to 1639), and P12 (1817 to 1794). Primers used in PCRs to analyze insert sequences included P13 (5′-TTTCAGTTGGCAGTGGCAGGGTTC), P14 (5′-TCAGACTTGGGGTGAGTTCGAAGC), P15 (5′-CATGTCTGAGT CACCTGCTAACCT), P16 (5′-GTGATACAAGCAGGTGGGCAGATG), P17 (5′-TGTCGAGTATTACCCTGTTATC), P18 (5-TGCGGCCGCGTCGAGAC GC), P19 (5′-GAGTTATACTGTTGCGCGAAGTAG), and P20 (5′-GCGAAG TTCCTATACTTTCTAGAGA). Combinations of these primers were used under standard or optimized conditions (Stratagene Opti-Prime PCR kit) with Promega *Taq* DNA polymerase. Each reaction mixture contained approximately 500 ng of genomic DNA. All PCRs were performed in a Perkin-Elmer thermal cycler with an initial denaturation of 94°C for 5 min and then 30 to 40 cycles of 94°C for 30 s, 55 to 65°C for 30 s, and 72°C for 1 to 2 min. This was followed by a final cycle of 74°C for 7 min.

Inverse PCR. To clone the sequence flanking the 581-nucleotide insert at its preexisting genomic location, inverse PCR was employed as described previously (47). Briefly, to prepare template DNA for PCRs, genomic DNA from cell line U1S36-TG^R was digested with outside cutter *Rsa*I, diluted, and then circularized with T4 DNA ligase. Ethanol-precipitated template DNA was subjected to 30 rounds of PCR with inverse PCR primers IP1 (5′-GAACCCTGCCACTGC CAACTGAAA) and IP2 (5′-CAGGAAGTTTTTAACAGCTGCTTCG), and a 1-μl aliquot was reamplified with nested primers IP3 (5′-AGGTGCCAGCTGC GAGTAAGAGG) and IP4 (5′-AAACTACAGAGATTCCCCTGCCTC).

Sequencing of PCR products. Individual PCR products were gel purified by Gene Clean or MERmaid (Biolabs 101) and sequenced directly with Stratagene's Exo(−) Pfu DNA Cyclist sequencing kit. Particular products were additionally subcloned into pBluescript or into pGEM-T (Promega) by the T-tailing method (32) and subjected to DNA sequencing in the Department of Molecular and Human Genetics nucleic acids core laboratory at Baylor College of Medicine.

Southern blot analysis. Genomic DNA digested with *Bcl*I was subjected to electrophoresis on low-percentage (0.4%) agarose gels. The digestion of DNA with restriction enzymes and Southern blot analysis were carried out as previously described (65). Gels were blotted onto Hybond nylon filters (Amersham). Gel purified DNA fragments were labelled with Amersham's Multiprime for use as probes.

RESULTS

Illegitimate reconstruction of a target gene. To establish a system in which random integration into a defined region of the genome could be selected for, the 3′ portion of the APRT gene (containing exons 3, 4, and 5) was electroporated into the APRT-deleted cell line U1S36-TG^R with the plasmid pDR27. Individual *gpt*$^+$ integrants were selected by growth in HAT medium and then expanded and analyzed by Southern blotting. Cell lines with one site of integration (DR15), two sites of integration (DR11), and multiple sites of integration (DR33) were chosen to test the feasibility of gene reconstruction by illegitimate integration. Each of these cell lines was shown by

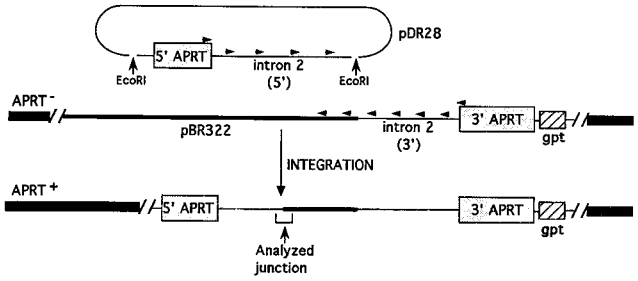


FIG. 1. General approach for nonhomologous reconstruction of the APRT gene. *Eco*RI-digested pDR28 was electroporated into APRT$^-$ cells having 3′ APRT target sites. As an example, a randomly integrated pDR27 target in cell line DR15 is illustrated. The integration of 5′ APRT at a nonhomologous chromosomal target reconstructs a functional APRT gene, allowing the cell line to survive ALASA selection. Also shown are relative positions of the primers (P1 to P5 and P6 to P12) used in PCRs and in sequence analyses of integration junctions.

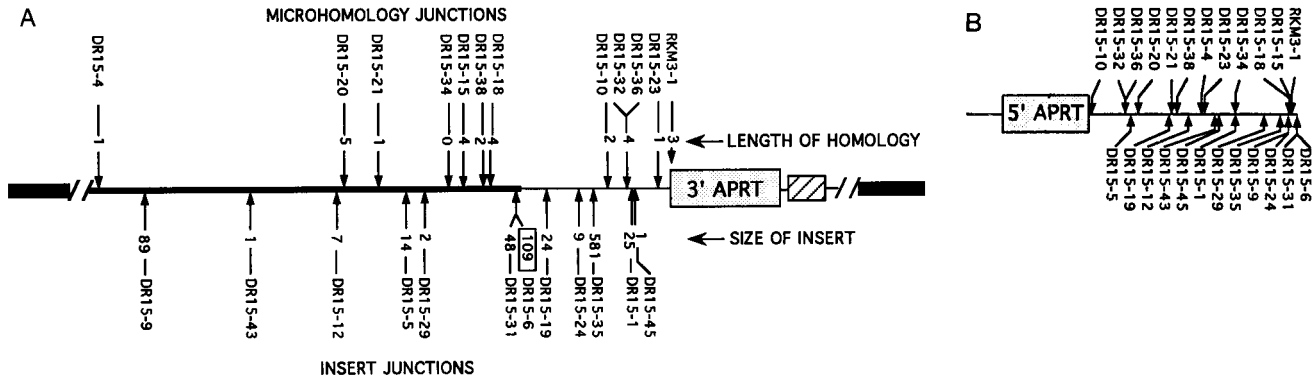


FIG. 2. Integration junctions at the gross level for 25 APRT reconstructants. (A) Relative integration sites and junctions for DR15-derived and RKM3-derived cell lines. For simplicity, integration events are shown in relation to only one target in DR15. Different reconstruction events in DR15 may have occurred at different targets in the tandem array. The illustrated target has greater than 1,600 nucleotides of the combined 3′ APRT intron sequence (thin line) and the pBR322 sequence as determined by sequence analysis (DR15-4 integration junction). The 109-nucleotide insert at the DR15-6 integration junction is boxed, since it does not represent a true insert. (B) Corresponding deletions from 5′ APRT intron sequences. The illustrated upstream portion of APRT intron 2 has 600 nucleotides of dispensable sequence. The site of deletion for each integration junction is marked by an arrow; all of the sequence to the right of the arrow was deleted.

Southern blotting to contain at least one intact copy of the 3′ portion of the APRT gene.

To test whether a functional APRT gene could be reconstructed by illegitimate integration, 2×10^7 cells of each of these cell lines were electroporated with 60 μg of plasmid pDR28 (Fig. 1) that had been digested with *Eco*RI, ethanol precipitated, and resuspended in electroporation medium. Plasmid pDR28 contains the 5′ portion of the APRT gene (exons 1 and 2 and the APRT promoter) but includes no sequence overlap with the integrated 3′ portion of the gene. Treated cells were selected for APRT function by growth in ALASA. Cell line DR33 gave no $APRT^+$ colonies, but DR15 and DR11 each yielded two colonies. Southern blotting confirmed the linkage of the 5′ and 3′ portions of the APRT gene for all four colonies (data not shown), giving a reconstruction frequency for these cell lines of 10^{-7} $APRT^+$ colonies per electroporated cell.

Cell line DR15 was chosen for subsequent experiments because its single site of integration promised a simpler analysis. Additional electroporations of *Eco*RI-digested pDR28 into DR15 gave reconstruction frequencies ranging from 1×10^{-7} to 10×10^{-7} with a mean of 4.8×10^{-7} (Table 1).

Reconstruction of the APRT gene at single-copy sites. To test whether APRT gene reconstruction was possible in cell lines containing a single copy of pDR27, two cell lines, RKM3 and RKM4, were isolated after electroporation of pDR27 into U1S36-TG^R. The electroporation of *Eco*RI-digested pDR28 into RKM4 gave no $APRT^+$ colonies, but four electroporations into RKM3 gave six colonies, for a reconstruction frequency of 7.5×10^{-8} (Table 1).

As an additional test for integration at single-copy sites, we used two deletions that encompassed the 5′ end of the gene at the native APRT locus: a spontaneous 3-kb deletion in cell line T2S24 TG^r-1 and a constructed 1.3-kb deletion in cell lines DEPO1 and DEPO2, which are independent clonal isolates that contain identical deletions. The deletion in T2S24 TG^r-1 leaves about 200 nucleotides adjacent to the *Eco*RI site in intron 2 that are homologous to the 5′ portion of the APRT gene used in the electroporation. The one $APRT^+$ colony that arose in this experiment proved to have arisen by homologous recombination in the segment of homology (data not shown).

The deletions in DEPO1 and DEPO2, which were constructed by site-specific gene targeting (34), correspond exactly to the 5′ portion of the APRT gene in pDR28; thus, they contain no homology with pDR28. No APRT gene reconstruction events were detected in several experiments (Table 1). Combining the results from the two deletions at the APRT locus gives an aggregate gene reconstruction frequency at the native APRT locus of less than 7.1×10^{-9} per electroporated cell.

Control experiments. Two controls were included in these experiments to rule out the possibility that the plasmid portion of pDR28, which is homologous to the plasmid at the integration site, was required for the observed reconstruction events. In one experiment, the 5′ portion of the APRT gene in plasmid pDR28 was digested with *Eco*RI and purified away from the plasmid backbone by agarose gel electrophoresis prior to electroporation into DR15. The electroporation of a molar amount of the purified fragment equal to that used in other experiments yielded a reconstruction frequency of 1.5×10^{-7}. In a second set of experiments, the 5′ APRT fragment was digested from plasmid pRM9, which has no homology to the target sites. The electroporation of equimolar amounts of the digested DNA yielded reconstruction frequencies of 5×10^{-8} for RKM3 cells and 1×10^{-7} for DR15 cells. Since these reconstruction frequencies were within the range observed in experiments with *Eco*RI-digested pDR28, we conclude that the observed reconstruction events do not depend on homology in the plasmid backbone.

An additional control was included to rule out possible effects of residual restriction enzyme. Although no *Eco*RI activity could be detected after ethanol precipitation, a sample of *Eco*RI-digested pDR28 was phenol extracted before it was subjected to ethanol precipitation. That experiment gave a reconstruction frequency of 10×10^{-7} per electroporated cell, indicating that residual *Eco*RI activity was not responsible for the observed reconstruction events.

Sequence analysis of integration junctions. Genomic DNA extracted from cell lines with APRT reconstruction events was subjected to PCRs with various pairs of primers specific to 5′ APRT and to target sequences. Genomic DNA preparations from 34 DR15-derived cell lines were examined. Of these, 90% (31 of 34) gave a PCR product that spanned the junction. Southern blotting demonstrated that 5′ and 3′ portions of the APRT sequence are linked for the three cell lines that failed to give a PCR product (data not shown). Twenty-five PCR products were sequenced to give a junction between the 5′ APRT segment and the chromosomal target. Twenty-three of these

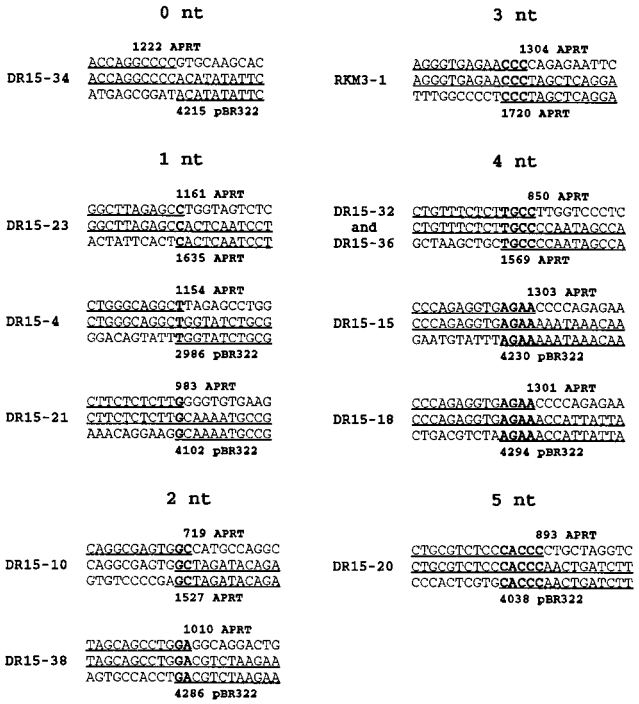


FIG. 3. Junction sequences with 0 to 5 nucleotides (nt) of homology. Each top line is the 5′ APRT sequence, and each bottom line is the target sequence. Middle lines represent the actual junction sequences, which are underlined. Junctional homologies are shown in boldface type. Nucleotide positions are based on GenBank maps for the APRT gene (46) and pBR322 (69). Reading from left to right, the first numeral of the nucleotide position is above or below the base to which it corresponds. (A few sequence differences from the published APRT sequence were detected in the APRT intron 2 sequence: an extra GCC CCG after APRT 1222, an extra C after position 1289, the lack of a C at position 1225, the lack of a GG at positions 1233 to 1234, and the lack of a G at APRT 1295).

junctions are presented (Fig. 2, 3, and 4). The remaining two junctions contain portions of the pUC sequence (pDR28 backbone). Since these two junctions have inserts composed of vector DNA and therefore represent special cases, they will not be considered further.

In addition to the 25 junction sequences determined for the DR15-derived cell lines, two junctions in the APRT$^+$ cell lines derived from RMK3 were sequenced. One of these junctions is presented (Fig. 2 and 3). The other junction also contained an insert composed of input DNA (a small portion of the 5′ APRT sequence) and will not be considered further.

As presented in Fig. 2A, integration events in DR15 and RKM3 occurred at many different sites within the target sequence. No preferential sites of integration were apparent. Although two cell lines (DR15-32 and DR15-36) had the same junction sequence, they were isolated in independent experiments. Of the 24 integration junctions considered here, 9 had small terminal deletions in the 5′ APRT intron sequence of 50 nucleotides or less. The remaining 15 junctions had a relatively even distribution of terminal deletions of 51 to 600 nucleotides (Fig. 2B).

Classes of integration junction sequences. A total of 24 integration junctions have been categorized. One of these (DR15-6) occurred by integration into chromosomal DNA that was linked to the target APRT sequences in the parent DR15 cell line (see below); this junction is not fully characterized. The remaining 23 integration junctions may be divided into two classes: 12 contained 0 to 5 nucleotides of homology (microhomology) at the junction (Fig. 3) and 11 contained inserted DNA of uncertain origin that ranged in size from 1 to 581 nucleotides (Fig. 4). None of the inserted sequences (excluding inserts of 1 to 9 nucleotides) demonstrate significant homology to those in the mammalian GenBank database (version 80). They also do not show any significant homology to one another or to the sequences used in these experiments.

PCR analysis of large inserts. In principle, the DNA inserts at the junctions in the integrants could have been created de novo or they could have been picked up in one or more pieces from the genome. We used a PCR analysis to distinguish among these possibilities for the larger inserts (49, 89, 109, and 581 nucleotides). Using primers located near the ends of the inserted sequences (primer pair I), we showed that each of

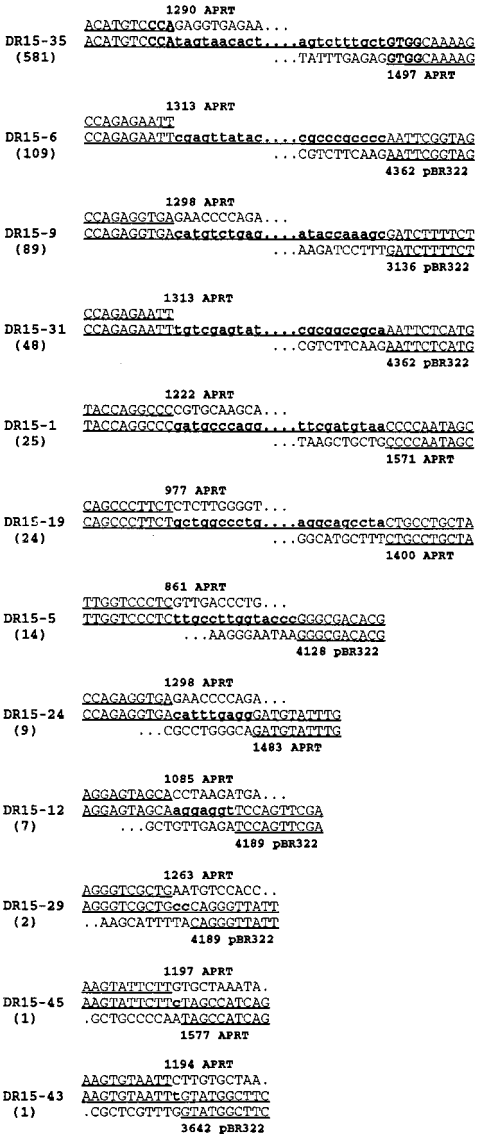
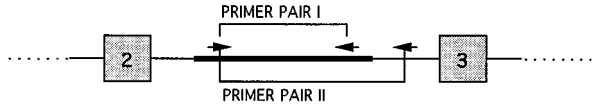


FIG. 4. Junction sequences containing inserts. The insert sequences are displayed in boldface lowercase letters, and the number of nucleotides in each insert is noted in parentheses; the 5′ APRT and target sequences are in uppercase letters. The nucleotides in boldface uppercase letters flanking the 581-nucleotide insert represent junctional homologies with the cloned sequence flanking the insert at its preexisting genomic site.



Cell line	Insert	Primer Pair	Reconstructed APRT+ cells	Parent lines DR15	Parent lines U1S36
DR15-35	581 nt	I (P13/P14)	+	+	+
		II (P13/P10)	+	-	-
DR15-9	89 nt	I (P15/P16)	+	+	+
		II (P15/P6)	+	-	-
DR15-31	48 nt	I (P17/P18)	+	+	+
		II (P17/P10)	+	-	-
DR15-6	109 nt	I (P19/P20)	+	+	+
		II (P19/P10)	+	+	-

FIG. 5. Approach and results of PCR analysis of large inserts. The thick line represents the insert sequence. Primer pair I is used to amplify the insert sequences. Primer pair II tests if the insert sequences are directly linked to the target site; one primer is to the insert sequence, and the other primer is to target sequence. Primers are described in Materials and Methods. A plus means that a PCR product of predicted size was detected on ethidium bromide-stained gels. A minus means that a product of predicted size was not detected. nt, nucleotide.

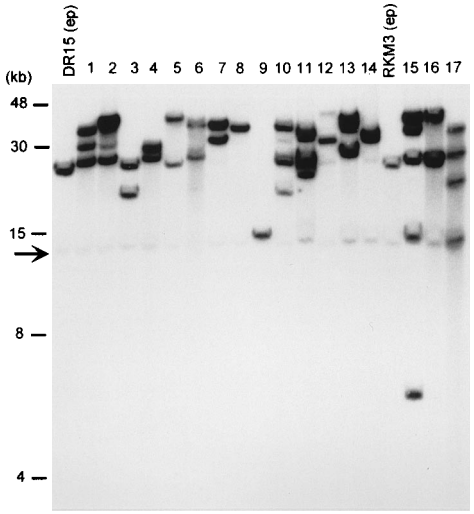


FIG. 6. Examples of rearrangements associated with integration events. Genomic DNAs from the early passage (ep) parent cell lines, DR15 and RKM3, and from reconstructed APRT$^+$ cell lines (lanes 1 to 17) were digested with *Bcl*I, which cuts outside the target sites, and analyzed by Southern blotting with the 3′ APRT sequence as a probe. The arrow marks the faint bands corresponding to the native APRT fragment. Lanes marked 1 to 14 correspond, respectively, to the following cell lines: DR15-24, DR15-5, DR15-23, DR15-8, DR15-18, DR15-31, DR15-4, DR15-21, DR15-10, DR15-35, DR15-36, DR15-37, DR15-34, and DR15-32. Lanes marked 15 to 17 are as follows: cell lines RKM3-2, RKM3-1, and RKM3-3.

these sequences preexisted in the U1S36 parent cell line from which DR15, the target cell line, was made (Fig. 5). In addition, by using a primer in the 3′ APRT target and one in the insert (primer pair II), we showed that the 109-nucleotide insert, but none of the others, was already linked to the target APRT sequences in the parent cell line, DR15 (Fig. 5). Thus, the 109-nucleotide sequence is not a true insert but is really part of the chromosomal sequences at the target site.

For the 581-nucleotide insert we confirmed that the entire insert preexisted in the parent DNA by using inverse PCR to clone a genomic segment that contained the insert. The sequence of the genomic clone corresponded exactly to the 581-nucleotide insert in the reconstructed cell line. In addition, the flanking sequences revealed a 3-nucleotide homology at the junction with the 5′ APRT sequences and a 4-nucleotide homology at the junction with the 3′ APRT target sequences (Fig. 4). Although the anonymity of the inserted sequences generally obscures information about how these junctions were generated, analysis of the 581-nucleotide insert suggests that microhomologies may be important in their formation.

Rearrangements associated with the integration events. To characterize the types of rearrangement, if any, associated with the integration events, we examined the DNA from reconstructed APRT$^+$ cell lines by Southern blotting. Genomic DNAs from 24 DR15-derived cell lines and 3 RKM3-derived cell lines were digested with *Bcl*I, which cuts outside the target sites, and analyzed by Southern blotting with the 3′ APRT sequence as a probe. Six of the DR15-derived DNAs, four of which (Fig. 6, lanes 8, 9, 12, and 14) are shown, gave a single band that was shifted with respect to the parental band, whereas all of the others gave a more complex pattern consisting of two to five bands (Fig. 6).

To determine whether the integrated 5′ APRT sequence was associated with one or more than one of the multiple bands, the filters were stripped and reprobed with the 5′ APRT sequence. In 20 of the 21 cell lines that gave multiple bands, only one of those bands was detected with the 5′ APRT probe; in the remaining cell line, RKM3-2, two of the bands contained 5′ APRT sequences (data not shown).

Stability of the target sites. Although the cell lines with a single shifted band could have resulted from deletion events associated with the integration of the 5′ APRT segment, the more complex patterns were unexpected and they raised questions about the stability of the target site itself. To test this, DR15 and RKM3 were subcloned and analyzed by Southern blotting as described above. For DR15, two of eight subclones had rearrangements at the target site, and for RKM3, eight of nine subclones had rearrangements (Table 2).

One of the subclones of DR15 that had a rearrangement (DR15sci4) and one that did not (DR15sci3) were subcloned again to test whether the rearrangements were ongoing. Four of 13 subclones of DR15sci3 and 2 of 10 subclones of DR15sci4 had rearrangements at the target sites (Table 2). Continuing rearrangement was also a property of the reconstructed cell lines, as shown for DR15-1, DR15-19, and DR15-43 in Table 2. Examples of the rearrangements at the target sites in different subclones are shown in Fig. 7.

To ensure that the observed rearrangements were not exclusive to single-colony isolates, cell lines DR15 and RKM3 were passed several times (an estimated 10 additional generations for DR15 and 20 additional generations for RKM3)

TABLE 2. Rearrangements at target sites

Cell line	No. of: Subclones examined	No. of: Rearrangements detected[a]
DR15	8	2
DR15sci3	13	4
DR15sci4	10	2
DR15-1	9	6
DR15-19	5	3
DR15-43	7	2
RKM3	9	8

[a] For a rearrangement to be detected, it has to give a band shift on a Southern blot; therefore, some small rearrangements will likely not be detected.

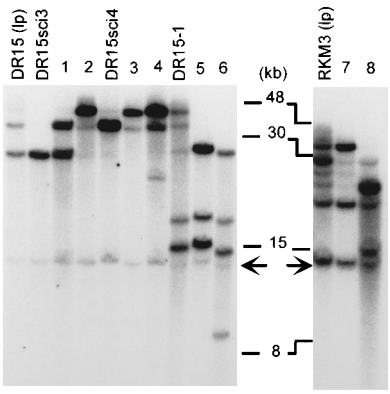


FIG. 7. Examples of late passage (lp) parent cell lines and subclones showing rearrangements at the target sites by Southern blotting. Numbered lanes are subclones of the cell line to their immediate left. Also shown are late passage parent cell lines DR15 and RKM3. The arrows mark the position of the native APRT fragment detected with the 3′ APRT probe.

before isolation of genomic DNA. Southern blot analysis of these late-passage DNAs showed extra bands for both DR15 and RKM3, indicating rearrangement of the target sequences during cell expansion (Fig. 7). This contrasts to the single bands detected in early-passage DNA from DR15 and RKM3 cells (Fig. 6).

In all Southern blots probed with the 3′ APRT sequence, a faint band corresponding to the undeleted portion of the native APRT gene remained constant (Fig. 6 and 7). This suggested that the instability associated with the target sites was not the result of high frequencies of rearrangement throughout the genome. This was confirmed by stripping blots and reprobing with fragments from three other cellular genes; in no case were rearrangements detected (data not shown).

The instability of the target sites was also examined by selection. Cell lines were preselected in HAT medium for *gpt* function and then allowed a 1-day recovery in HT medium. Each cell line was then replated to three dishes at 5×10^5 per plate for selection in 6-thioguanine for the loss of *gpt* function. Cell lines DR15, RKM3, and DR11 gave *gpt* mutant colonies at frequencies of 2×10^{-4} to 4×10^{-4}; cell lines RKM4 and DR33 gave frequencies of 7×10^{-5} to 9×10^{-5}.

DISCUSSION

By splitting the APRT gene in the second intron and placing the 3′ half in the genome, we have set up a system in which we can select for the illegitimate integration of the 5′ half of the gene into a particular genetic region by selecting for APRT gene function (Fig. 1). If all regions of the genome were equally accessible for illegitimate integration, that is, if integration were truly random, we would expect these gene reconstruction events to occur at a frequency of about 10^{-9}. This estimate is based on our measured rate of efficiency for stable transfection in cell line DR15 (2.8×10^{-3} per treated cell), our finding that reconstruction events occurred by integration within 2 kb of exon 3 (the first exon in the target), and the size of the diploid mammalian genome (6×10^6 kb): $(2.8 \times 10^{-3}) \times (2 \text{ kb}) \times (1/6 \times 10^6 \text{ kb}) = 10^{-9}$.

Compared with this genome-wide average, gene reconstruction in RKM3 (7.5×10^{-8}) and DR15 (4.7×10^{-7}) cells occurred, respectively, 75- and 470-fold more frequently than expected (Table 1). Expressed in a different way, these numbers correspond to one APRT$^+$ reconstructant per 6,000 integrants for DR15 and one reconstructant per 37,000 integrants for RKM3, compared with one reconstructant per 3×10^6 integrants expected on a random basis. Thus, the target sites in these cell lines represent "hot spots" for illegitimate integration. By contrast, illegitimate gene reconstruction at the native APRT locus was undetected in seven electroporations, for an aggregate frequency of $<7 \times 10^{-9}$ (Table 1). Thus, the native APRT locus may have only average or less than average accessibility for illegitimate integration.

Our results are analogous to those of a previous report in which preferential integration of transfected DNA occurred at an unstable site, in that case at an induced fragile site (57). We have shown by Southern blotting that 2 target sites representing hot spots for illegitimate integration are unstable. Nearly all RKM3 subclones and approximately 25% of all DR15 subclones had rearrangements of target sequences (Table 2). Thus, the frequency of integration at a particular genomic site may be related to the stability of that site. Although we have not yet tested this rigorously, it is noteworthy that cell lines DR15, DR11, and RKM3 gave higher frequencies of APRT reconstructants (Table 1) and higher frequencies of *gpt* mutant colony formation under selection than did the RKM4 and DR33 cell lines.

The basis for the instability of the target sites in DR15 and RKM3 is unclear. It may be influenced by the genotype of the parental U1S36 cell line, which contains a mutation in the ERCC2 gene rendering it deficient in nucleotide-excision repair; however, these cells integrate transfected DNA with the same level of efficiency as wild-type cells and loci outside the targets in DR15 and RKM3 cells were shown to be stable. If integration occurs preferentially at unstable sites, it may be that the original target plasmid also integrated at an unstable site in the genome and the ongoing rearrangement of the target is a consequence of that underlying instability. Conversely, the target plasmid itself may be responsible for the observed instability. The tandem repeat arrangement of plasmids in DR15 may contribute to its instability, but tandem copies are not a requirement since RKM3 has only a single integrated copy of the plasmid. A more likely possibility is the presence of the simian virus 40 origin of replication, which is contained within the promoter used to drive the *gpt* gene in the plasmid. Even in the absence of T antigen, the simian virus 40 origin of replication has been shown to promote an ongoing instability at sites where it is integrated in the genome of mouse cells (26). This may be due to some structural feature of the origin sequence or to the function of the sequence as an origin of replication, perhaps generating an "onion-skin" structure whose resolution leads to the instability (8, 26). We are testing this possibility by placing the simian virus 40 sequences at the native APRT locus and determining their effects on the stability and frequency of integration at that site.

Although the instability and complex rearrangements at the two analyzed target sites were unexpected, the integration junctions themselves are indistinguishable from the previously reported integration junctions generated by the transfection of DNA into mammalian cells. Of our 23 characterized junctions, 12 had 0 to 5 nucleotides of homology at the junction and 11 had inserted DNA ranging in size from 1 to 581 nucleotides. Of 13 such junctions in the literature, 7 had 0 to 6 nucleotides of homology (4, 23, 41) and 6 had inserted DNA ranging from 1 to 523 nucleotides (33, 41, 76). The nearly equal numbers of junctions with microhomology and junctions with inserts stand in contrast to their distributions in other types of illegitimate recombination junctions. The joining of DNA ends (59) and the generation of gene deletions (6, 7, 9, 11, 13–17, 19–22, 25, 28, 30, 36–39, 43–45, 49, 51, 54, 56, 64, 66–68, 70, 72, 77)

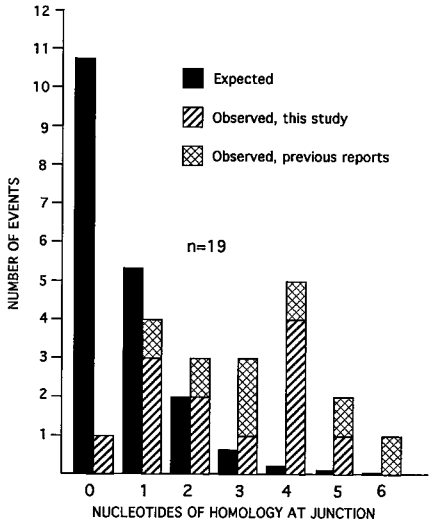


FIG. 8. Distribution of homologies at transfected DNA integration junctions. The distribution of 12 junctional homologies from this study (Fig. 3) is shown as hatched bars. Also included is the distribution of seven previously reported integration junctions from DNA transfections in mammalian cells (4, 23, 41), shown as diagonally crossed bars. The expected distribution of homologies (black bars) is based on the idea that the junctions arose in a way that was independent of homology (for example, by blunt-end joining, which can exhibit homology because of chance identities of flanking nucleotides). The expected distribution was determined by using the formula $P(x) = (x + 1)(1/4)^x(3/4)^2$, where $P(x)$ is the probability of finding x nucleotides of homology at the junction (60). For instance, the probability of finding 1 nucleotide of homology at the junction is given by $(2)(1/4)(3/4)^2$. This value (0.28) was then multiplied by the total number of junctions ($n = 19$) to give the expected distribution (5.3 junctions of 19 total junctions).

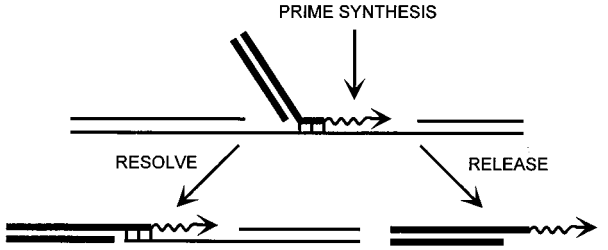


FIG. 9. Copy-join model for random integration events. Transfected DNA (thick lines) primes synthesis (wavy arrow) at a 3′-OH with a few homologous nucleotides (short vertical lines) within a single-stranded chromosomal region (gap). Resolution by endonuclease cleavage leaves the transfected DNA linked to the chromosome. The release of the invading DNA by a helicase leaves a new sequence at its end; integration elsewhere in the genome would result in inserted DNA at the junction. For simplicity, copying is shown for only one end of the transfected DNA; integration via a copy-join process would require priming at both ends.

typically produce junctions with microhomology 80 to 90% of the time and junctions with inserts only about 10 to 20% of the time.

As shown to be the case for other illegitimate junctions (60), the distribution of junctional microhomology is significantly skewed from that expected if the junctions arose in a way that did not depend on sequence homology (Fig. 8, expected levels), indicating that microhomology is important for the formation of integration junctions. The actual distribution of microhomology at integration junctions, however, seems to be different from that of other types of illegitimate junctions. Only 1 of 19 integration junctions had 0 nucleotides of homology, whereas 20 to 30% of gene deletions and extrachromosomal end joining events had 0 nucleotides of homology. In addition, in two recent studies, deletions that were induced by electroporating restriction enzymes, and therefore are likely to have arisen by the joining of cut ends, had junctions with 0 nucleotides of homology in about 30% of cases (53, 63). These comparisons suggest that the mechanism of integration of DNA into the chromosome may be more dependent on short homologies than are illegitimate recombination events that are thought to proceed by the joining of DNA ends.

The mechanism by which DNA becomes inserted at integration junctions is unknown. In this study the larger inserts (49, 89, and 581 nucleotides) were shown to preexist as discrete units elsewhere in the genome. These sequences could have become associated with the integrated DNA in several ways. First, they could have become linked to the target site as a consequence of a rearrangement that preceded the integration event. Although this is entirely reasonable given the ongoing instability at the target site, it does not readily account for the preponderance of short inserts (8 of 11 inserts were 25 nucleotides or less). Second, the input DNA could have picked up the inserts by end joining with extrachromosomal DNA fragments prior to integration. Extrachromosomal pieces of chromosomal DNA have been shown to exist in several cell types (74), and end joining of extrachromosomal DNA is a commonly observed phenomenon in transfection experiments with mammalian cells (62). Third, the input DNA could use short terminal homologies to prime DNA synthesis within an accessible region of the genome and in that way pick up terminal extensions prior to integration at the target site. An analogous mechanism, but involving more extensive homology, has been proposed for a common class of targeted recombinants that have picked up terminal extensions from the target locus prior to integration elsewhere in the genome (2, 5).

Our experimental results do not define a mechanism for the integration of transfected DNA, but they do put some constraints on the process. Any proposed mechanism must account for a seemingly enhanced integration at unstable sites, for the high frequency of inserted DNA, and for the relative abundance of 1- to 5-nucleotide homologies at integration junctions. Previously, we proposed that integration might occur by the end joining of transfected DNA to broken chromosome ends (62). Although our present results do not rule out such a mechanism, they do not support it either. The high frequency of inserted DNA and the low frequency of junctions with 0 nucleotides of homology were unexpected because of our previous characterization of extrachromosomal end joining (60–62) and the more recent characterization of the end joining of chromosomes cut by restriction enzymes (53, 63).

Many of the characteristics of the integration process could be accounted for by a mechanism of integration in which foreign DNA copies its way into the chromosome by invading a single-stranded region (Fig. 9). Exposed single strands are natural intermediates in DNA metabolism; they arise transiently during replication, repair, recombination, and transcription. If transfected DNA can prime DNA synthesis with short, terminal homologies, it could generate a branched structure that could be resolved by endonuclease cutting to leave the foreign DNA linked to the chromosome. Such a "copy-join" process might be expected to leave junctions with more extensive short homologies. An analogous priming process using short homologies has been demonstrated with extracts of *Xenopus* oocytes (52) and has also been suggested as a central step in slipped-mispairing models for gene deletions in bacteria (69). If the branched intermediate were resolved instead by a helicase, the invading DNA would be released with a terminal extension. A

subsequent integration event that used homology in the terminal extension would generate a junction with inserted DNA. As expected from this model, the fully characterized 581-nucleotide insert in DR15-35 has a 3-nucleotide homology with the transfected DNA at one end and a 4-nucleotide homology with the target DNA at the other end (Fig. 4). Such a copy-join process might also account for the high frequency of integration at the target sites in DR15 and RKM3 cells, if the embedded simian virus 40 origin promoted onion-skin replication and exposed extensive single-stranded regions. In addition, independent copy-join events at the two ends of the transfected DNA could contribute to the extensive alterations in the arrangement of chromosomal DNA that are commonly observed at sites of integration in mammalian cells. For instance, if the two ends of the transfected DNA each prime synthesis at separate sites on the same chromosome, the intervening chromosomal sequence would be deleted upon integration.

ACKNOWLEDGMENTS

We thank Geoff Sargent, Mark Brenneman, April Kilburn, and Karen Vasquez for valuable discussions and the Molecular Biology Group in the Baylor Biochemistry Department for helpful criticisms.

This work was supported by NIH grant GM38219 to J.H.W.

REFERENCES

1. **Adair, G. M., J. H. Carver, and D. L. Wandress.** 1980. Mutagenicity testing in mammalian cells. I. Derivation of a Chinese hamster ovary cell line heterozygous for the adenine phosphoribosyltransferase and thymidine kinase loci. Mutat. Res. **72:**187–205.
2. **Adair, G. M., R. S. Nairn, J. H. Wilson, M. M. Seidman, K. A. Brotherman, C. MacKinnon, and J. B. Scheerer.** 1989. Targeted homologous recombination at the endogenous adenine phosphoribosyltransferase locus in Chinese hamster cells. Proc. Natl. Acad. Sci. USA **86:**4574–4578.
3. **Adair, G. M., R. L. Stallings, R. S. Nairn, and M. J. Siciliano.** 1983. High-frequency structural gene deletion as the basis for functional hemizygosity of the adenine phosphoribosyltransferase locus in Chinese hamster ovary cells. Proc. Natl. Acad. Sci. USA **80:**5961–5964.
4. **Allen, M. J., A. J. Jeffreys, M. A. Surani, S. Barton, M. L. Norris, and A. Collick.** 1994. Tandemly repeated transgenes of the human minisatellite MS32 (D1S8), with novel mouse gamma satellite integration. Nucleic Acids Res. **22:**2976–2981.
5. **Aratani, Y., R. Okazaki, and H. Koyama.** 1992. End extension repair of introduced targeting vectors mediated by homologous recombination in mammalian cells. Nucleic Acids Res. **20:**4795–4801.
6. **Barsh, G. S., C. L. Roush, J. Bonadio, P. H. Byers, and R. E. Gelinas.** 1985. Intron-mediated recombination may cause a deletion in an α1 type I collagen chain in a lethal form of osteogenesis imperfecta. Proc. Natl. Acad. Sci. USA **82:**2870–2874.
7. **Bernatowicz, L. F., X.-M. Li, R. Carrozzo, A. Ballabio, T. Mohandas, P. H. Yen, and L. J. Shapiro.** 1992. Sequence analysis of a partial deletion of the human steroid sulfatase gene reveals 3 bp of homology at deletion breakpoints. Genomics **13:**892–893.
8. **Botchan, M., W. Topp, and J. Sambrook.** 1978. Studies on simian virus 40 excision from cellular chromosomes. Cold Spring Harbor Symp. Quant. Biol. **43:**709–719.
9. **Breimer, L. H., J. Nalbantoglu, and M. Meuth.** 1986. Structure and sequence of mutations induced by ionizing radiation at selectable loci in Chinese hamster ovary cells. J. Mol. Biol. **192:**669–674.
10. **Capecchi, M. R.** 1980. High efficiency transformation by direct microinjection of DNA into cultured mammalian cells. Cell **22:**479–488.
11. **Chen, S.-H., and C. R. Scott.** 1990. Recombination between two 14-bp homologous sequences as the mechanism for gene deletion in Factor $IX_{Seattle\ I}$. Am. J. Hum. Genet. **47:**1020–1022.
12. **Chu, G., H. Hayakawa, and P. Berg.** 1987. Electroporation for the efficient transfection of mammalian cells with DNA. Nucleic Acids Res. **15:**1311–1326.
13. **Chu, M.-L., V. Gargiulo, C. J. Williams, and F. Ramirez.** 1985. Multiexon deletion in an osteogenesis imperfecta variant with increased Type III collagen mRNA. J. Biol. Chem. **260:**691–694.
14. **De Jong, P. J., A. J. Grosovsky, and B. W. Glickman.** 1988. Spectrum of spontaneous mutation at the *APRT* locus of Chinese hamster ovary cells: an analysis at the DNA sequence level. Proc. Natl. Acad. Sci. USA **85:**3499–3503.
15. **Esumi, H., Y. Takahashi, S. Sato, S. Nagase, and T. Sugimura.** 1983. A seven-base-pair deletion in an intron of the albumin gene of analbuminemic rats. Proc. Natl. Acad. Sci. USA **80:**95–99.
16. **Figueiredo, M. S., M. H. Tavella, and B. P. Simoes.** 1994. Large DNA inversions, deletions, and *TaqI* site mutations in severe haemophilia A. Hum. Genet. **94:**473–478.
17. **Fodde, R., M. Losekoot, L. Casula, and L. F. Bernini.** 1990. Nucleotide sequence of the Belgian $^{G}\gamma^{+}(^{A}\gamma\delta\beta)^{o}$-thalassemia deletion breakpoint suggests a common mechanism for a number of such recombination events. Genomics **8:**732–735.
18. **Folger, K. R., E. A. Wong, G. M. Wahl, and M. R. Capecchi.** 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.
19. **Fuscoe, J. C., L. J. Zimmerman, K. Harrington-Brock, and M. M. Moore.** 1994. Deletion mutations in the *hprt* gene of T-lymphocytes as a biomarker for genomic rearrangements important in human cancers. Carcinogenesis **15:**1463–1466.
20. **Gilliam, A. C., A. Shen, J. E. Richards, F. R. Blattner, J. F. Mushinski, and P. W. Tucker.** 1984. Illegitimate recombination generates a class switch from Cμ to Cδ in an IgD-secreting plasmacytoma. Proc. Natl. Acad. Sci. USA **81:**4161–4168.
21. **Goldberg, S. Z., D. Kuebbing, D. Trauber, M. P. Schafer, S. E. Lewis, R. A. Popp, and W. F. Anderson.** 1986. A 66-base pair insert bridges the deletion responsible for a mouse model of β-thalassemia. J. Biol. Chem. **261:**12368–12374.
22. **Greger, V., E. Woolf, and M. Lalande.** 1993. Cloning of the breakpoints of a submicroscopic deletion in an Angelman syndrome patient. Hum. Mol. Genet. **2:**921–924.
23. **Hamada, T., H. Sasaki, R. Seki, and Y. Sakaki.** 1993. Mechanism of chromosomal integration of transgenes in microinjected mouse eggs: sequence analysis of genome-transgene and transgene-transgene junctions at two loci. Gene **128:**197–202.
24. **Hartwell, L.** 1992. Defects in a cell cycle checkpoint may be responsible for the genomic instability of cancer cells. Cell **71:**543–546.
25. **Henthorn, P. S., O. Smithies, and D. L. Mager.** 1990. Molecular analysis of deletions in the human β-globin gene cluster: deletion junctions and locations of breakpoints. Genomics **6:**226–237.
26. **Hunter, D. J., and E. G. Gurney.** 1994. The genomic instability associated with integrated simian virus 40 DNA is dependent on the origin of replication and early control region. J. Virol. **68:**787–796.
27. **Kato, S., R. A. Anderson, and R. D. Camerini-Otero.** 1986. Foreign DNA introduced by calcium phosphate is integrated into repetitive DNA elements of the mouse L genome. Mol. Cell. Biol. **6:**1787–1795.
28. **Kiba, T., H. Tsuda, C. Pairojkul, S. Inoue, T. Sugimura, and S. Hirohashi.** 1993. Mutations of the *p53* tumor suppressor gene and the *ras* gene family in intrahepatic cholangiocellular carcinomas in Japan and Thailand. Mol. Carcinog. **8:**312–318.
29. **Kootstra, A., L. K. Lew, R. S. Nairn, and M. C. MacLeod.** 1989. Preferential modification of GC boxes by benzo′a:pyrene-7,8-diol-9,10-epoxide. Mol. Carcinog. **1:**239–244.
30. **Laquerbe, A., E. Moustacchi, J. C. Fuscoe, and D. Papadopoulo.** 1995. The molecular mechanism underlying formation of deletions in Fanconi anemia cells may involve a site-specific recombination. Proc. Natl. Acad. Sci. USA **92:**831–835.
31. **Lieber, M. R.** 1992. The mechanism of V(D)J recombination: a balance of diversity, specificity, and stability. Cell **70:**873–876.
32. **Marchuk, D., M. Drumm, A. Saulino, and F. S. Collins.** 1991. Construction of T-vectors, a rapid and general system for direct cloning of unmodified PCR products. Nucleic Acids Res. **19:**1154.
33. **Mark, W. H., K. Signorelli, M. Blum, L. Kwee, and E. Lacy.** 1992. Genomic structure of the locus associated with an insertional mutation in line 4 transgenic mice. Genomics **13:**159–166.
34. **Merrihew, R. V., R. G. Sargent, and J. H. Wilson.** Efficient modification of the APRT gene by FLP/FRT site-specific targeting. Somatic Cell Mol. Genet., in press.
35. **Meuth, M.** 1989. Illegitimate recombination in mammalian cells, p. 833–860. *In* D. E. Berg and M. M. Howe (ed.), Mobile DNA. American Society for Microbiology, Washington, D.C.
36. **Miles, C., G. Sargent, G. Phear, and M. Meuth.** 1990. DNA sequence analysis of gamma radiation-induced deletions and insertions at the *APRT* locus of hamster cells. Mol. Carcinog. **3:**233–242.
37. **Molineaux, S. M., H. Engh, F. de Ferra, L. Hudson, and R. A. Lazzarini.** 1986. Recombination within the myelin basic protein gene created the dysmyelinating shiverer mouse mutation. Proc. Natl. Acad. Sci. USA **83:**7542–7546.
38. **Monnat, R. J., A. F. M. Hackman, and T. A. Chiaverotti.** 1992. Nucleotide sequence analysis of human hypoxanthine phosphoribosyltransferase (*HPRT*) gene deletions. Genomics **13:**777–787.
39. **Morris, T., and J. Thacker.** 1993. Formation of large deletions by illegitimate recombination in the *HPRT* gene of primary human fibroblasts. Proc. Natl. Acad. Sci. USA **90:**1392–1396.
40. **Mulligan, R. C., and P. Berg.** 1980. Expression of a bacterial gene in mammalian cells. Science **209:**1422–1427.
41. **Murnane, J. P., M. J. Yezzi, and B. R. Young.** 1990. Recombination events

during integration of transfected DNA into normal human cells. Nucleic Acids Res. **18:**2733–2738.
42. **Nairn, R. S., R. M. Humphrey, and G. M. Adair.** 1988. Transformation of UV-hypersensitive Chinese hamster ovary cell mutants with UV-irradiated plasmids. Int. J. Radiat. Biol. **53:**249–260.
43. **Nalbantoglu, J., D. Hartley, G. Phear, G. Tear, and M. Meuth.** 1986. Spontaneous deletion formation at the *aprt* locus of hamster cells: the presence of short sequence homologies and dyad symmetries at deletion termini. EMBO J. **5:**1199–1204.
44. **Nalbantoglu, J., C. Miles, and M. Meuth.** 1988. Insertion of unique and repetitive DNA fragments into the *aprt* locus of hamster cells. J. Mol. Biol. **200:**449–459.
45. **Nalbantoglu, J., G. Phear, and M. Meuth.** 1987. DNA sequence analysis of spontaneous mutations at the *aprt* locus of hamster cells. Mol. Cell. Biol. **7:** 1445–1449.
46. **Nalbantoglu, J., G. A. Phear, and M. Meuth.** 1986. Nucleotide sequence of the hamster adenine phosphoribosyl transferase (aprt) gene. Nucleic Acids Res. **14:**1914.
47. **Ochman, H., M. M. Medhora, D. Garza, and D. L. Hartl.** 1990. Amplification of flanking sequences by inverse PCR, p. 219–236. *In* M. A. Innis, D. H. Gelfand, J. J. Sninsky, and T. J. White (ed.), PCR protocols. Academic Press, Inc., San Diego, Calif.
48. **O'Gorman, S., D. T. Fox, and G. M. Wahl.** 1991. Recombinase-mediated gene activation and site-specific integration in mammalian cells. Science **251:** 1351–1355.
49. **Owens, J. D., Jr., F. D. Finkelman, J. D. Mountz, and J. F. Mushinski.** 1991. Nonhomologous recombination at sites within the mouse J_H-C_δ locus accompanies Cμ deletion and switch to immunoglobulin D secretion. Mol. Cell. Biol. **11:**5660–5670.
50. **Palmiter, R. D., T. M. Wilkie, H. Y. Chen, and R. L. Brinster.** 1984. Transmission distortion and mosaicism in an unusual transgenic mouse pedigree. Cell **36:**869–877.
51. **Peretz, H., N. Rosenberg, S. Usher, E. Graff, P. J. Newman, B. S. Coller, and U. Seligsohn.** 1995. Glanzmann's thrombasthenia associated with deletion-insertion and alternative splicing in the glycoprotein IIb gene. Blood **85:** 414–420.
52. **Pfeiffer, P., S. Thode, J. Hancke, and W. Vielmetter.** 1994. Mechanisms of overlap formation in nonhomologous DNA end joining. Mol. Cell. Biol. **14:** 888–895.
53. **Phillips, J. W., and W. F. Morgan.** 1994. Illegitimate recombination induced by DNA double-strand breaks in a mammalian chromosome. Mol. Cell. Biol. **14:**5794–5803.
54. **Pomykala, H. M., S. K. Bohlander, P. L. Broeker, O. I. Olopade, and M. O. Diaz.** 1994. Breakpoint junctions of chromosome 9p deletions in two human glioma cell lines. Mol. Cell. Biol. **14:**7604–7610.
55. **Porter, T., S. L. Pennington, G. M. Adair, R. S. Nairn, and J. H. Wilson.** 1990. A novel selection system for recombinational and mutational events within an intron of a eucaryotic gene. Nucleic Acids Res. **18:**5173–5180.
56. **Pulkkinen, L., A. M. Christiano, T. Airenne, H. Haakana, K. Tryggvason, and J. Uitto.** 1994. Mutations in the γ2 chain gene (LAMC2) of kalinin/laminin 5 in the junctional forms of epidermolysis bullosa. Nat. Genet. **6:** 293–297.
57. **Rassool, F. V., T. W. McKeithan, M. E. Neilly, E. van Melle, R. Espinosa, and M. M. Le Beau.** 1991. Preferential integration of marker DNA into the chromosomal fragile site at 3p14: an approach to cloning fragile sites. Proc. Natl. Acad. Sci. USA **88:**6657–6661.
58. **Robins, D. M., S. Ripley, A. S. Henderson, and R. Axel.** 1981. Transforming DNA integrates into the host chromosome. Cell **23:**29–39.
59. **Roth, D. B., X.-B. Chang, and J. H. Wilson.** 1989. Comparison of filler DNA at immune, nonimmune, and oncogenic rearrangements suggests multiple mechanisms of formation. Mol. Cell. Biol. **9:**3049–3057.
60. **Roth, D. B., T. N. Porter, and J. H. Wilson.** 1985. Mechanisms of nonhomologous recombination in mammalian cells. Mol. Cell. Biol. **5:**2599–2607.
61. **Roth, D. B., and J. H. Wilson.** 1986. Nonhomologous recombination in mammalian cells. Mol. Cell. Biol. **6:**4295–4304.
62. **Roth, D. B., and J. H. Wilson.** 1988. Illegitimate recombination in mammalian cells, p. 621–654. *In* R. Kucherlapati and G. R. Smith (ed.), Genetic recombination. American Society for Microbiology, Washington, D.C.
63. **Rouet, P., F. Smih, and M. Jasin.** 1994. Introduction of double-strand breaks into the genome of mouse cells by expression of a rare-cutting endonuclease. Mol. Cell. Biol. **14:**8096–8106.
64. **Rovinski, B., D. Munroe, J. Peacock, M. Mowat, A. Bernstein, and S. Benchimol.** 1987. Deletion of 5′-coding sequences of the cellular p53 gene in mouse erythroleukemia: a novel mechanism of oncogene regulation. Mol. Cell. Biol. **7:**847–853.
65. **Sambrook, J., E. F. Fritsch, and T. Maniatis.** 1989. Molecular cloning: a laboratory manual, 2nd ed. Cold Spring Harbor Laboratory, Cold Spring Harbor, N.Y.
66. **Sargent, R. G., G. Phear, and M. Meuth.** 1989. Deletion formation in mammalian cells: molecular analysis of breakpoints and junctions in the hamster *aprt* locus. New Biol. **1:**205–213.
67. **Schluter, G., and U. Wick.** 1994. An 87 bp deletion in exon 5 of the LDL receptor gene in a mother and her son with familial hypercholesterolemia. Clin. Genet. **45:**84–87.
68. **Shapiro, L. J., P. Yen, D. Pomerantz, E. Martin, L. Rolewic, and T. Mohandas.** 1989. Molecular studies of deletions at the human steroid sulfatase locus. Proc. Natl. Acad. Sci. USA **86:**8477–8481.
69. **Streisinger, G., Y. Okada, J. Emrich, J. Newton, A. Tsugita, E. Terzaghi, M. Inouye.** 1966. Frameshift mutations and the genetic code. Cold Spring Harbor Symp. Quant. Biol. **31:**77–84.
70. **Strobel, M. C., P. K. Seperack, N. G. Copeland, and N. A. Jenkins.** 1990. Molecular analysis of two mouse dilute locus deletion mutations: spontaneous dilute lethal^{20J} and radiation-induced dilute prenatal lethal Aa2 alleles. Mol. Cell. Biol. **10:**501–509.
71. **Sutcliffe, J. G.** 1979. Complete nucleotide sequence of the Escherichia coli plasmid pBR322. Cold Spring Harbor Symp. Quant. Biol. **43:**77–90.
72. **Thomas, J. W., B. C. Holdener, and T. Magnuson.** 1994. Sequence analysis of a radiation-induced deletion breakpoint fusion in mouse. Mamm. Genome **5:**518–519.
73. **Thomas, K. R., and M. R. Capecchi.** 1987. Site-directed mutagenesis by gene targeting in mouse embryo-derived stem cells. Cell **51:**503–512.
74. **Wiberg, F., P. Sunnerhagen, and G. Bjursell.** 1986. New, small circular DNA in transfected mammalian cells. Mol. Cell. Biol. **6:**653–662.
75. **Wichman, H. A., R. A. van den Bussche, M. J. Hamilton, and R. J. Baker.** 1992. Transposable elements and the evolution of genome organization in mammals. Genetica (Dordrecht) **86:**287–293.
76. **Wilkie, T. M., and R. D. Palmiter.** 1987. Analysis of the integrant in MyK-103 transgenic mice in which males fail to transmit the integrant. Mol. Cell. Biol. **7:**1646–1655.
77. **Woods-Samuels, P., H. H. Kazazian, and S. E. Antonarakis.** 1991. Nonhomologous recombination in the human genome: deletions in the human factor VII gene. Genomics **10:**94–101.