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

RAYMOND V. MERRIHEW, K. MARBURGER, SANDRA L. PENNINGTON, DAVID B. ROTH, and JOHN H. WILSON

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

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 cell AT3-2, which is hemizygous for APRT (1, 3). The U1S36-TG6 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-TG6 with pDR27 (uncut for DR15; EcoRI-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 confluence in two 100-mm-diameter plates for genomic DNA extraction and Southern blotting; cells at this point (about 25 total generations) were considered in...
TABLE 1. Illegitimate reconstruction of APRT

<table>
<thead>
<tr>
<th>Random integrant or native locus</th>
<th>Genomic Target sites</th>
<th>Electro-</th>
<th>APRT* colonies</th>
<th>Reconstruction frequency</th>
</tr>
</thead>
<tbody>
<tr>
<td>DR15</td>
<td>3–4</td>
<td>10</td>
<td>95</td>
<td>4.7 × 10⁻⁷</td>
</tr>
<tr>
<td>DR11</td>
<td>2</td>
<td>ND</td>
<td>1</td>
<td>2.0 × 10⁻⁷</td>
</tr>
<tr>
<td>DR33</td>
<td>4</td>
<td>ND</td>
<td>1</td>
<td>&lt;5.0 × 10⁻⁸</td>
</tr>
<tr>
<td>RKM3</td>
<td>1</td>
<td>1</td>
<td>4</td>
<td>7.5 × 10⁻⁸</td>
</tr>
<tr>
<td>RKM4</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>&lt;5.0 × 10⁻⁸</td>
</tr>
<tr>
<td>Native loci</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>T2S24</td>
<td>1</td>
<td>1</td>
<td>1*</td>
<td>&lt;5.0 × 10⁻⁸</td>
</tr>
<tr>
<td>DEPO1</td>
<td>1</td>
<td>1</td>
<td>3</td>
<td>&lt;1.7 × 10⁻⁸</td>
</tr>
<tr>
<td>DEPO2</td>
<td>1</td>
<td>1</td>
<td>3</td>
<td>&lt;1.7 × 10⁻⁸</td>
</tr>
</tbody>
</table>

* For each electroporation, 60 μg of pDR28 was digested with EcoRI, ethanol precipitated, and mixed with 2 × 10⁶ cells in electroporation buffer. Treated cells were selected the next day for APRT function in ALASA.

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%.

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

**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-TG³ 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 (DR11), two sites of integration (DR15), 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.** EcoRI-digested pDR28 was electroporated into APRT⁻ cells having 3’ APRT target sites. As an example, a randomly integrated pDR27 target (1G), 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.
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 × 10⁷ cells of each of these cell lines were electroporated with 60 μg of plasmid pDR28 (Fig. 1) that had been digested with EcoRI, 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 DR5 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 EcoRI-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-TGR. The electroporation of EcoRI-digested pDR28 into RKM4 gave no APRT+ colonies, but four electroporations into RKM3 gave six colonies, for 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 EcoRI-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-TGR. The electroporation of EcoRI-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-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-1 leaves about 200 nucleotides adjacent to the EcoRI 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 EcoRI 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 EcoRI-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 EcoRI activity could be detected after ethanol precipitation, a sample of EcoRI-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 EcoRI 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
Junctions are represented (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$^\prime$ 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$^\prime$ 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 1), we showed that each of

---

**FIG. 3.** Junction sequences with 0 to 5 nucleotides (nt) of homology. Each top line is the 5$^\prime$ 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 sequenced differences from the published APRT sequence were detected in the APRT intron 2 sequence: an extra GCC CCG after APRT1222, 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).

**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$^\prime$ 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.
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<sup>+</sup> cell lines by Southern blotting. Genomic DNAs from 24 DR15-derived cell lines and 3 RKM3-derived cell lines were digested with B<sub>cl</sub>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) and the stability of the target sites was evaluated again. The results were similar to those shown in Table 2. The stability of the target sites was also confirmed by Southern blotting of the 5′ APRT sequence as a probe for DR15 and RKM3.

**Table 2. Rearrangements at target sites**

<table>
<thead>
<tr>
<th>Cell line</th>
<th>Subclones examined</th>
<th>Rearrangements detected&lt;sup&gt;a&lt;/sup&gt;</th>
</tr>
</thead>
<tbody>
<tr>
<td>DR15sci3</td>
<td>13</td>
<td>2</td>
</tr>
<tr>
<td>DR15sci4</td>
<td>10</td>
<td>4</td>
</tr>
<tr>
<td>DR15-19</td>
<td>6</td>
<td>3</td>
</tr>
<tr>
<td>DR15-43</td>
<td>2</td>
<td></td>
</tr>
<tr>
<td>RKM3</td>
<td>8</td>
<td></td>
</tr>
</tbody>
</table>

<sup>a</sup> 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.
Each cell line was then replated to three dishes at 5 function and then allowed a 1-day recovery in HT medium.

Selection. Cell lines were preselected in HAT medium for

rearrangements detected (data not shown).

In all Southern blots probed with the 3' APRT sequence, a faint band corresponding to the undeleted portion of the native APRT fragment detected with the 3' APRT probe.

The instability of the target sites was also examined by selection. Cell lines were prescreened 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 x 10^6 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 x 10^-4 to 4 x 10^-5; cell lines RKM4 and DR33 gave frequencies of 7 x 10^-5 to 9 x 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^-7. This estimate is based on our measured rate of efficiency for stable transfection in cell line DR15 (2.8 x 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 x 10^9 kb): (2.8 x 10^-3) x (2 kb) x (1.6 x 10^6 kb) = 10^-9.

Compared with this genome-wide average, gene reconstruction in RKM3 (7.5 x 10^-6) and DR15 (4.7 x 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 x 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 x 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)
by (2)(1/4)(3/4)2. This value (0.28) was then multiplied by the total number of
stance, the probability of finding 1 nucleotide of homology at the junction is given
junctions).
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)(14)(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).

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 lev-
evels), 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,
itself a result of two recent studies, deletions that were
induced by electro-
porating restriction enzymes, and therefore are likely to have
arisen by the joining of cut ends, had junctions with 0 nu-
cleotides 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 integra-
tions 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 nucl-
etides or less). Second, the input DNA could have picked up
the inserts by end joining with extrachromosomal DNA frag-
ments prior to integration. Extrachromosomal pieces of chro-
mosomal DNA have been shown to exist in several cell types
(74), and end joining of extrachromosomal DNA is a com-
monly observed phenomenon in transfection experiments with
mammalian cells (62). Third, the input DNA could use short
term homologies to prime DNA synthesis within an access-
sible 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 con-
straints on the process. Any proposed mechanism must ac-
count 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 oc-
cur by the end joining of transfected DNA to broken chromo-
some 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 join-
ing 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 for-
eign 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, termi-
nal 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 homol-
ogies 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.

![Graph of nucleotides of homology at junction](image)

**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 \( P(x) = \frac{(x+1)(14)(3/4)^2}{(2)(14)(3/4)^2} \), 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
\( (2)(14)(3/4)^2 \). 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).

![Copy-join model for random integration events](image)

**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.
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 KKM3 cells, if the embedded simian virus 40 origin promoted oncoprotein-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 GM 38219 to J.H.W.

REFERENCES


Downloaded from http://mcb.asm.org/ on May 10, 2021 by guest


taneous 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.


64. Shapiro, L. J., P. Yen, D. Pomerantz, E. Martin, L. Rodelic, and T. Mohan


