A novel selection system for recombinational and mutational events within an intron of a eucaryotic gene

Thomas Porter, Sandra L.Pennington, Gerald M.Adair[1] Rodney S.Nairn[1] and John H.Wilson*
Verna and Marrs McLean Department of Biochemistry, Baylor College of Medicine, Houston, Texas 77030 and [1]The University of Texas M.D. Anderson Cancer Center, Science Park-Research Division, Smithville, TX 78957, USA

Received May 15, 1990; Revised and Accepted July 31, 1990

ABSTRACT

In order to identify a poison sequence that might be useful in studying illegitimate recombination of mammalian cell chromosomes, several DNA segments were tested for their ability to interfere with gene expression when placed in an intron. A tRNA gene and its flanking sequences (267 bp) were shown to inhibit SV40 plaque formation 100-fold, when inserted into the intron in the T-antigen gene. Similarly, when the same DNA segment was placed in the second intron of the adenosine phosphoribosyl transferase (APRT) gene from CHO cells, it inhibited transformation of APRT$^-$ CHO cells 500-fold. These two tests indicated that the 267-bp DNA segment contained a poison sequence. The poison sequence did not affect replication since the replication of poisoned SV40 genomes was complemented by viable SV40 genomes and poisoned APRT genes were stably integrated into cell chromosomes. Cleavage of the poison sequence in the SV40 T-antigen intron by restriction enzymes indicated that the tRNA structural sequences and the 5′ flanking sequences were not required for inhibition of SV40 plaque formation. Sequence analysis of viable mutant SV40, which arose after transfection of poisoned genomes, localized the poison sequence to a 35 bp segment immediately 3′ of the tRNA structural sequences.

INTRODUCTION

Genome rearrangements in mammalian cells comprise a diverse collection of genetic alterations, the majority of which—including deletions, inversions, and translocations—result from events that are neither site-specific nor homology-based. These rearrangements depend on little if any nucleotide sequence homology and are not under direct genetic control; they are commonly lumped together under the heading of illegitimate (or nonhomologous) recombination (1,2). Many such rearrangements arise by the joining of DNA ends, which result from errors of DNA metabolism, that is, as mistakes in replication, repair, recombination, and transcription (2–4). The efficient end-joining capabilities of vertebrate cells may represent a general defense mechanism in mammalian cells for dealing with broken chromosomes (5).

In previous studies we used the intron of the animal virus SV40 to study the end-joining step of illegitimate recombination because the intron is largely nonessential and no particular sequence must be recreated by the end-joining event (6–10). The advantages of using an intron as the site for studying illegitimate recombination led us to search for an analogous system in which to study illegitimate recombination on the chromosome. Our goal was to identify a segment of DNA that, when inserted into an intron in a selectable gene, would prevent expression of the gene. Such a 'poison' sequence would allow us to select for mutations that remove or modify the poison sequence, thus allowing gene expression. If such a poison sequence could be identified, then other DNA segments—potential hotspots for illegitimate recombination—could be placed adjacent to the poison sequence to test their influence on the frequency of deletion.

In this paper we describe the testing of several short DNA segments that carry transcription signals, which we thought might interfere with gene expression. Our strategy was to insert candidate sequences into the introns of the SV40 T-antigen gene or the Chinese hamster adenine phosphoribosyl transferase (APRT) gene to determine whether they adversely affected gene expression. We screened two polyadenylation signals, a pair of splicing signals, a polIII transcription unit in both orientations, and an internal antisense RNA. A 267 bp fragment containing a polIII transcription unit (X. laevis tyrosine tRNA^{sup} gene) was found to prevent gene expression. This fragment reduced plaque formation by SV40 when inserted into the intron in T-antigen gene; it also blocked expression of APRT when inserted into the second intron of the gene. Analysis of SV40 mutants that regained their ability to form plaques by alterations of the inserted DNA segment showed that the 35 bp immediately to the 3′ side of the tRNA coding sequence was critical for preventing gene expression. Whether the poison sequence interferes with gene expression by virtue of its polIII transcription signals or because of some other feature of the sequence is not yet known.

* To whom correspondence should be addressed

MATERIALS AND METHODS

Cell Lines, Viruses and DNAs

The established monkey kidney cell line, CV1, was grown according to standard procedures (6). The construction of the substitution mutant, su1903, was described previously (9). SV40 and plasmid DNA preparations, DNA transfections, plaque assays, and viral infections were carried out according to established procedures as described previously (6,11).

The CHO cell line U1S36 is a spontaneous APRT-deficient mutant derived from cell line UVL-1 (12). The mutation in U1S36 removed the entire structural gene for APRT (12). An HPRT-deficient derivative of this cell line (U1S36-TG^R), which was selected by exposure to thioguanine, was used in the present studies. Cells were grown and maintained as described previously (13,14). Selections for APRT$^+$ transformants and APRT$^-$ revertants have been described previously (12,13). HAT selection for bacterial gpt was as described previously (15).

Construction of Substrate Plasmids

Plasmids pSU1950 and pSU1951 (Figure 1) were constructed by inserting a 267 bp fragment in opposite orientations into pSu1903, a derivative of SV40 that contains a unique BglII site within the intron of the large-T antigen gene. The 267 bp HhaI fragment contains an amber suppressor tyrosine tRNA gene surrounded by short flanking sequences (67 bp on the 5′ side and 120 bp on the 3′ side). This fragment was subcloned from M13-33-3-Su$^+$ (16). The orientation and integrity of the inserts were confirmed by nucleotide sequencing. pSU1950 contains the transcribed strand of the tRNA gene on the SV40 early message coding strand. pSU1951 contains the 267 bp fragment in the opposite orientation.

Plasmids pAK1 and pAK3 (Figure 1) contain the 267 bp fragment inserted into the EcoRI site within the second intron of the APRT gene. To render the EcoRI site within the APRT intron unique, the 3.9 kb BamHI fragment containing the APRT gene was inserted into a modified pUC18 plasmid from which the EcoRI site in the polylinker had been removed by filling in the ends and religating. This treatment yielded pUCAPRT. The 267 bp fragment was removed from an intermediate vector and modified by addition of EcoRI linkers. The modified fragment was then inserted into the unique EcoRI site in the second intron of the APRT gene in pUCAPRT to give pAK1 and pAK3. pAK3 contains the insert oriented so that transcription of the tRNA gene is in the same direction as the APRT gene; pAK1 contains the tRNA gene in the opposite orientation.

Plasmids pSU1913E and pSU1913L were constructed by inserting the 237 bp BamHI/BclI fragment from SV40, which contains the early and late polyadenylation sites, into the unique BglII site in the intron of pSU1903. The insert is oriented in pSU1913E to test the SV40 early polyadenylation site; in pSU1913L it is oriented to test the SV40 late polyadenylation site.

Plasmids pADIR and pAINV were derived from the parental plasmid pGAL (17), which was digested with SmaI and EcoRV to liberate a 608 bp blunt-ended fragment containing the entire first exon and intron and a few nucleotides of the second exon. EcoRI linkers were added to the fragment and it was cloned directly into the unique EcoRI site of pUCAPRT. pADIR carries the insert as a direct repeat; pAINV carries the insert as an inverted repeat.

Plasmid pAGK1 was constructed by replacing the 3-kb SacI-EagI fragment of the APRT gene in pAG1 with the corresponding fragment from pAK1. pAGK1 comprises a complete APRT gene, which carries the tRNA gene-containing segment inserted in the second intron, and most of the sequences from pSV2gpt (18). pAG1 was derived from pSV2gpt in two steps. First, the 0.7-kb EcoRI-BamHI segment of pSV2gpt was replaced with a 2.6-kb EcoRI-BamHI fragment that was derived from pHaprt-1 (19) and contained the 3′ end of the APRT gene (15). Next, the complete APRT gene was reconstructed by inserting a 1.3-kb EcoRI fragment that contained the 5′ end of the gene into the unique EcoRI site in pAG7 (15). The 1.3-kb EcoRI fragment was derived from pGAL (17).

Transformation Assays

Calcium-phosphate transformations were performed according to standard procedures (20). Fourteen days after transformation, colonies were fixed and stained with methanol-acetic acid (90:10 v/v) plus 0.4% (w/v) methylene blue, and counted.

pAGK1 was introduced into U1S36-TG^R cells using a Promega Model 450 Electroporation System. A 1-ml cuvette containing 3.2×10^6 cells and 0.7 ug of pAGK1 DNA in HEBS buffer (20 mM HEPES, pH 7.05, 137 mM NaCl, 5 mM KCl, 0.7 mM Na_2HPO_4, 6 mM dextrose) was electroporated at 100 uF and 800 V for 100 msec. Cells were replated in normal medium and HAT selection was applied after 24 hours.

Mini-well Analysis and Nucleotide Sequencing

Analysis of individual plaques was determined by a mini-well labeling procedure followed by restriction digestion as described previously (21). DNA sequencing was performed by a modification of the Zagurski double-strand sequencing method (22), as previously described (9).

RESULTS

Strategy for Identifying a Poison Sequence and Initial Testing of Candidate DNA Segments

To screen potential poison sequences, we used introns in two different genes—the SV40 T-antigen gene and the CHO APRT gene. One viral gene and one cellular gene were used to try to identify a sequence that interfered with gene expression in a general rather than gene-specific fashion. The SV40 T-antigen gene and the CHO APRT gene offer particular advantages for identifying and analyzing poison sequences.

The intron of the SV40 large T-antigen gene has several features that recommend it. (i) The T-antigen intron can accommodate a variety of deletions and insertions with little effect on gene expression (9,10). Thus, any sequence that inhibits gene expression is likely to do so by a mechanism with general applicability. (ii) T-antigen is required for viral replication and plaque formation (23). Interference with T-antigen expression will result in a replication-defective virus that will be unable to form plaques. (iii) Once an interfering DNA segment is identified, mutations that disrupt the poison sequence can be isolated by selecting for virus that have regained the ability to form plaques. (iv) Viable viral mutants, which have lost a critical portion of the interfering DNA segment, can be readily amplified and sequenced to identify the nucleotides responsible for the interfering phenotype.

The CHO APRT gene is a cellular gene with a comparable set of advantages. (i) The natural gene is contained on a readily manipulable 3.9-kb BamHI fragment that has been cloned and sequenced (19,24). The gene contains a unique EcoRI site near the middle of the 1-kb-long second intron, which constitutes a

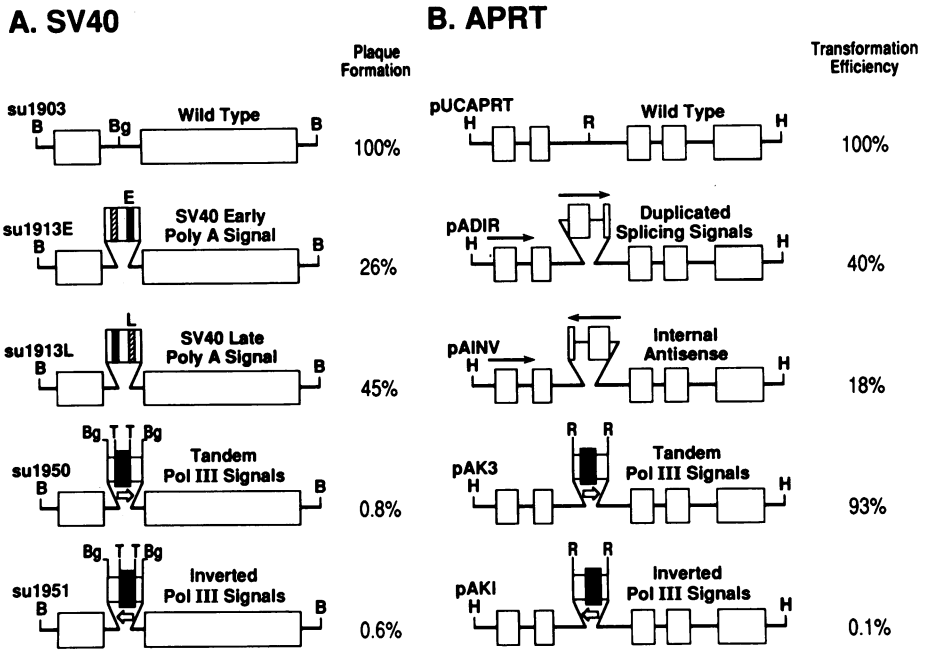

FIGURE 1. Structures of SV40 and APRT test substrates and their efficiencies of plaque formation and transformation. (A) SV40 substrates. The SV40 substrates were all derived from the parent virus su1903 by inserting DNA segments into the unique BglII (Bg) site in the T-antigen intron as described in Materials and Methods. Only the T-antigen gene in the SV40 genomes is shown to scale; the two open rectangles represent the T-antigen exons. Candidate test sequences are shown above the T-antigen intron. In su1913E and su1913L the inserted DNA contains the SV40 early (E) and late (L) polyadenylation sites, which are represented by solid and hatched bars respectively. In su1950 and su1951 the inserted DNA contains a polIII transcription unit. The location of the structural sequence for the tRNA gene is shown by the solid bar; the direction of polIII transcription is shown by the arrow. Relevant restriction sites are indicated: BglII (Bg), BamHI (B), TaqI (T). Efficiencies of plaque formation relative to su1903 are indicated at the right. Prior to transfection all substrates were cleaved from their plasmid backbones by digestion with BamHI, which liberates a linear SV40 genome. Numbers are the average of three or more transfections. (B) APRT substrates. The APRT substrates were all derived from the parent plasmid pUCAPRT by inserting DNA segments into the unique EcoRI (R) site in the second intron in the APRT gene. Only the APRT gene is shown; the five open rectangles represent the APRT exons. Candidate test sequences are shown above the second intron. In pADIR and pAINV the inserted DNA is derived from the 5′ end of the APRT gene; arrows show the orientation of the inserted segment relative to its normal orientation at the 5′ end of the gene. In pAK3 and pAK1 the inserted DNA contains the polIII transcription unit. The location of the structural sequence for the tRNA gene is shown by the solid bar; the direction of polIII transcription is shown by the arrow. Relevant restriction sites are indicated: HindIII (H), EcoRI (R). Efficiencies of transformation relative to pUCAPRT are indicated at the right. Prior to transformation all plasmids were linearized by digestion with HindIII, which cleaves the plasmid once. Numbers are the average of two or more transformations.

convenient site for testing potential poison sequences. (ii) There are defined selection schemes for and against APRT function (12,13). Thus, the ability of a DNA segment to interfere with APRT expression can be readily assessed by transforming the modified gene into APRT-deficient cells and selecting for APRT function. (iii) Once an interfering DNA segment is identified, mutations that disrupt the poison sequence can be isolated by depositing the modified gene in a chromosome and then selecting for APRT function.

Several candidate sequences were screened to find one that interfered with expression of both genes (Figure 1). The SV40 T-antigen gene was used to test polyadenylation signals and signals associated with polIII transcription. A 237-bp fragment carrying the SV40 early and late polyadenylation signals was inserted into the T-antigen intron in both orientations to test the two polyadenylation signals. The results of several transfections indicated that these polyadenylation signals reduced plaque formation only twofold to fourfold relative to wild-type virus (Figure 1). By contrast, a 267-bp fragment, which contained a polIII transcription unit for a X. laevis tyrosine tRNA gene, was shown to reduce plaque formation about 100-fold when inserted in either orientation into the T-antigen intron (Figure 1). As shown below, the plaques that arose after transfection with su1950 and su1951 carried virus with mutations in the T-antigen intron.

The APRT gene was used to test the effects of spurious splicing signals, internal antisense RNA, and polIII transcription signals. A 608 bp fragment that included the entire first exon and first intron of the APRT gene was inserted into the second intron as a direct repeat to test the effects of duplicated splicing signals. The ability of this construct to transform APRT-negative CHO cells to the APRT-positive phenotype was reduced only about twofold by the presence of these extra splicing signals (Figure 1). The identical 608 bp fragment was inserted into the second intron in the opposite orientation to test the effect of internal antisense RNA on the expression of the APRT gene. Transformation to the APRT-positive phenotype was reduced only about fivefold by these internally complementary sequences (Figure 1). The polIII transcription unit tested in the T-antigen intron was inserted into the APRT second intron to examine its effect on APRT expression. When the polIII unit was oriented so that its transcription was in the same direction as the APRT gene, there was no effect on transformation to the APRT-positive phenotype (Figure 1). However, when the polIII signals were in the opposite orientation, transformation to the APRT-positive phenotype was reduced more than 500-fold (Figure 1).

Since the DNA segment containing the polIII transcription unit interfered with expression of both test genes, it was analyzed in more detail as described in the following sections.

Table 1. Plaque Formation by SV40 Genomes

SV40 Genome	BamHI Linears[a]	Circles[a]	BglII Linears[a]
su1903	100%	100%	100%
su1950	0.8%	1.1%	64%
su1951	0.6%	2.4%	85%

[a]Data are the average of three or more transfections. Specific transfection efficiencies of the various control molecules were as follows: su1903 (BamHI linears) 138 plaques/ng; su1903 (circles) 541 plaques/ng; su1903 (BglII linears) 167 plaques/ng. Total number of control plaques counted: su1903 (BamHI linears) 3856 plaques; su1903 (circles) 3701 plaques; su1903 (BglII linears) 1954 plaques.

Table 2. Transformation Frequencies of APRT Substrates

Substrate	HindIII Linears[a]	EcoRI Linears[a]	Modified HindIII Linears[a]
pUCAPRT	100%	100%	100%
pAK3	93%	117%	125%
pAK1	0.13%	57%	52%

[a]Data are the average of three transformations with the exception of the modified HindIII linears, which are the average of two transformations. Specific transformation efficiencies for the various control substrates were as follows: pUCAPRT (HindIII linears) 24 transformants/ng; pUCAPRT (EcoRI linears) 1.4 transformants/ng; pUCAPRT (modified HindIII linears) 31 transformants/ng. Total number of control colonies counted: pUCAPRT (HindIII linears) 4400 colonies; pUCAPRT (EcoRI linears) 345 colonies; pUCAPRT (modified HindIII linears) 455 colonies.

Removal of the Inserted DNA Segment Restores Activity

To ensure that the apparent reduction in gene expression was due to the inserted segment and not, for example, the result of inadvertant damage to the SV40 and APRT backbones during cloning manipulations, the segment was removed and the activity of the construct was reassayed. For the SV40 constructs, the viral genomes were removed from the plasmid by digestion with BamHI and the resulting linear molecules were circularized. The circles were then digested with BglII, which cleaves on either side of the inserted segment and removes it from the SV40 sequences. The BamHI linears, circles, and BglII linears were transfected into CV1 cells and assayed for plaque formation. As shown in Table 1, the BamHI linears and the circles, which retained the inserted segment, formed plaques with low efficiency, whereas the BglII linears, from which the segment was removed, formed plaques with an efficiency comparable to the wild type control.

Analogous experiments with the APRT constructs also indicated that the apparent reduction in gene expression was due to the inserted segment. The inserted DNA segment was removed from the APRT constructs by digestion with EcoRI, which cleaves on either side of the insert. The resulting molecules contained a double-strand break within the second intron. As a result, they transformed cells to the APRT-positive phenotype only about 5% as efficiently as linear molecules with an intact APRT gene (see legend, Table 2). Nevertheless, molecules from which the insert was removed transformed cells with the same efficiency as control molecules in which the normal APRT gene was cleaved with EcoRI (Table 2). To confirm these results, the EcoRI linears were circularized and cleaved with HindIII, which liberates the APRT gene from the plasmid backbone. As shown in Table 2, these modified HindIII linears transformed cells with an efficiency comparable to identically treated control plasmids.

Since removal of the inserted DNA segment from the SV40 or APRT backbones restored activity to wild type levels, the interfering activity must be contained within the 267 bp segment of foreign DNA.

The Inserted DNA Segment Does Not Block Replication

A DNA sequence from the plasmid pBR322 was previously shown to inhibit SV40 replication when present in the same molecule (25–27). We did not anticipate that the DNA segment containing the tRNA gene would interfere with replication because of earlier results in which the identical DNA segment was inserted into the late region of SV40 to test the capacity of the suppressor tRNA gene to suppress amber mutations in trans (16). Stocks of this modified virus were readily prepared by complementation with SV40 virus that were defective in the early region, thereby indicating that the DNA segment does not block SV40 replication in *cis* or in *trans* (16).

Analysis of the rare plaques that arose after transfection are consistent with the above observations. Six of eight analyzed plaques consisted of obvious mixtures of genomes that differed in the length of the restriction fragment that carried the intron (Figure 2A). Mixtures R3 to R8 each have a variable set of bands that correspond to fragments with deletions (bracket) and a constant band that corresponds to the fragment carrying the inserted DNA (closed arrowhead) against a uniform background of wild type bands (open arrowheads). The mixed nature of these plaques was confirmed by replating mixtures R3, R4, and R7 at high dilution and isolating multiple plaques from each. Fourteen plaques were analyzed: none were mixed and none contained the 1211 bp fragment characteristic of the input DNA (Figure 2B). The intron-containing bands (bracket) varied among the plaques; however, in every case the band in an isolated plaque appeared to correspond to a visible band in the original mixture. The prominence of the 1211 bp band in the mixed plaques and its absence from the purified plaques suggest that the input DNA can replicate in the presence of viable virus but cannot replicate on its own. Thus, consistent with previous findings (16), the inserted DNA segment does not appear to block the replication of an SV40 genome that contains it.

A lethal block to replication on the chromosome could also account for the rarity of APRT-positive transformants from pAK1 transfection (Table 1). To test this possibility, we attempted to integrate the APRT gene carrying the poison sequence into a cellular chromosome in the absence of APRT selection. If the poison sequence blocked chromosome replication, it should not be possible to incorporate the modified APRT gene into the cell genome. To test this expectation, we transferred the poisoned APRT gene to pSV2gpt to generate pAGK1, electroporated the plasmid DNA into U1S36-TG^R cells, which are APRT$^-$ and HPRT$^-$, and selected for HAT resistant (gpt$^+$) colonies. Individual colonies were screened by Southern blotting and four independent sublines with single copies of the plasmid were analyzed in more detail. All four sublines were initially APRT-negative. Upon selection for APRT function, however, three sublines yielded APRT$^+$ cells at frequencies of 1 to 5×10^{-5}. (The mutations that generated these APRT$^+$ cells are currently being analyzed and will be described elsewhere.) The fourth subline yielded no detectable APRT$^+$ cells ($<10^{-7}$) and may have lost essential APRT information during transformation. The ability of three APRT$^-$ sublines to revert to the APRT$^+$ phenotype indicated that the modified APRT gene could be stably integrated in a chromosome and, therefore, that the poison sequence was unlikely to block chromosome replication.

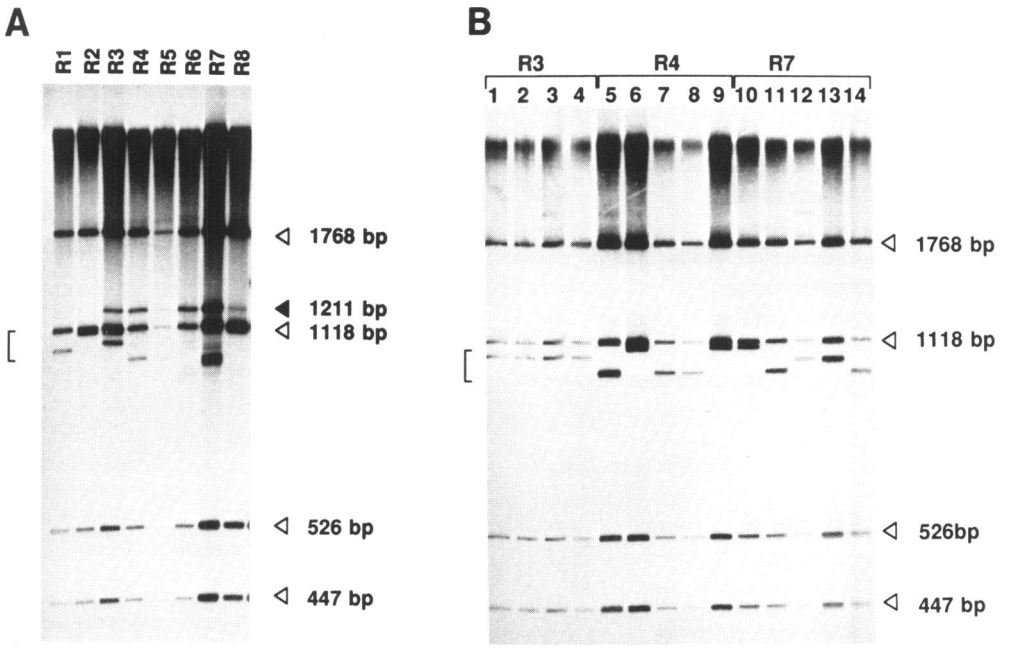

FIGURE 2. Restriction analysis of plaques that arose after transfection with poisoned SV40 genomes. (A) Mixed plaques. Eight plaques (R1 through R8), which formed after transfection with BamHI-linearized su1950, were isolated and amplified in the presence of ^{32}P-orthophosphate. Radiolabeled viral DNA was isolated, digested with HindIII, separated by electrophoresis through a 5% nondenaturing polyacrylamide gel, and visualized by autoradiography. Bands at positions indicated by open arrowheads are wild type SV40 HindIII fragments. The band indicated by the closed arrowhead migrates at the position of the HindIII fragment that spans the intron and contains the intact inserted DNA segment. The variety of bands encompassed by the bracket correspond to genomes that suffered deletions in the intron. (B) Pure plaques. Mixed plaques R3, R4, and R7 were diluted and replaqued on CV1 cells. Several individual plaques were isolated from each mixture, amplified in the presence of ^{32}P-orthophosphate, and treated as in part A. Symbols are as in part A. Each lane contains a single band corresponding to the intron fragment, indicating that each plaque contains a single predominant type of genome.

Restriction Analysis of the Poison Sequence in SV40 Genomes

The importance of the tRNA sequences in the DNA segment was tested by digesting circular su1950 and su1951 DNA with TaqI, which removes all of the tRNA coding sequences except for four bases at the 5′ end and seven bases at the 3′ end. Plaque formation by these TaqI linears of su1950 and su1951 were reduced 50-fold and 30-fold respectively, relative to circular su1930 genomes that had been linearized in the intron by BglII digestion. These results suggested that one or both of the flanking sequences, rather than the tRNA structural sequences, contained the inhibitory element.

The flanking sequences to the 5′ and 3′ sides of the tRNA gene were tested for interfering activity by cleaving BamHI-linearized su1951 with TaqI, which separated the flanking sequences onto two half-genome fragments. These fragments were purified from agarose gels, as were analogous wild type fragments generated by cleavage of wild type SV40 with BamHI and TaqI (which cleaves once in the T-antigen intron). Wild type and mutant genome fragments were mixed in appropriate combinations and transfected into CV1 cells to assay for plaque formation after intracellular ligation of the fragments, which is very efficient (28). Mixtures containing the 5′ flank gave 60% the number of plaques of the control mixture containing the wild type fragments, whereas mixtures containing the 3′ flank gave only 2% the number of plaques of the control mixture. SV40 genomes from two plaques that arose in mixtures containing the 5′ flanking sequences were isolated and their intron regions were sequenced. In both molecules the wild type and mutant fragments were linked at the TaqI site, as expected, and no sequence alterations were detected in the 5′ flanking region or in the surrounding intron sequences.

Table 3. Types of Mutations Among Viable SV40 Genomes

Input Genome	Total Plaques	Deletions	Insertions	Point Mutations
su1950 (BamHI Linears)	8	8	0	0
su1951 (BamHI Linears)	3	3	0	0
su1951 (TaqI Linears)	10	5[a]	2	4[a]

[a]One viable genome, the complex mutant pmdl1996, contains a deletion and a point mutation.

The viability of genomes containing only the 5′ flanking region and the inviability of genomes lacking the tRNA sequences indicate that the interfering activity is confined to the region of the DNA insert that flanks the 3′ end of the tRNA gene.

Mutational Analysis of the Poison Sequence in SV40 Genomes

To define the boundaries of the poison sequence more precisely, 21 viable SV40 genomes, which arose after transfection with BamHI and TaqI linears of su1950 and su1951, were amplified and sequenced in the region of the intron. (All of the sequenced DNAs were isolated from plaques that had been transfected at high dilution in order to ensure that they contained a single genome.) Twenty genomes contained a deletion, insertion, or point mutation in the intron and one genome carried both a deletion and a point mutation (Table 3).

In Figure 3, the 16 genomes that contained deletions are diagrammed. Deletions dl1991 and dl1993 from su1951 define a common overlap region beginning at the TaqI restriction site and extending 36 nucleotides into the 3′ flanking region. The

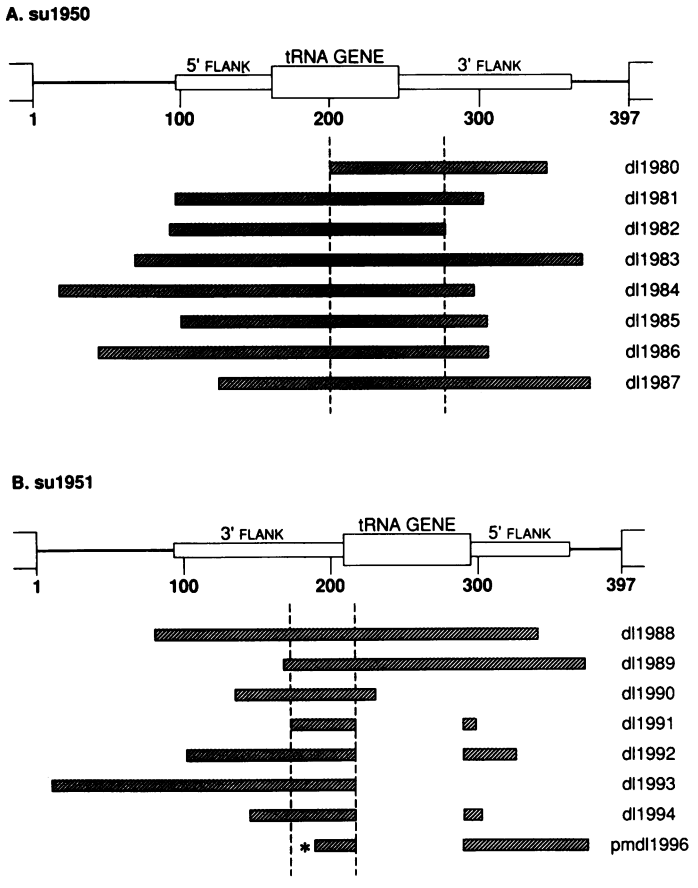

FIGURE 3. Locations of deletion mutations in the viable viruses that arose after transfection of poisoned SV40 genomes. (A) Deletions derived from su1950. All plaques were isolated after transfection with BamHI-linearized su1950. The position of the inserted DNA in the T-antigen intron is shown to scale and the positions of the tRNA structural sequences and 5′ and 3′ flanking sequences are indicated. Numbers count nucleotides from the 5′ splice site. Each deletion is represented by a horizontal solid bar. The common overlap region of the deletions is indicated by vertical dashed lines. (B) Deletions derived from su1951. Deletions dl1988, dl1989, and dl1990 arose after transfection of BamHI-linearized su1951; the rest of the deletions arose after transfection of TaqI-linearized su1951. Each deletion is represented by a solid bar; the gaps in deletions from TaqI-linearized genomes are present because the DNA between the TaqI sites (T) were not in the transfected genomes. The common overlap region is shown by the vertical dashed lines. The star in pmdl1996 indicates the position of the point mutation present in this double mutant.

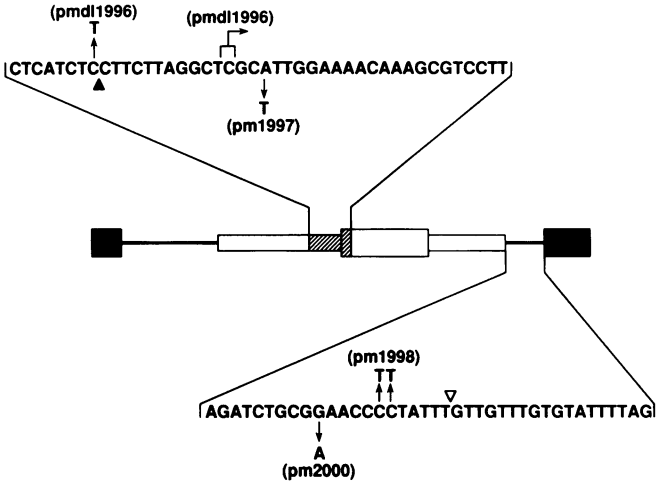

FIGURE 4. Location of point mutations that arose after transfection of TaqI-linearized su1951. su1951 is shown as in Figure 3B. The 43 nucleotides (corresponding to the transcribed RNA) from the TaqI site to the endpoint of dl1991 is shown above the inserted DNA. The endpoint of dl1982 (the most tRNA-proximal deletion generated after transfection with su1950) is indicated by the filled triangle. The nucleotide change in pm1997 is shown below the sequence; the deletion (arrow) and base change in pmdl1996 are shown above the sequence. The sequence of the intron from the end of the inserted DNA to the 3′ splice site is shown below the inserted DNA. The base change in pm2000 is shown below the sequence; the base changes in pm1998 are shown above the sequence. The junction of SV40 and pBR322 sequences that arose in the derivation of the parental virus su1903 are indicated by the open triangle.

sequence of these 43 nucleotides is shown in Figure 4. Both mutations in the complex mutant, pmdl1996, affected sequences within this interval, but were not used to delimit the interval further since it was not clear which mutation was responsible for inactivating the poison sequence. Deletions dl1980 and dl1982 from su1950 define a common overlap region beginning in the middle of the tRNA gene and extending 28 nucleotides into the 3′ flanking sequences. The endpoint of dl1982 is indicated by the closed triangle in Figure 4. Collectively, the deletions identify a 35 base pair sequence that includes 7 nucleotides of the tRNA gene and 28 nucleotides of the 3′ flanking region. A critical portion of the poison sequence presumably lies within these 35 nucleotides.

In addition to the point mutation in pmdl1996, three other mutants with nucleotide alterations were isolated. In pm1997 an A to T transversion inactivated the poison sequence; the change was located within the 35 nucleotides defined by the deletions (Figure 4). The other mutants, pm1998 and pm2000, were unexpected; they contained unaltered 5′ and 3′ flanking regions that were joined at the TaqI site. However, both genomes carried point mutations in a seven-nucleotide stretch around the branch point for T-antigen splicing (Figure 4). pm1998 contains two adjacent C to T transversion mutations; pm2000 contains a G to A transition mutation. These mutations, which lie outside the inserted DNA fragment, are able, in some fashion, to suppress the activity of the poison sequence. In the case of pm1998 the suppression is only partially effective, since plaque formation (expressed as plaques/ng) was reduced threefold to fourfold and delayed 48 hours relative to DNA from the parent virus (su1903) and from two deletion mutants (dl1992 and dl1993).

Two genomes, in1995 and in1999, carried insertions of about 120 bp and 230 bp, respectively, precisely between the input TaqI ends. Neither sequence is remarkable, although in1995 has 74 base pairs of alternating purines and pyrimidines. Neither insertion corresponds to sequences in the SV40 genome. It is not clear in either case how insertion reverses the effect of the adjacent poison sequence.

DISCUSSION

Our goal in these studies was to identify a segment of DNA that can prevent expression of a selectable gene when inserted into an intron. Such a poisoned gene would be useful in studies of illegitimate recombination on the chromosome for two reasons. It would allow selection of deletions whose endpoints are confined to a restricted region in a eucaryotic gene (an intron) and it would provide a test system to screen other DNA segments—potential hotspots for illegitimate recombination—for their influence on the frequency of deletion formation. In a survey of candidate sequences, we identified a short DNA fragment that inactivated

the large T-antigen gene of SV40 and the hamster APRT gene when placed within their introns.

Detailed analysis of viable SV40 mutants that arose after transfection of poisoned genomes into CV1 monkey cells supports the validity of this approach for studying illegitimate recombination. Of 22 mutations that restored T-antigen expression, 16 were deletions. The characteristics of these deletions were entirely consistent with our previous analyses of illegitimate recombination in the T-antigen intron (8,9,10). The deletion endpoints were not clustered and homology at the deletion junctions ranged from 0 to 4 nucleotides (9 genomes had 0 nucleotides of homology; 4 genomes had 1 nucleotide; 1 genome each had 2, 3, and 4 nucleotides of homology). Curiously, all 11 mutations isolated after transfection of BamHI linears were deletions, whereas 4 of 11 mutations identified after transfection of TaqI linears were point mutations. Although the sample size is small, this observation raises the possibility that point mutations are more frequent near input ends since the TaqI ends are adjacent to the poison sequence, while the BamHI ends are 2 kb away.

The poisoned APRT gene can be used to study illegitimate recombination of mammalian chromosomes in a way analogous to our extrachromosomal studies using SV40. We have established three APRT-negative cell lines that each carry a single copy of the poisoned APRT gene and revert to the APRT-positive phenotype at frequencies of 1 to 5×10^{-5}. The mutations that generated these APRT$^+$ cells are currently being analyzed and will be described elsewhere. Although these cell lines carry the poisoned APRT gene at random sites, it should also be possible to target the poison sequence to the endogenous APRT gene using homologous recombination (15).

We do not know how the poison sequence interferes with gene function or, for that matter, that the poison sequence interferes in the same way with the T-antigen gene and the APRT gene. Because a poisoned SV40 genome can replicate in the presence of viable genomes and a poisoned APRT gene can be stably integrated on the chromosome, the poison sequence presumably does not act by blocking replication of attached DNA sequences, but rather, by interfering in some way with gene expression. In addition, we know that the poison sequence can interfere with the adenovirus life cycle. When the poison sequence is placed in an inessential region of the adenovirus genome, downstream from the major late promoter, it severely reduces the viability of the resulting virus (Young and Porter, unpublished results).

Although we selected this DNA segment initially because it contained a complete tRNA gene, we have no evidence that polIII transcription or tRNA processing is involved in any way in the interference phenomenon. We showed that interference by the DNA segment in the T-antigen intron was independent of the tRNA structural gene and 5′ flanking sequences. Also, the inserted DNA segment interferes with APRT gene expression in a transcriptional orientation opposite that of the APRT gene, thereby eliminating tRNA processing (but not competing transcription) as the basis for interference with the APRT gene.

Restriction and mutation analysis of the poison sequence in the T-antigen intron implicated a 35 bp sequence at the 3′ end of the tRNA structural gene as a critical component of the poison sequence. Computer analysis of this sequence indicates that it contains no significant secondary structures and that there is no significant base pairing between the 35 bp sequence and other intron sequences. However, of 21 viable mutants of SV40 that were analyzed in detail, 17 carried mutations within this 35 bp sequence. The remaining 4 mutations did not alter the sequence of the inserted DNA fragment. Two mutants contained point mutations near the putative branch point for T-antigen pre-mRNA splicing. We do not know the precise branch point because the normal SV40 branch point was replaced with nucleotides from pBR322 during formation of the parent virus su1903 (9). Although su1903 forms plaques with the same efficiency as wild type (9), it is possible that these outside mutations improve the efficiency of splicing so that it can compete more effectively with whatever improper event is promoted by the poison sequence. The final two mutants carried insertions of about 120 and 230 base pairs at the TaqI restriction site that borders the poison sequence. Whether these insertions interfere with the poison sequence by virtue of their sequences, their effects on secondary structure, or for some other reason is not known. The presence of second-site revertants in these studies raises the possibility that such poisoned constructs might be useful for identifying *cis* or *trans* functions that affect RNA processing.

The other candidate sequences tested in this study all had relatively minor effects on the expression of the genes into which they were inserted. The early and late SV40 polyadenylation sites, which were inserted into the T-antigen intron, had little effect on viral viability (Figure 1) and did not cause detectable polyadenylation in the intron (29). More detailed studies indicate that polyadenylation may not occur in an otherwise normal intron (30,31). A functional pair of 5′ and 3′ splicing signals, which were arranged to define a new intron, had little effect on transformation efficiency when inserted into the second intron in the APRT gene (Figure 1). If exon definition precedes intron definition (32), it might be expected that reversing the order of the added splicing signals on the inserted DNA—that is, inserting a functional exon—may be a more effective way to interfere with gene expression. A 0.6 kb DNA fragment, which contained sequences extending from the 5′ upstream region of the APRT gene through the second exon of the APRT gene, was inserted into the second APRT intron in a reversed orientation so that the pre-mRNA would contain an extensive region of internal antisense RNA. This construct was reduced in transformation efficiency only about fivefold relative to the normal APRT gene. Although these candidate DNA segments did not interfere with gene expression sufficiently to be useful for our purposes, they did serve as important controls: the simple insertion of heterologous sequence information into introns in our test genes had little or no effect on gene expression.

ACKNOWLEDGMENTS

We thank Drs Susan Berget, C.S. Hamisch Young, and Xiu-Bao Chang for critical comments on the manuscript. De Dieu and Kathleen Marburger provided excellent technical assistance. We thank Dr Phil Sharp for his gift of tRNA clones. Finally, we thank David Roth, Neal Proctor, Hui Zheng, and the Baylor Molecular Biology Group for advice. This work was supported by grants from the National Institutes of Health (GM38219 and CA28711) from the American Cancer Society (CD-420) and from the Robert A. Welch Foundation (Q-977).

REFERENCES

1. Meuth,M. (1989) In Berg,D.E. and Howe,M.M. (eds.), Mobile DNA. American Society for Microbiology, Washington, pp. 833–860.
2. Roth,D.B. and Wilson,J.H. (1988) In Kucherlapati,R. and Smith,G.R. (eds.),

Genetic Recombination. American Society for Microbiology, Washington, pp. 621–653.
3. Franklin,N. (1971) In Hershey,A.D. (ed.), The Bacteriophage Lambda. Cold Spring Harbor Laboratory, Cold Spring Harbor, pp. 175–194.
4. Michel,B., and Ehrlich,S.D. (1986) EMBO J. 5, 3691–3696.
5. Wake,C.T., Vernaleone,F. and Wilson,J.H. (1985) Mol. Cell. Biol. 5, 2080–2089.
6. Wilson,J.H., Berget,P. and Pipas,J.M. (1982) Mol. Cell. Biol. 2, 1258–1269.
7. Roth,D.B. and Wilson,J.H. (1985) Proc. Natl. Acad. Sci. USA 82, 3355–3359.
8. Roth.D.B., Porter,T.N. and Wilson,J.H. (1985) Mol. Cell. Biol. 5, 2599–2607.
9. Roth,D.B. and Wilson,J.H. (1986) Mol. Cell. Biol. 6, 4295–4304.
10. Roth,D.B., Chang,X.-B. and Wilson,J.H. (1989) Mol. Cell. Biol. 9, 3049–3057.
11. Wilson,J.H., DePamphilis,M. and Berg,P. (1976) J. Virol. 20, 391–399.
12. Nairn,R.S., Humphrey,R.M. and Adair,G.M. (1988) Int. J. Radiat. Biol. 53, 249–260.
13. Adair,G.M., Stallings,R., Nairn,R.S. and Siciliano,M. (1983) Proc. Natl. Acad. Sci USA 80, 5961–5964.
14. Adair,G.M., Carver,J. and Wandres,D. (1980) Mutat. Res. 72, 187–205.
15. Adair,G.M., Nairn,R.S., Wilson,J.H., Seidman,M.M., Brotherman,K.A., MacKinnon,C. and Scheerer,J.B. (1989) Proc. Natl. Acad. Sci. USA 86, 4574–4578.
16. Laski,F.A., Belagaje,R., RajBhandary,U.L. and Sharp,P.A. (1982) Proc. Natl. Acad. Sci. USA 79, 5813–5817.
17. Kootstra,A., Lew,L.K., Nairn,R.S. and MacLeod,M.C. (1989) Mol. Carcinogenesis 1, 239–244.
18. Mulligan,R. and Berg,P. (1980) Science 209, 1422–1427.
19. Lowy,I., Pellicer,A., Jackson,J., Sim,G., Silverstein,S. and Axel,R. (1980) Cell 22, 817–823.
20. Gorman,C.M., Merlino,G.T., Willingham,M.C., Pastan,I. and Howard,B.H. (1985) Proc. Natl. Acad. Sci. USA 79, 6777–6781.
21. Wake,C.T., Gudewicz,T., Porter,T.N., White,A. and Wilson,J.H. (1984) Mol. Cell. Biol. 4, 387–398.
22. Zagursky,R.J., Baumeister,K., Lomax,N. and Berman,N.L. (1985) Gene Anal. Tech. 2, 89–94.
23. Tegtmeyer,P. (1972) J. Virol. 10, 591–598.
24. Nalbantoglu,J., Phear,G.A. and Meuth,M. (1986) Nucleic Acids Res. 14, 1914.
25. Peden,K.W.C., Pipas,J.M., Pearson-White,S. and D. Nathans (1980) Science 209, 1392–1396.
26. Lusky,M. and Botchan,M. (1981) Nature 293, 79–91.
27. Michaeli,T. and Prives,C. (1987) Nucleic Acids Res. 15, 1579–1594.
28. Wilson,J.H. (1977) Proc. Natl. Acad. Sci. USA 74, 3503–3507.
29. Porter,T.N. (1988) Ph.D. Thesis, Baylor College of Medicine, Houston, TX
30. Levitt,N., Briggs,D., Gil,A. and N.J. Proudfoot (1989) Genes and Devel. 3, 1019–1025.
31. Adami,G. and Nevins,J.R. (1988) EMBO J. 7, 2107–2116.
32. Robberson,B.L., Cote,G.J. and Berget,S.M. (1990) Mol. Cell. Biol. 10, 84–94.