Proc. Natl. Acad. Sci. USA
Vol. 82, pp. 3781–3784, June 1985
Genetics

Homologous recombination between a defective virus and a chromosomal sequence in mammalian cells

(DNA integration/simian virus 40)

YOSEF SHAUL*, ORGAD LAUB†, MICHAEL D. WALKER‡, AND WILLIAM J. RUTTER‡

Departments of *Virology and †Genetics, The Weizmann Institute of Science, Rehovot, Israel; and ‡Hormone Research Laboratory, University of California, San Francisco, CA 94143

Contributed by William J. Rutter, February 14, 1985

ABSTRACT **Replacement of the early region of simian virus 40 results in virus that cannot replicate in a normal host, CV-1 cells, but can replicate in COS cells, a derivative of CV-1 cells that constitutively express simian virus 40 tumor antigen (T antigen). However, passage of such an early replacement simian virus 40 mutant in COS cells results in the emergence of virus that can propagate in CV-1 cells. Analysis of this virus revealed that the mutant rescued the integrated T-antigen gene from the COS cell genome. Comparison of the sequence of the recovered virus with that of the viral DNA resident in COS cells (strain 776) and the mutant used in our studies (derived from strain 777) proves that the mutant virus acquired the T-antigen gene from the COS cell chromosome via homologous recombination. Most probably this process was mediated by a direct genetic exchange.**

Genetic exchange between identical DNA sequences (homologous recombination) occurs in most if not all organisms. In lower eukaryotes exogenous DNA can integrate into the chromosome and replace the endogenous cognate DNA sequence by homologous recombination (1). In mammalian cells, homologous recombination between extrachromosomal DNAs can be readily detected (2–4). However, when the cells are transfected with exogenous genes, the DNA is integrated at many different loci, apparently randomly, with no preference for the homologous site within the chromosome. This raises the possibility that in mammalian cells recombination between homologous donors and endogenous chromosomal sequences occurs very rarely or not at all. Although some cases of genetic exchange between exogenous and endogenous DNA in mammalian cells have been observed [mainly through the rescue of genetic markers from cellular genomes (5–10)], the mechanism of these exchanges is unknown. In some instances, homologous recombination is the most plausible explanation. However, there has been no direct evidence at the DNA sequence level that the recombination involved homologous donor and host sequences.

In this study, we have used a conditionally replication defective mutant of simian virus 40 (SV40) as a powerful probe to select for recombination events between extrachromosomal sequences and integrated SV40 sequences in a permissive host, the COS cell (11, 12). Selection for recombinant virus was accomplished by propagation of the mutant virus in COS cells followed by challenge of the progeny virus with a cell type (CV-1) nonpermissive for replication of the parental virus. By using two different strains of virus and DNA sequence analysis, we have been able to determine the site of recombination and to demonstrate that the recombination occurs via homologous sequences.

The publication costs of this article were defrayed in part by page charge payment. This article must therefore be hereby marked "*advertisement*" in accordance with 18 U.S.C. §1734 solely to indicate this fact.

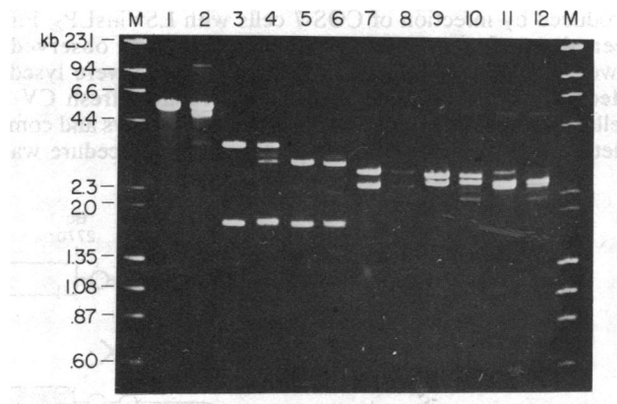

FIG. 1. Restriction pattern of the recombinant LSVinsLP$_2$ genome compared to that of wild-type SV40. Two 100-mm plates of confluent CV-1 cells were infected with 20 pfu per cell of either the recombinant virus LSVinsLP$_2$ or wild-type SV40 (strain 777). After 55 hr, extrachromosomal DNAs were extracted by the Hirt procedure (16), digested with restriction enzymes, and electrophoresed on a 1% agarose gel. Lanes 1, 3, 5, 7, 9, and 11 contain wild-type SV40 DNA; lanes 2, 4, 6, 8, 10, and 12 contain DNA of the recombinant LSVinsLP$_2$ virus. DNA samples were digested with the following enzymes: *Bgl* I (lanes 1 and 2), *Bgl* I and *Eco*RI (lanes 3 and 4), *Bgl* I, *Eco*RI, and *Taq* I (lanes 5 and 6), *Bgl* I and *Bcl* I (lanes 7 and 8), *Bgl* I and *Bam*HI (lanes 9 and 10), and *Bgl* I, *Bcl* I and *Bam*HI (lanes 11 and 12). The cleavage sites of these enzymes on the SV40 genome are indicated in Fig. 2. Lane M contains DNA size markers of λ DNA digested with *Hin*dIII and φX174 DNA cleaved with *Hae* III.

MATERIALS AND METHODS

Radioactive isotopes were purchased from New England Nuclear and Amersham. Restriction enzymes were from New England Biolabs and Bethesda Research Laboratories.

COS-7 and CV-1 cells were maintained in Dulbecco's modified Eagle's medium supplemented with 10% fetal bovine serum. Construction of plasmids and preparation of virus stock were as described (13). DNA sequence determination was carried out by chemical (14) and chain termination (15) methods.

RESULTS

Rescue of Tumor-Antigen (T-Antigen) Gene. The COS cell line is derived from a simian cell line (CV-1) permissive for lytic growth of SV40, by transformation with an origin-defective SV40 mutant (12). The cells contain no detectable episomal SV40 sequences. The genome of COS cells contains at least one copy of an intact T-antigen gene that is

Abbreviations: SV40, simian virus 40; kb, kilobase(s); T antigen, tumor antigen.

constitutively expressed. We have attempted to rescue this gene by using a defective SV40 virus ($LSVinsLP_2$) (13), in which the early gene region is absent. In $LSVinsLP_2$, the coding region for the SV40 early genes [between the *Bcl* I site (SV40 nucleotide 2770) and a *Hin*dIII site (nucleotide 5171)] is replaced by a fragment containing 1.6 kilobases (kb) of DNA including the entire human insulin gene, and 0.3 kb of SV40 late sequences. As we have previously described (13), since $LSVinsLP_2$ contains an intact SV40 origin of replication and a functional set of SV40 late genes, it can be propagated in COS cells to produce a virus stock, but it cannot replicate in CV-1 cells. Acquisition of a functional T-antigen gene via genetic exchange with the COS cells, however, should lead to production of a virus competent to replicate in CV-1 cells. To select for this putative viral recombinant, we challenged CV-1 cells with a virus stock produced by infection of COS-7 cells with $LSVinsLP_2$. Fifteen days after infection, cytopathology was observed. Twenty days after infection, most of the cells were lysed. Medium from this lysate was used to reinfect fresh CV-1 cells; a cytopathic effect was observed after 6 days and complete cell lysis occurred after 12 days. This procedure was repeated several times to dilute out parental $LSVinsLP_2$ and to increase the titer of the replicating virus. Since the only available source of the T-antigen gene in this system was the endogenous chromosomal gene from the COS cell genome, we infer that the replication-competent virus in CV-1 cells is a recombinant, which we term LSV-RI.

Genomic Composition of the Recombinant Virus. To determine the structure of LSV-RI, DNA was prepared from infected CV-1 cells by the Hirt procedure (16) and analyzed by restriction enzyme mapping. Cleavage with *Bgl* I yields two fragments (Fig. 1). The major fragment comigrates with the 5.3-kb band of wild-type SV40 DNA; the minor 4.8-kb fragment probably represents the parental $LSVinsLP_2$, which presumably replicates to some extent due to T-antigen production by the recombinant virus. The similarity of LSV-RI and wild-type SV40 is demonstrated by the similar patterns obtained after double digestion with *Bgl* I/*Eco*RI and *Bgl* I/*Bam*HI (Fig. 1). Significantly, the recombinant acquired the *Bcl* I site and the *Taq* I site, which are characteristic of wild-type SV40 DNA but are absent from the parental $LSVinsLP_2$ virus (13). These results suggest that the recombinant acquired the entire T-antigen gene.

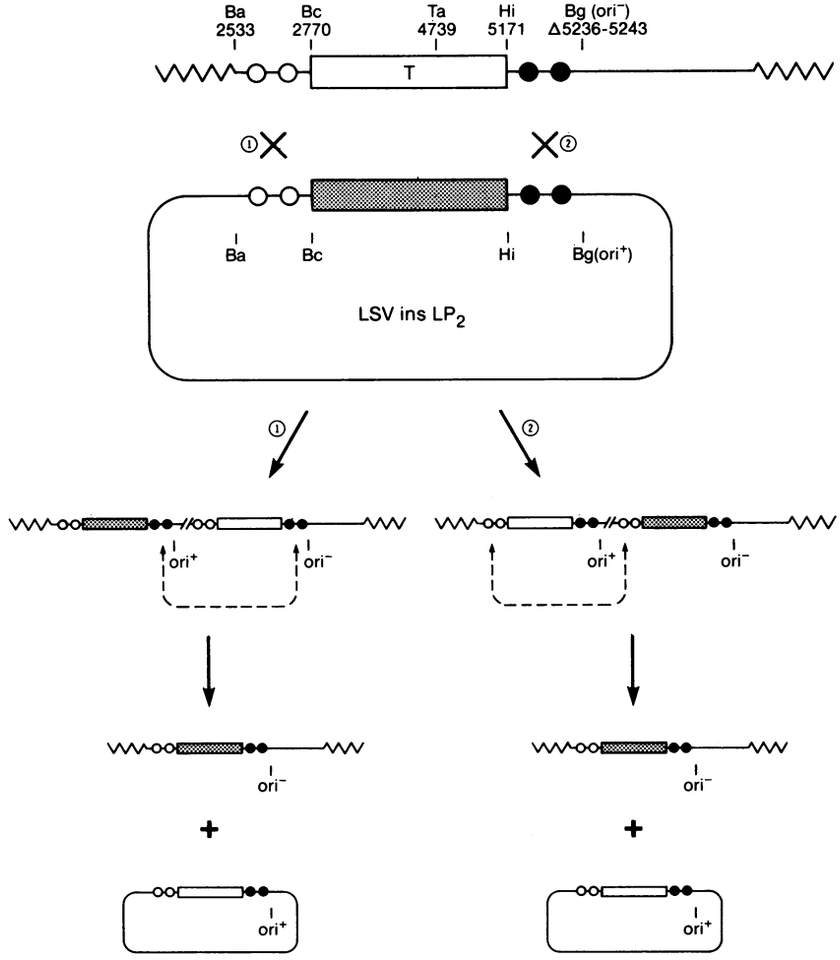

FIG. 2. Schematic representation of the proposed recombination between integrated SV40 DNA in the COS cell line and $LSVinsLP_2$ DNA. COS cells were generated (12) by using SV40 virus (strain 776) with a deleted origin of replication (Δ5236–5243). The structure of the integrated SV40 DNA in the COS genome was taken from Gluzman (12). Open rectangle marks the position of the T-antigen gene. Solid rectangle marks the location of the replaced portion of $LSVinsLP_2$ containing the human insulin gene. Zigzag lines represent host cell DNA; ori^-, defective origin of replication of the COS cell endogenous viral DNA. Ba, *Bam*HI; Bc, *Bcl* I; Bg, *Bgl* I; Ta, *Taq* I; Hi, *Hin*dIII. Numbers below the restriction enzyme abbreviation represent the location of the appropriate site in the wild-type viral DNA. The regeneration of intact wild-type SV40 DNA is proposed to occur via a double homologous recombination event. The portions of the viral genomes marked by open circles and solid circles delineate the regions at which recombination occurs. If recombination occurs initially at region 1 (open circles), the pathway shown on the left side of the figure would occur. Initial recombination at region 2 (solid circles) would lead to the pathway shown on the right. Broken arrows represent the second recombination events leading to generation of wild-type virus.

The acquisition of the T-antigen gene by the recombinant virus is most simply explained by two homologous recombination events (Fig. 2); one must occur between the *Hin*dIII site (map position 5171) and the *Bgl* I site (position 5235), because recombination beyond the *Bgl* I site would produce a virus containing the defective origin of replication. The other region of recombination must lie between the *Bam*HI (position 2533) and the *Bcl* I (position 2770) sites, because the LSVinsLP_2 genome contains nonviral sequences upstream from the *Bcl* I site, and the endogenous SV40 genome contains nonviral sequences beyond the *Bam*HI site (12).

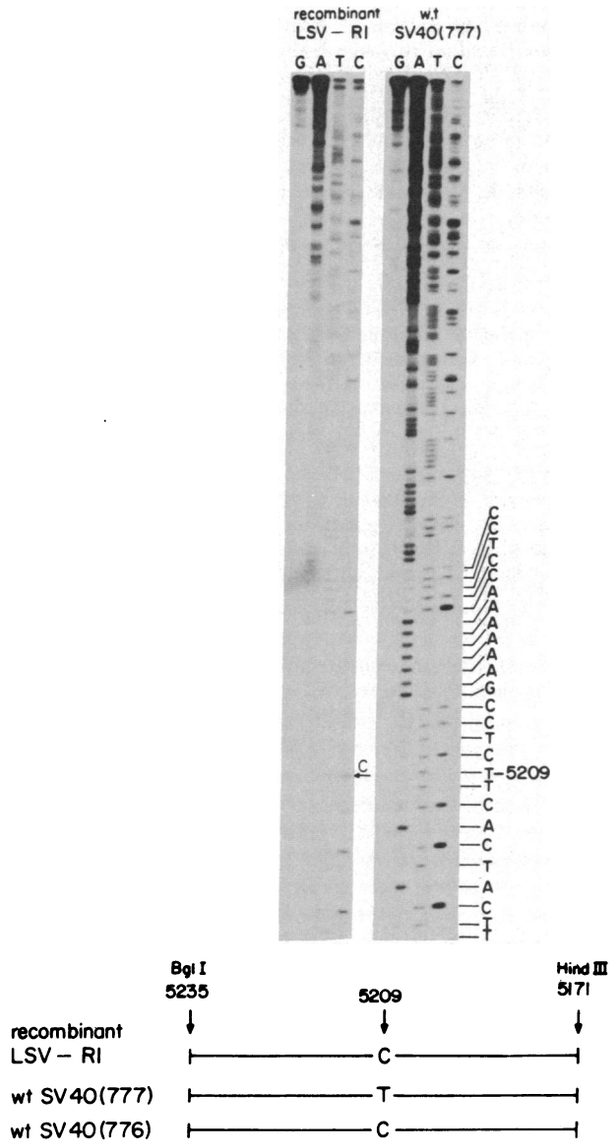

FIG. 3. Determination of DNA sequence in one of the regions of recombination of LSV-RI and the homologous region of wild-type SV40 DNA (strain 777). DNA of both viruses were prepared by the Hirt procedure (16). The *Bgl* I (nucleotide 5235)/*Bam*HI (nucleotide 2533) fragment was purified by electrophoresis on agarose gels and end-labeled using T4 polynucleotide kinase (14). The end-labeled fragments were further digested by *Taq* I (4739). The *Taq* I/*Bgl* I fragment was used for DNA sequencing, using an 8% polyacrylamide gel (14). SV40 strain 777 contains thymine at position 5209, while the recombinant LSV-RI contains cytosine. In the lower part of the figure, the sequence of the *Bgl* I/*Hin*dIII fragment of SV40 strain 777 and the LSV-RI were compared to the published sequence of SV40 strain 776, the COS cell resident virus (12, 17). A longer exposure time was used to visualize bands in the G lane.

DNA Sequence Analysis of the Regions at the Recombination Sites. The DNA sequences spanning both recombination sites were determined for the recombinant virus as well as the parental strain 777 and compared to the known sequence of SV40 strain 776, the resident virus of COS cells (Figs. 3 and 4). At the first recombination site (*Bgl* I/*Hin*dIII) (Fig. 3), the sequences of the three DNA fragments are identical, except that at nucleotide 5209 strain 777 contains thymine whereas strain 776 contains cytosine. Significantly, the recombinant LSV-RI contains cytosine, the same nucleotide as the COS cell resident virus. Due to the deletion of nucleotides 5236–5243, the resident virus does not contain the *Bgl* I site (11). However, the recombinant, like the parental virus (strain 777), retains this site at position 5235. Therefore, the site of recombination must lie between nucleotides 5209 and 5236. The perfect restoration of the sequence in this region is characteristic of homologous recombination.

The sequence of the second recombination site is presented in Fig. 4. The recombinant virus contains the genetic markers of both parental molecules: nucleotide 2957 derives from the parental virus and nucleotide 2751 originates from the COS cell genome. Unexpectedly, 6 base pairs (2769–2764) are missing from the recombinant viral genome, which predicts the loss of two amino acid residues from the T antigen.

DISCUSSION

We have shown that the chromosomally integrated T-antigen gene of COS cells can be rescued by a defective SV40 in which the early region of the virus is deleted. The rescue involves recombination events in two regions, which have been mapped and sequenced. Recombination between integrated SV40 sequences of transformed monkey cells and SV40 temperature-sensitive mutants has been reported (6, 18). However, the transformed cells used in those experiments contained SV40 sequences with an intact origin of replication. Therefore, the replication and excision of the resident virus prior to the recombination are not excluded. The absence of a functional SV40 origin of replication in COS cells excludes the possible involvement of such excised viral DNA in the experiments described here.

The mechanisms involved in generating these recombination events are not known. However, we envisage a two-stage interaction whereby the viral DNA first becomes integrated into the host genome and is subsequently excised (Fig. 2). This latter process is likely to be a relatively frequent event once the DNA containing an active origin of replication has become integrated and replicates, presumably via an "onion skin" mechanism (17, 19). As evidenced by the perfect restoration of sequence of the *Hin*dIII/*Bgl* I region, one stage in the process is a simple homologous recombination. In this instance, a stretch of 64 homologous bases is

		BclI							
		2770				2757		2751	
	ASP	HIS	ASN	GLN	PRO	TYR	HIS	ILE	CYS
wt SV40 (776)	GAT	CAT	AAT	CAG	CCA	TAC	CAC	ATT	TGT
LSV-RI	...	..Δ2769-2764.		...		..T	...	..T	...
wt SV40 (777)	...	...	...	...	...	..T	...	..C	...

FIG. 4. DNA sequence in the other region of recombination of LSV-RI compared with the sequence of the wild-type SV40 strains 776 and 777 in this region. The sequences of LSV-RI and SV40 strain 777 were determined by chain termination methods (15) and aligned with the published sequence of strain 776 (17).

sufficient for homologous recombination. This is 14 bases more than the limited amount of homology required for optimal homologous recombination in bacteria (20). Upon deletion of 20 more bases [removal of *Hin*dIII (nucleotide 5171) to *Stu* I (nucleotide 5190)], no recombinant virus was detected (data not shown) (21). The sequence in the other recombination region, which superficially appears to be the result of multiple crossover events, is in fact characteristic of the products of mismatch repair of heteroduplex intermediates formed during the process of homologous recombination (22, 23).

The recombination events described were reproducible: using three different deletion mutants we have isolated and defined three more recombinant viruses following passage on COS cells (data not shown). Most probably, the undefined wild-type virus, which was detected by Gluzman (12) after passage of early deletion mutants in COS-1 cells, was generated in a similar way. It is difficult to precisely determine the frequency of the recombination events. However, based on the time necessary for the viral stock to cause lysis of CV-1 cells, we estimate a titer of 1–10 plaque-forming units (pfu)/ml. Since we and others (12, 24) have determined that passage of defective virus on COS cells leads to titers of 10^7–10^8 pfu/ml, the likely frequency of the homologous recombination event is in the range of 10^{-6} to 10^{-8}.

Targeting exogenous DNA sequences to homologous sites on the genome requires isolating the reciprocal recombinant to those studied here. This may prove feasible by using appropriate selections or in some manner enhancing the rate of homologous recombination. A practical method for replacing an endogenous gene with an exogenous copy would be a powerful tool for studies of gene regulation and possible gene therapy.

We thank Ira Herskowitz and David N. Standring for valuable discussions and Leslie Spector for preparation of the manuscript.

1. Hinnen, A. H., Hicks, J. B. & Fink, G. R. (1978) *Proc. Natl. Acad. Sci. USA* **75**, 1929–1933.
2. Folger, K. R., Wong, E. A., Wahl, G. & Capecchi, M. R. (1982) *Mol. Cell. Biol.* **2**, 1372–1387.
3. Subramani, S. & Berg, P. (1983) *Mol. Cell. Biol.* **3**, 1040–1052.
4. De Saint Vincent, B. R. & Wahl, G. M. (1983) *Proc. Natl. Acad. Sci. USA* **80**, 2002–2006.
5. Hanafusa, H., Halpern, C. C., Buchhagen, D. L. & Kawai, S. (1977) *J. Exp. Med.* **146**, 1735–1747.
6. Gluzman, Y., Kuff, E. L. & Winocour, E. (1977) *J. Virol.* **24**, 534–540.
7. Vigne, R., Neil, J. C., Breitman, M. L. & Vogt, P. K. (1980) *Virology* **105**, 71–85.
8. Karess, R. E. & Hanafusa, H. (1981) *Cell* **24**, 155–164.
9. Ramsay, G. M., Enrietto, P. S., Graf, T. & Hayman, M. J. (1982) *Proc. Natl. Acad. Sci. USA* **79**, 6885–6889.
10. Goodenow, R. S., Stroynowski, I., McMillan, M., Nicolson, M., Eakle, K., Sher, B. T., Davidson, N. & Hood, L. (1983) *Nature (London)* **301**, 388–394.
11. Gluzman, Y., Sambrook, S. F. & Frisque, R. J. (1980) *Proc. Natl. Acad. Sci. USA* **77**, 3898–3902.
12. Gluzman, Y. (1981) *Cell* **23**, 175–182.
13. Laub, O. & Rutter, W. J. (1983) *J. Biol. Chem.* **258**, 6043–6050.
14. Maxam, A. M. & Gilbert, W. (1980) *Methods Enzymol.* **65**, 499–560.
15. Sanger, E., Nicklen, S. & Coulson, A. R. (1977) *Proc. Natl. Acad. Sci. USA* **74**, 5463–5467.
16. Hirt, B. (1967) *J. Mol. Biol.* **26**, 365–369.
17. Tooze, J. (1981) *Molecular Biology of Tumor Viruses* (Cold Spring Harbor Laboratory, Cold Spring Harbor, NY).
18. Vogel, T., Gluzman, Y. & Winocour, E. (1977) *J. Virol.* **24**, 541–550.
19. Botchan, M., Topp, W. C. & Sambrook, J. (1979) *Cold Spring Harbor Symp. Quant. Biol.* **43**, 709–719.
20. Singer, B. C., Gold, L., Gauss, P. & Doherty, D. H. (1982) *Cell* **31**, 25–33.
21. Rubnitz, J. & Subramani, S. (1984) *Mol. Cell. Biol.* **4**, 2253–2258.
22. Bullock, P. & Botchan, M. (1982) in *Gene Amplification*, ed. Schimke, R. T. (Cold Spring Harbor Laboratory, Cold Spring Harbor, NY), pp. 215–224.
23. Dressler, D. & Potter, H. (1982) *Annu. Rev. Biochem.* **51**, 727–761.
24. Laub, O., Rall, L., Truett, M., Shaul, Y., Standring, D. N., Valenzuela, P. & Rutter, W. J. (1983) *J. Virol.* **48**, 271–280.