Recombinational Joints in a Simian Virus 40 Variant Generated in a Persistent Infection

(Accepted 8 July 1982)

SUMMARY

SP1, a viable simian virus 40 (SV40) variant isolated from a persistent infection of rhesus monkey kidney cells, contains sequence rearrangements in the untranslated region of the SV40 genome which are transcribed into late mRNA leader sequences and in the region which encodes the large T antigen. Nucleotide sequences about the recombinational junctions in SP1 were determined. The sequence data show that in most instances there was not extensive homology between recombining sequences. The recombinant sequences are discussed with respect to the mechanisms by which they might have been generated.

Variants of simian virus 40 (SV40) which contain deletions, duplications, rearrangements and substituted cellular sequences arise when SV40 is serially propagated at high input multiplicity in permissive cells (for review, see Brockman, 1977). The nucleotide sequences across the recombinational joints in some of these variants have been compared with the parental viral and cellular sequences (Gutai & Nathans, 1978; Wakamiya et al., 1979; McCutchan et al., 1979; Woodworth-Gutai, 1981). A junction between adenovirus and SV40 sequences in an adenovirus-SV40 hybrid virus has been similarly examined (Zain & Roberts, 1978). In each of these instances, there was no extensive homology between recombining sequences. Thus, these virus variants were produced by an as yet unexplained 'illegitimate' recombinational process.

SV40 variants, similar to those isolated from serial passage stocks, also arise during persistent infection of permissive cells (Norkin, 1979a). This paper reports the nucleotide sequences across the recombinational junctions in SP1, a viable SV40 variant isolated from a persistent infection of rhesus monkey kidney cells (Norkin, 1979b). The results indicate that illegitimate recombinational processes are at least partly responsible for the generation of SP1. The recombinant sequences are discussed with respect to the mechanisms by which they might have been generated.

SP1 contains an insertion that was previously localized between the BglI (map location 0.67) and HpaII (map location 0.73) cleavage sites, in the untranslated region of the SV40 genome which is transcribed into the late SV40 mRNA leader sequences (Norkin, 1979b; Reddy et al., 1978a, b; Ghosh et al., 1978). The nucleotide sequence of this region of SP1 was determined by first isolating the restriction fragment bounded by the HpaII site at map location 0.73 and the HindIII site at map location 0.66. This fragment was labelled at its 5' ends with 32P and cut with BglI. The BglI/HpaII fragment was isolated and then sequenced by the method of Maat & Smith (1978). A typical sequencing gel is shown in Fig. 1.

The sequence data show that SP1 contains a perfect duplication of a 42 base pair parental wild-type (wt) sequence which extends from residue 190 to residue 249 (Fig. 2a, b). The numbering of the bases follows the convention of Reddy et al. (1978b). The duplicated sequence is bracketed at one end by the wt sequence ending at residue 249 and at the other end by the wt sequence beginning at residue 250. Note that residues 196 to 213 of the canonical wild-type sequence (Reddy et al., 1978a) are missing in both SP1 and its parent wt strain. This explains our earlier finding that both SP1 and its parent strain are resistant to restriction with KpnI (Norkin, 1979b). Another difference between SP1 and its parent wt strain is the insertion into SP1 of the hexanucleotide AGGTCA between residues 242 and 243 of one copy of the duplicated sequence (Fig. 2a, b).

The duplication in SP1 might be viewed as the result of a single recombinational event which joined residue 249 of one parental genome to residue 190 of another (Fig. 2b, c). The exact
position of the putative recombinational joint is ambiguous over several nucleotides because of a short region of homology at the junction sites (Fig. 2c). Because these nucleotide sequences display no long stretches of homology (Fig. 2c), this event would have to be accounted for by an as yet unexplained illegitimate recombinational process. The insertion of the hexanucleotide
Fig. 2. Sequence comparisons of SV40 strains showing presumptive recombinational joints in SP1 and its parental strain. (a) Sequence of wt strain 776 (Reddy et al., 1978b) and of the SP1 parent wt strain between 0.711 and 0.725 map units; the SP1 parent strain is missing residues 196 to 213. (b) SP1 sequence; the insertion (residues 190' to 249', excluding residues 196 to 213) is a duplication of the underlined parental sequence in (a). (c) Presumptive single-joint recombinational event which may have given rise to the large SP1 insert. Note lack of homology between recombining parental sequences. (d) Insertion into SP1 interpreted as a two-joint recombinational event; a sequence corresponding to residues 214 to 257 of the parental strain (e) is inserted between the contiguous residues 195 and 214 of the parental strain. (f) The canonical wt sequence (Reddy et al., 1978b) from residue 2620 to 2649: the underlined sequence is absent from SP1 based upon the Maxam & Gilbert (1980) sequence analysis of the short DNA fragment flanked by the restriction sites of BglI and BamHI. The same (underlined) sequence is presumed to be deleted from the SP1 parental strain, based upon the electrophoretic mobility of the BglI/BamHI restriction fragment of the parental strain.
AGGTCA between residues 242 and 243 of one copy of the duplicated sequence is also not explained.

The insertion might alternatively be viewed as the result of a two-joint recombinational event resembling that depicted in Fig. 2(d). The exact location of the 5' end of the insert is not determinable and this joint could have been the product of a legitimate recombinational process. In contrast, the junction at the 3' end of the insert would have been restricted to residues 250 to 257 on the donor chromosome and residues 190 to 195 on the recipient chromosome. This joint could not have involved long stretches of homology.

Short, discontinuous stretches of homology between otherwise heterologous parental strands have previously been implicated in the generation of an SV40 variant (Gutai & Nathans, 1978; Woodworth-Gutai, 1981). Also, it is known that homologous sequences of only 12 (or perhaps even fewer) nucleotides long are sufficient to promote unequal crossovers which generate tandem duplications in prokaryotic cells (Edlund & Normark, 1981). Consequently, the insertion event, as depicted here in the single-joint model (Fig. 2b, c), might be explained by an unequal crossover between the nearly homologous segments encompassed by residues 250 to 257 and 191 to 195. The CA doublet, or its complement, at positions 252 to 253 on one parental genome might have been looped-out from the putative heteroduplex recombinational intermediate.

The short homologous sequences in the putative single-joint event depicted in Fig. 2(b) resemble the octamer, GCTGGTGG, which is the longest common sequence at the Chi recombinational hot-spots of bacteriophage λ (Smith et al., 1981). These sites stimulate Rec BC-promoted recombination, presumably because they are recognized by one of the enzymes of the host's Rec BC pathway (for review, see Stahl, 1979a). Special sites which initiate recombination probably also act in eukaryotic cells, as inferred from the polarization of gene conversion (for review, see Stahl, 1979b). Furthermore, there is indeed suggestive evidence that Chi-like sequences function in eukaryotes to promote the rearrangements of immunoglobulin genes (Kentner & Birshtein, 1981). The Chi-like sequence might have promoted the SP1 duplication by either a single or a two-joint event since Chi can promote patch formation (even across regions of gross heterology) as well as splices (Stahl & Stahl, 1975; Stahl et al., 1980).

The same outcome might have resulted from an intramolecular event in which local unwinding and mispairing were followed by a gap-filling step, generating the insertion. In this case, residues 250 to 257 would mispair with residues 191 to 195 on the other strand of the same DNA molecule. A similar mechanism was suggested to explain frameshift mutagenesis (Streisinger et al., 1966). The putative Chi might also have facilitated this process by serving as the substrate for the generation of either an 'aggressive' free end or a gap.

Other virus recombinant joints appear to have been generated in the absence of any sequence complementarity (see Zain & Roberts, 1978; McCutchan et al., 1979; Woodworth-Gutai, 1981). Thus, a mechanism probably exists which can promote recombination in the complete absence of homology.

There are features of the two-joint model, with the junctions as shown in Fig. 2(d), which are reminiscent of a transposon-like process. For example, prokaryotic transposable elements are generally bracketed by a duplication of a segment of the target DNA (for review, see Grindley & Sherratt, 1978). The insert in SP1 (as interpreted in Fig. 2d) is bracketed by the sequence GCTGTTG which is present on the recipient, but not the donor chromosome. Thus, some aspect of a transposition-like process might have resulted in the deletion of two residues (CA) from positions 252 to 253 to yield the repeat at the 3' end of the insert (Fig. 2d, e). It should be noted that the generation of flanking duplications of five to nine base pairs is also a feature of insertion by mobile genetic elements in Drosophila and yeast (Cameron et al., 1979; Potter et al., 1979).

All prokaryotic transposable elements so far examined also have short inverted repeat sequences near their ends (Grindley & Sherratt, 1978). Thus, note that the sequence GGTTAGGT, beginning at position 25 of the insert (Fig. 2d) is an inverted repeat of the sequence ACCTAACC at position 2 of the insert. This sequence might also have been created during the insertion event since it too does not exist in the parental sequence (Fig. 2d, e).

The recombinational joints about the SP1 insert are similar to those which bound an insertion of cellular DNA in the recently described SV40 variant, in 1449 (Dhruva et al., 1980). Both inserts are bracketed by short repeated sequences and both include short inverted repeats near
their termini. Also, as a consequence of the insertion, a tetranucleotide was deleted from the SV40 parent of in 1449 at the insertion site. The resemblance of the in 1449 insert to a prokaryotic insertion element was also noted. Nevertheless, features suggestive of a transposon-like process are not found at the recombinational joints of other SV40 variants (see Gutai & Nathans, 1978; McCutchan et al., 1979). Thus, transposon-like processes are also not a general mechanism underlying the generation of these variants. Note that these virus insertions are considerably shorter than known prokaryotic transposable elements and that they lack the long terminal repeats characteristic of prokaryotic movable elements. However, eukaryotic transposable elements do not necessarily carry long terminal repeats (see David et al., 1981), although they too are generally longer than these virus insertions.

The mechanism which generated the deletion of residues 196 to 213 from SP1 and its parent strain (Fig. 2a) is also unknown. Because the deletion is flanked at its 5' end by the putative Chi which may have promoted the SP1 insertion at that site, it is tempting to suggest that Chi might have promoted the deletion as well. Alternatively, the shorter GGT sequence, repeated at positions 195 to 197 and 212 to 214 (Fig. 2a), may have played the primary role in generating the deletion. This is suggested by many examples in both prokaryotes and eukaryotes in which short repeated sequences act as endpoints of spontaneous deletions (see Farabaugh et al., 1978; Marotta et al., 1977). In this regard, another mutation in SP1 and its parent strain deletes residues 2634 to 2639 (or 2635 to 2640, Fig. 2f) from the region of the genome which encodes the C-terminal portion of the tumour (T) antigen (Reddy et al., 1978b). This deleted segment is flanked by the short repeating segment, GGT (or AGG). Short repetitive sequences are also apparent in the nucleotide sequences of other SV40 and polyoma virus deletion mutants (Seif et al., 1980; Smolar & Griffin, 1981).

The frequency with which short repetitive sequences are associated with the generation of deletions strongly suggests a causal relationship. Nevertheless, because of variations in the nucleotide composition and length of the repetitive sequences, and their position with respect to the deleted segments, it is difficult to suggest a single mechanism that might account for their action. Furthermore, other virus deletions are not bracketed by repetitive sequences (see Smolar & Griffin, 1981), indicating that the involvement of the repetitive sequences is still speculative and that more than one mechanism might underlie the generation of these mutants. Other SV40 deletions were probably generated by legitimate recombination events involving sequence homology (Van Heuverswyn & Fiers, 1979).

In summary, whereas several plausible mechanisms might account for the illegitimate recombination events which generated the lesions in SP1, the nucleotide sequences about the recombinant joints in SP1 do not indicate which, if any, of these mechanisms is indeed correct. Consideration of other SV40 and polyoma virus variants suggests that more than one mechanism might underlie the illegitimate recombination events by which they are often generated.

We thank Donna Mitchel and R. Bruce Register for their excellent technical assistance and Donna Tudryn for her expert preparation of the manuscript. We are grateful to Sherman Weissman and Maxine Singer for critically reading an earlier version of this manuscript. This investigation was supported by Public Health Service Research grant 1 R01 AI14049 from the National Institute of Allergy and Infectious Diseases, Biomedical Research Support grant RR07048, and a grant from the American Cancer Society. Michael Piatak was supported by Public Health Service Traineeship CA09159 and was a fellow of the Leopold Schepp Foundation.

Leonard C. Norkin
Michael Piatak

Department of Microbiology
University of Massachusetts
Amherst, Massachusetts 01003, U.S.A.

Department of Human Genetics
Yale University School of Medicine
New Haven, Connecticut 06510, U.S.A.

† Present address: Cetus, 600 Bancroft Way, Berkeley, California 94710, U.S.A.

REFERENCES


VAN HEUVERSWYN, N. & FIERS, W. (1979). Nucleotide sequence of the


REDDY, V. B., THIMMAPPAYA, B., DHAR, R., SUBRAMANIAN, K. N., ZAIN, B. S., PAN, J., GOSH, P. K., CELMA, M. L. &

REDDY, V. B., GAOSH, V K., LEBOWITZ, P. & WEISSMAN, S. M. (1978a). Gaps and duplicated sequences in the leaders of

NORKIN, L. C. (1979b). Small plaque variants of simian virus 40 from a persistent infection of rhesus monkey


KENTNER, A. L. & BERSHTEIN, B. K. (1981). Chi, a promoter of generalized recombination in \( \lambda \) phage, is present in

MAAT, J. & SMITH, A. J. H. (1978). Human \( \beta \)-globin messenger RNA. II. The structure of the variant 1103 and its comparison to variant


NORKIN, L. C. (1979a). The emergence of simian virus 40 variants in a persistent infection of rhesus monkey kidney

NORKIN, L. C. (1979b). Small plaque variants of simian virus 40 from a persistent infection of rhesus monkey kidney


REDDY, V. B., GHOSH, P. K., LEBOWITZ, P. & WEISSMAN, S. M. (1978a). Gaps and duplicated sequences in the leaders of

SV40 16S RNA.

Nucleic Acids Research 5, 4537-4545.

McCUTCCHAN, T., SINGER, M. & ROSENBERG, M. (1979). Structure of simian virus 40 recombinants that contain both host and viral DNA sequences. II. The structure of the variant 1103 and its comparison to variant


Nucleotide sequences derived from complementary DNA. Journal of Biological Chemistry 252, 5040-5053.


Nucleic Acids Research 5, 4195-4213.

REDDY, V. B., THIMMAPPAYA, B., DHAR, R., SUBRAMANIAN, K. N., ZAIN, B. S., PAN, J., GOSH, P. K., CELMA, M. L. &


STAHL, F. W. & STAHL, M. M. (1975). Rec-mediated recombinant host activity in bacteriophage \( \lambda \). IV. Effect of

heterology on Chi-stimulated crossing over. Molecular and General Genetics 140, 29-37.


orientation dependence and directionality of Chi? In Mechanistic Studies of DNA Replication and Genetic


VAN HEUVERSWYN, H. & FERS, W. (1979). Nucleotide sequence of the Hind-C fragment of simian virus 40 DNA:

comparison of the \( S^\prime \)-untranslated region of wild-type virus and some deletion mutants. European Journal of

Biochemistry 100, 51-60.


contain both host and viral DNA sequences. I. The structure of \( CV_{B(1)} / P_2(E)CoR \) res). Journal of Biological

Chemistry 254, 3584-3591.


virus Ad\(_2\)ND\(_1\). Journal of Molecular Biology 120, 13-31.

(Received 27 April 1982)