Recombination in Simian Virus 40-infected Cells

STRUCTURE OF NATURALLY ARISING VARIANTS ev-2114, ev-2102, AND ev-1110*

Lowell Shellin, Anthony Celeste, and Mary Woodworth-Gutai†

From the Department of Cell and Tumor Biology, Roswell Park Memorial Institute, Buffalo, New York 14263

The complete nucleotide sequence has been determined for three newly cloned evolutionary variants from two different independently generated evolutionary series (1100 and 2100 series) of simian virus 40 (SV40). These naturally arising variants, designated ev-1110, ev-2102, and ev-2114, were isolated after five high multiplicity serial passages. The structure of the variants consists of a monomeric unit tandemly repeated four times (ev-2102 and ev-2114) or six times (ev-1110) in the variant genome; the variants have four or six copies, respectively, of the viral origin signal for DNA replication. The DNA content in the three variants is vastly different in that the genome of variant ev-2114 contains only rearranged viral sequences, while variant ev-2102 contains a substitution with monkey DNA sequences consisting of a nearly complete dimeric unit of Alu family sequences as well as less repetitive sequences and variant ev-1110 contains monkey DNA sequences derived solely from repetitive α-component DNA. Recombination events, cellular sequences, and structural features of these and other naturally arising SV40 variants are compared.

The undiluted serial passage of SV40 in permissive monkey kidney cells results in the generation of evolutionary variants whose genomes represent the net result of recombination events that have occurred either among viral DNA sequences with the subsequent duplication and/or deletion of segments of the wild type SV40 genome or between viral and host DNA with the subsequent incorporation of host DNA and loss of additional viral sequences (for review see Ref. 1). We previously reported the structure of three variants, ev-1114, ev-1117, and ev-1119, from the 3rd serial passage (2, 3) which contain viral sequence rearrangements including two copies of the SV40 origin (ori) of replication and two viral-viral recombinant joints per variant genome, two variants from the 13th passage, ev-1101 and ev-1103, (4, 5), which are host-substituted variants each containing two unique viral-host junctions and five or nine copies of ori, respectively, and a variant from the 45th passage, ev-1104 (5, 6), which contains host DNA, six copies of ori arranged as inverted repeats, and six unique viral-host recombinant junctions. The host DNA in ev-1103 and ev-1104 consists of both repetitive α-component (7) and low reiteration frequency host DNA while only low reiteration frequency host DNA has recombined with viral DNA in ev-1101.

To compare recombination events in independently derived variants, we serially passed a second evolutionary series (2100 series) distinct from the variants isolated from the 1100 series. Since the initial recombination events are probably obscured by multiple recombination events in late passage variants, we isolated variants from the 8th undiluted serial passage in order to analyze recombination events that occur earlier in the evolution of SV40 variants.

One of these naturally arising variants, ev-2114, is missing the early region of SV40 as the result of recombination events which have generated two viral-viral recombinant junctions between the SV40 late region and the noncoding regulatory region located in HindIII-DIII restriction fragment Hin C which includes the origin for viral DNA replication. The other variants, ev-2102 and ev-1110, have approximately equal amounts of viral and host DNA with two viral-host recombinant junctions.

The present study is the first report of the incorporation into an evolutionary variant of SV40 of a monkey sequence that shares homology with the Alu family of dispersed repetitive sequences found in human and monkey DNA (8). Dhruva et al. (9) have also identified a monkey Alu sequence in a SV40-viable deletion mutant that acquired the monkey DNA insertion during transfection of monkey kidney cells with the viral mutant DNA. In both SV40 derivatives a conserved replication origin-like sequence is included in the monkey Alu family insert.

In contrast to other host-substituted variants of SV40 which contain only part of the 172-bp HindIII-cleaved repeat unit of α-component monkey DNA (5, 10, 11), variant ev-1110 contains a 334-bp monkey sequence homologous to nearly two tandem repeat copies of α-component DNA.

From our previous studies, the minimum Hin A sequence retained in naturally arising evolutionary variants was 31 bp. However, the sequence analysis of newly isolated ev-2114 shows that only 4 bp of Hin A sequence remain and 3 of the 4 bp are part of the recognition site for the HindIII enzyme.

This study discusses the structure of these and other SV40 variants including the nucleotide sequence analysis of the recombination events.

MATERIALS AND METHODS

1 The abbreviation used is: bp, base pair(s).
2 Portions of this paper (including "Materials and Methods," part of "Results," and Figs. 1-8) are presented in miniprint at the end of this paper. Miniprint is easily read with the aid of a standard magnifying glass. Full size photocopies are available from the Journal of Biological Chemistry, 9650 Rockville Pike, Bethesda, MD 20814. Request Document No. 93M-1121, cite the authors, and include a check or money order for $6.00 per set of photocopies. Full size photocopies are also included in the microfilm edition of the Journal that is available from Waverly Press.

* This research was supported by Grant CA-28250 from the National Cancer Institute, Department of Health and Human Services. The costs of publication of this article were defrayed in part by the payment of page charges. This article must therefore be hereby marked "advertisement" in accordance with 18 U.S.C. Section 1734 solely to indicate this fact.

† Recipient of Research Career Development Award CA-00449 from the National Cancer Institute.
RESULTS AND DISCUSSION

We have isolated and characterized three naturally arising evolutionary variants, ev-2114, ev-2102, and ev-1110, to learn additional information about recombination events in SV40-infected animal cells. All three variants were isolated from the 5th undiluted serial passage in order to compare variants from the same early passage of independently generated evolutionary series (1100 and 2100), to avoid the complexities of multiple recombination events seen in the late passage variants, and to separate independently generated recombination events from those that are the result of a particular recombination event being studied repeatedly as it is passed along through the evolution of progeny variants.

Recombinant ev-2114 was apparently generated when two large deletions removed the early region (Hin H, I, B, and all but 4 bp of Hin A) and part of the late region (Hin D and E, part of Hin C, and all but 36 bp of Hin K), with the result that the 346-bp noncoding segment of SV40 Hin C (plus 4 bp of Hin A) was joined to a 736-bp segment of late region (Hin F, J, and a portion of Hin K and G) (see Figs. 1 and 4 for comparison of wild type and variant maps). This recombinant segment was then copied to form four tandem repeats that comprise the variant genome. It is of interest to compare recombinant ev-2114 from the 5th passage to three variants isolated from the 3rd passage of evolutionary series 1100 (eu-1114, ev-1117, eu-1119 (3, 4)) which also contain no host sequences. The rearrangement of viral sequences in the 3rd passage variants reflects one deletion and one duplication per variant genome and each variant has two copies of the ori signal for viral DNA replication. In the five variants from two independently generated evolutionary series, there are a total of five recombinant joints that contain Hin C sequence and although the site of recombination in each variant has occurred within a 230-bp stretch of Hin C, in every instance recombination has occurred at a different nucleotide position in the Hin C sequence. Similarly different Hin A nucleotide residue occurs at the recombinant sites involving Hin A sequences. This data strongly points up the lack of any site specificity in the viral-viral recombination events taking place in SV40-infected monkey kidney cells as also previously noted (3, 22).

In recombinant ev-2102, a 565-bp viral segment derived from the Hin C-Hin A region of the SV40 genome has recombined with a 664-bp segment of monkey cell DNA. This recombinant sequence is tandemly repeated four times in the variant genome and the variant has four copies of ori. By computer analysis, the 664-bp segment of monkey cell DNA in the ev-2102 monomeric unit does not share homology with host-derived DNA sequences present in the 1100 series variants or any other evolutionary variants of SV40 isolated in our laboratory (5, 6) or other laboratories (11), nor with host sequences present in monkey cell repetitive a-component DNA (7). However, the 664-bp host sequence present in ev-2102 does contain a 233-bp sequence which is homologous to the Alu family of dispersed repetitive sequences found at a frequency of about 3% in human chromosomal DNA (8). Alu-type sequences have been identified in a number of different species (8) including purified preparations of small polydisperse circular DNA from the BSC-1 line of monkey kidney cells (32) and from a library of African green monkey genomic DNA (33). Dhruva et al. (9) have identified a 157-bp segment of monkey DNA corresponding to Alu family sequences which has been inserted into the 5' leader sequence of the early

region of an SV40 mutant, in1449. As shown in Fig. 7, the monkey insert of ev-2102 includes all but about 50 bp of a complete Alu sequence from the monkey genome and contains the representative features of human and primate Alu repeats (8), namely a conserved 9-bp palindrome, a dinucleotide structure consisting of 2 arms each approximately 130-bp long, an A-rich region between the 5' arm and 3' arm, a portion of the 31-bp insert in the 3' arm, and homology to human and monkey Alu repeat sequences. Unlike other reported monkey Alu-type sequences (9, 32, 33), the monkey Alu sequence insert in ev-2102 has an internally perfectly repeated 14-bp sequence which although it differs by two nucleotides from the 14-bp direct repeat found in the human consensus Alu sequence, a perfect repeat has nevertheless been conserved (nucleotide residues 213–200 and 78–65 of Fig. 7). The short direct repeats which usually flank the Alu family of mobile sequences are not identifiable in ev-2102 nor is the type of homology at the recombinant joints similar to that reported for in1449. The Alu equivalent monkey DNA insert is flanked by perfectly homologous decanucleotides (enclosed by boxes in Fig. 6); however, they occur on complementary strands. There is a 5-bp viral sequence adjacent to the decanucleotide in Hin A which is inverted and repeated in the Alu-type sequence at 16 bp from the viral-host recombinant junction (Fig. 5c). The precisely defined end, characteristic of Alu-type sequences, is intact in ev-2102 and is joined to a segment of single copy or low repetition frequency monkey DNA at residue 233; the other end of the Alu-type sequence has recombined with viral Hin A sequences. To determine whether this segment of less repetitive DNA was originally joined to the Alu family sequence in the cellular genome, the less repetitive DNA segment needs to be cloned from a monkey cell library.

It is unclear whether the conservation of the Alu family sequences in different species has any structural/functional significance. The replication ori-like sequence present in the Alu family sequences (34) which shares homology with the replication origin sequences of several papovaviruses has also been conserved in the ev-2102 Alu sequence insert (nucleotide residues 58–45 in Fig. 7). It remains to be determined if the Alu repeat sequences play a role in the initiation of DNA replication and whether the inclusion of a 233-bp segment of the Alu family of dispersed repeats in a SV40 evolutionary variant contributes to the selective replicative advantage of the variant. We plan to dissect and test the “enhancing” potential of the 664-bp segment as well as other host segments present in the fully characterized naturally arising variants of series 1100 and 2100 in terms of both replication and recombination.

Variant ev-1110 is similar to previously characterized host-substituted variants in series 1100, i.e. eu-1101 and eu-1103 (13th passage) (4, 5, 10) and eu-1104 (45th passage) (5, 6) in that the only viral sequences retained are derived from the Hin C-Hin A region of the SV40 genome and these viral sequences have recombined with monkey cell DNA. Whereas variant ev-1101 contains only nonrepetitive cell DNA, ev-1110 contains only the repetitive a-component monkey DNA, while ev-1103 and ev-1104 contain a portion of the repetitive monkey a-component DNA in addition to segments of low frequency monkey DNA which share no homology to the cellular sequence in ev-1101. Variant ev-1110 is particularly interesting because we know of no other instance where SV40 viral DNA has recombined exclusively with a-component monkey DNA. Furthermore, ev-1110 is the first variant which has incorporated a complete copy of the 172-bp repeat. In fact, ev-1110 has recombined with a concatenated 334-bp segment of the a-component sequence which is nearly two

copies of the basic 172-bp HindIII repeating unit. This is in contrast to ev-1103 whose cellular segment consists of 142 bp of the 172-bp α-component sequence inserted between a less repetitive cell sequence and a portion of the viral Hin C segment (5, 10). The 41-bp α-component sequence in ev-1104 is similarly inserted between cell and viral sequences (5), while the 66-bp α-component sequence in ev-1108 and the 156 bp of α-component sequence in CVP8/1/P2 (11) are flanked on both sides by less repetitive cell DNA and inserted between SV40 DNA sequences. Thus, the junction in ev-1110 between viral DNA and α-component DNA may reflect an initial recombinational event between the host and viral genomes in contrast to ev-1103, ev-1104, ev-1108, and CVP8/1/P2 where it seems more likely that the structures of these variants are indicative of additional recombinational events between repetitive and less repetitive cellular DNA.

A. T-rich sequences are often found at the recombinant joints in SV40 variants (3, 5, 10) including ev-2102 and ev-2114 and at the site of A integration in Escherichia coli (35). Such a preponderance of A- T residues could increase the potential for the recombination event to occur at a given site. Based on nucleotide sequence analysis of recombining parental DNAs contiguous to the recombinant joints in series 1100 variants, we have also postulated that the presence of short 4- to 7-bp stretches of homology between recombining DNAs could help stabilize the recombinant intermediate prior to joint formation in SV40 variants (3, 5, 22). However, in the case of ev-1110, the homology between the parental α-component sequence and parental SV40 sequence is not striking and the A. T content is about average. Perhaps the high reiteration frequency of the α-component sequence facilitates its recombination to viral DNA in the absence of rich A. T base content or patchy homology. Although several direct-repeat and inverted repeat sequences (5-6 bp in length) clustered around the two viral-host recombinant joints are noted in Fig. 5, c and d, what contribution if any they make to the recombination event is unknown. Nucleotide sequence analysis of the viral-host junctions reveals no single recognition sequence for recombination in either host or viral DNA. Indeed, analysis of both viral-viral and viral-host junctions in variants ev-2102 and ev-2114 from a new independently generated series (2100) of evolutionary variants points out the same lack of dependence on even minimal homology for recombination to occur. Furthermore, we have seen another example of the insertion of an extra nucleotide of unknown source at the crossover point (ev-2114).

The α-component sequence in ev-1110 is compared with that in variants ev-1103, ev-1104, ev-1108, CVP8/1/P2 (10) and with the known α-component monkey sequence (7) (see Fig. 8). It should be noted that because of the single common core nucleotide at the site of recombination between viral and host DNA, it is not known precisely whether the monomeric unit of ev-1110 consists of 334 bp or 335 bp of α-component DNA. The ev-1110 sequence differs by a few nucleotides from the α-component sequence present in the other four variants and in monkey cell DNA. As compared to monkey cell α-component DNA, there are five single base changes in the ev-1110 sequence. The observed divergence could be representative of different classes of α-component sequence present in the cell (as suggested by McCutchan et al. (10)) in which case each of the five variants has recombined with a different class of α-component sequence. However, since four of the variants were isolated from the same evolutionary series 1100 and nucleotide sequence analyses suggest that ev-1104 and ev-1108 are derived in part from ev-1103 or from a common ancestral molecule (5), divergence seen in the α-component sequence may have occurred after the evolutionary predecessor was formed.

One approach to understanding the functional significance of regulatory sequence signals is to study the incorporation and/or retention of a given sequence in naturally arising variants which become the dominant species presumably because they have a selective advantage over other defective and wild type helper virus during high multiplicity passage. For example, the 142-bp segment of α-component DNA present in ev-1103 (passage 13) is covalently joined at nucleotide 133 (refer to numbering of α-component DNA in Fig. 8) to a segment of less highly repetitive monkey DNA. At the 40th undiluted serial passage (same 1100 series), we cloned variant ev-1108 which has retained the segment of low reiteration frequency monkey DNA covalently linked to α-component DNA at nucleotide 133 but has lost 76 bp of the 142-bp α-component segment. Five passages later, we find that ev-1104 (passage 45) still retains the low frequency monkey DNA segment covalently joined to nucleotide 133 of α-component DNA but only 41 bp of α-component are retained. It remains to be tested, but probably the presence of α-component DNA in ev-1110 reflects the high reiteration frequency of α-component DNA rather than some selective advantage the sequence might confer on serially propagated variants. The data suggests that if the incorporation of a particular cellular sequence contributes to the selective advantage of the variant, that advantage most likely is contained in the segment of low frequency monkey DNA presumably retained intact from the 13th to 45th passage while at most only a 41-bp segment of the α-component sequence might have anything to contribute. It is interesting that the A. T base content of the retained segment of low frequency DNA is extremely high (83%).

Finally, it is worth noting that in comparing all variants isolated to date (series 1100 and 2100), the HindIII site at the Hin A/C junction is always preserved as if it is supplying some as yet unidentified cis-acting function. The 45th passage variant ev-1104 has multiple A/C junctions per monomer unit (5/monomer) whose Hin A content varies in size from 31 to 65 bp (5). That this sequence is not required to initiate replication was first demonstrated by Lai and Nathans (36) who constructed a variable deletion mutant, dl-1002, from which the entire Hin A segment has been deleted leaving Hin C joined to Hin H (see Fig. 4 for reference to wild type SV40 map) but which still can replicate in the presence of a helper virus. We also constructed a mutant which is missing 28 bp of Hin A immediately adjacent to the A/C junction and it replicates normally (6). A study by Lee and Nathans (37) indicates that this Hin A region does not play a role in encapsidation. Sequence analysis of newly isolated ev-2114 shows that only 4 bp of Hin A sequence remain and 3 of the 4 bp are part of the recognition site for the HindIII enzyme. It seems clear that Hin A sequences are indeed dispensable. It remains to be seen whether an evolutionary variant will be isolated that has lost the HindIII site at the Hin A/C junction but retained the adjacent sequence signal for initiating DNA replication.

Acknowledgments—We thank Dr. Kenneth Manly for writing the computer program used to search for sequence homologies and Marcia Held and Bonnie Becue for typing the manuscript.

REFERENCES

Structure of SV40 Variants ev-2114, ev-2102, and ev-1110

Lowell Shieh, Anthony Caiete and Mary Woodworth

Supplementary material to

Recombination in SV40 Infected Cells: Structure of Naturally Arising Variants ev-2114, ev-2102 and ev-1110

Excised DNA fragments larger than 50 nucleotides in length were electrophoresed on 10% polyacrylamide-urea gels. Eightout of 10 labeled terminal DNA fragments were isolated from polyacrylamide gels and eluted from the gel matrix using electrophoresis digestion with the enzyme Bal31 (16).

The restriction endonucleases used for DNA sequencing were EcoR, HindIII, BamHI, and SalI. The restriction endonuclease sites were determined by cleaving the DNA fragments with these enzymes and hybridizing them to the radiolabeled DNA fragments.

Computer analysis of DNA sequences present in variant DNAs. A computer program used to analyze the DNA sequence data provided by computer analysis is described in the data section.
The results of restriction enzyme analysis of ev-1110 are summarized in the physical map shown in Fig. 1A. Digestion of ev-1110 DNA with Eco RI produces a single 14.5-kb fragment. By electron microscopy, the virus-variant DNA is approximately 60% of wild type in length, and therefore the large fragment must be tightly repeated six times in the full-length variant genome.

Fig. 1. a) Restriction endonuclease cleavage map of a portion unit of ev-1110. b) Horizontal arrows denote the polarity and extent of sequences determined. c) Summary map of viral and host DNA contents. The region of viral DNA sequence is covered by host sequence, as indicated by host DNA sequences. The regions of viral DNA sequence are shaded. The insertions and deletions are due to the integration of viral DNA sequences into the host DNA sequences.

Fig. 2. The nucleotide sequences of the recombinant junctions in a: EV-2114 (c) and ev-2102 (b) are shown. The recombinant parental DNA sequences are denoted by the open box. The recombinant parental DNA sequences are shown by the filled box. The nucleotide sequence of the recombinant parental DNA sequences are shown by the shaded box. The nucleotide sequence of the recombinant parental DNA sequences are shown by the filled box.

Fig. 3. a) Restriction endonuclease cleavage map of a portion unit of ev-1110. b) Horizontal arrows denote the polarity and extent of sequences determined. c) Summary map of viral and host DNA contents. The region of viral DNA sequence is covered by host sequence, as indicated by host DNA sequences. The regions of viral DNA sequence are shaded. The insertions and deletions are due to the integration of viral DNA sequences into the host DNA sequences.

Fig. 4. The nucleotide sequences of the recombinant junctions in a: EV-2114 (c) and ev-2102 (b) are shown. The recombinant parental DNA sequences are denoted by the open box. The recombinant parental DNA sequences are shown by the filled box. The nucleotide sequence of the recombinant parental DNA sequences are shown by the shaded box. The nucleotide sequence of the recombinant parental DNA sequences are shown by the filled box.

Fig. 5. The nucleotide sequences of the recombinant junctions in a: EV-2114 (c) and ev-2102 (b) are shown. The recombinant parental DNA sequences are denoted by the open box. The recombinant parental DNA sequences are shown by the filled box. The nucleotide sequence of the recombinant parental DNA sequences are shown by the shaded box. The nucleotide sequence of the recombinant parental DNA sequences are shown by the filled box.
Structure of SV40 Variants ev-2114, ev-2102, and ev-1110

As depicted in Fig. 5, the monomer fragment of ev-1110 consists of a 471 bp viral sequence containing a portion of the Hae III-C fragment of SV40 which is recombined with a 114 bp segment of repetitive -component monkey DNA. The host cell DNA is inserted at nt 1304 of viral Hae III-X and nt 1296 of viral Hae III-C. The sequence at some viral-host recombinant junctions is shown in Fig. 5, and the structure of the viral-host recombinant junction is similar to that of the monomer in ev-2110. The viral-host recombinant junction is characterized by the presence of a 12 bp palindrome and a 4 bp direct repeat at the junction. The host sequence in ev-2110 which is entirely derived from -component monkey DNA is shown in Fig. 8.

Fig. 6. Composite of the monkey DNA sequences present in ev-2102. Shown are the 14 bp of direct repeat which recombine with the Hae III sequences in the 3' and 5' flanks of the viral Hae III-C. The Hae III sequence extends from nucleotide residues 1 to 235. A dashed line indicates that the Hae III sequence and the first of the complementary strand is inserted (212), the palindromes are indicated by solid lines. The dotted line above the arrowed sequences designates on (a) and (b) over the arrowed sequences over the arrowed sequences. The dot (-) overlines denotes sequence of alternating guanine and pyrimidine residues.

Fig. 7. Sequence of the portion of the monkey DNA inserted in ev-2102 which shares homology with the human consensus poly A sequence (11) and a monkey Alu-1 sequence. The poly A stretch consists of 12A (11), a dot indicates a poly A stretch while a letter indicates a nucleotide other than poly A. The boxed sequence is inserted adjacent to position 11 to keep the sequence in proper alignment. The insertions 14 bp direct repeats are denoted by horizontal lines (+), the conserved 9 bp palindrome by crosses (x), the 12A sequence by dashed line by asterisks, the A-rich sequence by a bracket (11) and a portion of the 2 bp insert of the 5' arm of the dimeric unit by dashes (---).

Fig. 8. Comparison of cellular DNA sequences derived from monkey -component DNA 17, present in ev-2110, ev-2102 (5,15), ev-2105 (5), ev-2106 (unpublished data), and CEP/TCP (28). The -component DNA sequence is written as two tandem repeats of the Hae III 172 bp monomer repeat unit. Underlined nucleotides indicate those nucleotides that differ from the + component repeat sequence, while dots indicate homologous residues.
Recombination in simian virus 40-infected cells. Structure of naturally arising variants ev-2114, ev-2102, and ev-1110.
L Sheflin, A Celeste and M Woodworth-Gutai


Access the most updated version of this article at http://www.jbc.org/content/258/23/14315

Alerts:
- When this article is cited
- When a correction for this article is posted

Click here to choose from all of JBC's e-mail alerts

This article cites 0 references, 0 of which can be accessed free at http://www.jbc.org/content/258/23/14315.full.html#ref-list-1