JOURNAL OF VIROLOGY, Nov. 1972, p. 317-325

Vol. 9, No. 2
Printed in U.S.A.

Sequence Heterogeneity in Closed Simian Virus 40 Deoxyribonucleic Acid

HENRY T. TAI, CHARLES ALLEN SMITH, PHILLIP A. SHARP,[1] AND JEROME VINOGRAD

Division of Biology and Division of Chemistry and Chemical Engineering, Norman W. Church Laboratory of Chemical Biology, California Institute of Technology, Pasadena, California 91109

Received for publication 12 October 1971

The heteroduplex molecules formed by self-annealing of denatured, singly nicked simian virus 40 (SV40) deoxyribonucleic acid (DNA) prepared from closed viral DNA were examined by formamide-protein film electron microscopy to test the DNA for sequence homogeneity. Sequence inhomogeneity appears in the heteroduplexes as single-strand loops. These result from sequence deletion or from sequence substitution, if regions greater than 50 nucleotides are involved. The undenatured DNA from viruses passaged twice at multiplicities of infection much less than 1 plaque-forming unit (PFU) per cell appeared to be homogeneous in size. The heteroduplexes formed by this DNA indicated that approximately 2% of the molecules carried deletions, but that substitutions were below the level of detection. In contrast, undenatured DNA from viruses grown by passaging undiluted lysates seven times or by infection with stock virus at a multiplicity of infection of 5 PFU per cell contained a large frequency of molecules shorter than the full length. The heteroduplex samples indicated that 12 and 7% of the undenatured material contained base substitutions, and 13 and 11% contained deletions. The deletions and substitutions appear to occur in separate molecules. Length measurements on heteroduplexes displaying the loop characteristic of substitutions have established that these molecules are from true sequence substitutions, and not from adjacent or overlapping deletions. More than 80% of the molecules carrying substitutions are shorter than the native SV40 length. On the average, the substituted sequence is about 20% of the length of SV40, but it replaces a sequence about 30% of the native length. The substituted sequences may be host cell nuclear DNA, possibly arising from integration of SV40 into the chromosome followed by excision of the SV40 DNA together with chromosomal DNA.

Simian virus 40 (SV40) is a small deoxyribonucleic acid (DNA)-containing virus which has been used extensively for the study of virus-cell interactions and as a model system for viral oncogenesis in animal cells. Infection of permissive host cells results in viral replication and culminates in cell death and lysis, releasing the newly made virus. Infection of nonpermissive host cells results in the alteration of cellular morphology and growth characteristics, a process termed *transformation*. That the viral DNA persists in the transformed cell has been shown by the production of whole virus after fusion of virus-free transformed cells and permissive cells (5-7), by the detection of viral DNA sequences in the transformed cell by nucleic acid hybridization (13, 15), and by the detection of viral-specific ribonucleic acid in such cells (1, 9, 11). Other experiments have suggested that the viral DNA persisting in the transformed cells is covalently linked with host cell nuclear DNA (15).

In the lytic infection with polyoma virus, a DNA virus very similar to SV40, it has been shown that some linear host cell DNA is encapsidated into virus particles (10, 16). These so-called "pseudovirions" have also been seen in SV40 infection of primary African green monkey kidney cells, but not with infection of BSC-1 cells, a permanent cell line derived from African green monkey cells (8).

Aloni et al. (2), investigating interactions between SV40 and the permissive BSC-1 cells, found hybridization between closed circular SV40 DNA and host cell nuclear DNA. They suggested that host cell DNA sequences might be

[1] Present address: Cold Spring Harbor Laboratory, Cold Spring Harbor, N.Y. 11724.

incorporated into SV40 viral DNA. Gelb, Kohne, and Martin (4) were unable to detect such hybridization using SV40 DNA from viruses grown by serial passage. However, the conditions used for passage were such that each infection was at low multiplicity (*personal communication*).

Yoshiike (17) has found that passage at high multiplicity produces noninfective particles containing closed circular DNA, heterogeneous in size and somewhat shorter than the DNA of infective particles. The foregoing results raise the possibility that host cell sequences are incorporated into DNA molecules heterogeneous in size.

To investigate this possibility, we have used electron microscopy to examine heteroduplex molecules artificially constructed from SV40 viral DNA recovered from viruses grown at low and high multiplicities. We have found that SV40 DNA from viruses grown at low multiplicity exhibits both length and sequence homogeneity. Like Yoshiike (17), we found that DNA from viruses grown at high multiplicity exhibits length heterogeneity. In addition, heteroduplexes constructed from such DNA contain both deletion loops and substitution loops. Measurements of molecules with substitution loops demonstrate that these loops are not the result of renaturation of single strands carrying overlapping or adjacent deletions. In most cases the substituted sequence is somewhat shorter than the sequence it replaces.

In a preliminary examination of heteroduplex molecules formed from laboratory stock SV40 DNA, R. Davis (Ph.D. dissertation, California Institute of Technology, 1970) observed molecules which contained deletion loops and apparent substitution loops.

MATERIALS AND METHODS

Cells. African green monkey kidney (BSC-1) cells were obtained from Flow Laboratories, Los Angeles, Calif. The cells were grown in Dulbecco's modified Eagle medium (Grand Island Biological Co., Berkeley, Calif.) supplemented with 10% calf serum.

Growth and isolation of low multiplicity SV40 virus. A non-plaque-purified laboratory stock of strain RH911 SV40 virus was banded in CsCl. The heavy end of the virus band, containing about 10% of the absorbance in the band, was collected and dialyzed against TD [0.14 M NaCl, 7 mM KCl, 7 mM Na_2HPO_4, 25 mM tris(hydroxymethyl)aminomethane (Tris), *p*H 7.5]. The dialyzed virus suspension was then titered by using the agar overlay procedure (14). BSC-1 monolayer cultures were infected at a multiplicity of infection of 10^{-1} plaque-forming units (PFU)/cell. After a full cytopathic effect (CPE) was observed, usually after 10 days, the viruses were harvested and purified as described below. The virus band in CsCl was collected, dialyzed, and titered. This virus was used to infect BSC-1 cells at a multiplicity of 10^{-2} PFU/cell, and, after a full CPE, the virus was purified and the DNA was extracted.

Growth and isolation of high multiplicity SV40 virus. Two independent stocks of high multiplicity virus were prepared. The undiluted lysate obtained with the low multiplicity virus (*see above*) was used to make the subsequent infection; this procedure was repeated six times to form high multiplicity stock (HM-A). In the other method, laboratory stock SV40 lysate was used to infect BSC-1 cells at a multiplicity of 5 PFU/cell, and the cells were harvested after 3 days (HM-B).

Purification of virus and preparation and purification of closed circular viral DNA. The infected cells were scraped from the plates and the entire suspension was frozen and thawed three times in plastic tubes. The suspension was homogenized with 10 strokes in a loose-fitting glass Dounce homogenizer. Sodium deoxycholate was added to a concentration of 1% and the solution was incubated at room temperature for 10 min. Crude deoxyribonuclease I and ribonuclease A were added to concentrations of 0.4 and 1.0 mg/ml, respectively, and the solution was incubated at room temperature for another 40 min. This mixture was centrifuged for 10 min at 10,000 rev/min in a Sorvall SS34 rotor, and 25-ml portions of the supernatant were layered over 15 ml of saturated KBr in Spinco SW27 cellulose nitrate tubes. These were centrifuged for 3 hr at 22,000 rev/min at 20 C. The opalescent virus bands, recovered by drop collection, were dialyzed against TD. The solution was madeto 10 mM $MgCl_2$ and crude deoxyribonuclease I was added to a final concentration of 5 mg/ml. After incubation at 37 C for 30 min, ethylenediaminetetraacetic acid (EDTA) was added to a concentration of 50 mM, and CsCl was added to a density of 1.3 g/ml. The solution was centrifuged for 20 hr at 36,000 rev/min in an SW50 rotor at 20 C. The virus band was recovered by drop collection. In some cases, the CsCl banding of the virus was repeated after another deoxyribonuclease treatment. After dialysis to remove CsCl, sodium dodecyl sulfate (SDS) was added to the virus suspension to a concentration of 1% and the solution was incubated at 37 C for 10 min. The solution was chilled to 0 C and CsCl was added to a concentration of 1 M. The mixture was maintained at 0 C for 10 min and then centrifuged for 10 min at 10,000 rev/min in an SS34 rotor at 4 C. The supernatant fluid was adjusted to contain 300 μg of ethidium bromide (EthBr) or propidium diiodide per ml, and CsCl was added to a density of 1.55 g/ml. The solution was centrifuged at 38,000 rev/min for 24 hr in an SW50 rotor at 20 C. The lower fluorescent band, which contains the closed circular DNA, was recovered by drop collection (12).

Preparation of singly nicked viral DNA. Three methods were employed. (i) Deoxyribonuclease: the nicking activity of a solution of deoxyribonuclease I with SV40 DNA as a substrate was assayed by analytical band velocity centrifugation through alkaline CsCl. Conditions were chosen such that 30% of the form I DNA was converted to form II. After nicking, the DNA was banded in CsCl-EthBr, and the two components were separated by drop collection. (ii)

X-irradiation: the DNA in 1 M CsCl, 1 mM EDTA, 2 mM histidine, 10 mM Tris (*p*H 8) was subjected to X rays (100 r/min) for 20 min. The form II DNA was separated from form I as above. (iii) Visible light-ethidium bromide: the DNA in 1 M CsCl, 1 mM EDTA, 10 mM Tris (*p*H 8), 70 μg of EthBr per ml was placed 8 cm from the lens of a slide projector containing a 500-w bulb and illuminated for 7 to 10 min. This converted 25% of the form I to form II. The components were separated as above and the process was repeated on the form I two times.

Formation of the heteroduplexes for electron microscopy. Approximately 0.2 μg of singly nicked SV40 DNA was denatured in 0.3 M NaOH for 10 min, neutralized with 2 M Tris-hydrochloride, and diluted to a total volume of 0.2 ml with 99% formamide. This sample was dialyzed for 2 hr against 50 ml of 99% formamide at room temperature. Unrenatured control samples were prepared by mixing 25 μliters of the above dialyzed solution with an equal volume of 0.2 M Tris, 0.02 M EDTA (*p*H 8.5) immediately prior to the spreading for preparation of the specimen grids. Heteroduplexes were formed by incubating a similar mixture for 5 hr at 25 C just prior to specimen grid preparation.

Electron microscopy. Both the aqueous and the formamide procedures for preparing samples for electron microscope examination have been described (3). Grids were stained and rotary-shadowed, and examined in a Phillips 300 electron microscope. Photographs on 35-mm film were traced on a Nikon 6F projection comparator. Lengths of molecules were determined with a map measurer with an accuracy of ±0.25 cm. Absolute lengths were determined by photographing a diffraction grating with each series of photographs.

Reagents. Ethidium bromide was a gift of the Boots Pure Drug Co. Ltd., Nottingham, England. Propidium diiodide was a gift of Calbiochem, Los Angeles, Calif. Optical grade CsCl was obtained from the Harshaw Chemical Co., Cleveland, Ohio. Deoxyribonuclease I and ribonuclease A were obtained from Sigma Chemical Co., St. Louis, Mo. Cytochrome *c* was obtained from Calbiochem. Formamide and SDS were obtained from Matheson, Coleman and Bell, Cincinnati, Ohio. All other chemicals were reagent grade.

RESULTS

Lengths of DNA from viruses grown at high and low multiplicities of infection. To compare our results with those of Yoshiike (17), length distributions of the DNA extracted from viruses grown at low multiplicity of infection (LM DNA) and at high multiplicity of infection (HM-A DNA) were determined (Fig. 1). The LM DNA was obtained as described earlier from infection at a multiplicity of 0.01 PFU/cell. The HM-A DNA was derived from serial passaging of undiluted lysates six times, starting with the low multiplicity virus stock. LM DNA exhibits a narrow length distribution with a mean of 1.7 μm. The length distribution of HM-A DNA is skewed

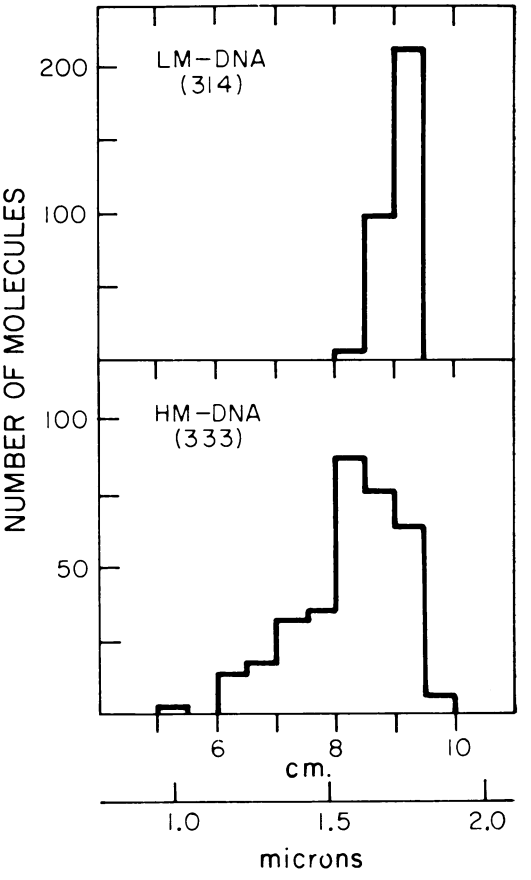

FIG. 1. *Length distributions of SV40 DNA from viruses grown at low multiplicity and at high multiplicity. Electron micrographs were traced and measured with a map measurer in centimeters, as shown. The micron scale is approximate; an external standard was used to determine the magnification. The number of molecules measured is given in parentheses.*

toward the short lengths. Both distributions cut off sharply at about the same upper length, presumably because of a limit on the size that can be encapsidated.

Molecular weight of LM DNA. The molecular weight of LM DNA, which has a narrow length distribution, was determined. This DNA was mixed with lambda-c_{26} DNA, and electron microscope grids were prepared by the aqueous technique. The lengths of 139 SV40 molecules and 15 lambda molecules were measured on photographs obtained with a single grid. The SV40 DNA had a mean length of 29.8 ± 1.3 cm; the lambda DNA had a mean length of 255.5 ± 7.4 cm. Based on the molecular weight of lambda, 30.8×10^6 daltons [N. Davidson and W. Szybalski. *In* A. D. Hershey and N. F. Dove (ed.), *The*

Bacteriophage Lambda, Cold Spring Harbor Laboratory, New York, *in press*], the molecular weight of the SV40 is $3.6 \pm 0.2 \times 10^6$ daltons. The uncertainty in the value for lambda brings our value to $3.6 \pm 0.3 \times 10^6$ daltons. The 4.4% standard deviation in SV40 length is that expected for a homogeneous DNA of this size (3).

Classification of the heteroduplexes. SV40 DNA, renatured so that about 35% of the DNA was in duplex form, was mounted for electron microscopy by the formamide technique and examined at a magnification of 60,000. The procedure allows both single- and double-stranded DNA to be visualized and distinguished. Both appear filamentous, but the single-stranded DNA is more irregular and thinner than duplex DNA and has been described as "knobby," "lumpy," and "kinky." In addition, its contrast is usually somewhat less than that of duplex DNA. The differences are readily seen at the electron microscope, where magnification and focus can be manipulated.

All molecules seen while systematically scanning a grid were classified into one of seven categories. The types of molecules placed into each category are diagrammed in Fig. 2 and some photographs of typical molecules are shown in Fig. 3. The majority of the DNA strands were undegraded. Forty per cent of the single strands were circular. The categories were: (a) linear single-stranded DNA; (b) circular single-stranded DNA; (c) heteroduplexes without single-stranded loops; (d) heteroduplexes with one or two deletion loops, a loop of single-strand DNA tangent at one point to a duplex region; (e) one or two substitution heteroduplexes with loops, regions where two single strands replace the duplex structure along a portion of its length; (f) heteroduplexes containing a forked structure at which the duplex separates into separate single-strands—these could arise from a single-strand scission in a substitution loop or in a deletion loop; (g) heteroduplexes containing both single- and double-stranded regions, which are knotted, tangled, or so complex as to make assignment impossible. These occur in both LM and HM DNA samples, which vary widely in their sequence heterogeneity.

The data might be significantly influenced by the failure of all the DNA to denature before renaturation was allowed to occur. In some cases, specimen grids were prepared with the material obtained just prior to the last incubation step as described earlier. Almost all of the DNA appeared to be single-stranded and, therefore, denatured. Furthermore, the small amount of renaturation which did occur with the HM-A DNA sample led to some heteroduplexes containing nonhomology regions. The numbers of molecules in each category in the renatured samples were corrected for the frequency of occurrence of these forms in the zero-time sample. These corrections were always small and well within the range of results found upon scoring different, renatured materials from the same sample. Nonrenatured samples contained 60% single-stranded linear and 40% single-stranded circular molecules, as expected for singly nicked DNA.

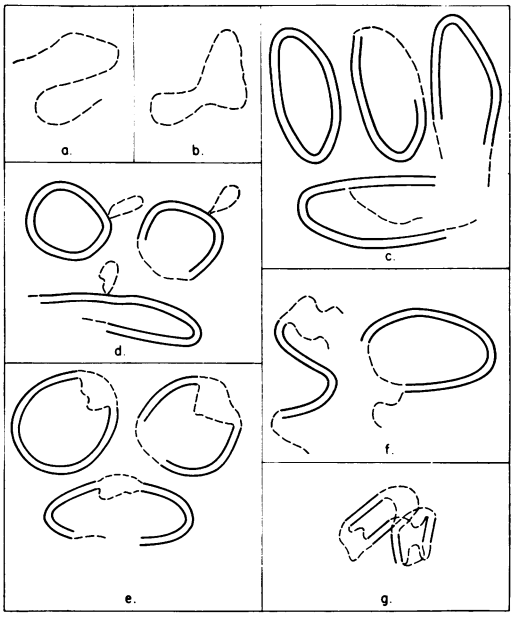

FIG. 2. *Categories of molecules used in classifying the DNA in the renatured samples. Solid lines, duplex DNA; dashed lines, single-stranded DNA.*

The frequencies of the various kinds of molecules containing duplex regions are presented in Table 1. Two separate denaturation-renaturations were carried out with LM DNA, and two spreadings were performed with each. The LM DNA heteroduplexes contained no detectable substitution loops in a total of more than 1,000 molecules examined. About 4% of the heteroduplexes contained deletion loops and 1% contained forked structures. A molecule containing a deletion gives rise to two heteroduplexes, each with a deletion loop. The probability of annealing two defective strands is small in this case. The above result indicates that 2.5% of the LM DNA molecules contain deletions.

Two samples of HM DNA were prepared, one by serial passage (sample A) and one by a single infection (sample B). The sequence heterogeneity

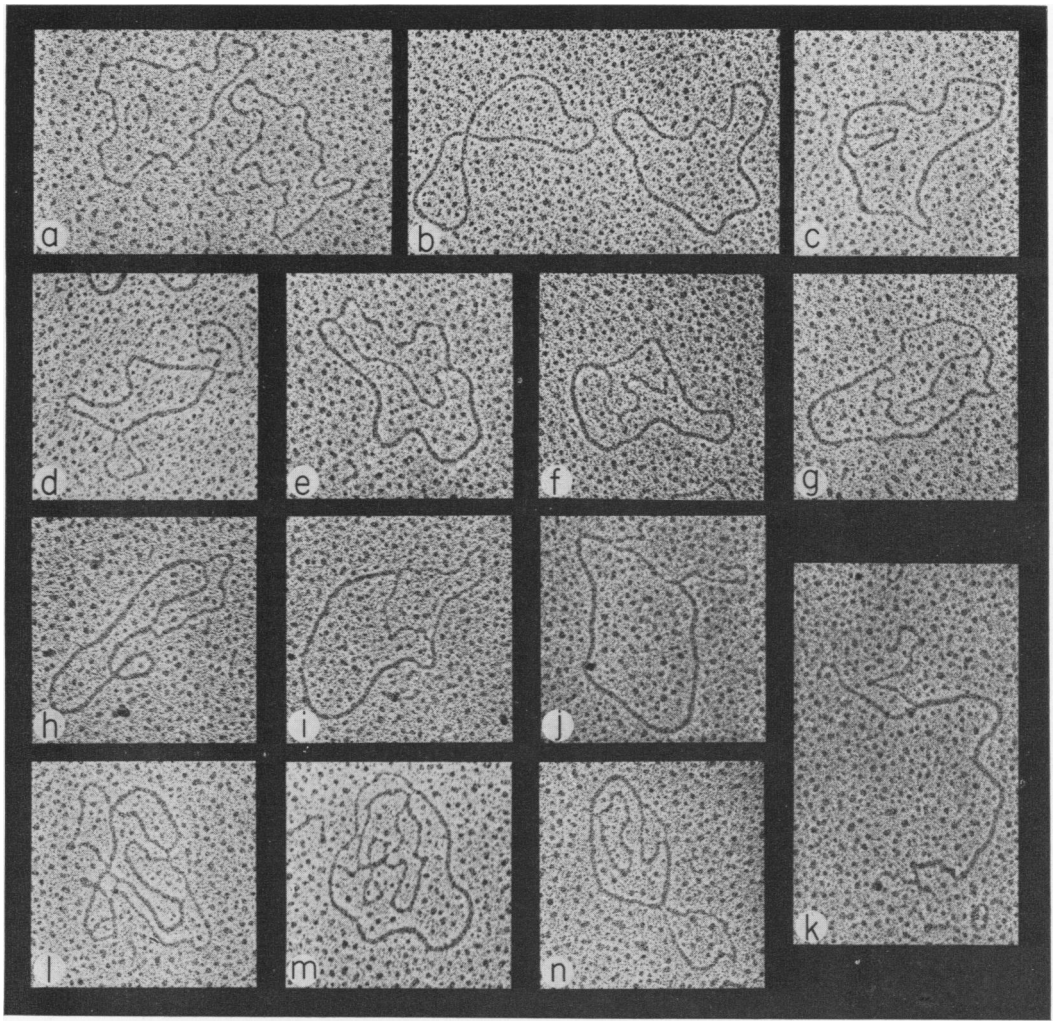

FIG. 3. *Electron micrographs of molecules in the renatured HM DNA sample. (a) Single-stranded circles. (b) Fully duplex circles. (c) Heteroduplex with a deletion loop. (d) Heteroduplex with two deletion loops. (e–i) Heteroduplexes with substitution loops. (j) A circular heteroduplex containing forks. (k) A linear heteroduplex with a fork. (l–m) Heteroduplexes with both single and double strands which are not interpretable. (n) A heteroduplex with two deletion loops and a substitution loop.*

TABLE 1. *Classification of heteroduplexes formed by self-annealing of high and low multiplicity simian virus 40 viral DNA*[a]

Sample[b]	Fully double-stranded (%)	Deletions (%)	Substitutions (%)	Forks (%)	Unscorable (%)	No. of molecules scored
LM DNA						
1a	95	5	0	0	0	75
1b	93	4	0	0	3	311
2a	91	2	0	2	5	324
2b	84	4	0	1	11	460
HM DNA, sample A						
3a	39	15	15	21	10	326
3b	41	18	17	13	11	306
HM DNA, sample B						
4a	52	16	8	13	11	243
4b	57	17	6	12	8	299
4c	59	13	12	8	8	134

[a] Each row in the table presents the results obtained with one spreading. Each number in the first column refers to one denaturation-renaturation experiment. Zero-time corrections were applied to experiments 1 and 4.

[b] LM DNA, DNA extracted from viruses grown at a low multiplicity of infection; HM DNA, high multiplicity of infection.

in the high-multiplicity samples, A and B, is shown by the high frequency of heteroduplexes containing deletions or substitutions. This is very apparent in comparison with the frequencies obtained with LM DNA. Disregarding the forked heteroduplexes, we calculate from the data in Table 1 that 8% of the HM-A molecules contained deletions, and 8% contained substitutions. Eight per cent of the HM-B molecules contained deletions and 4% contained substitutions. A better estimate is obtained by assigning half the forked molecules to deletions and half to substitutions. When this is done, HM-A DNA contains 13% deletions and 12% substitutions, and HM-B contains 11% deletions and 7% substitutions.

Nature of the substitution loops. A substitution loop seen in a heteroduplex molecule indicates that the two single-stranded regions unable to form duplex DNA have less than some critical amount of sequence homology. Such a structure could arise from a molecule which has undergone a deletion followed by an insertion of a segment of DNA with a different sequence. The two events need not be separate in time; for example, an inversion produced by an intramolecular recombination event would produce a substitution loop. A substitution loop could arise if each strand in the heteroduplex contains a deletion and the deletions are within a few base pairs of each other or overlap. The nature of the substitution loops encountered here was determined by measuring lengths of circular molecules that were completely double-stranded except for the loop. If the loop were due to a substitution, one strand of the heteroduplex would be of the native length. Therefore, the sum of the duplex part of the molecule and one of the single-stranded segments should equal that length. However, if the heteroduplex is formed by renaturation of two strands, each containing a deletion, the sum of the duplex and either of the single-stranded segments must be shorter than the native length. Also for overlapping deletions, the sum of the duplex and both single-stranded segments can never exceed the native length, and the sum becomes smaller as the size of the deletion-overlap becomes larger. The comparable sum for a substitution will always be longer than the native length.

To compute these sums, the lengths of single-stranded circles, duplex circles, and circles containing substitution loops on a single grid prepared from HM-A DNA were measured. Similar measurements were made of the single-stranded and the duplex circles in the LM DNA sample for comparison.

It would be expected that the complete duplex circles seen in the HM-A DNA would arise from the renaturation of two full-length native strands. That this is the case can be seen from the fact that the mean of the length distribution of these molecules (Fig. 4a) is very close to the mean of the distribution of lengths of complete duplex circles from the LM DNA. The small difference may be due to the fact that these measurements were necessarily made on grids from two separate spreading operations.

The distribution of lengths of single-strand

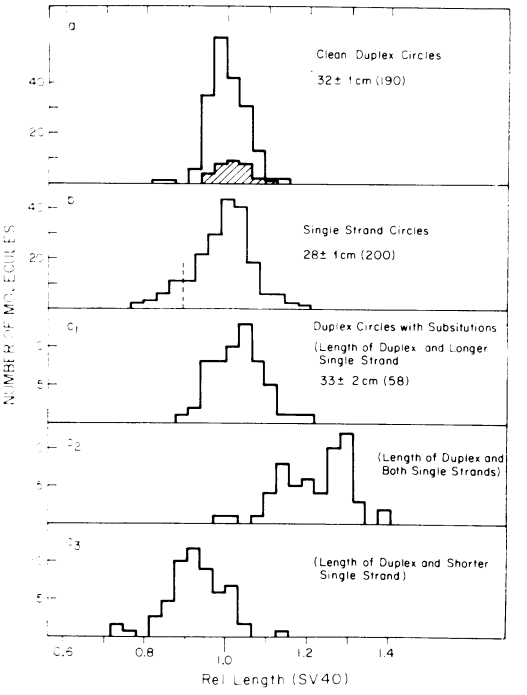

FIG. 4. *Length distributions of molecules in the renatured HM-A DNA samples. Lengths are expressed as fractions of the full SV40 length. Mean lengths of the distributions and the standard deviations are given in units obtained with the map measurer together with the number of molecules measured. The distribution was truncated by the dashed line in (b) for the calculation of the mean and standard deviation. The truncation reduces the effect of defective molecules in the calculation of the mass per unit length of single-stranded DNA relative to duplex DNA. (a) Length distribution of fully duplex circles. The shaded histogram is for the LM DNA sample. (b) Length distribution of single-stranded circles. (c_1 – c_3) Heteroduplexes with clean substitution loops. Lengths of the duplex portions and single strands in the substitution loop were measured. After the single-strand lengths were corrected for the molecular weight per unit length, the indicated sums were computed. (c_1) Sum of duplex and longer single strand. (c_2) Sum of duplex and both single strands. (c_3) Sum of duplex and shorter single strand.*

circles (Fig. 4a) peaks at a slightly smaller length, due to a smaller length per unit mass, and tails off to the short side. The tailing is due to the defectiveness of the sample (see Fig. 1b) and was not observed in the single-stranded preparation of LM DNA. The mean length of molecules greater than 25 cm was taken as the mean single-stranded length. (It compares well with the mean of lengths of the single-stranded circles for the LM DNA, 29.4 cm.) Single-stranded DNA lengths were multiplied by the ratio 32/28 to reduce the data to units of length per genome.

On a single specimen grid, the lengths of the duplex part and each single-stranded part of 58 heteroduplexes containing substitution loops were measured. The distribution for the sum of lengths of the duplex part and larger single-stranded part (Fig, $4c_1$) is similar to that of the lengths of duplex circles (Fig. 4a). The means of the distributions were within one standard deviation. When the lengths of both single strands are added to the duplex length, the distribution is wide (Fig, $4c_2$) and the lengths in the entire distribution are substantially longer than the mean value for the duplex lengths. This argues strongly that the loops are indicative of substitutions and not of overlapping deletions in the heteroduplexes.

Sequence lengths involved in the substitution process. Since the mean length of the longer strand in the heteroduplexes containing substitutions is the native length, we consider it to be native sequence, and the shorter strand to be the one containing the substituted sequence. The longer single-stranded segment length represents the size of the SV40 sequence removed from the substituted molecule prior to insertion of the new nonhomologous sequence. The shorter single-stranded segment length represents the length of the inserted sequence. The length distribution of the molecules containing substitutions is shown in Fig. $4c_3$. Here, the length of the duplex segment of the heteroduplex has been added to the length of the shorter single-stranded segment. The mean length of the substituted molecule is 11% shorter than native SV40.

The deleted sequence ranges from 10 to 60% of the SV40 length, with a mean of 31% (Fig. 5a). The inserted sequence ranges from 5 to 50% of the SV40 length, with a mean of 20% added (Fig. 5b). The distribution of the difference in lengths between the two single-stranded segments of each heteroduplex (Fig. 5c) shows that the difference is not constant, but varies widely: 83% of the molecules show some difference. Neither is the difference a constant fraction of the sequence deleted. The average difference in length between the sequence deleted and that inserted is 11% SV40. Thus, the reduced size of defective DNA is due both to simple deletions and to deletion-substitutions.

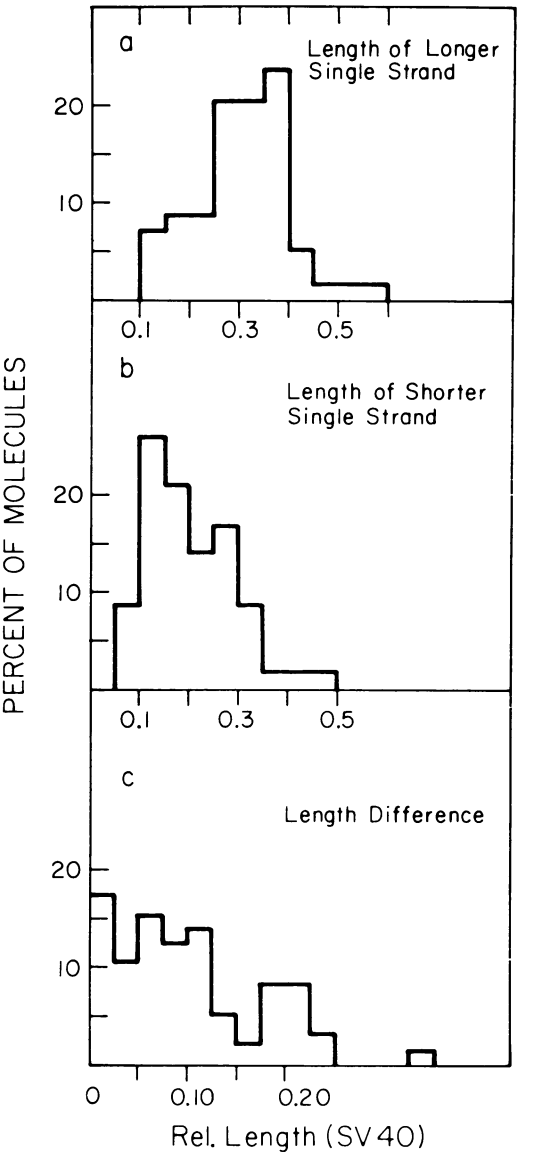

FIG. 5. *Distributions of the lengths of the substitution loops obtained from the measurements used to prepare the distribution in Fig. $4c_1$–c_3. The lengths are expressed as fractions of the full SV40 length. (a) Distribution of the lengths of the longer single-strand portions of the loops. (b) Distribution of the length of the shorter single strands in the loops. (c) Distribution of the difference in length between the two single strands in each heteroduplex.*

DISCUSSION

We have demonstrated that 7 to 12% of the closed circular DNA molecules from SV40 viruses grown at high multiplicity of infection contain a region in which a new base sequence has replaced the native one. These new sequences are detected as nonhomology regions in heteroduplexes. These sequences could be SV40 sequences from the same molecule, SV40 sequences from other SV40 DNA molecules, or sequences from host cell DNA.

A substitution might be generated intramolecularly by a reciprocal recombination event involving two presumably homologous sites in the molecule. Depending upon which pairs of strands were rejoined, the sequence between the sites would either be deleted, or would be inverted. Such an inversion would give rise to an apparent substitution in a heteroduplex. This explanation seems unacceptable, since the molecules with substitutions are found to be somewhat shorter than the native molecules and because the size of the substitutions varies, making multiple recombination sites necessary.

Two recombination events could delete a sequence from one SV40 molecule and replace it with a sequence from another SV40 molecule. If the two sequences bounded by the recombination sites were different, a substitution would result. The lengths of the sequences need not be identical.

Host cell DNA might be incorporated into SV40 DNA by a mechanism similar to production of lambda-transducing phage. During the lytic infection, SV40 viral DNA would become integrated at one or more sites into host cell DNA, as has been suggested for the transformation process. Subsequent excision would occur by recombination at a site within the viral DNA and a site somewhere on the cellular DNA. A molecule which has lost an SV40 sequence and gained a cellular DNA sequence results. The loss and gain need not be of the same size.

If integration of SV40 DNA into host cell chromosomes does exist and occurs at a common site on all SV40 DNA molecules, such a model predicts that the substitutions will all share a common end point. Because SV40 DNA is circular, the position of the deletions or substitutions in the genome cannot be determined in the absence of a marker. We do not know if the simple deletions that were seen are produced by a mechanism involving interaction with host cell DNA. Heteroduplexes have been observed which contain two non-overlapping deletion loops. If these heteroduplexes were formed from strands each containing a deletion, then at least two deletion sites are involved. We cannot rule out the possibility that these heteroduplexes are composed of one native strand and one strand containing both an inserted sequence and a deletion.

Lavi and Winocour (*manuscript submitted for publication*) have demonstrated that hybridization of SV40 viral DNA to BSC-1 host cell DNA is dependent on the conditions of infection. When non-plaque-purified virus was used at very low multiplicity of infection values, virtually no hybridization was observed. This was also the case when plaque-purified virus was used at a variety of multiplicity of infection values. Infection of non-plaque-purified virus at high multiplicity of infection led to viral DNA which hybridized significantly with host cell DNA. Plaque-purified virus grown by serial undiluted passage resulted in viral DNA which hybridized to a greater extent with each passage. These results suggest that the substituted sequences we have detected are host cell DNA.

If we take the frequency of molecules containing substitution to be 12% for HM-A DNA and the average substituted sequence to be 0.2 genomes, then substituted sequences comprise 2.4% of the DNA. It should be noted that our numbers are minimal estimates; molecules which had long substitutions would not have renatured to form heteroduplexes as fast as those with smaller substitutions. Large substitutions would also have been more susceptible to nicking, which would place them into the forked category. The rare heteroduplexes which appeared to have only a very short duplex region were not included in the molecules measured.

A quantitative comparison between the result obtained here for the proportion of substituted sequences, 2.4%, and the results of Lavi and Winocour is not possible because the proportion of chromosomal DNA is not directly calculable from their hybridization data.

Passaging the laboratory stock virus twice at low multiplicity effectively removed substituted molecules and substantially reduced the number of molecules containing deletions. This suggests that the substitutions, as well as the deletions, remove genes essential for a productive lytic infection. When the multiplicity of infection is low, any such defective molecules entering a cell are lost. The small number of deletions seen in the LM DNA may be due to deletions in nonessential regions of the genome or may represent the rate of production of deletions in the infection process.

The exact role of the high multiplicity of infection in the production of substitutions and deletions is not yet established. Certainly, infection at high multiplicity with a stock of virus containing

such molecules would allow them to replicate. Shorter molecules may even be selected, if their replication time is shorter or if their encapsidation is more easily accomplished than full-length molecules. The rate of production of substituted and deleted molecules may be constant with each infection, and independent of the multiplicity of infection. Alternatively, the rate may be a function of the multiplicity. A threshold multiplicity is perhaps needed for the substitution process to occur at all.

The suggestion from this study that cellular DNA is integrated into the viral genome during the lytic cycle makes it tempting to suggest that there may be fewer differences between the lytic and transformational events induced by SV40 than once thought. Our studies point out the desirability of examining SV40 stocks for substitutions when studying virus-cell interactions and viral transformations of cells.

ACKNOWLEDGMENTS

We thank S. Lavi and E. Winocour for sending us prior to publication a manuscript describing their study of hybridization of SV40 DNA to host cell DNA.

We thank Marlyn Teplitz and Jean Edens for technical assistance.

This investigation was supported by Public Health Service research grant CA08014 from the National Cancer Institute and GM15327 from the National Institute of General Medical Sciences. One of us, H. T. T., was supported by an American Cancer Society Postdoctoral Research Scholarship PRS-44. This is contribution no. 4347 from the Division of Chemistry and Chemical Engineering.

LITERATURE CITED

1. Aloni, Y., E. Winocour, and L. Sachs. 1968. Characterization of the simian virus 40-specific RNA in virus-yielding and transformed cells. J. Mol. Biol. **31**:129–145.
2. Aloni, Y., E. Winocour, L. Sachs, and J. Torten. 1969. Hybridization between SV40 DNA and cellular DNA's. J. Mol. Biol. **44**:333–345.
3. Davis, R. W., M. Simon, and N. Davidson. 1971. Electron microscope heteroduplex methods for mapping regions of base sequence homology in nucleic acids, p. 383–412. *In* L. Grossman and K. Moldave (ed.), Methods in enzymology, vol. XXI. Academic Press Inc., New York.
4. Gelb, L. D., D. E. Kohne, and M. A. Martin. 1971. Quantitation of simian virus 40 sequences in African green monkey, mouse and virus-transformed cell genomes. J. Mol. Biol. **57**:129–145.
5. Jensen, F. C., and H. Koprowski. 1969. Absence of repressor in SV40 transformed cells. Virology **37**:687–690.
6. Knowles, B. B., F. C. Jensen, Z. Steplewski, and H. Koprowski. 1968. Rescue of infectious SV40 after fusion between different SV40-transformed cells. Proc. Nat. Acad. Sci. U.S.A. **61**:42–45.
7. Koprowski, H., F. C. Jensen, and Z. Steplewski. 1967. Activation of production of infectious tumor virus SV40 in heterokaryon cultures. Proc. Nat. Acad. Sci. U.S.A. **58**:127–133.
8. Levine, A. J., and A. K. Teresky. 1970. Deoxyribonucleic acid replication in simian virus 40-infected cells. II. Detection and characterization of simian virus 40 pseudovirions. J. Virol. **5**:451–457.
9. Martin, M. A., and D. Axelrod. 1969. SV40 gene activity during lytic infection and in a series of SV40 transformed mouse cells. Proc. Nat. Acad. Sci. U.S.A. **64**:1203–1210.
10. Michel, M. R., B. Hirt, and R. Weil. 1967. Mouse cellular DNA enclosed in polyoma viral capsids (pseudovirions). Proc. Nat. Acad. Sci. U.S.A. **58**:1381–1388.
11. Oda, K., and R. Dulbecco. 1968. Regulation of transcription of the SV40 DNA in productively infected and in transformed cells. Proc. Nat. Acad. Sci. U.S.A. **60**:525–532.
12. Radloff, R., W. Bauer, and J. Vinograd. 1967. A dye-buoyant density method for the detection and isolation of closed circular duplex DNA: the closed circular DNA in HeLa cells. Proc. Nat. Acad. Sci. U.S.A. **57**:1514–1521.
13. Reich, P. R., P. H. Black, and S. M. Weissman. 1966. Nucleic acid homology studies of SV40 virus-transformed and normal hamster cells. Proc. Nat. Acad. Sci. U.S.A. **56**:78–85.
14. Rush, M. G., R. Eason, and J. Vinograd. 1971. Identification and properties of complex forms of SV40 DNA isolated from SV40-infected African green monkey (BSC-1) cells. Biochim. Biophys. Acta **228**:585–594.
15. Sambrook, J., H. Westphal, P. R. Srinivasin, and R. Dulbecco. 1968. The integrated state of viral DNA in SV40-transformed cells. Proc. Nat. Acad. Sci. U.S.A. **60**:1288–1295.
16. Winocour, E. 1967. On the apparent homology between DNA from polyoma virus and normal mouse synthetic RNA. Virology **31**:15–28.
17. Yoshiike, K. 1968. Studies on DNA from low-density particles of SV40. I. Heterogeneous defective virions produced by successive undiluted passages. Virology **34**:391–401.