Proc. Nat. Acad. Sci. USA
Vol. 71, No. 9, 3497–3501, September 1974

Sequence Arrangements in Clonal Isolates of Polyoma Defective DNA

(genetic deletions/electron microscopy/restriction endonuclease/DNA replication/episomes)

DONALD L. ROBBERSON* AND MIKE FRIED†

Department of Tumor Virus Genetics, Imperial Cancer Research Fund, Lincoln's Inn Fields, London WC2A 3PX

Communicated by Renato Dulbecco, May 20, 1974

ABSTRACT Five clonal isolates of purified polyoma defective DNA [Fried, M. (1974) *J. Virol.* 13, 939–946] have been examined by electron microscopy. Four isolates (D-92, D-74, D-50, and D-47) are largely homogeneous in sequence, whereas, the fifth isolate (D-80) is somewhat heterogeneous.

Polyoma nondefective DNA is cleaved in a unique region of the genome by the EcoR$_I$ endonuclease. Hybridization of the resulting linear molecules with randomly nicked defective DNA reveals distinguishable types of heteroduplex structures for each of the different defective DNAs. Although the defective DNAs are shorter than polyoma nondefective DNA, the heteroduplex experiments demonstrate that they are not simply deletion mutants containing only a portion of the viral genome. Three isolates (D-92, D-50, and D-47) contain regions of homology to polyoma DNA covalently linked to nonhomologous regions. One isolate (D-74) contained no regions of detectable homology to polyoma DNA. Another isolate (D-80) contained a large proportion of molecules with duplicated-inverted regions. Some of these isolates of defective DNA may contain specific host sequences at the site(s) of integration of the polyoma genome during the lytic cycle in mouse cells. A process we term "abortive replication" may explain the formation of different types of defective DNAs.

Infection of appropriate sensitive cells with polyoma virus or simian virus 40 (SV40) at high input multiplicities results in the intracellular formation of closed circular DNAs with shortened, heterogeneous lengths (1–3). Some of these shorter (defective) DNAs can be packaged by viral capsid proteins. These particular encapsulated DNAs come to predominate with repeated passage of the virus at high input multiplicity and are unable to form plaques (1, 4, 5). Application of the infectious center technique to a mixed population of defective and nondefective polyoma virus can yield clones of cells simultaneously infected with one or more classes of encapsulated defective DNA and at least one encapsulated nondefective DNA (1). Clonal purification of different members of the population of defective DNAs by this procedure allows analysis of their DNA sequences by electron microscopy.

MATERIALS AND METHODS

Polyoma defective DNAs from individual clonal isolates were separated from nondefective polyoma DNA as supercoiled molecules, by banding in ethidium bromide–CsCl after treatment with EcoR$_I$ (restriction endonuclease R$_I$ purified from *Escherichia coli*) as previously described (1). Those defective DNAs that are less than 80% the length of polyoma nondefective DNA (Py DNA) were further purified by sedimentation velocity in neutral sucrose gradients (1). The five clonal isolates of defective DNA (D-DNA) analyzed are designated D-47, D-50, D-74, D-80, and D-92 and possess average lengths which are 47, 50, 74, 80, and 92% of the length of Py DNA. These correspond, respectively, to clonal isolates PY-DEL-5-(21-4), PY-DEL-5-(21-3), PY-DEL-5-(22-3), TS-A-DEL-6-2, and PY-DEL-2-3, described in the earlier study (1).

The EcoR$_I$, a gift from Dr. W. R. Folk, was prepared by a modification (6) of the method of Yoshimori (7). Py DNA prepared as described (1), was treated with EcoR$_I$ under conditions which we refer to as standard digestion (1, 6) which convert 99% of the circular DNA to a linear form (Py/R$_I$-DNA). The excess digestion of Py DNA (Fig. 1) was achieved by increasing the ratio of enzyme to DNA by 20-fold in the reaction and conducting the digestion at reduced ionic strength (10 mM $MgCl_2$–0.01 M Tris·HCl at pH 7.7 for 15 min at 37°).

Samples of closed circular D-DNA dialyzed against STE [100 mM NaCl–10 mM ethylenediaminetetraacetate (EDTA) —50 mM Tris·HCl at pH 8.5] were irradiated with x-rays (8) such that 30–50% of the closed circles were nicked. The nicked DNA samples were then examined by the aqueous Kleinschmidt technique (9). Contaminating host DNA accounted for less than 2% of the mass of the D-DNA preparations. When the sizes of the DNAs were determined relative to Py DNA by inclusion of the internal length standard of nicked PM2 DNA (taken as 9900 nucleotide pairs), the length of Py DNA was determined to be 5200 ± 200 nucleotide pairs.

Heteroduplexes were constructed by mixing approximately 10 μl of a solution containing 0.015–0.030 μg of randomly nicked D-DNA in STE with 10 μl of a solution containing 0.0075–0.015 μg of Py/R$_I$DNA (standard digestion) and dialyzed into STE. The solution was thermally denatured by immersion in boiling water for 20 sec and quenched in ice water for 30 sec. The sample was then diluted by the addition of 70 μl of formamide and immediately dialyzed against a solution of 50% formamide, 20 mM EDTA, 0.2 M, Tris·HCl at pH 8.5, for 70–90 min at 21–23°. After dialysis, cytochrome *c* (1 mg/ml) was added to a final concentration of 0.05 mg/ml and the sample spread essentially as described (10) using the formamide modification of the Kleinschmidt technique (9). Self-renaturation of the randomly nicked D-DNAs was

Abbreviations: EcoR$_I$, restriction endonuclease R$_I$ purified from *Escherichia coli;* D-DNA, defective DNA resistant to cleavage by EcoR$_I$; Py DNA, circular nondefective polyoma DNA; Py/R$_I$DNA, Py DNA linearized by cleavage with EcoR$_I$.

* Present address: Department of Biology, The University of Texas System Cancer Center, M. D. Anderson Hospital and Tumor Institute, Houston, Texas 77025.

† To whom reprint request should be addressed.

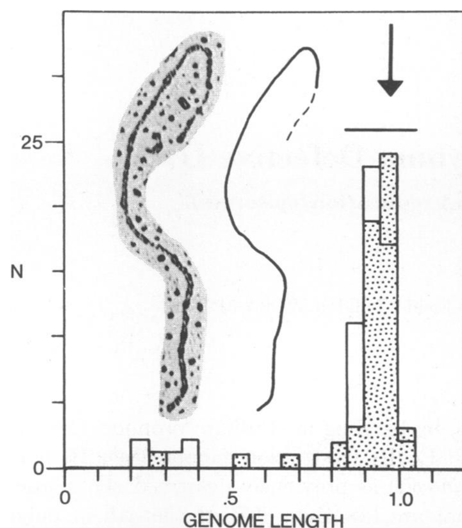

FIG. 1. Frequency distribution of linear duplex molecules after cleavage of Py DNA with EcoR$_I$ enzyme. Stippled, standard digestion conditions; unstippled, excess enzyme conditions; N, number of molecules; genome length is expressed as a fraction of the mean length of Py DNA. The average length of the linear DNA produced under standard conditions (indicated by arrow) is slightly larger than that produced under excess enzyme conditions. Denaturation-renaturation of the standard digestion linear DNA results in about half of the homoduplexes having one single-stranded terminus (micrograph inset and accompanying line drawing with single-strand segment indicated by dashed line and duplex segment indicated by solid line). The bar indicates the intervals over which the average lengths were computed.

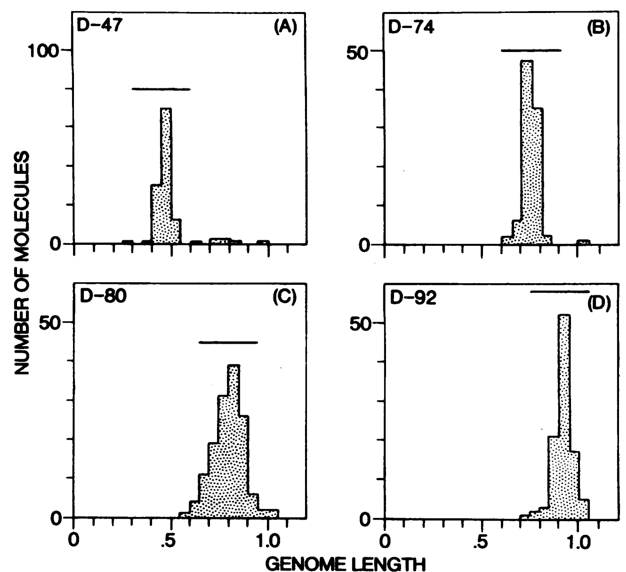

FIG. 2. Frequency distributions of clonal isolates of defective DNAs with average lengths of 47 ± 3% (A), 74 ± 3% (B), 80 ± 6% (C), and 92 ± 3% (D) of the length of Py DNA. An additional clonal isolate (D-50) possesses an average length of 50 ± 3% (measurement of 120 molecules). The expected standard deviations for homogeneous duplex DNAs with these average lengths would be ±2.5% (D-47), ±2.7% (D-50), ±3.2% (D-74), ±3.3% (D-80), and ±3.6% (D-92) (9). The bar indicates the intervals over which the average lengths were computed.

conducted in exactly the same manner with substitution of an equivalent volume of STE for the Py/R$_I$DNA solution. The hybridization conditions permit renaturation of approximately 40% of the linear single strands of Py R$_I$DNA, but probably do not permit the detection of sequence homologies of less than about 400 nucleotide pairs in length.

RESULTS

Cleavage of Py DNA by EcoR$_I$. Polyoma nondefective DNA is cleaved to duplex linear molecules by EcoR$_I$, whereas the clonal isolates of defective DNAs are resistant to cleavage by this enzyme (1). Standard digestion of Py DNA with EcoR$_I$ results in the production of linear molecules with an average length 98 ± 3% that of the circular genome (Fig. 1). This cleavage occurs in a unique region of the genome as demonstrated by the recovery after denaturation-renaturation of more than 90% of the renatured molecules as duplex linear forms. Analysis of the denatured-renatured molecules revealed that 47% contain a short single-stranded segment at one end (Fig. 1), whereas, only 8% contain single-stranded regions at both ends.

Further digestion of Py DNA with excess EcoR$_I$ results in the production of duplex linear molecules with an average length 94 ± 4% that of the circular genome (Fig. 1). Denaturation-renaturation of these molecules results in the formation of linear duplex forms with an average length 95 ± 4% that of the circular genome, but with no evidence of single-stranded ends.

An explanation for these results is that the EcoR$_I$ preparation contains an additional nuclease activity which cleaves Py DNA within a region which is 5% of the genome length from the EcoR$_I$ cleavage site. When standard digestion conditions are used, about half the molecules are shortened. Under conditions of excess enzyme and reduced ionic strength of the digestion medium, nearly all the molecules are shortened. This observation suggests a basis for the orientation of the linear single strands of Py/R$_I$DNA in heteroduplexes with defective DNAs (see below). An additional nuclease activity in an EcoR$_I$ preparation subsequently found to cleave near the EcoR$_I$ site in fragment 2 of the physical map (11), is probably the same activity as observed here.

Size and Homogeneity of Defective DNAs. The randomly nicked clonal isolates of defective DNAs examined in this study possess circular contour lengths which range from 47% (D-47) to 92% (D-92) of the contour length of Py DNA (Fig. 2). With the single exception of D-80 DNA (Fig. 2C), the length distributions of these DNAs have standard deviations that are no larger than would be expected for duplex DNAs with these lengths (9).

Except for D-80, homoduplexes of randomly nicked D-DNAs reveal few molecules with inhomologies. About 3% of the molecules in the self-renatured samples of D-47, D-50, and D-74 DNAs, and about 12% in D-92 showed a single inhomology region; single regions of inhomology were observed in 3% of the linear self-renatured molecules of Py/R$_I$DNA. Deletion loops (12) were found in only 1-2% of the renatured forms of each species of these D-DNAs. Partially renatured molecules with multiple inhomologies were also infrequently found (1-3%) and probably represent the incomplete renaturation of two intact circular single strands. These results indicate that D-47, D-50, D-74, and D-92 DNAs are each mostly homogeneous, but that D-92 DNA does contain at least one minority population of molecules with a different sequence. The average length of the base-

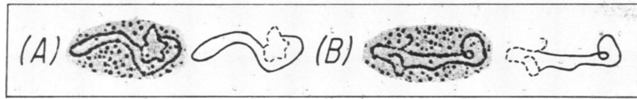

FIG. 3. Homoduplexes of D-92 DNA which showed a region of inhomology where both of the single strands of the inhomology were intact (A) and where one of the single strands was nicked (B). Single-stranded and duplex regions are as in Fig. 1.

paired region in the self-renatured molecules of D-92 DNA containing an inhomology region (Fig. 3) was 74 ± 7% of the length of Py DNA.

Heteroduplex Structures of Py and defective DNA. Heteroduplexes between the single strands of randomly nicked D-DNA and Py/R_IDNA can be specifically recognized as a circular single strand of D-DNA hybridized to a linear single strand of Py/R_IDNA. Such heteroduplex structures have been detected with the circular single strands of D-50 and D-92 DNAs (Fig. 4A and B). (The heteroduplex structure formed between D-47 DNA and Py/DNA is essentially equivalent to that between D-50 DNA and Py/R_IDNA.) Approximately 90% of all Py/R_I heteroduplex structures with D-50 DNA and D-92 DNA are the type shown in Fig. 4A and B, respectively. Few if any heteroduplexes were formed between D-74 and Py/R_I or D-80 and Py/R_I.

Because there is loss of DNA from one end of some molecules of Py/R_I DNA due to an additional nuclease present in our preparations of $EcoR_I$ (see above), it is possible to determine exactly which part of Py DNA is present in each class of D-DNA. Thus, D-50 × Py/R_I heteroduplexes have one long and one short single-strand tail (Fig. 4A); the distribution of lengths of the long tails is greater than expected (9), whereas, that of the short tails is not; therefore, the long tail must be the end of Py DNA that is variably subject to loss. Similarly, in D-92 × Py/R_I heteroduplexes, the single-stranded tail must be the other end because its length is not abnormally variable whereas the length of the duplex region is variable. Hence, in comparing D-50 × Py/R_I and D-92 × Py/R_I heteroduplex structures they should be oriented, with respect to each other, in the way shown in Figure 4.

Based on this orientation, D-92 would be expected to hybridize to a sequence of D-50 DNA which is approximately 6% of the length of Py DNA (over a region extending from 28 to 34% of the length of Py DNA measured from the left in Fig. 4A and B). This amount of homology is just under the level of detection for the heteroduplexing technique we have used. Heteroduplexes between D-50 DNA and D-92 which were detected (Fig. 4C) exhibit a region of homology which is 16 ± 2% the length of Py DNA. Only a low frequency (3%) of such heteroduplexes were found in an experiment in which 39% of the D-92 and 58% of the D-50 molecules formed homoduplexes. These rare heteroduplexes probably reflect a region of homology shared between D-50 and a minor population of D-92 DNA, possibly the same minor population as noted earlier (Fig. 3). A minor population of heteroduplexes is also found between D-92 and Py/R_IDNA (Fig. 5B) in which the region of homology is also approximately 16% the length of Py DNA.

It is possible that the regions of D-50 and D-92, that are not homologous to Py DNA, could represent an inverted or repeated sequence of viral DNA. With D-92, a few heteroduplexes are indeed detected which have the structure expected for hybridization of smaller inverted or repeated sequences of Py DNA (Fig. 5C and D); however, these structures occur at considerably lower frequencies than expected (13) for the major population of D-92 DNA. Minor populations of heteroduplexes also occur with D-50 DNA and possess structures expected for deletions of Py DNA (Fig. 5E and F).

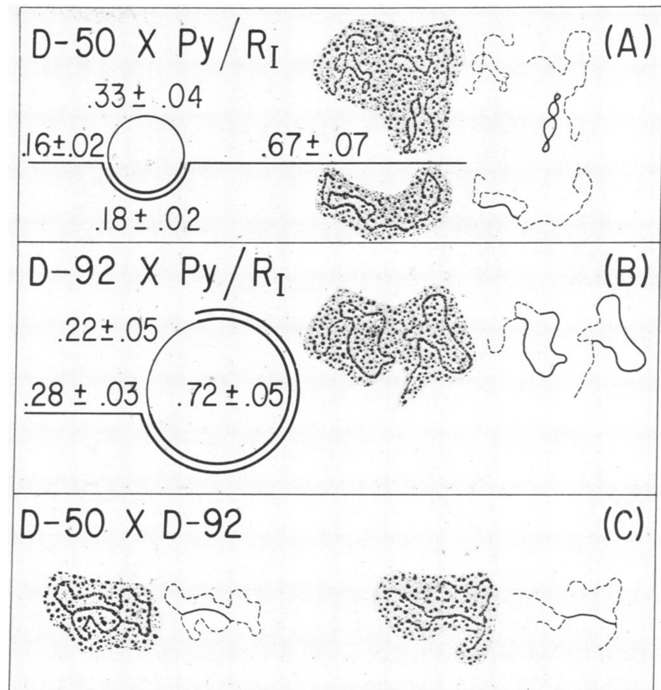

FIG. 4. Heteroduplex structures of D-DNAs. Each pane contains a drawing of the electron micrograph (inset) of single-stranded and duplex regions as in Fig. 1 and an accompanying digrammatic representation of the major heteroduplex with the observed measurements and standard deviations of the different segments. (A) D-50 × Py/R_I (39 molecules measured). The *calculated* standard deviations expected for the segments are 0.16 ± 0.020, 0.33 ± 0.028, 0.18 ± 0.016, and 0.67 ± 0.040 (8). The micrograph and drawing on the top are non-heteroduplexed single strands of D-50 and Py/R_I and a supercoil of D-50. (B) D-92 × Py/R_I (40 molecules measured). The *calculated* standard deviations expected for the segments are 0.28 ± 0.026, 0.22 ± 0.023, and 0.72 ± 0.31 (8). (C) D-50 × D-92 (27 molecules measured). Left, circular strands of each DNA appear intact, although a nick in one of the strands probably exists in the hybridized region; right, single strand of each DNA has been nicked. In (A) and (B), but not (C), heteroduplexes containing broken single-stranded circles were not used for length measurements.

Duplicated-Inverted Sequences of Polyoma Defective DNA. Denaturation of the nicked defective DNAs followed by a brief renaturation permits intramolecular hybridization of complementary regions (duplicated–inverted sequences) *within* the single strands. Of the D-DNAs studied, only D-80 contained a large proportion of molecules with self-complementary sequences. Self-complementary sequences were detected in 1–2% of the single-stranded molecules of D-47, D-50, D-74, and D-92 DNAs, but were not detected in Py/R_IDNA.

In D-80 DNA, self-complementary sequences greater than 15% of the length of Py DNA occur in 34% of all the single-stranded molecules. The most frequently occurring self-complementary structure (18% of all the single-stranded molecules) consists of a central duplex region with a length 20 ± 3% of Py DNA which is bounded by two single-strand

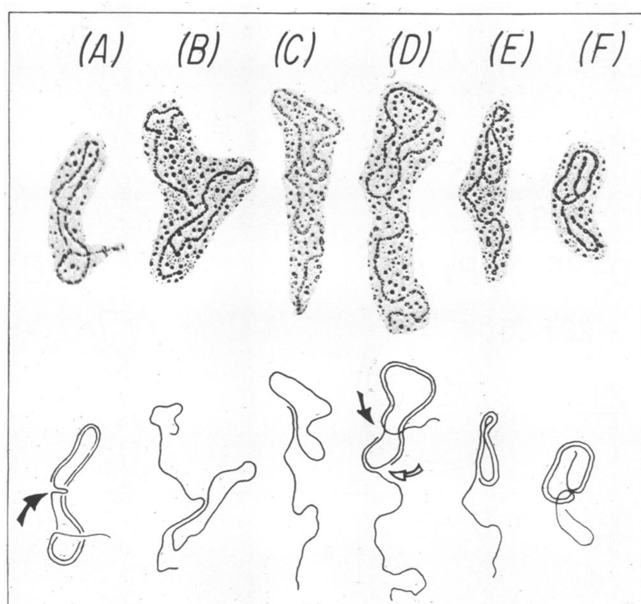

FIG. 5. Minor heteroduplex populations of either D-92 or D-50 DNAs with Py/R_IDNA. (A) A heteroduplex in which the circular D-92 DNA contains a small snap-back region near the EcoR_I site. (B and C) Single strands of Py/R_I DNA that have hybridized to a circular strand of D-92 DNA over a short region (16% and 10%, respectively, of the length of Py DNA). (D) A circular single strand of D-92 DNA that has hybridized to two strands of Py/R_I DNA. (E and F) Heteroduplexes in D-50 × Py/R_I with structures expected for clean deletions. Representations with equivalent radii of curvature are shown below each micrograph.

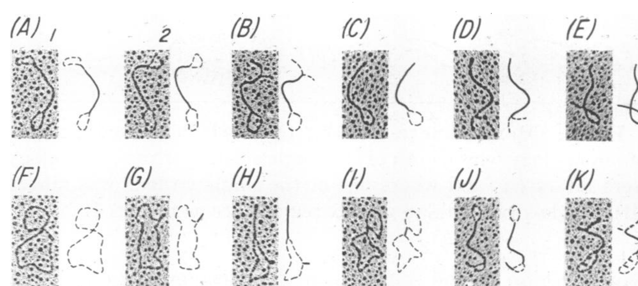

FIG. 6. Intramolecular hybridization within the single strands of D-80 DNA. ($A_{1,2}$) Most common snap-back structures with a duplex region terminated at both ends by a single-strand loop of DNA. (B) same as (A) structure but with a nick in a single-strand loop. (C) Snap-back structure with a duplex region terminating in only one single-strand loop of DNA. (D) same as (C) but with a nick in the single-strand loop. (E) Snap-back structure without single-strand loops. (F) A circular single strand with no evidence of snap-back regions. (G) and (I) Single strands with small snap-back regions. (H), (J), and (K) Small snap-back regions within the single-strand loops of other snap-back structures.

loops (Fig. 6A). The single-stranded loops have lengths 16 ± 3% and 19 ± 2% of Py DNA. The mean length of these D-80 DNA molecules which formed the commonest species of self-complementary structure, is 75 ± 4% the length of Py DNA (21 molecules measured) which is smaller than the mean length for the total population and may account for the heterogeneity observed in the length distribution of D-80 DNA (Fig. 2C).

Less frequently, the self-complementary region of single-stranded molecules of D-80 DNA terminates in one single-strand loop (Fig. 6C) or fork (Fig. 6D) (14% of all single-stranded molecules) or is devoid of single-strand portions (Fig. 6E) (2% of all single-stranded molecules). Small self-complementary regions are also detected in D-80 DNA single strands (Fig. 6G-K). A small self-complementary region is also detected in low frequency in some D-92 DNA single strands (Fig. 5A).

DISCUSSION

The results of our electron microscope analysis have demonstrated that most clonal isolates of polyoma defective DNAs are quite homogeneous in sequence. One isolate (D-80) is heterogeneous in size and contains a substantial population of molecules with duplicated-inverted sequences (Fig. 6).

Although each isolate of D-DNA has been found to be shorter in length than Py DNA, the heteroduplex experiments have revealed that D-DNAs are not simply deleted molecules of polyoma DNA but contain specific sequences which do not hybridize to Py DNA, covalently linked to limited portions of viral sequences. The regions in the D-DNAs which do not hybridize to Py DNA can be either host sequences or translocated, inverted, or repeated viral sequences. The inability to detect a substantial portion of heteroduplexes in the hybridization between D-92 and D-50 shows that the major populations of these D-DNAs contain no homologous regions with more than 400 base pairs in length. Thus, the DNA nonhomologous to Py DNA in D-92 and D-50 must be derived from different sources. If this DNA is host DNA, the two D-DNAs might differ in containing either host sequences from different sides of viral genomes integrated at one site or containing a sequence adjacent to viral genomes integrated at different sites in the host chromosome. Alternatively this nonhomologous DNA, in one or both D-DNAs, may be derived from the different translocated, inverted, or repeated viral sequences. The absence of heteroduplex structures between D-74 DNA and Py/R_I DNA is taken to indicate that D-74 DNA may be entirely composed of host DNA sequences or contains regions of Py DNA which are insufficient in length to provide detectable hybridization.

Exactly how defective DNA molecules are generated is not known. It has been suggested that they may be formed by incorrect excision of viral genomes integrated during the lytic cycle (12, 14). Another possible mechanism (abortive replication) is that defective DNA molecules can be generated as a result of incomplete or erroneous replication of either an integrated or an episomal viral genome. The excision of an erroneously replicated molecule could lead to formation of defective DNA molecules containing host sequences or inverted or translocated viral sequences.

The replication of mammalian chromosomal DNA is usually bidirectional with approximately equal rates of synthesis at the two replication forks of a particular replicon (15, 16). In a similar manner, the replication of SV40 (17, 18) and polyoma (19) viral DNA is initiated at a unique site on the genome and is predominantly bidirectional. If there were any delay in the synthesis of one of the two new strands in a fork, a region of single-strand template will be formed. Collapse of this single-strand segment could bring the origin and growing point of the new strand(s) into proximity for ligation (Fig. 7). Excision of this region could lead to the formation of a circular DNA containing an origin for DNA replication.

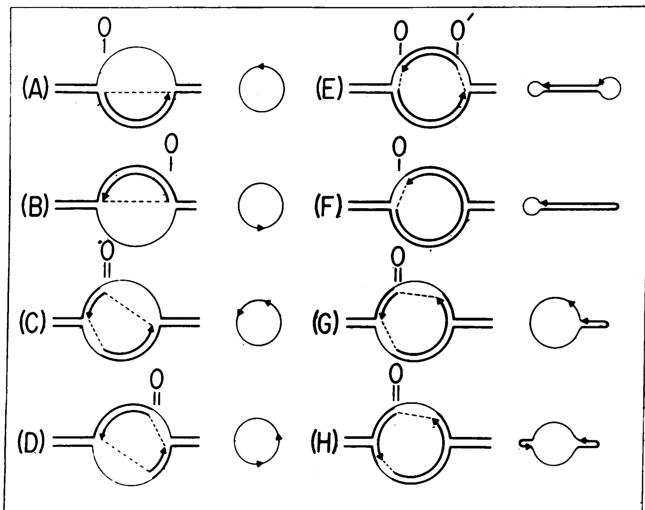

FIG. 7. Abortive replication. Asymmetric unidirectional replication (|) to the right (A) or left (B) of a given origin, 0, and bidirectional replication (||) with unequal rates of growth of the replication forks to the right (C) or left (D) of an origin, 0, can result in the formation of defective DNAs upon ligation of origin(s) and growing point(s) indicated by a dashed line and excision of the abortively synthesized sequences. Such abortive replication (E), initiated at origin, 0, which is accompanied by synthesis of the complementary strand initiated at origin, 0′, could lead to the formation of a defective DNA, the circular single strands of which will contain a self-complementary region which can snap-back to form a duplex segment bounded by single-strand loops (Fig. 6A). Abortive replication in which the growing chain(s) has crossed the replication fork(s) (21) leads to the formation of defective DNAs, the circular single strands of which contain one (F, G, and Fig. 6C) or two (H and Fig. 6H) self-complementary regions. The sequence(s) in the defective DNA that is duplicated and inverted by these abortive replication events can be adjacent to (F, G, and H) or separated (E) from its complement. The origins 0 or 0′ can be located in either viral or host sequences.

The particular form of the DNA produced by such a mechanism will depend on the exact state of replication at the moment of ligation of origin and growing point; Fig. 7 shows the way the various forms of D-DNA, observed in this study, could be generated. Thus, the particular patterns of self-complementary sequences apparent among the single strands of D-80 DNA are accommodated by patterns of DNA synthesis depicted in Fig. 7E–H. Obviously, when the region of junction between host and an integrated viral genome is involved in abortive replication of the sort shown (Fig. 7), the resulting circular DNA can contain both host and viral sequences.

If the nonhomologous portions of the defective DNAs do contain integration sites, then the study of the chemical nature of integration sites should now be feasible, as these defective DNAs can be produced in quantity without loss of the particular sequence arrangements described. Several of these isolates should also be useful in elucidating the biological activity and transforming ability of different viral genes.

After we had formulated the abortive replicon hypothesis, we were informed by N. Davidson that he and his colleagues, L. Chow and D. Berg, had completed heteroduplex studies of λ *dv* plasmids which suggest that aberrant patterns of viral DNA replication similar to those we propose can account for the sequence arrangements they observed (20). We thank these authors for receipt of their manuscript prior to publication. We thank Mrs. Moira Griffiths for excellent technical assistance and Dr. William R. Folk for the preparation of $EcoR_I$. D.L.R. is the recipient of a postdoctoral fellowship from the Damon Runyon Memorial Fund for Cancer Research, Inc.

1. Fried, M. (1974) *J. Virol.* **13**, 939–946.
2. Thorne, H. V. (1968) *J. Mol. Biol.* **35**, 215–226.
3. Yoshiike, K. (1968) *Virology* **34**, 391–401.
4. Thorne, H. V., Evans, J. & Warden, D. (1968) *Nature* **219**, 728–730.
5. Yoshiike, K. (1968) *Virology* **34**, 402–409.
6. Folk, W. R. & Wang, H. E. (1974) *Virology*, in press.
7. Yoshimori, R. N. (1971) Ph.D. Dissertation, University of California at San Francisco, Medical Center.
8. Sharp, P. A., Hsu, M. T., Otsubo, E. & Davidson, N. (1972) *J. Mol. Biol.* **71**, 471–497.
9. Davis, R. W., Simon, M. & Davidson, N. (1971) in *Methods in Enzymology*, eds. Grossman, L. & Moldave, K. (Academic Press, New York), Vol. XXI, pp. 413–428.
10. Robberson, D. L., Kasamatsu H. & Vinograd, J. (1972) *Proc. Nat. Acad. Sci. USA* **69**, 737–741.
11. Griffin, B., Fried, M. & Cowie, A. (1974) *Proc. Nat. Acad. Sci. USA* **71**, 2077–2081.
12. Tai, H., Smith, C. A., Sharp, P. A. & Vinograd, J. (1972) *J. Virol.*, **9**, 317–325.
13. Wetmur, J. G. & Davidson, N. (1968) *J. Mol. Biol.* **31**, 349–370.
14. Lavi, S. & Winocour, E. (1972) *J. Virol.* **9**, 309–316.
15. Huberman, J. A. & Riggs, A. D. (1968) *J. Mol. Biol.* **32**, 327–341.
16. Huberman, J. A. & Tsai, A. (1973) *J. Mol. Biol.* **75**, 5–12.
17. Danna, K. J. & Nathans, D. (1972) *Proc. Nat. Acad. Sci. USA* **69**, 3097–3100.
18. Fareed, G. C., Garon, C. F. & Salzman, N. P. (1972) *J. Virol.*, **10**, 484–491.
19. Crawford, L. V., Syrett, C. & Wilde, A. (1973) *J. Gen. Virol.* **21**, 515–521.
20. Chow, L. T., Davidson, N. & Berg, D. (1974) *J. Mol. Biol.* **86**, 69–89.
21. Guild, W. R. (1968) *Cold Spring Harbor Symp. Quan.t Biol.* **33**, 143.