JOURNAL OF VIROLOGY, Oct. 1979, p. 334-338
0022-538X/79/10-0334/05$02.00/0

Vol. 32, No. 1

Secondary Structures in Polyoma DNA†

MADELINE WU, HAIM MANOR,‡ AND NORMAN DAVIDSON*

Department of Chemistry, California Institute of Technology, Pasadena, California 91125

Received for publication 15 May 1979

Three reproducible secondary-structure features were observed on single strands of polyoma virus DNA mounted for electron microscopy by the T4 gene 32 protein technique: (i) a hairpin fold-back extending from 92.9 ± 0.8 to 95.0 ± 0.7 map units; (ii) a small loop extending from 63.2 ± 3.1 to 68.5 ± 2.8 map units; and (iii) a big loop extending from 51.9 ± 2.3 to 68.9 ± 2.1 map units. Both loops are bounded by inverted repeat stems of length 40 ± 20 base pairs. The stem sequences around 68.5 and 68.9 of the large and small loops overlap, either partially or completely. Several lines of evidence indicate that the inverted repeat stems of the two secondary-structure loops lie in the regions of polyoma virus DNA flanking and probably very close to the sequences that are spliced out in the formation of the late 16S and 18S messages, whereas the hairpin fold-back appears to map at a splicing point of an early message. These structures may therefore be important for the processing of the primary transcripts to form the early and late messages.

The T4 gene 32 protein method for mounting DNA for electron microscopy has proven to be particularly effective for recognizing interesting secondary-structure features in single-stranded regions of DNA (2, 4). We therefore thought it of interest, in connection with a recent study of the structure and splicing patterns of polyoma virus (Py) mRNA-DNA hybrids (2), to study those secondary-structure features on single-stranded Py DNA which are revealed by the gene 32 protein-mounting method. These observations are reported here.

The preparation of restriction endonuclease-cleaved Py DNA and of RNA-DNA hybrids has been described (2). Conditions for the spreading of denatured Py DNA after treatment with T4 gene 32 protein were as described (2, 4), except that in some cases the incubation with gene 32 protein was carried out at room temperature instead of at 37°C. Single-stranded circles of Py DNA were used as an internal length standard. On the basis of previous calibrations against simian virus 40 DNA, we take 5,300 nucleotides for the full length of Py DNA (100 map units [m.u.]).

The origin of coordinates for the conventional Py map is the single *Eco*RI site. When Py DNA was cleaved by digestion with *Eco*RI, denatured, and mounted for microscopy by the gene 32 protein method, three reproducible secondary-structure features were observed at frequencies of 5 to 20% (see Table 2): a small hairpin (Fig. 1a and b), a small loop (Fig. 1a), and a big loop (Fig. 1b and c). A second, less frequent hairpin at a position close to the more frequent one mentioned above is sometimes observed; it is shown in Fig. 1a but will not be considered further in the discussion below. The loops and hairpins are distinguishable from random crossovers and bumps by their reproducible dimensions and positions. Furthermore, the loops differ from crossovers in that they are usually connected to the rest of the single-stranded DNA by a short, double-stranded stem. This indicates that each loop is flanked by a short, inverted repeat sequence. The stems are indicated by arrows in Fig. 1a, b, and c.

The single-stranded length of the small hairpin was measured as 2.1 + 0.4 m.u., and it extended from 5.0 ± 0.7 m.u. to 7.1 ± 0.8 m.u. from an end of the *Eco*RI-cleaved molecules. Because of its proximity to an end, we denote this feature as the subterminal hairpin.

The subterminal hairpin feature appears to be stained with gene 32 protein; therefore, it is probably partially single stranded. It was never observed as a loop, possibly because of its short length.

The small loop was 5.3 ± 1.4 m.u. in length, and it began 24.4 ± 2.8 m.u. from the subterminal hairpin; therefore, it extended from 31.5 ± 2.8 to 36.8 ± 3.1 m.u. from that end of the molecule closest to the hairpin. The measured length of the big loop was 17.0 ± 1.0 m.u.; it began at 24.0

† Contribution no. 6004 from the Department of Chemistry, California Institute of Technology, Pasadena, CA 92215.

‡ Present address: Department of Biology, Technion Israel Institute of Technology, Haifa, Israel.

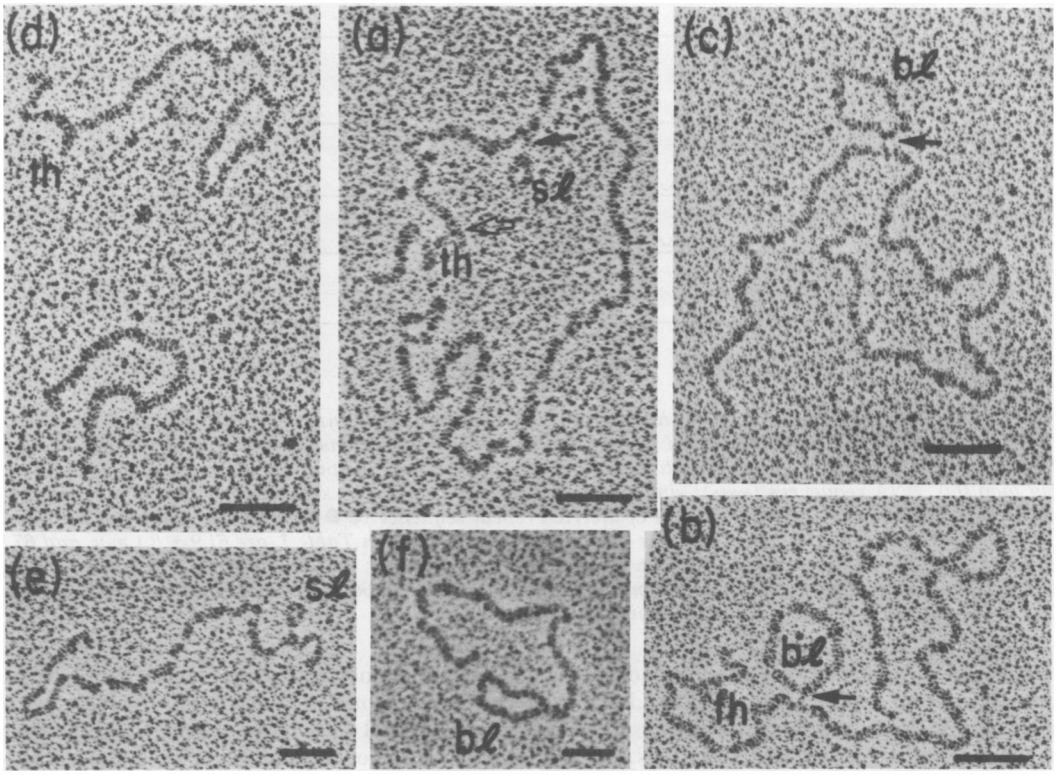

FIG. 1. *Electron micrographs of Py DNA mounted by the gene 32 protein method. a, b, and c show the secondary structure features observed on single strands of the EcoRI-cleaved molecules. (a) A subterminal hairpin (th) and a small loop (sl) with a solid arrow pointing to the short double-stranded stem of the small loop. In this micrograph, a second, less frequent hairpin (open arrow) was also observed. (b) A subterminal hairpin (th) and a big loop (bl) with an arrow pointing to the short, double-stranded stem. (c) A big loop (bl) with an arrow pointing to the stem. (d) A single strand of the EcoRI-cleaved Py DNA, hybridized to 16S mRNA showing the relative position of the subterminal hairpin fold-back (th), the DNA-RNA duplex region, and the spliced-out region. (e) The small loop (sl) on a single strand of the largest HhaI fragment. (f) The big loop (bl) on a single strand of the largest HhaI fragment.*

± 2.1 m.u. from the subterminal hairpin and therefore extended from 31.1 ± 2.1 to 48.1 ± 2.3 m.u. from the end of the molecule closer to the subterminal hairpin. These measurements and their standard deviations are given in Table 1. Therefore, all three of these secondary structure features are located on the same half of the *Eco*RI-cleaved, full-length molecule.

The twofold uncertainty as to the coordinates of these features on the Py map was resolved by several experiments. The structure and coordinates of the two most abundant late mRNA transcripts in Py-infected cells are shown in Fig. 2 (2). The 5′ ends are close to coordinate 68 m.u. There are short 5′ leader sequences of estimated lengths 1.2 m.u. The 16S spliced region extends from 67 to 49 m.u. The shorter 18S spliced region extends from 67 to 61 m.u. The messages then extend to their 3′ ends at 25 m.u.

*Eco*RI-cleaved Py DNA was denatured and

TABLE 1. *Length and position of the secondary-structure features and splicing loops on the Py genome*

Structure feature	Distance from the nearest EcoRI site[a]	Length[a]	No. of molecules
Subterminal hairpin[b]	5.0 ± 0.7	2.1 ± 0.4	11
Small loop[c]	31.5 ± 2.8	5.3 ± 1.4	18
Big loop[b]	31.1 ± 2.1	17.0 ± 1.0	11
Small splicing loop[d]	32.7 ± 3.1	5.4 ± 0.7	10
Large splicing loop[d]	32.0 ± 1.8	17.0 ± 1.2	21

[a] Each value indicates Py units ± standard deviation.

[b] Data were collected from intact DNA molecules with both the subterminal hairpin and the big loop.

[c] Data were collected from intact DNA molecules with both the subterminal hairpin and the small loop.

[d] From RNA-DNA hybrid structures.

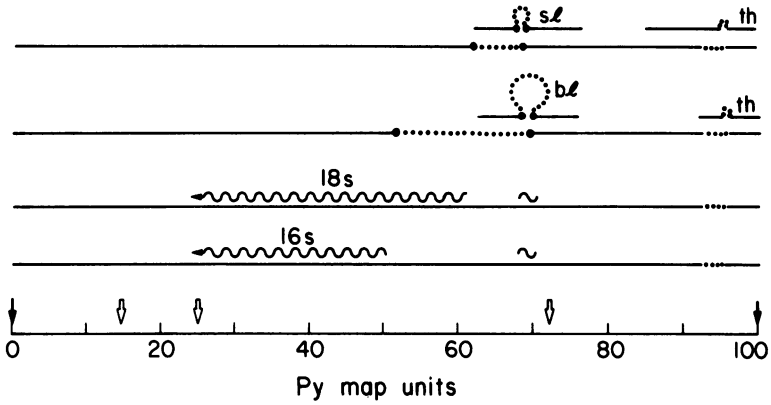

FIG. 2. *Diagrams of the Py genome showing the cleavage sites of the enzymes EcoRI (▼) and HhaI (▽); the coordinates of the DNA-RNA hybrid regions and the spliced-out regions for the 16S and 18S late message RNAs; and the various secondary-structure features, the subterminal hairpin (th), the small loop (sl), and the big loop (bl). RNA sequences (∿∿∿); single-stranded DNA (——); DNA sequences included in secondary-structure features (· · · ·); short segment of inverted repeat sequence (●●●). The coordinates of the splice points shown, which are based on the gene 32 protein measurements in Table 1, are 61.9 ± 3.1 m.u. and 67.3 ± 3.1 m.u. for the 18S message, and 51.0 ± 2.0 and 68.0 ± 2.0 m.u. for the 16S message. The corresponding values measured previously (2) in cytochrome c spreads, 59.4 ± 0.9 and 66.4 ± 0.5 m.u. (18S) and 49.4 ± 2.0 and 66.7 ± 1.0 m.u. (16S), are in agreement within measurement error.*

hybridized to late mRNA and mounted by the gene 32 protein method. Many molecules were seen of the type described previously (2). For example, in the hybrids with 16S mRNA, there is a short RNA-DNA duplex due to the leader sequence, a loop of single-stranded DNA of length 18.4 ± 1.7 m.u., corresponding to the regions spliced out in the mRNA, followed by a longer RNA-DNA duplex due to the main body of the message. In 11 of 100 of these molecules, the subterminal hairpin was observed on the 5′ side of the RNA. An example of such a molecule is shown in Fig. 1d. Therefore, the map coordinates of the subterminal hairpin are assigned as 92.9 ± 0.9 to 95.0 ± 0.7 m.u.

We then conclude that the coordinates of the small loop and the big loop are 63.2 ± 3.1 to 68.5 ± 2.8 m.u. and 51.9 ± 3.4 to 68.9 ± 2.1 m.u., respectively. Thus, the right-hand termini of the two loops are the same within experimental error.

A further observation bearing on the position and nature of the two internal secondary-structure loops is as follows. In the sample of RNA-DNA hybrids, there were some with full-length DNA hybridized to a short broken fragment of RNA derived from the 5′ end. These molecules showed a duplex segment due to the short 5′ leader sequence, a single-strand loop of DNA corresponding to the spliced region, and an RNA-DNA hybrid segment of variable length extending toward the 3′ end of the RNA. The subterminal hairpin was observed at the expected frequency and position in such molecules. However, among 50 molecules of this type, with either a small or a large splicing loop, neither the big-loop nor the small-loop secondary-structure feature was ever observed as an additional structure, either within the splice loop or outside of the region including the short RNA-DNA hybrid and the splice loop. These observations suggest that one or both of the stems of each secondary-structure loop are included in the RNA-DNA hybrid regions flanking and probably very close to the spliced-out regions.

The coordinates and dimensions of the several secondary-structure features are listed in Table 1. The coordinates of the spliced loops as measured by the gene 32 protein method are also presented. These values agree, within experimental error, with the values previously determined from cytochrome *c* spreads (2) (see legend to Fig. 2). Within experimental error, the spliced loops and the secondary-structure loops have the same lengths and positions. These quantitative data are consistent with the hypothesis that the inverted repeat stems of the secondary-structure loops lie in the RNA-DNA hybrid regions immediately flanking the spliced loops.

The secondary-structure loops were also observed in the expected position when single strands of the large *Hha*I fragment (Fig. 2) of Py DNA were mounted by the gene 32 method (Fig. 1e and f).

The double-strand length of the stem at the base of the big loop was measured as 0.92 ± 0.12

m.u. (n = 16); the double-strand length of the stem for the small loop was 0.6 ± 0.1 m.u. (n = 8).

Table 2 gives the frequencies of occurrence of the several secondary structure features when incubation of the single-stranded DNA with gene 32 protein was carried out at 37 and at 22°C. The several features are observed slightly more frequently at the lower temperature of incubation, presumably indicating a greater stability of the short duplex stems at the lower temperature.

The low frequencies of occurrence of the several inverted repeat features are presumably due to the short length of the inverted repeat duplex stems or to partial sequence mismatch in these duplex regions or both.

In Table 2, we have calculated from the observed total frequencies of the several secondary-structure features the predicted frequencies of molecules with two features, assuming that the occurrence of each feature is independent of the presence of another. The observed and predicted frequencies are comparable for the subterminal hairpin with either the small or big loop. However, in 400 molecules, none was seen with both the small and big loop; on the hypothesis of independent formation, 5.7 are expected. For a Poisson distribution, the probability of no events when 5.7 are expected is 0.003. Therefore, the data support the hypothesis that formation of the small loop and the big loop are mutually exclusive events, although the sample is too small to establish this conclusion at a very high confidence level.

In summary, by the gene 32 protein method, we have observed several secondary structure features on single strands of Py DNA with map coordinates and positions as given in Table 1 and as shown in Fig. 2. Two of the observed structures, the big and the small loop, are flanked by an inverted repeat duplex stem of estimated length 50 ± 6 and 32 ± 6 nucleotide pairs, respectively. Measurements of very short duplex regions bounded by single-stranded regions are not very accurate; therefore, we judge that the measurements indicate a length of 40 ± 20 base pairs for each stem. Formation of one loop precludes formation of the other. The data therefore suggest that there is a short sequence at about 68.7 m.u. which can pair either with an inverted complement at 63.2 to form the short loop or with a second inverted complement at 51.9 to form the large loop. The positions of these inverted repeat sequences are close to the ends of the regions removed by splicing in the formation of the 16S and 18S late messages (Table 1 and Fig. 2).

It is therefore possible that the same secondary-structure loops form in the primary nuclear transcript and that the duplex stems are cutting and joining sites for the processing of the primary transcript to the mature messages. The hairpin fold-back maps at a site which corresponds to a splicing point in early Py messages (93 m.u. [T. Favaloro, R. Treisman, and R. Kamen, personal communication]). Therefore, this secondary-structure feature, as well as the loop structures discussed above, may be related to the splicing phenomenon.

As a technical matter, we note that the T4 gene 32 protein method seems to be particularly effective, and more so than formamide, cytochrome c spreading, for identifying characteristic secondary-structure features defined by short flanking inverted repeat sequences. The loops and stems due to the short inverted repeat sequences flanking the *E. coli* 16S and 23S rRNA genes in the DNA of ϕ80*d3ilv* were first observed by the gene 32 protein method (4). The existence

TABLE 2. *Frequencies of secondary-structure features*

Structure	No. of molecules with structure/total molecules at:		Predicted frequencies at:[b]	
	37°C[a]	22°C[a]	37°C[a]	22°C[a]
Subterminal hairpin only	12/100	57/300		
Small loop only	9/100	41/300		
Big loop only	4/100	18/300		
Subterminal hairpin and small loop	5/100	13/300	2.8/100	14/300
Subterminal hairpin and big loop	3/100	8/300	1.4/100	6.8/300
Small loop and big loop	0/100	0/300	1.0/100	4.7/300

[a] Temperature of incubation of the DNA with gene 32 protein before spreading.

[b] From the data in the table, we calculate total frequencies of occurrence of the subterminal hairpin as f_{th} = 20/100 and 78/300 at 37 and 22°C, respectively; of the small loop as f_{sl} = 14/100 and 54/300 at 37 and 22°C, respectively; and of the big loop as f_{bl} = 7/100 and 26/300 at 37 and 22°C, respectively. The predicted frequencies, assuming that the occurrence of one feature is independent of the other, are $f_{th}\,f_{sl}$, $f_{th}\,f_{bl}$, and $f_{sl}\,f_{bl}$.

of inverted repeat stems of 26 base pairs in length (16S) (6) and 29 base pairs in length (23S) (R. A. Young and J. Steitz *In* D. Söll, J. Abelson, and P. Schimmel, ed., *Transfer RNA, part 2,* in press) were later demonstrated by DNA sequencing.

The γδ sequence of F which is also present on the DNA of *φ80d3ilv* forms a 5-kb loop bounded by short, inverted repeat stems (4). This structure is observed infrequently in cytochrome *c* spreads, but much more frequently in gene 32 spreads. In this case too, DNA sequencing has confirmed the existence of the inverted repeat and shown that it has a length of 34 base pairs (R. Reed, R. Young, J. Steitz, M. Guyer, and N. Grindley, personal communication).

There is a secondary-structure feature in single strands of adenovirus type 2 DNA. The loop and stem structure is much more clearly visualized by the gene 32 method than by cytochrome *c* spreading (5).

Several authors have observed secondary-structure features in single strands of SV40 DNA, by cytochrome *c* spreading from formamide-ammonium acetate solvents (1), by photochemical cross-linking with psoralen in dilute aqueous buffers followed by formamide-cytochrome *c* spreading (4), or by a combination of these methods. Some of these features correlate with the positions of splice points for SV40 early or late messages (J. C.-K. Shen and J. E. Hearst, Anal. Biochem., in press). In general, the secondary-structure patterns are more complex than those observed by us by the T4 gene 32 protein method. The latter method provides higher resolution for observing small features. Whether the differences in numbers of features observed are due mainly to sequence differences between the two genomes or to differences in the techniques used is not known at present.

This research was supported by grants from the U.S.A.-Israel Binational Science Foundation and the Leukemia Research Foundation, and by Public Health Service grant GM 20927 from the National Institutes of Health. H.M. was supported by an Eleanor Roosevelt-International Cancer Fellowship during the course of this work.

We thank Yaffa Bot for able technical assistance.

LITERATURE CITED

1. **Hsu, M.-T., and W. R. Jelinek.** 1977. Mapping of inverted repeated DNA sequences within the genome of simian virus 40. Proc. Natl. Acad. Sci. U.S.A. **74**:1631–1634.
2. **Manor, H., M. Wu, N. Baran, and N. Davidson.** 1979. Electron microscopic mapping of RNA transcribed from the late region of polyoma virus DNA. J. Virol. **32**:293–303.
3. **Shen, C.-K. J., and J. E. Hearst.** 1977. Mapping of sequences with 2-fold symmetry on the simian virus 40 genome: a photochemical crosslinking approach. Proc. Natl. Acad. Sci. U.S.A. **74**:1363–1367.
4. **Wu, M., and N. Davidson.** 1975. Use of gene 32 protein staining of single strand polynucleotides for gene mapping by electron microscopy: application to the φ80d3*ilv*$^{+}$7 system. Proc. Natl. Acad. Sci. U.S.A. **72**: 4506–4510.
5. **Wu, M., R. J. Roberts, and N. Davidson.** 1977. Structure of the inverted terminal repetition of adenovirus type 2 DNA. J. Virol. **21**:766–777.
6. **Young, R. A., and J. A. Steitz.** 1978. Complementary sequences 1700 nucleotides apart form a ribonuclease III cleavage site in *Escherichia coli* ribosomal precursor RNA. Proc. Natl. Acad. Sci. U.S.A. **75**:3593–3597.