JOURNAL OF VIROLOGY, May 1977, p. 520-526

Vol. 22, No. 2
Printed in U.S.A.

Sequence Analysis of the Nicks and Termini of Bacteriophage T5 DNA

BRIAN P. NICHOLS AND JOHN E. DONELSON*

Department of Biochemistry, University of Iowa, Iowa City, Iowa 52242

Received for publication 30 November 1976

Bacteriophage T5 DNA, when isolated from mature phage particles, contains several nicks in one of the two strands. The 5'-terminal nucleotides at the nicks were labeled with polynucleotide kinase and [γ-^{32}P]ATP, and the 3'-terminal nucleotides were labeled with *Escherichia coli* DNA polymerase I and [α-^{32}P]dGTP. The sequences around the nicks were analyzed by partial nuclease digestion followed by homochromatography fractionation of the resulting oligonucleotides. The nicks had at least the sequence -Pu_{OH} pGpCpGpC- in common. In addition, the two 5' external termini had the first seven nucleotides in common.

The chromosome of bacteriophage T5, when isolated from mature phage particles, contains several genetically defined nicks in one of the two strands (1, 4, 6, 8, 10), which can be sealed in vitro by DNA ligase (9) and thus contain adjacent 5'-phosphate and 3'-hydroxyl groups. Two classes of nicks, major and minor, occur in a population of T5 DNA molecules. Major nicks occur in most if not all molecules, and a given minor nick occurs in only a few molecules of a given population (7, 10). The biological function of these nicks is not known. Recently four different nucleases capable of introducing nicks into double-stranded DNA have been isolated from T5-infected *Escherichia coli*, but none alone or in combination with the others is capable of producing in vitro the pattern of nicks obtained from DNA extracted from mature phage particles (14). The sequence specificity of these nucleases was not determined. Nevertheless, the nicks may arise because of the sequence specificity of one or more putative nucleases capable of introducing a single-stranded break in duplex DNA. Jacquemin-Sablon and Richardson (9) have reported that dGMP is the 5'-terminal nucleotide at a majority (>60%) of nicks of T5 DNA. In this paper we investigate the sequence homologies at the nicks and at the 5' termini of T5 DNA. We find that the first four nucleotides at the 5' terminus of each nick are the same, after which the sequences diverge, and that the 3' terminus is a purine. Also, the first seven nucleotides at the two external 5' termini are the same.

MATERIALS AND METHODS

Bacteriophage purification and DNA isolation were as described before (10).

Oligodeoxyribonucleotides of defined sequence, purchased from Collaborative Research, Inc., were pGpC, pGpA, pGpT, pGpG, pGpCpGpC.

Labeling and identification of the 5'-terminal oligonucleotides. To remove the naturally occurring phosphate group from the 5' termini, 35 to 55 μg of T5 DNA in 100 μl of 10 mM Tris-hydrochloride (pH 8.1), 1 mM Na_3EDTA, and 20 mM NaCl was incubated for 30 min at 65°C, with the addition of 0.5 U of bacterial alkaline phosphatase (Worthington BAPF) at 0, 10, and 20 min. Incubation at this elevated temperature assures the removal of both internal and external phosphates (9). The reaction mixture was cooled to room temperature, 15 μl of 1 N KOH was added, and, after 10 min at 25°C, 15 μl of 1 M Tris-hydrochloride, pH 3.8, was added to bring the pH to approximately 8.5. The solution was then made 10 mM in $MgCl_2$ and 15 mM in β-mercaptoethanol, followed by addition of 20 μCi of [γ-^{32}P]ATP (specific activity, 1,000 mCi/μmol; prepared according to reference 16) and 3 U of T4 polynucleotide kinase (P-L Biochemicals). After 15 min at 37°C, the reaction was terminated with 10 μl of 0.5 M Na_3EDTA, pH 8.0, and the sample was applied to a Sephadex G-150 column (1 by 20 cm) and eluted with 50 mM triethylammonium bicarbonate, pH 7.8. The labeled DNA eluted from the column in the void volume, and the triethylammonium bicarbonate was removed by repeated evacuations to dryness. Typically, the specific activity was about 1,000 cpm per μg of DNA.

Complete digestion of labeled DNA was at 37°C for 2 h in 5 μl containing 5 μg of T5 DNA, 50 mM Tris-hydrochloride (pH 7.4), 10 mM $MgCl_2$, 5 mM β-mercaptoethanol, 0.2 μg of DNase I, and 0.4 μg of venom phosphodiesterase (the latter two from Worthington Biochemicals Corp.). Partial digestions were carried out under the same conditions, using one-tenth the amount of each enzyme (12). Trial digestions were necessary to determine the optimum partial conditions.

Separation of the four mononucleotides from a complete digestion was by electrophoresis on What-

man 3MM paper in pyridine acetate buffer, pH 3.5 (2). Oligonucleotides from partial digestions were separated by electrophoresis on Whatman DE-81 paper in 7% formic acid (pH 2) or by homochromatography (2, 3).

DNA ligase reaction. A mixture containing 45 μg of T5 DNA, 100 mM Tris-hydrochloride (pH 7.5), 50 mM NaCl, 10 mM $MgCl_2$, 1 mM dithiothreitol, 0.1 mM ATP, and 0.75 U of T4 DNA ligase (Miles) was incubated at room temperature for 90 min and then dialyzed at 4°C for 8 h against 10 mM Tris-hydrochloride (pH 8.1) and 1 mM EDTA. This repaired DNA migrated as a single band on agarose gels when denatured before application, showing that the internal nicks had been sealed by DNA ligase. The repaired DNA was then labeled at the 5'termini as described above.

Labeling and identification of the 3'-terminal nucleotides. The method for labeling the 3'-terminal nucleotides at internal nicks took advantage of the fact that the 5'-terminal dinucleotide at the nicks was found to be pGpC (see Results). Native T5 DNA was incubated with [α-^{32}P]dGTP and *E. coli* DNA polymerase I under conditions where the 5' → 3' exonuclease activity of the enzyme removes the 5'-terminal pG at the nicks and the polymerase activity incorporates a ^{32}pG, in effect translating each nick one residue down the strand (11). Further incorporation does not occur because dCTP (cytosine being the next nucleotide to be incorporated) is not present in the incubation.

T5 DNA (16 μg) was incubated in 500 μl containing 50 mM potassium phosphate (pH 7.3), 7 mM $MgCl_2$, 2 mM dithiothreitol, 0.05 mCi of [α-^{32}P]dGTP (New England Nuclear Corp.; specific activity, 100 to 130 mCi/μmol), and 10 U of *E. coli* DNA polymerase I (grade I from Boehringer-Mannheim). After 1.5 h at 16°C, the reaction was stopped by the addition of 15 μmol of Na_3EDTA, and the incubation mixture was applied to a Sephadex G-150 column and eluted as described above. The DNA was then subjected to either partial digestion or a nearest-neighbor analysis according to Sanger et al. (15).

RESULTS

Tetranucleotide sequence at the 5' termini. Treatment of T5 DNA with bacterial alkaline phosphatase at 65°C followed by incubation with polynucleotide kinase and [γ-^{32}P]ATP introduces a labeled phosphate group on both external and internal 5'-terminal nucleotides (9). After complete digestion of the labeled DNA with DNase I and venom phosphodiesterase, the nucleotides were separated by electrophoresis, and the radioactive nucleotide spots were cut out and quantitated (Table 1). Sixty-eight percent of the label migrates as dGMP, and the remaining label is distributed among the three other deoxynucleotides. Thus the majority of fragments were terminated at the 5' end by dGMP. This result is similar to that obtained previously (9).

When partial enzymatic digests of 5'-terminally labeled T5 DNA were run on the two-dimensional homochromatography system of Brownlee and Sanger (3), the resulting pattern of spots indicated the presence of two or more sequences (Fig. 1a). Those DNA fragments that did not terminate in d(pG) appeared to be heterogeneous in sequence, and appear in Fig. 1a as background. To determine the sequences present, partial enzymatic digests of 5'-terminally labeled T5 DNA were spotted on DEAE-cellulose thin-layer plates and developed one dimensionally with a 3%, 30-min homomixture (Fig. 1b). Spots corresponding to mono-, di-, tri-, and tetranucleotides were eluted and electrophoresed on Whatman DE-81 paper at pH 2 with known standards. In such an electrophoretic system, an M-value may be calculated as x/y, where y is the migration distance of an oligonucleotide and x is the difference in migration of the oligonucleotide and its first degradation product. The value thus obtained is characteristic of the particular nucleotide removed during the enzymatic degradation (2). M-values were calculated and compared with those of the standards. Nucleotide assignments were made by taking into account both M-values and the original pattern of the two-dimensional homochromatography. The M-values are shown in Table 2 and indicate one labeled mononucleotide (dGMP) and two each of the di, tri-, and tetranucleotides. The two dinucleotides co-electrophoresed with the standards d(pGpC) and d(pGpA). The third nucleotide of each trinucleotide could be assigned two possible M-values, depending on the dinucleotide with which it is compared. However, one set of assignments is not compatible with the observed mobility shifts on the two-dimensional homochro-

TABLE 1. *Mononucleotides labeled at the 5' termini and 3' termini*[a]

Nucleotide	% of labeled 5' termini[a]	% of nearest-neighbor analysis of labeled 3' termini[b]
dAMP	9.5	62.1
dGMP	68.1	31.8
dCMP	13.9	6.7
TMP	8.5	0.0

[a] Labeling was done as described in the text. Values are the average of four different labeling experiments with about 4,400 cpm sampled per experiment.

[b] Values are the average of ten different nearest-neighbor analyses from two different incorporation experiments in which samples were taken at 15, 30, 60, 120, and 180 min during the DNA polymerase I incubation. About 2,700 cpm was sampled at each time point, and the ratio of dAMP*/dGMP* = 2 did not vary with time. See text for further explanation.

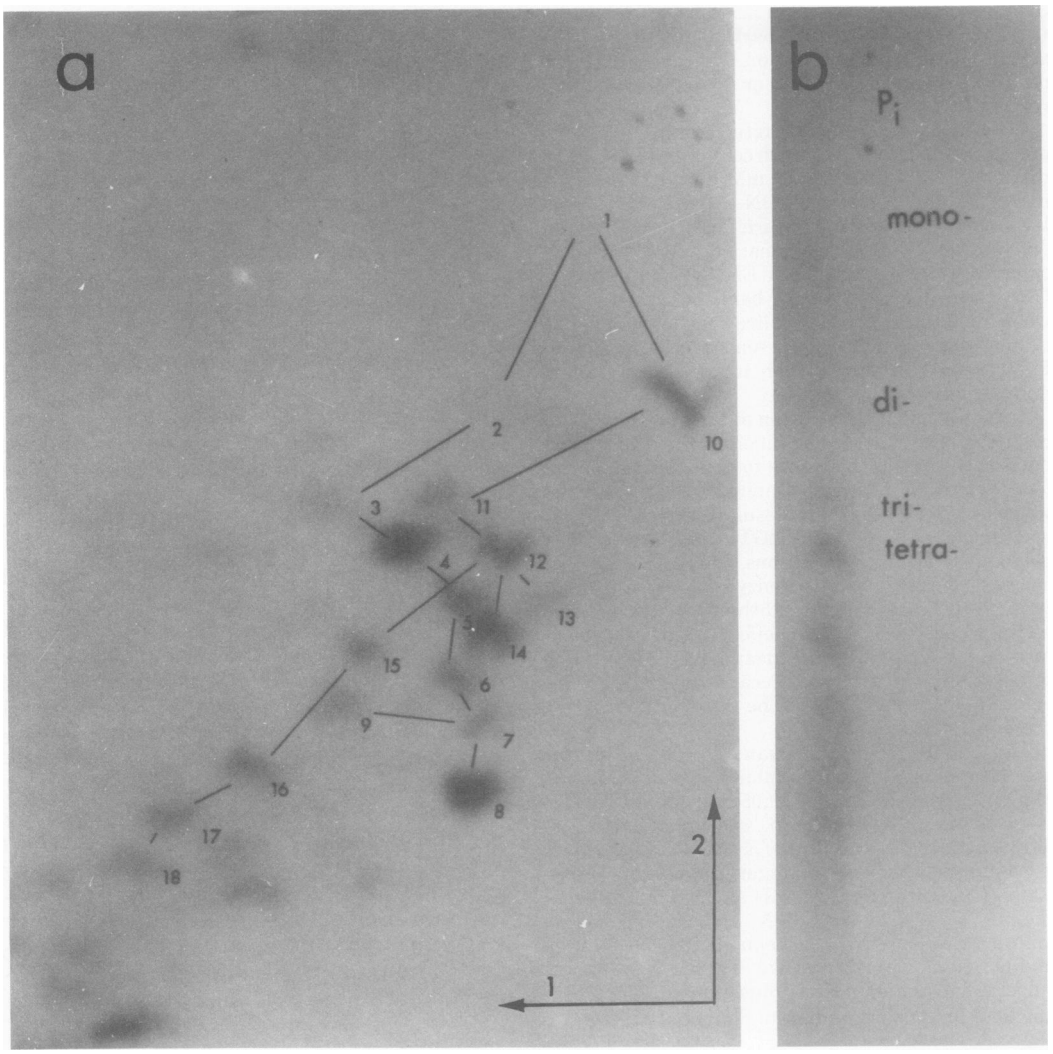

FIG. 1. *(a) Autoradiogram of a two-dimensional homochromatography fractionation of a partial digest of 5'-terminally labeled T5 DNA. (b) Autoradiogram of the same partial digest fractionated by one-dimensional homochromatography. The mononucleotide in (b) is equivalent to spot 1 in (a). The dinucleotide in (b) contains spots 2 and 10 of (a), the trinucleotide of (b) contains spots 3 and 11 of (a), and so forth.*

matography system. Thus the trinucleotides are d(pGpCpG) and d(pGpApG). Only one *M*-value assignment is possible for the fourth nucleotide, since the alternative assignment would yield a negative *M*-value. Thus the fourth nucleotide in each sequence is cytidine. This result is also consistent with the two-dimensional homochromatography pattern of partial digests. In addition, one of the tetranucleotides co-electrophoresed with the standard d(pGpCpGpC) on DEAE paper at pH 2 and at pH 3.5. Thus the two 5'-tetranucleotide sequences determined by this method are d(pGpCpGpC) and d(pGpApGpC).

Extension of 5'-terminal sequences. To extend further the two sequences thus far determined, oligonucleotide spots from a partial digest separated on two-dimensional homochromatography (Fig. 1a) were eluted, partially redigested with venom phosphodiesterase, and separated again either on two-dimensional homochromatography or on one-dimensional electrophoresis on DEAE paper. An example of a partial redigest of oligonucleotide 8 separated on two-dimensional homochromatography is shown in Fig. 2. This octanucleotide sequence was determined to be d(pGpApGpCpCpApCpA) from the mobility shifts (5) on homo-

TABLE 2. *M-values at pH 2 of standard (oligo)nucleotides and eluted nucleotides shown in Fig. 1b*

(Oligo)nucleotides	Experimental *M*-value	Sequence
Standards		
5'-dGMP		
d(pGpA)	0.14	
d(pGpG)	0.52	
d(pGpC)	0.08	
d(pGpT)	0.42	
Mononucleotide		
1	Co-electrophoreses with 5'-dGMP	dpG
Dinucleotides		
10	0.07	d(pGpC)
2	0.14	d(pGpA)
Trinucleotides		
11	0.57	d(pGpCpG)
3	0.85	d(pGpApG)
Tetranucleotides		
12	0.08	d(pGpCpGpC)
4	0.06	d(pGpApGpC)

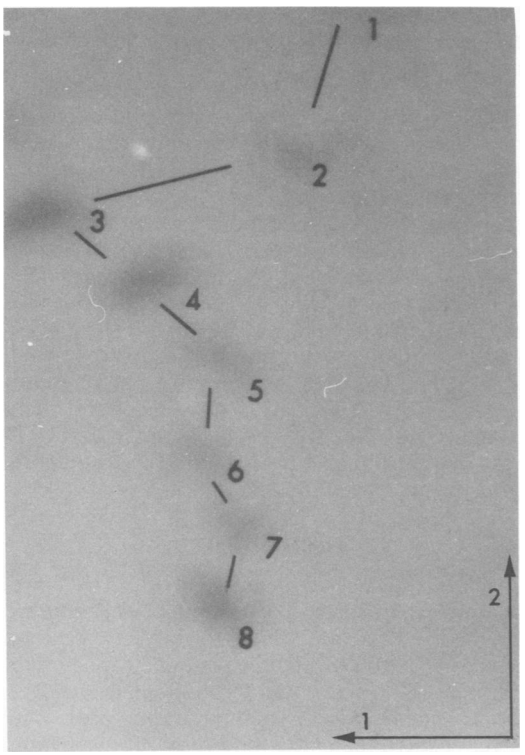

FIG. 2. *Autoradiogram of a partial redigest of the oligonucleotide in spot 8 of Fig. 1a fractionated by two-dimensional homochromatography. The numbers indicate the corresponding spots shown in Fig. 1a.*

chromatography and *M*-values on one-dimensional electrophoresis on DEAE paper at pH 2 and pH 3.5 (2). Experimental *M*-values and the sequence are shown in Table 3. The five sequences shown in Fig. 5 were determined by similar analyses of partial redigests on homochromatography and DEAE paper electrophoresis. Table 3 summarizes the data. Three of the sequences were homologous throughout the first four nucleotides, and two of the sequences were homologous throughout the first seven nucleotides.

Nucleotide sequence at external 5' termini. To distinguish the sequences present at the internal 5' termini from the sequences present at the external 5' termini, T5 DNA was treated with DNA ligase before labeling with polynucleotide kinase. Partial enzymatic digests separated on two-dimensional homochromatography (Fig. 3) yielded a pattern less complex than that shown in Fig. 1a. The sequence determined from the angles of the mobility shifts

TABLE 3. *Analyses of partial redigests of oligonucleotides shown in Fig. 1a*

Oligonucleotides redigested	Partial products[a]	*M*-value at pH 3.5 on DEAE paper electrophoresis	Indicated nucleotide from mobility on 2-D homochromatography[b]	Sequence
8	1	–		d*pG
	2	1.13 (0.14)[c]		A
	3	2.40 (0.82)	G	G
	4	0.47 (0.08)	C	C
	5	0.70 (0.05)	C	C
	6	1.50 (0.20)	A	A
	7	(0.05)	C	C
	8	(0.22)	A	A
9	–[d]			–[d]
	9		T	T
13	1	–		d*pG
	10	0.46		C
	11	2.6		G
	12	0.72	C	C
	13	0.63	C	C
14	–[e]			–[e]
	14	1.4	A	A
15–18	–[e]			–[e]
	15	2.6	G	G
	16	–	G	G
	17	–	T	T
	18	–	G	G

[a] Numbers indicate the oligonucleotides in Fig. 1a that correspond to the digestion products of the partial redigest.

[b] Nucleotides are determined from the angle of the mobility shift caused by the removal of one nucleotide from an oligonucleotide (see reference 5 for a description). 2-D, Two-dimensional.

[c] Numbers in parentheses indicate *M*-values at pH 2.0 instead of pH 3.5.

[d] Partial products 1 through 7 were obtained as with the redigestion of oligonucleotide 8. Thus, the sequence is *pGpApGpCpCpApC-.

[e] Partial products 1, 10, 11, and 12 were obtained as with the redigestion of oligonucleotide 13. Thus, the sequence is *pGpCpGpC-.

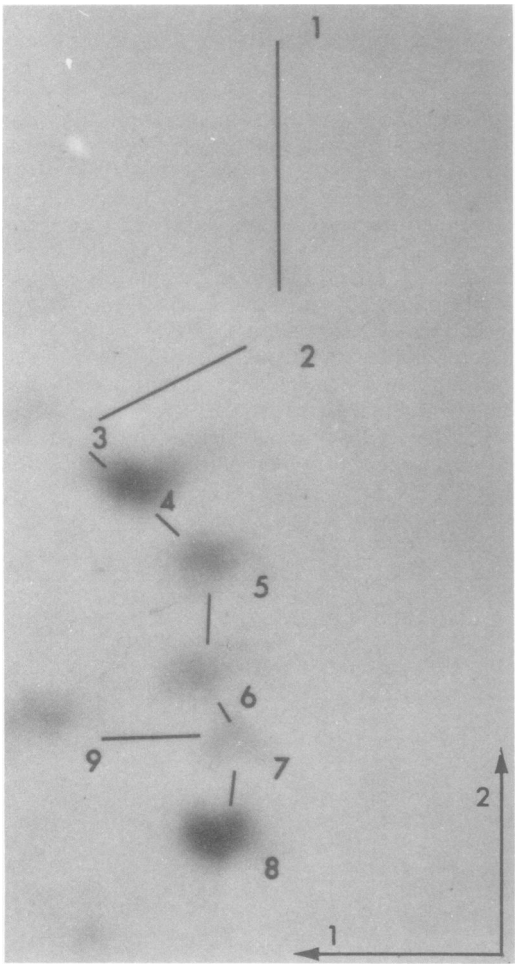

FIG. 3. *Autoradiogram of the fractionation (as in Fig. 1a) of a partial digest of 5′-terminally labeled T5 DNA that was repaired with DNA ligase before labeling.*

was confirmed by eluting the spots, partially redigesting the oligonucleotides, and resolving them a second time via one-dimensional electrophoresis on DEAE paper as described above. The two octanucleotide sequences d(pGpApGpCpCpApCpA) and d(pGpApGpCpCpApCpT) were determined, which differ only in the eighth residue and which must be the sequences at the two external termini of the chromosome.

Determination of the 3′-terminal nucleotide. Since the 5′-terminal nucleotide at the internal nicks of T5 DNA is 5′-dGMP (see previous section), *E. coli* DNA polymerase I in the presence of [α-^{32}P]dGTP will translate the nick one residue along the strand by removing the 5′-dGMP and adding a labeled dGMP to the 3′ nucleotide. Polymerization will then stop because the nucleoside triphosphate corresponding to the next residue is not present in the incubation mixture. When T5 DNA was labeled in this manner, the amount of radioactivity incorporated into acid-insoluble product increased with time until a plateau was reached at 60 min which corresponded to five dGMP residues added per DNA molecule. This correlates well with presence of four major nicks and one minor nick per molecule (10). Aliquots were removed from the incubation mixture at increasing times from 15 min until 3 h, and a nearest-neighbor analysis was performed on the labeled DNA. The ratio of Ap*/Gp* = 2 remained constant during the entire reaction (see Table 1). The 3′ external termini were not substrates for DNA polymerase I in this reaction because autoradiograms of agarose gels containing denatured labeled DNA (data not shown) showed that the largest fragment (i.e., the intact continuous strand) and one other fragment were not labeled during the incubation.

When 3′-labeled T5 DNA was partially digested with DNase II and spleen phosphodiesterase and run on two-dimensional homochromatography, a very complex pattern of oligonucleotides was obtained (Fig. 4). All spots corresponding to tetranucleotides or smaller were eluted, partially redigested, and run on a second two-dimensional homochromatography. However, the presence of sequence isomers in several spots precluded a complete sequence analysis at the 3′ termini of the nicks by this approach. Thus there appears to be much more sequence diversity on the 3′ side of the nick than on the 5′ side. Work is currently in progress to separate the labeled T5 DNA fragments prior to the 3′-terminal analysis.

DISCUSSION

The sequences determined in T5 DNA are summarized in Fig. 5. A common tetranucleotide sequence around the 5′ termini of the nicks is present, after which the sequence appears to diverge. However, one prominent sequence, [d(pGpCpGpCpGpGpTpG)], occurs past the level of tetranucleotide and may represent the sequence of more than one major fragment. In addition, the 3′-terminal residue is a purine with a ratio of adenosine/guanosine = 2. The minimum number of major and minor nicks per T5 molecule is five by electron microscope analysis (8), while this ratio suggests that the number per molecule is an integral of 3, the most likely of which would be 6. This in turn may suggest that the 3′-terminal residue of the four

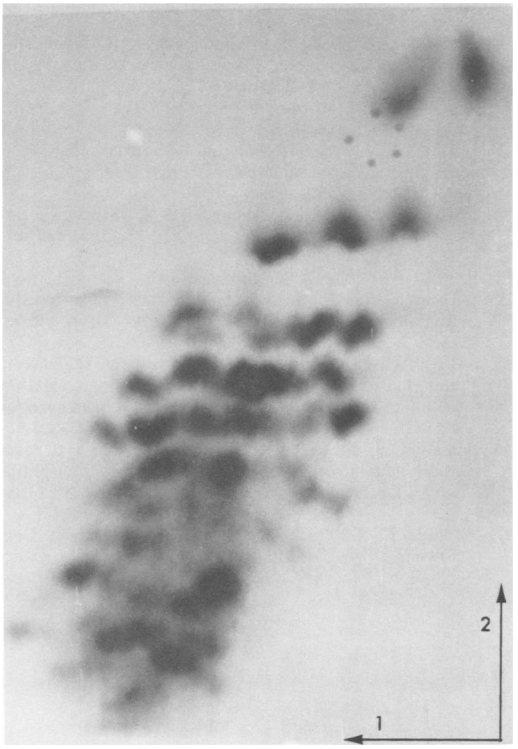

FIG. 4. *Autoradiogram of the fractionation (as in Fig. 1a) of a partial digest of 3'-terminally labeled T5 DNA.*

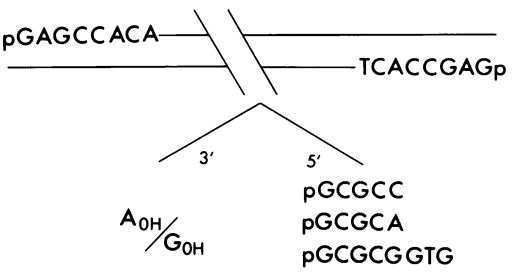

FIG. 5. *Summary of the sequence determined in T5 DNA. The orientation of the DNA molecule as depicted is arbitrary.*

major nicks is adenosine and that there are two minor nicks per molecule with guanosine as the 3' nucleoside.

No specific information was obtained about the sequence around the minor nicks. Some of the unanalyzed minor spots seen in Fig. 1a may arise from label being introduced into T5 DNA at the minor nicks. However, the sequences around the minor nicks may be very similar to those at the major nicks so that it would be very difficult to resolve the sequences for analysis.

The sequence homology found around the nicks suggests that at least the sequence -PupGpCpGpC- is recognized by a putative nuclease. This sequence would be expected to occur once every $2 \times 4^4 = 512$ base pairs, whereas a nick in T5 DNA occurs on the average every 24,000 base pairs (120,000 base pairs/5 nicks). Thus it seems likely that the entire recognition sequence is longer and involves specific base pairs still further removed from the actual cleavage site. For example, there may be as yet undetected sequence homology beyond the terminal nucleotide at the 3' side of the nicks.

At the external 5' termini, the first seven nucleotides of each strand are identical. This suggests the presence of a 14-nucleotide site of twofold rotational symmetry in a replicative concatamer. This sequence would be expected to occur once in every $4^{14} = 2.7 \times 10^8$ base pairs, and therefore is probably a unique sequence in T5 DNA. This site may be recognized by an endonuclease involved in the formation of monomer-length genomes from a replicative concatamer. If so, the recognition and

—AGTGGCTC | GAGCCACA—
—TCACCGAG | CTCGGTGT—

subsequent cleavage of a symmetrical site is reminiscent of the action of several restriction endonucleases.

Another model for generating monomer-length chromosomes from replicative concatamers takes into account the fact that about 10,000 base pairs (9% of the genome) adjacent to one end of the T5 chromosome is a terminal repetition of the other end (13). This model suggests that in T5 DNA, two single-stranded cleavages are introduced on opposite strands of the replicative concatamer about 10,000 base pairs apart. Monomer-length chromosomes with completely duplex termini are then formed by DNA polymerase-mediated repair either during or after separation from the rest of the DNA (see reference 17 for a review). If the model is correct, then the single-stranded cleavages occur at the above sequence and would be the first known case of a specific single-stranded cleavage occurring within a sequence possessing a twofold axis of symmetry.

ACKNOWLEDGMENTS

This work was supported by Public Health Service grant GM-21696 from the National Institute of General Medical Sciences and National Science Foundation grant PCM 76-13461.

LITERATURE CITED

1. **Abelson, J., and C. A. Thomas.** 1966. The anatomy of the bacteriophage T5 DNA molecule. J. Mol. Biol. 18:262-291.
2. **Barrell, B. G.** 1971. Fractionation and sequence analysis of radioactive nucleotides, p. 751-779. *In* G. L.

Cantoni and D. R. Davies (ed.), Procedures in nucleic acids research, vol. 2. Harper & Row, New York.
3. **Brownlee, G. G., and F. Sanger.** 1969. Chromatography of ^{32}P-labelled oligonucleotides on thin layers of DEAE cellulose. Eur. J. Biochem. 11:395–399.
4. **Bujard, H.** 1969. Location of single-strand interruptions in the DNA of bacteriophage T5$^+$. Proc. Natl. Acad. Sci. U. S. A. 62:1167–1174.
5. **Donelson, J. E., B. G. Barrell, H. L. Weith, H. Kossel, and H. Schott.** 1975. The use of primed synthesis by DNA polymerase I to study an intercistronic region of ϕX-174 DNA. Eur. J. Biochem. 58:383–395.
6. **Hayward, G. S.** 1974. Unique double-stranded fragments of bacteriophage T5 DNA resulting from preferential shear-induced breakage at nicks. Proc. Natl. Acad. Sci. U. S. A. 71:2108–2112.
7. **Hayward, G. S., and M. G. Smith.** 1972. The chromosome of bacteriophage T5. I. Analysis of single-stranded DNA fragments by agarose gel electrophoresis. J. Mol. Biol. 63:383–395.
8. **Hayward, G. S., and M. G. Smith.** 1972. The chromosome of bacteriophage T5. II. Arrangement of single-stranded DNA fragments in T5$^+$ and T5 st(0) chromosomes. J. Mol. Biol. 63:397–407.
9. **Jacquemin-Sablon, A., and C. C. Richardson.** 1970. Analysis of the interruptions in bacteriophage T5 DNA. J. Mol. Biol. 47:477–493.
10. **Johnston, J. V., B. P. Nichols, and J. E. Donelson.** 1977. Distribution of "minor" nicks in bacteriophage T5 DNA. J. Virol. 22:510–519.
11. **Kornberg, A.** 1969. Active center of DNA polymerase. Science 163:1410–1418.
12. **Ling, V.** 1972. Fractionation and sequences of the large pyrimidine oligonucleotides from bacteriophage fd DNA. J. Mol. Biol. 64:87–102.
13. **Rhoades, M., and E. A. Rhoades.** 1972. Terminal repetition in the DNA of bacteriophage T5. J. Mol. Biol. 69:187–200.
14. **Rogers, S. G., and M. Rhoades.** 1976. Bacteriophage T5-induced endonucleases that introduce site-specific single-chain interruptions in duplex DNA. Proc. Natl. Acad. Sci. U. S. A. 73:1576–1580.
15. **Sanger, F., J. E. Donelson, A. R. Coulson, H. Kössel, and D. Fischer.** 1975. Determination of a nucleotide sequence in bacteriophage f1 DNA by primed synthesis with DNA polymerase. J. Mol. Biol. 90:315–333.
16. **Schendel, P. F., and R. D. Wells.** 1973. The synthesis and purification of [γ-^{32}P] adenosine triphosphate with high specific activity. J. Biol. Chem. 248:8319–8321.
17. **Thomas, C. A., Jr.** 1967. The rule of the ring. J. Cell Physiol. 70(Suppl. 1):13–34.