THE 3'-TERMINAL NUCLEOTIDE SEQUENCES OF λ DNA

Donald P. Brezinski and James C. Wang

Group in Biophysics and Department of Chemistry
University of California
Berkeley, California 94720

Received November 20,1972

Summary: The base sequences of the 3'-termini of coliphage λ DNA have been analyzed by a new technique. Escherichia coli DNA polymerase I was used to add a single radioactive nucleotide to the 3'-OH terminus of one of the DNA strands. The DNA was then digested with pancreatic DNase I, and the resulting oligonucleotides were separated by two dimensional ionophoresis. Terminal oligonucleotides were identified by the presence of the radioactive label, and the base sequence of the labelled terminus was deduced from the base compositions of the terminal di-, tri-, tetra-, etc., oligonucleotides. It is found that the left 3'-terminus of λ DNA ends with the sequence d(pCpGpCpG) and the right 3'-terminus ends with the sequence d(pCpG).

In recent years several techniques for the sequencing of terminal regions of DNA have been developed (1-10). The procedures developed by Wu et al. (4-7) and by Englund (9-10) are capable of yielding sequences of considerable length. Both methods, however, are fairly tedious and require several consecutive experiments and careful quantitative analysis to sequence a single terminus. In this communication, we report a technique which allows the determination of a short terminal DNA sequence in a single, non-quantitative experiment. The technique could readily be extended to the sequencing of longer terminal sequences. Results on the 3' terminal sequences obtained for λ DNA are presented.

Experimental

λ DNA, unlabelled or uniformly labelled with ^{32}P, was extracted from the phage obtained by thermal induction of E. coli N831 (λcI857 S7 ind$^-$), a strain given to us to Dr. B. Sutherland. Tritiated nucleotides were added to the 3'-OH ends of λ DNA by a procedure similar to that used by Wu (5). Details are given in the legend to Figure 1. λ DNA was sheared into half molecules by repeated passage through a precision-bore glass capillary (11), and the halves were separated by Hg(II)-$CsSO_4$ density gradient centrifugation (12). λ DNA was digested into oligonucleotides by incubating 20 μg of DNA in 0.25 ml of 10^{-3} M $MgSO_4$, 10^{-3} M Tris (pH 7.4) with 40 μg bovine pancreatic DNase I (Worthington DPFF) for 24 hours at 30°C. The reaction was terminated by the addition of 2 μl of 0.2 M Na_3EDTA, and incubation for 5

min at 80°C. DNase I digests of a mixture of λ DNA terminally labelled with ^{3}H and λ DNA uniformly labelled with ^{32}P were concentrated by evaporation under a stream of N_2 and analyzed by two dimensional ionophoresis on cellulose acetate and DEAE paper. For procedural details see Sanger et al. (13). ^{32}P decay allowed the oligonucleotide spots on the DEAE-cellulose sheet to be located by radioautography against x-ray film (14). To locate ^{3}H-labelled oligonucleotides, all the spots were cut out and either counted directly in a liquid scintillation mixture or, for higher ^{3}H counting efficiency, the nucleotides were first eluted from the spots with 30% triethylamine carbonate, dried, and counted in a scintillation mixture containing aqueous solubilizer (see Table 1 for details).

For the purposes of base composition analysis, a DNase I digest of 20 μg of high specific activity ^{32}P-labelled λ DNA (3×10^5 cpm/μg) was subjected to two-dimensional ionophoresis. The various oligonucleotide spots of interest were cut out and eluted with 30% triethylamine carbonate. The base composition of each eluate was analyzed by digestion with venom phosphodiesterase I (Worthington) and ionophoresis on Whatmann #1 paper at pH 3.5 as described elsewhere (13,14). Cold deoxynucleotides were added as carriers.

Results

According to the results of Wu and Kaiser, in the presence of dATP under repair conditions, a single adenosine residue is added to the left 3' terminus of λ DNA by E. coli DNA polymerase I (6). This is confirmed by the results depicted in Figure 1. Shear breakage of λ DNA terminally labelled with ^{3}H-dA followed by Hg(II)-Cs_2SO_4 density gradient centrifugation showed that the ^{3}H activity is only associated with the left (A+T rich) half. The same procedure applied to λ DNA repaired with ^{3}H-dGTP showed that ^{3}H activity was only associated with the right (G+C rich) half.

Figure 2a depicts a typical two dimensional radioautogram of a DNase I digest of a mixture of ^{3}H-dA terminally-labelled λ DNA and uniformly ^{32}P-labelled λ DNA. Areas on the DEAE-cellulose sheet corresponding to the spots on the radioautogram were analyzed as described in the Experimental section, and the results are tabulated in Table I, column a. The spots which contain a significant level of ^{3}H activity are No. 12, 13, 49, and 50. Spot No. 12 contained a mixture of (AG) and (ACT), and spot No. 13 contained a mixture of (AGC) and (ACT). The appearance of (ACT) in both spots is probably due to sequence isomerization. These components were easily resolved by elution and subsequent ionophoresis on DEAE-cellulose wetted with 5% acetic acid, 0.5% pyridine, pH 3.5. The oligonucleotides (AG), (ACT), and (AGC) migrated

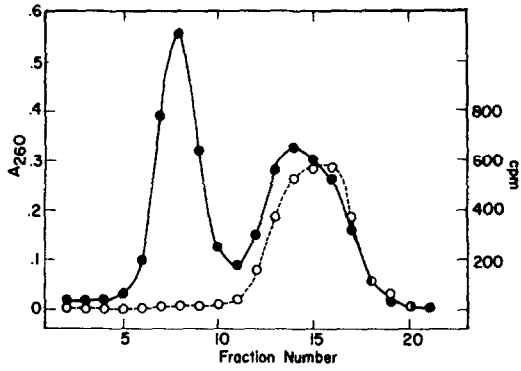

Figure 1. Banding pattern of sheared λ DNA halves in a Hg(II)-Cs_2SO_4 density gradient. The DNA had been terminally labelled with ^{3}H-dA prior to shear breakage by incubation with ^{3}H-dATP and DNA polymerase I as follows: A mixture of 20 µl of 11 mg/ml bovine serum albumin (nuclease free), 50 µl of 105 mM potassium phosphate (pH 6.9), and 50 µCi of ^{3}H-dATP (18.5 Ci/mmole, Schwarz/Mann) in 0.1 ml of 50% ethanol was dried at 0°C under a jet of nitrogen. A 0.6 ml sample of 0.6 mg/ml λ DNA in 105 mM potassium phosphate buffer (pH 6.9) was heated for 5 min at 74°C to disaggregate the cohesive ends and immediately quenched in a -100°C ethanol bath. The frozen DNA solution was melted in a 5°C water bath, cooled on ice, added to the dry ^{3}H-nucleoside triphosphate mix, and supplemented with 0.3 ml of a stock solution of 30 mM $MgSO_4$, 45 mM dithiothreitol, 240 mM NaCl. After thorough mixing, approximately 5 units of E. coli DNA polymerase I (18,000 u/mg, a gift from Dr. D. Brutleg) was added and the solution was incubated for 10 min on ice. The reaction was terminated by mixing in an equal volume of ice-cold 4 M NaCl, 0.05 M EDTA (pH 8) and heating at 80°C for 5-10 min. Unreacted ^{3}H-dATP was removed by exhaustive dialysis against 2 M NaCl, 0.01 M EDTA (pH 8). The same procedure was used to label the right 3'-OH terminus with ^{3}H-dGTP (13.4 Ci/mmole, Amersham/Searle), except that the incubation time was extended to 20 min.

After shearing (11), 50 µg of λ DNA halves were prepared for banding in 2 ml of $CsSO_4$ solution according to the procedure in (12). The solution was overlaid with silicone oil and centrifuged at 30,000 rpm and 20°C for two days in a Spinco Model L ultracentrifuge (SW 50 rotor). The tube was dripped by punching a hole at the bottom. Each fraction was two drops diluted with 0.2 ml of H_2O. (●), A_{260} mµ of fractions measured with a 1 cm light-path microcell. (○), ^{3}H counts with 50 µl of each fraction diluted to 0.6 ml with H_2O, added to 10 ml of scintillation mixture containing 10% Beckmann Biosolve BBS-3 aqueous solubilizer, and counted in a liquid scintillation counter.

---

with respective mobilities of ∿0.8, 0.7, and 0.5, relative to a xylene cyanol F.F. (blue) dye marker. ^{3}H activity was associated only with the oligonucleotides (AG) and (AGC). Spot 49 has base composition (AG_2C). Spot 50 was multicomponent but did not resolve further on DEAE paper at pH 3.5. Its position in the two-dimensional ionophoretogram indicates, however, that the ^{3}H-labelled component of this spot corresponds to the addition of a C to the ^{3}H-containing component of spot 46 (AG_2C). (See legend to Figure 2.) The ascending series of ^{3}H-dA terminally labelled oligonucleotides, (AG),

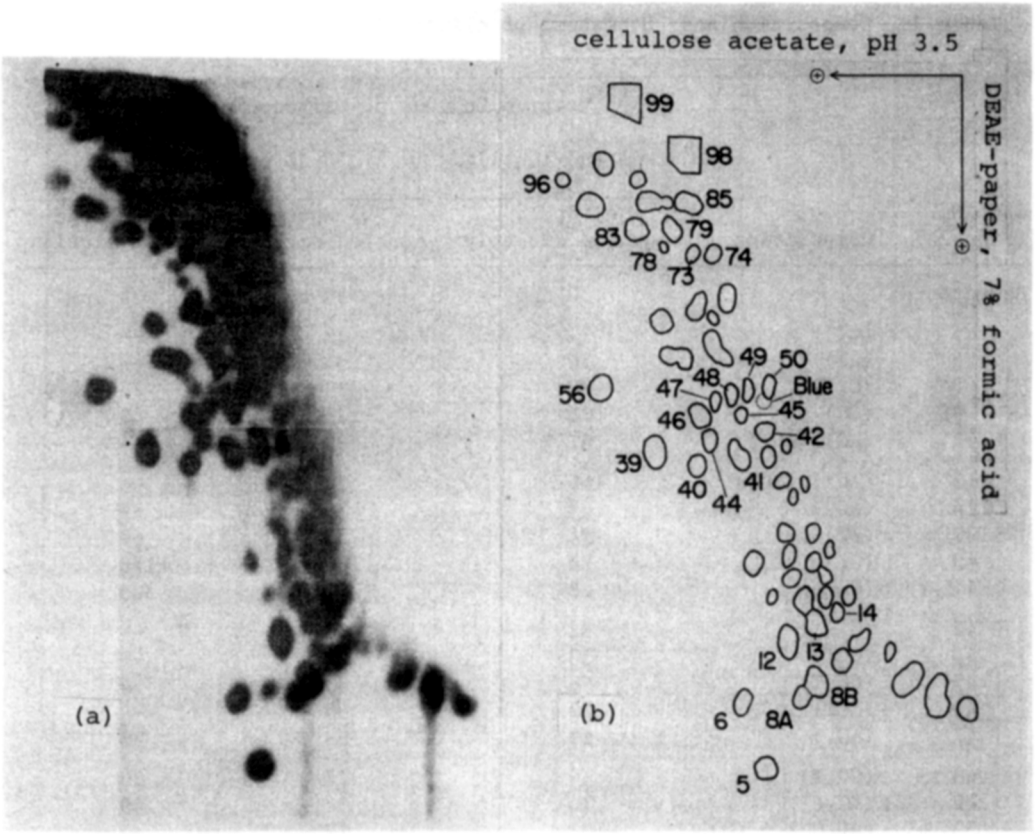

Figure 2. (a) Nucleotide "map" of a pancreatic deoxyribonuclease digest of λ DNA after two-dimensional ionophoresis. 20 μg of λ DNA terminally labelled with ^{3}H-dA was mixed with 3×10^{5} cpm of λ DNA uniformly labelled with ^{32}P, and digested with DNase. Ionophoresis of the digest was first carried out on a 2 × 22 inch cellulose acetate strip (Millipore) for 3.5 hr at 2500 volts. The buffer used was 7 M deionized urea, 10^{-3} M EDTA, adjusted to pH 3.5 with glacial acetic acid. Ionophoresis in the second dimension was on a 18 × 36 inch sheet of Whatmann DE81 DEAE-cellulose paper for 14 hours at 1300 volts. The buffer used was 7% (w/v) formic acid. Oligonucleotides on the DEAE cellulose paper were visualized by radioautography in folders against 14 × 17 inch sheets of Kodak No-Screen x-ray film. Exposure was ∿4 days. (b) A line drawing of some of the more prominent nucleotides in (a). The dotted zone denotes the position of the xylene cyanol FF (blue) dye marker. Analysis of the numbered nucleotides is given in Table 1. Note that near the center of the pattern, the addition of a C to an oligonucleotide shifts its position slightly to the right, e.g., Spot 40 gets shifted to 41 and 47 gets shifted to 49. Addition of the nucleotides G, A, or T causes larger shifts in an upward direction. These shifts were often consistent enough to predict base compositions of unknown spots in uncongested areas of the map from their positions relative to spots of known composition.

(AGC), (AG_2C), and (AG_2C_2), indicates a sequence of d(pCpGpCpGpA) for the left 3' terminus, because the nth base of the terminal n-mer must be the base which, when added to the base composition of the terminal (n-1)-mer,

Table 1. Composition and ^{3}H content of oligonucleotides shown in Figure 2

Spot No. (Fig. 2b)	Composition	Tritium Content of Oligonucleotides: ^{3}H-dA end-labelled DNA a) Spots counted directly	^{3}H-dG end-labelled DNA b) Spots counted directly	^{3}H-dG end-labelled DNA c) Counted after elution
Background	--	∿30 cpm	∿30 cpm	∿30 cpm
5	(CT)	31	29	--
6	(AT)	34	31	--
8A	(C_2T)	28	29	28
8B	(GC)	28	31	43
12	(AG),(ACT)	61	30	--
13	(ACT),(AGC)	160	29	--
14	(AGC)	33	29	--
39	(GT)	34	36	67
40	(G_2)	26	160	1173
41	(G_2C)	28	60	303
42	(G_2C_2)	32	32	--
44	(GCT)	30	32	--
45	(GC_2T)	--	33	34
46	(GCT)	32	30	--
47	(AG_2)	29	28	28
48	(GC_2T)	--	27	30
49	(AG_2C)	74	34	49
50	(AG_2C_2) + (?)	101	29	--
56	(T_2)	31	32	--
73	(G_3C)	29	46	174
74	--	30	35	62
78	(G_3)	31	349	3172
79	--	31	40	119
83	(G_2T)	28	39	63
85	--	32	46	170
96	(T_3)	30	29	--
98	--	--	133	--
99	--	--	111	--

All the spots on a DEAE-cellulose sheet (e.g. Figure 2a) were cut out, placed in glass scintillation vials containing 10 ml of 8 g/l butyl PBD, 0.5 g/l PBBO toluene-based scintillation mixture (Beckman) and counted in a Beckman LS-250 Liquid Scintillation Counter. Spill-over of ^{32}P counts into the ^{3}H channel was low (1.3%) and the presence of ^{3}H in most spots could easily be determined. This procedure was used to obtain the data in columns a and b. For ∿10 times greater tritium efficiency, the spots were eluted by soaking them in 3 ml of 30% triethylamine carbonate (pH 10) containing 60 µg/ml of DNase-digested calf thymus DNA in glass scintillation vials for 48 hrs. The spots were removed, the vials were taken to dryness in a partial vacuum, and the residue was dissolved in 20 µl of 0.2 M NaCl. 10 ml of scintillation mixture containing 0.5% Biosolve BBS-3 (Beckman) was added to each vial. The vials were shaken until clear and counted for both ^{3}H and ^{32}P. This procedure was used to obtain the data in column c. All spots not listed in Table 1 had tritium counts below 35 cpm. All counts were for 10 minutes. A background of ∿30 cpm was registered by the counter in the absence of ^{3}H.

yields the base composition of the terminal n-mer. Since the dA nucleotide was added by the polymerase, the left 3' terminal sequence of native λ DNA is d(pCpGpCpG).

The right 3' terminus of λ DNA was sequenced in a similar way. The fact that as many as three G's can be added to the right end, however, introduces some complications. Tritium was observed in spots No. 40 (G_2), 41 (G_2C), 78 (G_3), 73 (G_3C), 79, 85, and in two areas vertically above spot 85 where the resolution was poor (Table 1, columns b and c).

The possible tritiated oligonucleotides that could be found in Spot 41 are d(pCpGpG*), d(pCpG*pG*), and d(pGpCpG*), where pG* designates a ^{3}H-dGMP residue. d(pCpG*pG*) and d(pGpCpG*) would imply 3'-OH terminus sequences ...pC and ...pGpC respectively. Note, however, that either of these terminal sequences should lead to the formation of the species d(pCpG*) in the digest, which would be found in spot 8B. The fact that no tritiated d(pCpG*) was found indicates that the tritiated species in spot 41 must be d(pCpGpG*), implying a native right 3'-OH terminal sequence of d(...pCpG). This is in agreement with the findings of Wu and Kaiser (4) that the 3'-OH terminus of each strand of λ DNA is a G. Spots 78 and 73 could correspond to the addition of one more terminal dG* to spots 40 and 41, respectively, and therefore do not yield additional sequence information. Oligonucleotide spots 42 (G_2C_2) and 49 (AG_2C) contained very little tritium, so the third nucleotide of the terminal sequence is most likely neither a C nor an A. Note that the presence of d(pGpCpGpG*) cannot be excluded, since its position may coincide with that of d(pCpGpG*pG*). Spot 85 appeared to be composed of several poorly-resolved components. Subsequent ionophoresis at pH 3.5 on DEAE-cellulose paper provided no additional separation. Spot 79 has not been analyzed.

Based on the above information, the right 3' terminus of native λ DNA is d(pXpCpG), where X is probably either a G or a T, to be determined by further experimentation.

With both the ^{3}H-dA and ^{3}H-dG end-label experiments, ∿45% of the ^{3}H counts in the digest were recovered from the DEAE paper.

Discussion

The method of sequencing presented in this paper has some inherent advantages. None of the procedures require much quantitative precision. Only those oligonucleotides which contain tritium label need be analyzed for base composition, and once most of the oligonucleotides have been identified in the two-dimensional pattern, it is unnecessary to work with high levels of ^{32}P. Under favorable circumstances, a terminal sequence of 3 or 4 bases

can be elucidated in a single short experiment. It should be possible to extend the technique to larger terminal sequences through the judicious use of a DNA polymerase "proofreading function" to degrade the 3'-OH terminus to a new fixed endpoint (15).

The base sequences of the 3' termini are interesting in connection with the recognition of the cohesive end sites of λ DNA by the enzyme system which generates the cohesive ends (16,17). One of the simplest schemes for this recognition process would have both nicking sites recognized and nicked by the same enzyme system acting separately on the two sites (18). This scheme predicts two identical sites flanking or straddling the generated nicks. Even if nicking is accomplished by a single act of recognition accompanied by a coordinated nicking of both sites, the joint recognition of two identical sequences separated by a fixed number of base pairs would halve the complexity of the protein devoted to site recognition with no loss of specificity. The present results are intriguing in that the 3' terminal sequence of the left end, ...pCpGpCpG, and that of the right end, ...pCpG, might be very similar. It remains to be seen whether this is coincidental or is of functional significance.

Acknowledgement

This work has been supported by grants from the Public Health Service (GM 14621) and the National Science Foundation (GB 29331). D. P. B. has been supported by a National Science Foundation Predoctoral Fellowship and U. S. Public Health Service Training Grant 5 TO1-GM-00829-11.

References

1. Richardson, C. C. (1966). *J. Mol. Biol.* 15, 49.
2. Wu, R. and A. D. Kaiser. (1967). *Proc. Natl. Acad. Sci. U.S.* 57, 170.
3. Weiss, B. and C. C. Richardson. (1967). *J. Mol. Biol.* 23, 405.
4. Wu, R. and A. D. Kaiser (1968). *J. Mol. Biol.* 35, 523.
5. Wu, R. (1970). *J. Mol. Biol.* 51, 501.
6. Wu, R. and E. Taylor. (1971). *J. Mol. Biol.* 57, 491.
7. Padmanabhan, R. and R. Wu. (1972). *J. Mol. Biol.* 65, 447.
8. Salser, W., K. Fry, C. Brunk, and R. Poon. (1972). *Proc. Natl. Acad. Sci. U.S.* 69, 238.
9. Englund, P. T. (1971). *J. Biol. Chem.* 246, 3269.
10. Englund, P. T. (1972). *J. Mol. Biol.* 66, 209.
11. Yew, F. F. H. and N. Davidson. (1968). *Biopolymers* 6, 659.
12. Wang, J. C., U. S. Nandi, D. S. Hogness, and N. Davidson. (1965). *Biochemistry* 4, 1697.
13. Sanger, F., G. C. Brownlee, and B. G. Barrell. (1965). *J. Mol. Biol.* 13, 373.
14. Murray, K. (1970). *Biochem. J.* 118, 831.
15. Brutleg, D. and A. Kornberg. (1972). *J. Biol. Chem.* 247, 241.
16. Mousset, S. and R. Thomas. (1969). *Nature* 221, 242.
17. Yarmolinsky, M. B. (1971). In *The Bacteriophage Lambda*, ed. by A. D. Hershey, p. 97. Cold Spring Harbor Laboratory, N.Y.
18. Wang, J. C. and N. Davidson. (1968). *Cold Spr. Harb. Symp. Quan. Biology* 33, 409.