NUCLEOTIDE SEQUENCE ANALYSIS OF DNA

XI[+]. The 3' Terminal Sequences of Bacteriophage λ and φ80 DNA

Gurdev S. Ghangas, Ernest Jay, Robert Bambara and Ray Wu

Section of Biochemistry, Molecular and Cell Biology
Cornell University
Ithaca, New York 14850

Received August 11, 1973

SUMMARY: Two procedures have been developed and applied to the determination of the 3' terminal sequences of λ DNA and φ80 DNA. In the first procedure, each 3' terminus was specifically labeled with a single ^{32}P-nucleotide. Radioactive oligonucleotides of different lengths were obtained by partial pancreatic deoxyribonuclease digestion. From the characteristic mobilities of these oligonucleotides in two dimensional fractionation systems, the 3' terminal sequence -ACCCGCG for the r-strand and -GGTTACG for the l-strand of λ DNA have been determined. In the second procedure, approximately six nucleotides were removed from each 3' terminus with exonuclease III, and they were replaced with radioactive nucleotides by partial repair synthesis. After enzymatic digestion and sequence analysis, the above sequences have been confirmed. The 3' terminal sequences in φ80 DNA are identical to those in λ DNA at least up to the fifth nucleotide from the 3' ends.

After replication of bacteriophage λ DNA, linear monomeric DNA molecules with cohesive ends (1) are produced presumably by action of a specific endonuclease (2,3). This specific endonuclease probably recognizes a unique sequence of nucleotides near the cohesive ends and introduces one nick for each recognition sequence. Two nicks, 12 base-pairs apart, are made on opposite strands to create the cohesive ends of the mature phage DNA. The recognition sequence is assumed to be present at both sides of the cleavage point. One part of the sequence is within the cohesive ends (segments SI and SIII in Figure 1), and the other part is at the 3' termini adjacent to the cohesive ends (segments SII and SIV). The cohesive end sequences of λ DNA and φ80 DNA molecules are known (4,5), and recently nucleotide sequences corresponding to

---

[+] Paper X in this series is Bambara, Padmanabhan and Wu (5). This work was supported by Grants GB-31485X from N.S.F. and GM-CA 18887 from N.I.H.

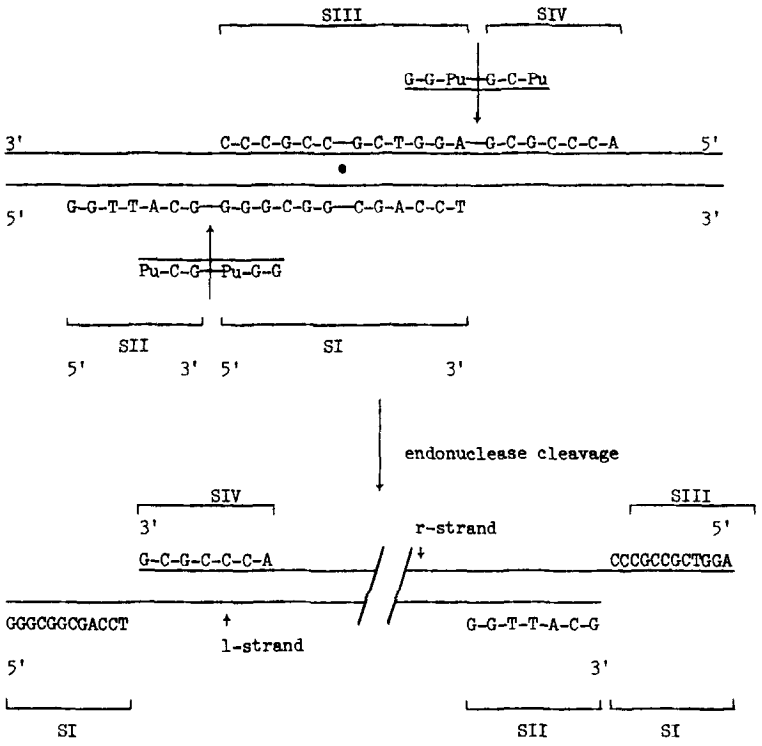

Figure 1 The Cleavage of Circular or Dimeric λ DNA to Produce Cohesive Ends.

A segment of a circular or dimeric λ DNA molecule is given in the upper panel. This is cleaved by a specific endonuclease to produce linear monomeric λ DNA with cohesive ends as shown in the lower panel. Segments SI and SIII represent the two cohesive ends, and segments SII and SIV represent sequences at the 3' termini of the DNA. Arrows in the upper panel indicate the points of cleavages to be made by the specific endonuclease.

The 3' terminal nucleotide in segment SII is joined to the 5' nucleotide in segment SI in the linear dimer, concatemer as well as circular monomer. The nucleotide sequence we have found at the 3' terminus of the l-strand is pGpGpTpTpApCpG-OH (Segment SII), that of the r-strand is pApCpCpCpGpCpG-OH (Segment SIV). The endonuclease recognition sequences are proposed to be the hexanucleotides pu-C-G-pu-G-G which are present in each strand (where pu stands for purine). The two hexanucleotides in the upper panel have a two-fold axis of symmetry. The axis of symmetry is perpendicular to the page and is represented by the dot.

part of the sequence in segments SII and SIV in λ DNA have been reported (6,7). In this communication, we report the successful development of two independent procedures for DNA sequence analysis. These procedures have been applied to the analysis of the sequence in segments SII and SIV of both λ DNA and ϕ80 DNA. A preliminary report has appeared (8).

RESULTS AND DISCUSSION

Sequence analysis through labeling of a 3' terminus - In this method, only

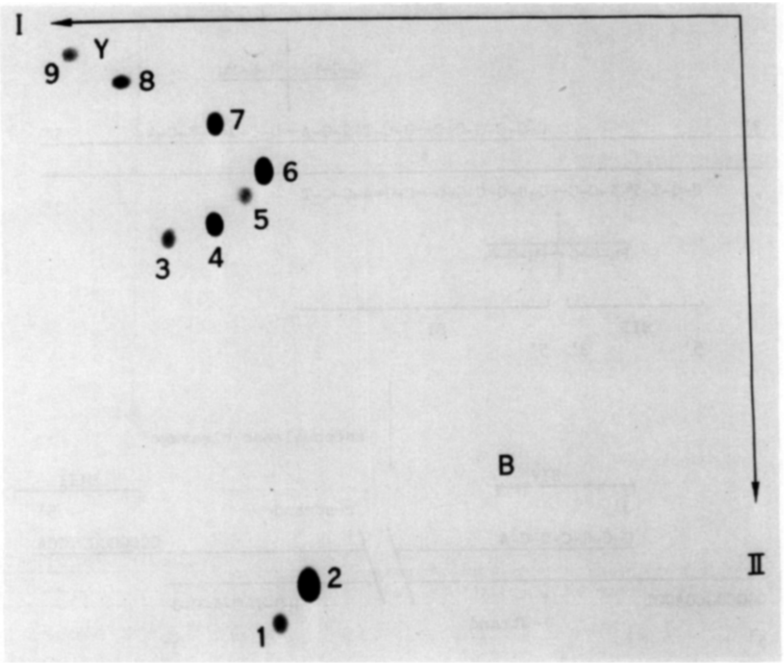

Figure 2 Two Dimensional Fingerprint of ^{32}P-A Labeled Pancreatic DNase Fragments from λ DNA Labeled at the Left-Hand 3' End with ^{32}P-dAMP.

Dimension I: electrophoresis on cellulose acetate strip in pyridine-acetate at pH 3.5; Dimension II: electrophoresis on DEAE cellulose paper in acetic acid-formic acid buffer at pH 1.9 (11). One hundred μg of λ DNA was labeled at the 3'-terminus of the r-strand with one *pA by partial repair synthesis with α-^{32}P-dATP and DNA polymerase (9). The labeled λ DNA was purifed by alkaline sucrose gradient centrifugation followed by digestion with different levels of pancreatic DNase (0.4, 1.0, 2.0 and 6.0 μg) in 0.2 ml volume containing 0.1 M NH_4HCO_3, pH 8.0, 0.01 M $MgCl_2$ and 12 μg of carrier sperm salmon DNA. After 3 hours at 37°C, the digests were combined, dried and fractionated by 2-dimensional electrophoresis. All the oligonucleotides 1 through 9 gave the correct G*pA nearest neighbor affirming that they are of the same family derived from the left-hand 3' end. The identity of the spot 1 was confirmed by co-electrophoresis with a synthetic pGA in both pH 1.9 and pH 3.5 systems. Identical results were obtained when φ80 DNA was used in place of λ DNA. (B) and (Y) mark the positions of the blue and yellow dye markers, respectively (11).

---

one of the 3' terminal nucleotides of a DNA chain is specifically labeled with a ^{32}P-nucleotide using a repair reaction (9). This nucleotide is directed by the nucleotide in the complementary strand which serves as a template. In this procedure, the labeled nucleotide is always added to the natural 3' terminus of a DNA chain. Therefore, sequence information derived from this point already has its exact location on the DNA molecule defined.

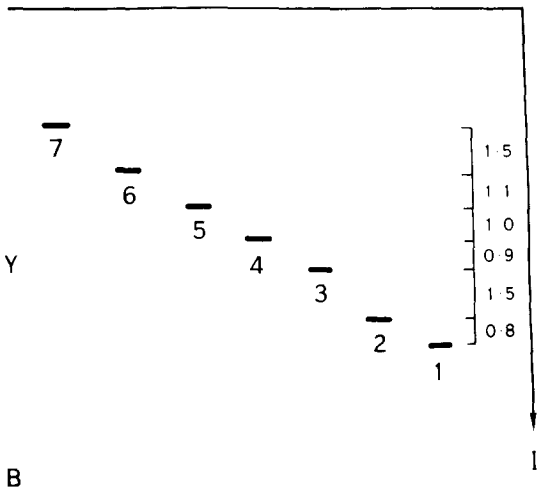

Figure 3 One Dimensional Homochromatography of ^{32}P-A Labeled Pancreatic DNase Fragments from λ DNA Labeled at the Left-Hand 3' End with ^{32}P-dAMP.

Oligonucleotides 1 through 7 shown in Figure 2 were eluted from the DE 81 paper and then applied onto thin layer DEAE cellulose plate together with known oligonucleotides as markers (see Figure 4 and legend). One dimensional chromatography was developed with a 2% partially hydrolyzed yeast RNA containing 7M urea. The lines indicate the center of each radioactive spot.

In the next step, the 3' labeled DNA molecule is degraded with different levels of pancreatic DNase to produce labeled oligonucleotides of varying lengths whose sequences overlap progressively from the 3' terminus (7,10). This family of oligonucleotides is then fractionated on two dimensional electrophoresis (11), or electrophoresis followed by homochromatography (12,13). From the characteristic mobilities of oligonucleotides in the two dimensional fractionation systems, the sequence of the different oligonucleotides in this family can be deduced.

Two fractionation schemes have been used to illustrate the usefulness of this method. In the first scheme, the family of 3'-p$\overset{*}{A}$ labeled oligonucleotides from the left-hand 3' end were fractionated on two dimensional electrophoresis (11). As shown in Figure 2, by comparing the mobilities of oligonucleotides 1 and 2, an additional pC in oligonucleotide 2 is indicated. Addition of either a pG or pT causes the mobility shift from oligonucleotides 2 to 3. To

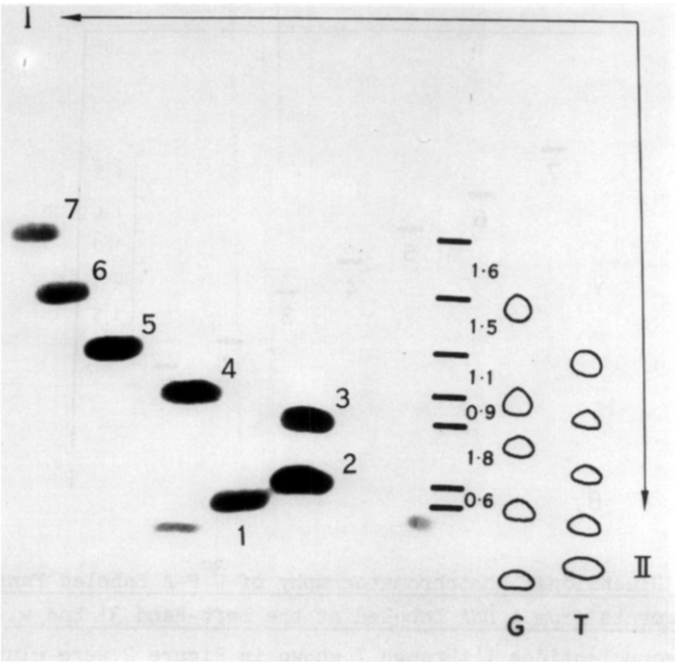

Figure 4 Two Dimensional Fingerprint of Pancreatic DNase Fragments from λ DNA Labeled at the Right-Hand 3' End with ^{32}P-rGMP.

Dimension I: electrophoresis on cellulose acetate strip in pyridine-acetate at pH 3.5; Dimension II: homochromatography on DEAE-cellulose plate in 2% partially hydrolyzed yeast RNA containing 7M urea (12,13). Twenty-five μg of λ DNA was labeled at the 3' terminus of the l-strand with p*-rG by α-^{32}P-rGTP and DNA polymerase (14,15). Under these conditions, one or two p*-rG are incorporated (15). The second p*-rG is removed by alkali and phosphatase treatment. (If dGTP were used for the repair reaction, 3 molecules of pG* would have been incorporated which would be undesirable for sequence analysis.) The DNA, terminally labeled with a single p*-rG, was subjected to pancreatic DNase digestion as in the legend of Figure 2. After the two dimensional fractionation, each oligonucleotide was eluted and subjected to phosphatase treatment and nearest neighbor analysis (4). All the major spots 1 through 7 gave Gp* as the only labeled nucleotide as expected. The other minor spots do not have Gp*G nearest neighbor and must be impurities. The identity of spot 1 was verified by co-electrophoresis with a synthetic pGG on DEAE-cellulose paper in both pH 1.9 and 3.5 systems. Markers (run in dimension II only) in series G represent the products from partial snake venom phosphodiesterase digestion of the synthetic oligomer pGGGCG* and series T those of the oligomer pTTTTT*. It is clearly evident that the mobility change for the loss or gain of a purine is about 1.6 times as large as that of a pyrimidine; this is also true for the oligonucleotides derived from labeled λ DNA. Identical results were obtained when φ80 DNA was used in place of λ DNA.

distinguish between pG and pT, as well as to confirm other nucleotides, one dimensional homochromatography was carried out as the second step in this analysis. From analysis of known oligonucleotides by homochromatography (Figure 3), the purine shifts are larger than the pyrimidine shifts by a fac-

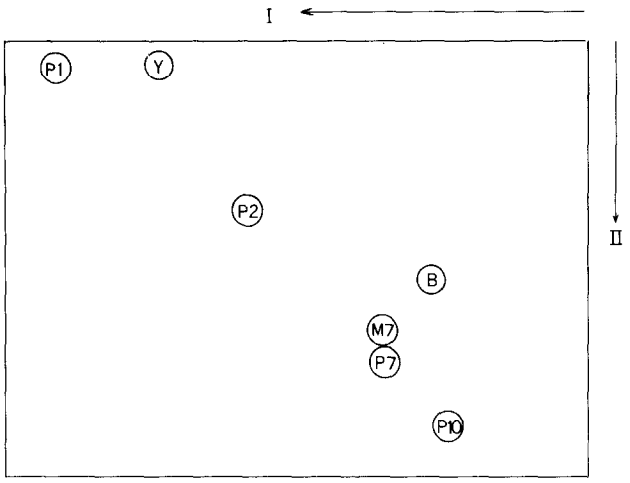

Figure 5 Two Dimensional Electrophoresis of Digests of DNA Labeled after Exonuclease III Treatment.

λ DNA (160 µg in 7 ml) was treated with exonuclease III (6 units) at 5°C to remove approximately 6 nucleotides from each 3' end (16). Labeled DNA was prepared with the use of 2 or 3 labeled nucleotides in partial repair reactions catalyzed by DNA polymerase. The labeled nucleotides used for repair synthesis are given under legend of Table II. In experiments PA and PB the DNA was digested with low levels of pancreatic deoxyribonuclease (2, 4 and 5 µg per 50 µg DNA in a final volume of 0.2 ml each. The buffer was 10 mM Tris-HCl, pH 7.8, and 5 mM $MgCl_2$). After incubation at 37°C for 2 hours, the samples were boiled for 5 minutes, combined and dried. In experiment M, the DNA was digested with micrococcal nuclease (0.2, 0.3 and 0.4 µg per 50 µg DNA in a final volume of 0.2 ml each. The buffer was 60 mM NH_4HCO_3, pH 8.6 and 0.33 mM $CaCl_2$). After incubation at 37°C for 20 minutes, the samples were boiled for 5 minutes, combined and dried. For electrophoresis, the first dimension was at pH 3.5 for 2.5 hours and the second dimension was at pH 1.9 for 15 hours (11). This figure is a composite figure of experiments M, PA and PB and shows only the heavily labeled oligonucleotides selected for further analysis.

---

tor of 1.6. The mobility shift from oligonucleotide 2 to 3 is large and is definitely a purine nucleotide shift. Therefore, it must be pG when the information from Figure 2 is also considered. Additional details on mobility shifts in homochromatography systems will be published elsewhere. Information given in these two figures thus established the sequence -pACCCGCG$\overset{*}{p}$A for oligonucleotide 7.

In the second scheme, the family of 3'-$\overset{*}{p}$rG labeled oligonucleotides from the right-hand 3' end were fractionated on a two dimensional system with electrophoresis as the first dimension and homochromatography as the second dimen-

TABLE I

Number of Additional Nucleotides Incorporated into Bacteriophage φ80 and λ DNA Cohesive Ends After Limited Exonuclease III Digestion

DNA from	Exonuclease III used (units)	Additional Nucleotides[a] Incorporated per DNA pG	pC	pA	pT	Total
φ80	0.3	2	3	0.9	2.1	8.0
	0.6	3	4	1.0	3.2	11.2
	0.9	4	6	1.4	3.2	14.6
λ	0.6	4	4	2.0	2.2	12.2
	0.8	5	5	2.4	2.6	15
Expected number[b]		4	5	1.0	2.0	12

DNA (15 μg in 0.6 ml) was incubated at 5°C for 30 minutes with varying amounts of exonuclease III under the conditions as described previously (6). The exonuclease III was inactivated after the addition of 2 mM EDTA by heating the solution to 70°C for 20 minutes. The cohesive ends and the newly exposed single-stranded regions were completely repaired with deoxynucleotides using DNA polymerase I at 5°C (9). In each of the 4 experiments, one nucleotide was labeled and the other 3 unlabeled.

[a] The table shows nucleotides incorporated per DNA molecule in excess of the 24 nucleotides incorporated into the cohesive ends of λ and φ80 DNA (4,5). Since 10 pG, 10 pC, 2 pA and 2 pT are incorporated into the cohesive ends of these DNA molecules, and these numbers are subtracted, the calculated values for additional pG and pC incorporation are somewhat less accurate. The symbols G, C, A, T are used throughout to stand for deoxynucleosides. When a ribonucleotide is mentioned, it is prefixed with r (e.g. prG = ribo-pG).

[b] These nucleotides are expected from the 3' terminal hexanucleotide sequences shown in Figure 1.

sion (12,13). As shown in Figure 4, oligonucleotide 1 is pG-$\overset{*}{p}$rG, and the addition of a pC causes it to shift to position 2. The larger mobility difference between oligonucleotides 2 and 3 in the second dimension and a very small difference in the first indicates the addition of a pA. By comparing oligonucleotides 3 and 4, the addition of a pT is indicated (a small shift in the second dimension). The structure of oligonucleotides 5, 6, 7 can be deduced in a similar manner. In this scheme, all the information necessary for establishing the entire sequence, -GGTTACG$\overset{*}{p}$rG, was obtained in a single fractionation scheme making this a faster method than those reported by others (4,6,7,10).

TABLE II

Sequence Analysis of Oligonucleotides from Both Ends of λ DNA Molecules Labeled After Exonuclease III Treatment

Expt.	Nucleoside triphosphates used[a]	Oligonucleotide number[b]	Products found after enzyme digestion[c]			Sequence deduced
M	$\dot{G}TP$, $\dot{T}TP$, $\overset{*}{C}TP$	M7	$1\dot{G}$,	$1\dot{G}\overset{*}{p}$,	$2\overset{*}{C}p$	$\overset{*}{C}p\overset{*}{C}p\overset{*}{C}p\dot{G}p\overset{*}{C}p\dot{G}$-OH[d]
PA	$\dot{G}TP$	P2	$1\dot{G}$,	$1\dot{G}\overset{*}{p}$,	$1\dot{C}p$	(p)$\dot{G}p\overset{*}{C}p\dot{G}$-OH
	$\dot{C}\overset{*}{T}P$	P7	$1\dot{C}$,	$1\dot{G}\overset{*}{p}$,	$2\dot{C}\overset{*}{p}$	(p)$\dot{C}p\dot{C}p\overset{*}{C}p\dot{G}p\dot{C}$-OH[e]
		P10	$1\dot{C}$,	$1\overset{*}{A}p$,	$1\dot{C}\overset{*}{p}$	(p)$\overset{*}{A}p\dot{C}p\dot{C}$-OH
PB	$\dot{A}TP$, $\dot{T}TP$, $\dot{C}\overset{*}{T}P$	P1	$1\dot{C}$,	$1\dot{A}\overset{*}{p}$,	$2\dot{T}p$	(pNpN)$p\dot{T}pTp\dot{A}p\dot{C}$-OH[f]

[a] The symbols * and · over the phosphate and nucleosides denote P^{32} and ^{3}H-labeled compounds, respectively.

[b] The oligonucleotides from pancreatic DNase or micrococcal nuclease digests were isolated from two dimensional electrophoresis as shown in Figure 5.

[c] The oligonucleotides P2, P7, P10 and P11 were treated with semen phosphatase to remove the 5' phosphate (9), boiled, and digested by spleen phosphodiesterase. The resulting nucleosides and nucleotides were separated and analyzed as previously described (4). Oligonucleotide M7 was digested with spleen phosphodiesterase without prior semen phosphatase treatment.

[d] The structure of this hexanucleotide is supported by that of oligonucleotides P2 and P7. It is also supported by $C\overset{*}{p}C\dot{p}G\overset{*}{p}Cp$ and $C\overset{*}{p}C\dot{p}G\overset{*}{p}$ (not shown here) found from the same micrococcal nuclease digest of Expt. M. Since oligonucleotide M7 came from a micrococcal nuclease digest and has a $\dot{G}$-OH at the 3' end, this oligonucleotide must be at the 3' end of the partially repaired DNA.

[e] By overlapping $\dot{G}\overset{*}{p}\dot{C}$ of oligonucleotides P2 and P7, the position of the internal pG in oligonucleotide P7 is defined.

[f] This oligonucleotide is likely to be a hexanucleotide as judged from its mobility on homochromatography (not shown). The nucleotides pN are probably pG as judged by the low mobility of this oligonucleotide in the second dimension of electrophoresis (Figure 5) and the lack of tritium counts in pN.

Sequence analysis after partial digestion of DNA with exonuclease III followed by partial repair synthesis - For this method, approximately six nucleotides were removed from each 3' terminus of the DNA molecule with exonuclease III. They were replaced with radioactive nucleotides by partial re-

pair synthesis. In order to remove only six nucleotides rather specifically from each 3' terminus, the exact amount of the enzyme necessary was established in a pilot run using several levels of exonuclease III. As shown in Table I, at 5°C, with 0.6 units of exonuclease III per 15 μg of DNA, approximately 12 nucleotides were removed as revealed by the total number of nucleotides incorporated into each molecule of φ80 DNA and λ DNA in a subsequent repair synthesis step (16). The number of each kind of additional nucleotide incorporated per DNA is generally in agreement with the expected values as judged from the 3' terminal hexanucleotide sequence shown in Figure 1. In order to produce labeled segments specifically from the region attacked by exonuclease III (segments SII and SIV in Figure 1) and not from the cohesive end region (segments SI and SIII), partial repair synthesis was then used instead of complete repair synthesis. Results from several experiments using different labeled nucleotides for partial repair synthesis are given in Figure 5 and Table II. Oligonucleotide M7, CpCpCpGpCpG-OH, appeared to be the hexanucleotide which replaced the segment removed by exonuclease III from the left-hand end of the DNA molecule. The sequences of oligonucleotides P2 and P7 support that of M7. The sequence of P10 indicates a pA as the seventh nucleotide at the 5' end of M7. Oligonucleotide P1 was probably incorporated into the right-hand end of the exonuclease III treated DNA. Thus, the sequence information obtained with the exonuclease III method is in full agreement with that obtained with the 3' terminal labeling method.

We propose that the nucleotide sequences recognized by the specific endonuclease which produces the cohesive ends of λ DNA and φ80 DNA are two identical hexanucleotides purine-C-G-purine-G-G, as shown in Figure 1 (8). These two hexanucleotides could be recognized by two identical subunits of a molecule of dimeric enzyme, and the hexanucleotides are cleaved right in the center. This type of recognition sequence is similar to the hyphenated symmetrical sequence proposed by Weigel et al. (7). Other types of recognition sequence are certainly possible.

REFERENCES

1. Hershey, A. D., Burgi, E. and Ingraham, L., Proc. Nat. Acad. Sci. USA, 49, 748 (1963).
2. Mousset, S. and Thomas, R., Nature 221, 242 (1969).
3. Wang, J. C. and Kaiser, A. D., Nature New Biology 241, 16 (1973).
4. Wu, R. and Taylor, E., J. Mol. Biol. 57, 491 (1971).
5. Bambara, R., Padmanabhan, R. and Wu, R., J. Mol. Biol. 75, 741 (1973).
6. Brezinski, D. and Wang, J. C., Biochem. Biophys. Res. Commun. 50, 398 (1973).
7. Weigel, P., Englund, P. T., Murray, K. and Old, R. W., Proc. Nat. Acad. Sci. USA 70, 1151 (1973).
8. Bambara, R., Ghangas, G. S., Jay, E. and Wu, R., Fed. Proc. 32, 664 Abs. (1973).
9. Wu, R., J. Mol. Biol. 51, 501 (1970).
10. Murray, K. and Murray, N. E., Nature New Biology 243, 134 (1973).
11. Sanger, F., Brownlee, G. G. and Barrell, B. G., J. Mol. Biol. 13, 373 (1965).
12. Brownlee, G. G. and Sanger, F., Europ. J. Biochem. 11, 395 (1969).
13. Ling, V., J. Mol. Biol. 64, 87 (1972).
14. Berg, P., Fancher, H. and Chamberlin, M., in "Informational Macromolecules," Academic Press, New York, p. 467 (1963).
15. Wu, R., Donelson, J., Padmanabhan, R. and Hamilton, R., Bull. L'Institut Pasteur 70, 203 (1972).
16. Donelson, J. D. and Wu, R., J. Biol. Chem. 247, 4661 (1972).
17. Wu, R. and Kaiser, A. D., J. Mol. Biol. 35, 523 (1968).