Proc. Nat. Acad. Sci. USA
Vol. 69, No. 11, pp. 3448–3452, November 1972

DNA Nucleotide Sequence Restricted by the RI Endonuclease

(symmetry/cohesive ends/RSV DNA polymerase/DNA–protein recognition/R-factor product/coliphage λ)

JOE HEDGPETH, HOWARD M. GOODMAN, AND HERBERT W. BOYER

Departments of Microbiology and Biochemistry, University of California at San Francisco Medical Center, San Francisco, Calif. 94122

Communicated by Paul Berg, September 18, 1972

ABSTRACT **The sequence of DNA base pairs adjacent to the phosphodiester bonds cleaved by the RI restriction endonuclease in unmodified DNA from coliphage λ has been determined. The 5′-terminal nucleotide labeled with ^{32}P and oligonucleotides up to the heptamer were analyzed from a pancreatic DNase digest. The following sequence of nucleotides adjacent to the RI break made in λ DNA was deduced from these data and from the 3′-dinucleotide sequence and nearest-neighbor analysis obtained from repair synthesis with the DNA polymerase of Rous sarcoma virus**

$$5'....A/T_pG \downarrow {}_pA_pA_pT_pT_pC_pT/A....3'$$
$$3'....T/A_pC_pT_pT_pA_pA_p \uparrow G_pA/T....5'$$

The RI endonuclease cleavage of the phosphodiester bonds (indicated by arrows) generates 5′-phosphoryls and short cohesive termini of four nucleotides, $_pA_pA_pT_pT$. The most striking feature of the sequence is its symmetry.

The restriction and modification enzymes of several host specificities (1–7), including the endonuclease and methylase of the host-specificity (RI) controlled by the *fi*$^+$ R-factor (Yoshimori, Roulland-Dussoix, Aldridge & Boyer; Yoshimori, Roulland-Dussoix, Goodman & Boyer, to be published) have been characterized. The restriction endonuclease of a given host specificity interacts with a specific sequence of DNA base-pairs and produces a double-strand break in the molecule (1–6), while the modification methylase methylates a base in each strand of this sequence (refs. 3 and 7; Yoshimori, Roulland-Dussoix, Goodman & Boyer, to be published). Methylation of a base in either strand is sufficient to prevent the endonuclease from attacking the sequence (3). Therefore, in molecular terms, a given DNA restriction and modification host-specificity can be defined by the sequence of base pairs recognized by the restriction endonuclease and modification methylase. In this paper, we present an analysis of the sequence of base pairs adjacent to the phosphodiester bond cleavages made in bacteriophage λ DNA by the RI restriction endonuclease.

MATERIALS AND METHODS

Enzyme. The RI restriction endonuclease, purified as will be described by Yoshimori, Roulland-Dussiox, Aldridge, and Boyer (to be published), was free of detectable nonspecific exo- or endonuclease activities and migrated as one band in native or Na dodecyl SO_4–polyacrylamide gels. Purified RNA-directed DNA polymerase of Rous Sarcoma virus, free of detectable exo- or endonucleolytic activities (8, 9), was a generous gift from Dr. A. J. Faras. Polynucleotide kinase (10) was purified as described elsewhere. *Escherichia coli* alkaline phosphatase rechromatographed as described elsewhere (11), pancreatic DNase, micrococcal nuclease, venom phosphodiesterase, and spleen phosphodiesterase were obtained from Worthington Biochemical Co.

Phage λ DNA. A λ ($C_{I857}sus S_7$) lysogen of *E. coli* 1100 was used as a source of λ phage. Tryptone broth cultures of lysogens (2×10^8 cell per ml) were heated to 42° for 0.5 hr (isotope was added here for labeled DNA) and incubated at 37° for 3 hr. The cells were collected by low-speed centrifugation and resuspended in 5–10 ml of 50 mM Tris·HCl (pH 7.5)–10 mM $MgCl_2$. 0.5 ml of $CHCl_3$ was added and the suspension was incubated at 37° for 15 min. After low-speed centrifugation, the viscosity of the supernatant was reduced by addition of pancreatic DNase (10–50 μg). The supernatant was centrifuged at 79,000 × *g* at 4°, for 1 hr. The pellet was resuspended in 50 mM Tris·HCl (pH 7.5) and layered on a CsCl step-gradient of the following composition: 1.7, 1.5, 1.3, 1.2 g/cm^3 CsCl and a top layer of 36% sucrose, all in 50 mM Tris·HCl (pH 7.5). The layered gradient was centrifuged for 2 hr at 41,000 × *g* and 10°. The phage, which banded at 1.5 g/cm^3 CsCl, was aspirated and recentrifuged to equilibrium in CsCl. The purified phage was dialyzed and extracted with phenol. After ether extraction of the phenol, the DNA preparation was dialyzed against 50 mM Tris·HCl (pH 7.5)–0.1 mM EDTA (TE buffer). Usually 90% of the labeled DNA prepared this way sedimented as one band in alkaline sucrose gradients. However, terminal labeling data (see *Results*) indicated that the unlabeled DNA used here had 2–3 internal nicks per λ molecule. DNA concentrations were determined spectrophotometrically; 6.6 A_{260} = 1 mM DNA nucleotide (12).

Terminal Labeling of DNA. RI endonucleolytic digestions of unmodified λ DNA were performed in 100 mM Tris·HCl (pH 7.5)–5 mM $MgCl_2$ for 30 min at 37°. After exhaustive endonucleolytic digestion the DNA was extracted with phenol and dialyzed against 10 mM Tris·HCl (pH 7.5)–1 mM EDTA. Terminal phosphoryl groups were removed by alkaline phosphatase at 37° (11), and the DNA was extracted with phenol and dialyzed against TE buffer. The 5′-terminal hydroxyls were phosphorylated with polynucleotide kinase (10), and [γ-^{32}P]rATP [synthesized by a modification of the procedure of Glynn and Chappell (13) and purified as described by Wehrli *et al.* (14), 2–3 Ci/mmol]. The terminally-

Abbreviations: RSV, Rous sarcoma virus; nucleosides without a prefix (A, T, G, C) are *deoxy*nucleosides.

labeled [^{32}P]DNA was extracted with phenol, dialyzed against TE buffer–0.5 M NaCl, and precipitated with ethanol.

Generation and Purification of Terminally Labeled Oligonucleotides. The terminally labeled [5′-^{32}P]DNA was mixed with uniformly labeled λ [^{33}P]DNA and digested with pancreatic DNase and venom phosphodiesterase (15). The digest was adsorbed to a DEAE-Sephadex A-25 column (0.9 × 20 cm) equilibrated with 5 mM Tris·HCl (pH 7.5)–7 M urea, and maintained at 65° (16). Oligonucleotides of increasing chain length were eluted with a linear gradient of (0–0.4 M) NaCl, desalted on small DEAE-Sephadex columns, and evaporated to dryness by flash evaporation.

RSV DNA Polymerase Reactions. RI-digested and undigested λ DNA were used as primer–templates for the RSV DNA polymerase with various combinations of [α-^{32}P]deoxyribonucleotide triphosphates (8–20 Ci/mmol, ICN Corp.) as substrates. A typical reaction contained 0.1 M Tris·HCl (pH 8.0), 2 mM 2-mercaptoethanol, 10 mM $MgCl_2$, 10 μM NTP, and 0.0002 units of RSV DNA polymerase per μg of DNA. The enzyme specific activity was 0.1 unit/μg of enzyme protein (8). Incubations were for 2 hr at 37°. An aliquot of the reaction was precipitated with perchloric acid to determine the total incorporation. The remainder was extracted with phenol, and dialyzed for 2 days against at least three changes of 4 liters of 0.3 M sodium acetate–10 mM Tris·HCl (pH 7.5). The DNA was precipitated with ethanol and resuspended in a minimal volume of water. The labeled DNA was completely digested for nearest-neighbor analysis [40 units of micrococcal nuclease in 10 mM sodium borate buffer (pH 8.6)–20 mM $CaCl_2$ for 30 min at 37°, then with 0.2 units of spleen phosphodiesterase, 50 mM Tris·HCl (pH 7.5), 10 mM $MgCl_2$, 5 mM rAMP].

Electrophoresis. Oligonucleotides were separated by two-dimensional electrophoresis on DEAE-paper (Whatman DE-81); the first dimension was pyridine-acetate (pH 3.5) and the second dimension was formate-acetate (pH 1.9) (17, 18). At times, 7% formic acid was used for the second dimension. Mononucleotides were separated by electrophoresis on Whatman 3MM paper in pyridine–acetate (pH 3.5). Appropriate absorbance markers were added where necessary, the pyridine was removed from the paper in an NH_3 atmosphere, and the marker was visualized under a UV lamp. All the other microtechniques for nucleotide sequence analysis are essentially the same as described elsewhere (17, 18).

Radioautography and Radioactive Counting. Radioactive spots on electropherograms were located by placing three sheets of x-ray film (Kodak RPR 14) over the electrophoresis paper. Spots containing ^{32}P and ^{33}P, or spots containing ^{33}P alone, were located by exposure of the first film adjacent to the paper. Spots containing ^{32}P could be distinguished from spots containing only ^{33}P by exposures on the third film, since a given amount of ^{33}P was only 1–2% as effective at exposing the third film as was an identical amount of ^{32}P. Radioactivity was quantitated by counting in a Nuclear-Chicago gas-flow counter, with or without appropriate screens, or by differential counting in a liquid scintillation counter.

RESULTS

The 5′-terminal mononucleotide at the RI endonuclease break

Preliminary experiments established that λ DNA treated with endonuclease RI and alkaline phosphatase before phosphorylation with polynucleotide kinase incorporated six to twelve more molecules of ^{32}P per λ genome than did λ DNA treated only with phosphatase. DNA treated only with endonuclease RI accepted about the same amount of ^{32}P as did untreated DNA. We take these results to indicate that the RI endonuclease generates 5′ phosphoryls and three to six double-strand breaks per λ genome. Two other independent estimates indicate about five sites per λ genome, the number we have used for subsequent calculations (R. Davis, personal communication; Yoshimori, Roulland-Dussoix, Goodman, and Boyer, to be published). To determine the mononucleotide at the 5′ terminus, an aliquot of DNA that had been treated with endonuclease RI and alkaline phosphatase was phosphorylated at the 5′-terminus with ^{32}P and digested to 5′-mononucleotides by successive treatment with pancreatic DNase and venom phosphodiesterase. Analysis of the mononucleotides after separation by paper electrophoresis demonstrated that the amount of ^{32}P occurring in pA increased 6- to 7-fold when the DNA had been treated with endonuclease RI before dephosphorylation (see Table 1). Only pA showed a significant increase in ^{32}P, and the increase did not occur in DNA treated only with endonuclease RI. This result indicates that the 5′-mononucleotide at each strand of an endonuclease RI break is pA.

Although the increase (7–8 molecules per genome) in total ^{32}P incorporated into endonuclease RI–phosphatase-treated DNA (compared to DNA treated with phosphatase only) appears to be accounted for by the increase in pA, we cannot say with certainty that this accounts for all of the RI-created termini. We were unable to convincingly demonstrate incorporation of ^{32}P above background into the normal λ termini. These are probably obscured by the presence of two to three presumably internal random breaks per λ molecule, which evidently occurred during purification and/or storage (see Table 1, Exp. II). Therefore, we cannot exclude the possibility that one in ten breaks had a different 5′-terminus. However, these data, along with the RSV DNA polymerase

TABLE 1. *The 5′-terminal nucleotide at the RI endonuclease break**

Exp. no.	Treatment before phosphorylation		^{32}P atoms incorporated/λ genome	[^{32}P]cpm			
				$_pC$	$_pA$	$_pG$	$_pT$
I	None		—	170	470	524	213
	Endonuclease RI		—	160	1056	415	209
	Endonuclease RI + phosphatase		—	231	7854	690	314
II	Phosphatase	(*a*)	3.6	402	780	518	630
		(*b*)	4.8	290	625	427	416
	Endonuclease RI + phosphatase	(*a*)	12.4	455	4165	657	474
		(*b*)	12.2	389	4016	509	528

* See note added in proof.

TABLE 2. *The 5'-terminal sequence at the RI endonuclease break*

Chain length	% of input cpm	Base composition of ^{32}P-containing oligonucleotide	Value of M*	Terminal nucleotide from M	Deduced sequence
Mono-	6.6	A_1	—	—	$_pA$
Di-	15.2	A_2	—	—	$_pA_pA$
Tri-	19.7	A_2, T	1.2	T	$_pA_pA_pT$
Tetra-	24.7	A_2, T_2	2	T	$_pA_pA_pT_pT$
Penta-	14.3	A_2, T_2, C	0.03	C	$_pA_pA_pT_pT_pC$
Hexa-	8.9	A_2, T_3, C	2.3	T	$_pA_pA_pT_pT_pC_pT$
		A_3, T_2, C	0.3	A	$_pA_pA_pT_pT_pC_pA$
Hepta-	5.1	—	—	—	—

* The value of M is characteristic of the nucleotide removed (see text), and the range of M- values for each nucleotide are C (0.08–0.1), A (0.3–0.5), G (1.4–3), and T (1.4–3).

experiments discussed below, make it highly likely that we have accounted for nearly all (nine out of ten) or probably all of the 5'-termini at the RI endonuclease breaks in λ DNA.

The 5'-terminal sequence

RI endonuclease-treated λ DNA labeled at the 5'-termini with ^{32}P (10) was added to λ DNA uniformly labeled with ^{33}P, and the mixture was digested exhaustively with pancreatic DNase and briefly with venom phosphodiesterase to generate 5'-phosphoryl-terminated oligonucleotides of average chain length three to five (Table 2 and ref. 15). After separation according to chain length (16), the oligonucleotide mixtures of identical chain length were separated according to base composition and sequence by two-dimensional high-voltage electrophoresis (17, 18). Spots containing ^{32}P and/or ^{33}P were located by radioautography, and both ^{32}P and ^{33}P were quantitated. After elution from the paper, digestion with venom phosphodiesterase, and separation of the 5'-mononucleotides, the base composition of each oligonucleotide spot was established by determination of the relative amount of ^{33}P in each mononucleotide (^{32}P was only in pA). The sequences of the di- through hexanucleotide ^{32}P-containing oligonucleotides deduced from these data are shown in Table 2. Two other means were also used to confirm these data: (*i*) The position of an oligonucleotide in these two-dimensional electropherograms is extremely characteristic, and the positions of each of the oligonucleotides deduced here was in agreement with previous maps (17, 18). (*ii*) The isolated ^{32}P-containing tetra- to heptanucleotides and, separately, a partial venom phosphodiesterase digest of the pentanucleotide were re-run on DEAE-paper at pH 1.9. The ratio of the distance between an oligonucleotide and its first degradation product (X) to the distance of the oligonucleotide from the origin (Y) gives a value ($M = X/Y$) that is characteristic of the 3'-terminal nucleotide that has been removed (17, 18). The M values listed in Table 2 agree with the deduced sequences.

A single ^{32}P-labeled spot predominated in each size class from the di- through the pentanucleotide. In each case, this spot contained at least 60% of the recovered ^{32}P, no other spot contained more than 10%. Since most of the ^{32}P occurred in one spot, the same pentanucleotide sequence probably occurs at each 5'-terminus generated by the endonuclease.

Two spots labeled with ^{32}P, each containing only A, C, and T, were detected in the hexanucleotide fingerprint. The amount of ^{33}P recovered in these spots was insufficient to accurately determine the relative proportion of each of the nucleotides, but the hexanucleotide base sequences were determined from their electrophoretic mobilities and M values on DEAE-paper at pH 1.9 (18). These data indicate that only A or T was found at the sixth position internal from the 5'-terminus. The heptanucleotide fraction gave at least three ^{32}P-containing spots after two-dimensional electrophoresis, two of which contained G in addition to the other nucleotides. Although eight oligonucleotide spots are to be expected if the seventh position is completely degenerate, we probably would not have been able to detect them all because as the chain length (n) of an oligonucleotide increases, it becomes more difficult to separate the various isomers due to their increase in number 4^n and the decrease in relative electrophoretic mobility differences. We estimate that a maximum of five heptanucleotides could be resolved under our conditions.

RSV DNA polymerase experiments

The 5'-terminal sequence described above can be arranged in four ways (Fig. 1, *a–d*), depending on the topology of the phosphodiester-bond cleavages. The finding of Mertz and Davis (19) that the RI endonuclease makes a short cohesive end suggested that if the break consisted of a 3'-hydroxyl end and a protruding 5'-single-strand end (possibilities 1*a* or 1*b*), it would serve as a primer–template for a DNA polymerase. The recent characterization of the DNA polymerases from RNA tumor viruses (e.g., Rous Sarcoma Virus) showed that they catalyze repair-like reactions on such templates, and are free of the two exonuclease activities associated with *E. coli* DNA polymerase I (8, 9).

RSV DNA polymerase (8) was used to study the incorporation of various combinations of [α-^{32}P]- and nonradioactive-nucleoside triphosphates into λ DNA cleaved with RI endonuclease; the products were analyzed by nearest-neighbor transfer experiments (Table 3). Increasing the time of incubation or the amount of enzyme in the standard reaction did not appreciably change the incorporation nor the nearest-neighbor data. Since significant incorporation was detected for RI-treated λ DNA above untreated λ DNA controls (Expts. 1 and 3–6), the two arrangements of the RI ends shown in Fig. 1*c* and *d* are not possible because they would not serve as primer–templates for RSV DNA polymerase.

It is possible to detect incorporation into the RI cohesive ends, even though there are two λ cohesive ends (twelve nucleotides long) and ten RI cohesive ends (four nucleotides long) per λ DNA molecule. The reason is evident from inspection of the nucleotide sequences of the λ cohesive ends (20), which indicate that in the absence of GTP and CTP only one A residue can be incorporated per λ genome. Therefore, for Expts. 1–6 the expected pmol of [^{32}P]NTP incorporated per pmol of λ DNA from the λ cohesive ends would be 1, 0, 1, 0, 1, and 4, respectively, compared to 20, 0, 20, 20, 40, and 40 for the RI cohesive ends. This prediction is consistent with the data in Table 3 where the shorter RI cohesive ends, which contain only A and T but which occur more frequently, accounted for 85–97% of the incorporation in the absence of CTP and GTP.

The incorporation and nearest-neighbor data of Table 3 (Expts. 1–5) confirm the facts that the RI restriction endonuclease makes a 3'-hydroxyl 5'-phosphoryl break, forms

TABLE 3. *Nearest-neighbor analysis of nucleotides incorporated into RI endonuclease-treated λ DNA by RSV DNA polymerase*

Labeled substrate	Unlabeled substrate	pmol of [^{32}P]NTP incorporated per pmol λ DNA: RI-treated λ DNA	pmol of [^{32}P]NTP incorporated per pmol λ DNA: λ DNA	pmol of [^{32}P]-NTP incorporated per pmol RI End†: Exp.	Theory	Products of micrococcal nuclease & spleen phosphodiesterase digestion‡: C_p^*	A_p^*	G_p^*	T_p^*	Deduced sequence
1. $_{ppp}^{*}A$	—	14.7	2.6	1.2	2	<0.02	1.2	1.0	<0.02	G↓$_{p}^{*}A_p^*A$
2. $_{ppp}^{*}T$	—	1.4	0.81	0.06	0	—	—	—	—	—
3. $_{ppp}^{*}A$	$_{ppp}T$	25.2	0.69	2.4	2	<0.02	1.3	1.0	<0.04	G↓$_{p}^{*}A_p^*A$
4. $_{ppp}^{*}T$	$_{ppp}A$	15.6	1.3	1.4	2	<0.02	1.1§	<0.02	1.0	G↓$_{p}A_p^*A_p^*T_p^*T$
5. $_{ppp}^{*}A$ $_{ppp}^{*}T$	— —	31.0	1.4	3.0	4	<0.02	1.9	1.0	1.0	G↓$_{p}^{*}A_p^*A_p^*T_p^*T$
6. $_{ppp}^{*}A$ $_{ppp}^{*}T$	$_{ppp}C$ $_{ppp}G$	68.7	9.1	5.9	4	<0.02	1.6	1.0	1.1	G↓$_{p}^{*}A_p^*A_p^*T_p^*T$
7. $_{ppp}^{*}C$ $_{ppp}^{*}G$	$_{ppp}A$ $_{ppp}T$	8.9	10.5	<0	0	—	—	—	—	—
				Conclusion:		G↓$_{p}A_pA_pT_pT$				

RI endonuclease-treated and untreated λ DNA were used as primer-templates for the incorporation of [α-^{32}P]-and/or nonradioactive-nucleoside triphosphates with RSV DNA polymerase. The *asterisks* show the position of the ^{32}P atom. The *arrow* indicates the position of the RI endonuclease break.

† The ratio of pmoles was calculated on the assumption of 10 RI ends per λ genome after subtracting the incorporation into untreated λ DNA. The value listed under '*Theory*' is deduced from the final sequence.

‡ The nearest-neighbor analyses are listed for the RI-treated λ DNA.

§ In Exp. 4, the ratio of A_p^* to T_p^* varied between 1.1 and 1.9 in different experiments.

a cohesive end, and that the sequence of the first four nucleotides at the 5′ end are $_pA_pA_pT_pT$, as already determined from the polynucleotide kinase experiments described above. Furthermore, the nucleotide on the 3′ side of the break is unique since [α-^{32}P]ATP transferred label only to G_p (other than to the A_p known to be incorporated into the cohesive end).

These experiments do not distinguish between the two possible configurations depicted in Fig. 1*a* and *b*. However, in Expts. 6 and 7 (Table 3), the incorporation data are more compatible with those predicted from configuration (*a*), and the nearest-neighbor analysis of Exp. 6 provides additional support for this configuration. In addition, the T_m (5–6°) determined for the RI cohesive termini is compatible with this arrangement (19). Therefore, the combined data prove that the topology of the breaks is as shown in Fig. 1*a*.

The 3′-terminal dinucleotide sequence

The dinucleotide at the 3′-hydroxyl side of the break was analyzed by labeling RI endonuclease-treated λ DNA with [α-^{32}P]ATP and unlabeled TTP by the use of RSV DNA polymerase. The 3′-terminally labeled DNA was exhaustively digested with micrococcal nuclease, and the ^{32}P-labeled dinucleotides were isolated by column chromatography or paper electrophoresis. The dinucleotides $N_pG_p^*$, $G_p^*A_p^*$, and $A_p^*A_p$ are to be expected from the previous labeling data and the known sequence (. . .$N_pG_p^*A_p^*A_pT_pT$). We detected $A_p^*A_p$, which occurs in relatively high proportion (8.5%) in a limit micrococcal nuclease digest, but did not detect $G_p^*A_p^*$ because it occurs in extremely low amount (0.09%). There is a large difference in occurrence, especially between isomeric pairs, of each dinucleotide in an exhaustive micrococcal nuclease digest due to the specificity of the nuclease (our unpublished results). Of the four possible dinucleotides at the 3′ end of the RI endonuclease break, which all occur in relatively high frequency in the digest A_pG_p (21%), T_pG_p (19%), C_pG_p (9%), and G_pG_p (2%), only $A_pG_p^*$ and $T_pG_p^*$ were detected in reasonable amounts. The ^{32}P counts in $C_pG_p^*$ and $G_pG_p^*$ together accounted for less than 5% of the counts in $A_pG_p^*$ plus $T_pG_p^*$. The amounts of ^{32}P in $A_pG_p^*$ and $T_pG_p^*$ were almost equal. If the known background incorporation into sites other than the RI cohesive ends is corrected for the occurrence of each

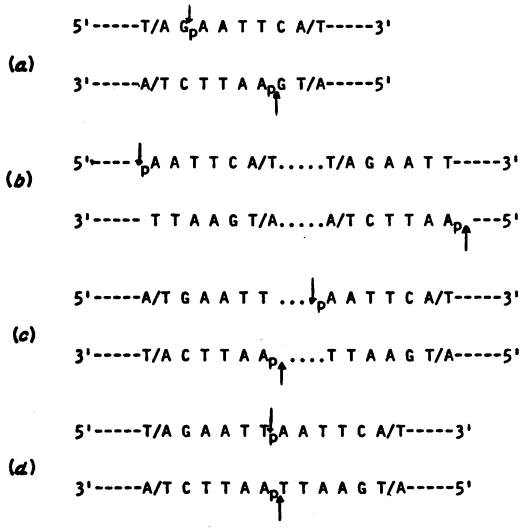

FIG. 1. The four possible arrangements of the RI endonuclease breaks are shown. The known sequence from the polynucleotide kinase labeling is indicated by the *symbols*, the possible unknown by *dots*, and the remainder of the double-stranded DNA structure by two parallel *dashed lines*.

dinucleotide in the micrococcal nuclease digest and subtracted from the ^{32}P found in $C_pG_p^*$ and $G_pG_p^*$, the difference is zero. Therefore, of the ten RI endonuclease single-strand breaks in λ DNA, less than 0.5 have a C_pG or G_pG at the 3′ side of the break. The corrected data reduces this number to zero, so that there are five breaks with T_pG and five breaks with A_pG at the 3′-hydroxyl side of the RI endonuclease breaks in λ DNA. The sequence is therefore only partially degenerate (A or T) at this position, in agreement with the data from the 5′-terminally labeled hexanucleotides and the topology of the breaks (Fig. 1*a*).

DISCUSSION

The 5′-terminal hexanucleotide sequences ($_p$AATTCA and $_p$AATTCT) at the RI endonuclease breaks in λ DNA were determined by labeling the ends with ^{32}P by the use of polynucleotide kinase, followed by digestion and separation of the oligonucleotides. The 5′-terminal tetranucleotide ($_p$AATT) and 3′-hydroxyl terminal dinucleotides (. . .$_pA_pG$ and . . .$_pT_pG$) were determined from nearest-neighbor analysis of DNA labeled by RSV DNA polymerase-directed incorporation of [α-^{32}P]nucleoside triphosphates into the short cohesive ends generated in λ DNA by RI endonuclease cleavage. Considering the anti-parallel complementarity of double-helical DNA, these two independent methods give the same sequence and provide evidence for the nucleotide sequence adjacent to the phosphodiester bonds in λ DNA cleaved by the RI endonuclease (Fig. 1*a* and *Abstract*). The sequence is unique for six base pairs, and the outside bases can be either A or T.

Although we could not determine with absolute certainty whether one, to a maximum of two, out of ten RI endonuclease breaks in λ DNA had a different sequence by either method above, the results obtained by these two methods together make it extremely likely that the sequences around all ten breaks in λ DNA are identical. Also, the 5′-terminal hexanucleotide data and 3′-hydroxyl dinucleotide data are independent determinations of whether C and/or G can occur adjacent to the six unique base pairs at each site. As both estimates indicate less than one out of ten (and probably zero) breaks has a C or G at this position, it appears that in λ DNA all five sites cleaved by RI endonuclease are degenerate for only A and T at the outside bases.

We conclude that in λ DNA G or C does not occur on either side of the symmetrical hexanucleotide. The sample size (ten) was large enough to have detected G or C at the outside positions if they occur randomly in λ DNA and with the expected frequencies, i.e., 25% (confidence level >95% by chi-square test). This conclusion can be interpreted in one of two ways: either the specific sequence cleaved by the endonuclease is a symmetrical hexanucleotide, but in λ DNA this hexanucleotide is not flanked by G or C, or the specific sequence cleaved is eight nucleotides long and the enzyme cannot discriminate between A and T, but can distinguish either from G or C at the outside positions. Although additional experiments are required to decide between the two interpretations, we favor the latter.

The RI sequence exhibits 180-degree rotational symmetry for either of the possibilities just discussed, a result also found in the sequence of six base pairs restricted by the *Hemophilus influenzae* restriction endonuclease (15). However, the ambiguity of the RI sequence is different in two respects from that of the *H. influenzae* sequence. (*i*) It occurs at the outside of the sequence rather than centrally, and (*ii*) It is a purine–pyrimidine (A to T) rather than a purine–purine (A to G) degeneracy. It has been suggested that the substrate symmetry in these cases is related to the subunit structure of the enzymes that interact with the sequence (2, 15). It is of interest that the purified RI endonuclease is composed of two subunits of identical molecular weight (37,000), presumably coded for by one cistron (Yoshimori, Roulland-Dussoix, Aldridge, and Boyer, to be published). The other important feature of the RI endonuclease break is that the staggered cleavages give rise to short cohesive termini, which afford the possibility of reconstructing DNA molecules *in vitro* from any two DNA fragments generated by RI endonuclease digestion (19).

NOTE ADDED IN PROOF

It should be noted that for the concentrations of λ DNA used in the 5′-terminal labeling experiments (about 500 μg/ml), although the DNA solutions were heated to 75° and quickly cooled on ice before alkaline phosphatase treatment, the very rapid reannealing of λ cohesive ends at 37° ($t_{1/2}$ is a few minutes) [Wang, J. C. & Davidson, N.(1966) *J. Mol. Biol.* **19**, 469–482] would have rendered the 5′-terminal phosphates resistant to alkaline phosphatase at 37°, and therefore insusceptible to labeling with polynucleotide kinase. This explains in part our inability to recover ^{32}P in those nucleotide sequences corresponding to λ 5′-termini.

We thank Ms. Judy Aldrich for her invaluable technical assistance. This investigation was supported by Public Health Service Grants GM14378 and CA14026, and by American Cancer Society Grant P-591 and A.C.S. California Division Grant 599. J. H. is a fellow of the Giannini Foundation.

1. Arber, W. & Linn, S. (1969) *Annu. Rev. Biochem.* **38**, 467–500.
2. Boyer, H. W. (1971) *Annu. Rev. Microbiol.* **25**, 153–176.
3. Meselson, M. & Yuan, R. (1968) *Nature* **217**, 1110–1114.
4. Linn, S. & Arber, W. (1968) *Proc. Nat. Acad. Sci. USA* **59**, 1300–1306.
5. Roulland-Dussoix, D. & Boyer, H. W. (1969) *Biochim. Biophys. Acta* **195**, 219–229.
6. Smith, H. O. & Wilcox, K. W. (1970) *J. Mol. Biol.* **51**, 379–391.
7. Kühnlein, U. & Arber, W. (1972) *J. Mol. Biol.* **63**, 9–19.
8. Faras, A. J., Taylor, J. M., McDonnell, J. P., Levinson, W. E. & Bishop, J. M. (1972) *Biochemistry* **11**, 2334–2342.
9. Hurwitz, J. & Leis, J. P. (1972) *J. Virol.* **9**, 116–129.
10. Richardson, C. C. (1965) *Proc. Nat. Acad. Sci. USA* **54**, 158–165.
11. Weiss, B., Live, T. R. & Richardson, C. C. (1968) *J. Biol. Chem.* **243**, 4530–4542.
12. Padmanabhan, R. & Wu, R. (1972) *J. Mol. Biol.* **65**, 447–467.
13. Glynn, I. M. & Chappell, J. B. (1964) *Biochem. J.* **90**, 147–149.
14. Wehrli, W. E., Verbeyden, D. & Moffatt, J. G. (1965) *J. Amer. Chem. Soc.* **87**, 2265–2277.
15. Kelly, T. J., Jr. & Smith, H. O. (1970) *J. Mol. Biol.* **51**, 393–409.
16. Junowicz, E. & Spencer, J. H. (1970) *Biochemistry* **9**, 3640–3648.
17. Sanger, F., Brownlee, G. G. & Barrell, B. G. (1965) *J. Mol. Biol.* **13**, 373–398.
18. Murray, K. (1970) *Biochem. J.* **118**, 831–841.
19. Mertz, J. E. & Davis, R. W. (1972) *Proc. Nat. Acad. Sci. USA* **69**, 3370–3374.
20. Wu, R. & Taylor, E. (1971) *J. Mol. Biol.* **57**, 491–511.