J. Mol. Biol. (1970) **51**, 393–409

A Restriction Enzyme from *Hemophilus influenzae*

II. Base Sequence of the Recognition Site

THOMAS J. KELLY, JR.† AND HAMILTON O. SMITH

Department of Microbiology
Johns Hopkins University, School of Medicine
Baltimore, Md. 21205, U.S.A.

(*Received 18 February 1970*)

Hemophilus influenzae strain Rd contains an enzyme, endonuclease R, which specifically degrades foreign DNA. With phage T7 DNA as substrate the endonuclease introduces a limited number (about 40) double-strand breaks (5′-phosphoryl, 3′-hydroxyl). The limit product has an average length of about 1000 nucleotide pairs and contains no single-strand breaks. We have explored the nucleotide sequences at the 5′-ends of the limit product by labeling the 5′- phosphoryl groups (using polynucleotide kinase) and characterizing the labeled fragments released by various nucleases. Two classes of 5′-terminal sequences were obtained: pApApCpNp . . . (60%) and pGpApCpNp . . . (40%), where N indicates that the base in the 4th position is not unique. The dinucleoside monophosphates at the 3′-ends were isolated after micrococcal nuclease digestion of the limit product and identified as TpT(60%) and TpC (40%). We conclude that endonuclease R of *H. influenzae* recognizes the following specific nucleotide sequence:

5′ . . . pGpTpPy | pPupApCp . . . 3′
3′ . . . pCpApPup | PypTpGp . . . 5′

The implications of the twofold rotational symmetry of this sequence are discussed.

1. Introduction

The accompanying paper (Smith & Wilcox, 1970) describes the purification and properties of endonuclease R of *Hemophilus influenzae*, an enzyme which appears to be capable of recognizing and degrading foreign DNA molecules. The purified enzyme has no detectable nucleolytic activity against either duplex or single-stranded *H. influenzae* DNA molecules but produces a limited number of duplex cleavages (5′-phosphoryl, 3′-hydroxyl) in a variety of native foreign DNA molecules. The limit product of endonuclease R digestion has an average chain length of the order of 1000 nucleotide pairs and contains no single-strand breaks. These properties are qualitatively similar to those of the *Escherichia coli* restriction enzymes studied by Meselson & Yuan (1968) and Linn & Arber (1968).

The fact that endonuclease R hydrolyzes only about 0·1% of the phosphodiester linkages potentially available to it strongly suggests that the enzyme "recognizes" a relatively small number of specific sites within a foreign DNA molecule. The experiments described in this paper were motivated by the assumption that the specificity of this recognition process is determined by the local nucleotide sequence at these

† Present address: National Institute of Allergy and Infectious Diseases, National Institutes of Health, Bethesda, Md. 20014, U.S.A.

sites. Using end-labeling techniques, we have explored the sequences at the ends of the fragments released by digestion with endonuclease R and succeeded in identifying the complete base sequence of its recognition site.

2. Materials and Methods

(a) *Enzymes*

Electrophoretically pure pancreatic DNase (2000 units/mg) was obtained from Sigma Chemical Co. Bacterial alkaline phosphatase (20 units/mg), snake venom phosphodiesterase (potency 0·3), micrococcal nuclease (1000 units/mg) and spleen phosphodiesterase were obtained from Worthington Biochemical Corp.

Polynucleotide kinase (fraction VI of Richardson (1965), (5000 units/ml.) and rechromatographed bacterial alkaline phosphatase, 20 units/ml.(Weiss, Live & Richardson, 1968). were supplied by Dr Bernard Weiss.

Exonuclease I from *E. coli* was received from Dr Paul Englund as the ammonium sulfate I fraction of a DNA polymerase purification and was reprecipitated with ammonium sulfate and chromatographed on DEAE-cellulose as described by Lehman & Nussbaum (1964). The final specific activity was 20,200 units/mg.

The purification of endonuclease R of *H. influenzae* has been described (Smith & Wilcox 1970). The final 200-fold purified preparation contained 16 units/ml.

(b) *Preparation of ^{33}P-labeled phage T7 DNA*

Phage T7 (from B. Weiss) was grown on *E. coli* B in a synthetic medium containing 1·5 μg phosphorus/ml. (Smith, 1968) and 2·5 μc ^{33}P/ml. (New England Nuclear Corp.; supplied as ^{33}P-labeled phosphoric acid containing 0·02 N-HCl). Following lysis the phage were purified by differential centrifugation and by banding twice in a preformed CsCl gradient (Thomas & Abelson, 1966). Phenol extraction was carried out as described by Kelly & Thomas (1969). After phenol extraction the DNA solution was exhaustively dialyzed against 0·05 M-NaCl, 0·01 M-Tris, pH 7·4. The preparation of ^{32}P-labeled T7 DNA was carried out by the same methods used for ^{33}P-labeled T7 DNA except that 5μc of carrier-free [^{32}P]orthophosphate (New England Nuclear Corp.) /ml. was substituted for the [^{33}P]phosphoric acid in the growth medium.

(c) *Fractionation of nucleotides and oligonucleotides*

Electrophoresis of nucleotides and oligonucleotides was performed on cellulose thin-layer sheets (Eastman Kodak Co.) at a potential gradient of 25 v/cm in a Warner–Chilcott model E-800-2B migration chamber cooled to 5°C. For routine separations 0·075 M-ammonium formate buffer, pH 3·55, was used. When a second round of electrophoretic fractionation was required, 8·7% acetic acid, 2·5% formic acid (v/v), pH 1·9, (Sanger, Brownlee & Barrell, 1965) was used.

Two-dimensional chromatographic separation of nucleotides on 6 cm × 6 cm PEI-cellulose thin-layers (Brinkman Instruments, Inc., Cel MN 300 PEI) was carried out by methods similar to those described by Randerath & Randerath (1967). The chromatograms were developed in the first dimension with 0·5 M-LiCl and in the second dimension with 0·5 M-acetic acid. Between dimensions the chromatograms were air dried and then desalted by immersion in anhydrous methanol for 15 min. Following chromatography the nucleotide spots were located with an ultraviolet lamp and cut from the thin layer for assay of radioactivity in a scintillation counter. With this procedure a nucleotide analysis can be performed in about 1 hr.

Fractionation of oligonucleotides according to net charge was carried out by the method of Tomlinson & Tener (1963). The sample containing 0·02 M-Tris, pH 7·6, and 7 M-urea was loaded onto a 1 cm × 10 cm column of DEAE-cellulose (Serva; in bicarbonate form) which had been equilibrated with the same solvent. The column was then developed with a linear gradient of NaCl (0 to 0·3 M) in 0·02 M-Tris, pH 7·6, 7 M-urea. The total gradient volume was generally 1000 ml. and the flow rate was 50 ml./hr. The oligonucleotides were subsequently freed from salt and urea by adsorption to DEAE-cellulose (Rushizsky & Sober, 1962) or partially inactivated Norit (Threlfall, 1957; Josse & Moffatt, 1965).

(d) *Autoradiography*

The distribution of radioactivity on a thin-layer sheet following fractionation of a collection of oligonucleotides by electrophoresis was generally determined by autoradiography. In a number of the experiments to be described in this paper the oligonucleotides were uniformly labeled with ^{33}P. However, those oligonucleotide species which originated from the 5′-termini of an endonuclease R limit product also contained ^{32}P (see section (e) below). In order to monitor the distribution of both labels the usual autoradiographic procedure was modified as follows: a sheet of X-ray film (Eastman Kodak Co., NS-54T) was placed in direct contact with the thin-layer matrix (unscreened autoradiograph). Both ^{33}P and ^{32}P disintegrations contributed to grain exposure on this film. A second sheet of X-ray film (screened autoradiograph) was separated from the thin-layer matrix by a thin screen. (The Mylar backing of commercial thin layers was found to be suitable.) Because of absorption of the lower energy ^{33}P disintegrations by the screen, grain exposure on this film was primarily due to ^{32}P disintegrations. Using this method it was possible to tell at a glance which of a number of oligonucleotide species contained ^{32}P and therefore originated from the 5′-ends of an endonuclease R limit product. Densitometer tracings of the autoradiographs were made with a Joyce–Loebl recording microdensitometer.

(e) *^{32}P-labeling of the 5′-end of uniformly ^{33}P-labeled endonuclease R limit product*

A preparation of uniformly ^{33}P-labeled phage T7 DNA was digested to a limit product with endonuclease R (Smith & Wilcox, 1970). The reaction mixture (1·2 ml.) contained 500 mμmoles of ^{33}P-labeled T7 DNA (spec. act. of the order of 10^4 cts/min/mμmole), 50 mM-NaCl, 40 mM-Tris, pH 7·4, 7 mM-$MgCl_2$, 10 mM-mercaptoethanol and 0·6 unit of endonuclease R. After incubation for 15 min at 37°C an additional 0·6 unit of endonuclease R was added to the reaction mixture and the incubation was continued for another 15 min at 37°C. The resulting digest was then dialyzed twice against 1 liter of 0·05 M-NaCl, 0·01 M-Tris, pH 8·0.

The dialyzed preparation of endonuclease R limit product (1·0 ml.) was incubated for 30 min at 37°C with 2 units of rechromatographed alkaline phosphatase. One additional unit of phosphatase was then added and incubation was continued for 30 min at 37°C. The resulting digest was adjusted to 0·05 M in EDTA, extracted twice with redistilled phenol (saturated with 0·1 M-Tris, pH 7·4), and exhaustively dialyzed against 0·05 M-NaCl, 0·01 M-Tris, pH 7·4.

The resulting preparation of dephosphorylated endonuclease R limit product was rephosphorylated using [γ-^{32}P]ATP in the polynucleotide kinase reaction as described by Weiss *et al.* (1968). The reaction mixture (1·1 ml.) contained 500 mμmoles of dephosphorylated ^{33}P-labeled limit product, 50 mM-NaCl, 50 mM-Tris (pH 7·4), 10 mM-$MgCl_2$, 15 mM-mercaptoethanol, 1·5 mM-potassium phosphate buffer (pH 7·4) 15 μM-[γ-^{32}P]ATP, and 150 units of polynucleotide kinase. The reaction mixture was incubated for 1 hr at 37°C with the addition of another 75 units of kinase at 30 min. Precipitation of the 5′-^{32}P-end labeled limit product was accomplished by addition of 1 vol. of 0·1 M-sodium pyrophosphate and 1·33 vol. of 12·5% trichloroacetic acid. The precipitate was collected on a glass filter (Whatman GF/C) and washed extensively with 6% trichloroacetic acid, 0·1 M-sodium pyrophosphate followed by 95% ethanol. The DNA was recovered by macerating the filter in 1 ml. of 0·2 N-NH_4OH and then filtering the resulting suspension through a semi-micro, medium-sintered glass funnel. After washing the funnel twice with 1 ml. of 0·2 N-NH_4OH the filtrate containing the dissolved DNA was evaporated to the desired volume under a warm air jet.

The [γ-^{32}P]ATP used in this labeling procedure was provided by B. Weiss and by P. Englund. The specific activities used were in the range of 10^5 to 10^6 cts/min/mμmole.

(f) *Preparation of 5′-terminal oligonucleotides from an endonuclease R limit digest of phage T7 DNA*

(i) *Mononucleotides*

Digestion of 5′-^{32}P-terminally labeled endonuclease R limit product to mononucleotides was carried out by consecutive hydrolysis with pancreatic DNase and snake venom phosphodiesterase. A reaction mixture (0·1 ml.) containing 150 μM-DNA (prepared as

described in section (e)), 10 mM-Tris (pH 7·4), 5 mM-$MgCl_2$ and 20 μg pancreatic DNase/ml. was incubated at 37°C for 30 min to yield oligonucleotides. The pancreatic DNase was then inactivated by heating to 95°C for 5 min. A 10-μl. sample of the digest was digested to mononucleotides by adding 1 μl. of 1 M-glycine buffer, pH 9·2, and 1 μl. of snake venom phosphodiesterase (4 mg/ml.) and incubating at 37°C for 30 min.

(ii) *Dinucleotides*

A 10-μl. sample of the above pancreatic DNase digest was adjusted to 0·1 M in glycine buffer, pH 9·2. Exonuclease I was added to a final concentration of 4×10^3 units/ml. and the reaction mixture incubated at 37°C for 30 min. This procedure is similar to that described by Weiss & Richardson (1967).

(iii) *Trinucleotides*

A portion of 5′-^{32}P-terminally labeled endonuclease R limit product was digested to an average chain length of 5 residues with pancreatic DNase. The reaction mixture (1 ml.) contained 300 μM-DNA, 10 mM-Tris (pH 7·4), 5 mM-$MgCl_2$, 50 mM-NaCl and 20 μg pancreatic DNase/ml. After incubation at 37°C for 45 min the reaction mixture was heated to 95°C for 5 min. An average of 1 to 2 nucleotides was then removed from the 3′-ends of the oligonucleotides in this digest by partial digestion with snake venom phosphodiesterase (Holley, Madison & Zamir, 1964). The pancreatic DNase digest was equilibrated at 20°C and 0·25 ml. of snake venom phosphodiesterase (40 μg/ml. in 0·06 M-$MgCl_2$) was added. Incubation at 20°C was continued for 15 min and then the digest was quickly frozen in a salt–ice bath. The average chain length of the final product was 2·1 residues. Preparatory to fractionation by DEAE-cellulose chromatography, 0·84 g of urea was added to the frozen oligonucleotide mixture, which was then allowed to thaw at room temperature. After addition of 0·04 ml. of 1 M-Tris, pH 7·6, the volume of the solution was adjusted to 2·0 ml. Fractionation according to chain length was carried out as described in section (c) above. The trinucleotide fraction was desalted and concentrated to 20 μl. for electrophoretic analysis.

(iv) *Tetranucleotides*

A solution containing 150 mμmoles of 5′-^{32}P-terminally labeled endonuclease R limit product was digested with pancreatic DNase under the conditions described in section (i) above. The tetranucleotides were isolated from the resulting digest by methods similar to those described for the isolation of trinucleotides.

(g) *Determination of oligonucleotide chain length*

The number-average chain length of a given collection of ^{33}P- or ^{32}P-labeled oligonucleotides was determined by measuring the ratio of total radioactivity to radioactivity released as P_i by alkaline phosphatase treatment. The reaction mixture for phosphatase treatment (0·04 ml.) contained less than 1 mμmole of oligonucleotide, 0·125 M-Tris (pH 8·0) and 3 units of alkaline phosphatase/ml. Incubation was carried out at 37°C for 30 min. A sample of the reaction mixture was removed for determination of total radioactivity (T). A second sample was applied 1 cm from the bottom edge of a 1 cm × 5 cm strip of PEI-cellulose thin layer which had been banded with a Norit suspension (20% packed volume) in a 1-cm width, 2 cm from the bottom edge. The chromatographic strip was dried and then developed with 1 N-HCl for 15 min. In this procedure acid-insoluble DNA is bound at the origin, soluble oligonucleotides are quantitatively absorbed by the Norit band, and only inorganic phosphate migrates to the upper 1·5 cm of the strip. This upper 1·5 cm segment was dried, cut from the strip, and assayed for radioactivity (P). The average chain length is given by T/P.

(h) *Isolation of 3′-terminal dinucleotides of endonuclease R limit product*

Uniformly ^{32}P-labeled phage T7 DNA was digested to a limit product with endonuclease R in a reaction mixture (1·7 ml.) containing 375 mμmoles of DNA, 6 mM-$MgCl_2$, 6 mM-mercaptoethanol, 50 mM-NaCl, 10 mM-Tris, pH 7·4, and 0·4 unit of endonuclease R. After incubation at 37°C for 20 min an additional 0·4 unit of endonuclease R was added and the incubation was continued for another 10 min. The limit digest was extracted once with

phenol (saturated with 0·1 M-Tris, pH 7·4) and then exhaustively extracted with ether to remove the phenol. The endonuclease R limit product was then digested to completion with micrococcal nuclease. The reaction mixture was constructed by addition of 0·16 ml. of dGMP (20 mM), 0·05 ml. $CaCl_2$ (0·1 M), and 0·015 ml. micrococcal nuclease (7500 units/ml.) to the phenol-extracted endonuclease R limit digest. After incubation at 37°C for 15 min an additional 5 μl. of micrococcal nuclease was added and incubation was continued for another 15 min at 37°C. The dinucleoside monophosphates from the 3′-ends of the endonuclease R limit product were the only species in the micrococcal digest which possessed a single negative charge at pH 7·6. They were recovered by fractionating the digest according to net charge by the method described in section (c). (Note: the dGMP was included in the micrococcal nuclease reaction mixture to reduce the possibility of production of spurious dinucleoside monophosphate species by any residual phosphatase activity.)

3. Results

(a) *Nucleotide sequences at the 5′-ends of the limit product of endonuclease R digestion of T7 DNA*

As described in Smith & Wilcox (1970), endonuclease R of *H. influenzae* introduces approximately 40 double-strand breaks in phage T7 DNA. The cleavage is such as to produce 5′-phosphoryl and 3′-hydroxyl end groups. Figure 1 shows the general scheme used in analyzing the nucleotide sequences at the 5′-termini of the limit product of endonuclease R digestion of T7 DNA. Uniformly ^{33}P-labeled T7 DNA was

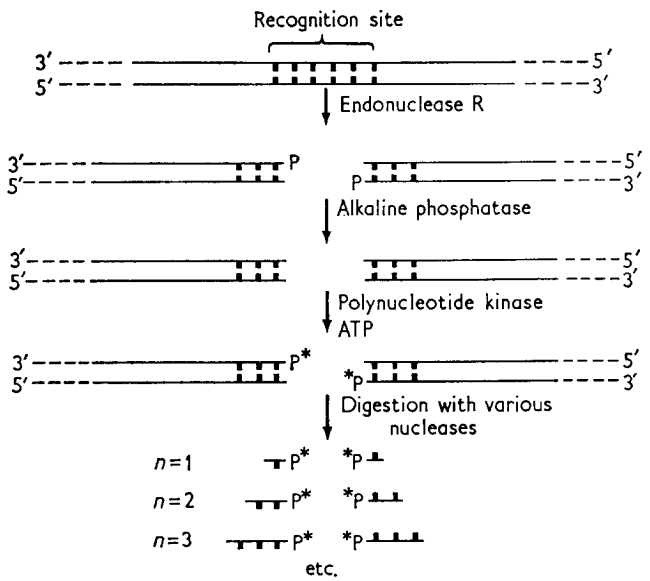

FIG. 1. General method for analysis of the nucleotide sequences at the 5′-termini of endonuclease R limit product.

Uniformly ^{33}P-labeled phage T7 DNA was digested to a limit product with endonuclease R of *H. influenzae*. The 5′-phosphoryl end-groups produced during this digestion were removed with alkaline phosphatase. ^{32}P-labeled phosphoryl groups were then transferred from [γ-^{32}P]ATP to the dephosphorylated 5′-termini of the limit product using polynucleotide kinase (Richardson, 1965). The 5′-terminal mono-, di- and trinucleotides were isolated following degradation of the terminally labeled limit product with various nucleases and identified as described in the text. Note: in the above Figure endonuclease R is pictured as producing an "even" duplex break at the exact center of its recognition site. It is possible to imagine a number of other possible patterns of cleavage. This particular pattern is shown only to simplify the illustration and does not imply any *a priori* assumptions concerning the mechanism of action of the enzyme.

digested to completion with endonuclease R. The 5′-phosphoryl groups produced during this digestion were removed with alkaline phosphatase. ^{32}P-labeled phosphoryl groups were then transferred from [γ-^{32}P]ATP to the dephosphorylated 5′-termini using polynucleotide kinase (Richardson, 1965). The resulting product (referred to below as 5′-^{32}P-terminally labeled endonuclease R limit product) was digested with various nucleases and the ^{32}P-labeled fragments released were isolated and characterized.

(i) *5′-Terminal nucleotides*

A preparation of ^{33}P-labeled T7 DNA was digested to a limit product with endonuclease R. The 5′-ends were labeled with ^{32}P-phosphoryl groups as described above. A portion of this terminally labeled DNA was hydrolyzed to mononucleotides by sequential digestion with pancreatic DNase and snake venom phosphodiesterase, and the resulting digest was analyzed by electrophoresis on thin-layer cellulose at pH 3·55. Figure 2 shows densitometer tracings of unscreened (Fig. 2(a)) and screened (Fig. 2(b))

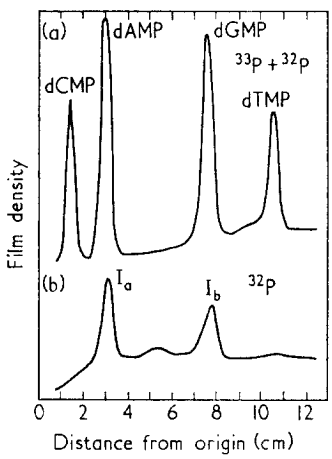

FIG. 2. Electrophoresis of the 5′-terminal mononucleotides of endonuclease R limit product. Endonuclease R limit product labeled at its 5′-termini with ^{32}P and internally with ^{33}P was hydrolyzed to mononucleotides by sequential digestion with pancreatic DNase and snake venom phosphodiesterase as described in Materials and Methods. The mononucleotides were separated by electrophoresis on a cellulose thin-layer sheet in 0·075 M-ammonium formate, pH 3·55, at a potential gradient of 25 v/cm. Following electrophoresis, screened and unscreened autoradiographs of the thin-layer sheet were prepared as described in Materials and Methods section (d).

(a) Microdensitometer tracing of unscreened autoradiograph (^{32}P and ^{33}P); (b) microdensitometer tracing of screened autoradiograph (^{32}P).

autoradiographs of the thin-layer sheet prepared following electrophoresis (see Materials and Methods). Examination of the screened autoradiograph revealed that dAMP and dGMP were the only mononucleotides which contained ^{32}P. However, the radioactivity was not divided equally betweeen the two mononucleotides, dAMP containing about 63% of the total ^{32}P radioactivity and dGMP containing about 37% (see Table 1). This result appears to reflect a real difference in the relative frequency of dAMP terminals *versus* dGMP terminals rather than a difference in the recovery of the two nucleotides, since the distribution of ^{33}P radioactivity closely paralleled the known base composition of T7 DNA. Furthermore, the relative frequency of dAMP and dGMP seems to be a function of the particular DNA substrate

TABLE 1

Analysis of 5′-terminal nucleotides

Nucleotide	^{33}P (cts/min)	^{32}P (cts/min)
d(pC)	720	0
d(pA)	710	170
d(pG)	690	100
d(pT)	700	5

The areas containing the four 5′-nucleoside monophosphates were cut from the cellulose thin-layer strip of Fig. 2. and assayed for ^{32}P- and ^{33}P-radioactivity in a liquid-scintillation spectrometer.

used, since analysis of the 5′-terminal nucleotides of an endonuclease R limit digest of phage P22 DNA yielded 85% dAMP and 15% dGMP. The simplest interpretation of these facts is that the nucleotide immediately adjacent to a site of phosphodiester bond cleavage by endonuclease R is not absolutely specified, i.e. the enzyme cleaves both strands of the DNA duplex on the 5′-side of a purine nucleotide but does not unambiguously distinguish between the two possible purine nucleotides, dAMP and dGMP.

(ii) *5′-Terminal dinucleotides*

A portion of 5′-^{32}P-terminally labeled endonuclease R limit product was digested with pancreatic DNase. The resulting collection of oligonucleotides was then digested with exonuclease I. Exonuclease I acts from the 3′-hydroxyl end of a single-stranded

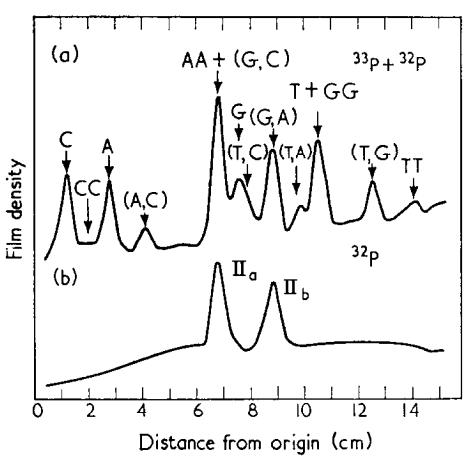

FIG. 3. Electrophoresis of the 5′-terminal dinucleotides of endonuclease R limit product. Endonuclease R limit product labeled at its 5′-termini with ^{32}P and internally with ^{33}P was digested sequentially with pancreatic DNase and exonuclease I as described in Materials and Methods. The resulting collection of mono- and dinucleotides was fractionated by electrophoresis at pH 3·55 as described in the legend to Fig. 2. The known positions of the various dinucleotides are indicated on the tracing of the unscreened autoradiograph (a). Since isomeric dinucleotides were not resolved under the conditions used for electrophoresis, the symbol (X, Y) is used to indicate the position of both pXpY and pYpX.

(a) Microdensitometer tracing of unscreened autoradiograph (^{32}P and ^{33}P); (b) microdensitometer tracing of screened autoradiograph (^{32}P).

TABLE 2

Nucleotide analysis of 5′-terminal dinucleotides

Nucleotide	Species IIa ^{32}P (cts/min)	Species IIb ^{32}P (cts/min)
d(pC)	2	1
d(pA)	1485	7
d(pG)	9	857
d(pT)	4	0

The areas containing species IIa and IIb of Fig. 3 were cut from the thin layer, and the radioactive dinucleotides eluted with distilled water. A portion of each was hydrolyzed to mononucleotides with snake venom phosphodiesterase. The reaction mixtures contained 0·1 M-glycine buffer (pH 9·2), 0·01 M-$MgCl_2$ and 0·2 mg snake venom phosphodiesterase/ml. After incubation at 37°C for 30 min, each reaction mixture was fractionated by electrophoresis at pH 3·55. The ^{32}P radioactivity in each mononucleotide was determined by liquid-scintillation counting.

DNA moledule releasing 5′-mononucleotides in stepwise fashion, but is unable to hydrolyze the 5′-terminal dinucleotide (Lehman, 1960). Electrophoretic analysis (pH 3·55) of the exonuclease I digest revealed two distinct ^{32}P-containing dinucleotide species (Fig. 3(b)). Species IIa contained about 62% of the total ^{32}P radioactivity and had an electrophoretic mobility compatible with three possible dinucleotides: pApA, pGpC and pCpG (Fig. 3(a)). Species IIa was eluted from the thin layer and hydrolyzed to 5′-mononucleotides with snake venom phosphodiesterase. The ^{32}P radioactivity was found only in dAMP (Table 2) indicating that species IIa was pApA. Species IIb contained about 38% of the total ^{32}P radioactivity and migrated at a rate compatible with either pGpA or pApG. Upon digestion to mononucleotides the ^{32}P radioactivity was found associated exclusively with dGMP (Table 2) indicating that species IIb was pGpA. This identification of the 5′-terminal dinucleotides of an endonuclease R limit digest of T7 DNA as pApAp and pGpA was confirmed by thin-layer cellulose chromatography against known marker dinucleotides in the solvent system of Markham & Smith (1952) as modified by Weiss & Richardson (1967) in experiments not reported here. On the basis of these results it was concluded that endonuclease R cleaves both strands of the DNA duplex on the 5′-side of the sequence d(pPupA).

(iii) *5′-Terminal trinucleotides*

Extending the sequence analysis of the 5′-ends of endonuclease R limit product beyond the dinucleotide level presented certain special problems. In particular, no method has been described for obtaining quantitatively higher order terminal oligonucleotides. The following method was employed to obtain a reasonably representative sample of the 5′-terminal trinucleotides. A preparation of 5′-^{32}P-terminally labeled endonuclease R limit product was digested with pancreatic DNase to an average chain length of about five residues/oligonucleotide. An average of about 1·4 residues was then removed from the 3′-end of each chain by partial digestion with snake venom phosphodiesterase. The resulting oligonucleotide mixture was fractionated according to chain length by DEAE-cellulose chromatography in the presence of 7 M-urea (Fig. 4) (Tomlinson & Tener, 1963). The trinucleotide fraction, containing about 20% of the original ^{32}P radioactivity, was collected and analyzed by electrophoresis on cellulose thin-layer at pH 3·55. Two ^{32}P-containing trinucleotide species

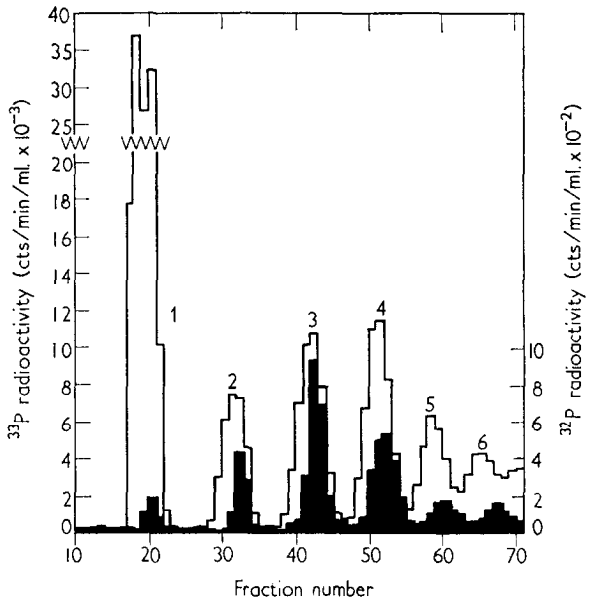

FIG. 4. Fractionation of 5′-terminal oligonucleotides of endonuclease R limit product according to chain length by DEAE-cellulose chromatography in the presence of 7 M-urea.

A preparation of endonuclease R limit product labeled at its 5′-termini with ^{32}P and internally with ^{33}P was digested to an average chain length of 5 residues with pancreatic DNase. The oligonucleotides thus formed were further degraded from the 3′-end with snake venom phosphodiesterase to a final average chain length of 2·1 residues (see Materials and Methods). The resulting digest was loaded onto a 1 cm × 10 cm DEAE-cellulose column, and the oligonucleotides were eluted according to chain length with a linear gradient of NaCl (0 to 0·3 M) in 0·02 M-Tris (pH 7·6) and 7 M-urea. The total gradient volume was 1 l. and the fraction volume was 8 ml. The chain lengths corresponding to the various oligonucleotide fractions are indicated in the Figure.

Shaded area, ^{32}P radioactivity; unshaded area ^{33}P radioactivity.

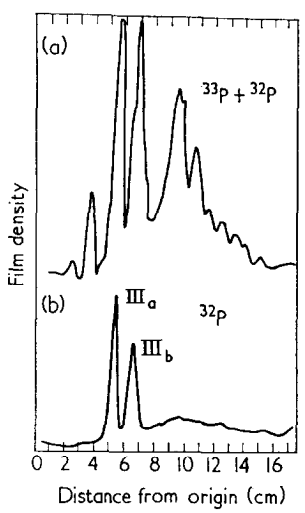

FIG 5. Electrophoresis of the 5′-terminal trinucleotides of endonuclease R limit product.

The trinucleotide fractions of Fig. 4 were pooled, desalted and concentrated as described in Materials and Methods. Fractionation by electrophoresis was carried out under the conditions described in the legend to Fig. 2.

(a) *Microdensitometer tracing of unscreened autoradiograph* (^{32}P and ^{33}P); (b) *microdensitometer tracing of screened autoradiograph* (^{32}P).

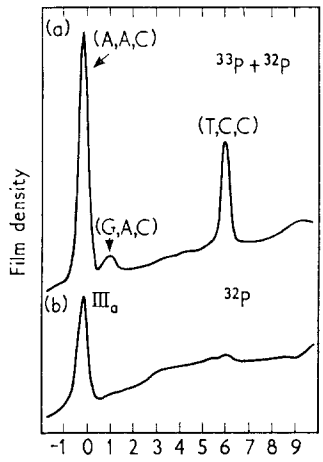

FIG. 6

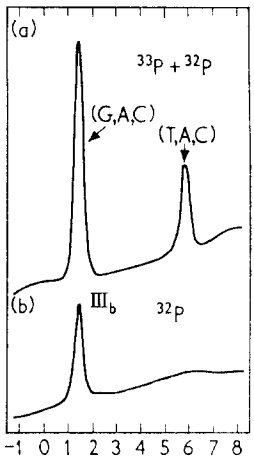

FIG. 7

FIG. 6. Separation of species IIIa from contaminating trinucleotides by electrophoresis at pH 1·9.

The area of the thin-layer cellulose sheet corresponding to species IIIa (Fig. 5) was cut out and the labeled trinucleotides were eluted with 0·02 M-ammonium bicarbonate. Electrophoretic fractionation of this material was carried out on cellulose thin-layer in 8·7% acetic acid, 2·5% formic acid (pH 1·9) at a potential gradient of 25 v/cm. Autoradiography of the thin layer was performed in the usual manner (see Materials and Methods section (d)). The base composition of the various trinucleotides in the eluate is indicated on the tracing of the unscreened autoradiograph.

(a) Microdensitometer tracing of unscreened autoradiograph (^{32}P and ^{33}P); (b) microdensitometer tracing of screened autoradiograph (^{32}P).

FIG. 7. Separation of species IIIb from contaminating trinucleotides by eletrophoresis at pH 1·9.

The area containing species IIIb of Fig. 5 was cut from the cellulose thin-layer sheet and the labeled trinucleotides were eluted with 0·02 M-ammonium bicarbonate. Electrophoretic fractionation was carried out as described in the legend to Fig. 6.

(a) Microdensitometer tracing of unscreened autoradiograph (^{32}P and ^{33}P); (b) microdensitometer tracing of screened autoradiograph (^{32}P).

TABLE 3

Nucleotide analysis of 5′-terminal trinucleotides

Nucleotide	Species IIIa		Species IIIb	
	^{33}P (cts/hr)	^{32}P (cts/hr)	^{33}P (cts/hr)	^{32}P (cts/hr)
d(pC)	3496	14	3410	0
d(pA)	6371	3020	2950	0
d(pG)	98	45	2783	2050
d(pT)	0	14	0	0

Species IIIa and IIIb of Figs 6 and 7, respectively, were eluted from the cellulose thin-layer with 0·02 M-ammonium bicarbonate. Separate samples of each species were hydrolyzed to mononucleotides as described in the legend to Table 2. The mononucleotides were separated by two-dimensional chromatography on a PEI-cellulose thin-layer sheet as described in Materials and Methods section (c). The ^{32}P- and ^{33}P-radioactivity in each of the four deoxyribonucleotides was assayed in a liquid-scintillation spectrometer.

were observed (Fig. 5). These species were individually eluted from the cellulose thin-layer and separated from contaminating trinucleotides by a second dimension of electrophoresis at pH 1·9 (Figs 6 and 7). Nucleotide analyses of the isolated 5′-terminal trinucleotides are shown in Table 3. In the case of species IIIa the ^{32}P radioactivity was associated exclusively with dAMP. The ^{33}P radioactivity was associated with dAMP and dCMP in a ratio of about 2:1. In the case of species IIIb the ^{32}P radioactivity was associated with dGMP. The ^{33}P radioactivity was found in dGMP, dAMP and dCMP in roughly equal amounts. On the basis of these results together with those described in the preceding section, species IIIa was identified as d(pApApC) and species IIIb was identified as d(pGpApC). It was concluded that endonuclease R cleaves both strands of the DNA duplex on the 5′-side of the sequence d(pPupApC).

(b) *Nucleotide sequences at the 3′-ends of the limit product*

Endonuclease R cleaves the two strands of the T7 DNA duplex at points in close enough proximity to one another to allow the molecule to fall apart into two essentially duplex fragments (Smith & Wilcox, 1970). During this process no "extra" nicks are introduced into either strand nor is there any detectable release of acid-soluble material (Smith & Wilcox, 1970). As indicated in the preceding section, both strands are cleaved on the 5′-side of the sequence pPupApC. These facts are compatible with three possible structures for the endonuclease R substrate region (Fig. 8). In structure (a) the two strands of the substrate are broken at exactly opposite points ("even"

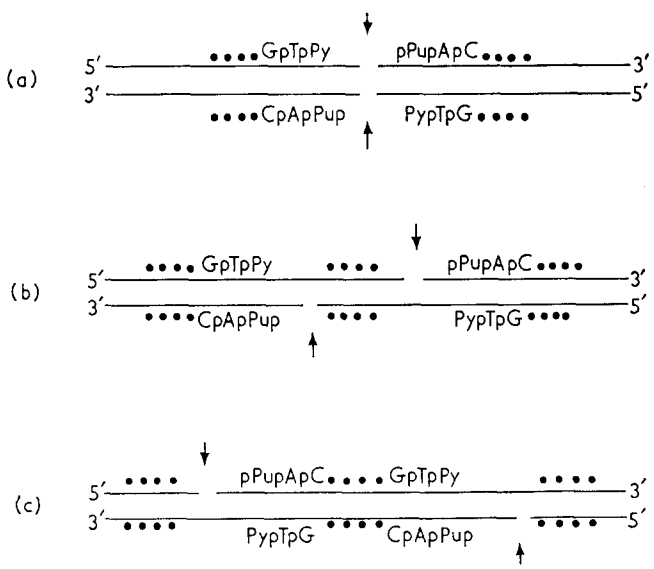

FIG. 8. Possible structures for the endonuclease R substrate region.

Analysis of the sequences at the 5′-ends of endonuclease R limit product has indicated that the enzyme nicks both strands of the DNA duplex on the 5′-side of the sequence pPupApC. The Figure shows 3 structures for the endonuclease R substrate region which are consistent with this finding. In structure (a) the nicks (shown by arrows) are exactly opposite one another so that the 3′-ends of the fragments produced are complementary to the 5′-ends. In structures (b) and (c) the nicks are staggered so that either the 3′-ends (b) or the 5′-ends (c) of the product fragments extend beyond their complementary partner strands as short single-stranded tails. In both of these structures the 3′-ends of the product fragments are not necessarily complementary to the 5′-ends.

break), while in structures (b) and (c) a short stretch of nucleotide pairs intervenes between the two breaks ("staggered" break). Structure (a) can easily be distinguished from the others because only in the case of an "even" break are the 3′-ends of the limit product necessarily complementary to the 5′-ends. To investigate this latter possibility an analysis of the 3′-dinucleotides of the endonuclease R limit product was undertaken.

Uniformly ^{32}P-labeled T7 DNA was digested to a limit product with endonuclease R. The resulting digest was then incubated with micrococcal nuclease. Micrococcal nuclease produces 3′-phosphoryl, 5′-hydroxyl endonucleolytic cleavages and reduces DNA molecules (native or denatured) to a limit product consisting of about 50% each of nucleoside-3′-phosphates (Xp) and dinucleoside-3′-phosphates (XpYp) (Sulkowski & Laskowski, 1962). The ends of the DNA molecules yield unique nucleotide species. The 5′-ends yield pXp and pXpYp-type compounds, while the 3′ ends yield XpY's and nucleosides. In the present case the dinucleoside monophosphates (XpY's) from the 3′-ends of the endonuclease R limit digest were separated from all other species in the micrococcal digest by chromatography on DEAE-cellulose in the presence of 7 M-urea (Tomlinson & Tener, 1963). The isolated dinucleoside monophosphates were fractionated by electrophoresis at pH 3·55. Two major radioactive species were observed (Fig. 9). Species I released pC when digested with snake venom

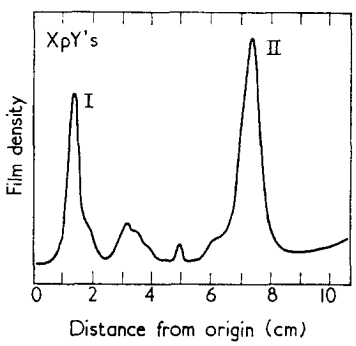

FIG. 9. Electrophoresis of the 3′-terminal dinucleoside monophosphates of endonuclease R limit product.

Uniformly ^{32}P-labeled endonuclease R limit product was digested to completion with micrococcal nuclease and the 3′-terminal dinucleoside monophosphates were isolated as described in Materials and Methods. Electrophoretic fractionation on a cellulose thin layer-sheet at pH 3·55 was performed as described in the legend to Fig. 2. The Figure shows a densitometer tracing of an unscreened autoradiograph of the thin-layer sheet.

phosphodiesterase and Tp when digested with spleen phosphodiesterase (Fig. 10) indicating that it was TpC. Species II yielded pT when digested with snake venom phosphodiesterase and Tp when digested with spleen phosphodiesterase indicating that it was TpT. This was the expected result for an "even" break. It was concluded that structure (a) of Figure 8 most closely describes the endonuclease R substrate region.

(c) *Size of the recognition sequence*

Assuming for simplicity that the base sequence of T7 DNA is random, the sequence of six nucleotide pairs depicted in Figure 8(a) would be expected to occur about once

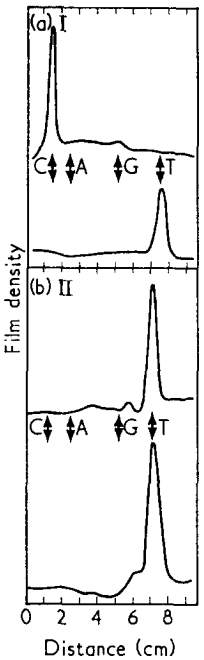

FIG. 10. Sequence analysis of the 3′-terminal dinucleoside monophosphates of endonuclease R limit product.

Species I and II of Fig. 9 were eluted from the cellulose thin-layer with 0·02 M-ammonium bicarbonate. Separate portions of each species were hydrolyzed (37°C for 30 min) with snake venom phosphodiesterase in reaction mixtures containing 0·1 M-glycine buffer, pH 9·2, 0·01 M-$MgCl_2$, and 0·2 mg phosphodiesterase/ml. Separate fractions of each were also hydrolyzed (37°C for 30 min) with spleen phosphodiesterase in reaction mixtures containing 0·1 M-Tris (pH 7·1), 0·02, M-$MgCl_2$, and 0·5 unit phosphodiesterase/ml. (Snake venom phosphodiesterase produces 5′-mononucleotides and spleen phosphodiesterase produces 3′-mononucleotides.) Marker mononucleotides were added to the 4 digests and electrophoretic fractionation was carried out on a cellulose thin-layer sheet in 0·075 M-ammonium formate, pH 3·55, at a potential gradient of 25 v/cm. Following electrophoresis a single unscreened autoradiograph of the thin-layer sheet was prepared.

(a) Nucleotides released by digestion of species I with snake venom phosphodiesterase (upper curve) and spleen phosphodiesterase (lower curve).

(b) Nucleotides released by digestion of species II with snake venom phosphodiesterase (upper curve) and spleen phosphodiesterase (lower curve.)

in every 1024 nucleotide pairs. The average length of the limit product of endonuclease R digestion of T7 DNA is about 1000 nucleotide pairs (Smith & Wilcox, 1970). Similarly, for phage P22 DNA the average length of the limit product is about 1300 nucleotide pairs. These facts suggest that the sequence of Figure 8(a) contains sufficient information to account fully for the observed degree of specificity of endonuclease R. However, since strictly speaking neither the T7 nor the P22 DNA molecule represents a truly random collection of nucleotides, the possibility remains that the recognition region for endonuclease R might be larger than the above calculations suggest. In order to rule out this possibility an analysis of the 5′-terminal tetranucleotides of the endonuclease R limit product was undertaken.

A 5′-^{32}P-terminally labeled limit digest of T7 DNA was hydrolyzed with pancreatic DNase to an average chain length of about 4·4. The digest was then fractionated

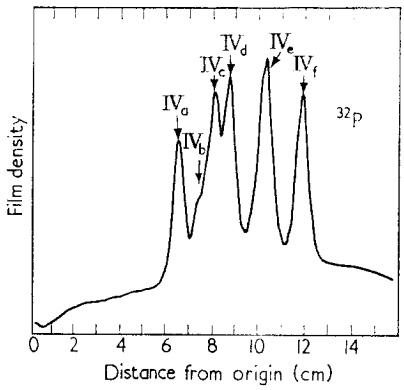

FIG. 11. Electrophoresis of 5′-terminal tetranucleotides of endonuclease R limit product. Endonuclease R limit product labeled at its 5′-termini with ^{32}P was digested to an average chain length of 4·4 with pancreatic DNase. The tetranucleotide fraction was isolated from this digest by ion-exchange chromatography in the presence of 7 M-urea (see Fig. 4) and fractionated by electrophoresis at pH 3·55 under the conditions described in the legend to Fig. 2. A screened autoradiograph of the thin-layer sheet was prepared as described in Materials and Methods section (c). The Figure shows a microdensitometer tracing of this autoradiograph.

according to chain length by ion-exchange chromatography in the presence of 7 M-urea (Tomlinson & Tener, 1963). The tetranucleotide fraction, containing about 30% of the original ^{32}P radioactivity, was collected and analyzed by electrophoresis at pH 3·55 (Fig. 11). Six distinct ^{32}P-containing tetranucleotides were observed, clearly indicating that the fourth nucleotide from the 5′-ends of the limit product is not uniquely specified. It was concluded that the sequence of six nucleotide pairs illustrated in Figure 8(a) represents the complete recognition region for endonuclease R.

It should be noted that if the fourth position from the cleavage point can be occupied by any nucleotide pair, then eight tetranucleotides (rather than six) would be expected. It is possible that the two tetranucleotides not accounted for in the present experiment occur at such a low frequency in the T7 DNA molecule as to escape detection, or that two of the bands observed during electrophoresis contained more than one species.

4. Discussion

Endonuclease R of *H. influenzae* and other restriction enzymes belong to a class of proteins which act at a limited number of specific sites within a polynucleotide. Other members of this class (e.g. repressors, RNA polymerases, aminoacyl-tRNA synthetases) appear to have important roles in a number of basic cellular processes, including transcription, translation and the control of gene expression (for a recent review see Yarus, 1969). It is generally agreed that the site-specific action of these proteins is a consequence of their ability to recognize specific nucleotide sequences. However, at present the nature of these sequences is largely unknown.

The experiments described in this paper represent an attempt to reconstruct the complete nucleotide sequence of the recognition site of endonuclease R by studying the nucleotide sequences at the ends of the fragments produced by the action of the

enzyme. The results indicate that endonuclease R recognizes the following nucleotide sequence:

$$\begin{array}{c} \qquad\downarrow \\ 5' \ldots \text{GpTpPypPupApC} \ldots 3' \\ 3' \ldots \text{CpApPupPypTpG} \ldots 5' \\ \qquad\uparrow \end{array}$$

where the arrows indicate the sites of phosphodiester bond hydrolysis.

Since the opportunities for interaction between the enzyme and the bases of this sequence are rather limited when DNA is in the double helical configuration, it seems reasonable to suppose that the helix is disrupted locally during the recognition process. It is not known at present which of the 12 bases in the sequence are involved in specific interaction with the enzyme. However, the fact that endonuclease R does not degrade single-stranded DNA suggests the possibility that bases on both strands are involved.

Reasoning by analogy to the *E. coli* restriction system, we assume that endonuclease R does not degrade *H. influenzae* DNA because this recognition sequence is modified in some way. Nevertheless, since we do not yet have direct evidence for the existence of a modification activity in *H. influenzae*, we must, for the sake of logical completeness, consider the possibility that *H. influenzae* DNA does not contain the sequence. However, the widespread occurrence in bacteria of host-controlled modification renders this latter alternative unlikely (Arber, 1965).

The most striking feature of the recognition site of endonuclease R is its symmetry. Disregarding for the moment the central ambiguity, the sequence possesses a 2-fold rotational axis of symmetry perpendicular to the helix axis, i.e. when read with the same polarity, the sequence of bases on one strand is the same as that on the other. Of course, the presence of the ambiguity in the two central positions means that the symmetry of the sequence may not always be exact in the geometric sense, i.e. one purine site may be occupied by A and the other by G. From an operational point of view, however, this distinction is probably not important since the enzyme does not distinguish between A and G in these positions. From the enzyme's point of view the sequence is always symmetrical and the molecular environment about each of the two cleavage points is the same.

It is unlikely that the symmetry of this sequence is fortuitous, since the number of possible asymmetrical sequences of this type is about 30 times the number of possible symmetrical sequences (See Arber & Linn (1969) for a general discussion of the expected properties of symmetrical and asymmetrical recognition sequences.) It is probable, therefore, that the observed symmetry is a reflection of some fundamental feature of the mechanism of action of endonuclease R. Figure 12 presents two possible models for the action of the enzyme which take advantage of symmetry.

In model I a single enzyme molecule recognizes six bases on one strand and introduces a single "nick". The presence of a symmetrical recognition site allows the same enzyme to recognize and nick the opposite strand in exactly the same place to produce a duplex break. If the sequence were asymmetrical and both the recognition and the nucleolytic processes involved only one strand as depicted in model I, two enzymes with different recognition specificities would be required.

A more interesting possibility is that the symmetry of the recognition site may reflect underlying symmetry in the enzyme. There is good reason to believe that closed oligomeric proteins composed of identical or nearly identical subunits display

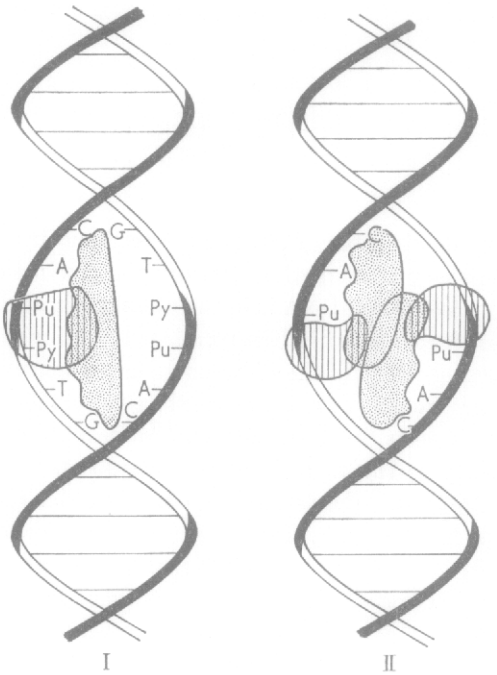

FIG. 12. Recognition of a symmetrical nucleotide sequence by endonuclease R. Two possible models.

Model I. A single enzyme molecule specifically binds to the base sequence GpTpPypPupApC. The active site of the enzyme (shaded area) catalyzes 5′-phosphoryl, 3′-hydroxyl cleavage of the phosphodiester bond between Py and Pu. The enzyme then detaches and attacks the other strand which (because of symmetry) contains the same base sequence. An even duplex break results. (Note: since endonuclease R is inactive on single-stranded DNA, it is necessary to assume in this model that the enzyme in some way "senses" the bihelical configuration of the substrate).

Model II. The enzyme is composed of two identical subunits (related by a 2-fold rotational axis of symmetry) which bind to the sequence pPupApC on opposite strands of the DNA duplex. Each subunit is actually a dimer constructed from a "recognition" subunit (stippled) and a "nuclease" subunit (shaded). The recognition subunits bind to A and C, while the nuclease subunits bind to Pu. Hydrolysis of the phosphodiester bonds between Py and Pu by the nuclease subunits results in a duplex break.

rotational symmetry (for a discussion of this point see Monod, Wyman & Changeux, 1965). In particular, a closed dimer is expected to possess a 2-fold rotational axis. These considerations suggest the possibility, depicted in model II, that recognition is accomplished by two identical subunits related by a 2-fold axis of symmetry. In this model each subunit recognizes the same sequence of three bases on opposite strands of the DNA duplex. From the standpoint of economy of genetic information it is much cheaper to specify two identical subunits each capable of recognizing three bases in a symmetrical sequence of six than to specify a larger protein capable of recognizing the entire sequence.

In model II each of the two identical subunits is constructed from two non-identical subunits, a "recognition" subunit and a "nuclease" subunit. We suggest the possibility that this division of labor occurs for two reasons. In the first place there is genetic evidence in the *E. coli* system that at least three gene products are involved in restriction and modification. One product carries out the restriction function, another

the modification function, and the third determines the site specificity (Arber & Linn, 1969). In the second place the recognition sequence of endonuclease R contains two distinct levels of specificity. The two central base pairs are only partially specified, whereas the remainder of the sequence is absolutely specified. Many general nucleases are known to possess the limited type of specificity displayed in the central part of the sequence, e.g. pancreatic DNase preferentially cleaves phosphodiester linkages of the type PupPy (Laskowski, 1959). Thus, the partial specificity of endonuclease R at the cleavage point could be determined by a nuclease subunit which binds in this region, while the absolute specificity of the enzyme for the remainder of the sequence could be determined by a recognition subunit. This would mean that the recognition subunit need recognize only two bases. Of course, confirmation of these speculations will depend upon further structural studies on the enzyme.

In retrospect, it is probably not very surprising that endonuclease R recognizes a symmetrical sequence since the enzyme carries out an essentially symmetrical operation, namely the cleavage at equivalent points of two DNA strands of opposite polarity. This consideration suggests that symmetrical recognition sequences may be the rule for restriction enzymes in general. On the other hand, many of the proteins that are known to possess the ability to recognize a specific nucleotide sequence carry out fundamentally asymmetrical operations (e.g. RNA polymerase). The recognition sequences for these proteins would be expected to be asymmetrical.

This work was supported by U.S. Public Health Service grant number AI-07875. One of us (H.O.S.) was the recipient of U.S. Public Health Service Career Development Award no. AI 17902.

REFERENCES

Arber, W. (1965). *Ann. Rev. Microbiol.* **19**, 365.
Arber, W. & Linn, S. (1969). *Ann. Rev. Biochem.* **38**, 467.
Holley, R. W., Madison, J. T. & Zamir, A. (1964). *Biochem. Biophys. Res. Comm.* **17**, 389
Josse, J. & Moffatt, J. G. (1965). *Biochemistry*, **4**, 2825.
Kelly, T. J. & Thomas, C. A. (1969). *J. Mol. Biol.* **44**, 459.
Laskowski, M., Sr. (1959). *Ann. N.Y. Acad. Sci.* **81**, 776.
Lehman, I. R. (1960). *J. Biol. Chem.* **235**, 1479.
Lehman, I. R. & Nussbaum, A. L. (1964). *J. Biol. Chem.* **239**, 2628.
Linn, S. & Arber, W. (1968). *Proc. Nat. Acad. Sci., Wash.* **59**, 1300.
Markham, R. & Smith, J. D. (1952). *Biochem. J.* **52**, 552.
Meselson, M. & Yuan, R. (1968). *Nature*, **217**, 1110.
Monod, J., Wyman, J. & Changeux, J. (1965). *J. Mol. Biol.* **12**, 88.
Randerath, K. & Randerath, E. (1967). In *Methods in Enzymology*, ed. by L. Grossman & K. Moldave, vol. 12, part A, p. 323. New York: Academic Press.
Richardson, C. C. (1965). *Proc. Nat. Acad. Sci., Wash.* **54**, 158.
Rushizski, G. W. & Sober, H. A. (1962). *Biochem. biophys. Acta*, **55**, 217.
Sanger, F., Brownlee, G. G. & Barrell, B. G. (1965). *J. Mol. Biol.* **13**, 373.
Smith, H. O. (1968). *Virology*, **34**, 203.
Smith, H. O. & Wilcox, K. (1970). *J. Mol. Biol.* **51**, 379.
Sulkowski, E. & Laskowski, M., Sr. (1962). *J. Biol. Chem.* **237**, 2620.
Thomas, C. A. & Abelson, J. A. (1966). In *Procedures in Nucleic Acid Research*, ed. by G. Cantoni & D. Davies, p. 553. New York: Harper & Row.
Threlfall, C. J. (1957). *Biochem. J.* **65**, 694.
Tomlinson, R. V. & Tener, G. M. (1963). *Biochemistry*, **2**, 697.
Weiss, B., Live, T. R. & Richardson, C. C. (1968). *J. Biol. Chem.* **243**, 4530.
Weiss, B. & Richardson, C. C. (1967). *J. Mol. Biol.* **23**, 405.
Yarus, M. (1969). *Ann. Rev. Biochem.* **38**, 841.