Volume 2 number 10 October 1975 Nucleic Acids Research

Sequences spanning the EcoRI substrate site

David E. Garfin, Herbert W. Boyer and Howard M. Goodman*

Departments of Microbiology and Biochemistry & Biophysics*,
University of California Medical Center, San Francisco, CA 94143, USA

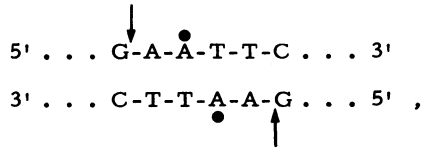
Received 19 August 1975

ABSTRACT

Substrate recognition by the EcoRI restriction endonuclease was investigated by analysis of the nucleotide sequences at the sites of enzymatic cleavage in various DNA molecules. 5'-end labeling and homochromatographic fingerprinting led to the determination of a 17-base-pair sequence spanning the EcoRI site of simian virus 40 DNA and a 15-base-pair sequence overlapping the EcoRI site of Col El plasmid DNA. Three other DNAs were similarly tested, although extended sequences were not determined in these cases. The EcoRI site was shown to be the symmetric, double-stranded equivalent of -N-G-A-A-T-T-C-N-.

INTRODUCTION

The Type II restriction endonucleases[1,2] are widely used in the analysis and manipulation of DNA molecules. The sequence-specific cleavage properties of these enzymes have been used to physically map viral genomes [3], to segment DNA for nucleotide sequence analysis [4,5,6], and to produce defined, cohesive termini for the in vitro construction of recombinant DNA molecules[7,8]. The nucleotide sequences at the cleavage sites of a number of Type II endonucleases have been determined[2,9,10]. All of the determined sites are relatively small (4-6 nucleotide base pairs) sequences possessing axes of twofold rotational symmetry. One of these enzymes, the EcoRI restriction endonuclease[11], makes two staggered, single-strand scissions in DNA, four base pairs apart, to produce DNA fragments with cohesive, phosphorylated 5'-termini. The cleavage site of the EcoRI endonuclease has been determined to be[12]:



where the arrows indicate the positions of the two broken internucleotide bonds. In addition, the EcoRI modification enzyme[1,2] has been demonstrated to act on the same physical sequence[13], methylating the adenine residues indicated by dots to produce N^6-methyl adenine. We report here a further analysis of the EcoRI restriction-modification site (EcoRI site).

Sequence analyses of restriction-enzyme cleavage sites have in general not been carried much beyond the first appearance of degeneracy[2,9,10]. This is true in the case of the EcoRI endonuclease where for technical reasons in the initial analysis only the first six nucleotides from the 5'-end of the EcoRI break were determined. When the EcoRI site was originally characterized, bacteriophage λ DNA was used as a substrate for the enzyme and only A-T base pairs were detected in the positions immediately adjacent to the cleavage site. We therefore wished to investigate the extent of symmetry in the vicinity of the EcoRI site and whether the enzyme can function with G-C nucleotide pairs flanking the site. Detection of a hyphenated symmetry around the EcoRI site would have implications with regard to substrate recognition.

End-labeling methods[9] are particularly well suited to the sequence analyses of restriction sites. For our purposes, 5'-terminal labeling[9,14] was especially advantageous because of the nature of the EcoRI site. EcoRI cleavage produces DNA fragments with protruding, single-stranded 5'-termini of known sequence[12] which provide reference oligonucleotides useful in sequence analyses. To determine whether G-C base pairs can occur adjacent to the EcoRI site, we analyzed the sequences at the EcoRI cleavage sites on the DNAs of bacteriophages λ and φ 80, Micrococcus lysodeikticus, simian virus 40 (SV40), and the Col E1 plasmid. The latter two molecules were ideal for extended sequence analyses. EcoRI digestion converts the covalently-closed, circular DNAs of both SV40[15,16] and Col E1[17] to full-length, linear molecules with the EcoRI restriction-site sequence at the 5'-end of each strand. Technically, the analyses in these cases were identical to the sequencing of the termini of the lambdoid phages[18]; only two 5'-sequences for each DNA had to be distinguished.

Our sequencing scheme was based on two-dimensional homochromatographic analysis[19,20] of the labeled termini. The procedure proved to be an easy and rapid method for determining the terminal sequences of DNA. We were able to obtain by direct analysis the sequence of 17 nucleotide pairs spanning the EcoRI site of SV40 DNA and 15 base pairs overlapping the EcoRI site of Col E1 plasmid DNA. These two sequences and the more

limited ones obtained from the EcoRI sites of the bacteriophage and micrococcal DNAs display no unusual features outside the six-base-pair sequence shown above. We find both G-C and A-T base pairs neighboring the hexamer and we find no elements of symmetry beyond it. We conclude that the EcoRI recognition site is the central hexamer of the previously determined sequence[12], and that this is substrate for the endonuclease regardless of the identity of the base pairs outside this sequence.

MATERIALS AND METHODS

Materials

Enzymes. EcoRI endonuclease, prepared by a published procedure[11], was generously provided by P. J. Greene. The enzyme was homogeneous as determined by SDS-polyacrylamide gel electrophoresis and was free of contaminating nucleolytic activities. Polynucleotide kinase was purchased from P-L Biochemicals, Inc. P-L lot number PR1 did not contain DNase or significant phosphatase activity. Bacterial alkaline phosphatase was code BAPF from the Worthington Biochemical Corporation and was used as supplied without further purification. Pancreatic DNase (DPFF) and snake venom phosphodiesterase were both also from Worthington. They were stored frozen at 1 mg/ml in their respective reaction buffers, thawed for use as needed, and refrozen.

DNAs. The preparations of SV40 DNA[13] and the bacteriophage DNAs[12] have been described. Col E1 plasmid DNA was purified from cleared lysates[21] by centrifugation to equilibrium in CsCl-propidium diiodide gradients. M. lysodeikticus DNA was purchased from Miles Laboratories, Inc., and was repeatedly treated with phenol before use.

Fractionation Media. Cellulose acetate strips for electrophoresis were cut to size (2.5 x 52 cm) from 5 x 300 cm rolls obtained from the Millipore Corp. Thin-layer chromatography plates (1 part DEAE-cellulose to 7.5 parts cellulose; 20 x 40 cm x 250 μm) were purchased from Analtech, Inc. BDH yeast RNA for the preparation of homomixture c was purchased from the Gallard-Schlesinger Chemical Mfg. Corp. DEAE-cellulose paper was Whatman DE-81 obtained from H. Reeve Angel and Company. The various types of Sephadex were from Pharmacia Fine Chemicals.

Methods

Preparation of [γ-^{32}P] rATP. [γ-^{32}P]rATP was routinely prepared by a modification of the method of Glynn and Chapell[14, 22, 23]. The modified procedure detailed below is based on an unpublished method of J. E. Dahlberg(personal communication) used in the preparation of [α-^{32}P]NTPs. Fifty millicuries of carrier-free ^{32}P-orthophosphate (ICN Pharmaceuticals,

Inc.) in 0.02N HCl were taken up in water and completely dried in a siliconized tube on an apparatus resembling the Buchler Evapomix. The dried phosphate was dissolved in 200 µl of a freshly prepared solution of 0.1M Tris-HCl, pH 8, 0.0001M EDTA, 0.014M β-mercaptoethanol, 0.001M sodium-3-phosphoglycerate (Sigma Chemical Company) containing 0.1-0.5 µmole of rATP and 6 units each of rabbit-muscle glyceraldehyde-3-phosphate dehydrogenase (Worthington) and 3-phosphoglyceric phosphokinase (Sigma). Two µmoles of $MgCl_2$ were added and the exchange reaction was allowed to proceed for 1 hour at room temperature (about 22°C). The reaction mixture was then diluted with 5-10 ml of water and loaded on a small (0.5 ml) column of QAE-Sephadex to purify the labeled nucleotide. The column was packed in water in a pasteur pipette loosely plugged with glass wool and washed first with about 2 ml of 0.1N acetic acid and then with about 2 ml of water before the diluted reaction mixture was applied. After washing the loaded column with about 5-10 ml of water, $^{32}PO_4^{3-}$ was eluted with 0.02N HCl and [γ-^{32}P]rATP with 0.2N HCl. To prevent phosphorolysis, the labeled rATP was collected into a tube containing about 1 ml of concentrated ammonium hydroxide. This solution was dried to approximately 0.2 ml and the nucleoside triphosphate was desalted on a G-10 Sephadex column (0.8 x 30 cm) in 0.005M Tris-HCl, pH 8, 0.0001M EDTA. Elutions were monitored by passing the column effluents in front of the detector tube of a radiation counter. The desalted rATP was dried and redissolved in 100 µl H_2O, mixed with 100 µl of 100% ethanol, and sampled for quantitation. The concentration of nucleotide was measured spectrophotometrically ($\epsilon_{259\ nm}$ = 15,400) and the radioactivity by liquid scintillation counting. When assayed by electrophoresis at pH 3.5 on DEAE-cellulose and Whatman 540 papers, more than 98% of the radioactivity was found in the position of rATP. Approximately 50% of the input $^{32}PO_4^{3-}$ and 80% of the input rATP were recovered as [γ-^{32}P]rATP. The specific activity of the labeled product was dependent on the amount of rATP in the reaction: 0.5 µmole of rATP led to specific activities of 50-100 mCi/µmole; and 0.1 µmole rATP gave 200-300 mCi/µmole. The entire procedure took 4-5 hours. The [γ-^{32}P]rATP was stored at -20°C in 50% ethanol and was suitable for kinase labeling reactions for 1-2 months.

Terminal Labeling of EcoRI-digested DNA. Intact DNA (100-200 µg) was digested to completion with EcoRI endonuclease at 37°C in a buffer consisting of 0.1M Tris-HCl, pH 7.5, 0.005M $MgCl_2$, and 0.05M NaCl [11]. Sample aliquots of the digestion mixture were electrophoresed into 1.2% Agarose gels [11] to check that the hydrolysis was complete. At the conclu-

sion of the digestion, EDTA was added to a concentration of 0.01M and the solution was dialyzed into 0.02M Tris-HCl, pH 8, 0.002M EDTA. Terminal phosphates were removed from the EcoRI-digested DNA by incubation with bacterial alkaline phosphatase. Phosphatase was added to the dialyzed DNA solution to a concentration of 1-2 units/ml and the solution was incubated at 37°C for 1 hour. Enzymes were extracted with phenol and the DNA solution was dialyzed against 0.2M NaCl, 0.001M EDTA. The dephosphorylated DNA was recovered by centrifugation following precipitation from cold 70% ethanol. The DNA was then dissolved in 200 μl of 0.07M Tris-HCl, pH 7.4, 0.01M $MgCl_2$, 0.014M β-mercaptoethanol containing a 500- to 1000-fold molar excess of [γ-^{32}P]rATP over DNA 5'-ends. The 5'-termini of the EcoRI-digested DNA were rephosphorylated by incubating the preceding mixture with 4-5 units of polynucleotide kinase for 1-2 hours at 37°C. 100 μl of 0.1M NaCl were added and the 5'-labeled DNA was separated from unreacted [γ-^{32}P]rATP by passage through a 1.2 x 45 cm column of G-75 Sepahdex in 0.1M NaCl. The DNA in the excluded volume of the column was recovered by precipitation from 70% ethanol.

Digestion of DNA to Oligonucleotides. EcoRI-digested, 5'-labeled DNA, prepared as above, was hydrolyzed to random-sized oligonucleotides with pancreatic DNase (0.1 mg/ml in 200 μl of 0.01M Tris-HCl, pH 7.4, 0.01M $MgCl_2$). Incubation with DNase was usually for 30 minutes, but trial digestions of from 10 to 60 minutes were initially carried out to make certain that all possible 5'-labeled oligonucleotides were being generated. The limitations imposed by cellulose acetate electrophoresis (see below) on sample volume made it necessary to desalt the oligonucleotide mixtures pH 8.5, 0.01M $MgCl_2$. Sample aliquots from the room-temperature (about 22°C) reactions were taken at 15-minute intervals for from 60 to 90 minutes. These were applied to the origin of a 110-cm sheet of DEAE-cellulose paper for separation by high-voltage (2 kV) electrophoresis in a pH 1.9 electrolyte buffer[25] (2.5% formic acid, 8.7% acetic acid). In early trials, the sample aliquots were divided in two for electrophoresis at pH 3.5[25] (5% acetic acid, 0.5% pyridine) as well as at pH 1.9 for additional confirmation of the identities of the smallest molecules. The labeled digestion products were located by radioautography and the relative mobilities of the oligonucleotides in the sample lanes were measured for "M-value" determinations [9, 14, 25, 26] (M-values are diagnostic of the 3'-terminal residues exonucleolytically removed by the snake venom enzyme. The M-value for a particular split is the ratio of the difference in mobilities between an oligonucleotide and its first degradation product to the mobility of the oligonucleo-

tide[25]). To verify sequence assignments, the partial-digestion products of selected oligonucleotides were fractionated two-dimensionally on the homochromatography system. In these cases, the timed aliquots were squirted on polyethylene sheets and immediately dried in a vacuum desiccator. These digests were pooled by drying each aliquot on top of the preceding ones. The final dried samples were dissolved in 2-3 μl of 7M urea-pH 3.5 buffer for cellulose acetate electrophoresis.

RESULTS

We have used 5'-end labeling and homochromatographic fingerprinting to determine the nucleotide sequences in the vicinities of the EcoRI sites of five different DNAs. These procedures are very simple and rapid, and are quite reliable for the direct analysis of short DNA sequences.

Purified DNA was digested to completion with EcoRI endonuclease. The 5'-terminal phosphates remaining after EcoRI hydrolysis were removed with alkaline phosphatase and replaced with ^{32}P-phosphate from high-specific-activity [γ-^{32}P]rATP in the polynucleotide kinase reaction. The resultant end-labeled DNA was purified by gel filtration chromatography and then digested to small, random-sized oligonucleotides with pancreatic DNase. The oligonucleotides were fractionated in two dimensions by cellulose-acetate electrophoresis and homochromatography. Labeled oligonucleotides were located on the homochromatograms by radioautography, their positions before concentrating them for fractionation. This was done by first diluting the DNase-digestion solution with 5 ml of 0.01M triethylamine bicarbonate, pH 8 (TEAB), then adsorbing the oligonucleotides to a small (0.5 ml) column of DEAE-Sephadex A-25 (bicarbonate form) in 0.01M TEAB. The column was washed with approximately 20 ml of 0.01M TEAB. Then, oligonucleotides were eluted with 4.5 ml of 2M TEAB and dried in an air stream in a siliconized tube at 37°C. The dried oligonucleotides were dissolved in a small volume of water, transferred to a polyethylene sheet, and dried in a vacuum desiccator.

Fractionation of Oligonucleotides[19,24]. The dried pancreatic DNase digestion products were dissolved in 2-3 μl of water and applied to a strip of cellulose acetate wetted with 7M urea, 10% acetic acid, 0.001M Na_2-EDTA, pH 3.5, for electrophoretic fractionation. High-voltage electrophoresis at 6 kV was continued until the distance between the blue (xylene cyanol FF) and major pink (acid fuchsin) components of the tracking dye[25] was about 16 cm. Nearly all of the radioactivity migrated between these two dye markers, and this region was transferred by blotting[19] to a 1:7.5 DEAE:cellulose thin-layer plate. pA and pA-A were

generated in very low yield. They migrated more slowly than the blue dye and were not included in the region transferred to the thin layer. After the transfer, the thin layer was dried and washed with 95% ethanol to remove residual urea. Second-dimensional fractionation was by ascending homochromatography at 60°C in homomixture c. (Homomixture c was prepared by the recipe given by Brownlee and Sanger [19] and consisted of a neutralized and extensively dialyzed 3% solution of 30-minute-hydrolyzed RNA in 7M urea, 0.001M EDTA. It was stored frozen, thawed for use, and reused until resolution had noticeably deteriorated.) After chromatography, the thin-layer plate was dried and radioautographed. The labeled oligonucleotides from the 5'-termini of the EcoRI-digested DNA were located by means of the radioautograph, eluted as described [19,25] with three additions of 200 µl of 2M TEAB, dried under streams of air, and washed three times with water.

Partial Digestion with Snake Venom Phosphodiesterase. Samples eluted from homochromatography thin layers were dissolved in 25 µl of a 0.25 mg/ml solution of snake venom phosphodiesterase in 0.02M Tris-HCl, used as indicators of their sequences, and the sequences confirmed by analysis of the products of partial exonucleolytic hydrolysis of the eluted molecules.

A representative fingerprint obtained from 5'-labeled, EcoRI-cleaved SV40 DNA is shown in Fig. 1. Cleavage of covalently-closed, supercoiled SV40 DNA by EcoRI endonuclease converted the DNA to the linear form terminated at both ends with the nucleotide sequence of the EcoRI substrate site. After labeling and fingerprinting, only those oligonucleotides arising from the 5'-termini of the two strands of the EcoRI-cleaved DNA were detectable by radioautography. Moreover, the first five nucleotides from the 5'-end of each strand were the components of the EcoRI substrate sequence. The two sequences at the 5'-ends of the two labeled DNA strands were well resolved on the homo chromatograms and this allowed the sequences to be essentially "read off" from the spots on the radioautograph[20]. The relative intensities of the spots in the fingerprints may reflect the cleavage preferences of pancreatic DNase[27].

It has been recognized for some time [6,9,20,24,28,29,30] that the position of an oligonucleotide on a two-dimensional electrophoretic-homochromatographic fingerprint depends on its composition; the relative positions of two oligonucleotides originating from a common terminus depends on the sequence differences between the two compounds. The pattern shown in Fig. 1 represents two families of sequences originating

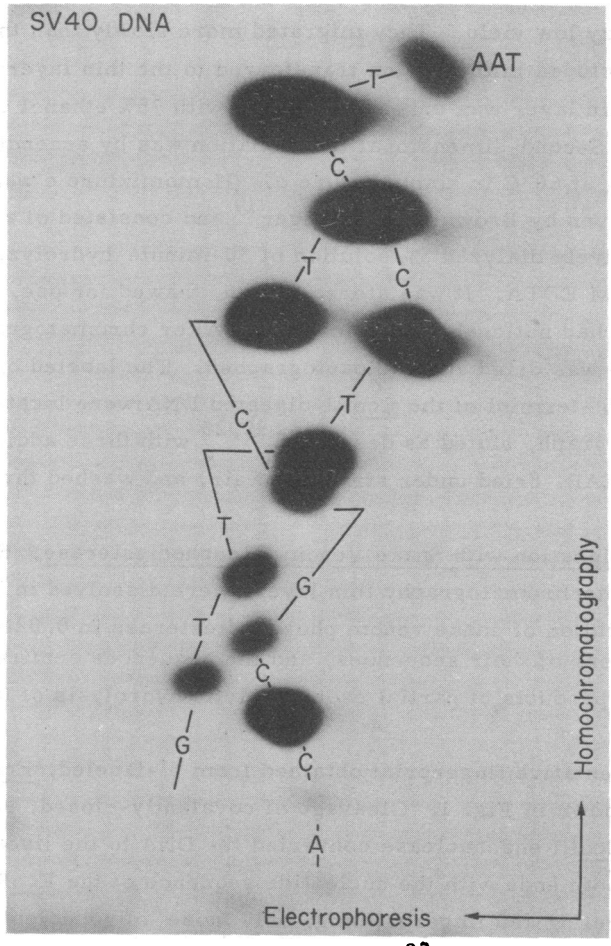


Fig. 1. Pancreatic DNase fingerprint of [5'-^{32}P]labeled, EcoRI-cleaved SV40 DNA. EcoRI cleavage, end labeling, and DNase digestion are described in the text. Fractionation in the first dimension was by electrophoresis at pH 3.5 on cellulose acetate and in the second direction by thin-layer homochromotography in homomixture c. Sequences originating from the 5'-termini of both DNA strands are traced. The letters represent the 3'-terminal residues differentiating the oligonucleotides connected by the lines. Symbols for phosphates have been omitted.

from the 5'-ends of the two strands of SV40 DNA. By multiple repetitions and by varying pancreatic DNase digestion times from 10 to 60 minutes, we have ascertained that every labeled molecule shown in Fig. 1 differs from its sequence neighbors by only one nucleotide. When two oligonucleotides differ from one another by a single nucleotide, they migrate relative to one another to positions characteristic of that nucleotide. Thus,

once we had distinguished the two families of oligonucleotides from each other, we could determine the two terminal sequences by tracing out the relative positions of adjacent oligomers.

Each labeled oligonucleotide was eluted from the thin-layer and subjected to partial exonucleolytic digestion with snake venom phosphodiesterase. Radioautographs of the electrophoretic fractionation patterns of the partial-digestion products revealed the stepwise exonucleolytic removal of nucleotides from the 3'-ends of the 5'-labeled oligomers. The appearance of "key" oligonucleotides on the electrophoretograms of the partial exonuclease digests allowed us to place each parent oligonucleotide in one or the other of the two 5'-terminal sequences at the EcoRI site. In addition, by calculating M-values for each residue removed by the venom enzyme, we were able to confirm and determine by an independent means the sequences of each oligonucleotide. The M-values for the electrophoretic fractionation at pH 1.9 on DEAE-paper of the partial exonuclease digestion products (Table I) all fall within the ranges set by others for this type of analysis[9,14,25,26]. Because each oligonucleotide was labeled only at its 5'-terminus, extraneous breaks did not produce any extra spots on the radioautographs. This meant that the M-value calculated between two adjacent spots on the radioautograph accurately identified the 3'-terminal residue by which the two partial-digestion products differed. In analyses not shown, we also fractionated some partial exonuclease digestion products on DEAE-paper at pH 3.5 to further confirm the sequences of the smallest 5'-labeled oligonucleotides. From the M-values and absolute mobilities in the two electrophoresis buffers, we have confirmed the sequence of the first five nucleotides at the 5'-ends of the EcoRI site[12].

To make certain that the two 5'-sequences were properly assigned, we used the two-dimensional fractionation system to identify the snake venom phosphodiesterase partial digestion products of selected oligonucleotides. The homochromatograms of the partial exonuclease digestion products of pA-A-T-T-C-C-T-T-T and pA-A-T-T-C-T-C-G-C-C (Figs. 2a and 2b, respectively) are characteristic of the two sequences.

Col El plasmid DNA was used as substrate for EcoRI endonuclease in a parallel analysis to test for structural features at the EcoRI site that might be significant but overlooked with use of only one substrate DNA. Col El DNA is also a closed-circular molecule with a single, unique EcoRI site. Nucleotide sequence analysis at the EcoRI site of Col El plasmid DNA was done exactly as described above for the analysis of SV40 DNA. A representative fingerprint is shown in Fig. 3 and the M-values of the partial

TABLE I

A. SV40 DNA

1. Undecanucleotide

Oligonucleotide	M-value
AAT	-
AATT	2.10 ± 0.04 (5)
AATTC	0.06 ± 0.01 (7)
AATTCT	2.10 ± 0.11 (8)
AATTCTC	0.17 ± 0.06 (6)
AATTCTCG	1.60 ± 0.30 (6)
AATTCTCGC	0.24 ± 0.04 (6)
AATTCTCGCC	0.22 ± 0.03 (5)
AATTCTCGCCA	0.70 ± 0.20 (2)

2. Decanucleotide

Oligonucleotide	M-value
AAT	-
AATT	2.10 ± 0.04 (5)
AATTC	0.06 ± 0.01 (7)
AATTCC	0.08 ± 0.02 (7)
AATTCCT	2.30 ± 0.18 (6)
AATTCCTT	2.10 ± 0.20 (6)
AATTCCTTT	2.20 ± 0.16 (5)
AATTCCTTTG	1.60 ± 0.20 (2)

B. Col E1 Plasmid DNA

1. Undecanucleotide

Oligonucleotide	M-value
AAT	-
AATT	2.20 ± 0.07 (3)
AATTC	0.07 ± 0.01 (4)
AATTCT	2.10 ± 0.10 (4)
AATTCTC	0.20 ± 0.05 (4)
AATTCTCT	2.00 ± 0.40 (4)
AATTCTCTG	1.60 ± 0.20 (4)
AATTCTCTGC	0.30 - (2)
AATTCTCTGCT	2.00 - (2)

2. Octanucleotide

Oligonucleotide	M-value
AAT	-
AATT	2.20 ± 0.07 (3)
AATTC	0.07 ± 0.01 (4)
AATTCC	0.08 ± 0.01 (4)
AATTCCT	2.20 ± 0.08 (4)
AATTCCTG	1.80 - (1)

M-values for the oligonucleotides from the 5'-ends of EcoRI-cleaved SV40 DNA (A) and Col E1 plasmid DNA (B). The M-values shown were measured on the radioautographs of the electrophoretograms (DEAE-paper, pH 1.9) of the products of the partial exonucleolytic degradation of oligonucleotides eluted from homochromatography thin-layers (Figs. 1 and 3). They characterise the 3'-terminal residues which differentiate one oligonucleotide from the next smallest molecule in each sequence. The partial degradation products, themselves, were used to assign oligomucleotides to sequences and to directly sequence each oligomer by its individual M-values. The M-values for the partial degradation products of each oligonucleotide were identical to those tabulated above. In each section of the table, the first three entries correspond to components of the EcoRI site. The identity of AAT was verified by its electrophoretic mobilities on DEAE-paper at pH 1.9 and pH 3.5 and by its degradation products. Each molecule shown was labeled with ^{32}P-phosphate at its 5'-terminus. The symbols for all phosphates have been omitted. The M-values are listed as (Mean M-value) ± (Mean Deviation from the Mean). The figures in parentheses in the table are the numbers of observations not counting those made from the exonuclease degradation products of each oligonucleotide.

exonuclease digestion products are listed in Table IB. Electrophoresis of the snake venom phosphodiesterase digestion products of the 5'-labeled oligonucleotides was only done at pH 1.9, since the smallest of them migrated identically to the analogous SV40 DNA products.

We were able to determine the sequences of eleven nucleotides on one strand of SV40 DNA and ten nucleotides on the other by the methods presented above. By overlapping the nucleotides of the unique, symmetrical EcoRI substrate sequence, we arrived at a double-stranded structure of 17 base pairs (upper sequence, Fig. 4). Similarly, overlap of an undecamer and an octamer at the EcoRI site of Col E1 plasmid DNA led to a 15-base-pair structure (lower sequence, Fig. 4).

The analyses of the other three DNAs (bacteriophages λ and φ 80 and M. lysodeikticus) were complicated by the sensitivity of our approach to extraneous breaks in the DNA preparations. The phage DNAs and the

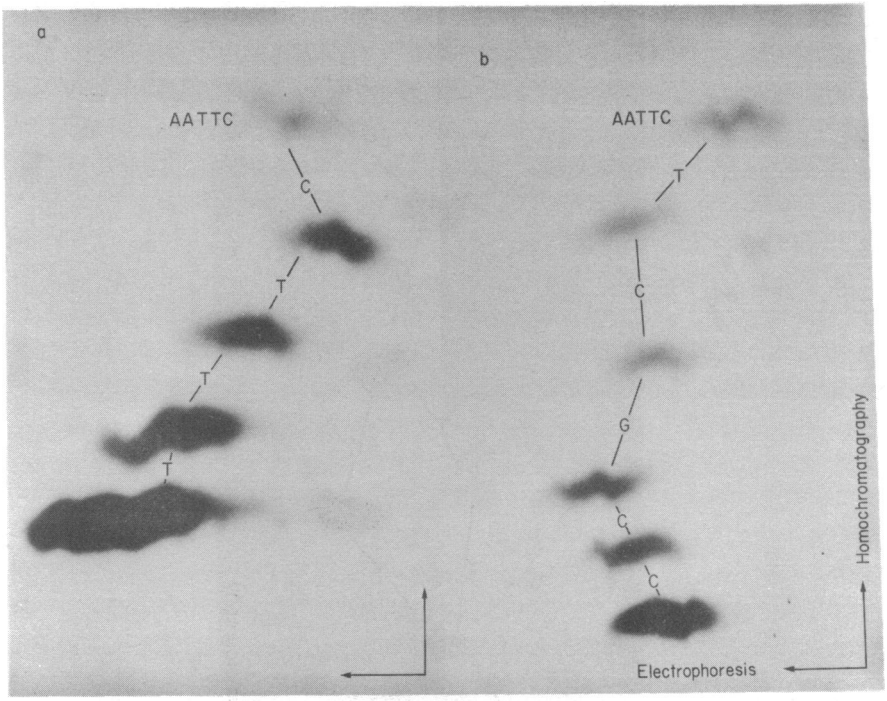


Fig. 2. Two dimensional fractionation of the products of the partial degradation with snake venom phosphodiesterase of pA-A-T-T-C-C-T-T-T (a) and pA-A-T-T-C-T-C-G-C-C (b). The stepwise reconstruction of the two sequences from the exonuclease digestion products is illustrated. The elongated shapes of the spots resulted from an overloading of the cellulose acetate. The composite pattern of these two fingerprints is identical to the corresponding portion of the fingerprint shown in Fig. 1.

commercial micrococcal DNA were all badly nicked during purification. Dephosphorylation and rephosphorylation led to the incorporation of ^{32}P-phosphate groups into the 5'-termini of these nicks resulting in the appearance of extraneous oligonucleotides on the homochromatograms. The major spots on the patterns, nevertheless, corresponded to the EcoRI sequences. By analyzing a number of them, we were able to obtain useful sequence information. The five EcoRI sites on phage λ DNA[11] yield the 5'-hexamers pA-A-T-T-C-A, pA-A-T-T-C-T, and pA-A-T-T-C-G. The first two of these are those previously detected[12]. For technical reasons it is difficult to quantitate the third sequence with respect to the other two. It probably does not occur in more than one or two of the ten strands at the five EcoRI sites. This low level of occurrence would have been missed with the techniques used in the previous analysis[12]. The 5'-hexanucleotides

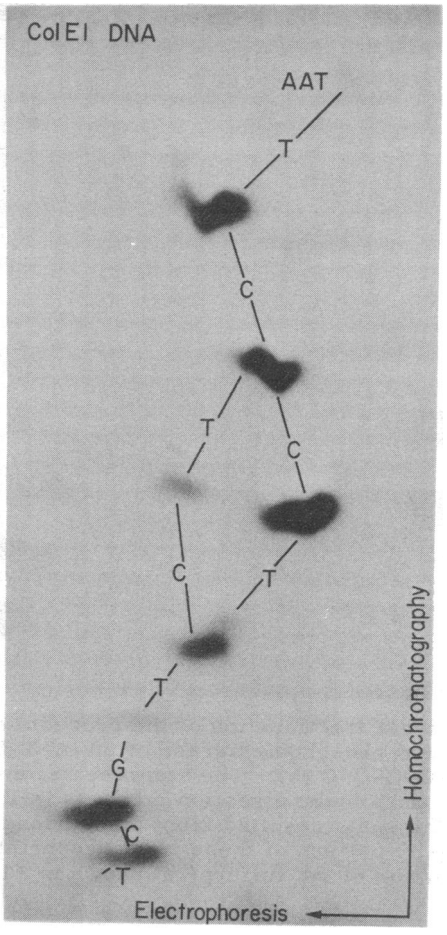


Fig. 3. Pancreatic DNase fingerprint of [5'-^{32}P]labeled, EcoRI-cleaved Col E1 plasmid DNA. Details of the fingerprinting are the same as for Fig. 1. pA-A-T-T-C-C-T-G is missing from this fingerprint.

at the eight EcoRI splits in phage ϕ80 DNA[11] are pA-A-T-T-C-A, pA-A-T-T-C-C, and pA-A-T-T-C-T, while the micrococcal DNA gives evidence for all four nucleotides in the sixth position: pA-A-T-T-C-N. Our general conclusion is that the EcoRI substrate site is the symmetrical, double-stranded equivalent of the hexanucleotide -G-A-A-T-T-C-.

DISCUSSION

Our analyses demonstrate that the sequence specificity of the EcoRI restriction endonuclease, and, presumably, the modification methylase, does not extend past the symmetrical, central hexamer in the sequence initially determined by Hedgpeth et al.[12]. This conclusion rests on the observa-

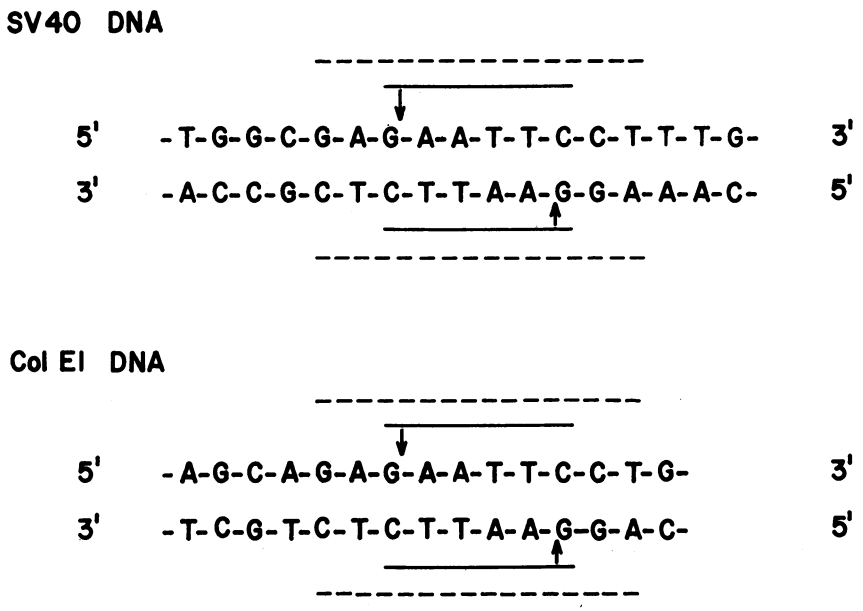


Fig. 4. Sequences spanning the EcoRI sites of SV40 DNA (upper) and Col E1 plasmid DNA (lower). The two sequences are aligned at the EcoRI sites which are indicated by the solid lines. The dashed lines indicate a region of unexplained homology between the two sequences. The arrows show the cleavage points of the EcoRI restriction endonuclease.

tions that we can find all four nucleotides in the sixth position from the 5'-terminus of the EcoRI break-point in one or another of the substrate DNAs tested and that there does not appear to be any prominent structural feature outside this hexamer. The failure of the previous investigation[12] to detect G-C nucleotide pairs flanking the EcoRI site is a consequence of both the choice of λ DNA as substrate and the techniques used in that analysis. The EcoRI sites in SV40 and Col E1 DNAs (Fig. 4) have identical sequences of ten base pairs spanning the site. This common decameric sequence is not symmetric outside the central hexamer and its occurrence is probably just a coincidence.

Recently, Greene et al. (J. Mol. Biol., in press) have demonstrated with a synthetic oligonucleotide that the EcoRI enzyme will act on a double-stranded molecule containing only two additional base pairs beyond the EcoRI hexamer. In addition, Polisky et al. (Proc. Nat. Acad. Sci. USA, in press) have shown that by altering digestion conditions, the substrate specificity of the EcoRI endonuclease can be reduced to the central tetranucleotide of the canonical sequence. These observations, together with those of

Dugaiczyk et al.[13] and those presented here, argue strongly that the restriction, modification, and recognition sites of the EcoRI system are one and the same and consist of the symmetrical, double-stranded equivalent of the sequence -G-A-A-T-T-C-. The finding that the different EcoRI sites on bacteriophage λDNA are cleaved at varying rates[31] suggests that sequences outside the hexamer may influence the rate of cleavage at a particular site.

The method of analysis we have used provides a rapid and simple procedure for the determination of the nucleotide sequences of restriction sites in particular and the termini of DNA molecules in general. The major limitation in the technique as we have used it is in the electrophoretic fractionation of the partial exonuclease products for verification of sequences. Even with extended electrophoretic runs, oligonucleotides much longer than the ones we have studied do not move far enough from the origin of the DEAE-paper for accurate determination of mobilities. Homochromatographic fractionation has to be used to detect the splits made in large molecules and this increases both the amount of labor and the time required for the analysis. The range of the primary fingerprinting can, however, be easily increased to longer oligonucleotides by use of short DNase digestion times and alternative homomixtures. For example, through the use of homomixtures c (3%) made with RNA hydrolyzed for 10 and 15 minutes we have been able to make the next size class of oligonucleotides available for analysis.

ACKNOWLEDGMENTS

We thank F. M. DeNoto for preparing the plasmid DNA and K. Austin for helping with the figures. This work was supported by U.S. Public Health Service Grants AI 00299, GM 14378, and CA 14026.

REFERENCES

1. Boyer, H. W. (1971) Ann. Rev. Microbiol. 25, 153-176.
2. Boyer, H. W. (1974) Fed. Proc. 33, 1125-1127.
3. Subramanian, K. N., Pan, J., Zain, S., and Weissman, S. M. (1974) Nucleic Acids Research 1, 727-752.
4. Zain, B. S., Weissman, S. M., Dhar, R., and Pan, J. (1974) Nucleic Acids Research 1, 577-594.
5. Maniatis, T., Ptashne, M., Barrell, B. G., and Donelson, J. (1974) Nature 250, 394-397.
6. Maniatis, T., Jeffrey, A., and Kleid, D. G. (1975) Proc. Nat. Acad. Sci. USA 72, 1184-1188.

7. Morrow, J. F., Cohen, S. N., Chang, A. C. Y., Boyer, H.W., Goodman, H. M., and Helling, R. B. (1974) Proc. Nat. Acad. Sci. USA 71, 1743-1747.
8. Hershfield, V., Boyer, H. W., Yanofsky, C., Lovett, M. A., and Helinski, D. R. (1974) Proc. Nat. Acad. Sci. USA 71, 3455-3459.
9. Murray, K. and Old, R. W. (1974) Prog. Nucleic Acid Res. and Mol. Biol. 14, 117-185.
10. Salser, W. A. (1974) Ann. Rev. Biochem. 43, 923-965.
11. Greene, P. J., Betlach, M. C., Boyer, H. W., and Goodman, H. M. (1974) in Methods in Molecular Biology, Vol. 7, (R. B. Wickner, ed.) pp. 87-111, Marcel Dekker, New York.
12. Hedgpeth, J., Goodman, H. M., and Boyer, H. W. (1972) Proc. Nat. Acad. Sci. USA 69, 3448-3452.
13. Dugaiczyk, A., Hedgpeth, J., Boyer, H. W., and Goodman, H. M. (1974) Biochemistry 13, 503-512.
14. Murray, K. (1973) Biochem. J. 131, 569-583.
15. Mulder, C. and Delius, H. (1972) Proc. Nat. Acad. Sci. USA 69, 3215-3219.
16. Morrow, J. F. and Berg, P. (1972) Proc. Nat. Acad. Sci. USA 69, 3365-3369.
17. Lovett, M. A., Guiney, D. G., and Helinski, D. R. (1974) Proc. Nat. Acad. Sci. USA 71, 3854-3857.
18. Murray, K. and Murray, N. E. (1973) Nature New Biology 243, 134-139.
19. Brownlee, G. G. and Sanger, F. (1969) Eur. J. Biochem. 11, 395-399.
20. Sanger, F., Donelson, J. E., Coulson, A. R., Kossel, H., and Fischer, D. (1973) Proc. Nat. Acad. Sci. USA 70, 1209-1213.
21. Guerry, P., LeBlanc, D. J., and Falkow, S. J. (1973) J. Bacteriol. 116, 1064-1066.
22. Glynn, I. M. and Chappell, J. B. (1964) Biochem. J. 90, 147-149.
23. Chamberlin, M. J. and Ring, J. (1972) J. Mol. Biol. 70, 221-237.
24. Garfin, D. E. and Goodman, H. M. (1974) Biochem. Biophys. Res. Commun. 59, 108-116.
25. Barrell, B. G. (1971) Proced. Nucleic Acid Res. 2, 751-779.
26. Murray, K. (1970) Biochem. J. 118, 831-841.
27. Ehrlich, S. D., Bertazzoni, U., and Bernardi, G. (1973) Eur. J. Biochem. 40, 143-147.
28. Ling, V. (1972) J. Mol. Biol. 64, 87-102; Proc. Nat. Acad. Sci. USA 69, 742-746.
29. Rensing, U. F. E. and Schoenmakers, J. G. G. (1973) Eur. J. Biochem. 33, 8-18.
30. Jay, E., Bambara, R., Padmanabhan, R., and Wu, R. (1974) Nucleic Acids Research 1, 331-353.
31. Thomas, M. and Davis, R. W. (1975) J. Mol. Biol. 91, 315-328.