DNA Site Recognition and Reduced Specificity of the Eco RI Endonuclease*

(Received for publication, November 27, 1979, and in revised form, July 24, 1980)

Charles P. Woodbury, Jr., Otto Hagenbüchle, and Peter H. von Hippel
From the Institute of Molecular Biology and Department of Chemistry, University of Oregon, Eugene, Oregon 97403

It has been shown previously (Polisky, B., Green, P., Garfin, D. E., McCarthy, B. J., Goodman, H. M., Boyer, H. W. (1975) Proc. Natl. Acad. Sci. U. S. A. 72, 3310-3314; Hsu, M., and Berg, P. (1978) Biochemistry 17, 131-138) that the cleavage sequence specificity of Eco RI endonuclease can be "relaxed" by various means. In this paper this phenomenon is explored in detail, in order to obtain further insight into the nature and selectivity of sequence recognition patterns between proteins and double-stranded nucleic acids. Using conditions of low ionic strength and alkaline pH, we have mapped the positions of potentially cleavable sites in the (completely sequenced) replicative form of the bacteriophage φX174 genome, and have deduced their sequence. The time course of digestion of φX174 DNA suggests that double-stranded sequences reading GAATT, GAAATT, GAAATTT, and GAATTA (only "top" strands, written 5' → 3', are shown) are cleaved readily under these conditions, while sequences reading CAATTN (N = A, T, G) resist attack. Cleavages at (at least) the more labile sites result in cohesive ends that are religatatable. End group analysis of cleaved φX174 DNA fragments indicates the presence of a 5'-terminal adenine residue on most of the fragments; some fragments may carry a 5'-terminal guanine residue, consistent with the cleavage site sequences suggested above. Addition of Mn++ to cleavage reactions carried out at moderate salt concentrations and near-neutral pH induces the same pattern of cleavage seen at low ionic strength and alkaline pH. These results are combined with those from other studies, and are interpreted in terms of a model for the site-specific interaction of the Eco RI endonuclease with its substrate, considering both the effects of changes in DNA sequence and of environmental alterations. The resulting model is compared with data developed on similar grounds for Eco RI methylase (see Woodbury, C. P., Downey, R. L., and von Hippel, P. H. (1980) J. Biol. Chem. 255, 11526-11533), and attempts are made to define both common and differing molecular facets of the DNA recognition specificity of this companion (but genetically distinct) enzymes.

In the preceding paper (Woodbury et al., 1980), we examined the relaxation in the specificity of methylation by Eco RI methylase induced by various alterations in the solvent environment. Our results were consistent with a compensation of the loss of certain specific favorable (presumably hydrogen-bonding) recognition contacts by an increase in the strength of nonspecific electrostatic interactions, resulting in methylation at sites of less than optimal base pair sequence. In this paper we report parallel studies with the Eco RI endonuclease, focusing on the details of the sequences of the reduced specificity sites and stressing particularly the effects of base replacements and extended sequence symmetries in and around these sites.

MATERIALS AND METHODS

Reagents and Buffers—Most of the reagents and buffers used have been described previously (Woodbury et al., 1980).[^1] (as carrier-free orthophosphate) was purchased from New England Nuclear. ATP was purchased from P-L Biochemicals. [γ-32P]ATP was synthesized as described by Maxam and Gilbert (1977). [γ-32P]ATP was purchased from New England Nuclear (specific activity > 300 Ci/mmol) for use in "nick translation" of DNA.

Standard electrophoresis buffers were TBE buffer (89 mM Tris/borate, 2.5 mM EDTA, pH 8.5) and E buffer (40 mM Tris, 245 mM sodium acetate, 1 mM EDTA, pH 7.8). Enzymes and Proteins—Polynucleotide kinase (from phage T4 XF-l-infected Escherichia coli) was obtained from P-L Biochemicals. Bacterial alkaline phosphate was the generous gift of J. Chlebowski. Bovine pancreatic DNase and venom exonuclease were from Worthington.

Restriction endonucleases HindIII, Hae III, Pst I, and Xho I were purchased from New England BioLabs. Restriction endonuclease Eco RI was prepared by the method of Modrich and Zabel (1976) from E. coli strain RY13 cell paste (purchased from the Microbiological Research Establishment, Porton, England). Final purity of the enzyme, as judged by SDS-polyacrylamide gel electrophoresis,[^1] was > 98%, with a specific activity in excess of 10,000 units/mg (1 unit is defined as the cleavage of 2 pmol of phosphodiester bonds/min).

DNA and Polynucleotides—φX174 am3 cs70 (RP1) DNA was purchased from New England BioLabs and stored at 4°C. The preparation of phage λ (cI857 and Ab2) DNA has been described (Woodbury et al., 1980). Ab2 DNA was labeled internally with [γ-32P]ATP using the technique of nick translation as described by Maniatis et al. (1975). 5'-End-32P-labeled φX174 DNA for mapping restriction enzyme cleavage sites was prepared from about 9.4 µg φX174 DNA cleaved by Xho I, ethanol-precipitated, then resuspended in 100 µl of 10 mM Tris, pH 8.0, 1 mM EDTA. Phosphatase treatment and kinase-mediated incorporation of [γ-32P]TP phosphatase was carried out as described by Maxam and Gilbert (1977). A one-third milliliter portion of ethanol-precipitated, end-labeled φX174 DNA was resuspended in 6 mM Tris, pH 7.4, 50 mM NaCl, 6 mM MgCl2, 6 mM 2-mercaptoethanol, and 100 µg/ml of bovine serum albumin, and incubated with 1 unit of Pst I at 37°C for 1 h. The reaction was quenched by addition of SDS to 1%, and EDTA to 10 mM levels. To separate the two φX174 DNA fragments generated, the ice-chilled sample was electrophoresed on

[^1]: The abbreviations used are: SDS, sodium dodecyl sulfate; AdoMet, S-adenosyl-L-methionine; ColEl, colicin E1.

11534
a 1% agarose slab gel (0.3 cm x 12 cm x 16 cm) for 3 h at 35 mA using E buffer, and the positions of the labeled DNA bands determined by autoradiography. DNA fragment bands were excised and the DNA eluted in 4M lithium hydroxide buffer. The DNA was phenol-extracted once, ether-extracted three times, concentrated by ethanol precipitation, resuspended in 10 mM Tris, pH 8.0, 1 mM EDTA, and stored at 4°C.

Eco RI nuclease digests of whole or 32P-labeled fragments of φX174 DNA under conditions of reduced Eco RI nuclease specificity were carried out with 1 to 5 µg of φX174 DNA in 20 mM Tris, pH 8.0, 2 mM MgCl2 at 37°C, and with varying amounts of enzyme. Reactions were quenched with SDS and EDTA as described above, the DNA was ethanol-precipitated and resuspended in 10-fold dilute TBE buffer containing the tracking dyes xylene cyanol FF (0.025%; Matheson, Coleman, and Bell) and bromphenol blue (0.025%; Matheson, Coleman, and Bell) plus glycerol to 5% (w/v). Samples of Hpa II- and Hae III-digested φX174 DNA, and of Ab2 DNA digested with Eco RI and HindIII, were prepared similarly. Electrophoresis of the digests on agarose or polyacrylamide slab gels was performed at room temperature at 30 mA for approximately 3½ h. Positions of the bands were determined by autoradiography using standard methods. The developed films were examined in some cases with a Grant Instruments scanning comparator-microphotometer. For determination of 5' terminal residual cleavage, 13 µg of φX174 DNA was digested at 37°C for 24 h with 50 units of Eco RI endonuclease in 125 µl of the reduced specificity buffer described above. After phenol and ether extractions, followed by ethanolic precipitation, a 0.6-volume aliquot of the resuspended DNA (approximately 8 µg) was phosphorylated with 32P at the 5'-end by the method of Maxam and Gilbert (1980). A 0.1-volume aliquot of the ethanol-precipitated, resuspended DNA (in TE buffer) was electrophoresed on a 1.5% agarose gel slab; six distinct DNA fragment bands were recovered and purified as described above. Following ethanol precipitation, the samples were digested thoroughly with bovine pancreatic DNase and venom exonuclease and the resulting mononucleotide phosphates were separated by paper electrophoresis (Brownlee, 1972). After autoradiography, nucleotide assignments were made by comparison to the relative mobility of co-electrophoresed standards of dAMP, dCMP, dGMP, and dTMP. Relative incorporation of 32P into the four nucleotides was determined by excising the respective spots on the paper and counting by liquid scintillation (Woodbury et al., 1980).

"Overmethylated" φX174 (RF) DNA, for the demonstration of blockage of noncanonical attack by Eco RI endonuclease, was prepared by incubating 8 µg of DNA with 4 µg of Eco RI methylase in 200 µl of 25 mM Tris, pH 9.0, 2 mM EDTA, 2.5 mM dithiothreitol, 4 µM [3H]AdoMet, and 400 µg/ml of bovine serum albumin for 45 h at 37°C, leading to an average incorporation per +X174 molecule of 1250 dpm/µg. Following extraction with phenol and ether, the DNA was concentrated by precipitation with ethanol and resuspended in 4M lithium hydroxide buffer. Two micrograms of the alcohol-precipitated, resuspended DNA were then digested for 1 or 4 h at 37°C with 50 units of Eco RI endonuclease in 100 µl of 20 mM Tris, pH 8.0, 2 mM MgCl2, and prepared for electrophoresis as described above.

RESULTS

Previously, Polisky et al. (1975) had demonstrated that under low salt and high pH conditions Eco RI endonuclease shows reduced site specificity; i.e. it can cleave many double-stranded DNA sequences in addition to the canonical double-stranded Eco RI-specific G^AATTG (only top strand of sequence shown) site. This reduced specificity activity, termed Eco RI* by these workers, was reported to be specific for sequences reading N^AAATTG. Here we examine this reduced specificity in more detail.

The double-stranded (RF) DNA of bacteriophage φX174 provides a very convenient substrate for these studies. The entire nucleotide sequence of φX174 has been determined (Sanger et al., 1977; revised, 1978), and has been shown to contain no canonical Eco RI recognition sites. The φX174 double-stranded genome does, however, contain a large number of sequences which are more or less closely related to the canonical site, including 25 Eco RI* sequences (defined as NAAATTG; Polisky et al., 1975).

We planned to determine the preferred sites of cleavage by generating a series of partial digests of φX174 (RF) DNA, under conditions that stimulate the reduction of site specificity of the endonuclease. By mapping these cleavage sites on φX174, we could then compare the sequence about these preferred sites to sequences digested subsequently (less favorable sites), and perhaps to sites where cleavage occurs only very slowly or not at all. Previous workers (Goodman et al., 1977) have shown that there seems to be a hierarchy of preference in cleavage sites for the Eco RI* activity, and have suggested, on the basis of relative nucleotide frequencies, that sequences reading GAATTC might be preferred over sequences reading GAATTG, which in turn seemed to be preferred over all other NAAATTG* sequences. Using φX174 (RF) DNA we could examine this hypothesis further, as well as investigate cleavage sites for other features of interest, including those which might lie outside the hexameric canonical sequence region.

Mapping Cleavage Sites in φX174 DNA—We followed two procedures in the preparation of partially digested φX174; the first was suggested by the technique developed by Smith and Birnstiel (1976) for mapping cleavage sites of restriction endonucleases. This technique uses a series of limited digests of 32P-terminally labeled DNA molecule to develop an overlapping set of partially digested molecules with a common 32P-base at the 3'-end of the molecule. The resulting maps for various times of digestion, together with our map of highly labile cleavage sites developed from the first approach, provided a means to deduce the positions of the more refractory sites.

To prepare φX174 DNA with a 32P-terminus for use in the Smith-Birnstiel technique, the DNA was first digested with restriction endonuclease Xho I. This enzyme cuts at a single site on the φX174 genome and produces protruding 5'-termini, suitable for label-alignment labeling. Following 32P-label incorporation at these termini (as detailed under "Materials and Methods"), we digested the labeled DNA with restriction endonuclease Pst I, which cleaves φX174 DNA at a single site. This resulted in two terminally labeled fragments, a short and a very long (5226 base pairs) fragment. These were easily resolved by electrophoresis. Following purification of the large fragment, we incubated it for varying (short) periods with sufficient Eco RI endonuclease under conditions designed to stimulate the reduced specificity activity of the enzyme (20 mM Tris, pH 8.0, 2 mM MgCl2), electrophoresed the resulting partial digests, and examined the DNA band pattern of the resulting autoradiogram. To deduce the size of the various fragments represented by these bands, we co-electrophoresed a sample of Ab2 DNA, labeled with 32P-containing nucleotides via the nick translation procedure and doubly digested with both Eco RI and HindIII. The sizes of the resulting fragments from this cross-digestion of Ab2 are well known (Robinson and Landy, 1977), and provided a calibration standard for determining the sizes of the φX174 DNA fragments.

Fig. 1 is an autoradiogram of a series of limited digests of the large 32P-labeled φX174 fragment. Scanning the autoradiogram by densitometer allowed us to determine quite precisely the relative mobilities of the various DNA fragments; a
plot of the relative mobilities of the \( \lambda b2 \) fragments versus the logarithm of their molecular weight was linear over the range from 520 base pairs to 5700 base pairs, and provided an accurate calibration curve for determining the sizes of the \( \phi X174 \) DNA fragments (data not shown). Densitometer scans of the gel band patterns of the partially digested large \( \phi X174 \) fragment showed several prominent bands, as well as several bands of lower intensity; we suggest that the prominent bands represent fragments resulting from cleavages at very labile sites, and that the less intense bands represent fragments from subsequent cleavages at less favorable sites.

In Fig. 1 we see five intense bands, resulting from cleavage at five very labile sites in the large \( \phi X174 \) fragment. The intensity of four of these bands decreases with time, while that of the fifth (and lowest in molecular weight) increases.

![Figure 1](image1.png)

**Fig. 1.** Digestion of \( ^{32}P \)-labeled \( \phi X174 \) (RF) DNA with Eco RI endonuclease under low site specificity conditions. Samples were prepared as described under "Materials and Methods," then layered on a 12% agarose slab gel and electrophoresed at 30 mA (constant current) for 3 h, using TBE as the running buffer. Track A, 5-min digestion; B, 30-min digestion; C, 60-min digestion; D, \( ^{32}P \)-labeled \( \lambda b2 \) DNA, digested with Eco RI and HindIII.

The fifth band thus represents the fragment resulting from cleavage at the site nearest the \( ^{32}P \)-labeled \( Xho I \) cleavage site. Moreover, judging from the intensity of the low molecular weight band in very limited digests (Tracks A and B of Fig. 1) this site represents the most favorable cleavage locus in the entire \( \phi X174 \) fragment.

The sizes of four of the five "early" fragments are easily established by reference to the HindIII/Eco RI-digested \( \lambda b2 \) DNA fragment series (Fig. 1). The apparent sizes (in base pairs) of these early fragments are 1100 \( \pm 100 \), 1950 \( \pm 200 \), 3250 \( \pm 300 \), and 3600 \( \pm 300 \). The smallest of these five fragments lies outside the calibration range of the \( \lambda b2 \) DNA fragments, but linear extrapolation of the calibration assigns to it a size of 310 \( \pm 30 \) base pairs. (Fragment separation in gels of this type follows the logarithmic dependence of fragment size to below 140 base pairs.) Based on these fragment sizes we have assigned provisionally the points of cleavage to (approximate) positions 470, 1300, 2100, 3400, and 3800 (relative to the \( Pst I \) cleavage site) in the \( \phi X174 \) sequence (see Fig. 3 below).

The Smith-Birnstiel technique thus allowed us to determine the positions of a total of eight cleavage sites on the large \( \phi X174 \) fragment. As pointed out above, this approach could not be applied to the mapping of very refractory cleavage sites, and so we supplemented this technique by examining the fragment patterns in prolonged digests of unlabeled whole \( \phi X174 \) (RF) DNA. Fig. 2, A and B, are photographs of typical slab gels of \( \phi X174 \) (RF) DNA digested with varying amounts of enzyme and for various times, to generate a spectrum of partial digestion products. Sizes of the various fragments have been determined by comparison to the mobilities of \( Hpa II \) and \( Hae III \)-generated fragments of \( \phi X174 \) (Sanger et al., 1977).

As can be seen from these figures, attack on \( \phi X174 \) DNA by Eco RI endonuclease occurs first at a limited number of sites, generating only a few very large fragments. As digestion proceeds, these large fragments are cleaved into a host of smaller ones; transient species of intermediate size appear and disappear, with a general shift in intensity of the bands to fragments of lower molecular weight. In the most extended digestion achieved (Track 2 of Fig. 2B), we can resolve fragments of approximately 145, 150, 200, 235, 275, 290, 310,
470, 575, 690 to 715 (multiplet band), 850, 890, and 950 base pairs. Not all of these fragments can represent limit digests (i.e., complete cleavage at all available sites), since the sum of the base pairs of all of the observed fragments in this sample gives 7040 base pairs, exceeding the 5386 base pairs of the whole φX174 DNA (RF) molecule. In fact, certain of these bands can be identified as partially digested intermediates. For example, the progressive loss in intensity for the bands corresponding to 310, 575, and 890 base pairs indicates that they are most likely intermediate species containing one or more potential cleavage sites. Subtracting the contribution of these putative intermediates from the previous sum (7040 base pairs), we obtain 5275 base pairs, a value comparable to the 5386 base pairs expected for whole φX174.

The various partially digested fragments observed can be utilized to formulate a hierarchy of cleavage site labilities. For example, using the Smith-Birnstiel technique we found that cleavage occurs rapidly at roughly positions 470, 1300, 2100, 3400, and 3800 on φX174. Thus, a spectrum of fragment sizes corresponding to cleavages at these sites might be expected to appear in very limited digests of whole φX174. In particular, if cleavage of φX174 occurred only at the above five sites, a complete digest should generate fragments of approximately 400, 800, 830, 1300, and 2100 base pairs, and a partial digest would generate fragments of roughly 1600, 1700, 2100, 2500, 2900, 3300, 3700, and 3800 base pairs. Examination of Fig. 2A, Track c, reveals bands corresponding to fragments of roughly 570, 900, 1150, 1500, 1800, 2100, 2500, 2900, and 3900 base pairs (as well as a number of other bands), in reasonable agreement with the predicted pattern. Track b of that figure shows faint bands corresponding to fragments of approximately 570, 900, 1700, and 2500 base pairs, also consonant with early attack at the five labile sites. A consistent pattern of cleavage thus emerges, with cleavages occurring first at the five sites mapped out by the Smith-Birnstiel technique, followed by subsequent cleavage of the large precursor fragments so generated. A cleavage map summarizing the assignment of the most readily cleaved noncanonical sites, as well as the sequence-numbering system, is shown in Fig. 3.

**Effects of Divalent Ions**—Hsu and Berg (1978) have shown that Mn\(^{++}\) stimulates the reduced specificity activity of the Eco RI endonuclease, and Goodman et al. (1977) have reported that Mn\(^{++}\) (at concentrations of 100 to 150 μM) can replace Mg\(^{++}\) (2 mM) in the Eco RI* reaction. We ask if the Mn\(^{++}\)-stimulated activity follows the same pattern of attack on φX174 as that found under conditions of alkaline pH and low ionic strength. As before, we used the Smith-Birnstiel technique to compare partial digests in 20 mM Tris, pH 8.0, 2 mM MgCl\(_2\) (partial digests in 100 mM Tris, pH 7.5, 2 mM MnCl\(_2\)). Fig. 4 is an autoradiogram of the resulting slab gel band pattern. In the main, the pattern of attack is the same for both sets of conditions. The low salt, pH 8.0, reaction yields (in a brief digest) fragments of roughly 300, 1070, 2000, 3400, and 3800 base pairs (Track e, Fig. 4), in good agreement with the pattern seen before (cf. Fig. 1). The reaction carried out in the presence of Mn\(^{++}\), with the same amount of enzyme

---

**Fig. 3.** Cleavage map of highly labile noncanonical Eco RI sites on φX174 (RF) DNA. The relative positions of cleavage sites are drawn to scale on a circular map divided into units of 1000 base pairs (inner tick marks); following the convention of Sanger et al. (1977), the Pst I cleavage site is taken as the origin while noncanonical Eco RI cleavage sites (A through E) are numbered according to the position of their respective first base pairs in the φX174 sequence (Sanger et al., 1977; revised, 1978).

**Fig. 4.** Equivalence of Mn\(^{++}\)- and low salt/alkaline pH-stimulated noncanonical digestion of φX174 (RF) DNA by Eco RI endonuclease. Samples were prepared as described under "Materials and Methods," then layered on a 1.2% agarose slab gel and electrophoresed at 50 V (constant voltage) for 16 h, using TBE as the running buffer. (a) λ P2 DNA, digested with Eco RI and HindIII calibration standard; (b) Mn\(^{++}\)-Eco RI digest of φX174, 30 units, 3 min; (c) Mn\(^{++}\)-Eco RI digest of φX174, 50 units, 5 min; (d) Mn\(^{++}\)-Eco RI digest of φX174, 100 units, 60 min; (e) alkaline low salt-Eco RI digest of φX174, 30 units, 3 min; (f) alkaline low salt-Eco RI digest of φX174, 50 units, 5 min; (g) alkaline low salt-Eco RI digest of φX174, 100 units, 60 min.

---

**Effects of Divalent Ions**—Hsu and Berg (1978) have shown that Mn\(^{++}\) stimulates the reduced specificity activity of the Eco RI endonuclease, and Goodman et al. (1977) have reported that Mn\(^{++}\) (at concentrations of 100 to 150 μM) can replace Mg\(^{++}\) (2 mM) in the Eco RI* reaction. We ask if the Mn\(^{++}\)-stimulated activity follows the same pattern of attack on φX174 as that found under conditions of alkaline pH and low ionic strength. As before, we used the Smith-Birnstiel technique to compare partial digests in 20 mM Tris, pH 8.0, 2 mM MgCl\(_2\) (partial digests in 100 mM Tris, pH 7.5, 2 mM MnCl\(_2\)). Fig. 4 is an autoradiogram of the resulting slab gel band pattern. In the main, the pattern of attack is the same for both sets of conditions. The low salt, pH 8.0, reaction yields (in a brief digest) fragments of roughly 300, 1070, 2000, 3400, and 3800 base pairs (Track e, Fig. 4), in good agreement with the pattern seen before (cf. Fig. 1). The reaction carried out in the presence of Mn\(^{++}\), with the same amount of enzyme
On the Specificity of the Eco RI Endonuclease

and for the same period of time (Track b, Fig. 4), shows the same high molecular weight bands, but at reduced intensity and with many additional (some very faint) bands at intermediate positions, corresponding to fragments of roughly 410, 440, 540, 710, 850, 1050, 1600, 1700, 1900, and 2000 base pairs. These same bands also appear in the digests carried out in the absence of Mn\(^{2+}\), but seem to require more time for their appearance. The approximate positions of the cleavage sites of these fragments are roughly 570, 600, 700, 870, 1000, 1200, 1760, 1860, 2100, and 2200 on the φX174 map relative to the PstI cleavage site. A most striking feature of Fig. 4 is the intensity of the band corresponding to the 300-base pair fragment, which strongly suggests preferential cleavage at about position 470 (Position A of Fig. 3) early in all the Mn\(^{2+}\)-stimulated reactions. The early appearance of the many partial digest bands in the Mn\(^{2+}\) incubation indicates a general stimulation of activity toward the less labile sites, as well. In separate experiments (not shown), we surveyed transition metal ions other than Mn\(^{2+}\) for their possible stimulation of any reduced specificity activity for the Eco RI endonuclease; we were unable to show any effect with 2 mM Co\(^{2+}\), Zn\(^{2+}\), or Cu\(^{2+}\) dissolved in 20 mM to 100 mM Tris buffers, at pH 7.5 or pH 8. It was observed previously (Hedgpeth et al., 1972) that Cu\(^{2+}\) and Zn\(^{2+}\), with similar (negative) results.

Nature of DNA Fragment Termini Generated in Noncanonical Site Cleavage—In their work on the Eco RI* activity, Polisky et al. (1975) found that double-stranded SV40 or pM89 DNAs were cleaved in a staggered fashion to generate cohesive termini. Such cohesive ends could be rejoined by E. coli ligase, and could serve as template-primers for the DNA polymerase from avian sarcoma virus (which has a strict requirement for internal 3'-ends and protruding 5'-ends). The pattern of attack was found to match that seen for the canonical Eco RI activity (Hedgpeth et al., 1972). As a check on our procedures, we have used the same approach to examine our φX174 (RF) DNA digests. An extensively digested sample of φX174 (RF) DNA was tested for religation capacity by treating it for 72 h at 4°C with Escherichia coli ligase under the conditions defined by Modrich et al. (1973). Agarose gel electrophoresis of ligase-treated and untreated digests showed that a substantial portion of the fragments present could be rejoined covalently (data not shown). Similar experiments with ColE1 plasmid DNA gave parallel results. Moreover, samples of either DNA that had been digested in the presence of Mn\(^{2+}\) (2 mM MnCl\(_2\), 100 mM Tris, pH 7.5) were just as easily ligated as those generated under conditions of low ionic strength and alkaline pH (data not shown), in agreement with the findings of Hsu and Berg (1978).

Polisky et al. (1975) reported that attack by the Eco RI* activity yields primarily adenine residues at the 5'-terminus; we wished to determine whether the termini generated by our enzyme preparation followed this pattern. Using a 32P-terminally labeled preparation of φX174 DNA that had been digested previously with Eco RI endonuclease, we electrophoresed the DNA on an agarose slab gel and selected six distinct fragment bands for further purification and analysis (see “Materials and Methods” for details of the purification and end group analysis).

Table I summarizes the results; the samples are ranked in order of decreasing fragment size with Sample 1 representing a band several thousand base pairs in length, while Sample 6 (the smallest) represents a band less than 200 base pairs in length. (It is possible that a single apparent band may represent two or more poorly resolved fragment species.) We find that the bulk of 32P-label (70% to 95% of total 32P incorporated) is found in pA for every sample examined, in agreement with the results of Polisky et al. (1975). The low level of incorporation into pT, pC, and pG most probably represents labeling at nicks or nonspecific breaks in the DNA induced by the various manipulations preceding the kinase-exchange incorporation of 32P.

An interesting trend toward increasing incorporation into pG is found in Samples 5 and 6. Assuming that incorporation into pT and pC in these samples represents a nonspecific labeling background of about 3% and 7%, respectively, for the experiments, then labeling of pG for these two samples is apparently at about twice the background level (7.2% and 14.8%, respectively). We return to this point under “Discussion.”

Protection of Noncanonical Cleavage Sites by “Overmethylation” with the Eco RI Methytransferase—Berkner and Folk (1978) reported that overmethylation of polyoma DNA by the Eco RI methylase partially protects DNA against attack by the Eco RI* activity of the endonuclease, suggesting that the methylase acts in or near Eco RI* sequences. Protection of the DNA was incomplete, however, because of the rather low level of overmethylation obtained in those experiments (approximately 4 times the expected level of methyl incorporation, based on the single canonical site in polyoma DNA molecules). With the much higher levels of overmethylation we have achieved using the Eco RI methylase and various DNAs (see preceding paper [Woodbury et al., 1980]), we wished to determine whether full protection against the noncanonical Eco RI endonuclease activity could be attained, and if not, which noncanonical endonuclease sites remained unmodified by the Eco RI methylase.

We prepared overmethylated φX174 (RF) DNA (as detailed under “Materials and Methods”) which had, on the average, over 80 methyl groups incorporated per φX174 molecule. Digestion of this DNA under Eco RI* conditions results in very little fragmentation, compared to unmethylated φX174 controls (Fig. 5), confirming that overmethylation confers protection against the noncanonical cleavage activity of Eco RI endonuclease.

In Fig. 5 we can detect eight fragment bands for the overmethylated-digested φX174 DNA. Two of these bands, approximately 4500 and 5900 base pairs in length, contain the bulk of the digested DNA; however, six minor fragment bands, approximately 2000, 1500, 1200, 940, 710, and 300 base pairs in length, are also apparent. We suggest that most of the methylase-treated molecules resist endonuclease attack, and are susceptible at only one or two sites; this resistant population constitutes the two major high molecular weight bands seen in Fig. 5.

A comparison of the cleavage map established by the Smith-Birnstiel technique (Fig. 3) to the fragmentation pattern seen here reveals a surprising degree of agreement. This cleavage
TBE as gel and electrophoresed at DNA, digested with (RF) DNA, calibration standard; refer to as Site F) would generate fragments of about 1500 base pairs, while attack at Sites A and B (Fig. 3), while attack at Site C and a site mapping at around position 5200 in the +X174 sequence (which we shall refer to as Site F) would generate fragments of about 1500 and 3900 base pairs. (Attack at Site F might well have gone undetected by the Smith-Birnstiel technique, for reasons discussed above.) Additionally, cleavages at Sites A and F would yield fragments of roughly 700 and 4700 base pairs. This rather neatly accounts for the two major fragments bands as well as three of the six minor bands seen in Fig. 5 for the "protected" DNA. Furthermore, attack at Sites C and D would account for the observed fragment bands of 1200 and 900 base pairs, while a partially digested fragment of 2100 base pairs might also result.

This scheme, however, does not account completely for all the bands observed in the digest of overmethylated φX174 (RF) DNA. In particular, the fragment of about 300 base pairs cannot be accommodated within the limited set of cleavage sites proposed here. The simplest explanation for this incongruity is the presence of one or more additional sites of cleavage, left unmodified and unprotected by the methylase treatment and not included in the original set of highly labile cleavage sites.

**DISCUSSION**

The following general conclusions can be drawn from these results: (i) there are (at least) five mapped sites in φX174 that are very susceptible to cleavage by the Eco RI endonuclease under conditions of low ionic strength and alkaline pH; (ii) there are a number of more refractory sites whose positions (and sequences) are yet to be deduced; (iii) addition of Mn²⁺ produces the same pattern of cleavage as does low salt and alkaline pH; (iv) cleavage at these sites generates cohesive ends which carry principally 5'-adenine termini (although some may carry 5'-guanine); (v) protection of most, but not all, such sites is afforded by overmethylation with Eco RI methylase, implying a substantial overlap in specificity of the two enzymes in their reduced specificity mode of action. (The lack of complete protection may provide additional insight into differences between the sequence recognition mechanisms of the two enzymes.)

**Possible Contamination by Other Endonucleases**—As in connection with the noncanonical methylation considered in the previous paper (Woodbury et al., 1980), we must ask here whether an additional nuclease, which might be co-purified with Eco RI in our preparative procedure, could be responsible for the noncanonical cleavage activity observed. The following lines of evidence argue against a contaminating enzyme hypothesis: (i) more than 98% of the protein in our preparation migrates (at the expected molecular weight of the endonuclease) as a single band in SDS-polyacrylamide gel electrophoresis experiments; (ii) in common with Hsu and Berg (1978), we point out that any putative contaminant purported to be responsible for noncanonical cleavage must be inactive with Mn²⁺, but active with Mn³⁺ (under standard endonuclease assay conditions), and active with either ion under conditions of low ionic strength and alkaline pH; this represents a most unusual pattern of divalent ion activation; (iii) in common with the canonical cleavage products of Eco RI endonuclease, our enzyme preparation generates (ligatable) DNA fragments with cohesive ends carrying primarily a 5'-terminal adenine residue; (iv) all the noncanonical sites cleaved share most of the canonical site sequence, and not the sequences recognized by most other known endonucleases; (v) most of the noncanonical cleavage sites can be protected by prior overmethylation with the Eco RI methylase; and (vi) a single gene is responsible for both the canonical Eco RI and the Eco RI* activities (Polisky et al., 1975; Goodman et al., 1977). Taken in aggregate, these results argue strongly against any significant contribution from a contaminating nuclease.

**Noncanonical Cleavage Sequences**—In what follows, we define the position of a candidate hexanucleotide recognition sequence by the number of the first base pair of that sequence in the overall φX174 base pair sequence of Sanger et al. (1978) (see Fig. 9). We consider first hexameric, rather than shorter sequences, since the endonuclease is quite specific for such sequences under "standard" conditions; moreover, work on the nicking of noncanonical sites (Bishop, 1979) as well as on noncanonical double strand cleavage (Polisky et al., 1975; Goodman et al., 1977) suggests a requirement for four or five base pairs of the canonical sequence is retained under most conditions. Thus, in considering sequences in φX174 (RF) DNA which might be cleavage candidates, we search first for hexanucleotide blocks that resemble the canonical sequence at four or more positions, and that yield 5'-terminal adenine (or possibly guanine) residues when cut in the staggered fashion characteristic of the canonical specificity.

As described under "Results," five early cleavage sites were mapped to the approximate positions 470, 1300, 2100, 3400, and 3800. An examination of the sequence of φX174 around these positions shows one or more candidates for noncanonical cleavage near each mapped site. There are several possibilities for cleavage sites around the 2100 and the 3800 map positions; these possibilities will be sorted out below. First we consider the candidates for recognition sequences at the other three early cleavage sites. These three sites include the cleavage site where endonuclease activity was so markedly stimulated by

FIG. 5. Protection of φX174 (RF) DNA by overmethylation with Eco RI methylase. Samples were prepared as described under "Materials and Methods," then layered on a 5% polyacrylamide slab gel and electrophoresed at 15 mA (constant current) for 3 h, using TBE as the running buffer. (a) untreated whole φX174 (RF) DNA; (b) Eco RI-digested φX174 (RF) DNA, 50 units, 1 h; (c) Eco RI-digested φX174 (RF) DNA, 50 units, 4 h; (d) Hpa II-digested φX174 (RF) DNA, calibration standard; (e) Hpa II-digested φX174 (RF) DNA, calibration standard; (f) overmethylated φX174 (RF) DNA, digested with Eco RI, 50 units, 4 h; (g) overmethylated φX174 (RF) DNA, digested with Eco RI, 50 units, 1 h; (h) overmethylated φX174 (RF) DNA, undigested.

Possible Contamination by Other Endonucleases—As in connection with the noncanonical methylation considered in the previous paper (Woodbury et al., 1980), we must ask here whether an additional nuclease, which might be co-purified with Eco RI in our preparative procedure, could be responsible for the noncanonical cleavage activity observed. The following lines of evidence argue against a contaminating enzyme hypothesis: (i) more than 98% of the protein in our preparation migrates (at the expected molecular weight of the endonuclease) as a single band in SDS-polyacrylamide gel electrophoresis experiments; (ii) in common with Hsu and Berg (1978), we point out that any putative contaminant purported to be responsible for noncanonical cleavage must be inactive with Mn²⁺, but active with Mn³⁺ (under standard endonuclease assay conditions), and active with either ion under conditions of low ionic strength and alkaline pH; this represents a most unusual pattern of divalent ion activation; (iii) in common with the canonical cleavage products of Eco RI endonuclease, our enzyme preparation generates (ligatable) DNA fragments with cohesive ends carrying primarily a 5'-terminal adenine residue; (iv) all the noncanonical sites cleaved share most of the canonical site sequence, and not the sequences recognized by most other known endonucleases; (v) most of the noncanonical cleavage sites can be protected by prior overmethylation with the Eco RI methylase; and (vi) a single gene is responsible for both the canonical Eco RI and the Eco RI* activities (Polisky et al., 1975; Goodman et al., 1977). Taken in aggregate, these results argue strongly against any significant contribution from a contaminating nuclease.

Noncanonical Cleavage Sequences—In what follows, we define the position of a candidate hexanucleotide recognition sequence by the number of the first base pair of that sequence in the overall φX174 base pair sequence of Sanger et al. (1978) (see Fig. 9). We consider first hexameric, rather than shorter sequences, since the endonuclease is quite specific for such sequences under "standard" conditions; moreover, work on the nicking of noncanonical sites (Bishop, 1979) as well as on noncanonical double strand cleavage (Polisky et al., 1975; Goodman et al., 1977) suggests a requirement for four or five base pairs of the canonical sequence is retained under most conditions. Thus, in considering sequences in φX174 (RF) DNA which might be cleavage candidates, we search first for hexanucleotide blocks that resemble the canonical sequence at four or more positions, and that yield 5'-terminal adenine (or possibly guanine) residues when cut in the staggered fashion characteristic of the canonical specificity.

As described under "Results," five early cleavage sites were mapped to the approximate positions 470, 1300, 2100, 3400, and 3800. An examination of the sequence of φX174 around these positions shows one or more candidates for noncanonical cleavage near each mapped site. There are several possibilities for cleavage sites around the 2100 and the 3800 map positions; these possibilities will be sorted out below. First we consider the candidates for recognition sequences at the other three early cleavage sites. These three sites include the cleavage site where endonuclease activity was so markedly stimulated by
Mn\(^{2+}\) (i.e. position 470), as well as positions 1300 and 3400. For each of these three cleavage sites the sequence, GGATTT, appears to be the only reasonable choice. This sequence is found at positions 470, 1349, and 3369. The agreement between mapped cleavage position and the position of these sequences is quite good, certainly well within the experimental error in map position. Other candidate sequences are too far removed (generally by 300 or more base pairs) to be considered. In addition, the presence of G residues in the "second" position of this sequence is compatible with the observation that G residues seem to occur at the 5'-termini of some noncanonically cleaved phX174 fragments (Table I).

We note that these are the only three sites in the phX174 genome containing this particular hexameric sequence (although there are many other possible cleavage sites). This sequence does match the canonical Eco RI recognition sequence at four (underlined) points: GGATTT. It should also be pointed out that this sequence, unlike the canonical cleavage sequence, is not palindromic (centrisymmetric); in the 5' → 3' direction along the "lower" strand it reads AAATCC.

To distinguish between these candidates we must consider the data from digests of whole phX174 (Fig. 2A). First, consider the candidate sites for cleavage around position 2100, and another five sites around position 3800. To distinguish between these candidates we must consider the data from digests of whole phX174 (Fig. 2A). First, consider the candidate sites for cleavage around position 2100 (Site C; listed in Table II). The five candidate sites fall at positions 1896, 1983, 2222, 2324, and 2330, in the revised phX174 sequence. Facile cleavage at positions 1359 and 3369, in conjunction with a cleavage at position 2222, would result in fragments of 862 and 1148 base pairs, and indeed strong bands appear in Fig. 2A at about 900 ± 100 and 1150 ± 100 base pairs in brief digests. Similar cleavages at positions 2324 or 2330, in conjunction with cleavage at position 1359 and 3369, would yield fragments of around 1050 base pairs; only a faint corresponding band is seen in gels of the partial digests of whole phX174, arguing against early cleavage at these points. Cleavages at positions 1896 or 1983, and at position 1359, would yield fragments of 537 or 624 base pairs, respectively; no such fragment sizes are observed in partial digests, arguing against positions 1896 and 1983 as candidates.

Next, we consider the five candidate sites for cleavage around position 4000 (Site E). Their sequences begin at positions 3860, 3899, 3939, 3962, and 4147 (Table II). We include as a candidate sequence here the one beginning at position 3962, although it has no more than a three-base pair homology with the canonical sequence, because it does share a number of key recognition features which will be elucidated below.) The first of these candidates may be dismissed, since cleavage at that point (3860) and at position 3369 would generate a fragment of ~490 base pairs early in the digestion. Such a fragment is not seen. Early cleavage at position 3899 is argued against by the lack of a 530 base pair fragment, expected for early attack at this point in conjunction with attack at position 3369. A fragment or sets of fragments produced early in the digests of whole phX174 runs around 570 to 590 base pairs; early cleavage at positions 470, 1359, 2222, or (most particularly) 4147, cannot yield a fragment in this size range by subsequent attack at any other candidate site, and so we can eliminate position 4147. We are left with the two remaining possible sites, at positions 3939 and 3962. Because of the very short separation between these sites (23 base pairs), and the experimental errors in determining fragment sizes from slab gel mobilities, we are unable to state definitely which of these sites is preferred. Cleavage at either, in combination with cleavages at nearby sites, would fit the observed gel band patterns. We tend to be biased against the Eco RI site at position 3939 since its sequence reads CAATTT, and other evidence presented here, as well as the work of Goodman et al. (1977), suggests that the Eco RI sequences beginning with cytosine are strongly discriminated against. We cannot, however, exclude this possibility completely.

Temporarily dropping the latter site from consideration, we compare the sequences of the five remaining noncanonical cleavage sites we have defined for points of similarity. Three of these share the sequence, GGATTT; the two other sites read AAATTT (position 2222) and AAATCT (position 3962). All five sites (reading in at least one direction) contain the sequence, AAT, and four of the five sites share the sequence ATTT. The tetrameric sequence, AATT, is also shared by most of the candidates for the cleavage sites deduced by the Smith-Birnstiel techniques to lie at approximately positions 690, 960, and 1960. These candidates read, respectively, GAATTT (position 693), CAATTT (position 956), TAAATT (position 962), AAATTA (position 996), and AAATTT (position 1983). Moreover, the AATT site is shared by a great number of other possible sites, at which cleavage would be consistent with the observed gel band patterns.

There are, however, two special Eco RI sites we must consider; those which contain C (or G) in the apparently favorable first (or sixth) position in the hexameric sequence (e.g. positions 2988 and 3860). Cleavage at either site would yield fragments in either partial or complete digests which are not observed. In this connection, we recall the refractory site at position 4269 (GAATTT) and note the apparently unfavorable terminal guanine residue. A similarly resistant site of the same sequence is found at position 744; facile cleavage at this site should have been readily detected, but was not seen with the Smith-Birnstiel technique. An apparent contradiction to this rule regarding the position of guanine or cytosine residues is the cleavage site or sites at positions 956 to 962. Here the close proximity of two candidate recognition sequences (AATT) for the reduced specificity activity may overcome the inhibiting influence of terminal guanine or initial cytosine residues. Another possibility is that introduction of a single nick in each of the two sites, but on opposite DNA strands, would generate a full double strand cut in this fashion:

\[
\text{CAATTTT|AATTG}
\]
\[
\text{GTTAA|AATTAC}
\]

(note the apparent reversed polarity in strand scission). There
On the Specificity of the Eco RI Endonuclease

is precedent for such "one-sided nicking" by Eco RI (Bishop 1979). Moreover, the preference against terminal guanines is not absolute, since bands are observed (with extended digestion) that are consistent with eventual cleavages of the refractory sites at positions 744 and 4269; the initial guanine residues in the sequences for these two sites may possibly exert a counterbalancing favorable influence.

This brings us to the hypothesis advanced by Goodman et al. (1977); i.e. that Eco RI* activity involves preferential cleavage at sites containing guanine as the initial residue (GAATTN; where N is T, A, or G). There are a number of such sequences in the φX174 genome (positions 81, 102, and 139 would, of course, not have appeared since these sites lie on the small DNA fragment generated by Xho 1-Pst I cleavage; Eco RI endonuclease attack on this fragment was not studied). It does appear likely that cleavage occurs at some of these Eco RI* sites rather soon after the initial cuts have been made at the early sites, judging from the digests of whole φX174, and that eventually most or all of these sites are cleaved. We conclude that a guanine residue preceding the core tetramer, AATT, has a favorable effect for recognition and cleavage by the endonuclease, but that this is not necessarily the most favorable combination of bases (short of the canonical sequence) since our evidence points toward the sequence, GGATTT, as the most readily cleaved under our conditions. In any case, cleavages seem to occur more readily at sites with a purine in the first position and a pyrimidine in the sixth.

Other possibly important features of the sequences at cleavage sites are the symmetry and the base composition of immediately neighboring regions. We are unable to discern any particular bias in cleavage activity toward sites flanked by regions rich in either A-T or G-C base pairs, or rich in purine or pyrimidine residue on any given strand. Symmetry in base sequence that extends beyond the core tetramer, AATT, for Eco RI* sites seems to have no especially favorable influence on enzyme activity at a number of sites (positions 1963, 2650, 3939, 4436, or 5340), but conceivably, symmetry plays a role in stimulating cleavage at positions 2222, 956 to 362, and 2324 to 2330. The "tandem" nature (close proximity of two potential Eco RI* sites) of the last two pairs of sites may also play a role.

*Mn**+-induced Noncanonical Cleavage Sequences—As indicated under "Results," we have observed that replacement of Mg**+ by Mn**+ in moderate ionic strength buffer promotes Eco RI cleavage at many of the same noncanonical sites attacked by the endonuclease under conditions of low salt and alkaline pH. Comparison of gel patterns for the two systems suggest that much the same hierarchy of relative site lability is followed in those two sets of noncanonical cleavages as well. Furthermore, this effect is Mn**+-specific; Co**+, Zn**+, or Cu**+ are not effective in reducing the specificity of the enzyme. This result was also obtained by Hsu and Berg (1978) in examining Eco RI digests of SV40 DNA in the presence of Mn**+.

These workers found six highly labile sites of Mn**+-stimulated Eco RI cleavage in SV40 DNA, in addition to the single canonical site present in the molecule, and located these sites on the SV40 restriction map. At the time their results were published, the sequence of SV40 had not been determined fully; subsequently, Fiers and co-workers (1978) and Weiseman and colleagues (Reddy et al., 1978) have published the complete sequence of this DNA. We have compared the DNA sequence with the approximate map positions of the six labile sites observed by Hsu and Berg (1978). Even though there are a number of candidate sequences near each mapped point of cleavage, in every case there is a candidate sequence reading GAATTN (N = A, T, or G). The occurrence of this sequence around all six cleavage loci goes beyond happenstance; it suggests that GAATTN is the favored recognition sequence for the Mn**+-stimulated noncanonical cleavage activity in these digests.

We do find one significant discrepancy between our results with φX174 DNA, and those of Hsu and Berg (1978) with SV40. As indicated above, we find the sequence GGATTT to be the most labile under noncanonical cleavage conditions. Although SV40 DNA contains four such sequences, none lie near the map positions of the six early cleavage sites assigned by Hsu and Berg (1978). While we can provide no definitive interpretation of this discrepancy at present, it may reflect the possibility that flanking sequences play a part in controlling the relative lability of noncanonical, as well as canonical sequences (see below).

Hierarchy of Recognition Elements for Eco RI Endonuclease—As a point of departure for consideration of site recognition mechanisms, we summarize the relative rates of attack of the enzyme on the various potential Eco RI endonuclease cleavage sites observed. The full canonical sequence, GAATTC, is most reactive, although the rate of cleavage of GGATTT (see below) appears to depend in addition on the nature of the flanking base pairs (Thomas and Davis, 1975; Forsblom et al., 1976). The sequence GGATTT seems to be next most reactive. A check of the whole φX174 sequence shows that possible cleavages at the sequences GGATTA or GGATTG will not fit the observed patterns of attack (φX174 contains no sequences GGATTA), and so we discard these latter two possibilities as candidates for recognition sites. Following GGATTT in the hierarchy, we place AATTTC as well as GAATTN (where N is either A or T), basing this assignment again on their relative reactivity in φX174 (RF) (as well as SV40) DNA. After this we suggest that the general Eco RI sequence NAATTN (both N and N' are unspecified nucleotides) provides the bulk of the (slowly reacting) noncanonical cleavage sites (Polsky et al., 1975). We conclude the hierarchy with the central base pair dimer, AT, which is found in all the above suggested recognition sequences and contains the adenine residue which lies closest to the center of the recognition sequence and which, when methylated, is responsible for blocking the cleavage reaction.2

2To investigate further the similarity of the relaxed specificity of the Eco RI endonuclease with the companion methylase, we carried out preliminary experiments to determine whether the nuclease can cleave poly[d(A-T)]. To this end we compared nuclease-treated and control samples of poly[d(A-T)] using a kinase technique to look for the incorporation of new 5’ ends that could incorporate 32P in a kinase-mediated exchange reaction with γ-32PATP. Double-stranded poly[d(A-T)] (5 μg) was digested for 24 h at 37°C in 100 μl of 20 mM Tris, pH 8.0, 2 mM MgCl₂ with ~200 units of Eco RI endonuclease. The reaction was quenched by addition of 1 volume of neutralized phenol. The phenol was decanted and the aqueous phase extracted three times with ether to remove the last traces of phenol. The sample was then ethanol-precipitated and resuspended in 100 μl of "kinase" buffer: 25 mMimidazole, pH 6.5, 1.8 mM MgCl₂, 45 mM KCl, 5 mM diethiothreitol, 300 μM ADP, and 10 μM ATP plus [γ-32P]ATP. Six units of polynucleotide kinase (P-L Biochemicals) were added and the mixture incubated at 37°C for 10 min. The reaction was quenched and the DNA precipitated by addition of sodium pyrophosphate and trichloroacetic acid; the precipitate was collected on a wetted GF/C filter and washed, then dried and counted. As expected, after 25-h treatment with restriction enzyme, approximately 1 of every 2 poly[d(A-T)] molecules had been cleaved once, as evidenced by a 43% rise in incorporated [32P]phosphate (11.6 x 10³ cpm versus 8.1 x 10³ cpm in the control sample). Arguments ruling out possible contamination of the endonuclease preparation (see above) strongly suggest that these cleavages are due to the Eco RI endonuclease. Thus, the reduction in specificity of the Eco RI endonuclease appears to parallel that of the companion methylase, even
Recognition Elements Utilized by the Endonuclease—There now exists a moderate amount of information on the effects of alterations in the canonical sequence with respect to the activity of both the Eco RI methylase and the endonuclease: here we attempt to combine this body of information with our own results on the specificity of the noncanonical activity in order to deduce specific points of contact on the DNA which may be important for the binding specificity of the endonuclease.

The bulk of the available evidence suggests strongly that the endonuclease makes most, if not all, of its recognition contacts with the initial G-C base pair of the canonical sequence in the major groove of the DNA helix. Thus, Berkner and Folk (1977) and Ito et al. (1975) have shown that glycosylated hydroxymethylcytosine residues in the canonical recognition sequence will block cleavage activity completely, implicating contacts in the major groove; however, the presence of the C5 position of cytosine of an hydroxymethyl group does not abolish cleavage activity (Kaplan and Nierlich, 1975), suggesting that the cytosine C5 position itself is not an intrinsically important point of contact. Furthermore, substitution of inosine for guanine in the canonical sequence has no appreciable effect on endonuclease activity (Modrich and Rubin, 1977), demonstrating that the N2 amino group of guanine (in the minor groove) is not required by the endonuclease for site recognition.

As pointed out by Seeman et al. (1976), a single directed hydrogen bond to a given base will not suffice to distinguish that base pair from all other possibilities; a minimum of two directed interactions are required for unambiguous recognition. For recognition here of the initial G-C (or the terminal C-G) base pair two interactions are required which are sufficiently specific to discriminate these pairs from the "wrong" pairs: A-T, T-A, and C-G. The strong prejudice shown against cleavage at sequences beginning with C (or ending with G) suggests that at least one of these specific interactions is lost (and perhaps unfavorable contact points are introduced) in these cases. Hydrogen bond contacts in the minor groove are quite similar for G-C and C-G (Seeman et al., 1976); in addition the evidence of Modrich and Rubin (1977; discussed above) argues against important contacts in this groove. In the major groove, for G-C pairs, we can eliminate the C5 position on cytosine as an important contact point, since hydroxymethylation at this position does not block cleavage.

Remaining sites for possible hydrogen bonds are the N7 position on the purine, the N4 amino group of cytosine, and the O6 carbonyl of guanine.

Seeman et al. (1976) present a very useful set of diagrams illustrating the stereochemistry of double-helical A-U (or A-T) and G-C base pairs. By inspection, it is apparent that G-C and C-G base pairs are readily distinguished via the major groove by contacts at any of the donor or acceptor groups mentioned above. Ambiguous points of contact in the major groove are completely absent, and this probably explains the strong discrimination shown against the replacement of G-C base pairs by C-G pairs in the canonical sequence. By comparison, G-C and T-A base pairs would be difficult to resolve if the endonuclease probed only the center of the major groove for hydrogen-bonding groups, since the O4 and O6 carbonyl of thymine and guanine differ in position by only 1.1 Å (as do the N4 and N6 amino groups of cytosine and adenine). Unambiguous resolution of T-A from G-C could be provided, however, by the N7 imino group on the purine; any pyrimidine replacement for guanine would suffer from the lack of a comparably positioned hydrogen bond acceptor group.

In turn, A-T and G-C base pairs are unambiguously distinguished by contacts in the center of the major groove of the helix; this discrimination is ambiguous only if contact is made solely with the N7 position of the purine ring, at the sides of the major groove. We note here that a single hydrogen-bonding donor probe from the protein probably could not distinguish between the two different purines by bonding to the N7 position; moreover, a thymine replacement for guanine would place an acceptor group (the O4 carbonyl) rather close to the original N7 donor group (approximately 2 Å toward the center of the groove and 1 Å farther out), and so a slight displacement of the probing moiety might still enable it to make sufficient contact for the enzyme to bind and cleave the altered sequence.

Turning now to recognition of the four A-T or T-A base pairs of the canonical sequence, we note that the endonuclease shows a moderate sensitivity to alterations of the C5 substituent on thymine. Berkner and Folk (1977) have shown that while a uridine replacement has no effect on $V_{max}$ or $K_m$ for the endonuclease, substitution by hydroxymethyluridine (CH$_2$OH replacing CH$_3$ at the C5 position of thymidine) lowers $V_{max}$ 20-fold, while showing little or no effect on $K_m$. More recently, Marchionni and Roufa (1978) have reported that 5-bromouridine-containing DNA is digested more slowly by the endonuclease than unsubstituted DNA. The effects of these various substituents at the C5 position may reflect either steric interference by the bulkier bromine or hydroxymethyl group, or the altered polarizability of the C5 substituent, with possible consequences for hydrophobic interactions (Berkner and Folk, 1977).

A central major groove interaction involves the methylation or non-methylation of the N6 group of the central adenine residues in the canonical recognition sequence. The endonuclease is highly sensitive to methylation of these adenine residues, since half-modified sites are as well protected against cleavage as fully modified sites (Modrich and Zabel, 1976). Previous work from this laboratory on the conformation of N6-methyladenine residues in polynucleotides has demonstrated that such residues destabilize the DNA helix by -0.4 to -0.9 kcal/mol (Engel and von Hippel, 1978). The helical structure, however, is still intact; bases are fully paired, with no "eversion" of the modified residues of the double helical conformation. The N6-methyl groups project into the major groove of the helix, not inward to the interior of the helix, and so they can provide an obvious recognition feature. It is not clear exactly how introduction of the methyl group blocks cleavage. Perhaps it interferes with a hydrogen bond to the same adenine moiety, at either the N6 or N7 positions, or it might interfere sterically with groups on the protein and block either cleavage or binding. Protection of half-modified sites may reflect the dimeric structure of the endonuclease and a possible symmetry of interaction with the basically symmetric recognition sequence (Modrich and Zabel, 1976); interference with a single subunit by a single N6-methyladenine might be sufficient to protect the site fully. An intriguing possibility is that the N6-methyladenine, through its destabilizing effect on the DNA helix locally might affect binding or cleavage by the enzyme. With regard to this last possibility it is interesting to note that a double-stranded substrate is necessary for enzyme activity; Greene et al. (1975) observed neither cleavage by the nuclease, nor modification by the methylase, of a self-complementary oligonucleotide containing the canonical recognition sequence at temperatures above the melting point of the oligomer.

These observations, taken together, indicate that the en-
donuclease makes most or all of its critical recognition contacts in the central region of the canonical sequence via the major groove of the helix. We can bolster this assertion further by considering the effects on sequence recognition within this region by various base pair replacements: (i) replacement of A-T by T-A presents much the same sort of array of hydrogen-bonding donor groups in the minor groove for either base pair, but differences in the arrangement of such groups is readily apparent in the major groove; (ii) replacement of A-T by C-G is readily apparent through hydrogen-bonding contacts in either the major or the minor groove, which easily explains the strong discrimination shown against this particular substitution; (iii) replacement of A-T by G-C could be detected through contacts in either groove; however, the endonuclease does accept a G-C replacement for A-T in some noncanonical sites (e.g. GGATTT). This last fact suggests that minor groove contacts are relatively unimportant, as G-C and C-G base pairs are not readily distinguished from the minor groove. This forces us to consider contact points in the major groove where G-C base pairs might bear some similarity to A-T pairs. A leading candidate is the N7 hydrogen bond acceptor present on both purine rings; this readily discriminates against T-A and C-G replacements, while allowing some ambiguity in distinguishing A-T from G-C base pairs. More stringent discrimination of A-T base pairs could be provided by a second hydrogen-bonding contact, for example, at the N6-amino group of adenine, or by hydrophobic or steric interactions around the C5-methyl group of thymine.

A Model for Endonuclease Site Recognition—The points of contact within the recognition sequence that have been deduced can be summarized as follows: (i) the leading G-C base pair of the sequence is probably recognized by contact at the N7 position of guanine, and also possibly by contacts at the 06 carbonyl of guanine or the N4-amino group of cytosine; (ii) the innermost A-T (or T-A) base pair is probably recognized by contact with the N7 amino group and the N6-amino group of adenine, and possibly by contact with the C5 methyl group of thymine; (iii) the outer A-T (or T-A) base pair is probably recognized by contact at the N7-imino group of adenine, the 04-carbonyl of thymine, or the C5-methyl group of thymine.

Using Fig. 6 (top), we show (in a highly schematic form) the set of contacts in the canonical sequence just enumerated for the endonuclease. Briefly, each base pair is represented by a line from which extend (vertically out into the major groove of the helix) a set of hydrogen bond acceptor and donor groups on the bases which represent relative rotational displacement of the base pairs in the double helix. A AkT represents a single base pair, with the symbols A and D representing hydrogen bond acceptor and donor groups on the bases which extend into the major groove (upward from the base pair line) or into the minor groove (downward from the line). Base pair lines are offset proportionately to represent relative rotational displacement of the base pairs in the double helix.

Recalling that the sequences GGATTT, AAATTT, and GAATTA represent especially favorable noncanonical cleavage sequences, we can compare in this fashion the trimer (half-site) sequences GAA, GGA, AAA, and TAA (Fig. 6, bottom). A very intriguing feature, common to all four half-sites, emerges; i.e. an alignment along the axis of the helix of two hydrogen bond acceptor-donor pairs which can easily be bridged by a bifunctional hydrogen-bonding amino acid moiety such as glutamine or asparagine. Recognition of specific base pairs by such “bridging” probes on a protein has been previously proposed (Seeman et al., 1977) but consideration was limited to contacts made in the plane of the base pairs (perpendicular to the helix axis). We propose (see also von Hippel, 1979) that bridging contacts are equally likely to be made across base pairs in the recognition sequence (parallel to the helix axis), as illustrated by the shaded rectangles in Fig. 6, drawn about the proposed hydrogen bond donor-acceptor pairs. This model is entirely consistent with previous deductions concerning specific base contacts, it accounts for the preferences enumerated above in noncanonical cleavage site sequences, and it predicts that sequences where such bridging contacts are very distorted or totally absent will resist cleavage even under noncanonical conditions.

One of the four noncanonical half-site sequences, GGA, exhibits a lateral displacement within the major groove of one of the two cross-base bridges; the resulting distorted recognition element could be responsible for the reduced rate of cleavage at the sequence GGATTT, as compared to canonical
site cleavage. (A comparable distortion is required to form the double cross-bridge structure for the half-site sequences AGA and CAA, which have been suggested for the noncanonical site at ϕX174 position 3962.) We note that other possible candidate half-site sequences at which cleavage has not been observed (GTA or GCA) can be eliminated a priori by the total absence of the proposed set of double bridge contacts (GCA) or by the need to grossly distort both bridging contacts (GTA).

Sequences such as GGATTT are not palindromic (symmetric); since cleavage of the two strands appears to involve independent binding and catalytic acts (see Modrich, 1979), it may be that in such cases the "better aligned" half-site sequence (e.g. here AAA) is cleaved first, introducing additional flexibility for the enzyme to better align (and better bind to?) the other half-site sequence. The single strand nicking of the nonsymmetrical sequence GAATTA (Bishop, 1979) under noncanonical conditions is certainly consistent with such a hypothesis; the TAA half-site sequence here is somewhat distorted (Fig. 6B) and thus would be expected to be cleaved only under noncanonical conditions.

Recognition Elements Utilized by the Methylase—Based on these results for the endonuclease, we consider now a parallel analysis of the recognition mechanism of the Eco RI methylase. From the results of Modrich and Rubin (1977) on the effects of inosine-substituted recognition sites, the methylase apparently uses the N2 amino group of guanine (in the minor groove) to distinguish the initial G-C base pair of the canonical sequence. This is supported by the observation that the methylase acts readily on phage T4 DNA (Berkner and Folk, 1977) which contains substantial amounts of glucosylated hydroxymethylcytosine; one might expect the bulky sugar substituent to block contacts not only at the C5 position of cytosine, but at other nearby points in the major groove of the helix as well (e.g. at the N4-amino group of cytosine or the O6-carbonyl of guanine). The fact that the methylase apparently disregards such potentially blocked contacts implies their lack of significance in the site-recognition mechanism of the methylase. However, based on our observation that in overmethylated ϕX174(RF) DNA sequences such as AAATTT are protected against noncanonical attack by the endonuclease, we conclude that the N2 amino group of the initial guanine residue of the canonical sequence is not absolutely essential for the methyltransfer reaction. We speculate further that the methylase might make some other contact in either the major or minor groove with the initial G-C base pair, since the enzyme obviously discriminates between G-C and C-G base pairs at this position in the recognition sequence; these base pairs differ only slightly in terms of contact at the N2 of guanine (Seeman et al., 1976), and a second, independent contact-interaction (in either groove of the helix) would seem necessary for the precise discrimination seen. Perhaps it is the interaction with this (postulated) second contact point, even in the absence of the guanine N2 donor group, that allows methylation of such noncanonical sequences as AAATTT, GAATTT, or GAATTA, with concomitant protection against attack by the endonuclease.

The methylase is quite sensitive to substitution at the C5 position on thymine, since replacement of the methyl group here by either hydrogen or an hydroxymethyl group will essentially eliminate canonical site methylation under standard conditions (Berkner and Folk, 1977). In contrast to this, replacement of this methyl group by bromine allows both canonical and noncanonical reactions to proceed with little apparent effect on the rate of reaction. It is not clear whether these differing results are due to steric effects on binding or catalysis, or to some more subtle effect of the varying electro-negativities of the substituents.

Our results on the apparent lack of protection of sequences such as GGATTT imply a strong discrimination by the methylase of A·T from G·C base pairs at the second position of the recognition sequence. The observed protection afforded sequences such as GAATTT argues against the lack of methylation being solely due to replacement of the terminal (canonical) C·G base pair by T·A. Possibly the G·C for A·T substitution (in GGATTT) does not provide a proper contact on the pyrimidine (see above), since the C5 hydrogen substituent of cytosine would mimic the disfavored uridine substitution in this respect; many other possible contacts in this base pair substitution are distorted or absent as well, of course, and may play a role here also.

Concerning recognition of the central A·T base pair to be methylated, we suggest that the sensitive discrimination of the (proper) A·T base pair from a C·G base pair (which offers a similarly positioned exocyclic amino group in the major groove) may be due in part to contacts at the C5 position of the thymine residue; the C·G base pair would, of course, completely lack a comparably positioned substituent.

Methylation of canonical sites is apparently insensitive to the difference between half-modified and unmodified sites, implying an asymmetry in recognition contacts in the canonical sequence (Modrich and Rubin, 1977). More particularly, the methylase must "ignore" the state of modification of the central adenine residue on one strand while it operates on the corresponding adenine of the other strand of the recognition sequence. This implies that the N6-amino group of the former adenine residue (modified or not) is not a contact point for the methylase during this process.

A Model for Methylase Site Recognition—We can draw these various results and speculations together in the form of a template model for recognition contacts made by the methylase (Fig. 7), including only those contacts which appear to be important. New features are introduced into this template representation because of the possible importance of contacts in the minor groove and with methyl groups in the major groove. To represent possible hydrogen-bonding groups in the

![Schematic template of canonical site recognition contacts for Eco RI methylase. Thymine C5-methyl groups are represented by me; a monomethylated N6-amino group of adenine is represented by (D).](image)
minor groove, these acceptors or donors are drawn downwards from the line representing the basic pair under consideration (major groove contacts still extend upward). Thymine C5 methyl groups appropriately positioned on the template are represented as me, and extend upward from the base-line into the major groove; monomethylated N6-amino groups of adenine are represented by a bracketed donor symbol (D).

Comparison of Figs. 6 (top) and 7 brings out the differences in recognition contacts made by the Eco RI endonuclease and methylase. The asymmetry of site recognition by the methylase appears in the center of Fig. 7, where the one N6 adenine contact is methylated and "ignored" (bracketed "D"). By contrast, the array of contacts made by the endonuclease is fully symmetric. Another difference which is apparent in these representations is the predilection of the endonuclease for contacts in the center and to the "purine" side of the major groove, while contacts of the methylase in the major groove are on the opposite ("pyrimidine") side. There is, of course, the important contact made in the minor groove by the methylase; no good evidence exists for any important minor groove contacts for the endonuclease. With the present information on Eco RI methylase recognition contacts we are unable to discern any pattern of intra- or cross-base pair bridging contacts such as those suggested for the Eco RI endonuclease.

Effects of Environmental Alterations on Site Specificity of Interaction—The dependence of Eco RI endonuclease site specificity on pH, low ionic strength, and polar organic solutes could be due to titration of groups on the protein that are responsible for recognition contacts. Alternatively, titration of certain groups on the protein (not necessarily directly involved in interactions with the nucleic acid) could change the protein's conformation at key points sufficiently to alter the specificity of interaction by other groups on the protein. The mechanism(s) by which Mn$^{2+}$ ions affect the specificity of the endonuclease must be quite similar to the mechanism(s) for pH, ionic strength, etc., since the same trends in site specificity are observed. It seems unlikely that Mn$^{2+}$ produces its effects by simply binding to the nucleic acid and somehow altering recognition contacts on the bases. Although Mn$^{2+}$ does show an appreciable affinity for the N7 position on purines, relative to binding to phosphate groups (reviewed by Eichhorn, 1973), transition metal ions such as Cu$^{2+}$ and Co$^{2+}$, which have even greater base versus phosphate affinities, do not stimulate Eco RI activity by the endonuclease. Thus, it seems more likely that the reduction in sequence specificity effected by Mn$^{2+}$ is due to protein-Mn$^{2+}$ interactions. Possibly the Mn$^{2+}$ ion acts as a bridge between groups on the protein and on the nucleic acid (at the phosphate backbone, at the sugars, or at the bases themselves). Another possibility, consistent with a possible titration of groups on the protein, is that the Mn$^{2+}$ ion forms a complex with the titrated form of this group or groups, thus shifting its (their) pK$_a$ value(s) downward (Freeman, 1973). This could explain why Mn$^{2+}$ effects on nucleic site specificity are seen under considerably less alkaline conditions, and at higher ionic strengths, than the other perturbations examined.

The overall rate of reaction at noncanonical sites exhibited by the endonuclease or methylase is only 1% to 10% of that seen at canonical sites. We assume that this difference is due to a change in the stability of the enzyme-nucleic acid complex, and in particular, is manifested in a low rate of dissociation of the enzyme from "good" sites, as compared to "poor" sites.

Noncanonical sites presumably lack one or more necessary contacts for a stable complex (or they have gained new destabilizing interactions) and are thus generally "rejected" by a (relatively) rapid dissociation of the putative enzyme-substrate complex before the catalytic step can take place. If, for example, a 100-fold difference in rate reflects a 100-fold difference in binding affinity, at 37°C (310 K) this corresponds to a difference in binding free energy between the complexes of roughly 2.8 kcal/mol, a quantity easily accounted for by minor alterations in the hydrogen-bonding interactions responsible for the extra specificity of the canonical sequence. (We note that a destabilization of 2.8 kcal/mol is about 1½ times the destabilization of repressor-operator complexes observed for single base pair changes in operator-constitutive mutations of the lac operator sequence (see Goeddel et al., 1978; von Hippel, 1979).)

To illustrate these notions in a semiquantitative fashion, we conclude with a heuristic calculation on the relative effects of ionic and base-specific interactions on the stability of the enzyme-substrate complex. Various results (Goppelt et al., 1980; Woodhead and Malcolm, 1980; Modrich, 1979) suggest that at least the Eco RI endonuclease can bind specifically to potential cleavage sites in the absence of divalent cations (and thus of catalysis). This result also supports the assumption that it is binding, rather than catalysis, which is nucleotide sequence-specific. In addition, preliminary results in our laboratory$^3$ and by Goppelt et al. (1980) indicate that nuclease binding decreases appreciably with increasing salt concentration. Following the analysis of Record et al. (1976, 1978), these results can be used to calculate the contribution of electrostatic interactions to the binding free energy of complex formation as follows:

$$\Delta G_{\text{ionic}} = m' \psi R T \ln [\text{Na}^+]$$

where $m'$ is the number of protein-nucleic acid charge-charge interactions formed in the complex and $\psi$ is the fraction of counterions bound thermodynamically (per phosphate) to the unliganded DNA. Using a value of $m' = 2$ charge-charge interactions (Goppelt et al., 1980), and $\psi = 0.88$ (Record et al., 1976), we calculate that for the noncanonical sites at 37°C $\Delta G_{\text{ionic}}$ changes (become more favorable) by $\sim -2.5$ kcal/mol with an 10-fold decrease in salt concentration. This increase in electrostatic binding free energy just about offsets the loss of nonelectrostatic binding free energy associated with the opening of one to two base pair specific interactions in the complex, as estimated from the lac repressor-operator studies (see above), and suggests that the general approach outlined here may indeed have quantitative validity.

Acknowledgments—We are grateful to our colleagues in the Institute, particularly Drs. Gerald Smith and Daryl Faulds, for many helpful discussions. O. H. also thanks Dr. Karen Sprague for financial support (United States Public Health Service Research Grant GM-25388, K. U. Sprague, Principal Investigator).

REFERENCES

Berkner, K. L., and Folk, W. R. (1977) J. Biol. Chem. 252, 3185-3193
Berkner, K. L., and Folk, W. R. (1978) Nucleic Acids Res. 5, 435-450
Bishop, J. O. (1979) J. Mol. Biol. 128, 545-559
Brownlee, G. G. (1972) Determination of Sequences in RNA, pp. 54-66 (North Holland) American Elsevier, New York
Eichhorn, G. L. (1973) in Inorganic Biochemistry (Eichhorn, G. L., ed) pp. 1191-1209, 1210-1243, Elsevier, New York
Engel, J. D., and von Hippel, P. H. (1978) J. Biol. Chem. 253, 927-934
Fiers, W., Contreras, R., Haegeman, G., Rogiers, R., van de Voorde, A., van Heuswelyn, H., van Herweghe, J., Voltkaert, G., and Ysebaert, M. (1978) Nature 273, 113-120
Forsslom, S., Rigler, R., Ehrenburg, M., Pettersson, U., and Philipson, L. (1976) Nucleic Acids Res. 3, 3293-3299

$^3$ Also C. P. Woodbury, unpublished work.
