Accuracy of the *Eco*RV Restriction Endonuclease: Binding and Cleavage Studies with Oligodeoxynucleotide Substrates Containing Degenerate Recognition Sequences[†]

Jürgen Alves,*[,‡] Ursel Selent,[§] and Heiner Wolfes[‡]

Zentrum Biochemie, Medizinische Hochschule Hannover, Konstanty-Gutschow-Strasse 8, D-30623 Hannover, Germany, and Institute für Biochemie, Justus-Liebig-Universität Giessen, Heinrich-Buff-Ring 58, D-35392 Giessen, Germany

Received March 21, 1995; Revised Manuscript Received June 23, 1995[⊗]

ABSTRACT: In order to investigate the accuracy of the *Eco*RV restriction endonuclease, we have synthesized a set of double-stranded oligodeoxynucleotides comprising the canonical recognition sequence, the 9 star sequences (i.e., sequences deviating by one base pair from the canonical sequence), and the 18 mismatch sequences (i.e. sequences deviating by one base from the canonical sequence). For each individual single strand of all these 28 substrates we have measured the rate of phosphodiester bond cleavage under normal buffer conditions. Double-strand cleavage of star substrates is at least 5 orders of magnitude slower than cleavage of the canonical substrate. In contrast, most of the mismatch substrates are accepted more readily. In the absence of the essential cofactor Mg^{2+}, *Eco*RV binds weakly but equally to the canonical and degenerate substrates, (i.e., K_{Diss} is in the micromolar range). However, the inactive catalytic site mutant D90A in the presence of Mg^{2+} binds the canonical substrate 1−2 orders of magnitude better than degenerate substrates. Therefore, the *EcoRV* endonuclease needs the essential cofactor Mg^{2+} to create thermodynamic discrimination between degenerate and canonical sites. But the main discrimination is kinetically controlled and takes place during cleavage. While in the canonical substrate both single strands are cleaved with an equal velocity, in all other substrates one single strand is cleaved faster than the other one, resulting in a dissociation of the enzyme from the DNA between the two cuts. *In vivo* this may lead to a repair of the erroneous cleavage site by DNA ligases. The order of single-strand nicking together with the division of base contacts on both subunits suggests that correct recognition by one subunit triggers cleavage by the other one.

Restriction endonucleases constitute a bacterial defense system against foreign DNA (e.g., bacteriophage DNA) by cleaving it within a defined recognition sequence (Heitman, 1993; Roberts & Halford, 1993). Their corresponding modification enzymes protect the chromosomal DNA by methylation of this sequence. An essential requirement of restriction enzymes is their high accuracy: cleavage at sites protected by methylation or at sites which deviate from the canonical sequence would induce recombination and repair events which may lead to mutations. Nevertheless, restriction enzymes exhibit a very low degree of DNA cleavage activity at noncanonical sequences as described for *Eco*RI (Rosenberg & Greene, 1982), *Bam*HI (George & Chirikjian, 1982), *Hin*dII (Nasri & Thomas, 1986), *Eco*RV (Halford et al., 1986), *Pvu*II (Nasri & Thomas, 1987), *Taq*I (Barany, 1988), and *Ceq*I (Izsvak & Duda, 1989). In all these studies, the erroneous cleavage of the enzymes was accelerated by changing the buffer to the so-called "star" conditions. Star cleavage is also detectable at high enzyme concentration over prolonged incubation times. Therefore, it indicates the normal errors made by these enzymes.

The *Eco*RV restriction endonuclease recognizes the sequence -GAT↓ATC- and cleaves it as indicated to produce blunt ends (Schildkraut et al., 1984). The amino acid sequence of the enzyme was deduced from the gene (Bougeleret et al., 1984), and cocrystal structures with oligodeoxynucleotides were solved (Winkler et al., 1993; Kostrewa & Winkler, 1995). Under normal buffer conditions, *Eco*RV cleaves plasmid DNAs at sites deviating by one base pair from the canonical recognition sequence at least 10^4 times more slowly (Taylor & Halford, 1989). *EcoRV* methylase does not prevent noncanonical cleavage (Taylor et al., 1991). At noncanonical sites, a change in mechanism from simultaneous double-strand cleavage to a preferred single-strand nicking is seen, which allows the DNA ligase to repair erroneous *Eco*RV cleavage *in vivo* (Taylor et al., 1990). The cleavage rate is modulated by the DNA sequences flanking the cleavage site (Taylor & Halford, 1992) as was also shown for *Eco*RI (Thomas & Davis, 1975; Alves et al., 1984). Therefore, we decided to use a set of oligodeoxynucleotide substrates with identical flanking sequences to study the accuracy of the *Eco*RV restriction endonuclease. The set consists of one canonical substrate, nine star substrates with all possible one base pair deviations of the cognate site, and 18 mismatch substrates with all possible mismatch base pairs within the recognition sequence. For these 28 substrates, binding affinity and cleavage rates by *Eco*RV were determined.

[†] This work has been supported by grants from the Deutsche Forschungsgemeinschaft (Pi 122/2-3).

* Address correspondence to this author. Phone: +49-511-532 3671. Fax: +49-511-532 5966.

[‡] Medizinische Hochschule Hannover.

[§] Justus-Liebig-Universität.

[⊗] Abstract published in *Advance ACS Abstracts,* August 15, 1995.

0006-2960/95/0434-11191$09.00/0

EXPERIMENTAL PROCEDURES

EcoRV Restriction Endonuclease Purification and Characterization. The *Eco*RV restriction endonuclease was isolated in homogeneous form from an overproducing strain as described (Thielking et al., 1991). Prior to the experiments, the enzyme was dialyzed against 10 mM Tris·HCl, pH 7.5, 300 mM NaCl, and 1 mM EDTA in order to remove glycerol, which is an ingredient of the normal storage buffer. This enzyme stock solution had a specific activity of approximately 2×10^6 units/mg of protein, as assayed with bacteriophage λ DNA, and is stable for at least 2 weeks if stored at 4 °C.

Purification of the D90A Mutant. The catalytic site mutant D90A was purified as described in Selent et al. (1992).

Oligodeoxynucleotide Synthesis. Oligodeoxynucleotides were synthesized, purified, and characterized as described (Alves et al., 1989b).

Melting Curves. The thermal stabilities of double-stranded oligodeoxynucleotides were determined in a Zeiss DMR 10 spectrophotometer. For this purpose equimolar amounts of single-stranded oligodeoxynucleotides (2 μM) were annealed in cleavage buffer (10 mM Tris·HCl, pH 7.5, 50 mM NaCl, and 10 mM $MgCl_2$). Melting curves were then recorded between 10 and 90 °C with a heating rate of 10−15 °C/h. To record melting curves of a nicked mismatch substrate, 150 μM of an analytically ^{32}P-labeled canonical single strand was incubated in cleavage buffer with 2.5 μM *Eco*RV for 24 h at room temperature. Complete cleavage was assessed by homochromatography and autoradiography. The solution was heated to 95 °C for 10 min to denature *Eco*RV irreversibly, mixed with the noncanonical single strand, and annealed.

Cleavage Experiments. Cleavage experiments were carried out with 1 μM double-stranded oligodeoxynucleotide radioactively labeled in both strands in 10 mM Tris·HCl, pH 7.5, 50 mM NaCl, and 10 mM $MgCl_2$ at 20 °C. The concentration of *Eco*RV endonuclease was 1 μM in most experiments. Only for the canonical substrate and those mismatch substrates with $k_{rel} > 0.1$ min^{-1} was it necessary to use 0.02 μM *Eco*RV. The asymmetric design of the oligodeoxynucleotides, with nine base pairs 5′ to the recognition sequence and four 3′ flanking base pairs, allows simultaneous detection of the cleavage products of both single strands. After separation by homochromatography and visualization by autoradiography, the radioactive spots were excised and counted as described (Thielking et al., 1990).

Nitrocellulose Filter Binding Experiments. Binding experiments with wild-type *Eco*RV were carried out at 20 °C in 10 mM Tris·HCl, pH 7.5, 50 mM NaCl, 0.5 mM EDTA, 50 μM DTT, and 100 μg/mL bovine serum albumin. Binding experiments with the D90A mutant were performed at 20 °C in 10 mM Tris·HCl, pH 7.5, 50 mM NaCl, 10 mM $MgCl_2$, 50 μM DTT, and 100 μg/mL bovine serum albumin. Binding of the enzymes to the canonical substrate was measured directly as described (Alves et al., 1989a). The binding to the noncanonical substrates was determined in competition with the canonical substrate as described (Thielking et al., 1990).

RESULTS

Substrates: Design and Stability. The comparison of cleavage rates at the canonical recognition sequences and the star sequences in plasmid DNA can be misleading, since cleavage rate is influenced by neighboring sequences (Taylor & Halford, 1989) and modulated by different sequence contexts. Therefore, we have chosen 19-mer oligodeoxynucleotides with identical flanking sequences as substrates in order to study the accuracy of the *Eco*RV restriction endonuclease. These oligodeoxynucleotides differ in one base pair or one base from the -GATATC- sequence (Figure 1). The recognition sequence is positioned asymmetrically in the substrates to allow for simultaneous quantitation of cleavage products of each individual single strand. Four base pairs on the shorter side are sufficient to create phosphate contacts extending over a 12−14 bp long enzyme−DNA interface as observed in the *Eco*RV−oligodeoxynucleotide cocrystal structures (Winkler et al., 1993; Kostrewa & Winkler, 1995). The sequences on both sides are designed such that double strands formed by annealing of identical single strands are destabilized by Pu−Pu[1] and Py−Py base pairing while the heteroduplex double strand is stable. Sequences of different substrates vary only in one-half of the symmetrical recognition sequence. Melting curves for one example of each kind of mismatch base pair were recorded in cleavage buffer to confirm that the substrates are double stranded under cleavage conditions. The AA and GA mismatch substrates are almost as stable as the fully base paired canonical substrate (Table 1). The largest destabilizing effect is observed with the GG and TC mismatch substrates, their T_M being 14 °C lower than the

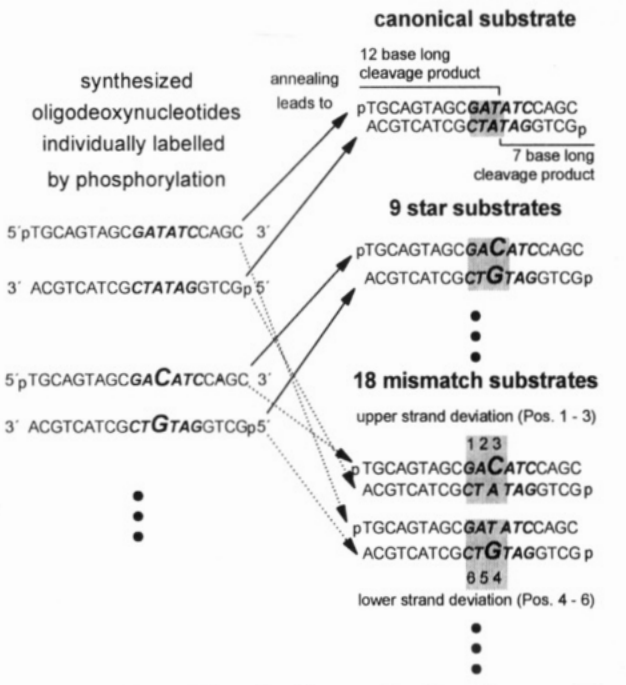

FIGURE 1: Design of the oligodeoxynucleotide substrates. Ten complementary pairs of 19-mer oligodeoxynucleotides (two examples are shown on the left side with the 5′ and 3′ ends of the single strands indicated) were synthesized. Annealing (indicated by arrows) generates the canonical (upper right) and the nine star substrates (one typical example represented) with all possible base pair variations. When the canonical single strands are annealed to the complementary single strands of the star substrates (dotted arrows), 18 mismatch substrates are created. p indicates a ^{32}P-label of each single strand. The length of the labeled cleavage products is shown for both single strands of the canonical substrate.

[1] Abbreviations: Pu, purine; Py, pyrimidine; T_M, melting temperature.

Table 1: Thermal Stabilities of Some Oligodeoxynucleotides Used as Substrates for the *Eco*RV Endonuclease

oligodeoxynucleotide	mismatched bases	T_M (°C)[a]
fully base paired		
TGCAGTAGC GATATC CAGC ACGTCATCG CTATAG GTCG	none	64
mismatch		
TGCAGTAGC **C**ATATC CAGC ACGTCATCG CTATAG GTCG	CC	58
TGCAGTAGC **A**ATATC CAGC ACGTCATCG CTATAG GTCG	AC	56
TGCAGTAGC GATATC CAGC ACGTCATCG **G**TATAG GTCG	GG	50
TGCAGTAGC G**T**TATC CAGC ACGTCATCG CTATAG GTCG	TT	60
TGCAGTAGC GA**A**ATC CAGC ACGTCATCG CTATAG GTCG	AA	62
TGCAGTAGC GA**G**ATC CAGC ACGTCATCG CTATAG GTCG	GA	62
TGCAGTAGC GATATC CAGC ACGTCATCG CT**C**TAG GTCG	TC	50
TGCAGTAGC GATATC CAGC ACGTCATCG CT**G**TAG GTCG	TG	59
mismatch, nicked		
TGCAGTAGC **C**ATATC CAGC ACGTCATCG CTATAG GTCG	CC	31
TGCAGTAGC GATATC CAGC ACGTCATCG CT**G**TAG GTCG	TG	46

[a] Values are accurate within ±2 °C.

Table 2: Cleavage and Binding of Star Substrates by the *Eco*RV Endonuclease

canonical substrate	k_{rel}[min^{-1}][a]	K_{ass}[M^{-1}][b] wt*Eco*RV -Mg	K_{ass}[M^{-1}][b] D90A +Mg
TGCAGTAGC GATATC CAGC ACGTCATCG CTATAG GTCG	15,0 ± 3,8 15,0 ± 4,0	4 x 10^5	6 x 10^7
star-substrate	k_{rel}can/k_{rel}[c]	K_{ass}[M^{-1}][d] wt*Eco*RV -Mg	K_{ass}[M^{-1}][d] D90A +Mg
TGCAGTAGC **C**ATATC CAGC ACGTCATCG **G**TATAG GTCG	650 000 ± 150 000 1 300 000 ± 200 000	4 x 10^5	7 x 10^5
TGCAGTAGC **A**ATATC CAGC ACGTCATCG **T**TATAG GTCG	940 000 ± 150 000 190 000 ± 10 000	6 x 10^5	9 x 10^5
TGCAGTAGC **T**ATATC CAGC ACGTCATCG **A**TATAG GTCG	9 400 000 ± 5 700 000 580 000 ± 80 000	7 x 10^5	7 x 10^5
TGCAGTAGC G**T**TATC CAGC ACGTCATCG C**A**ATAG GTCG	230 000 ± 10 000 150 000 ± 10 000	6 x 10^5	1 x 10^6
TGCAGTAGC G**G**TATC CAGC ACGTCATCG C**C**ATAG GTCG	540 000 ± 200 000 110 000 ± 20 000	7 x 10^5	1 x 10^6
TGCAGTAGC G**C**TATC CAGC ACGTCATCG C**G**ATAG GTCG	160 000 ± 20 000 30 000 ± 5 000	4 x 10^5	4 x 10^6
TGCAGTAGC GA**A**ATC CAGC ACGTCATCG CT**T**TAG GTCG	450 000 ± 50 000 9 400 ± 400	6 x 10^5	1 x 10^6
TGCAGTAGC GA**G**ATC CAGC ACGTCATCG CT**C**TAG GTCG	480 000 ± 40 000 200 000 ± 50 000	6 x 10^5	1 x 10^6
TGCAGTAGC GA**C**ATC CAGC ACGTCATCG CT**G**TAG GTCG	210 000 ± 20 000 16 000 ± 1 000	8 x 10^5	7 x 10^6

[a] k_{rel} values are the average of four experiments. [b] K_{ass} value is the result of eight titrations at different concentrations of the canonical substrate. [c] k_{rel} values are the result of a minimum of two experiments. [d] K_{ass} values are the result of a minimum of two titrations using the unphosphorylated substrate in competition with the phosphorylated canonical substrate and are accurate within at least ±30%.

fully base paired double strand. The stability of a nicked cleavage intermediate (corresponding to these substrates) was also measured. Even in a CC mismatch substrate, where the mismatch is two bases from the nick, the melting temperature declines to only 31 °C. Therefore, it can be assumed that at a cleavage temperature of 21 °C most of each substrate is in double-stranded cleavable form even after single-strand nicking.

Interaction with the Canonical Substrate. In the canonical substrate both strands are cleaved with the same rate of 15 min^{-1} (Table 2). Binding of the wild-type enzyme in the absence of Mg^{2+} is quite low. It was already shown that under these conditions no sequence-specific binding can be detected (Taylor et al., 1991). Therefore, we decided to

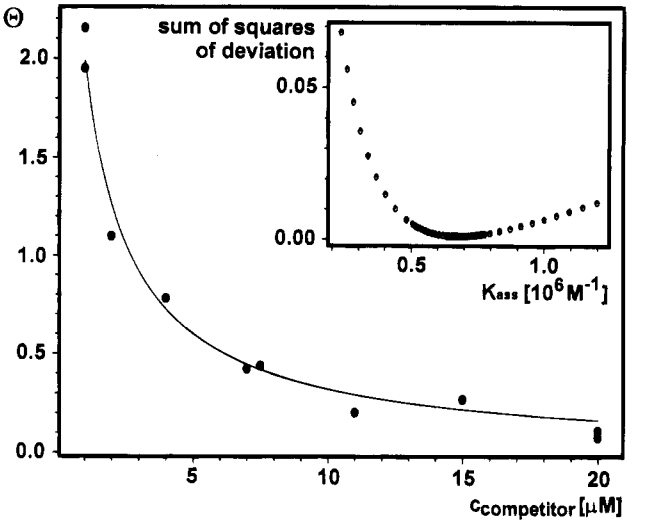

FIGURE 2: Example of a binding competition experiment. ^{32}P-Labeled canonical substrate and D90A mutant (each 40 nM) were incubated with different amounts of the GACATC star substrate in the presence of Mg^{2+}. θ (the concentration of the D90A−canonical substrate complex divided by the total concentration of the canonical substrate) corresponds to the radioactivity retained on the nitrocellulose filters. The solid line represents the best fit of $K_{ass} = 6 \times 10^7$ M^{-1} for the canonical substrate and $K_{ass} = 7 \times 10^6$ M^{-1} for the GACATC star substrate [fitting procedure: cf. Thielking et al. (1990)]. The insert illustrates the confidence limit for this determination: An increase or decrease of the binding constant for the star substrate by more than 30% leads to an unacceptable large deviation of measured and calculated data points.

determine the binding affinity of the catalytic site mutant D90A in the presence of Mg^{2+}. Specific binding to the recognition site was shown for this mutant (Thielking et al., 1992). The binding constant of 6×10^7 M^{-1} is 2 orders of magnitude higher than the one for wild-type *Eco*RV in the absence of Mg^{2+} (Table 2).

Interaction with Star Substrates. These substrates with deviations of one base pair from the recognition sequence are nicked in one strand at least 4 orders of magnitude more slowly. Cleavage in the second strand (double-strand cleavage) occurs at least 5 orders of magnitude more slowly compared with the canonical sequence (Table 2). The maximum discrimination of *Eco*RV is observed for substrates with substitutions of the GC base pair followed by those of the outer AT base pair, while substitutions of the inner TA base pair lead to the fastest nicking rates. One important difference from the cleavage of the canonical substrate is the discontinuous cleavage of the double strand. Nicking rates of the two individual single strands of the different star substrates deviate by a factor of 2−50.

In the absence of Mg^{2+} wild-type *Eco*RV is unable to discriminate between the canonical and star substrates (Table 2): Binding to all substrates is identical within the experimental error margin, as already demonstrated by Taylor et al. (1991) for nonspecific sequences. The D90A mutant, however, is able to distinguish between canonical and star substrates. Under Mg^{2+} buffer conditions, binding to star substrates is between 1 and 2 orders of magnitude lower as compared with binding to the canonical substrate (Figure 2). As can be seen from Table 2, discrimination in binding strength corresponds roughly to discrimination in cleavage rates: A star substrate which is cleaved faster is also bound better. But the extent of reduction of cleavage rates

Table 3: Cleavage and Binding of Mismatch Substrates by the *Eco*RV Endonuclease

canonical substrate	k_{rel}[min^{-1}][a]	K_{ass}[M^{-1}][b] wt*Eco*RV -Mg	K_{ass}[M^{-1}][b] D90A +Mg
TGCAGTAGC GATATC CAGC ACGTCATCG CTATAG GTCG	15,0 ± 3,8 15,0 ± 4,0	4 x 10^5	6 x 10^7
mismatch-substrate	k_{rel}can/k_{rel}[c]	K_{ass}[M^{-1}][d] wt*Eco*RV -Mg	K_{ass}[M^{-1}][d] D90A +Mg
TGCAGTAGC **C**ATATC CAGC ACGTCATCG CTATAG GTCG	180 000 ± 60 000 11 000 ± 700	1 x 10^5	9 x 10^5
TGCAGTAGC **A**ATATC CAGC ACGTCATCG CTATAG GTCG	430 000 ± 60 000 32 000 ± 7 000	7 x 10^5	2 x 10^6
TGCAGTAGC **T**ATATC CAGC ACGTCATCG CTATAG GTCG	88 000 ± 13 000 16 000 ± 4 000	7 x 10^5	1 x 10^6
TGCAGTAGC G**T**TATC CAGC ACGTCATCG CTATAG GTCG	750 ± 320 68 ± 8	8 x 10^5	5 x 10^5
TGCAGTAGC G**G**TATC CAGC ACGTCATCG CTATAG GTCG	45 000 ± 17 000 880 ± 110	6 x 10^5	2 x 10^6
TGCAGTAGC G**C**TATC CAGC ACGTCATCG CTATAG GTCG	36 ± 5 3,4 ± 0,3	7 x 10^5	2 x 10^6
TGCAGTAGC GA**A**ATC CAGC ACGTCATCG CTATAG GTCG	330 000 ± 20 000 12 000 ± 400	7 x 10^5	1 x 10^6
TGCAGTAGC GA**G**ATC CAGC ACGTCATCG CTATAG GTCG	15 000 000 ± 10 000 000 1 700 000 ± 300 000	6 x 10^5	6 x 10^5
TGCAGTAGC GA**C**ATC CAGC ACGTCATCG CTATAG GTCG	2 600 ± 200 110 ± 10	3 x 10^5	3 x 10^6
TGCAGTAGC GATATC CAGC ACGTCATCG CT**T**TAG GTCG	380 ± 70 130 ± 10	1 x 10^6	1 x 10^7
TGCAGTAGC GATATC CAGC ACGTCATCG CT**C**TAG GTCG	3400 ± 1200 4,5 ± 0,3	9 x 10^5	3 x 10^7
TGCAGTAGC GATATC CAGC ACGTCATCG CT**G**TAG GTCG	4 800 ± 1 000 13 000 ± 2 000	1 x 10^6	1 x 10^7
TGCAGTAGC GATATC CAGC ACGTCATCG C**A**ATAG GTCG	130 ± 20 200 ± 10	1 x 10^6	1 x 10^6
TGCAGTAGC GATATC CAGC ACGTCATCG C**C**ATAG GTCG	140 ± 30 63 ± 8	1 x 10^6	2 x 10^6
TGCAGTAGC GATATC CAGC ACGTCATCG C**G**ATAG GTCG	31 000 ± 4 000 9 400 ± 600	6 x 10^5	4 x 10^6
TGCAGTAGC GATATC CAGC ACGTCATCG G**T**ATAG GTCG	10 000 ± 1 000 4 500 ± 300	8 x 10^5	1 x 10^6
TGCAGTAGC GATATC CAGC ACGTCATCG T**T**ATAG GTCG	500 ± 50 270 ± 60	7 x 10^5	1 x 10^6
TGCAGTAGC GATATC CAGC ACGTCATCG A**T**ATAG GTCG	1 800 ± 300 830 ± 110	5 x 10^5	2 x 10^6

[a] k_{rel} values are the average of four experiments. [b] K_{ass} value is the result of eight titrations at different concentrations of the canonical substrate. [c] k_{rel} values are the result of a minimum of two experiments. [d] K_{ass} values are the result of a minimum of two titrations using the unphosphorylated substrate in competition with the phosphorylated canonical substrate and are accurate within at least ±30%.

compared with the canonical substrate is much higher than the reduction in binding strength. Therefore, binding contributes only in part to the overall discrimination seen for cleavage.

Interaction with Mismatch Substrates. Cleavage rates of mismatch substrates are very heterogeneous (Table 3; Figure 3). In general, cleavage occurs faster than in star substrates with one exception: GA**G**ATC is the most unfavorable substrate for *Eco*RV and nearly devoid of cleavage (7 orders of magnitude slower enzymatic activity). On the other hand, GCTATC is cleaved nearly as fast as the canonical substrate. The order of double-strand cleavage velocities is as follows. Upper strand: position 1, G >>>> T ≥ C ≥ A; position 2, A > C > T >> G; position 3, T >>> C >> A >> G. Lower strand: position 4, A >> T > C > G; position 5, T >> C ≥ A >> G; position 6, C >> T ≥ A ≥ G. (≥ indicates the same order of magnitude; >, >>, >>>, and >>>> correspond to a differences of 1, 2, 3, and 4 orders of magnitude, respectively.) This ordering of double-strand cleavage velocities reveals that pyrimidines as the incorrect base are more readily accepted than purines. Among the purines, adenine is better tolerated than guanine. Furthermore, a change in the lower strand leads to faster cleavage than a change in the upper strand with two exceptions: T and C substitutions of A at the second position are cleaved

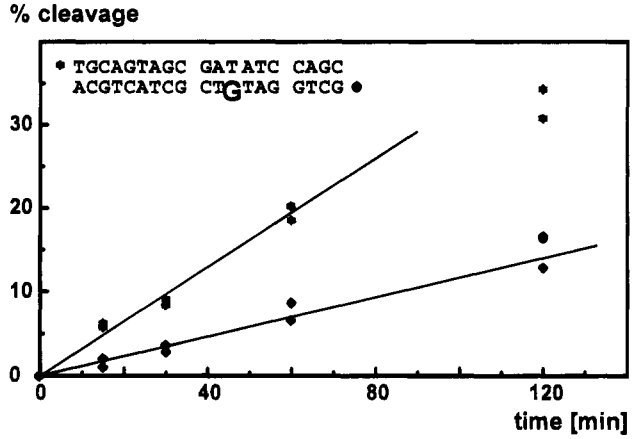

FIGURE 3: Example of a cleavage experiment. The kinetics of cleavage of 1 μM mismatch substrate indicated by 1 μM *Eco*RV restriction endonuclease was measured at 21 °C in cleavage buffer. The amount of cleavage of both single strands at each time point could be measured in one experiment because the products are separated by the denaturing homochromatography. The solid lines denote the initial cleavage rates deduced. For substrates with larger deviations in the nicking rates of the two single strands, experiments with different time scales were used for determinations of individual initial cleavage rates of both strands.

extraordinarily fast. Again, nicking rates of the individual single strands are different as for the star substrates.

Binding of wild-type *Eco*RV to mismatch substrates in the absence of Mg^{2+} is as weak as to the canonical substrate and has no correlation to cleavage rates. There may be a tendency for an enhanced binding of substrates with base changes in the lower strand, but these data are within the error margin.

Binding of the D90A mutant to mismatch substrates in the presence of Mg^{2+} is about 50 times lower than to the canonical substrate. An exception are the three substrates in which the A at the fourth position of the lower strand is substituted. Binding affinity of these substrates is only 2−6 times lower than that of the canonical substrate. In contrast to the star substrates, D90A binding strength does not reflect cleavage rates; e.g., the worst substrate, GA**G**ATC, is bound as poorly as GTTATC, which is cleaved 4−5 orders of magnitude faster.

DISCUSSION

We have investigated the interaction of the *Eco*RV restriction endonuclease with DNA substrates which deviate from the canonical sequence in one base pair (star substrates) or one base only (mismatch substrates). Oligodeoxynucleotides were chosen as substrates because the sequence context of the recognition sites which influences cleavage (Tayler & Halford 1989) is unvarying.

The cleavage rate of the canonical substrate is about twice as fast as the cleavage rates measured with shorter, self-complementary oligodeoxynucleotides [7 min^{-1} for AAA-*GATATC*TT, Fliess et al. (1988); 6.9 min^{-1} for GAC-*GATATC*GTC, Newman et al. (1990a,b)]. These differences are easily explained by the differences in length and sequence of the substrates compared. The binding constant of 6 × 10^7 M^{-1} is slightly smaller than the 4 × 10^8 M^{-1} measured for the AT-rich 20-mer CATTGTTA*GATATC*ATAAC by Thielking et al. (1992). This difference may originate from the influence of the sequence context on binding strength.

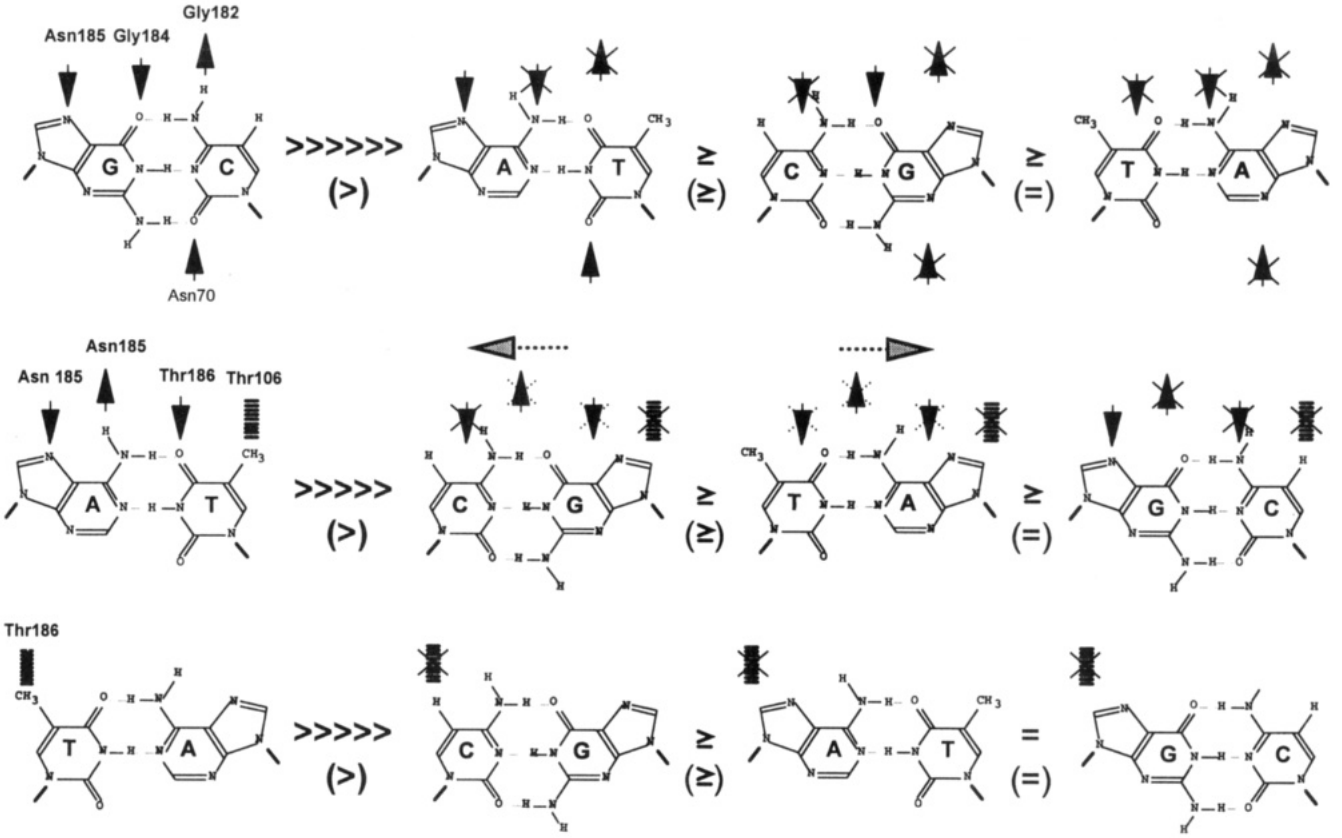

FIGURE 4: Disturbance of direct base contacts by changing base pairs in one half of the recognition sequence. *Eco*RV−DNA contacts as deduced from the X-ray structure analysis (Winkler et al., 1993; Kostrewa & Winkler 1995) are depicted in the left panel. Hydrogen bonds of the base pairs are symbolized by arrows, while hydrophobic contacts to amino acids are drawn as shadowed boxes. Invalidated contacts to noncanonical base pairs are marked by a cross (right panel). The horizontal arrows reflect the fact that already a slight movement of base pairs substituting for the AT base pair would create some of the required contacts. Comparisons of double-strand cleavage velocities (binding strength) resulting from a base pair change are indicated as follows: ≥ corresponds to the same order of magnitude; >, >>>>>, and >>>>>> correspond to differences of 1, 5, and 6 orders of magnitude, respectively (the exact data are in Table 2).

Furthermore, Thielking et al. determined the binding constant with electrophoretic mobility shift assays, which may lead to slight differences from the value measured with nitrocellulose filter assays here.

For all star substrates, we see an overall reduction in double-strand cleavage rate of at least 5 orders of magnitude. On this level, substitutions of the GC base pair are more readily discriminated than substitutions of the AT and TA base pairs. This seems to reflect the number of base contacts detected in the crystal structure. While the outer two base pairs are hydrogen bonded by four and three amino acids, respectively, the inner TA base pair is not contacted by hydrogen bonds at all. Furthermore, the methyl groups of the thymidines are located in a loose hydrophobic pocket consisting of Thr106, Thr186, and Asn188 (Wenz et al., 1994). A simple tally of possible remaining contacts to the changed base pair explains the order of cleavage velocities at the GC position but not at the AT or TA positions (Figure 4).

Although this analysis is straightforward, it does not take into account structural changes at the neighboring positions induced by the new base pair which may have influences on base or phosphate contacts, or both, at these positions. But even more important are the interconnections of most of the base contacts in the major groove, which are raised from the recognition loops of residues 182−186 in both subunits [Winkler et al., 1993; also discussed in Thielking et al. (1991)]. Hydrogen bonds to the GC base pair arise exclusively from the peptide backbone of residues Gly182, Gly184, and Asn185, while the AT base pair is contacted by hydrogen bonds from Asn185 and Thr186 and hydrophobic interactions with Asn188 and Thr106 (Wenz et al., 1994). The recognition loop of one subunit is mainly responsible for recognition of one GAT half of the sequence. The two loops are connected by bridging water molecules between the Asn185 residues of both subunits (Kostrewa & Winkler, 1995). Given this structural framework, it is easy to imagine that by influencing the loop geometry the exchange of one base pair disturbs not only enzyme contacts to its own position but also contacts to the neighboring base pair. This may even influence the recognition loop of the second subunit contacting the other half of the sequence via the interactions between the two loops. This interconnection of base contacts in the major groove correlates with the drastic overall reduction of cleavage rates of 5 orders of magnitude upon changing the GC or the AT base pair regardless of the nature of the new base pair.

The TA base pair does not form hydrogen bonds with the protein. Only a hydrophobic interaction with Thr186 is possible. Therefore, the strong discrimination observed for substrates with changes at this position can be described only in part by disturbance of direct contacts. Upon complex formation *Eco*RV induces a central kink of approximately 50° resulting in a narrower and deeper major groove (Winkler et al., 1993). This enables the described interconnection of the recognition loops and is necessary for a close approach

of the catalytic centers of both *Eco*RV subunits and the scissile phosphodiester bonds, which are located between the TA base pairs of both half-sites. The kink is stabilized by three center hydrogen bonds between the two central TA base pairs which may compensate the loss of stacking interactions. It is not clear how strong the impact of these hydrogen bonds is, but they are impeded if one of the two TA base pairs is changed. As a result, kinking becomes energetically more unfavorable and may disturb the cross-talk of the recognition loops or catalysis, or both.

Beyond direct contacts to the bases of the recognition sequence, phosphate contacts play an important role for strong binding and selectivity, as was shown for the *Taq*I restriction endonuclease (Mayer & Barany, 1994). In the crystal structure of *Eco*RV, several amino acid residues are in close vicinity to phosphate groups along the contact interface, which encompasses more than the recognition sequence (Winkler et al., 1993). Even more contacts are possible by bridging DNA phosphates and the protein by divalent cations. Two are obvious in the catalytic centers of both subunits, but further contacts are likely (Jeltsch et al., 1995). Since the overall geometry of base pairs is very similar, it is hard to estimate to what extent phosphate positions are changed by changing base pairs. Although it is well known that DNA structure varies with the sequence (Joshua-Tor & Sussman, 1993), we cannot predict the structure of a given sequence, especially when it is altered by complex formation with a protein. Therefore, one can only expect that disturbance of phosphate contacts may contribute to sequence discrimination.

It is rather tedious to explain the observed cleavage rates of all mismatch substrates. Although there is much X-ray and NMR data about the structure of base mismatches available, it does not cover all possible mismatches. Mismatch structure may also depend on the sequence context [for a review and references, see Joshua-Tor and Sussman (1993)]. Generally it is assumed that mismatched base pairs are stacked within the helix without global distortion of the DNA (Bhattacharyya & Lilley, 1989). But mismatches can also influence flexibility of DNA deduced from significantly faster base pair opening rates up to the third base pair flanking the mismatch (Patel et al., 1984). More flexible DNA can improve substrate proportion as shown for the Trp repressor (Mazarelli et al., 1992) and the *Eco*RI endonuclease (Lesser et al., 1993). This may explain the unexpected high cleavage rates for some mismatch substrates.

Binding strength of the *Eco*RV restriction endonuclease in the absence of Mg^{2+} does not tell anything about recognition. Under these conditions the recognition process of *Eco*RV is disturbed, and sequence selectivity is almost lost. This was already demonstrated by Taylor et al. (1991) for restriction fragments with many nonspecific sequences and verified here once more for sequences which in the presence of Mg^{2+} are recognized by the enzyme and are cleaved slowly. Therefore, we used the catalytically inactive mutant D90A to determine binding to the different substrates. Thielking et al. (1992) have demonstrated that DNA binding of this mutant in the presence of Mg^{2+} is specific for the recognition sequence, while binding to nonspecific sequences is weak. For the different substrates, variations in binding strength within the range of 2 orders of magnitude are found. Binding of star substrates correlates well with disturbance of direct base contacts as shown in Figure 4 and, therefore, roughly with cleavage rates. In the case of the mismatch substrates, the differences in binding vary by a factor of 100. But there is no correlation at all to the differences in cleavage rates. One should be aware that a catalytically inactive mutant provides only a rough understanding of the binding process of the wild-type enzyme. It might be possible that wild-type binding is stronger, since one amino acid coordinating a Mg^{2+} ion was mutated. On the other hand, it could also be weaker as a result of the negative charge of the canonical aspartate at this position. In line with the latter interpretation is the fact that the double mutant D74A/D90A (both negatively charged amino acids of the catalytic center exchanged for alanine) specifically binds even stronger to the canonical sequence than the D90A single mutant (Köhler et al., 1994), as does the triple mutant E45A/D74A/D90A (Jeltsch et al., 1995).

The nicking rates of the canonical substrate in both single strands are equivalent. This is an indication of consecutive cleavage of both strands. The situation is different for star and mismatch substrates. Here significant differences in nicking rates of both single strands are seen. The low level of cleavage rates results in a dissociation of the enzyme after the first cut and the need of reassociation before the second cut. This is even more pronounced with longer DNA substrates where the high concentration of nonspecific binding sites retards reassociation. Therefore, *in vivo* a ligase can repair the nicks before the appearance of a second cleavage event (Taylor et al., 1990). The importance of an active ligase protecting cells against *Eco*RV cleavage in other sites than the methylated canonical sites was already shown by Taylor et al. (1990). The same phenomenon of discontinuous cleavage of the two single strands was also seen in the *Eco*RI system in experiments analogous to the ones presented here (Lesser et al., 1990; Thielking et al., 1990). To explain this phenomenon, one should consider that changing one base pair or one base breaks the symmetry of the recognition site. The two subunits of the dimer are unable to develop the same symmetric contacts, and therefore, the two catalytic centers are differently activated. The order of cleavage of the two single strands is indicative of the cooperation of the two subunits in the recognition and cleavage processes. In the substrates investigated here the base changes all are located in one half-site of the palindromic sequence. A subunit which forms most of the contacts to this half-site cleaves the lower strand. Out of the 27 degenerate substrates tested, 24 are nicked faster in this lower strand. Two possibilities exist to explain this behavior: First, the disturbance of contacts between a subunit and the DNA may be favorable to trigger its own active center. However, considering the drastic reduction in cleavage rates, this is unlikely. Second, recognition of contact points by one subunit could trigger strand cleavage by the active center of the other subunit. It is not clear at the moment how this information transfer is mediated. The main interactions between the two subunits are hydrophobic contacts which are flexible, as the comparison of structures of *Eco*RV with and without DNA shows (Winkler et al., 1993).

The discrimination of *Eco*RV versus star substrates is much stronger than against most of the mismatch substrates. While star sequences occur in the normal background of cellular DNA, the enzyme is optimized to spare these sequences. Base mismatches, however, are very rarely

encountered. Furthermore, if a mismatch occurs in recombination intermediates or as a replication error, the preferred nicking by *Eco*RV will initiate repair of the mismatch, although nicking is not directed to the newly replicated strand. Therefore, evolutionary pressure was high to prevent cleavage of star sequences but not of mismatch sequences.

ACKNOWLEDGMENT

We thank Kornelia Zander for expert technical assistance and Edeltrud Schuchardt for typing the manuscript. We are grateful to Dr. Alfred Pingoud for critical comments on the manuscript, to Drs. Stephen Halford and Fritz Winkler for making available manuscripts prior to publication.

REFERENCES

Alves, J., Pingoud, A., Haupt, W., Langowski, J., Peters, F., Maass, G., & Wolff, C. (1984) *Eur. J. Biochem. 140*, 83−92.

Alves, J., Rüter, T., Geiger, R., Fliess, A., Maass, G., & Pingoud, A. (1989a) *Biochemistry 28*, 2678−2684.

Alves, J., Urbanke, C., Fliess, A., Maass, G., & Pingoud, A. (1989b) *Biochemistry 28*, 7879−7888.

Barany, F. (1988) *Gene 65*, 149−165.

Bhattacharyya, A., & Lilley, D. M. J. (1989) *J. Mol. Biol. 209*, 583−597.

Bougueleret, L., Scharzstein, M., Tsugita, A., & Zabeau, M. (1984) *Nucleic Acids Res. 12*, 3659−3676.

Fliess, A., Wolfes, H., Seela, F., & Pingoud, A. (1988) *Nucleic Acids* 4Res. *16*, 11781−11793.

George, J., & Chirikjian, J. (1982) *Proc. Natl. Acad. Sci. U.S.A. 79*, 2432−2436.

Halford, S., Lovelady, B., & McCallum, S. (1986) *Gene 41*, 173−181.

Heitman, J. (1993) in *Genetic Engineering* (Setlow, J. K., Ed.) Vol. 15, pp 57−108, Plenum Press, New York.

Izsvak, Z., & Duda, E. (1989) *Biochem. J. 258*, 301−303.

Jeltsch, A., Maschke, H., Selent, U., Wenz, C., Köhler, E., Connolly, B. A., Thorogood, H., & Pingoud, A. (1995) *Biochemistry 34*, 6239−6246.

Joshua-Tor, L., & Sussman, J. L. (1993) *Curr. Opin. Struct. Biol. 3*, 323−335.

Köhler, E., Wende, W., & Pingoud, A. (1994) in *Structural Biology: The State of the Art* (Sarma, R. H., & Sarma, M. H., Eds.) pp 217−224, Adenine Press, New York.

Kostrewa, D., & Winkler, F. K. (1995) *Biochemistry 34*, 683−696.

Lesser, D. R., Kurpiewski, M. R., & Jen-Jacobson, L. (1990) *Science 250*, 776−786.

Lesser, D. R., Grajkowski, A., Kurpiewski, M. R., Koziolkiewicz, M., Stec, W. J., & Jen-Jacobson, L. (1992) *J. Biol. Chem. 267*, 24810−24818.

Lesser, D. R., Kurpiewski, M. R., Waters, T., Connolly, B. A., & Jen-Jacobson, L. (1993) *Proc. Natl. Acad. Sci. U.S.A. 90*, 7548−7552.

Mayer, A. N., & Barany, F. (1994) *J. Biol. Chem. 269*, 29067−29076.

Mazzarelli, J. M., Rajur, S. B., Iadarola, P. L., & McLaughlin, L. W. (1992) *Biochemistry 31*, 5925−5936.

Nasri, M., & Thomas, D. (1986) *Nucleic Acids Res. 14*, 811−821.

Nasri, M., & Thomas, D. (1987) *Nucleic Acids Res. 15*, 7677−7687.

Newman, P. C., Nwosu, V. U., Williams, D. M., Cosstick, R., Seela, F., & Connolly, B. A. (1990a) *Biochemistry 29*, 9891−9901.

Newman, P. C., Williams, D. M., Cosstick, R., Seela, F., & Connolly, B. A. (1990b) *Biochemistry 29*, 9902−9910.

Patel, D. J., Kozlowski, S. A., Ikuta, S., & Itakura, K. (1984) *Fed. Proc. 43*, 2663−2670.

Roberts, R. J., & Halford, S. E. (1993) in *Nucleases* (Linn, S. M., Lloyd, R. S., & Roberts, R. J., Eds.) 2nd ed., pp 35−88, Cold Spring Harbor Laboratory Press, Cold Spring Harbor, NY.

Rosenberg, J. M., & Greene, P. (1982) *DNA 1*, 117−124.

Schildkraut, I., Banner, C. D. B., Rhodes, C. S., & Parekh, S. (1984) *Gene 27*, 327−329.

Selent, U., Rüter, T., Köhler, E., Liedtke, M., Thielking, V., Alves, J., Oelgeschläger, T., Wolfes, H., Peters, F., & Pingoud, A. (1992) *Biochemistry 31*, 4808−4815.

Taylor, J. D., & Halford, S. E. (1989) *Biochemistry 28*, 6198−6207.

Taylor, J. D., & Halford, S. E. (1992) *Biochemistry 31*, 90−97.

Taylor, J. D., Goodall, A. J., Vermote, C. L., & Halford, S. E. (1990) *Biochemistry 29*, 10727−10733.

Taylor, J. D., Badcoe, I. G., Clarke, A. R., & Halford, S. E. (1991) *Biochemistry 30*, 8743−8753.

Thielking, V., Alves, J., Fliess, A., Maass, G., & Pingoud, A. (1990) *Biochemistry 29*, 4681−4692.

Thielking, V., Selent, U., Köhler, E., Wolfes, H., Pieper, U., Geiger, R., Urbanke, C., Winkler, F. K., & Pingoud, A. (1991) *Biochemistry 30*, 6416−6422.

Thielking, V., Selent, U., Köhler, E., Landgraf, A., Wolfes, H., Alves, J., & Pingoud, A. (1992) *Biochemistry 31*, 3727−3732.

Thomas, M., & Davis, R. W. (1975) *J. Mol. Biol. 91*, 315−328.

Wenz, C., Selent, U., Wende, W., Jeltsch, A., Wolfes, H., & Pingoud, A., (1994) *Biochim. Biophys. Acta 1219*, 73−80.

Winkler, F. K., Banner, D. W., Oefner, C., Tsernoglou, D., Brown, R. S., Heathman, S. P., Bryan, R. K., Martin, P. D., Petratos, K., & Wilson, K. S. (1993) *EMBO J. 12*, 1781−1795.

BI9506381