Sequence Context Influencing Cleavage Activity of the K130E Mutant of the Restriction Endonuclease *Eco*RI Identified by a Site Selection Assay†

S. Windolph,‡ A. Fritz,§ T. Oelgeschläger,|| H. Wolfes,⊥ and J. Alves*,⊥

Zentrum Biochemie, Medizinische Hochschule Hannover, D-30623 Hannover, Germany, Anatomie, Technische Universität München, Biedersteiner Strasse 29, D-80802 München, Germany, Pathologie, Gesellschaft für Umwelt und Gesundheit (GSF), Ingolstädter Landstrasse 1, D-85758 Oberschleissheim, Germany, and Biochemistry & Molecular Biology Lab, Rockefeller University, 1230 York Ave, New York, New York 10021

Received January 13, 1997; Revised Manuscript Received May 14, 1997⊗

ABSTRACT: We have generated several *Eco*RI mutants which exhibit a decreased cleavage rate on one of the five specific cleavage sites in bacteriophage λ-DNA. To study the influence of the sequence context on the cleavage rate in more detail, we developed a site selection assay. From a complete set of 4096 plasmid substrates, differing in three bases on both sides of a recognition sequence, optimal (best cut) and unfavorable (worst cut) sequences were selected by repeated limited digestion, separation, and in vivo amplification of cleaved and uncleaved plasmids. In order to compare the sequence preferences of the inner arm mutant K130E and the wild type enzyme, the cleavage rates and sequences of individual plasmids from the resulting pools were determined. The inner arm mutant K130E selected pools with clearly defined consensus sequences and a high amount of palindromic sequences. The cleavage rates of the selected sequences are specific for the K130E mutant as is shown by their cleavage with other mutants. In contrast, wild type *Eco*RI does not lead to a selection in this assay. Pre-steady state kinetics show that preferences for a certain sequence context are a result of differences in the dissociation rates of the wild type enzyme. *Eco*RI is evolved to efficiently recognize and cleave each nonmethylated DNA invading the cell. Therefore, a fast dissociation after cleavage is not mandatory.

The *Eco*RI restriction enzyme cleaves double-stranded DNA within its recognition sequence GAATTC with high specificity in the presence of Mg^{2+}. For a review of structure and function of type II restriction endonucleases see Roberts and Halford (1993) and Pingoud and Jeltsch (1997). The cleavage rate of *Eco*RI at a specific site can be modulated by its sequence context. This was first reported by Thomas and Davis (1975) and investigated in more detail later by Halford et al. (1980) and Berkner and Folk (1983) for bacteriophage λ-DNA. Similar results were obtained also for bacteriophage T4 DNA (Goldstein et al., 1975), for adenovirus DNA (Forsblom et al., 1976), and for short oligodeoxynucleotides (Alves et al., 1984). Furthermore, variations in cleavage rate, depending on the sequence context, were also detected for other type II restriction enzymes like *Rsa*I, *Alu*I, *Hae*III (Majumber, 1989), *Bam*HI (Nardone et al., 1990a; Nardone et al., 1990b), *Eco*RV (Taylor & Halford, 1992), and *Nae*I (Yang & Topal, 1992).

For several mutants of *Eco*RI [P90A (Grabowski et al., 1995), I94W, Q115A (Jeltsch et al.,1993), K130E (Windolph & Alves, 1997), D133A, D135H, N141A, N141S (Fritz et al., manuscript in preparation), N149A, and R203K] we have detected an even more pronounced difference in cleavage rates for different cleavage sites. They all show a significantly decreased cleavage rate at the λ-site B. Surprisingly, the mutated amino acid residues do not share equivalent functions in the activity of *Eco*RI. Pro90 and Ile94 are close to the catalytic center without any direct participation in catalysis. Asp133 and Asp135 are in the inner arm and may possibly form a binding site for divalent cations (Windolph & Alves, 1997). Lys130, Asn149, and Arg203 may be involved in phosphate contacts. Gln115 and Asn141 establish direct contact to bases of the recognized DNA sequence (Rosenberg, 1991). Furthermore, the effect of a particular mutation is unpredictable, as the Q115E, K130A, or D135A mutant does not show a decreased cleavage rate at λ-site B.

For a detailed examination of this phenomenon we chose the K130E mutant. It is characterized by a high cleavage activity with a reduced Mg^{2+} optimum under standard buffer conditions. Differing from most other known *Eco*RI variants, binding is only possible in the presence of divalent cations like the catalytic inactive Ca^{2+} (Windolph & Alves, 1997). Two other mutants in the inner arm (K130A and R131E) show the same binding behavior. But only DNA cleavage by the K130E mutant is stimulated by the addition of low amounts of Ca^{2+} to the standard reaction buffer.

We established a site selection assay allowing us to test the influence of the three base pairs next to the recognition sequence on both sides. Best cut and worst cut sequences were selected.

MATERIALS AND METHODS

Enzyme Preparation. Mutagenesis was carried out using the gapped duplex method (Kramer et al., 1984) with site-specific mismatch primers for the introduction of amino acid substitutions. The wild type enzyme was purified by anion

† This work has been supported by a grant from the Deutsche Forschungsgemeinschaft (MA 465/17-1).

* Address correspondence to this author. Phone: +49-511-532-3671. FAX: +49-511-532-5966. E-mail: alves@bpc.mh-hannover.de.

‡ TUM.

§ GSF.

|| Rockefeller University.

⊥ Medizinische Hochschule Hannover.

⊗ Abstract published in *Advance ACS Abstracts,* July 15, 1997.

S0006-2960(97)00076-7 CCC: $14.00

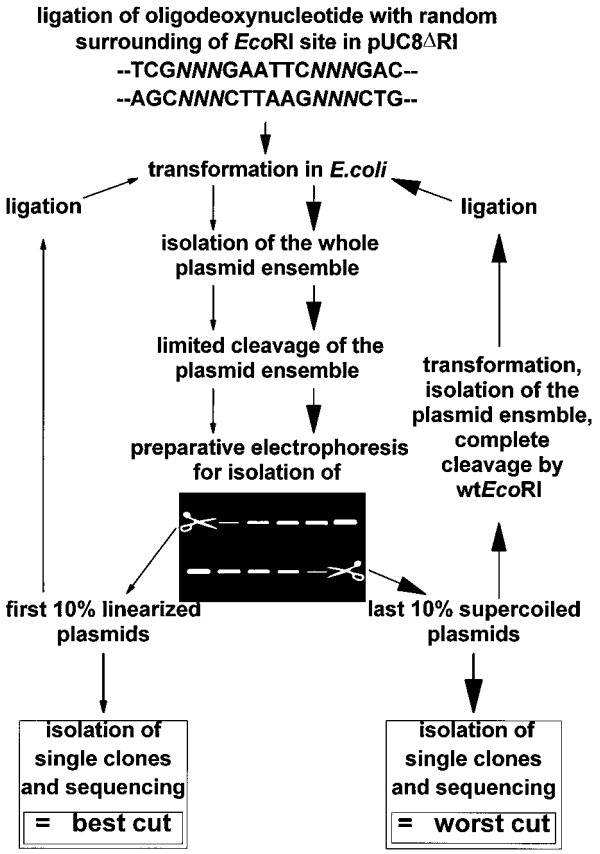

FIGURE 1: Site selection assay. The procedure of the selection of best cut sequences (left side, small arrows) and of worst cut sequences (right side, large arrows) is shown schematically.

exchange (D52, Whatman) and hydroxylapatite chromatography (Geiger et al., 1989). The K130E mutant, which carries a C-terminal His_6-affinity tag, was prepared using NiNTA-agarose (Qiagen) and phosphocellulose (Grabowski et al., 1995).

Site Selection Assay. The *Eco*RI site of the pUC8 plasmid was destroyed by *Eco*RI restriction and fill-in polymerization. The resulting product (pUC8ΔRI) was cleaved with *Hin*dIII and *Pst*I and ligated with a 51 bp oligodeoxynucleotide containing one *Eco*RI restriction site flanked by three random bases on each of its 5′ and 3′ sides. While most of the transformed cells were left in liquid culture, an aliquot was plated to determine the multiplicity of this culture. Transformation resulted in about 500 000 clones. The resulting plasmid pool for the site selection assay pUCEcoRND (2708 bp) consists of 4096 different substrates (Figure 1).

A 400 ng/15 μL solution of this plasmid DNA was subjected to a limited digestion by (1) 0.3 nM wild type *Eco*RI in standard buffer (20 mM Tris/HCl, pH 7.5, 50 mM NaCl, and 10 mM $MgCl_2$) or (2) 2.3 nM K130E mutant in a standard buffer with a Mg concentration reduced to 5 mM. The first 10% of the substrate, cleaved after a short reaction time, was prepared quantitatively from 0.8% agarose gels and taken for the enrichment of best cut substrates. The linear plasmid DNAs were ligated and transformed into LK111λ cells. After a long reaction time the last 10% of the supercoiled plasmids were prepared from agarose gels as worst cut substrates. These supercoiled plasmids were amplified in LK111λ cells, purified as a pool and completely cleaved with the wild type enzyme to exclude substrates missing an *Eco*RI restriction site due to a coincidental mutation. The linear plasmids were prepared from an agarose gel, ligated and transformed into *E. coli* cells. Each of these transformations resulted in 30 000−50 000 clones. For an efficient selection it was necessary to use several rounds of enrichment.

The resulting pools were sequenced with pUC8 primer d(GCCAGGGTTTTCCCA), and the amount of each base at all flanking positions was determined. The final best cut and worst cut pools were retransformed in recA deficient HB101 cells to avoid catenane formation. Single clones were sequenced and characterized kinetically.

Cleavage velocities of the K130E mutant found for best cut and worst cut sequences deviate by a factor of about 100. In order to avoid very long reaction times, worst cut cleavage experiments with the K130E mutant were carried out at 37 °C. The temperature was reduced to 15 °C in best cut cleavage studies with the mutant to keep a constant ratio of enzyme and substrate.With pUC8 plasmid DNA as substrate, a 10-fold difference of cleavage rates between these two reaction temperatures was determined. This factor was used to normalize the cleavage rates of the best cut sequences measured at 15 °C to 37 °C.

Pre-Steady State Kinetics. Pre-steady state kinetics measurements for the wild type enzyme were performed in a pulsed quenched-flow apparatus at 20 °C (Langowski et al., 1981; Urbanke, 1982). 9 nM selected plasmid substrates and 180 nM wild type *Eco*RI in standard buffer were mixed. After 10 msec to 100 s the reaction was quenched by addition of 0.1 M EDTA, pH 8.5. The mixture was treated with proteinase K (100 μg/mL end concentration in 10 mM Tris/HCl, pH 7.8; 5 mM EDTA; 0.5% SDS) for 1 h at 37 °C to digest the enzyme for better resolution of the DNA in the gel electrophoresis. DNA was then precipitated with EtOH at a final concentration of 70%. The pellet was lyophilized and dissolved in 15 μL of H_2O. 5 μL of agarose gel loading buffer (0.1 M EDTA; 25% Ficoll; 0.1% bromphenol blue; 0.1% xylene cyanol; pH 8.0) was added. The reaction products were separated on 0.8% agarose gels and quantified with a Herolab gel documentation system and EASY software. For the K130E mutant pre-steady state kinetics were performed under exactly the same conditions but the reaction times were much longer (12 s to 2 h).

Stopped-Flow. Stopped-flow experiments were performed at 20 °C in a modified version of a Durrum-Gibson stopped-flow apparatus. Protein fluorescence was recorded as described previously (Alves et al., 1989), in order to measure pre-steady state as well as steady state kinetics in a single experiment. Experiments were carried out with 0.5−4 μM enzyme and 0.5−16 μM oligodeoxynucleotides which resemble sequences found to be cleaved quickly d(AAA GAATTC GCA) and slowly d(ACA GAATTC CAG) by wild type *Eco*RI.

Oligodeoxynucleotide Cleavage with the K130E Mutant. The 18-mer double-stranded oligodeoxynucleotides were synthesized using β-cyanoethyl phosphoamidites (MWG Biotech, Ebersberg, Germany) on a Milligen Cyclone DNA synthesizer. They represent a best cut sequence d(TTCG AAA GAATTC TCT GA) and a worst cut sequence d(TTCG TTC GAATTC CAG GA). To study influences from each half-site, oligodeoxynucleotides consisting of one best cut and one worst cut flanking half-site d(TTCG AAA GAATTC CAG GA) and d(TTCG TTC GAATTC TCT GA) were also synthesized. Both strands of the asymmetric substrates were

Table 1: Sequence Context of the Five *Eco*RI Sites in Bacteriophage λ-DNA

site	position	sequence
A	21226	GG TGA GAATTC GGC CT
B	26104	GA AAT GAATTC TAA GC
C	31747	GA AGT GAATTC AAA CA
D	39168	TC AGA GAATTC TGG CG
E	44972	GT CCT GAATTC ATT AG

5′ labeled using [γ-^{32}P]ATP and polynucleotide kinase. Oligodeoxynucleotides (0.1−1 μM) were cleaved with 0.1 μM K130E mutant at 37 °C in (20 mM Tris/HCl, pH 7.5, 10 mM $MgCl_2$, and 50 mM NaCl). The substrates and the cleavage products of both single strands (8- and 6-mer) were separated by homochromatography at 65 °C (Jay et al., 1972) on DEAE cellulose sheets (Polygram CEL 300 DEAE/HR-2/15, Macherey & Nagel). Dried cellulose sheets were documented with a Fuji BAS1000 PhosphorImager and quantitatively analyzed by MacBAS 2.0 (Fuji).

RESULTS

Cleavage Activity of the K130E Mutant. Lys130 was mutated to study the importance of possible phosphate contacts for cleavage activity of the *Eco*RI restriction endonuclease. Like other *Eco*RI mutants, it cleaved most slowly at λ-site B which resulted in a 100-fold reduced specific cleavage activity (Windolph & Alves, 1997). An inspection of the five recognition sites in λ-DNA (Table 1) shows no obvious reason for differences in cleavage rate. The 16-mer oligodeoxynucleotides comprising λ-sites B and E also show quite large differences in cleavage rates for the mutants tested (data not shown), which rules out influences from sequences far away as being the cause for the modulation of cleavage rates. Therefore, we developed the site selection assay described below to systematically study preferences for flanking sequences.

Site Selection Assay with the K130E Mutant. The pUCEcoRND plasmid pool was generated, which represents a set of 4096 different substrates deviating systematically in three base pairs on both sides of the single *Eco*RI recognition sequence. Several rounds of limited cleavage were used to select for best cut and worst cut substrates (Figure 1). The first 10% of the cleaved product was isolated and amplified to enrich best cut substrates. The cleavage rate of the reduced plasmid pool increased regularly during seven rounds of selection. Two additional rounds were tested but only a very small further increase of cleavage rate was recorded. The 7th pool (termed best cut pool) was chosen for a detailed examination of isolated plasmids, because an additional selection resulted in a loss of sequences rather than in an enhancement of cleavage rate.

The last 10% of uncleaved supercoiled plasmid was utilized to generate worst cut substrates. To remove plasmids missing an *Eco*RI recognition site, the isolated supercoiled plasmid DNA was cleaved by wild type *Eco*RI. The linear form was isolated, re-ligated, and transformed in *E. coli* cells. The amount of enrichment of the resulting pools was tested

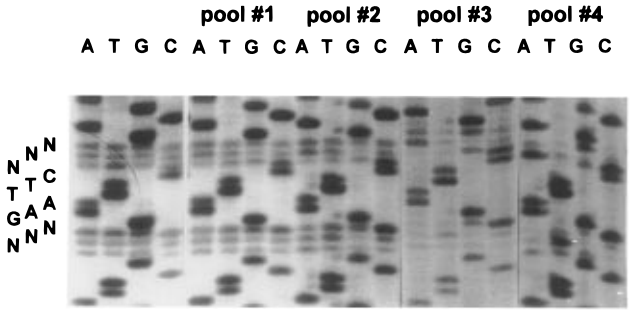

FIGURE 2: Sequencing of plasmid pools during worst cut selection. Starting with the random pool pUCEcoRND on the left side the sequences for the four worst cut selection pools are shown. At the positions marked with N on the left margin the enrichment from nondefined to defined sequences can be traced.

kinetically. After four rounds of selection no further decrease of cleavage rate for worst cut sequences was detected.

Sequencing of all pools obtained during the selection gave a first insight into the change of sequence spectrum compared to the pUCEcoRND pool (Figure 2). The distribution of each base at the three positions flanking the recognition sequence was documented with a Fuji BAS1000 PhosphorImager and quantitatively analyzed by MacBAS 2.0 (Fuji). The intensity of single bases was internally standardized by the intensity of a base in the recognition sequence. This resulted in the percentage of each base at each position shown in Table 2 (upper part). pUCEcoRND was confirmed as a homogeneous starting pool with 25% of each base at all flanking positions.

Cleavage by the K130E mutant resulted in rate constants of 3.2 min^{-1} for the best cut and 0.02 min^{-1} for the worst cut pool, compared with 0.6 min^{-1} for the whole ensemble of pUCEcoRND. Subsequently, individual plasmids isolated from clones of both pools were selected for cleavage rate determination. The 100 examined best cut plasmids varied in cleavage rate between 2 and 6 min^{-1} while the 100 worst cut plasmids showed cleavage rates between 0.01 and 0.06 min^{-1}. This confirms the homogeneity in cleavage rate of the selected pools.

After 73 of the kinetically characterized best cut plasmids had been sequenced, it was noted that 11 sequences were found twice and two sequences were found three times (Table 3). Therefore, further sequencing was discontinued. The number of sequences found more often than once was used to estimate the total number of different sequences in the pool. Under the assumption of a homogeneous plasmid pool the probabilitiy of obtaining distinct numbers of 2- and 3-fold detections of the same sequence was determined by a computer simulation. We used a random experiment in which samples of size 73 were repeatedly drawn out of pools of various sizes n. For a pool size of 150 sequences exactly eleven 2-fold and two 3-fold drawn sequences were found. Therefore, one should expect the best cut pool to consist of approximately 150 different sequences. In comparison, sequencing of 40 worst cut sequences (Table 3) resulted in five sequences found twice, one 3-fold, one 5-fold, and one six-fold. Using the same computer simulation as above, the resulting number of sequences was distinctly less than the 24 different sequences already determined. This implies that the sequences are probably not equally distributed in the pool, which is a basic requirement of the computer simulation. With no indication of the relative amounts of sequences in

Table 2: Proportion of the Occurrence of Bases at Positions Flanking the Recognition Sequence and the Resulting Consensus Sequences[a]

N−3	*N*−2	*N*−1	*N*+1	*N*+2	*N*+3	% base	*N*−3	*N*−2	*N*−1	*N*+1	*N*+2	*N*+3
Best Cut Pool Sequencing							Worst Cut Pool Sequencing					
41	36	44	27	16	23	A	25	15	9	8	44	29
17	13	22	40	19	23	T	19	38	17	6	10	9
13	19	8	13	14	21	G	32	14	12	17	18	41
29	32	26	20	51	33	C	24	33	62	69	28	21
A/c	A/C	A/c	T/a	C	N	consensus sequence	N	T/C	C	C	A/c	G/a
Sequences of Best Cut Single Plasmids							Sequences of Worst Cut Single Plasmids					
41	31	56	42	15	22	A	29	15	0	0	81	17
19	17	22	41	17	32	T	39	71	0	0	12	10
20	43	5	10	31	22	G	15	2	15	27	0	68
20	9	17	7	37	24	C	17	12	85	73	7	5
A	A/G	A	A/T	G/C	N	consensus sequence	T/a	T	C	C	A	G

[a] In the upper part the percentage of each of the four bases at the indicated positions on the left (*N*−) or on the right (*N*+) side of the recognition sequence was quantified from sequencing the whole pools as described in the text. In the lower part the occurrence of the four bases at the flanking positions in the individual sequenced plasmids was used for calculating the proportions of each base on a percentage basis for comparison. Capital letters in the consensus sequences represent the main base at this position, and lower case letters represent the subordinate bases.

Table 3: Sequences Flanking the Recognition Sequences in Plasmids Isolated from the Best Cut and the Worst Cut Pools[a]

best cut pool sequences				worst cut pool sequences			
N+1 = A	*n*	*N*+1 = T	*n*	*N*+1 = A	*n*	*N*+1 = T	*n*
AAA...AAA		TGC...TAA	2	none		none	
AAA...AAG	2	AGA...TAT					
TAT...AAT		CGA...TAT	2	*N*+1 = G	*n*	*N*+1 = C	
TAT...AAC		AGA...TAG		TAC...GAA		AGG...CAG	
ACT...AAT	2	ATA...TTA		CCC...GAA		ATG...CAG	2
GGT...ATA		GGG...TTA		ACC...GAA		GAG...CAG	
AAA...AGG		AAA...TTT		CTC...GAT		CAC...CAG	
GAA...AGT		AGA...TTT		TTC...GAT	2	ATC...CAG	6
CAA...AGT		AGT...TTT	2	TTC...GAG		TTC...CAG	5
TGA...AGT		GGT...TTT		GTC...GAG		GTC...CAG	2
ATA...AGG		GAC...TTG		TTC...GTA	2	CTC...CAG	3
AGA...AGG		CGA...TTC				TCC...CAG	2
AAG...AGG	2	CTT...TTC				ATC...CAT	
ATA...AGC		TGC...TGT	2			TAC...CTG	
AGA...AGC		TGC...TGG				CTC...CTG	
GGT...AGC		ATA...TGC				ACC...CTG	
CAA...ACA		GGC...TGC				ATC...CCA	
TTT...ACA		TGA...TCA				GTC...CCC	
AGA...ACT		GCT...TCA	3			TTC...CCC	
CTG...ACT	2	GCT...TCT					
GGC...ACC		AAA...TCT					
CAA...ACC		AAA...TCC					
ATA...ACC		TAC...TCC					
TAA...ACC							
N+1 = G	*n*	*N*+1 = C	*n*				
GTA...GCA		GAA...CGT	2				
AGT...GCA	2	ACA...CGG					
TGC...GCA	3	CGC...CCG					
ACT...GCT							
AGC...GCT							
CTA...GCG							

[a] Sequences are given in 5′ → 3′ orientation according to the standard numbering of pUC8 DNA. They were arranged in the order A, T, G, and C at position *N*+1 which is one of the more decisive positions. *n* represents multiple independent occurrences of the same sequence.

Table 4: Proportion of Complementary Sequences in the Different Pools[a]

flanking complementary positions	starting pool (*n* = 4096)		best cut pool (*n* = 56)		worst cut pool (*n* = 24)	
	number	%	number	%	number	%
one (*N*1 or *N*2 or *N*3)	1728	42.2	17	30.4	11	45.8
two (*N*1/*N*2 or *N*2/*N*3 or *N*1/*N*3)	576	14.1	9	16.1	9	37.5
three (*N*1/*N*2/*N*3)	64	1.6	4	7.1	0	0
total	2368	57.8	30	53.6	20	83.3

[a] The proportions of complementary sequences in the starting pool resulted from a probability calculation. For the example with one flanking complementary position: the chance to find complementary bases at position *N*1 is 0.25 multiplied with the chances to have simultaneously noncomplementary bases at positions *N*2 and *N*3 (0.75 each) times the total number of sequences: $0.25 \times 0.75 \times 0.75 \times 4096 = 576$. The sum at all three postions is 1728. For the best and worst cut pools the individual sequences were taken only once regardless how often they were found during sequencing.

the pool, we can only estimate that more than half of the sequences in the pool have been determined and that, therefore, the pool contains roughly 40 different sequences.

The best cut sequences found for single clones show a high preference for AT base pairs at positions directly flanking the recognition sequence (*N*−1 and *N*+1). G or C bases prevail as GC or CG dinucleotides at positions *N*−2/*N*−1 and *N*+1/*N*+2. In contrast, in worst cut plasmids G and C bases are mainly located at positions N−1 and *N*+1. However, the combination GC or CG at *N*−2/*N*−1 and *N*+1/*N*+2 is not detected. Furthermore, a characteristic feature of 50% of the worst cut plasmids is a CAG trinucleotide at the 3′ side of the recognition sequence. Consensus sequences were deduced from single plasmid sequences from the worst cut and best cut pools (Table 2). As expected, these consensus sequences show only small differences compared with those determined from sequencing the whole pools.

Another characteristic feature of both pools selected by the K130E mutant was the preference for complementary flanking sequences. Taking the sum of all possible single, double, or totally palindromic sequences at the three positions, the starting pool already possesses 57.8% complementary sequences. This number is roughly equivalent to the 53.6% of the best cut single sequences and increases significantly to 83.3% for worst cut single sequences (Table 4). In best cut plasmids, the proportion of palindromic sequences increased distinctly compared to the starting pool for 12 bp long palindromes with complementary bases at all three positions. In contrast, complementary bases at only one of the three positions are under-represented, mainly due to a drop at the *N*2 position. In worst cut plasmids 10 bp palindromic sequences were favored, but no 12 bp palindromic sequences were detected. The latter observation is

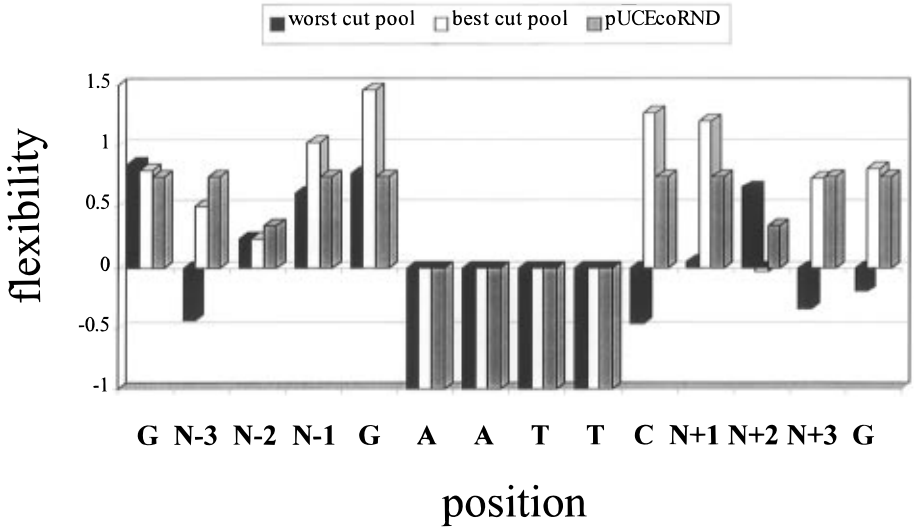

FIGURE 3: Comparison of the flexibility of best cut and worst cut sequences using the algorithm by Satchwell and Travers (1989). Flexibility scores were computed for trinucleotides of single sequences out of the three pools. The mean values of these are shown at the position of the nucleotide in the middle of the trinucleotide used for calculation. Positive values correspond to flexibility while negative values correspond to rigidity.

not significant given the low number of sequences in the pool.

Structural Parameters of Flanking Sequences. From the co-crystal structure of *Eco*RI (Kim et al., 1990) it is obvious that no direct base contacts exist outside the recognition sequence. In order to understand the modulating effect of the flanking sequences, we evaluated a variety of algorithms describing DNA structural parameters. These algorithms rely on the local base conformation in regard to stacking energies, flexibility, or the positioning of base pairs to each other. The nucleotide conformation varies with the sequence context. Therefore, these descriptions focus basically on next-neighbor interactions. Datasets for stacking energies based on melting experiments (Delcourt & Blake, 1991; Santa Lucia et al., 1996) were taken for dinucleotide analysis. The stacking energies differ only slightly to the starting pool and do not change significantly for sequences from the worst cut and the best cut pool.

The algorithms developed by Satchwell and Travers (1989), which are based on the nucleosome positioning, and by Brukner et al. (1995), based on DNA bending by DNase I, were used to estimate the flexibility of the flanking sequences. Both algorithms assign flexibility scores to the nucleotide in the middle of trinucleotides. The recognition sequence shows a relative rigidity, while the flanking positions of sequences of the best cut pool are flexible (Figure 3). The highest flexibility prevails at positions directly flanking the recognition sequence. Sequences of the worst cut pool show less flexibility at flanking positions except for position *N*+2.

Site Selection Assay with the Wild Type EcoRI. As a control, the site selection assay was also performed with the wild type enzyme, although the cleavage rates at specific recognition sites differ only about 10-fold (Alves et al., 1984). After seven rounds of selection for the best cut pool and four rounds for the worst cut pool, cleavage rates under steady state conditions in standard buffer differed only by this factor of 10 between both pools. However, isolated plasmids showed randomly distributed cleavage rates, regardless of whether they originated from the best cut or worst cut pool. This result was verified in a second independent selection.

Pre-Steady State Kinetics. Selected plasmids from different pools were cleaved by a 16-fold excess of enzyme over DNA. The K130E mutant shows the same 100-fold difference in cleavage rates for best cut and worst cut substrates as detected in steady state experiments. Therefore, this difference depends neither on temperature, which was 20 °C in these experiments, nor on dissociation rate. In contrast, quenched-flow kinetics for the wild type enzyme demonstrated that out of all sequences tested, the intrinsic cleavage rates of the first and the second single-strand cut were identical within the limits of error. Stopped-flow kinetics for selected oligodeoxynucleotides corresponding to sequences from different pools also show identical association rate constants, while the dissociation rate differed roughly by a factor of 10, as found also in steady state experiments. It has to be pointed out that differences in dissociation rates are not selected in this site selection assay. Therefore, the resulting pools from selection with wild type *Eco*RI accidentally harbor selected clones with a random distribution of steady state cleavage rates.

Comparison of Cleavage of Plasmid Substrates by Wild Type EcoRI and Several Mutants. Steady state cleavage rates were determined for the selection pools as well as isolated single plasmids with wild type *Eco*RI and the mutants P90A, K130E, N141A, and N141S under buffer conditions optimal for each enzyme (Table 5). All five enzymes cleaved the best cut pool faster than the worst cut pool. However, only the K130E mutant, which was used to generate the pools, cleaves the isolated plasmids with comparable cleavage rates. As expected, the wild type enzyme also shows fast cleavage rates for the plasmids out of the worst cut pool, as does the N141A mutant. The P90A mutant appears to exhibit a preference similar to the K130E mutant, although its cleavage rates differ merely by a factor of 4. One plasmid is cleaved remarkably fast by the N141S mutant. Although the four *Eco*RI mutants cleave the λ-site B more slowly than the other sites, their preferences for flanking sequences are not equivalent and differ for each mutant individually.

Table 5: Cleavage of Selected Plasmid Substrates by Wild Type *Eco*RI and Several Mutants Which Cleave Slowly at λ-Site B[a]

	wild type *Eco*RI	P90A	K130E	N141A	N141S
pUCEcoRND	5.4	0.8	0.6	2.3×10^{-3}	0.1
best cut pool	6.1	0.8	3.2	2.3×10^{-3}	0.5
ACT GAATTC GCT	2.1	0.6	3.5	2.6×10^{-3}	5.0
GGT GAATTC AGC	1.5	0.6	5.4	4.3×10^{-3}	0.5
worst cut pool	0.4	0.6	0.02	7.2×10^{-4}	0.3
TCC GAATTC CAG	4.9	0.2	0.05	8.7×10^{-4}	0.1
ATG GAATTC CAG	2.2	0.2	0.06	1.4×10^{-3}	0.1

[a] Cleavage rates are given for double strand cleavage in min^{-1}. Optimal buffer conditions were used for each enzyme.

Cleavage of Oligodeoxynucleotides with Varying Flanking Sequences by the K130E Mutant. Oligodeoxynucleotide substrates allow the detection of cleavage products of each end-labeled single strand and exclude effects of linear diffusion and unspecific interaction with the substrate on cleavage velocity. Therefore, we designed 18-mer double-stranded oligodeoxynucleotides which represent the consensus sequences for the best cut pool and the worst cut pool which were also found as isolated plasmid sequences. To study the influence of the 3′ and 5′ flanking sequences individually, we also synthesized oligodeoxynucleotides with 5′ best and 3′ worst half-sites and vice versa (Table 6).

The cleavage rate of the best cut oligodeoxynucleotide corresponds to the values determined for the plasmid substrate with identical flanking sequences. In contrast, the worst cut oligodeoxynucleotide was not cleaved within a time span of 24 h, although the enzyme was in excess. Longer incubation times were not possible because cleavage activity of the enzyme started to diminish. The wild type enzyme cleaved both substrates with nearly the same fast cleavage rate. Given that 1% cleavage would be detectable by the assay used, the factor of 600 between the cleavage rates of best and worst cut sequences in plasmid substrates increased to more than 2.5×10^6 with oligodeoxynucleotides. The hybrid substrates with half-sites from both selection pools show the differing importance of the 5′ and 3′ half-site for cleavage rates. Both worst cut half-sites are necessary for the very slow cleavage rate detected with the worst cut oligodeoxynucleotide and the 3′ half-site is more effective in diminishing cleavage activity than the 5′ half-site.

DISCUSSION

During the last five years many *Eco*RI mutants were produced in our lab to study completely different phenomena, like catalysis, base contacts, or phosphate contacts. Several of these mutants showed a decreased cleavage rate at λ-site B in comparison to the other λ-sites. The λ-sites differ in the sequence context of the recognition sequence GAATTC (Table 1). Therefore, differences in cleavage rate seem to be a more general phenomenon that depends on the DNA structure given by the λ-site B sequence (AAT GAATTC TAA). To determine DNA sequences, which exhibit positive or negative influences on cleavage activity of *Eco*RI mutants in general, we decided to develop the site selection assay described here. In this assay, a complete set of 4096 plasmid substrates differing in three base positions directly flanking the *Eco*RI recognition sequence on both sides was used for a limited digestion by the mutants. In several rounds of selection the first 10% cleaved and the last 10% of uncleaved plasmids were isolated for the enrichment of best cut and worst cut substrates. The resulting plasmid pools were characterized by cleavage rate and sequence determination for isolated plasmids.

The selection with the K130E mutant was accomplished successfully. A pool of 150 best cut sequences was cleaved ten times faster than the random starting pool. About 40 worst cut sequences were selected by the assay. They were cleaved 60-fold slower than the starting pool. Cleavage rates of individual plasmids vary by a factor of 2−3 from the cleavage rate of the whole pool, which confirms the homogeneity of the pools. The pool sizes, the number of selection rounds necessary to aquire the pool, and also the deviation of cleavage rate from that of the random pool demonstrate that retardation of cleavage is a more characteristic feature of only a low number of sequences, while most sequences are cleaved reasonably fast. This is confirmed by the finding that cleavage of oligodeoxynucleotide substrates is delayed only if both half-sites of the flanking sequences fit to the worst cut consensus sequence.

The wild type *Eco*RI does not result in any meaningful accumulation of flanking sequences in the site selection assay. Although cleavage rates of the selection pools deviated by a factor of 10, examination of the individual sequences showed that these pools were grouped accidentally. The differences in cleavage rates are based on different dissociation constants, while the initial cleavage rates of all sequences are identical. This was demonstrated by quenched-flow experiments with many plasmid substrates. It was verified in stopped-flow experiments with selected oligodeoxynucleotides that the association rate was identical, too. In contrast, the K130E mutant also exhibits the differences in cleavage rate between best cut and worst cut sequences under pre-steady state conditions. Dissociation as the rate-determining step in the cleavage reaction catalyzed by wild type *Eco*RI was already shown by Rubin and Modrich (1978). The site selection assay described here cannot select dissociation rates because the cleavage products used for the selection are isolated from agarose gels after denaturation of the enzyme by SDS and electrophoresis. Therefore, independent of the time the enzyme requires for dissociation, once a plasmid is cut, it joins the best cut pool and not the worst cut pool. A slow dissociation rate only diminishes the amount of enzyme engaged in the multiple turnover of the cleavage reaction. In our selection assay nicked sequences are sorted neither to best cut nor to worst cut sequences. Substrates, which admit only one phosphodiester cleavage, would be an interesting target for further investigation. However, these substrates are almost inaccessible as long as they are not accumulating because, at each time point used for selection, cleavage may be stopped accidentally after the first single cut.

The selection carried out for the K130E mutant resulted in best cut and worst cut sequences with clearly defined consensus sequences. The amino acid residue Lys130 contacts the phosphate 3′ of the C of the recognition sequence possibly not by a hydrogen bond but by a charge interaction. It is conceivable that a glutamic acid at this position is unable to make this contact. The DNA binding capacity of this mutant depends on divalent cations, which are suspected to bind in close vicinity (Windolph & Alves, 1997). Like the other mutants, which show more or less stringent dependence

Table 6: Cleavage of 18-mer Oligodeoxynucleotide Substrates by Wild Type *Eco*RI and the K130E Mutant[a]

	sequences	k (min^{-1})	
		K130E mutant	wild type *Eco*RI
best cut sequence	TTCG AAA GAATTC TCT GA	5.8	4.6
worst cut sequence	TTCG TTC GAATTC CAG GA	$< 2.3 \times 10^{-6}$	8.8
5′ best cut half-site	TTCG AAA GAATTC CAG GA	0.3	8
5′ worst cut half-site	TTCG TTC GAATTC TCT GA	1.9	3.6

[a] Cleavage of both single strands of the double-stranded substrates was simultaneous; therefore, the rate constants of double-strand cleavage are shown.

of cleavage activity on surrounding sequences, the K130E mutant also has no newly formed capacity of direct contacts to bases outside the recognition sequence. The most plausible explanation for the analyzed sequence effects is that some properties of the DNA structure itself influence the cleavage rates of *Eco*RI mutants. In the wild type enzyme these influences come into play during dissociation only after both single strands are cleaved. Is wild type *Eco*RI able to actively block any influence from outside? The loss of a phosphate contact directly outside the recognition sequence in the K130E mutant supports this view, as does the R203K mutation, which disturbs the corresponding contact to the phosphate 5′ to the G in the other strand and also shows retardation of cleavage at λ-site B. However, mutations of base contacts (Q115A, N141A, N141S) or mutations next to the catalytic center (P90A, I94W) will not directly influence phosphate contacts, but all of these mutations may well impair conformational changes necessary for the recognition and for the triggering of the cleavage reaction.

As is known from the cocrystal structure of *Eco*RI restriction endonuclease in the complex with its recognition sequence, both the conformation of the DNA substrate and of the enzyme are changed in the specific complex (induced fit). Compared to the structure of unbound DNA, *neokinks* at both ends of the recognition sequence and a central unwinding by 30° are induced by the protein (McClarin et al., 1986; Thomas et al., 1989). The energy cost of these distortions of the DNA may vary depending on the surrounding sequences (Lesser et al., 1993). The wild type enzyme compensates these energetic costs by establishing a network of numerous contacts within the recognition sequence and further phosphate contacts. Disturbance of this network impairs the ability to counterbalance unfavorable influences from sequence context. Therefore, it is plausible that sequences leading to high energy costs will cause a notable reduction of the cleavage rate of the K130E mutant as detected for the worst cut sequences out of the selection.

In order to find out whether a general feature characterizes the sequences found in the site selection assay, we tested different algorithms describing next neighbor DNA structural parameters for differences between best cut and worst cut sequences according to base stacking (Delcourt & Blake, 1991) and flexibility (Satchwell & Travers, 1989). The recognition sequence is characterized by low stacking energies and a relative rigidity. Stacking energies of the flanking sequences show no significant differences between the starting pool and both enrichment pools. Therefore, stacking energies seem not to be a structural feature of the DNA which modulates enzyme activity. The best cut sequences seem to be more flexible than the starting pool and the worst cut sequences, except at the position $N+2$. The position $N+2$ also showed distinct differences in complementarity between both pools. The differences in flexibility are more pronounced at the 3′ half-site, which also had more effect on cleavage rate of the oligodeoxynucleotide substrates. This leads to a hypothesis which implies that the induction of *neokinks* is disturbed for the mutant enzyme. Therefore, flexibility of the sequences next to the recognition sequence is favorable in advancing into the transition state. For *Eco*RV a similar dependence on flexibility of flanking sequences was detected for the cleavage at the star site GTTATC (Taylor & Halford, 1992). Although only a few sequence variations were made, DNA flexibility up to four base pairs outside seemed to be necessary for fast cleavage. Again, cleavage of the canonical sequence by *Eco*RV did not depend on the sequence context. To observe this effect, the recognition complex must be thrown out of balance by using a noncanonical star site as substrate for wild type *Eco*RV, while, in this study, flanking sequences showed influence on the cleavage of canonical sequences by *Eco*RI mutants. It is hard to decide whether rigidity is the main cause of slow cleavage in both systems. Other, perhaps subtle, differences in the DNA structure may be important, but they are hard to deduce from the given sequences (Hunter, 1996).

ACKNOWLEDGMENT

We thank Dr. G. Maass for stimulating discussions; J. Greipel for computoral assistance; and U. Kaysser for expert technical assistance.

REFERENCES

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

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

Berkner, K., L., & Folk, W., R. (1983) *Anal. Biochem. 129*, 446−456.

Brukner, I., Sanchez, R., Suck, D., & Pongor, S. (1995) *EMBO J. 14*, 1812−1818.

Delcourt, S. G., & Blake, R. D. (1991) *J. Biol. Chem. 266*, 15160−15169.

Forsblom, S., Rigler, R., Ehrenberg, M., Pettersson, U., & Philipson, L. (1976) *Nucleic Acids Res. 3*, 3255−3269.

Geiger, R., Rüter, T., Alves, J., Fliess, A., Wolfes, H., Pingoud, V., Urbanke, C., Maass, G., Pingoud, A., Düsterhöft, A., & Krüger, M. (1989) *Biochemistry 28*, 2667−2677.

Goldstein, K., Thomas, M., & Davis, R. W. (1975) *Virology 66*, 420−427.

Grabowski, G., Jeltsch, A., Wolfes, H., Maass, G., & Alves, J. (1995) *Gene 157*, 113−118.

Halford, S. E., Johnson, N. P., & Grinsted, J. (1980) *Biochem. J. 191*, 581−592.

Hunter, C. A. (1996) *BioEssays 18*, 157−162.

Jay, E., Bambara, R., Padmanabhan, R., & Wu, R. (1972) *Nucleic Acids Res. 1*, 331.
Jeltsch, A., Alves, J., Oelgeschläger, T., Wolfes, H., Maass, G., & Pingoud, A. (1993*) J. Mol. Biol. 229*, 221−234.
Kim, Y., Grable, J. C., Love, R., Greene, P. J., & Rosenberg, J. M. (1990) *Science 249*, 1307−1309.
Kramer, W., Drusta, V., Jansen, H. W., Kramer, B., & Fritz, H. J. (1984) *Nucleic Acids Res. 12*, 9441−9456.
Langowski, J., Urbanke, C., Pingoud, A., & Maass, G. (1981) *Nucleic Acids Res. 9*, 3483−3490.
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.
Majumber, K. (1989) *Biophys. J. 55*, 48a.
McClarin, J. A., Frederick, C. A., Wang, B.-C., Boyer, H. W., Grable, J., & Rosenberg, J. M. (1986) *Science 234*, 1526−1541.
Nardone, G., Wastney, M. E., & Hensley, P. (1990a) *J. Biol. Chem. 265*, 15308−15315.
Nardone, G., Connaughton, J. F., Kaloss, W., & Chirikjian, J. G. (1990b) *FASEB J. 4*, A1794.
Pingoud, A., & Jeltsch, A. (1997) *Eur. J. Biochem. 246*, 1−22.
Roberts, R. J., & Halford, S. E. (1993) in *Nucleases* (Linn, S. M., & Roberts, R. J., Eds.) pp 35−88, Cold Spring Harbor Laboratory Press, Plainview, NY.
Rosenberg, J. M. (1991) *Curr. Opin. Struct. Biol. 1*, 104−113.
Rubin, R. A., & Modrich, P (1978) *Nucleic Acids Res. 5*, 2991−2997.
Satchwell, S. C., & Travers, A. A. (1989) *EMBO J. 8*, 229−238.
Santa Lucia, J., Allawi, H. T., & Seneviratne, P. A. (1996) *Biochemistry 35*, 3555−3562.
Taylor, J. D., & Halford, S. E. (1992) *Biochemistry 31*, 90−97.
Thomas, G. A., Kubasek, W. L., Peticolas, W. L., Greene, P., Grable, J., & Rosenberg, J. M. (1989) *Biochemistry 28*, 2001−2009.
Thomas, M., & Davis, R. W. (1975) *J. Mol. Biol. 91*, 315−328.
Urbanke, C. (1982) *Fresenius Z. Anal. Chem. 311*, 364−365.
Windolph, S., & Alves, J. (1997) *Eur. J. Biochem. 244*, 134−139.
Yang, C. C., & Topal, M. D. (1992) *Biochemistry 31*, 9657−9664.

BI970076G