Science Center at Houston, for help in performing the mass spectroscopic analyses.

REFERENCES

Ama, T., Kawaguchi, H., Kanekiyo, M., & Yasui, T. (1980) *Bull. Chem. Soc. Jpn. 53*, 956.
Barker, D. G., Ebel, J.-P., & Burton, C. J. (1982) *Eur. J. Biochem. 127*, 449.
Blanquet, S., Iwatsubo, M., & Waller, J.-P. (1973), *Eur. J. Biochem. 36*, 213.
Blanquet, S., Fayat, S., & Waller, J.-P. (1975) *J. Mol. Biol. 94*, 1.
Bradford, M. M. (1976) *Anal. Biochem. 72*, 248.
Brunie, S., Mellot, P., Zelwer, C., Risler, J.-L., Blanquet, S., & Fayat, G. (1987) *J. Mol. Graphics 5*, 18.
Brunie, S., Zelwer, C., & Risler, J.-L. (1990) *J. Mol. Biol. 216*, 411.
Burbaum, J. J., Starzyk, R. M., & Schimmel, P. (1990) *Proteins: Struct. Funct., Genet. 7*, 99.
Cassio, D., & Waller, J.-P. (1974) *Eur. J. Biochem. 20*, 283.
Fayat, G., & Waller, J.-P. (1974) *Eur. J. Biochem. 44*, 335.
Fayat, G., Fromant, M., & Blanquet, S. (1977) *Biochemistry 16*, 2570.
Fee, J. A., Hegeman, G. D., & Kenyon, G. L. (1974a) *Biochemistry 13*, 2528.
Fee, J. A., Hegeman, G. D., & Kenyon, G. L. (1974b) *Biochemistry 13*, 2533.
Fersht, A. R., & Dingwall, C. (1979) *Biochemistry 18*, 1250.
Heinrikson, R. L., & Hartley, B. S. (1976) *Biochem. J. 105*, 17.
Lenz, G. R., & Martell, A. E. (1964) *Biochemistry 3*, 745.
Mayaux, F., Kalogerakos, T., Brito, K. K., & Blanquet, S. (1982) *Eur. J. Biochem. 128*, 41.
McAuliffe, C. A., Quagliano, J. V. & Vallarino, L. M. (1966) *Inorg. Chem. 5*, 1996.
Monteilhet, C., Brick, P., & Blow, D. M. (1984) *J. Mol. Biol. 173*, 477.
Old, J. M., & Jones, D. S. (1977) *Biochem. J. 165*, 367.
Posorske, L. H., Cohn, M., Yanagisawa, N., & Auld, D. S. (1979) *Biochim. Biophys. Acta 576*, 128.
Rosevear, P. R. (1988) *Biochemistry 27*, 7931.
Schimmel, P. R., & Soll, D. (1979) *Annu. Rev. Biochem. 48*, 601.
Terrill, J. B., & Reilley, C. N. (1966) *Inorg. Chem. 5*, 1988.
Vallee, B. L., & Auld, D. S. (1990) *Proc. Natl. Acad. Sci. U.S.A. 87*, 220.
Williams, J. S., & Rosevear, P. R. (1991) *J. Biol. Chem. 266*, 2089.

Site-Directed Mutagenesis Studies with *Eco*RV Restriction Endonuclease To Identify Regions Involved in Recognition and Catalysis[†,‡]

Vera Thielking, Ursel Selent, Eleonore Köhler, Heiner Wolfes, Uwe Pieper, Robert Geiger, Claus Urbanke, Fritz K. Winkler,[§] and Alfred Pingoud*

Zentrum Biochemie, Medizinische Hochschule Hannover, Konstanty-Gutschow-Strasse 8, D-3000 Hannover 61, Germany

Received October 31, 1990; Revised Manuscript Received March 11, 1991

ABSTRACT: Guided by the X-ray structure analysis of a crystalline *Eco*RV–d(GGGATATCCC) complex (Winkler, in preparation), we have begun to identify functionally important amino acid residues of *Eco*RV. We show here that Asn70, Asp74, Ser183, Asn185, Thr186, and Asn188 are most likely involved in the binding and/or cleavage of the DNA, because their conservative substitution leads to mutants of no or strongly reduced activity. In addition, C-terminal amino acid residues of *Eco*RV seem to be important for its activity, since their deletion inactivates the enzyme. Following the identification of three functionally important regions, we have inspected the sequences of other restriction and modification enzymes for homologous regions. It was found that two restriction enzymes that recognize similar sequences as *Eco*RV (*Dpn*II and *Hin*cII), as well as two modification enzymes (M·*Dpn*II and, in a less apparent form, M·*Eco*RV), have the sequence motif -SerGlyXXXAsnIleXSer- in common, which in *Eco*RV contains the essential Ser183 and Asn188 residues. Furthermore, the C-terminal region, shown to be essential for *Eco*RV, is highly homologous to a similar region in the restriction endonuclease *Sma*I. On the basis of these findings we propose that these restriction enzymes and to a certain extent also some of their corresponding modification enzymes interact with DNA in a similar manner.

Type II restriction endonucleases catalyze the site-specific cleavage of double-stranded DNA [reviews: Modrich and Roberts (1982), Malcolm and Snousnou (1987), Bennett and Halford (1989), and Pingoud et al. (1990)]. Over 1200 different restriction enzymes have been discovered in a wide variety of prokaryotes (Roberts, 1990), among them many isoschizomers which recognize the same DNA sequence. Altogether, about 130 different specificities have been described.

*Eco*RV is one of the best characterized restriction enzymes [review: Luke et al. (1987)]. Its sequence is known (Bougueleret et al., 1984). It has been purified to homogeneity (d'Arcy et al., 1985) from an overproducing strain (Bou-

[†] This work has been supported by grants from the Deutsche Forschungsgemeinschaft (Pi 122/5-1 and Ma 465/11-5) and the Fonds der Chemischen Industrie.

[‡] Crystallographic coordinates for the complex between *Eco*RV and a noncognate DNA fragment have been submitted to the Brookhaven Protein Data Bank.

* To whom correspondence should be addressed.

[§] Present address: F. Hoffmann-LaRoche, Zentrale Forschungseinheiten, Grenzacher Strasse 124, CH-4002 Basel, Switzerland.

0006-2960/91/0430-6416$02.50/0

gueleret et al., 1985) and characterized biochemically (Kuzmin et al., 1984). It is a dimer of two identical subunits (d'Arcy et al., 1985) which recognizes the double-stranded DNA sequence -GAT/ATC- (Kholmina et al., 1980) and cleaves DNA as indicated (Schildkraut et al., 1984) with very high specificity in the presence of Mg^{2+}. Substrates with sites differing in only one base pair from the canonical sequence (Luke et al., 1987) and most substrates with a single mismatch in the canonical site (Alves and Pingoud, unpublished results) are cleaved several orders of magnitude more slowly. Sequences flanking the recognition site have an influence on the rate of cleavage as well as on the mechanism of the reaction, i.e., whether both strands are cleaved in a single binding event or not (Taylor and Halford, personal communication). It is of great importance to investigate what makes the recognition process so precise. This question must be addressed both by studying the structural elements of the DNA needed for recognition as well as by identifying the amino acid residues of the protein involved in DNA contacts. Some of the structural elements of the DNA important for recognition have been identified by introducing chemical modifications into the DNA substrate and analyzing their effect on the rate of cleavage by *Eco*RV (Fliess et al., 1986, 1988; Mazzarelli et al., 1989; Newman et al., 1990a,b; Cosstick et al., 1990). More recently, site-directed mutagenesis studies have been initiated to identify amino acid residues essential for recognition and for catalysis. These studies have been guided by preliminary results of the X-ray structure analysis of crystals of *Eco*RV and *Eco*RV–DNA complexes (Winkler, in preparation), the coordinates of which have been deposited with the Protein Data Bank (Brookhaven National Laboratories).

We describe here the genetic engineering of several *Eco*RV mutants, their isolation, and their characterization with respect to DNA cleavage activity. This work was carried out to identify regions important for the function of *Eco*RV; it will be followed by a detailed characterization of the individual mutants in both structural and functional terms. The results obtained so far demonstrate that regions of the protein implicated by the X-ray structure analysis to be located in close proximity to the DNA are essential for the function of *Eco*RV. We show, furthermore, that these regions have counterparts in some other restriction and modification enzymes, in as much as sequence motifs within these regions can be found in the sequences of *Dpn*II, *Eco*RI, *Hin*cII, and *Sma*I, as well as in M·*Dpn*II and, in a less obvious form, also in M·*Eco*RV.

Experimental Procedures

Bacterial Strains and Plasmid Vectors. The following *Escherichia coli* strains were used: the repair-deficient WK6mutS(λ) for transformation of DNA after gapped-duplex mutagenesis (Kramer et al., 1984), LK111(λ) for propagation of DNA after cassette mutagenesis, and TGE900 for overexpression of mutant proteins. These strains produce the wild type [WK6mutS(λ), LK111(λ)] or a thermosensitive (TGE900) bacteriophage λ repressor and, encoded by the plasmid pLBM4422, the *Eco*RV methylase. Plasmid pLBM4422 is the 4422 bp *Bgl*I/*Bam*HI fragment of pLBM (Bougueleret et al., 1985) that carries the *Eco*RV methylase and a chloramphenicol resistance gene. For mutagenesis, sequencing and overexpression of *Eco*RV proteins, the plasmid pRVF03 and its derivatives were used. pRVF03 carries the *Eco*RV gene under control of the bacteriophage λ p_L-promoter, allowing the induction of overexpression in TGE900 by a temperature shift from 30 to 42 °C. It is derived from the plasmid pRIF309+ (Wolfes et al., 1986) and contains the *Eco*RV gene instead of the *Eco*RI gene.

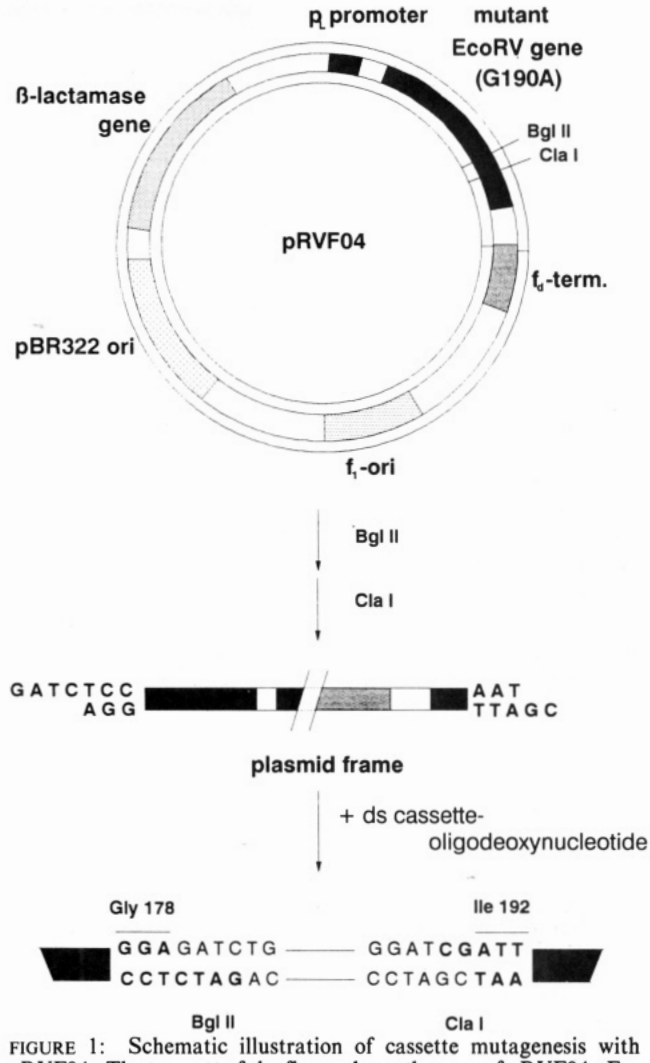

FIGURE 1: Schematic illustration of cassette mutagenesis with pRVF04. The top part of the figure shows the map of pRVF04. For the purpose of cassette mutagenesis, the plasmid was cleaved in the *Eco*RV gene with *Bgl*II and *Cla*I. After removal of the small fragment, a synthetic double-stranded oligodeoxynucleotide ("cassette") carrying the mutation was ligated with the large fragment ("plasmid frame"), thereby restoring the gene and the circular plasmid.

Mutagenesis and Sequencing. The pRVF03 derivative pRVF04 was used for cassette mutagenesis of amino acids 183–190. It contains a single *Bgl*II site at codon 178 in the *Eco*RV gene and a single *Cla*I site at codon 190. Both these sites were used for the preparation of the mutagenesis frame for cassette mutagenesis. In order to prevent dam methylation of the *Cla*I site, the codon GGA for Gly 190, which is part of the site (*GG*ATCGAT), had to be changed by site-directed mutagenesis according to the gapped-duplex protocol. We chose GCA, resulting in the *Eco*RV mutant G190A. This mutation was reversed to wild type in the course of cassette mutagenesis (Figure 1). Details of the mutagenesis following the gapped-duplex protocol have been described by Geiger et al. (1989).

For cassette mutagenesis, typically 0.2 pmol of double-stranded oligodeoxynucleotide was mixed with 0.06 pmol of plasmid frame and ligated overnight at 14 °C in a total volume of 20 μL of ligation buffer (66 mM Tris-HCl, pH 7.5, 5 mM $MgCl_2$, 1 mM DTE, 1 mM ATP, and 1 unit of T4 DNA ligase). The ligation mixture was then transformed into competent LK111(λ) cells containing the plasmid pLBM4422. Cells were plated on LB agar with the appropriate antibiotics. Screening for positive clones was carried out first by restriction analysis with *Cla*I (positive clones had a methylated *Cla*I site

and could therefore not be cleaved by this enzyme), and then by sequencing (Geiger et al., 1989). All synthetic oligodeoxynucleotides used for site-directed mutagenesis were synthesized on a Milligen Cyclone DNA synthesizer and purified as described by Wolfes et al. (1986).

Fermentation. TGE900 cells were grown in L-broth containing 100 μg/mL ampicillin and 30 μg/mL chloramphenicol to an optical density of 2 A_{600nm} (1 cm) at 30 °C. The culture flasks were then transferred to 42 °C and incubated for another 2 h. Cells were harvested by centrifugation and washed once with ice-cold STE buffer (10 mM Tris-HCl, pH 8.0, 0.1 mM EDTA, and 0.1 M NaCl).

Purification of EcoRV Mutants. Cell pellets were resuspended in an equal volume of freshly prepared lysis buffer (25 mM sucrose, 10 mM Tris-HCl, pH 7.8, and 0.75 mg/mL lysozyme) and incubated on ice for 10 min. After addition of 2 volumes of buffer A (30 mM potassium phosphate, pH 7.2, 0.5 mM EDTA, 0.1 mM DTE, 0.1% Lubrol, and 0.1 M NaCl), the incubation was continued for another 10 min. After removal of cell debris by centrifugation, the supernatant was mixed with half a volume of a phosphocellulose suspension in buffer A. The suspension was kept on ice for 5 min to allow binding of the enzyme to the phosphocellulose material. After centrifugation the phosphocellulose was washed three times with buffer A. The protein was eluted with three volumes of buffer B (buffer A containing 0.5 M NaCl instead of 0.1 M). This procedure which can be carried out in parallel with several mutants within 2 h yields a protein preparation which contains at least 30% *Eco*RV and is free of nonspecific nucleases. The mutants T186S, N188A, N188Q, G190A, and the deletion mutant Del 183–188 were purified to homogeneity as judged by SDS–PAGE. For this purpose an additional chromatographic step, consisting of adsorption of appropriately diluted material from the ion-exchange chromatography step to Blue Sepharose CL-6B (Luke et al., 1987) and elution with a 0.15–1.7 M NaCl gradient, was included into the purification scheme.

Bacteriophage λ DNA Cleavage Assay. To measure the catalytic activity of *Eco*RV mutants in cleaving λ DNA, appropriate dilutions were mixed with assay buffer (0.5 μg of bacteriophage λ DNA in 10 mM Tris-HCl, pH 7.5, 10 mM $MgCl_2$, and 50 mM NaCl, final volume 15 μL). Samples were incubated at 37 °C for 15 min and mixed with 5 μL of agarose gel loading buffer; reaction products were separated by agarose gel electrophoresis and stained with ethidium bromide.

Sequence Comparisons, Secondary Structure Prediction, and Hydrophobicity Plots. Sequence comparisons were in part carried out by using the computer program MOTIF developed by Smith et al. (1990) for rapidly finding patterns of conserved amino acid residues in a group of related proteins. Alignments of pairs of sequences were done by using the computer program BESTFIT based on the algorithm of Smith and Waterman (1981) from the Heidelberg Unix Sequence Analysis Resources (HUSAR) program package. Secondary structure predictions describing the propensity of a polypeptide sequence to adopt an α-helix, β-sheet, or nonperiodic structure were made according to the rules established by Chou and Fasman (1978). Hydrophobicity plots displaying the distribution of polar and apolar residues along a given sequence were constructed according to Kyte and Doolittle (1982) with a window of five residues.

Results and Discussion

According to the X-ray structure analysis of the *Eco*RV–d(GGGATATCCC) complex, a loop comprising amino acids 180–190 is in close contact with the major groove of the DNA (Winkler, in preparation). On the basis of the preliminary assignments of contacts between this protein segment and the DNA bases, we have begun to investigate by site-directed mutagenesis the contributions of individual amino acid residues within this loop for the catalytic activity of *Eco*RV. For this purpose, we have first introduced two unique restriction sites into the *Eco*RV gene, viz., a *BglII* site by changing codon 180 from TTG to CTG and a *ClaI* site by changing codon 190 from GGT to GCA and codon 191 from AGC to TCG. While the first and the last mutations are silent, the second mutation leads to a replacement of Gly by Ala. The two new restriction sites define a cassette of 39 base pairs which has been replaced by synthetic double-stranded oligodeoxynucleotides for the purpose of introducing various mutations between codons 180 and 190. The oligodeoxynucleotides were designed such that during the process of cassette mutagenesis the mutation Gly190 to Ala was reversed and the *ClaI* site converted to a dam methyltransferase sensitive sequence. The absence of a cleavable *ClaI* site allowed a simple screening procedure for positive clones by restriction analysis. Mutations were confirmed by sequencing which was extended over the entire coding region of the gene to verify that no mutations other than the desired one(s) had occurred. Isolation of *Eco*RV mutants was carried out by using phosphocellulose chromatography, which yields protein preparations sufficiently pure (at least 30%) for a semiquantitative analysis of the activity of these mutants. Some mutants (T186S, N188A, N188Q, G190A, Del 183–188) were purified to homogeneity by an additional chromatographic step using Blue Sepharose CL-6B. The specificity and activity of the mutants were analyzed in cleavage experiments with bacteriophage λ DNA. The results are given in Table I. They demonstrate that Ser183, Asn185, Thr186, and, to a lesser extent, also Asn188 (Figure 2) are essential for the function of *Eco*RV, since their substitution by similar amino acids leads to an inactive or a much less active enzyme. The substitution of Ser183 by Ala and Thr leads to a decrease in activity by 2–3 orders of magnitude. Presumably, the hydroxyl function of Ser is needed, but there is no space for an extra CH_3 group. A large hydrophobic amino acid residue as supplied by Ile in the S183I mutant leads to a totally inactive enzyme, most likely because it lacks the functional group and disrupts the local structure due to its size and hydrophobic nature. Asn185, which has a hydrogen bond acceptor as well as a donor function, cannot be replaced by the smaller Ala which lacks the $CONH_2$ group without complete loss of activity, nor by the larger Gln. A very low activity is observed after substitution by the nearly isosteric Asp which could supply a hydrogen bond acceptor function, not, however,

Table I: Activity of *Eco*RV Mutants in Cleaving Bacteriophage λ DNA[a]

mutant	activity	mutant	activity
N70Q	<wt	T187S	≈wt
D74E	≪wt	T187N	≈wt
S183A	≪wt	N188A	<wt
S183T	≪wt	N188Q	≪wt
S183I	none	N188T	≪wt
N185D	≪wt	N188D	none
N185A	none	G190A	≈wt
N185Q	none	Del 183–188	nonspecific phosphodiesterase
T186S	none		
T186N	none	Del 216–245	none

[a] The relative magnitude is indicated by ≈wt (within a factor of 10 as active as the wild-type *Eco*RV), <wt (within a factor of 10–100 less active than wild-type EcoRV), ≪wt (within a factor of 100–1000 less active than wild-type *Eco*RV), or none (no detectable cleavage).

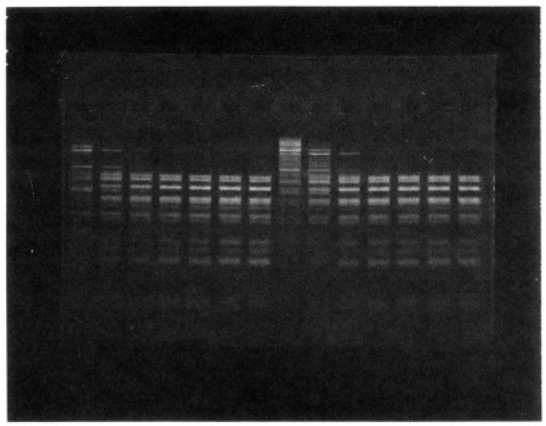

FIGURE 2: Bacteriophage λ DNA cleavage assay: 2.5 μg of bacteriophage λ DNA was incubated in 75 μL of cleavage buffer (20 mM Tris-HCl, pH 7.5, 10 mM $MgCl_2$, and 50 mM NaCl) at 37 °C with the *Eco*RV mutant N188A or the wild-type enzyme. After 2, 4, 6, 8, 10, 15, and 20 min aliquots were withdrawn and the reaction was stopped by mixing with agarose gel loading buffer. The enzyme concentration in the reaction mixtures was 330 nM N188A or 6.7 nM wild-type *Eco*RV, respectively. Left half: N188A. Right half: Wild-type *Eco*RV. This experiment demonstrates that the N188A mutant has an approximately 50-fold lower specific activity than wild-type *Eco*RV.

the donor function when it is ionized. Thr186 seems to fulfil a very delicate role; its substitution by Ser, which is slightly smaller but has the same functional group, or by Asn, which is of similar size and polarity, but different shape and functionality, renders the enzyme totally inactive. Asn188 seems to be as important for the catalytic activity of *Eco*RV as Asn185 but shows a different sensitivity toward different amino acid substitutions, indicating that it has a different role; it is noteworthy that the substitution by Ala is not as deleterious as the same substitution in position 185. This may indicate that in position 188 the hydrogen-bonding capacity of Asn is not needed. The inactivity of the N185A and the reduced activity of the N188A mutant in cleaving DNA were also demonstrated by Halford and co-workers (Vermote and Halford, personal communication). The identification of Gly184 to be essential for the activity of *Eco*RV is based on our finding that several double mutants involving positions 184 and 187, which arose spontaneously during mutagenesis experiments at position 187, were inactive (data not shown). Since Thr187 can be replaced by other polar amino acids without effect on cleavage activity, it is very likely that the substitution of Gly by other amino acids is responsible for the defect. The replacement of the same amino acid in various positions of this region has different effects, viz., Gly in positions 184 (essential) and 190 (nonessential) and Thr in positions 186 (essential) and 187 (nonessential). This means that only some of the functional groups within this region are needed for recognition. As expected on the basis of the results obtained with the single-site mutants, the complete deletion of amino acids 183–188 leads to a mutant that exhibits a weak nonspecific phosphodiesterase activity, presumably because of an uncoupling of recognition and catalysis. This fact may be correlated with the finding that this mutant is defective in dimerization as shown by gel filtration experiments (data not shown).

The structure of the *Eco*RV–d(GGGATATCCC) complex also shows that Asn70 contacts the DNA in the minor groove (Winkler, in preparation). We have exchanged this residue by site-directed mutagenesis using the gapped-duplex protocol. In accordance with the presumed specific function of this residue, the N70Q mutant is by a factor of 100 less active than wild-type *Eco*RV (Table I). A residue nearby, Asp74, which because of its negative charge might be involved in Mg^{2+} binding and/or catalysis by abstracting a proton from a water molecule, was substituted by Glu; this substitution leads to an almost inactive enzyme (Table I). Both these substitutions are conservative ones; the effects observed, therefore, argue for a specific role of these amino acid residues, which cannot be fulfilled by a larger amino acid residue with the same functional group.

*Eco*RV presumably makes contacts to the DNA backbone in regions outside of the recognition sequence (Luke et al., 1987). These may serve to bind the substrate more firmly and presumably are nonspecific. In *Eco*RV, such contacts are formed by some amino acid residues of segment 221–226 which is near the C-terminus (Winkler, in preparation). Our mutagenesis data show that a deletion mutant of *Eco*RV, Del 216–245, is inactive (Table I). We are currently engaged in narrowing down the essential parts of this region by producing a series of mutants with short truncations at the C-terminal end of *Eco*RV.

In general, type II restriction enzymes do not share extended stretches of homology. There are a few exceptions regarding several pairs of isoschizomers, as became apparent during the 2nd New England Biolabs Workshop on Biological DNA Modification (Berlin, Sept 2–7, 1990). As in the previously published case of *Eco*RI and *Rsr*I (Greene et al., 1981; Newman et al., 1981; Greene et al., 1988: Aiken & Gumport, 1988; Stephenson et al., 1989), these pronounced homologies are of obvious significance and indicate that these enzymes are evolutionary closely related. Very short stretches of homology have been detected before between *Bsu*RI, *Eco*RI, *Eco*RV, and *Pae*R7 (Lauster, 1989). The significance of this weak homology among restriction enzymes that recognize *different* sequences and have a *dissimilar* mode of cleavage by producing either blunt or sticky ends is not clear at present, because it cannot be correlated as yet with a particular common function, such as DNA recognition, Mg^{2+} binding, or nucleolytic activity. We have wondered whether other, not readily apparent homologies can be detected among restriction enzymes that recognize *similar* sequences or have a *similar* mode of cleavage. Such weak homologies, not necessarily significant by themselves, may become meaningful when considered in the light of a known structure or function. Accordingly, we have analyzed the sequences of other restriction enzymes for homologies with those regions in *Eco*RV that have been identified to be important for the function of this enzyme by our site-directed mutagenesis studies.

We have detected such a homology with *Dpn*II, *Eco*RV, and *Hinc*II which show no overall sequence homology. These enzymes not only are of similar size, their subunit molecular weights being around 30K, but also have recognition sequences with similar base composition which are cleaved to give blunt ends. Among the more than 20 restriction enzymes whose sequences are known these are the only ones that cleave DNA within a sequence G--/--C or G-/-C (the dashes represent adenine or thymidine residues) and produce blunt ends. Our sequence comparison shows that one short region of homology is shared by four of these (Table II). It contains a core segment comprising nine amino acids which conform to the consensus sequence -SerGlyXXXAsnIleXSer- and in *Eco*RV is part of the region spanning amino acids 180–190. Among over 15 000 proteins whose sequences have been deposited in the Swiss protein data bank (release 15, August 1990), only nine other proteins contain this sequence, namely, *Cochliobolus* ATP synthase α-subunit, cattle plague virus hemagglutinin

Table II: Alignment in the Common Motif Region for the *Dpn*II, *Eco*RV, and *Hin*cII Restriction Enzymes[a]

Enzyme	Recognition sequence	
R·EcoRV	GAT/ATC	^{177}A G D L A G S G N T T N I G S I H A H
R·DpnII	GA/TC	^{215}V N F Y S G S G S K L N E T A R S Y K
R·HincII	GTY/RAC	^{130}R R N I S K S A Q A P N I I S A Y K L
M·DpnII	GATC	^{88}N G T V W I S G S L H N I Y S V G M A
Most common amino acids		- - - - - - S G - - - N I - S - - - -

[a] Identical amino acid residues at a given position are shown in boldface. Additionally, the sequence of the *Dpn*II methyltransferase (dpn A gene product) is shown, which also contains this motif. Sequences were obtained from the following references: R·*Dpn*II and M·*Dpn*II, Lacks et al. (1986); R·*Eco*RV and M·*Eco*RV, Bougueleret et al. (1984); *Hin*cII, Ito et al. (1990).

Table III: Alignment of the Sequences of the *Eco*RV Restriction and Modification Enzymes within the Common Motif Region[a]

R·EcoRV	^{177}A G D L A G **S** G N T T N I G **S** I H A **H**
M·EcoRV	^{205}L F E T L S **S** L N A T F I T **S** T W H **H**

[a] Identical amino acids are shown in boldface. Sequences were obtained from the following references: R·*Eco*RV and M·*Eco*RV, Bougueleret et al. (1984).

Table IV: Alignment of the C-Terminal Region of the *Eco*RV Restriction Enzyme with a Sequence in *Sma*I[a]

R·EcoRV	^{188}**N** I G S I H A **H** Y K D F V **E** - **G K G**
R·SmaI	^{114}**N** I S K S Y I **H** V S K W M **E** L **G K G**
	^{205}I F D S **E** D E - - - **F L** D Y W R **N** Y
	^{132}**E** W I L **E** L L L E R **F L** E H L **E N** Y
	^{222}**E** R T S Q **L** R N D K Y N N I **S E** Y R
	^{152}**E** R I F T **L** R - - - Y F K I **S E** Y K

[a] Identical amino acids are shown in boldface. The chance occurrence of this homology is less than 1% for two peptide sequences of the same amino acid composition (BESTFIT). Sequences were obtained from the following references: R·*Eco*RV, Bougueleret et al. (1981); R·*Sma*I, Heidmann et al. (1989).

Table V: Homologous Amino Acid Residues in the Presumptive Mg^{2+} Binding Site and/or Catalytic Center of *Eco*RV, *Eco*RI, and *Rsr*I[a]

EcoRV		^{73}**P D**	^{90}**D** I **K**
EcoRI	isoschizomers	^{90}**P D**	^{111}**E** A **K**
RsrI	isoschizomers	^{95}**P D**	^{116}**E** S **K**

[a] Identical amino acids are shown in boldface. Sequences were obtained from the following references: R·*Eco*RV, Bougueleret et al., (1984); R·*Eco*RI, Newman et al. (1981) and Greene et al. (1981); R·*Rsr*I, Stephenson et al. (1989).

neuraminidase, *Drosophila* homeotic sex combs reduced protein, cowpox host range protein, *Pneumococcus* modification methylase (M·*Dpn*II), shark myoglobin, dengue virus polyprotein, *Pichinde arena* virus polyprotein, and *Vaccinia* virus 27.4-kDa protein. While eight of these nine proteins presumably contain this motif by chance, M·*Dpn*II most likely does not, because it has one function in common with the *Dpn*II restriction enzyme, i.e., to recognize a specific DNA sequence. The *Dpn*II methyltransferase otherwise shows no significant homology with the *Dpn*II restriction endonuclease or any other restriction enzyme, which supports the assumption of a specific function of this stretch of amino acids also in M·*Dpn*II. Prompted by these findings, we have looked more closely at the *Eco*RV methyltransferase sequence and found an interesting extended homology between the *Eco*RV restriction and modification enzymes (Table III) which encompasses the consensus sequence of the common motif region of *Dpn*II, *Eco*RV, and *Hin*cII. The secondary structure prediction shows that the predominant part of the common motif is likely not to be part of an extended α-helix or β-pleated sheet in all five proteins. The hydrophobicity plot indicates that it occurs in a slightly hydrophilic region. The X-ray structure analysis of an *Eco*RV crystal has shown that these assignments are essentially correct for *Eco*RV (Winkler, in preparation). On the basis of mutational analysis with *Eco*RV, we propose that this sequence motif is involved in the recognition of a particular feature of the DNA conformation within the recognition sequences of these enzymes, possibly—although this is highly speculative—the common GC base pair. While it is evident from our data that this sequence is part of a region which is essential for the function of *Eco*RV, it remains to be demonstrated, of course, whether our proposal is correct for the other enzymes.

In addition to the sequence homology which concerns the region spanning amino acids 180–190 in *Eco*RV, there is another hitherto unnoticed homology regarding *Eco*RV and *Sma*I (Table IV). According to our site-directed mutagenesis results, this region is essential for the enzymatic activity of *Eco*RV, probably because it is involved in nonspecific contacts to the DNA. This would explain why two restriction enzymes which have very dissimilar recognition sequences, -GAT/ATC- vs -CCC/GGG-, have homologous regions.

The comparison of the structure of *Eco*RI (Kim et al., 1990) and *Eco*RV (Winkler, in preparation) has revealed another interesting homology which presumably concerns the Mg^{2+} binding site and/or the active center of the two enzymes (Table V). In the current model of the *Eco*RI structure (Kim et al., 1990), Glu111 is placed near the scissile bond of the DNA and is part of a β-strand running parallel to the DNA backbone. Mutations at this locus have been identified (Wolfes et al., 1986; Kim et al., 1989) that strongly reduce the cleavage activity of the enzyme. Nearby there is Lys113, and vis a vis as part of a loop there are Pro90 and Glu91. Homologous residues in structurally similar locations are found in the *Eco*RV structure with Asp90 and Lys92 in a β-strand parallel to the DNA backbone near the scissile bond and Pro73 and Asp74 in a nearby loop. Site-directed mutagenesis experiments show that Asp74 is essential for the enzymatic activity of *Eco*RV (Table I). We have found the same motif, -ProAsp...Glu(Asp)XLys-, in many other restriction endonucleases including *Dpn*II and *Hin*cII. This is an example of a homology which is not necessarily significant by itself but may become so in the light of structural and functional studies. Thus, this motif may represent one version of a Mg^{2+} binding site and/or an active center of restriction enzymes which could be formed by two acidic amino acids as in the aspartate proteases [review: James and Sielicki (1987)].

The observation that regions identified in *Eco*RV to be important for its catalytic activity have counterparts in other restriction enzymes is not totally unexpected because type II restriction enzymes constitute a very large family of functionally related proteins. It can be anticipated that their

common function will be reflected in common—albeit not necessarily readily apparent—structural features, as was shown in recent years for a similarly large family, the DNA methyltransferases (Lauster et al., 1987, 1989; Chandrasegaran & Smith, 1988; Posfai et al., 1989; Lauster, 1989; Ingrosso et al., 1989; Smith et al., 1990).

Conclusions

We have described here a set of *Eco*RV mutants which were constructed to identify functionally important regions of this enzyme. So far, we have identified three regions which seem to be necessary for binding and/or cleavage of the DNA, the criteria being that conservative amino acid substitutions in these regions lead to *Eco*RV mutants of very low or no activity. While our experiments were guided by the preliminary data from the X-ray structure analysis, their results constitute independent evidence for the importance of these regions.

A sequence comparison suggests that several other restriction and modification enzymes have counterparts to these regions, a suggestion that must be tested by similar site-directed mutagenesis experiments as carried out for *Eco*RV. With more sequences of restriction and modification enzymes being published, we soon will have an independent means to test the validity of our suggestions.

Added in Proof

In agreement with our assignment that the sequence motif PD...DXK may represent part of the catalytic center of *Eco*RV, we have found that the mutants P73A, P73G, D74E, D90A, D90N, D90E, D90T, and K92E are inactive or almost inactive.

Acknowledgments

We are indebted to Dr. M. Zabeau for two plasmids harboring the genes of the *Eco*RV endonuclease and methyltransferase. We thank Drs. J. S. Benner, V. Butkus, M. Kröger, T. A. Trautner, and P. A. Waite-Rees for supplying us with sequence information prior to publication, Drs. B. A. Connolly, R. I. Gumport, and S. E. Halford for sending us unpublished manuscripts, and S. Chandrasegaran for making computer programs available to us. We acknowledge the cooperation of the Deutsche Krebsforschungszentrum for giving us access to their computing facilities to use the HUSAR program package. We thank our colleagues Drs. G. Maass and J. Alves for many helpful discussions. We gratefully acknowledge the technical assistance of Mrs. U. Kaysser and Mr. R. Mull-Grotefend.

Registry No. *Eco*RV, 83589-02-0; *Eco*RV *Dpn*II, 81295-28-5; *Eco*RV *Hin*cII, 81811-55-4; M·*Dpn*II, 98002-56-3; M·*Eco*RV, 91448-95-2; *Eco*RV *Sma*I, 82391-42-2; Asn, 70-47-3; Asp, 56-84-8; Ser, 56-45-1; Thr, 72-19-5.

References

Aiken, C., & Gumport, R. I. (1988) *Nucleic Acids Res. 16*, 7901–7916.

Bennett, S. P., & Halford, S. E. (1989) in *Current topics in cellular regulation*, Vol. 30, pp 57–104, Academic Press, New York.

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

Bougueleret, L., Tenchini, M. L., Botterman, J., & Zabeau, M. (1985) *Nucleic Acids Res. 13*, 3823–3839.

Chandrasegaran, S., & Smith, H. O. (1987) in *Structure and Expression: From Proteins to Ribosomes* (Sarma, R. H., & Sarma, M. H., Eds.) Vol. I, pp 149–156, Adenine Press, Schenectady, NY.

Chou, P. Y., & Fasman, G. D. (1978) *Adv. Enzymol. 47*, 45–148.

Cosstick, R., Li, X., Tuli, D. K., Williams, D. M., & Connolly, B. A. (1990) *Nucleic Acids Res. 18*, 4771–4778.

D'Arcy, A., Brown, R. S., Zabeau, M., Wijnaendts van Resandt, R. & Winkler, F. K. (1985) *J. Biol. Chem. 260*, 1987–1990.

Fliess, A., Wolfes, H., Rosenthal, A., Schwellnus, K., Blöcker, H., Frank, R., & Pingoud, A. (1986) *Nucleic Acids Res. 14*, 3463–3474.

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

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

Greene, P. J., Gupta, M., Boyer, H. W., Brown, W. E., & Rosenberg, J. M. (1981) *J. Biol. Chem. 256*, 2143–2153.

Greene, P. J., Ballard, B. T., Stephenson, F., Kohr, W. J., Rodriguez, J. M., Rosenberg, J. M., & Boyer, H. W. (1988) *Gene 68*, 43–52.

Heidmann, S., Seifert, W., Kessler, C., & Domdey, H. (1989) *Nucleic Acids Res. 17*, 9783–9796.

Ingrosso, D., Favler, A. V., Bleibaum, J., & Clarke, S. (1989) *J. Biol. Chem. 264*, 20131–20139.

Ito, H., Sadaoka, A., Kotani, H., Hiraoka, N., & Nakamura, T. (1990) *Nucleic Acids Res. 18*, 3903–3911.

James, M. N. G., & Sielecki, A. R. (1987) in *Biological Macromolecules and Assemblies* (Jurnak, F. A., & McPherson, A., Eds.) Vol. III, pp 413–482, Wiley, New York.

Kholmina, G. V., Rebentish, B. A., Skoblov, Y. S., Mironov, A. A., Yankovskii, Y., Kozlov, Y. I., Glatman, L. I., Moroz, A. E., & Debabov, V. G. (1980) *Dokl. Akad. Nauk. SSSR 253*, 495–497.

Kim, K., Benkovic, S. J., & Modrich, P. (1989) *J. Biol. Chem. 264*, 11807–11815.

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., Pflugfelder, M., & Fritz, H. J. (1984) *Nucleic Acids Res. 12*, 9441–9456.

Kuz'min, N. P., Loseva, S. P., Belyaeva, R. Kh., Kravets, A. N., Solonin, A. S., Tanyashin, V. I., & Baev, A. A. (1984) *Mol. Biol. (Moscow) 18*, 197–204.

Kyte, J., & Doolittle, R. F. (1982) *J. Mol. Biol. 157*, 105–132.

Lacks, S. A., Mannarelli, B. M., Springhorn, S. S., & Greenberg, B. (1986) *Cell 46*, 993–1000.

Lauster, R. (1989) *J. Mol. Biol. 206*, 313–321.

Lauster, R., Kriebardis, H., & Guschlbauer, W. (1987) *FEBS Lett. 220*, 167–176.

Lauster, R., Trautner, T. A., & Noyer-Weidner, M. (1989) *J. Mol. Biol. 206*, 305–312.

Luke, P. A., McCallum, S. A., & Halford, S. E. (1987) in *Gene Amplification and Analysis* (Chirikjian, J. G., Ed.) Vol. V, pp 183–205, Elsevier, New York.

Malcolm, A. D. B., & Snousnou, G. (1987) in *Topics in molecular and structural biology* (Neidle, S., Ed.) Vol. IX, pp 193–222, Macmillan Press, London.

Mazzarelli, J., Scholtissek, S., & McLaughlin, L. W. (1989) *J. Biol. Chem. 28*, 4616–4622.

Modrich, P., & Roberts, R. J. (1987) in *Nucleases* (Linn, S. M., & Roberts, R. J., Eds.) pp 109–154, Cold Spring Harbor Laboratory, Cold Spring Harbor, NY.

Newman, A. K., Rubin, R. A., Kim, S.-H., & Modrich, P. (1981) *J. Biol. Chem. 256*, 2131-2139.
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.
Pingoud, A., Alves, J., & Geiger, R. (1991) in *Enzymes in molecular biology* (Burrell, M. M., Ed.) Humana Press, Clifton, NJ (in press).
Posfai, J., Bahwat, H. S., Posfai, G., & Roberts, R. J. (1989) *Nucleic Acids Res. 17*, 2421-2435.
Roberts, R. J. (1990) *Nucleic Acids Res. 18*, Suppl., 2331-2365.
Schildkraut, I., Banner, C. D., Rhodes, C. S., & Parekh, S. (1984) *Gene 27*, 327-329.
Smith, H. O., Annau, T. M., & Chandrasegaran, S. (1990) *Proc. Natl. Acad. Sci. U.S.A. 87*, 826-830.
Smith, T. F., & Waterman, M. S. (1981) *Adv. Appl. Math. 2*, 482-489.
Stephenson, F. H., Ballard, B. T., Boyer, H. W., Rosenberg, J. M., & Greene, P. (1989) *Gene 85*, 1-13.
Vershon, A. K., Youderian, P., Weiss, M. A., Susskind, M. M., & Sauer, R. T. (1985) in *Sequence specificity in transcription and translation*, pp 209-218, Alan R. Liss, New York.
Wolfes, H., Alves, J., Fliess, A., Geiger, R., & Pingoud, A. (1986) *Nucleic Acids Res. 14*, 9063-9080.

The NADH-Binding Subunit of the Energy-Transducing NADH-Ubiquinone Oxidoreductase of *Paracoccus denitrificans*: Gene Cloning and Deduced Primary Structure†,‡

Xuemin Xu, Akemi Matsuno-Yagi, and Takao Yagi*
Division of Biochemistry, Department of Molecular and Experimental Medicine, Research Institute of Scripps Clinic, La Jolla, California 92037

Received February 4, 1991; Revised Manuscript Received April 3, 1991

ABSTRACT: The NADH dehydrogenase complex isolated from *Paracoccus denitrificans* is composed of approximately 10 unlike polypeptides and contains noncovalently bound FMN, non-heme iron, and acid-labile sulfide [Yagi, T. (1986) *Arch. Biochem. Biophys. 250*, 302-311]. The NADH-binding subunit (M_r = 50000) of this enzyme complex was identified by direct photoaffinity labeling with [^{32}P]NADH [Yagi, T., & Dinh, T. M. (1990) *Biochemistry 29*, 5515-5520]. Primers were synthesized on the basis of the N-terminal amino acid sequence of this polypeptide, and these primers were used to synthesize an oligonucleotide probe by the polymerase chain reaction. This probe was utilized to isolate the gene encoding the NADH-binding subunit from a genomic library of *P. denitrificans*. The nucleotide sequence of the gene and the deduced amino acid sequence of the entire NADH-binding subunit were determined. The NADH-binding subunit has 431 amino acid residues and a calculated molecular weight of 47 191. The encoded protein contains a putative NAD(H)-binding and an iron-sulfur cluster-binding consensus sequence. The deduced amino acid sequence of the *Paracoccus* NADH-binding subunit shows remarkable similarity to the α subunit of the NAD-linked hydrogenase of *Alcaligenes eutrophus* H16. When partial DNA sequencing of the regions surrounding the gene encoding the NADH-binding subunit was carried out, sequences homologous to the 24-, 49-, and 75-kDa polypeptides of bovine complex I were detected, suggesting that the structural genes of the *Paracoccus* NADH dehydrogenase complex constitute a gene cluster.

The energy-transducing NADH-quinone (Q)[1] oxidoreductase (complex I or NDH-1) in the respiratory chains of mitochondria and bacteria contains a noncovalently bound FMN and multiple iron-sulfur clusters. The enzyme catalyzes electron transfer between NADH and Q and leads to the concomitant formation of a proton gradient across the membrane (Hatefi et al., 1985; Hatefi, 1985; Ragan, 1987; Yagi, 1989, 1991). The NADH-Q oxidoreductase of mitochondria is composed of more than 25 unlike polypeptides (Tuschen et al., 1990; Hatefi, 1985; Ragan, 1987) and appears to have the most intricate structure of any known membrane-bound protein complex. This complexity has hampered progress in studies on many aspects of the structure and mechanism of action of this enzyme complex (Yagi, 1991). However, such studies have taken on a greater significance in recent years since there has been an increasing number of reports on human mitochondrial diseases involving structural and functional defects at the level of this enzyme complex (Morgan-Hughes et al., 1990; Wallace et al., 1988). It has been recently suggested that Parkinson's and Huntington's diseases might also

† This work was supported by US Public Health Service Grant GM33712 to T.Y. and MO1 RR00833 to General Clinical Research Center. This is publication 6657-MEM from the Research Institute of Scripps Clinic, La Jolla, CA.
‡ The nucleotide sequence of the *Paracoccus* NADH-binding subunit gene given in this paper has been submitted to GenBank under Accession Number M64432.
* To whom correspondence should be addressed.

[1] Abbreviations: Q, quinone; UQ, ubiquinone; NDH-1 or complex I, energy-transducing NADH-quinone oxidoreductase; NDH-2, NADH-quinone oxidoreductase lacking energy coupling site; complex III, ubiquinol-cytochrome *c* oxidoreductase; bp, base pairs; FP, IP, and HP, respectively, the flavoprotein, the iron-sulfur protein, and the hydrophobic protein fractions of complex I; SDS-PAGE, sodium dodecyl sulfate-polyacrylamide gel electrophoresis; PCR, polymerase chain reaction; PVDF, poly(vinylidene difluoride).

0006-2960/91/0430-6422$02.50/0