Regions of Endonuclease EcoRII Involved in DNA Target Recognition Identified by Membrane-bound Peptide Repertoires*

Monika Reuter†, Jens Schneider-Mergener¶, Dagmar Kupper, Andreas Meisel§, Petra Mackeldanz, Detlev H. Krüger, and Cornelia Schroeder

From the Institutes of Virology and Medical Immunology, Humboldt University Medical School (Charité), D-10098 Berlin, Germany

Target sequence-specific DNA binding regions of the restriction endonuclease EcoRII were identified by screening a membrane-bound EcoRII-derived peptide scan with an EcoRII recognition site (CCWGG) oligonucleotide duplex. Dodecapeptides overlapping by nine amino acids and representing the complete protein were prepared by spot synthesis. Two separate DNA binding regions, amino acids 88–102 and amino acids 256–273, which share the consensus motif KRXKXK, emerged. Screening 570 single substitution analogues obtained by exchanging every residue of both binding sites for all other amino acids demonstrated that replacing basic residues in the consensus motifs significantly reduced DNA binding. EcoRII mutant enzymes generated by substituting alanine or glutamic acid for the consensus lysine residues in DNA binding site I expressed attenuated DNA binding, whereas corresponding substitutions in DNA binding site II caused impaired cleavage, but enzyme secondary structure was unaffected. Furthermore, Glu96, which is part of a potential catalytic motif and also locates to DNA binding site I, was demonstrated to be critical for DNA cleavage and binding. Homology studies of DNA binding site II revealed strong local homology to SsoII (recognition sequence, CCNGG) and patterns of sequence conservation, suggesting the existence of functionally related DNA binding sites in diverse restriction endonucleases with recognition sequences containing terminal C:G or G:C pairs.

Type II restriction endonucleases (ENases) are ideal models for investigating the molecular basis of specificity in the interaction of proteins with their specific DNA recognition sites. More than 2900 type II ENases isolated from different sources have been described, representing more than 200 individual DNA sequence specificities. They form one of the most comprehensive groups of functionally similar proteins with distinct DNA binding specificities. As phylogenetically diverse enzymes coded by eubacteria, archaeabacteria, and viruses evolved to recognize identical DNA sequences, a variety of structures and mechanisms involved in DNA recognition can be expected.

EcoRII belongs to the IIE type of ENases characterized by their essential interaction with two copies of the recognition site for DNA cleavage. This cooperative mode of action limits their efficiency of DNA cleavage, but may reflect additional biological functions besides the defense of the host cell against invading foreign DNA (for reviews, see Refs. 2 and 3). In bridging two DNA recognition sites in cis or trans, type IIE ENases resemble proteins involved in DNA replication and recombination, as well as in transcription control, in pro- and eukaryotes (2–6). The dimeric EcoRII ENase associates with two DNA sites (7–9) in a distance-dependent fashion through a DNA bending/looping mechanism (10). Up to now, functional domains of EcoRII have not been identified, and it is not known how DNA recognition is realized at the amino acid level. The crystal structure of an EcoRII-DNA complex remains to be determined.

Restriction ENases are remarkably recognition site-specific enzymes. This is borne out in a high affinity (association constants, $K_d \approx 10^4$–$10^5$ M$^{-1}$) to their recognition sequence, a low affinity to sites differing by only 1 base pair, and the precise coupling of DNA recognition and site-specific cleavage. The structural elements that make specific contacts with DNA target sites differ in all ENase-DNA co-crystal structures analyzed so far, whereas their tertiary and quaternary structure and the residues involved in catalysis show a degree of conservation (cf. Refs. 11 and 12 for review). The structural data base of ENases is still too limited to infer rules of protein-DNA recognition and, in particular, to decide whether these may define a protein-DNA recognition code. The theoretical interest and practical impact of such recognition rules make faster methods of identifying DNA binding sites in proteins most desirable.

We used spot synthesis (13) to prepare EcoRII-derived peptide scans bound C-terminally to continuous cellulose membranes. Peptide scans have been applied successfully to the investigation of linear or discontinuous protein-protein (14–16), or protein-metal (17) contact sites. We have now adapted the peptide scan approach to investigate protein-DNA contacts.

The study presented here led to the identification of two EcoRII peptides capable of sequence-specific DNA binding. EcoRII mutants constructed with mutations in these DNA binding sites had altered DNA binding and scission properties. The sequence of one of the two DNA binding sites led to the discovery of a putative family of DNA binding peptides in diverse restriction enzymes.

* This work was supported by Deutsche Forschungsgemeinschaft Grant Re 8792/1, Bundesministerium für Bildung und Forschung (BMBF FKZ 0311014), Fonds der Chemischen Industrie, and Universität für Wissenschaft und Technologie Westfalen-Lippe.

† To whom correspondence should be addressed. Tel.: 49-30-2802-2388 or 49-30-2802-2357; Fax 49-30-2802-2180.

‡ Present address: Department of Neurology, Humboldt University Medical School (Charité), D-10098 Berlin, Germany.

§ The abbreviations used are: ENase, restriction endonuclease; MTase, methyltransferase.

This paper is available on line at http://www.jbc.org

Received for publication, November 5, 1998, and in revised form, November 30, 1998

Printed in U.S.A.
EXPERIMENTAL PROCEDURES

Determination of the EcoRII-encoding Nucleotide Sequence—At the onset of this study, there were two GenBank entries for the complete nucleotide sequence of the ecoRII ENase gene, X61025 (18) and M26404 (19), which differ in six nucleotides and include two frameshifts. Before synthesizing peptide scans, it was necessary to clarify these uncertainties. We sequenced our EcoRII expression plasmid (10) originating from pBR322 in both directions. By using the Thermo Sequenase dye- AzA-G-Kit (Amersham Pharmacia Biotech), it was possible to resolve a track of four guanines, otherwise appearing as three, corresponding to nucleotide position 676–678 of the sequence M26404. Except for this one additional G and the deletion of a C at nucleotide position 710, our results match sequence M26404. The revised nucleotide sequence was submitted to GenBank (accession number AF224995).

DNA-Peptide Binding Experiments—The cellulose-bound peptide scan was rinsed for 5 min in methanol and then preincubated for 1 h in 100 mM maleic acid, 150 mM NaCl, 1% blocking reagent (Boehringer GmbH). To compare data from different experiments, a serial dilution of a 13C standard was included with each experiment, and the absolute amount of radioactivity was taken into account. Peptide scans were stripped of bound radioactivity either by electroblotting and/or extended washing steps in EcoRII binding buffer supplemented with 2× NaCl or 1× K HPO4, pH 8.0.

Site-specific Mutagenesis—Oligonucleotide-directed mutagenesis was carried out according to VandeYar et al. (21). The 1294-base pair EcoRII fragment from expression vector pQE-30 containing the coding region for ENase EcoRII (10) was cloned in M13mp18. Potential cloning sites, as well as the mutations themselves, were verified twice by sequencing, after mutagenesis and again after re-cloning.

RESULTS

Sequence-specific DNA Binding Regions Identified in EcoRII Peptide Scans—The entire amino acid sequence presented as a set of 132 covalently cellulose-bound dodecapeptides overlapping by 9 amino acids was screened for its ability to bind EcoRII recognition site-containing oligonucleotide duplexes. The length and overlap of the peptides in the scan were chosen because individual segments of discontinuous DNA binding sites generally do not exceed 9–12 amino acids. Fig. 1a shows that peptides from two regions distant in the primary sequence
exhibit high affinity to the oligonucleotide. The influence of Mg\(^{2+}\) ions on DNA-peptide binding is evident by comparing Fig. 1a and Fig. 1b. The quantitative effect of Mg\(^{2+}\) is shown in Fig. 1c. Because Mg\(^{2+}\) ions clearly diminished unspecific DNA binding to several EcoRII peptides, all subsequent DNA binding experiments were carried out in the presence of 10 mM Mg(OAc)\(_2\).

To examine whether the interaction with the DNA substrate was specific to the EcoRII recognition sequence CCWGG, the following series of competition experiments was carried out: (i) binding of the labeled specific oligonucleotide duplex alone, (ii) binding as in i with saturating amounts of an unlabeled, unspecific oligonucleotide, and (iii) binding as in i with saturating amounts of an unlabeled, specific oligonucleotide. The unspecific competitor was designed so as to avoid any homology with the EcoRII recognition sequence at the dinucleotide level. Competing oligonucleotide duplexes were saturating with respect to total peptide in the spots and represented a 10\(^{-3}-10\(^{-5}\) molar excess of unlabeled over labeled substrate. Data from series of competition experiments were compared quantitatively on the basis of a 14C standard.

Under stringent competitive conditions, only five dodecapeptide spots specifically bound oligonucleotides containing the EcoRII DNA recognition site (Fig. 2). The binding of labeled specific oligonucleotide duplex to the peptides was not reduced by saturating concentrations of unlabeled unspecific oligonucleotides. In contrast, the same concentration of unlabeled specific competitor DNA decreased the binding of labeled specific substrate below 1%. In comparison to an EcoRII-specific substrate, the labeled unspecific oligonucleotide duplex itself bound to these peptides with an efficiency between 1% and, maximally, 10% (data not shown).

Two specific DNA binding regions spanning 15 and 18 amino acids, located between EcoRII amino acid positions 88–102 (binding site I) and 256–273 (binding site II) were thus identified under highly competitive binding conditions. Both potential binding regions share the minimal consensus motif KXXRXK (Fig. 2), which does not occur elsewhere in the EcoRII sequence.

Substitution Analysis Reveals Residues Critical for Peptide-DNA Binding—Substitution analogues of both potential DNA binding sites of EcoRII were synthesized in which every amino acid of the original sequence is replaced by all others. Fig. 3 depicts the layout of the peptide scan, where each binding site is represented by 15 rows and 21 columns. The original (wild-type) sequence occurs twice in each row: in the left-most spot and in the column of the substituent corresponding to the original residue. To evaluate the influence of single substitutions on DNA binding, we calculated the initial binding efficiency (i.e. to the wild-type peptide) as the average over all 30 original peptide spots plus or minus the 3-fold S.D. All binding efficiencies outside this range were considered significant. It is evident that certain basic amino acids had a significant influence on DNA binding to both sites. Considerable effects of amino acid exchanges on DNA binding are seen at the consensus motif lysine and arginine residues at positions 92, 94, 97, and 98, as well as tryptophan 102 in the first binding region and, more pronounced, at positions 263, 265, and 268 in the second binding region (cf. Fig. 2).

Replacement of Critical Lysine Residues Alters DNA Binding Properties but Not the Secondary Structure of EcoRII—Following the leads from the substitution experiment, both lysines (boldface letters) in each potential binding site (site I, RHFGKTRNEKRIRTW; site II, NSVSNRRKSRAGKSLRKSRAKGSLELH) were replaced by neutral alanine or acidic glutamic acid residues, yielding the double mutants K92A/K97A, K92E/K97E, K263A/K268A, and K263E/K268E and the quadruple mutants K92A/K97A/K263A/K268A, K92A/K97A/K263E/K268E, K92E/K97E/K263A/K268A, K92E/K97A/K263E/K268E, K92E/K97E/K263E/K268E.
K263A/K268A, and K92E/K97E/K263E/K268E. On expression in *E. coli*, wild-type and mutant EcoRII appeared in both the soluble and insoluble fractions (cf. Ref. 9). The quadruple mutants K92E/K97E/K263E/K268E and K92E/K97E/K263A/K268A produced the lowest yields of soluble endonuclease proteins. All EcoRII derivatives eluted from nickel-nitrilotriacetic acid columns under the same conditions as the wild-type protein. Western blot analysis of the purified wild-type and the eight DNA binding site mutant proteins confirmed their reactivity toward polyclonal EcoRII antibodies and the presence of one predominant band at the molecular weight of the wild-type protein (data not shown).

Three CD spectra for each EcoRII mutant enzyme were accumulated, and the secondary structure was calculated according to Chen et al. (25). Fig. 4 shows that the CD spectra of wild-type and the DNA binding site mutants do not reveal obvious differences in secondary structure of the proteins. Because of the lower protein concentrations available for the spectrometric measurements of the quadruple mutants K92E/K97E/K263E/K268E and K92E/K97E/K263A/K268A, their spectra exhibited a lower signal-to-noise ratio than those from the other proteins.

The DNA binding behavior of wild-type EcoRII and its derivatives mutated in one or both binding sites was compared by gel retardation assays in the absence of Mg²⁺ ions to prevent DNA cleavage (Fig. 5). Constant low DNA concentration and varying protein concentrations, covering at least 2 orders of magnitude, were employed to determine the apparent dissociation constants *K*_D(app) of EcoRII-DNA complexes. For the calculation of *K*_D(app) values, we made the simplifying assumption that given the excess of enzyme over DNA, only one substrate (S) binding site of the enzyme dimer (E) was occupied. The obtained constants shown in Fig. 5a were therefore calculated according to *K*_D = [S]/[E]/[SE].

The replacement of both K residues by A in DNA binding site I (K92A/K97A) led to a 6-fold decrease in DNA binding and the formation of slower migrating complexes in comparison to wild-type EcoRII, whereas the introduction of E in place of K (K92E/K97E) nearly abolished DNA binding. In contrast, both binding site II mutants showed dissociation constants comparable to wild-type. The *K*_D values of the quadruple mutants substituted in both binding sites exceeded those of the individual binding site mutants from which they were composed. Furthermore, the gel mobility shift assay with the quadruple mutant enzymes reproducibly separated a number of complexes distinct in size and/or conformation from the EcoRII wild-type-DNA complex (Fig. 5b). It was evident from competition experiments with a 1000-fold molar excess of unspecific or specific unlabeled DNA that the residual binding capacity of the EcoRII mutants was still recognition sequence-specific (data not shown).

**Impaired Restriction Activity of EcoRII DNA Binding Site Mutants**—The catalytic activity of the eight mutant enzymes was tested on linearized pBR322 DNA and compared with the wild-type enzyme (Fig. 6a). Alanine substitutions in DNA binding site I at positions 92 and 97 did not reduce cleavage activity compared with the EcoRII wild-type ENase. However, lysine exchanges to glutamic acid caused a strong decrease in DNA cleavage activity, without a concomitant change in recognition site specificity (Fig. 6a). This was apparently correlated to the extremely weak substrate binding of the mutant enzyme K92E/K97E (cf. Fig. 5), consistent with a repulsion of DNA phosphates by the introduced negative charges. Monitoring the time dependence of pBR322 cleavage confirmed the low but still measurable cleavage rate of K92E/K97E (data not shown).

Both mutations in DNA binding site II and all of the quadruple mutations resulted in catalytically inactive enzymes (Fig. 6a). Even after 2 h, no DNA hydrolysis could be detected under standard EcoRII reaction conditions (data not shown).

Table 1 summarizes the properties of all constructed EcoRII mutants and additionally includes the ability of the enzymes to be activated by site-containing oligonucleotides to cleave resistant DNA sites, as well as their efficiency of restricting phage λ.
reproduction in *E. coli* host cells. EcoRII can be activated in trans to cleave single, resistant recognition sites, e.g., on viral T3 and T7 DNA, by co-incubation with another, susceptible DNA (26) or even by short synthetic oligonucleotide duplexes carrying the canonical recognition sequence (22).

The alanine substitutions in binding site I yielded a mutant with a 6-fold increased $K_{D(app)}$ value for DNA that behaved like the wild-type in the other investigated parameters. This is not surprising, because the assays were run under optimal conditions. For example, the ratio of enzyme dimers to recognition sites of 1:2 employed in the cleavage reactions was sufficient to counteract the reduced substrate affinity of K92A/K97A. For the glutamic acid mutant in the same binding site, DNA binding was hardly detectable, and catalytic activity was very weak. These in vitro features of K92E/K97E match a significant decrease of phage $\lambda$ restriction to only 1 order of magnitude, as opposed to 4 orders of magnitude by the EcoRII wild-type and the K92A/K97A mutant. Both binding site II mutants were catalytically inactive in vitro and could not be transactivated but were still capable of restricting phage $\lambda$ by approximately 1–2 orders of magnitude. Conceivably, the physiological environment offers more appropriate conditions for the mutant enzymes, because K263A/K268A can be stimulated to cleave DNA in vitro in the presence of 5–50 mM Mn$^{2+}$. Restriction in vivo may also be related to tight repressor-like binding of the binding site II mutants to their very frequent recognition sites in the $\lambda$ genome. The four enzyme variants with mutations in both binding sites exhibited in vitro characteristics similar to those of the binding site II mutants but with significantly reduced DNA binding, and they no longer restricted phage $\lambda$ in the host cell.

**DNA Binding Site I Overlaps a Potential Catalytic Motif**—A search for catalytically relevant amino acid residues based on

---

2 M. Reuter and P. Mackeldanz, unpublished results.
the crystal structures of ENase-DNA complexes (cf. Refs. 11, 27, and 28 for review) detected at least three potential catalytic consensus motifs P(D/E)X3(D/E)X(K/R), two of which are in the vicinity of binding site I (Fig. 7a). Both acidic amino acid residues in the consensus motif were found to be important for cleavage in the ENases EcoRI and EcoRV (cf. review in Ref. 11). We have constructed mutants K96A, D130A, and E234A in the three potential catalytic motifs of EcoRII. Substituting glutamic acid at position 96 by alanine eliminated DNA cleavage (Fig. 6b) and binding (data not shown), whereas when alanine was introduced into equivalent positions of the two other potential catalytic motifs, 117PEX3DCK137 and 214DXXEL238, it was tolerated, and DNA cleavage activity was almost unaffected (Fig. 6b). These results clearly implicated Glu96 of the first potential catalytic motif located in DNA binding site I as critical for catalysis.

Sequence Alignments Identify DNA Binding Site II Analogues in Other Restriction Endonucleases—A BLAST search of EcoRII DNA binding sites I and II produced three ENases with similarities to binding site II, SsoI, NgoPII, and LlaII. All available amino acid sequences of their isoschizomers (1) were then screened to yield ScrFI, MthTI, and DpnII (Fig. 7b). Further close relatives of LlaII (MboI and MjO600) and of NgoPII (FnuDI) are not shown here because the respective sequence elements are nearly identical. Sequences related to DNA binding site II were also found in CglII and CviAI (Fig. 7b). The core sequences of the alignment (Fig. 7b, between vertical lines) were compared for every pair of enzymes, revealing similarities not only among truly homologous enzymes with identical recognition sites, such as LlaII and DpnII, but also, for example, between NgoPII and DpnII, the recognition sites of which share only the outer G:C pairs (Fig. 7c). However, the greatest similarity between EcoRII DNA binding site II and all other analyzed ENase sequences was observed to SsoI (Fig. 7c). It appears especially significant that the homology to SsoI as well as ScrFI is focal to DNA binding site II, as both enzymes are specific to CCNGG, a degenerate version of the EcoRII DNA recognition sequence.

To see whether the identified DNA binding site II analogues would be juxtaposed by multiple alignment of the whole set of enzymes (Fig. 7b), their total sequences were submitted to DIALIGN 2.0, a program designed to identify local similarities in functionally related proteins, the sequences of which are not necessarily globally related. DIALIGN 2.0 is particularly suitable for the task as it was the only of five alignment programs to identify all DNA binding sites in a set of 11 diverse helix-turn-helix proteins of vastly different size (24). DIALIGN 2.0 aligned the DNA binding site II analogues in EcoRII, SsoI, ScrFI, NgoPII, MthTI, LlaII, and DpnII as the regions of maximal similarity of these enzymes. The same maximal alignment score for the binding site II analogues of these enzymes, as well as of FnuDI and MboI (see above), resulted when a larger set of sequences that additionally included LlaI, NgoFVII, HpaII, and Cfr10I was analyzed, enzymes selected for resemblances to the enzymes shown in Fig. 7b. NgoFVII is a CglII homolog with a known recognition sequence (GCSGC) but with less similarity to DNA binding site II than CglII itself; HpaII is another type IIE enzyme (1) and LlaI (open reading frame 3) encodes a subunit of a complex restriction endonuclease (29) that, like CviAI, contains the KRXXXK motif.

Surprisingly, three different peptides of the type IIE ENase NaeI were identified by independent alignment routines. NaeI 58–75 (Fig. 7b) emerged from a FASTA homology search (cf. the procedure described under “Experimental Procedures”) of the C-terminal half of the EcoRII sequence, NaeI 188–205 emerged from the DIALIGN 2.0 analysis of the enzyme set shown in Fig. 7b, and NaeI 147–164 emerged from the alignment of the enlarged set of enzymes. All three NaeI peptides include a leucine corresponding to Leu9274 and a glycine corresponding to Gly247 of EcoRII DNA binding site II, the most conserved residue of the alignment in Fig. 7b. By random mutagenesis, Holtz and Topal (30) located functionally essential residues in all three of these NaeI peptides, including Thr93, Glu97, Gly155, and Gly196, and a recent domain analysis (31) placed the latter two peptides in the C-terminal DNA binding domain of NaeI, whereas peptide 58–75 is in the N-terminal part of the enzyme predicted to be involved in catalysis and dimerization. NaeI 58–75 was included in Fig. 7b because the Thr and Glu residues (boxed) implicated in catalysis (30) belong to the consensus residues of the alignment.

Although the BLAST search of EcoRII DNA binding site I did not pick up any other ENases, a similar peptide in SfiI was located by a Higgins-Sharp alignment (cf. the procedure described under “Experimental Procedures”) (Fig. 7b). SfiI is functionally related to EcoRII in requiring, and simultaneously cleaving, two copies of its recognition site; however, it does this not as a dimer but a tetramer (6). A mutational analysis of SfiI is not available.

**DISCUSSION**

**Deconstruction of the EcoRII-DNA Interaction**—We were interested in the structural basis of target recognition by the
restriction ENase EcoRII. Because crystallographic information was not available, we explored an alternative approach to the study of DNA-protein affinity, synthetic peptide scans. The procedure that allowed the identification of sequence-specific DNA binding to matrix-bound peptides comprised the following elements: unspecific DNA binding was lowered in the presence of Mg$^{2+}$ ions (Fig. 1, b and c) and suppressed in the presence of an excess of unlabeled unspecific competitor DNA, designed to differ from the target site-containing DNA at the dinucleotide level. Specific binding was only suppressed by an excess of unlabeled specific competitor DNA (Fig. 2). The peptide scan approach allowed us to delineate two potential DNA binding sites of EcoRII.

Type II DNA restriction ENases generally bind palindromic double-stranded DNA recognition sites in a symmetrical fashion (normally as dimers); a number of residues within one or two 5–15-amino acid regions, and shorter peptides or individual amino acids from different parts of the ENase monomer establish direct or water-mediated contacts with bases and phosphate groups of the DNA target site. Can any degree of specificity of such a complex DNA–protein interaction be preserved after breaking up a three-dimensional binding site into short peptide modules? Attempts to model DNA binding specificity of a protein by peptides in solution often fail (cf. Ref. 32). One exception is the dodopeptide WDGMAAGNAEIER comprising the extended chain region (underlined) of EcoRII. This peptide binds specifically to GAATTC with a $K_d = 3 	imes 10^4$ M$^{-1}$, i.e. about 5000 times weaker than EcoRI itself. Judging by inhibition studies of enzymes with related recognition sequences the target specificity of this peptide was retained (33).

In general, incorporation of DNA binding peptides into macromolecular frameworks appears to be necessary to overcome the loss of conformational and translational entropy accompanying sequence-specific association of a peptide with DNA (cf. Ref. 32). Thus, incorporation of peptides involved in sequence-specific DNA recognition, e.g. of zinc fingers into a phage display coat protein (34), or of a recognition helix from a helix-loop-helix protein into an antibody Fab domain (35), reproduced the DNA sequence specificities of the original proteins. Stanojevic and Verdine (32) introduced a new experimental approach to the quest they defined as the deconstruction of sequence-specific DNA–protein interactions by tethering DNA binding domains covalently to DNA. The close proximity of the peptide to its target sequence overcomes the entropic and energetic barriers to peptide DNA recognition.

Johnson et al. (36) extended the concept of induced fit from purely macromolecular to peptide-DNA interactions by demonstrating that the intrinsically flexible DNA binding peptides of basic helix-loop-helix and leucine zipper proteins assume α-helical conformation after binding DNA. Compared with peptides in solution, peptide scans offer two features that potentially facilitate sequence-specific induced fit to oligonucleotides: 1) the terminal covalent attachment of the peptide to the membrane reduces conformational entropy, and 2) the high local

**Fig. 7.** DNA binding sites in the restriction ENase EcoRII: conserved residues and similar motifs in other ENases. a, location of DNA binding sites and of potential catalytic motifs in the sequence of EcoRII. Amino acid residues marked with a “c” belong to potential catalytic consensus motifs identified by computer search. Homology studies of EcoRII DNA binding sites I and II. Alignments of the two binding sites (between vertical lines) and nine flanking residues are shown. All EcoRII residues and similar residues of other enzymes are boxed. Boldface indicates the occasional presence of homologous (* above the predominant residue) amino acids of its homology group(s), defined as follows: {K;R;H}, {D;N}, {E;Q}, {N;Q}, and similar residues (in capital letters) are shown. All residues known (at the left) and residues where >80% of the residues are homologous (+ above the predominant residue) indicates the occasional presence of other amino acids of its homology group(s).
concentration of each peptide may counterbalance the loss of several orders of magnitude of affinity associated with the transition from the macromolecular to the peptide/oligonucleotide level.

Both DNA binding sites identified in the EcoRII peptide scan are hydrophilic and characterized by an abundance of positively charged residues. Conversely, not all positively charged peptides in the peptide scan specifically bound to DNA, for example, LAVKTTCKDRWR (amino acids 262–273), the peptide in spot F8. Another relatively basic region, G3–G6, only bound to the specific probe in the absence of Mg$^{2+}$ (Fig. 1). The substitution experiment confirmed the importance of the basic residues of the discovered conserved motif, KXRXXK. A critical role seems to be played by Lys$^{268}$ because Arg was the only tolerated replacement (Fig. 3). At certain positions (Thr$^{39}$, Lys$^{97}$, Arg$^{261}$, and Arg$^{262}$), binding of the specific oligonucleotide was only abrogated by introducing acidic residues (Fig. 3).

Replacement of several amino acids by lysine and/or arginine was only abrogated by introducing acidic residues (Fig. 3). At certain positions (Thr$^{93}$, Lys$^{97}$, Arg$^{261}$, and Arg$^{262}$), binding of the specific oligonucleotide was only abrogated by introducing acidic residues (Fig. 3). Replacement of several amino acids by lysine and/or arginine in the peptides actually enhanced binding, indicating an increase in the strength of nonspecific interactions. However, where the substitution by basic amino acids had no effect, the contribution of the original residues (Thr$^{93}$, Trp$^{102}$, and Leu$^{272}$) to specific binding can be inferred to exceed the gain in strength of nonspecific electrostatic interactions.

Of course, the function of individual residues for sequence-specific DNA binding by the protein cannot be directly extrapolated from the results at the peptide level. In particular, we cannot predict whether the lysines of the motif KXRXXK form base or backbone contacts or are important for the geometry or the charge distribution at the DNA binding sites. However, we have demonstrated that these lysines play an important role in DNA binding, as all eight mutant enzymes in which these lysines were replaced by alanine or glutamic acid exhibited impaired DNA binding and/or cleavage (Table 1).

Although the substitution peptide scan experiment may emphasize the contribution of electrostatic binding at the expense of other types of interaction, basic amino acids do play a direct role in sequence-specific interactions of all six restriction ENases of which the target recognition sequence has been analyzed crystallographically (for a review, see Refs. 11, 27, and 28). To cite two examples of sequence-specific DNA binding predominantly involving lysine and arginine residues, the red domain of p50 (37, 38) and the tumor suppressor factor p53 (39), contain seven (p50) or six (p53) basic amino acids in their target recognition sites, of which four (p50) or two (p53) form hydrogen bonds to DNA bases, and the others of which bind to the phosphate groups. These latter interactions are termed the indirect readout, as they reflect the sequence-specific conformation of the DNA backbone (40, 41).

**Implications for the Structure of the EcoRII-DNA Complex**

How do the two DNA binding sites identified in the primary sequence of EcoRII relate to the substrate binding sites of the enzyme? EcoRII operates as a dimer binding two DNA recognition sequences at two substrate binding sites (2, 3, 7–10, 42), the functional equivalence of which is indicated by cleavage kinetics (43), by the capacity for simultaneous cleavage at both DNA recognition sequences independent of the orientation of the internal A/T pairs (10). Hence, neither substrate-binding site exhibits DNA strand preference in the recognition and cleavage process.

Type II ENases with palindromic recognition sites generally form rotationally symmetrical dimer-DNA complexes with the monomers establishing nonsymmetrical contacts to both strands of the recognition site and cleaving each strand separately (11). By this logic, the EcoRII dimer should require four catalytic sites, one for each strand of each of the two DNA target sequences. It is reasonable to propose that these are formed by two catalytic centers from each monomer. We have presented evidence (Fig. 6b) that Glu$^{96}$ in DNA binding site I is essential for cleavage and could be part of a classical catalytic motif P/D/E$X_{-4/5}$D/E$X_{+1/4}$K/R (Fig. 7a). Similar to the case of FokI (27, 12) a second catalytic motif is as yet unaccounted for, and in view of the 5-base pair distance between the scissile phosphodiester bonds, it is difficult to envisage cleavage of both DNA strands by a single catalytic center, as this would necessitate a rearrangement of the whole enzyme-substrate complex.

We hypothesize that both of the predicted DNA binding sites I and II participate in the formation of each of the two equivalent substrate-binding sites of EcoRIII. This configuration is consistent with the behavior of the DNA binding site mutants. For example, the substitutions K92E/K97E in binding site I nearly abolished binding of single site substrates to the enzyme dimer, so both EcoRII substrate binding sites must have been affected. In the quadruple mutants, binding site II governed catalytic activity, whereas binding site I determined the stability of the enzyme-DNA complex. Glu$^{96}$, which is essential for catalysis is also located in binding site I (Figs. 5–7; Table 1). In summary, changes in either of the identified DNA binding sites always affected binding or cleavage at both substrate-binding sites of the EcoRII dimer. It still remains to be shown whether the substrate binding sites of the EcoRII dimer are sequestered to the monomers or involve DNA binding sites from both monomers.

In general, the organization of EcoRII functional regions (Fig. 7a) derived from our studies resembles that proposed for NaeI, another member of the type IIE ENases. For NacI, one functional domain located by random mutagenesis near the N terminus includes acidic, possibly catalytic amino acid residues, and a basic region where amino acid substitutions attenuate DNA binding. The second, more C-terminal domain can be expressed separately and includes a larger basic region required for DNA binding (30, 31).

**Related DNA Binding Motifs in Diverse Restriction ENases**

The great similarity of DNA binding site II of EcoRII (C/C-WGG) to a sequence in SsoII (C/CNCGG) and its considerable resemblance to a ScrFI (CC/NGG) sequence lends independent support to a role of this site in sequence-specific DNA recognition (Fig. 7b), especially because EcoRII is a type IIE enzyme, whereas SsoII and ScrFI are mutually unrelated type II enzymes. Their distinct cleavage preferences necessitate differences in DNA recognition sequence specificity, which may be reflected in the sequence differences of the EcoRII DNA binding site II analogues.

Unlike DNA methyltransferases (MTases), type II ENases have not been subtyped according to patterns of secondary structure, catalytic and DNA binding motifs (cf. Ref. 44), with the exception of a group of ENases homologous to EcoRI, in which this assignment was based on the EcoRII 3D structure (45, 46). Janulaaties (46) also mentioned the existence of islands of sequence homology in other groups of ENases with related recognition sites but as yet unknown DNA binding sites. A different approach was taken by Jeltsch et al. (44), who attempted to reconstruct evolutionary trees of type II ENases by progressive, multiple sequence alignments and to assess the significance of the resulting groups of pairwise related enzymes by a multistep Monte-Carlo analysis of their recognition sequences. Although enzymes within the same group usually had similar recognition sequences, EcoRII and NpoPII (GG/CC) emerged as the closest relatives. This result agrees with the local sequence similarity of the two enzymes discovered by the BLAST search and confirmed by DIALIGN 2.0 (cf. the proce-
due described under “Results”).

The target-recognition domains of (cytosine-C5) DNA MTases have a conserved general structure (47, 48). Several attempts have been made to find common DNA recognition principles in cognate ENases and MTases. Sequence similarities between the region of EcoRII encompassing Arg265 and Lys268 (part of KRXRRK in binding site II) and MTases specific to CCWGG or CCNGG, in an RXXK motif of their predicted target recognition domains (49), were already discovered by Kossykh et al. (50). In earlier, related studies, Janulaitis et al. (51) identified conserved motifs shared by other cognate ENases and MTases, whereas Swaminathan et al. (52) found scattered, short fragments of homology between several ENases and MTases sharing recognition sites with central CG.

Our homology study (Fig. 7, b and c) identified sequences related to DNA binding site II in several ENases with terminal G:C or C:G base pairs as the only common feature of their diverse recognition sequences (Fig. 7b). These peptides may constitute modules for sequence-specific DNA recognition, specialized to read C:G or G:C in different contexts. Their overall similarity is striking, and it will be interesting to establish how the differences between them relate to their specific DNA targets and whether other families of DNA recognition modules exist. Peptide scans, which have proved their utility for identifying DNA binding regions in proteins specifically interacting with DNA, could provide further clues.

Acknowledgments—We are very grateful to Andrea Kunz for help in DNA sequencing of EcoRII mutants and to Ursula Schernke for skilled technical assistance. We thank Jürgen Alves (Zentrum Biochemie, Medizinische Hochschule Hannover, Institut für Biophysikalische Chemie) for his advice and help with CD spectrometry. We are obliged to Bayer, E., and Pingoud, A. (1995) Nature Struct. Biol. 2, 450–457.

Pein, C.-D., Reuter, M., Cech, D., and Krüger, D. H. (1989) J. Biol. Chem. 264, 581–586.

McLean, K. E., Burton, D. R., and Ghazal, P. (1995) Proc. Natl. Acad. Sci. U. S. A. 92, 5214–5218.

Johnson, H. P., Lindstrom, J., Basse, W. A., and van Hipped, P. H. (1994) Proc. Natl. Acad. Sci. U. S. A. 91, 11163–11167.

Ghosh, G., Van Duyne, G., Ghosh, S., and Sigler, P. B. (1995) Nature 373, 303–310.

Müller, C. W., Hey, F. A., Sedeoka, M., Verdine, G. L., and Harrison, S. C. (1995) Nature 373, 311–317.

Cho, Y., Gorina, S., Jeffrey, P. D., Pavletich, N. P. (1994) Science 265, 346–355.

Otwonowski, Z., Schwitz, R. W., Zhang, R. G., Lawson, C. L., Joachimiak, A., Marmorstein, R. Q., Luisi, B. F., and Sigler, P. B. (1988) Nature 335, 321–329.

Wenz, C., Jeltsch, A., and Pingoud, A. (1996) J. Biol. Chem. 271, 5565–5573.

Pein, C.-D., Reuter, M., Meisel, A., Cech, D., and Krüger, D. H. (1991) J. Biol. Chem. 266, 15315–15322.

Ghosh, G., Van Duyne, G., Ghosh, S., and Sigler, P. B. (1995) Nature 373, 303–310.

Müller, C. W., Hey, F. A., Sedeoka, M., Verdine, G. L., and Harrison, S. C. (1995) Nature 373, 311–317.

Cho, Y., Gorina, S., Jeffrey, P. D., Pavletich, N. P. (1994) Science 265, 346–355.

Otwonowski, Z., Schwitz, R. W., Zhang, R. G., Lawson, C. L., Joachimiak, A., Marmorstein, R. Q., Luisi, B. F., and Sigler, P. B. (1988) Nature 335, 321–329.

Wenz, C., Jeltsch, A., and Pingoud, A. (1996) J. Biol. Chem. 271, 5565–5573.

Pein, C.-D., Reuter, M., Meisel, A., Cech, D., and Krüger, D. H. (1991) J. Biol. Chem. 266, 15315–15322.

Ghosh, G., Van Duyne, G., Ghosh, S., and Sigler, P. B. (1995) Nature 373, 303–310.

Müller, C. W., Hey, F. A., Sedeoka, M., Verdine, G. L., and Harrison, S. C. (1995) Nature 373, 311–317.

Cho, Y., Gorina, S., Jeffrey, P. D., Pavletich, N. P. (1994) Science 265, 346–355.

Otwonowski, Z., Schwitz, R. W., Zhang, R. G., Lawson, C. L., Joachimiak, A., Marmorstein, R. Q., Luisi, B. F., and Sigler, P. B. (1988) Nature 335, 321–329.

Wenz, C., Jeltsch, A., and Pingoud, A. (1996) J. Biol. Chem. 271, 5565–5573.

Pein, C.-D., Reuter, M., Meisel, A., Cech, D., and Krüger, D. H. (1991) J. Biol. Chem. 266, 15315–15322.

Ghosh, G., Van Duyne, G., Ghosh, S., and Sigler, P. B. (1995) Nature 373, 303–310.

Müller, C. W., Hey, F. A., Sedeoka, M., Verdine, G. L., and Harrison, S. C. (1995) Nature 373, 311–317.

Cho, Y., Gorina, S., Jeffrey, P. D., Pavletich, N. P. (1994) Science 265, 346–355.

Otwonowski, Z., Schwitz, R. W., Zhang, R. G., Lawson, C. L., Joachimiak, A., Marmorstein, R. Q., Luisi, B. F., and Sigler, P. B. (1988) Nature 335, 321–329.

Wenz, C., Jeltsch, A., and Pingoud, A. (1996) J. Biol. Chem. 271, 5565–5573.
Regions of Endonuclease EcoRII Involved in DNA Target Recognition Identified by Membrane-bound Peptide Repertoires

Monika Reuter, Jens Schneider-Mergener, Dagmar Kupper, Andreas Meisel, Petra Mackeldanz, Detlev H. Krüger and Cornelia Schroeder

J. Biol. Chem. 1999, 274:5213-5221.
doi: 10.1074/jbc.274.8.5213

Access the most updated version of this article at http://www.jbc.org/content/274/8/5213

Alerts:
• When this article is cited
• When a correction for this article is posted

Click here to choose from all of JBC's e-mail alerts

This article cites 52 references, 17 of which can be accessed free at http://www.jbc.org/content/274/8/5213.full.html#ref-list-1