

Changing endonuclease *EcoRII* Tyr308 to Phe abolishes cleavage but not recognition: possible homology with the Int-family of recombinases

Michael D. Topal* and Michael Conrad

Lineberger Comprehensive Cancer Center, Department of Pathology and Curriculum in Genetics, University of North Carolina Medical School, Chapel Hill, NC 27599-7295, USA

Received March 12, 1993; Accepted April 20, 1993

ABSTRACT

Endonuclease *EcoRII* is one of a group of type II restriction enzymes, including *NaeI*, *NarI*, *BspMI*, *HpaII*, and *SacII*, that require binding of an enhancer sequence to cleave DNA. Comparison of the *EcoRII* amino-acid sequence with the amino-acid consensus motifs that differentiate between recombinase families uncovered similarity between a 29 amino-acid sequence in the carboxyl end of *EcoRII* and the motif defining the integrase family of recombinases. This similarity implied that *EcoRII* tyrosine 308 should be involved in catalyzing hydrolysis of the scissile bond. Site-directed mutagenesis was used to mutate Tyr308 to Phe. The phenylalanine-substituted enzyme could not cleave T5 DNA under conditions in which wild-type enzyme completely cleaved this DNA. The Tyr308 to Phe mutation abolished cleavage activity but not specific binding to DNA. No evidence was found for the existence during the cleavage reaction of a covalent linkage between Tyr308 and DNA.

INTRODUCTION

The *NaeI*-type restriction endonucleases, *BspMI*, *HpaII*, *NaeI*, *NarI*, *SacII* [1, 2], and *EcoRII* [3, 4], exhibit recognition sites that are resistant to cleavage; this resistance can be overcome by the addition of an enhancer sequence either cis or trans to the cleavage site. Studies of the interactions of *NaeI* with DNA define cleavage-enhancer as the cognate recognition sequence plus flanking sequences that direct binding to the enhancer-binding site on the enzyme [5]. Occupation of this site stabilizes an active conformation of the *NaeI* dimer [6].

As expected for an enzyme that has two DNA binding sites that require the same recognition sequences but recognize different families of flanking sequences, *NaeI* has been shown to induce loops with *NaeI* bound at their base in DNAs containing multiple recognition sites [7]. This ability to juxtapose and cleave distant recognition sites is reminiscent of site-specific recombination and transposition. Therefore, we compared the

sequence for *EcoRII* endonuclease [8] with the amino-acid consensus sequences (motifs) that differentiate between recombinase families (reviewed in [9]); we report here possible homology between a 29 amino-acid sequence in the carboxyl end of *EcoRII* and the motif defining the integrase-family of recombinases. To our knowledge this represents some of the first evidence of an evolutionary relationship between a restriction enzyme and a protein family.

One amino acid of the *EcoRII* endonuclease, tyrosine 308, was predicted, based on its position within the *EcoRII* amino-acid sequence showing similarity with the integrase motif, to be involved in catalyzing hydrolysis of the phosphodiester backbone. Changing Tyr308 to Phe by site-specific mutagenesis abolished the DNA-cleavage activity of the *EcoRII* endonuclease, but not its specific binding to DNA.

MATERIALS AND METHODS

Amino-acid sequence alignments

Sequence comparisons were made using the Sequence Analysis Software Package by Genetics Computer Group, Inc. Matches were maximized using BestFit, which is based on the local homology algorithm of Smith and Waterman [10, 11] and Needleman and Wunsch [12]. The significance of the matches found were determined by a Monte Carlo analysis [10, 11]: one of the sequences was jumbled multiple times and the average quality of match to the jumbled sequences compared to the quality of the putative match.

Site-specific mutation of pR224

Ten μg supercoiled pR224 was nicked using 80 U *EcoRI* (Promega) (as described elsewhere [13]). The nicked DNA was extracted 2 \times with phenol, extracted 3 \times with ether, and desalted through Sephadex G-25. 50 μg of DNA was reacted with 100 U Exo III (New England Biolabs) for 20 min at 37°C. Gapped DNA was extracted 3 \times with 10 μl of Strataclean resin (Stratagene) and alcohol precipitated. Gap-size was determined on agarose gels after digestion of single-strand DNA with S1

* To whom correspondence should be addressed at: The Lineberger Comprehensive Cancer Center, University of North Carolina Medical School, Chapel Hill, NC 27599-7295, USA

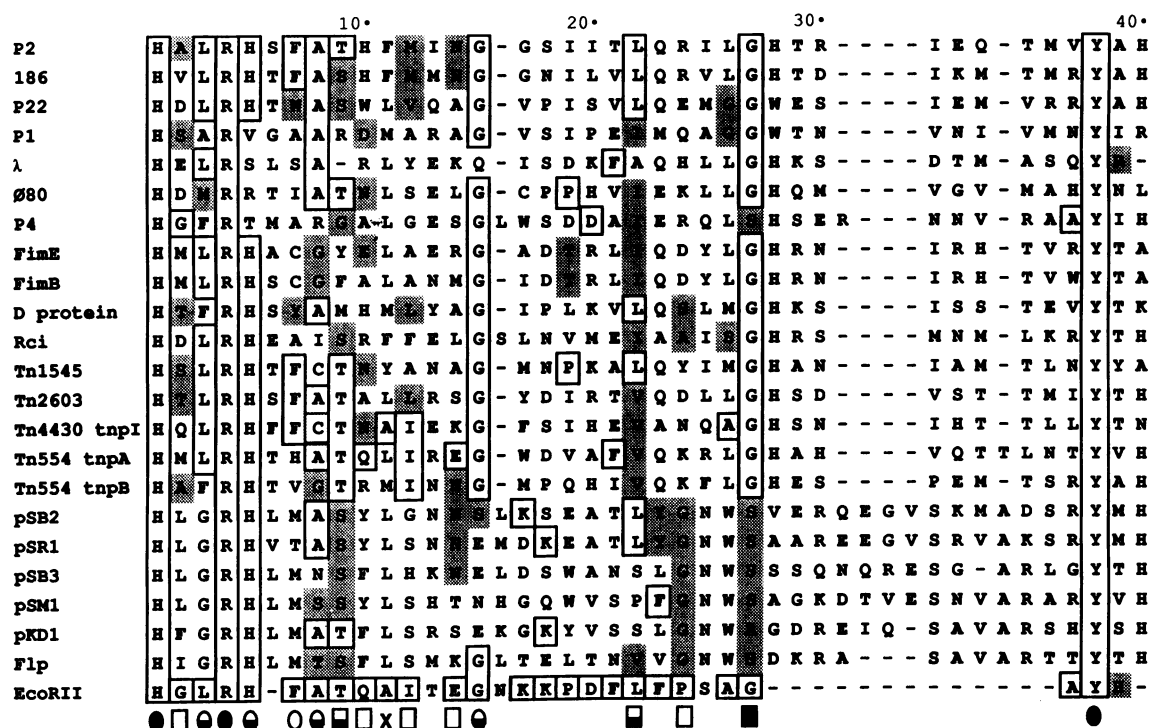


Figure 1. Comparison of the 40-residue motif sequences of *EcoRII* endonuclease with the integrase family of recombinases (single letter amino-acid code). Boxed residues indicate conservation between *EcoRII* and any of the other proteins. Shaded residues refer to residues that fall within the same exchange group [25]. Extents of conservation are indicated below *EcoRII* residue positions filled symbol, 100%, half filled, >50%, open symbol, >30%; circles and squares exclude and include residues in the same exchange group, respectively. An X indicates a residue that is conserved among the recombinases but not by *EcoRII*. Integrase sequences are from: phages P2, 186, P22, P1, λ , ϕ 80, and P4 [17]; *E. coli* switch proteins FimE and FimB [26]; D protein [27]; Rci protein [28]; transposons Tn554 (tpA and tpB) [29], Tn1545 [30], Tn2603 [31], Tn4430 [32]; yeast plasmids pSB2, pSR1, pSB3, pSM1, pKD1 [33], Flp protein [34].

nuclease (not shown). 1.0 μ g of gapped DNA was annealed to a molar excess of primer pCGCGGGGGCTTTCCACGATACT or pCAGTATCGTGGAAAGCCCCCGC, synthesized as described elsewhere [5], in 10 mM MgCl₂ and 5 mM β -mercaptoethanol (pol buffer) at 60°C for 1 min and cooled to 4°C over 10 min in 12 μ l. 2 μ l of 5 mM rATP and 4 μ M dNTPS and then 1 μ l containing 1 U of Klenow pol I (BRL) and 1000 U of phage T4 ligase (NEB) were added. After 15 min at 25°C, 7 μ l 0.7 mM rATP and 0.6 mM dNTPs in pol buffer containing 1 U Klenow pol I and 1000 U T4 ligase were added and the reactions continued for 7 hrs. Formation of closed-circular DNA was followed by gel electrophoresis in the presence of ethidium bromide (results not shown). The reactions were heated to 60°C for 20 min, desalted through sephadex G-25, and treated with dam methylase (NEB) according to manufacturer's protocol, to reduce mismatch repair. Methylated DNA was desalted and transfected into permeabilized DH5 α cells (BRL). Plasmid DNA was isolated from transfected cells grown in 25 mls media containing ampicillin (60 μ g/ml). 10 μ g of DNA was cut with 75 U of *EcoNI*, the DNA was desalted and 2 μ g retransfected; this was repeated. Plasmid DNA from 10 colonies were analyzed and all were *EcoNI* resistant. Six of these, from oligonucleotides 1 and 2, had acquired a *SacII* site. The region of interest was sequenced using the chain termination method and found to contain the desired mutation (not shown).

RESULTS AND DISCUSSION

The site-specific recombinases fall into at least two families, based on their protein structure and chemistry of strand breakage

(reviewed in [9]). The resolvase/invertase family is characterized by transient attachment of the protein to the broken DNA through a 5' phosphodiester linkage to a serine residue located within a highly conserved region of sequence in the N-terminal catalytic domain (reviewed in [14]). The integrase family is characterized by transient attachment of the protein to the broken DNA through a 3' phosphodiester linkage to a tyrosine residue located within a conserved region of sequence in the C-terminal catalytic domain [15]. This tyrosine residue is located within a 40-residue motif that is conserved among the integrase family [16] (Fig. 1). Three of the forty amino acids (His, Arg, and Tyr) are perfectly conserved and alteration of these amino acids block recombination without affecting recombinase binding to DNA [15, 17, 18]. We compared the *EcoRII* endonuclease sequence to the sequence motifs that define the different DNA-cleaving protein families; similarity was found with the integrase-motif.

EcoRII endonuclease is an ~45 kDa protein [8, 19] that recognizes the sequence CC(A or T)GG and cleaves the phosphodiester bond preceding the first C leaving a 5-base 5' overhang [20, 21]. This type of overhang is consistent with that generated by double-strand cleavage by the Int-family proteins. *EcoRII* contained a sequence similar to the 40-residue motif of the integrase family and conserved the His1, Arg4, and Tyr38 (of the 40-residue motif), which are perfectly conserved among all 22 integrase-family proteins so far sequenced (Fig 1). In addition, the integrase-family has 8 highly (>50%) conserved amino acids, Leu3, His5, Ala8, Leu11, Gly15, Gly27, His28, and His40; *EcoRII* conserved all of these except Leu11, His28, and His40. Defining residues in the same exchange group as conserved, 16 amino acids in the discovered integrase-motif

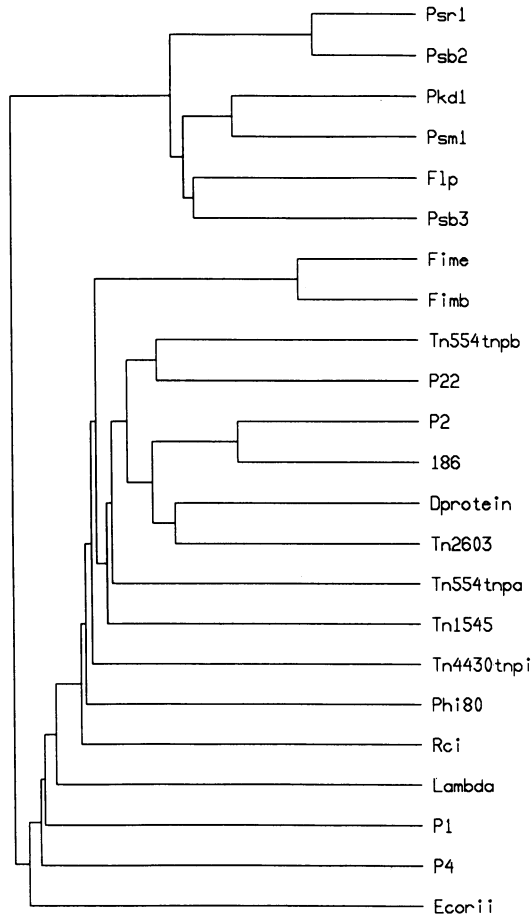


Figure 2. A dendrogram indicating similarity relationships among the motifs defining the Int-family of recombinases and the putative Int-motif in *EcoRII* endonuclease. The horizontal branch lengths are proportional to the similarity between the sequences. The dendrogram was generated as described in Material and Methods using PileUp from the GCG package of sequence analysis software running on a VAX 6620 computer.

```

EcoRII 281 HGLRHFAATQAITTEGNKPDFLFPSA.....GAYHDTFFPVENLRML 321
          ||:| :| .|:|.:. .| :... :|| .| .: :|: |:
P4       348 HGFRTMARGALGESGLWSDDAIERQLSHSERNNVRAAYIHTSEHLDERRLM 398
          †
    
```

Figure 3. Similarity between the putative Int-motif of *EcoRII* and the Int-motif from the phage P4 integrase. Alignment was maximized by computer using BestFit as described in Materials and Methods. The alignment shows 54% similarity and 27% identity between aligned amino acids. Comparison value between amino acids ≥ 0.5 , ::; comparison value ≥ 0.1 , .; comparison value = 1, |.

region of *EcoRII* are >30% conserved among all the 22 proteins; the motif in *EcoRII* is only 29 residues long.

Of the 22 integrase-family proteins, the putative Int-motif of *EcoRII* was found by pairwise alignment by computer to be most similar to the Int motif contained within the integrase protein from phage P4 (Fig. 2). Using BestFit, from the Sequence Analysis Software Package by Genetics Computer Group, Inc., to maximize alignment showed that 7 of the first 19 amino acids of the two motifs are identical and another 4 are conserved without allowing any gaps for alignment purposes (Fig. 3). Using Monte Carlo statistical analysis, this similarity was found to be

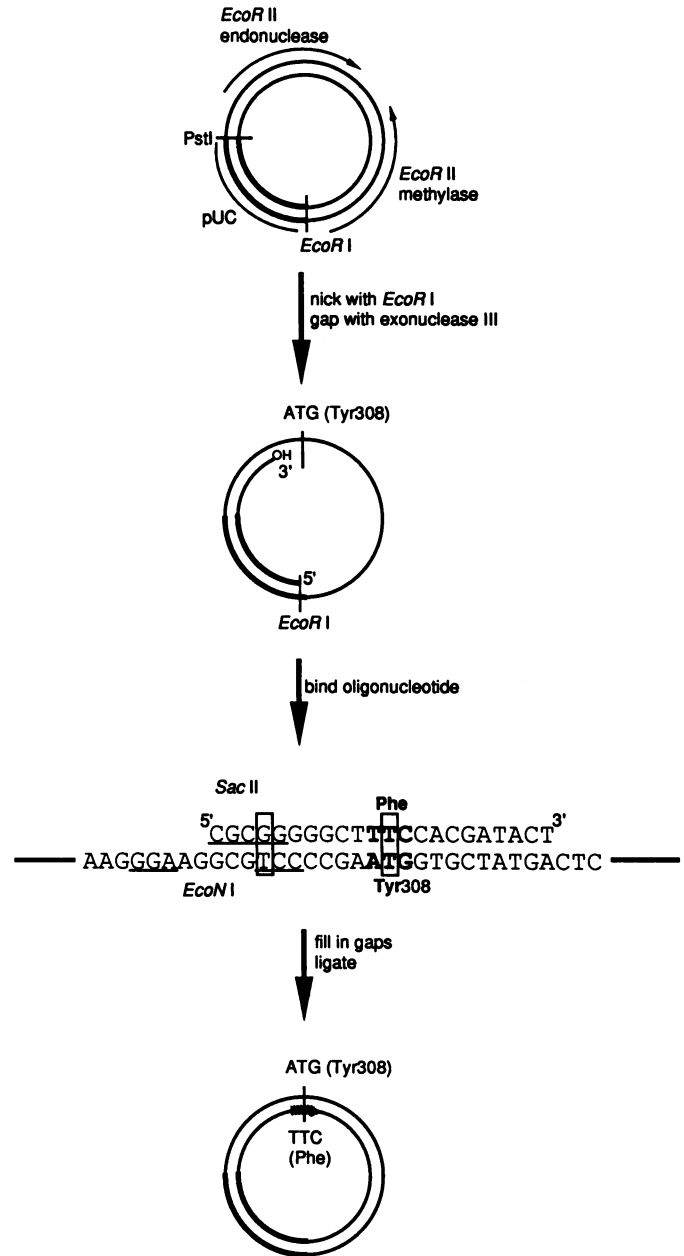


Figure 4. Site-specific mutation of pR224. Tyr308 in the *EcoRII* endonuclease gene was changed to Phe by nicking the plasmid with *EcoRI* in the presence of ethidium bromide; creating a gap by extending the nick using Exo III; binding a complementary oligonucleotide that creates mismatches to change Tyr to Phe concomitantly eliminating an *EcoNI* restriction site and creating a *SacII* restriction site; fill-in and ligate to produce closed-circular DNA. These DNA molecules were used to transfect *E. coli* followed by repeated isolation, restriction with *EcoNI*, and transfection to select only mutated DNA molecules.

6.5 standard deviations (based upon 100 randomizations) above the probability expected from matching *EcoRII* to jumbled P4 Int sequences [10, 11].

The last sixteen amino acids of the alignment shown in Fig. 3 match with a similarity score approximately 5 SDs above random. Tyr308 is found in this region and was aligned by the computer analysis opposite the tyrosine in P4 integrase that forms a phosphotyrosyl intermediate (†). The middle part of the integrase motif is the most variable among members of the Int-family. This

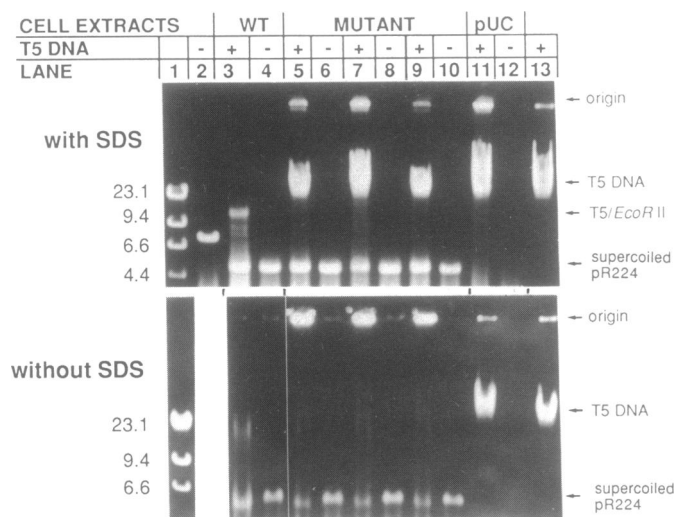


Figure 5. *EcoRII* activity of extracts from cells containing wild-type and mutant pR224. Substrate T5 DNA was incubated for 15 min with cell extracts from cells grown with wild-type pR224, pR224 containing mutant *EcoRII*, and only pUC, as indicated. **a**, Reaction products were incubated with hot SDS (equal volume of 0.5% SDS, 50 mM NaOH, and 5 mM EDTA at 68°C for 30 min [34]) to release protein. Lane 1, λ -*Hind* III digest molecular-weight markers; Lane 2, *EcoRI* cleavage of wt pR224 cell extract; the 7.5 kb linear plasmid band was present. Lanes 3 & 4, cleavage of T5 DNA by extract containing wild type *EcoRII* protein; Lanes 5 through 10, lack of cleavage and recovery of DNA from extracts from cells containing 3 independent pR224 mutants. Lanes 11 & 12, lack of cleavage of T5 DNA by pUC extracts, which lack *EcoRII* endonuclease activity; Lane 13, T5 DNA alone. **b**, Repeat of the reactions in the corresponding (a) lanes but without treatment with hot SDS. Mutant *EcoRII* endonuclease still does not cleave T5 DNA, but now product does not enter the gel; see origin in Lanes 5 through 10. Wild type and pUC extracts give same results as in (a). Indicated alongside the figure is the location of uncleaved T5 DNA, T5 DNA cleaved by *EcoRII* (see Lane 3), and the position of the pR224 plasmid DNA. Note the absence of the latter in the pUC lanes (11 & 12) and in lane 2, in which pR224 has been cleaved by *EcoRI*.

region is deleted in the *EcoRII* enzyme. The overall match between the region of *EcoRII* containing the putative Int motif and the analogous region of P4 integrase (Fig. 3) is 4.4 SDs above a random match. This level of similarity is significantly better than the two to three SDs above randomness that generally characterize the relationships among the integrases in Table 1 [16], which have diverged significantly from each other, but still maintain similar function.

The homology between *EcoRII* protein and the integrase-family of recombinases suggests that Tyr308 should be required for *EcoRII* activity but not for its binding to DNA. Plasmid pUC18 containing the *EcoRII* endonuclease and methylase genes was, therefore, obtained (generously provided by A. Bhagwat, Wayne State Univ.) and Tyr308 changed to Phe using oligonucleotide site-directed mutagenesis techniques (Fig. 4). This is a conservative change: only a hydroxyl group is removed that may be required for nucleophilic attack at DNA phosphate.

To mutate Tyr308 to Phe, oligonucleotides were synthesized against both strands of pUC18. These oligonucleotides introduced two basepair changes (Fig. 4) that simultaneously mutated Tyr308 to Phe, eliminated a unique *EcoNI* restriction site, and created a unique *SacII* site. These changes in restriction character of the DNA were used for unambiguous selection of mutant phage. DNA sequencing confirmed that the expected base changes were the only ones produced (not shown).

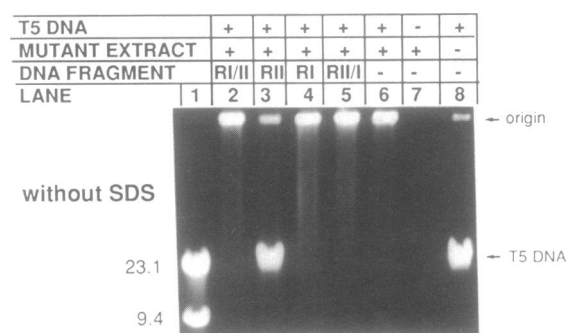


Figure 6. Mutant *EcoRII* protein binding to DNA requires an *EcoRII* site. Lane 1, λ -*Hind* III digest molecular-weight markers; Lane 2, cell extract of *EcoRII* mutant reacted with 0.2 μ g of T5 DNA as in Fig. 5b (Lanes 5–9): a 32 bp DNA fragment was added before the addition of T5 DNA. The DNA fragment was heteroduplex in restriction site (*EcoRI* and *EcoRII* sites oppose each other). Lane 3, same as Lane 2 except fragment contains *EcoRII* site; Lane 4, fragment with *EcoRI* site; Lane 5, fragment was heteroduplex complement to fragment in Lane 2; Lane 6, mutant extract with no added fragment, note product at origin; Lane 7, cell extract from mutant *EcoRII* showing that bands are not contributed by extract in the absence of added T5 DNA; Lane 8, phage T5 DNA.

To determine the effect of the Tyr308 to Phe mutation on *EcoRII* endonuclease cleavage and binding activity, extracts were prepared from cells containing the mutant and wildtype genes. Extracts from cells containing wild-type *EcoRII* endonuclease completely cleaved phage T5 DNA substrate (Fig. 5, Lanes 3 & 4). Under the same conditions, extracts from cells containing Phe308 mutant *EcoRII*, however, were unable to cleave substrate T5 DNA (Fig. 5, Lanes 5–10) and required hot SDS to release mutant *EcoRII* protein from uncleaved T5 DNA (Fig. 5, compare Lanes 5–10 with and without SDS); in the absence of SDS the protein-DNA complex could not enter the gel. Thus, substitution of Phe for Tyr at position 308 eliminated cleavage activity but not *EcoRII* binding to DNA.

The mutant *EcoRII* endonuclease bound DNA specifically at *EcoRII* sites: addition of enzyme-saturating amounts of a DNA fragment containing an *EcoRII* site released T5 DNA from the mutant enzyme as judged by the reappearance of the uncleaved T5 DNA band upon electrophoresis (Fig. 6, Lane 3). Addition of a DNA fragments lacking a properly base paired *EcoRII* recognition site were not able to inhibit protein binding (Fig. 6, Lanes 2, 4, 5).

A hallmark of the integrase proteins is their formation of a covalent intermediate through a 3' phosphodiester linkage to a tyrosine residue [15]. We have been unable to detect a covalent intermediate between labeled-DNA and *EcoRII* using the following methods: *EcoRII*-mediated DNA cleavage was interrupted with various agents, including, hot SDS, NaOH, and EDTA, tried alone and in combination, and the reaction products either electrophoresed on denaturing gels or hydrolyzed in acid and analyzed by thin layer chromatography to visualize the presence of phosphotyrosine. This negative result implies that either a covalent intermediate is not formed during cleavage or it is readily hydrolyzed making its isolation in an aqueous environment very difficult.

Thus, we were unable to isolate a covalent intermediate between *EcoRII* and DNA using the same methods that were used to demonstrate covalent intermediates between λ -integrase and DNA [15] and between topoisomerases and DNA [22, 23]. Yet,

the Tyr308 -OH was found to be required for DNA cleavage but not for specific DNA recognition by *EcoRII*. These results lead us to speculate that, whereas the integrases have evolved to nick and rejoin DNA, *EcoRII* has apparently evolved to cleave DNA efficiently. Thus *EcoRII* may have evolved to use the high concentration of water (55 M) present in the reaction environment, rather than another DNA strand, to displace the tyrosyl-DNA intermediate. Amino-acid differences between *EcoRII* and the integrase proteins that facilitate water access to the active site of *EcoRII* could lead to more efficient hydrolysis of the phosphotyrosyl bond and explain our inability to isolate the *EcoRII*-DNA covalent intermediate.

The presence of a putative Int-motif in the *EcoRII* amino-acid sequence implies an evolutionary relationship between *EcoRII* and the integrases. We have no evidence as to which, endonuclease or integrase, is the evolutionary precursor, but we would not be surprised to find that the cell, to defend itself from invading DNAs, recruited an integrase activity that could specifically cleave DNA and that could be inhibited from host-DNA cleavage by DNA modification. Restriction, however, could have preceded integration: restriction enzymes have been found to induce genomic rearrangements when introduced into cells (e.g., [24]), but a role in genetic recombination has not been shown.

ACKNOWLEDGEMENTS

We thank Bonnie Baxter for help trying to isolate the covalent *EcoRII*-DNA intermediate and Ashok Bhagwat for providing the *EcoRII* clone. This work was supported by University of North Carolina institutional funds.

REFERENCES

- Conrad, M. & Topal, M. D. *Proc. Natl. Acad. Sci. U. S. A.* **86**, 9707–9711 (1989).
- Oller, A. R., Broek, W. V., Conrad, M. & Topal, M. D. *Biochem.* **30**, 2543–2549 (1991).
- Kruger, D. H., Bareak, G. J., Reuter, M. & Smith, H. O. *Nucleic Acids Res.* **16**, 3997–4008 (1988).
- Gabbara, S. and Bhagwat, A. S. (1992) *J. Biol. Chem.* **267**, 18623–18630.
- Yang, C. C. and Topal, M. D. (1991) *Biochem.* **31**, 9657–9664.
- Baxter, B. K. and Topal, M. D. (1993) *Biochem.* **33**, 000–000.
- Topal, M. D., Thresher, R. J., Conrad, M. & Griffith, J. *Biochem.* **30**, 2006–2010 (1991).
- Bhagwat, A. S., Johnson, B., Weule, K. & Roberts, R. J. *J. Biol. Chem.* **265**, 767–773 (1990).
- Craig, N. L. *A. Rev. Genet.* **22**, 77–105 (1988).
- Smith, T. F. and Waterman, M. S. (1981) *Ad. Appl. Math.* **2**, 482–489.
- Lipman, D. J., Wilbur, W. J., Smith, T. F., and Waterman, M. S. (1984) *Nucl. Acids Res.* **12**, 215–226.
- Needleman, S. B. and Wunsch, C. D. (1970) *J. Mol. Biol.* **48**, 443–453.
- Müller, W., Weber, H., Meyer, F. & Weissman, C. *J. Mol. Biol.* **124**, 343–358 (1978).
- Sherratt, D. in *Mobile DNA* (Berg, D. E. & Howe, M. M., eds.) Am. Soc. Microbiol. (Wash. D. C., 1989) pp.163–184.
- Pargellis, C. A., Nunes-Düby, S. E., Moitosa de Vargas, L. & Landy, A. *J. Biol. Chem.* **263**, 7678–7685 (1988).
- Argos, P., Landy, A., Abremski, K., Egan, J. B., Haggard-Ljungquist, E., Hoess, R. H., Kahn, M. L., Kalionis, B., Narayana, S. V. L., Pierson III, L. S., Sternberg, N. & Leong, J. M. *EMBO J.* **5**, 433–440 (1986).
- Wierzbicki, A., Kendall, M., Abremski, K. & Hoess, R. *J. Mol. Biol.* **195**, 785–794 (1987).
- Prasad, P. V., Young, L.-J. & Jayaram, M. *Proc. Natl. Acad. Sci. U. S. A.* **84**, 2189–2193 (1987).
- Kosykh, V. G., Puntezhis, S. A., Buryanov, A. A. *Biochem. (U. S. S. R.)* **47**, 524–529 (1982).
- Bigger, C. H., Murray, K. & Murray, N. E. *Nature (New Biol.)* **244**, 7–10 (1973).
- Boyer, H. W., Chow, L. T., Dugaiczzyk, A., Hedgpeth, J. & Goodman, H. M. *Nature (New Biol.)* **244**, 40–43 (1973).
- Tse, Y.-C., Kirkegaard, K. & Wang, J. C. *J. Biol. Chem.* **255**, 5560–5565 (1980).
- Rowe, T. C., Tewey, K. M. & Liu, L. F. *J. Biol. Chem.* **259**, 9177–9181 (1984).
- Schiestl, R. H. and Petes, T. D. (1991) *Proc. Natl. Acad. Sci. USA* **88**, 7585–7589.
- Dayhoff, M. O., Schwartz, R. M. & Orcott, B. L. In Dayhoff, M. O. (ed.) *Atlas of Protein Sequence and Structure*. National Biomedical Research Foundation, Wash., D.C., Vol. 5, Suppl. 3, pp. 345–352 (1978).
- Klemm, P. *EMBO J.* **5**, 1389–1393 (1986).
- Lane, D., de Feyter, R., Kennedy, M., Phua, S.-H. & Semon, D. *Nucl. Acids Res.* **14**, 9713–9728 (1986).
- Kubo, A., Kusakawa, A. & Komano, T. *Mol. Gen. Genet.* **213**, 30–35 (1988).
- Murphy, E. in *Mobile DNA* (Berg, D. & Howe, M. M. eds.) Am. Soc. Microbiol., Wash. D.C. (1989) pp. 269–288.
- Poyart-Salmeron, C., Trieu-Cuot, P., Carlier, C. & Courvalin, P. *EMBO J.* **8**, 2425–2433 (1989).
- Hall, R. M. & Vockler, C. *Nucl. Acids Res.* **15**, 7491–7501 (1987).
- Mahillon, J. & Lereclus, D. *EMBO J.* **7**, 1515–1526 (1988).
- Utatsu, I., Sakamoto, S., Imura, T. & Toh-E, A. *J. Bact.* **169**, 5537–5545 (1987).
- Hartley, J. L. & Donelson, J. E. *Nature* **266**, 860–864 (1980).
- Maniatis, T., Fritsch, E. F. & Sambrook, J. *Molecular Cloning. A Laboratory Manual*. Cold Spring Harbor Laboratory Press, Cold Spring Harbor, NY, pp. 370 (1982).