Crystal Structures of Type II Restriction Endonuclease EcoO109I and Its Complex with Cognate DNA*

Received for publication, October 14, 2004, and in revised form, December 6, 2004
Published, JBC Papers in Press, December 7, 2004, DOI 10.1074/jbc.M411684200

Hiroshi Hashimoto‡, Toshiyuki Shimizu‡, Tsyuoshi Imasaki‡, Matsuri Kato‡, Naoki Shichijo‡, Keiko Kitai‡, and Mamoru Sato‡¶
From the ‡Graduate School of Integrated Science, Yokohama City University, 1-7-29 Suehiro-cho, Tsurumi-ku, Yokohama 230-0045 and §Graduate School of Agriculture, Kyoto University, Gokasho, Uji, Kyoto, 611-0011, Japan

EcoO109I is a type II restriction endonuclease that recognizes the DNA sequence of RGGNCCY. Here we describe the crystal structures of EcoO109I and its complex with DNA. A comparison of the two structures shows that the catalytic domain moves drastically to capture the DNA. One metal ion and two water molecules are observed near the active site of the DNA complex. The metal ion is a Lewis acid that stabilizes the pentavalent phosphorus atom in the transition state. One water molecule, activated by Lys-126, attacks the phosphorus atom in an SN2 mechanism, whereas the other water interacts with the 3′-leaving oxygen to donate a proton to the oxygen. EcoO109I is similar to EcoRI family enzymes in terms of its DNA cleavage pattern and folding topology of the common motif in the catalytic domain, but it differs in the manner of DNA recognition. Our findings propose a novel classification of the type II restriction endonucleases and lead to the suggestion that EcoO109I represents a new subclass of the EcoRI family.

Bacteria have evolved a mechanism to protect themselves from viral infection. Restriction endonucleases (REases)1 provide an anti-viral protection for bacteria by degrading the foreign DNA of invading bacteriophages. The enzymes recognize specific nucleotide sequences and cleave both strands of DNA. To date, more than 3,500 REases have been characterized and classified into four types, I, II, III, and IV (1). Of these types, type II REases are widely used in genetic technology and are the most well studied.

Type II REases generally recognize 4–8 base pairs of double-stranded DNA and hydrolyze phosphodiester bonds within the recognition sequences. The amino acid sequences are not homologous with the exception of the active site sequence. The active sites of the type II enzymes have a signature sequence P(D,E)X,DXX, in which four residues (Pro, Asp, Asp, and Lys; n = 1 (PvuII) to ~49 (Bse634I) (2)) are weakly conserved and two separated acidic residues are usually followed by a basic residue. The active site structures thus share a common structural motif consisting of a five-stranded β-sheet flanked with α-helices (3–5). Specific divalent metal cations such as Mg2+ are required to express the enzymatic activities and are coordinated to the conserved acidic residues during the catalytic reaction. However, details of the mechanism of the catalytic reaction are not fully understood (2).

Type II REases have been classified into two families, EcoRI and EcoRV, based on their DNA cleavage pattern. The EcoRI family produces 5′-overhang DNA, whereas the EcoRV family produces blunt-end DNA. From a structural viewpoint, the EcoRI family approaches DNA from the major groove side, whereas the EcoRV family approaches from the minor groove side. Moreover, the folding topology of the five-stranded β-sheet in the common structural motif in the active site differs between the two families (7) where the topology of four β-strands are absolutely conserved in the common motif but the fifth β-strand is oppositely oriented (8).

To date, 15 three-dimensional structures have been determined for type II REases by x-ray crystallography (REBASE; rebase.neb.com/cgi-bin/crylist). They mostly recognize the continuous 6-bp-long palindromic sequence but no three-dimensional structures for enzymes that recognize the discontinuous 7-bp-long palindromic sequence. In the type II REases that recognize degenerate base pairs, BsoBI (C↓YCGRG, where ↓ indicates cleavage position) (9), Bae634I (R↓CCGGY) (10), Cfr10I (R↓CCGGY) (11), and HincII (GTY↓RAC) (12) have been subjected to the x-ray crystallographic analyses and the three-dimensional structures of BsoBI and HincII have been determined in the form of complexes with their respective cognate DNAs. For the type II enzymes that recognize discontinuous sequences, only the three-dimensional structure of BglI (GCCNNNN↓NGGC) (13) has been determined as a complex with its cognate DNA. However, so far, no three-dimensional structure of a REase that recognizes both degenerate and discontinuous sequence has been determined.

EcoO109I is a type II REase isolated from Escherichia coli H709c and recognizes double-stranded DNAs with a seven-base pair motif of both degenerate and discontinuous sequence, RG↓GNCCY (where R = A or G and Y = T or C), and cleaves the phosphodiester bond between the second and the third bases to produce 5′-overhang DNA. Here we describe the crystal structures of both DNA-free EcoO109I and DNA-bound EcoO109I and show the structural basis for the catalytic mechanism at the atomic level. Furthermore, we suggest that EcoO109I represents a new subclass of the EcoRI family.

MATERIALS AND METHODS
Site-directed Mutagenesis, Expression, Purification, and Enzymatic Assay—EcoO109I was overexpressed and purified as described previ-
Crystal Structure of Restriction Endonuclease EcoO109I

Values in parentheses are for the highest resolution shell.

| Parameters | DNA-free EcoO109I | DNA-bound EcoO109I |
|------------|-------------------|--------------------|
| Wavelength (Å) | 1.0000 | 1.0082 |
| Space group | I4 | I4 |
| Cell dimensions | 5.50 | 5.60 |
| a (Å) | 175.5 | 176.6 |
| b (Å) | 175.5 | 176.6 |
| c (Å) | 144.6 | 146.6 |
| Resolution (Å) | 2.4 | 2.6 |
| Total observations | 186,285 | 178,049 |
| Unique reflections | 21,557 | 21,548 |
| Completeness (%) | 99.1 (99.3) | 99.1 (99.3) |
| Rmerge (%)* | 15.3 (7.7) | 15.3 (7.7) |
| Wilson B value (Å²) | 53.0 | 56.9 |

*Rmerge = \sum h|I(h) - \langle I(h) \rangle|/\sum hI(h).*

| TABLE II | Refinement statistics |
|-----------|-----------------------|
| Parameters | DNA-free EcoO109I | DNA-bound EcoO109I |
| Resolution (Å) | 2.4 | 1.9 |
| Reflections used | 25,421 | 53,090 |
| R (%)* | 25.7 | 17.1 |
| Rfree (%)* | 29.7 | 22.8 |
| No. of residues | 551 | 544 |
| No. of atoms |
| Protein | 4131 | 4432 |
| DNA | 0 | 1049 |
| Na⁺ | 45 | 654 |
| Water | 10 | 0 |
| Glyceraldehyde | 0.014 | 0.025 |
| Bond angle (°) | 1.424 | 2.189 |
| Mean B value (Å²) | 55.5 | 24.3 |

*R and Rfree = \sum |Fo - Fc|/\sum |Fc|, where the free reflections (5% total used) were held aside for Rfree throughout refinement.

RESULTS AND DISCUSSION

Overall Structure—EcoO109I forms a homodimer (molecules A and B) in the crystallographic asymmetric unit of both its DNA-free form and DNA-bound forms (Fig. 1, a and b). Ab initio low resolution structure analysis of x-ray solution scattering data has shown that DNA-free EcoO109I in solution has a shape and a size similar to those of the dimer formed in the crystal (Fig. 2), indicating that the dimer formed in the crystal is the functional unit of the enzyme in solution. The monomer of EcoO109I consists of two domains, the dimerization and the catalytic domains (Fig. 1, c and d). As shown in Fig. 3a, the asymmetric unit of EcoO109I together with an assignment of the secondary structures in the DNA-free and DNA-bound enzymes, the dimerization domain consists of only a beta-helices (residues 1–86 and 210–272), which interact with those of the other molecule to form the functional dimer. Buried accessible surface areas of the dimer, calculated with the program MSMS (26), are 5,608 and 7,139 Å² for DNA-free and DNA-bound enzymes, respectively, showing that the dimerization domain of the DNA-bound enzyme interacts more tightly than that of the DNA-free enzyme.

The catalytic domain of the DNA-free enzyme adopts a flexible conformation with relatively high average temperature factors (74 Å²) as compared with the dimerization domain (46 Å²). A
comparison of the structures between the DNA-free and DNA-bound enzymes shows that a large conformational change occurs in the catalytic domain. Indeed, α6 and a loop region between α7 and β4 move into the major groove by 10 Å to wrap around the DNA (Fig. 1, c and d). The regions from residues 93 to 100 in molecule A and from residues 178 to 182 in molecule B are disordered in the DNA-free enzyme, whereas they are well ordered in the DNA-bound enzyme. The average temperature factor of the catalytic domain (22 Å²) is now lower than that of the dimerization domain (26 Å²) in the DNA-bound enzyme. Furthermore, binding of DNA induces new helical structures of α5 and two 3₁₀ helices (Fig. 3a).

**DNA Duplex Is Distorted by a π-π-Interaction with EcoO109I**—The DNA sequence used in this study is not palindromic (Fig. 3b); thus, DNA is bound to EcoO109I in two opposite directions in the crystallographic asymmetric unit of the DNA complex. These two indistinguishable DNA structures in which the recognition pattern of EcoO109I is identical are well superimposed on each other. Therefore, we based our discussion on only one of these structures.

The DNA adopts a structure typical of the B form DNA where some base pairs are distorted upon complex formation with EcoO109I. Of these, significant distortions of DNA duplex occur in the base pairs of Gua₅(X):Cyt₉(Y) and Cyt₉(X):Gua₅(Y) because of a π-π-interaction between the indole ring of Trp-130 and the pyrimidine ring of Cyt₉ (Figs. 1d and 4a). In fact, although the respective buckle and propeller angles of normal B-DNA are 0.0 and 3.8°, those of the base pair of Gua₅(X):Cyt₉(Y) are 21.6 and 25.1°, respectively, and those of the base pair of Cyt₉(X):Gua₅(Y) are -21.1 and 26.9°, respectively. Furthermore, intercalation by Trp-130 gives rise to roll angles -33.9 and -35.0° of Gua₄(X):Cyt₁₀(Y)-Gua₅(X):Cyt₉(Y) and Cyt₉(X):Gua₄(Y)-Thy₁₀(X):Ade₄(Y), respectively. The width of the major groove (15.0 Å) is also wider than that...
Because replacement of Trp-130 with alanine decreases endonuclease activity, the structural changes induced by the interaction are responsible for the sequence specific hydrolysis activity of this enzyme (Fig. 5).

Fig. 3. **Amino acid and DNA sequences.** *a*, amino acid sequence of EcoO109I together with the assignment of the secondary structures of the A molecules for DNA-free and DNA-bound EcoO109I. The dimerization and catalytic domains are colored in pale blue and pink, respectively. Disordered residues are indicated by dashed lines (residues 83<sub>A</sub>-100<sub>A</sub> and 178<sub>B</sub>-182<sub>B</sub>). Signature sequences of type II restriction endonucleases are highlighted by a black background. *b*, sequence of the DNA duplex in complex with EcoO109I. The recognition sequence is highlighted by a blue background. Cleavage positions are indicated by the red arrows.

Fig. 4. **Base pair recognition made by EcoO109I.** *a*, recognition of the outer GC pairs. *b*, recognition of the inner GC pairs. The black arrow indicates the C5 position of Cyt8, which is the site of methylation by EcoO109I methyltransferase. *c* and *d*, recognition of degenerate R:Y pairs. In all of the panels, the upper side is the major groove side.

of normal B-DNA (11.6 Å). Because replacement of Trp-130 with alanine decreases endonuclease activity, the structural changes induced by the π-π-interaction are responsible for the sequence specific hydrolysis activity of this enzyme (Fig. 5).
the recognition of Gua6(X):Cyt8(Y) and that of Cyt8(X):Gua6(Y) are also identical. The outer GC pairs of Gua5(X):Cyt9(Y) and Cyt9(X):Gua5(Y) are recognized from only the major groove side of the DNA duplex (Fig. 4a) where the main-chain carbonyl oxygen of Trp-130 forms a hydrogen bond with N4 of Cyt9 and the main-chain amido nitrogen of Leu-134 forms a hydrogen bond with N7 of Gua5. Furthermore, N7 of Lys-173 is hydrogen-bonded to O6 of Gua5 through a water molecule. In the outer GC base pairs, a base-specific interaction occurs between the main-chain carbonyl oxygen of Trp-130 and N4 of Cyt9 by which EcoO109I precisely recognizes Cyt9 from the major groove side.

By contrast, the inner GC base pairs of Gua6(X):Cyt8(Y) and Cyt8(X):Gua6(Y) are recognized from both the major and the minor groove sides (Fig. 4b) where O1 of Thr-66 interacts with O6 of Cyt8 through two water molecules from the minor groove side, N2 of Gln-133 interacts with O6 of Gua6 from the major groove side, and O1 of Thr-70 interacts with N2 of Gua6 from the minor groove side. In the inner base pairs, base-specific interactions occur between N2 of Gln133 and O6 of Gua6 as well as between O1 of Thr-70 and N2 of Gua6 by which EcoO109I precisely recognizes Gua6 from both the major and the minor groove sides. The C5 position of cytosine in the inner GC pairs has been shown to be methylated by EcoO109I methyltransferase (27). The C5 atom of Cyt8 corresponds to this position, and the distance between the C5 atom and the Ca atom of Thr-131 corresponds to the normal van der Waals distance (3.6 Å). Therefore, methylation of the C5 position gives rise to steric hindrance and is likely to result in a considerable loss of the protein-DNA interaction to the extent that the DNA is protected from cleavage.

Both of the degenerate base pairs of Gua4(X):Cyt10(Y) and Thy10(X):Ade4(Y) are recognized from only the major groove side (Fig. 4, c and d). N7 of Lys-173 (in molecule A) interacts with N7 of Ade4(Y) directly and with N6 of Ade4(Y) through a water molecule (Fig. 4c), and N7 of Lys-173 (in molecule B) interacts with N7 of Gua4(X) directly and with O6 of Gua4(X) through a water molecule (Fig. 4d). Of the type II REases (BsoBI (C \ YCGRG), Bse634I (R \ CCGGY), Cfr10I (R \ CCGG), and HincII (GTY \ RAC)) that recognize degenerate sequences, the structures of BsoBI and HincII have been determined as a complex with their cognate DNAs. Direct and water-mediated indirect hydrogen bonding occurs between protein and DNA in the nonspecific base pair of Cyt7(X):Gua7(Y).

In addition to the degenerated base pair R:Y, DNA with the sequence RG \ GNNCCY includes a non-recognized base, N. As expected, no direct and/or water-mediated indirect hydrogen bonding occurs between protein and DNA in the nonspecific base pair of Cyt7(X):Gua7(Y).

**Active Site Structure**—Among the type II REases, the sequence PDDE \ DEEXK is conserved in the active site, although the overall amino acid sequences of the enzymes are not homologous. In regard to EcoO109I, the conserved sequence corresponds to 109Ile-Asp(X) \ Ser-Leu-Lys126 (Fig. 3a). The electron density map around the active site of the DNA-EcoO109I complex clearly shows that the phosphodiester bond is not cleaved and that a metal ion is located near the active site (Fig. 5a). A REase requires a divalent metal cation such as a Mg2+ or Mn2+ ion to cleave DNA. Given that the crystallization solution contains Na+ and K+ ions and that the metal ion is coordinated square-bipyramidally with six oxygen atoms (the side-chain carboxyl oxygen of Asp-110, a phosphate oxygen of DNA, the main-chain carboxyl oxygen of Leu-125, and three water oxygens), with bond distances ranging from 2.3 to 2.5 Å,
A Proposed Catalytic Mechanism—The hydrolysis of a phosphodiester bond catalyzed by EcoO109I has been shown to require a general base to activate the NW water molecule, a Lewis acid such as a divalent metal cation to stabilize the negatively charged pentavalent phosphorus in the transition state, and a general acid to donate a proton to the 3'-hydroxyl-leaving group (5). On the basis of these three requirements, the hydrolysis reaction starts with the binding of a water molecule to the metal ion near the active site, resulting in a suitable positioning of the water molecule for nucleophilic attack on the phosphorus atom (Fig. 6). Asp-110 is a key residue for localizing a metal ion near the active site, and Lys-126 changes the water molecule to a hydroxyl ion, which is more nucleophilic than a water molecule, through a general base activation mechanism found to have lost endonuclease activity (Fig. 5). The general base activation is also aided by polarization of the water molecule through its coordination to the divalent metal ion. The putative nucleophilic water molecule was also observed in the structure of HincII and was within hydrogen-bonding distance of the conserved active site lysine (Lys-129) as well as the oxygen of the 3'-phosphate group of scissile phosphate (6). The importance of 3'-phosphate during the activation of water molecule was discussed (6). NW in the EcoO109I also interacts with the oxygen of 3'-phosphate group of scissile phosphate. Therefore, the interaction might be involved in the activation of the water molecule.

Nucleophilic attack by the hydroxyl ion on the phosphorus atom releases a 3'-hydroxyl group through the pentavalent phosphorus found in the transition state, showing that the water molecule is the nucleophile for hydrolysis of the phosphodiester bond and that the hydrolysis of DNA proceeds by an $S_2$2 mechanism.

REFERENCES

1. Roberts, R. J., Belfort, M., Bestor, T., Bhagwat, A. S., Bickle, T. A., Bitinaite, J., Blumenthal, R. M., Defgyarav, S., Kr, D., Dybdvig, K., et al. (2000) Nucleic Acids Res. 28, 1805–1812
2. Galbur, E. A., and Stoddard, B. L. (2002) Biochemistry 41, 13851–13860
3. Yudina, C., Timinskas, A., and Siksny, V. (1994) Proteins 20, 279–282
4. Koval, R. A., and Matthews, B. W. (1998) Proc. Natl. Acad. Sci. U. S. A. 95, 7893–7897
5. Koval, R. A., and Matthews, B. W. (1999) Curr. Opin. Chem. Biol. 3, 578–583
6. Elskorn, C., and Horton, N. C. (2004) Biochemistry 43, 12356–12370
7. Cheng, X., Balandin, K., Schildkraut, I., and Johnson, J. B. (1994) EMBO J. 13, 3927–3935
8. Pintou, A., and Kiehls, A. (2001) Nucleic Acids Res. 29, 3705–3727
9. Van der Weerd, M. J., Pelleter, J. J., Xu, S., and Friedman, M. (2001) Structure 9, 133–144
10. Grazulis, S., Deibert, M., Rimslieeni, R., Skirgaila, R., Sasaunakus, G., Lagunnavistus, A., Reip, V., Urbanke, C., Huber, R., and Siksny, V. (2002) Nucleic Acids Res. 30, 876–885
11. Bozic, D., Grazulis, S., Siksny, V., and Huber, R. (1996) J. Mol. Biol. 255, 176–186
12. Horton, N. C., Dorner, L. F., and Perona, J. J. (2002) Nat. Struct. Biol. 9, 42–47
13. Newman, M., Lunnan, K., Wilson, G., Greci, J., and Phillips, S. E. V. (1998) EMBO J. 17, 5466–5476
14. Imasaki, T., Hashimoto, H., Shimizu, T., Kato, M., Tsuda, J., Kita, K., and Sato, M. (2004) Acta Crystallogr. Sect. D Biol. Crystallogr. 60, 1165–1166
15. Otwonowski, Z., and Minor, W. (1997) Methods Enzymol. 267, 307–326
16. Flughart, J. W. (1999) Acta Crystallogr. Sect. D Biol. Crystallogr. 55, 1718–1725
17. Tverwilliger, T. C., and Berendzen, J. (1999) Acta Crystallogr. Sect. D Biol. Crystallogr. 55, 849–861
18. Jones, T. A., Zou, J.-Y., Cowan, S. W., and Kjeldgaard, M. (1991) Acta Crystallogr. A 47, 110–119
19. Bruggen, A. T., Adama, P. D., Clere, G. M., DeLano, W. L., Glos, P., Grossenheider, H., and Jansen, J.-S. (1998) Acta Crystallogr. Sect. D Biol. Crystallogr. 54, 905–904
20. Murshudov, G. N., Vagin, A. A., and Dodson, E. J. (1997) Acta Crystallogr. Sect. D Biol. Crystallogr. 53, 240–255
21. Tenno, T., Fujisawa, K., Tsuchi, H., Iwai, K., Morita, E. H., Hayashi, H., Murata, S., Hiraoki, H., Sato, M., Tanaka, K., and Shirakawa, M. (2004) Genes Cells 9, 685–697
22. Svergun, D. I. (1999) Biophys. J. 76, 2879–2886
23. Kraulis, P. J. (1991) EMBO J. 10, 6706–6710
24. Merrit, E. A., and Murphy, M. E. P. (1994) Acta Crystallogr. Sect. D Biol. Crystallogr. 50, 869–875
25. Laverly, R., and Sklenar, H. (1988) J. Biol. Struct. Dyn. 6, 63–91
26. Sanner, M. F., Olson, A. J., and Spehner, J. C. (1996) Biopolymers 38, 305–329
27. Kita, K., Tsuda, J., and Nishigaki, R. (2001) Biosci. Biotechnol. Biochem. 65, 2512–2518
28. Harding, M. (2002) Acta Crystallogr. Sect. D Biol. Crystallogr. 58, 872–874
Crystal Structures of Type II Restriction Endonuclease EcoO109I and Its Complex with Cognate DNA
Hiroshi Hashimoto, Toshiyuki Shimizu, Tsuyoshi Imasaki, Matsuri Kato, Naoki Shichijo, Keiko Kita and Mamoru Sato

*J. Biol. Chem. 2005, 280:5605-5610.*
doi: 10.1074/jbc.M411684200 originally published online December 7, 2004

Access the most updated version of this article at doi: 10.1074/jbc.M411684200

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 28 references, 2 of which can be accessed free at
http://www.jbc.org/content/280/7/5605.full.html#ref-list-1