Structure of the metal-independent restriction enzyme BfiI reveals fusion of a specific DNA-binding domain with a nonspecific nuclease

Saulius Grazulis*, Elena Manakova*, Manfred Roessle†, Matthias Bochtler†, Giedre Tamulaitiene*, Robert Huber†, and Virginijus Siksnys‡

*Laboratory of Protein–DNA Interaction, Institute of Biotechnology, Graiciuno 8, LT-02241 Vilnius, Lithuania; †Department of Structural Biology, International Institute of Molecular and Cell Biology, Skłodowska J. Trojdena Street 4, 02-109 Warsaw, Poland; ‡European Synchrotron Radiation Facility, B.P. 220, F-38043 Grenoble Cedex, France; §Department of Structural Biology, International Institute of Molecular and Cell Biology, Ksie˙cja J. Trojdena Street 4, 02-109 Warsaw, Poland; and ‡Abteilung Strukturforschung, Max-Planck-Institut für Biochemie, Am Klopferspitz 18a, D-82152 Martinsried, Germany

Contributed by Robert Huber, September 20, 2005

Among all restriction endonucleases known to date, BfiI is unique in cleaving DNA in the absence of metal ions. BfiI represents a different evolutionary lineage of restriction enzymes, as shown by its crystal structure at 1.9 Å resolution. The protein consists of two structural domains. The N-terminal catalytic domain is similar to Nuc, an EDTA-resistant nuclease from the phospholipase D superfamily. The C-terminal DNA-binding domain of BfiI exhibits a β-barrel-like structure very similar to the effector DNA-binding domain of the Mg2+-dependent restriction enzyme EcoRII and to the B3-like DNA-binding domain of plant transcription factors. BfiI presumably evolved through domain fusion of a DNA-recognition element to a nonspecific nuclease akin to Nuc and elaborated a mechanism to limit DNA cleavage to a single double-strand break near the specific recognition sequence. The crystal structure suggests that the interdomain linker may act as an autoinhibitor controlling BfiI catalytic activity in the absence of a specific DNA sequence. A PSI-BLAST search identified a BfiI homologue in a Mesorhizobium sp. BNC1 bacteria strain, a plant symbiont isolated from an EDTA-rich environment.

Restriction endonucleases (REases) protect bacteria by hydrolyzing invading viral or other foreign DNA. Type II REases perform their function by catalyzing the sequence-specific cleavage of double-stranded DNA molecules in the presence of Mg2+ ions within or close to their recognition sites (1). Surprisingly, orthodox REases and bacteriophage λ-exonuclease that binds a free end of double-stranded DNA and processively degrades one strand in the 5′ to 3′ direction share a similar catalytic mechanism and a common structural ancestor (2). A λ-exonuclease-like domain has been identified in many Mg2+-dependent nucleases involved in DNA recombination and repair (3–5), suggesting that it has been remodeled during evolution to perform different functions. To constrain cleavage at specific sites, REases had to develop effective means to control nucleolytic activity of λ-exonuclease-like catalytic domain.

A variety of mechanisms have evolved to maintain REases in inactive configuration to avoid uncontrolled DNA cleavage and couple it to the recognition of specific nucleotide sequence. Orthodox REases, like EcoRI or BamHI, are homodimers that make largely symmetrical interactions with palindromic DNA sequences and contain two distinct sites each responsible for catalyzing cleavage in one DNA strand (1). In orthodox restriction enzymes, structural elements involved in sequence recognition are grafted on the conserved λ-exonuclease-like scaffold that makes a catalytic core (6). Structural analysis indicates that EcoRI amino acid residues involved in specific DNA binding are coupled to the catalytic residues through the intricate network of interactions (7). Therefore, any perturbation in the recognition site (for example, an incorrect base pair) would propagate via the network to the cleavage site and thus prevent DNA cutting at the incorrect site.

Different mechanisms of the activity control are used by type IIS REases. Most type IIS REases are monomers and recognize asymmetric nucleotide sequences. The archetypal type IIS enzyme FokI exhibits a modular architecture and consists of two functional domains: the N-terminal DNA recognition domain and the C-terminal catalytic domain (8). FokI binds DNA as a monomer, with the N-terminal recognition domain making all of the base-specific contacts at the recognition site (8). The catalytic C-terminal domain of FokI exhibits a λ-exonuclease-like fold similar to the orthodox REases (8) and possesses a weak nonspecific nuclease activity (9). Control of nuclease activity in FokI is achieved by two different mechanisms. First, the catalytic domain is sequestered by the recognition domain in a “piggy-back” fashion preventing contact with DNA at the cleavage site (8). Second, FokI has only one catalytic site per monomer; therefore, dimerization of two C-terminal domains is required to achieve a double-strand break in DNA (10–12). Association of the catalytic domains of FokI is much more effective when FokI monomers are bound to the separate DNA sites on the same DNA molecule (10). Thus, rearrangement of FokI triggered by specific DNA binding promotes DNA cleavage and limits it to the specific sites.

Yet, another type of mechanism of DNA cleavage control is used by the EcoRII enzyme that belongs to the type IIE subtype of REases. EcoRII requires two copies of the CCWGG recognition sequence for the optimal activity but cleaves only one copy before the first C; the second copy acts as an allosteric activator for the cleavage of the first DNA copy. Structural and biochemical studies demonstrate a modular architecture of EcoRII (13, 14). Proteinase treatment of EcoRII generates the N-terminal allosteric DNA-binding domain and C-terminal domain dimer that cleaves DNA in a site-specific manner (14). Apo-structure of EcoRII suggests (13) that N-terminal allosteric DNA-binding domain sterically blocks DNA access to the catalytic C-terminal domain. Binding of the first DNA copy at the allosteric site triggers enzyme rearrangement and promotes binding and cleavage of the second DNA copy at the C-terminal domain.

BfiI is a type IIS REase that acts at an asymmetric sequence, 5′-ACTGGG-3′, and cuts top and bottom strands at fixed
positions downstream of this site (15). In this respect, it is similar to FokI but unique among all of the restriction enzymes found to date by its metal-independence (16). BfiI must therefore hydrolyze phosphodiester bonds by a novel mechanism.

The N-terminal domain of BfiI shows sequence similarity to Nuc, an EDTA-resistant nuclease from the phospholipase D (PLD) superfamily (16, 17). BfiI presumably evolved through the domain fusion of a DNA-recognition element to a nonspecific nuclease akin to Nuc and acquired specificity and function of a restriction enzyme. To gain insight into the architecture of the unique Mg\(^{2+}\)-independent restriction enzyme, we have determined the crystal structure of BfiI at 1.9-Å resolution.

**Materials and Methods**

**Protein Crystallization.** Wild-type BfiI protein was purified as described in ref. 18. The purified protein was stored at −20°C in a storage buffer containing 50% glycerol, 200 mM KCl, 10 mM Tris-HCl (pH 7.5), 1 mM DTT, and 1 mM EDTA at concentration of ∼3 mg/ml. Immediately before crystallization an aliquot of protein was rebuffered by adding an equal volume of saline buffer (400 mM NaCl/20 mM Tris-HCl, pH 8.5/1 mM EDTA) and concentrated in Centricrons (10-kDa cutoff; Millipore) to the final concentrations of ∼3–4 mg/ml. Crystals were prepared by mixing in a 1:1 ratio of the concentrated BfiI solution with the precipitating solution containing 1.2–1.6 M Na tartrate and 75 mM Na-MES (pH 6.5). Drops were equilibrated against the reservoir solution that was prepared by mixing in a 1:1 ratio of the concentrated BfiI solution with Cns again, allowing just cofactors to change. This yielded a crystal that was rigid-body-refined in the new data. The resulting maps, although not traceable with ARP/WARP, showed new details that gave a new map where some parts of the C-terminal domain were visible. By using iterative procedure of manual building, most of the C-terminal domain main chain was built. After several iterations of the building cycle at 3.0-Å resolution, no new density appeared, and the model was repositioned for the 1.9-Å synchrotron data set. The model itself, however, did not produce sufficiently good phases to continue with the model building. Therefore, a procedure suggested in the ARP/WARP (26) manual was undertaken. Phases were taken from the MIR data at 3.0 Å (the data had a slightly different cell), merged with the 1.9-Å synchrotron data, and solvent-flattened with DM. The resulting phases were combined with the phases from the partial model that was rigid-body-refined in the new data. The resulting maps, although not traceable with ARP/WARP, showed new details that enabled to complete the model manually to good levels of correspondence with the experimental data (Table 1). Initially, the structure was refined by using the CNS suite, but at the final stages REFMAC (27) gave better R factors and better fit to the maps. However, there were still problems with the refinement of some of our cofactors, because too many bond angles and lengths deviated severely from ideal values if REFMAC was used with default settings. Therefore, at the last stage, the REFMAC bond restraints were made stricter, and then the model was refined with CNS again, allowing just cofactors to change. This yielded a model almost without geometry violations and good R factors. In the refined structure, all protein residues except one fall into the most favored and additionally allowed Ramachandran plot regions (28). Figures of the molecules were produced by using the programs MOLSCRIPT (29), RASTER3D (30), and MOLMOL (31).

**Results**

**Overall Structure: Domain Architecture of BfiI.** Electron densities calculated at 1.9-Å resolution (see Fig. 5, which is published as supporting information on the PNAS web site) show all of the protein chain except for a loop between residues 246 and 249 in the chain B (see Fig. 6, which is published as supporting information on the PNAS web site, for the topological diagram of BfiI). The crystal structure of the BfiI reveals two chains in the asymmetric unit (Fig. 1). Each chain is folded into two distinct globular domains (Fig. 1). Both domains are connected by a
The N-Terminal Domain of BfiI: Similarities to Nuc Nuclease. The N-terminal domain of BfiI extends from Met-1 to Met-170. It is comprised of two β-sheets that are interconnected by β13 and β14 strands, shared by both sheets (Fig. 6). The two β-predictors are sandwiched together into a barrel-like structure, composed of eight β-strands (β10, β11, β13–β17, and β19). The “barrel,” however, is not closed because strands β11 and β19 are not hydrogen-bonded. Two additional short β-strands (β12 and β18) and three α-helices are located outside the pseudo-barrel.

A DALI database search revealed that the C-terminal domain of BfiI is significantly similar (Z = 8.8) to the N-terminal domain of Mg²⁺-dependent REase EcoRII that belongs to the type IIE subtype. Another significant similarity (Z = 7.1) is to the B3 DNA-binding domain of Arabidopsis cold-responsive transcription factor RAV1 (38) (PDB entry 1WID), specific for the CACCTG sequence (39). A structural superposition of the C-terminal domain of BfiI, the N-terminal domain of EcoRII REase, and the RAV1 transcription factor of Arabidopsis (Fig. 3a) clearly shows that all three domains represent the same fold. The structures of BfiI, EcoRII (13), and the B3-domain of RAV1 transcription factor of Arabidopsis (38) were solved in a DNA-free form. Structural similarities between the C-terminal domain of BfiI, the N-terminal domain of EcoRII REase, and the B3-domain of RAV1 (Fig. 3a), however, provide some clues regarding the putative DNA-binding surface. Topologically similar concave clefts laid by the positively charged residues and large enough to accommodate a DNA molecule could be identified in all domains (see Fig. 7, which is published as supporting information on the PNAS web site). Cross-linking and mutational studies of EcoRII demonstrated the importance of His-36, Tyr-41, Lys-92, Arg-94, Glu-96, Lys-97, and Arg-98
sequences (ACTGGG in the case of BfiI and CCWGG for EcoRII) is intriguing in view of the different recognition sequences so far has not yet been tested. The functional importance of these residues for DNA binding (41). Their side chain rotamers differ but may change in the presence of DNA. In addition, Arg-272 and Ser-254 residues of BfiI coincide with Arg-81 and Ser-63 residues of EcoRII. The functional importance of these residues so far has not yet been tested. The conservation of some residues in the DNA-binding clefts of BfiI and EcoRII is intriguing in view of the different recognition sequences (ACTGGG in the case of BfiI and CCWGG for EcoRII).

Interdomain Interface. N- and C-terminal domains in a BfiI monomer make an extensive set of hydrophobic contacts between the “inner” faces of the β-sheet. Approximately 1,700 Å² are buried at the interdomain interface in BfiI monomer. Contacts of the C-terminal domain to the N-terminal domain of the neighboring subunit (i.e., C terminus of chain A with the N terminus of the chain B) contribute an additional 500 Å² to the interdomain interface. Thus, the two domains have a relatively large interdomain interface of ~2,200 Å², characteristic of specific protein–protein interactions (42). In comparison, the main dimer interface of the BfiI formed by the N-terminal domains buries 3,500 Å² (Fig. 1).

The C-terminal DNA-binding domain of BfiI and the N-terminal effector DNA-binding domain of EcoRII are similar to the B3-like DNA-binding domain of plant-specific transcription factors (Fig. 3a). Hence, the transcription factor-like DNA-binding domain has been used not only to confer DNA specificity to a PLD family nuclease, to give the metal-free BfiI enzyme as shown here, but also to create allosteric site(s) in an orthodox PD . . . EXK family enzyme, to generate the metal-dependent type IIIE restriction enzyme EcoRII. Although the C-terminal DNA-binding domain of BfiI is structurally similar to the N-terminal effector DNA-binding domain of EcoRII, they employ different interfaces to interact with the corresponding catalytic domains (see Fig. 8, which is published as supporting information on the PNAS web site). In BfiI, the DNA-binding domain makes contacts predominantly with the catalytic N-terminal domain of the same chain, making hydrogen bonds and hydrophobic contacts through residues coming from the edge of the first barrel-forming β-sheet (strands β11 and β12). In contrast, in EcoRII the effector DNA-binding N-terminal domain contacts the catalytic domain through the residues located on α-helices and their flanking loop regions (13). Thus, DNA-binding domains of BfiI and EcoRII possess a similar DNA-binding cleft but are oriented differently with respect to their catalytic cores. Moreover, whereas the DNA-binding domains of BfiI are spatially separated, the N-terminal domains of EcoRII contact each other and contribute to the dimer interface.

Interdomain Peptide Linker: Possible Role in Activity Control. The linker connecting the N- and C-terminal domains of the BfiI contains a stretch of 28 amino acid residues (171–198) that predominantly has an extended conformation (Fig. 4). The...
Discussion

The crystal structure of BfiI (Fig. 1) reveals a novel domain architecture consistent with previously reported limited proteolysis (44) data, where the N terminus is shown to be an unspecific nuclease and the C terminus is a DNA-binding domain. The crystal structure of BfiI demonstrates how nature can reconfigure protein domains for new functions. In BfiI, a DNA-recognition element is fused to a nonspecific nuclease core family to generate a REase. In the asymmetric unit, two N-terminal domains of BfiI are arranged into a dimer structurally similar to the nonspecific EDTA-resistant Nuc nuclease (Fig. 2) that belongs to the PLD superfamily. Structural similarities between BfiI, Nuc nuclease (17), tyrosyl-DNA phosphodiesterase Tdp1 (33), and PLD (34) provide a direct link between BfiI and the PLD superfamily, a diverse group of proteins that includes phospholipases, phospholipid synthases, bacterial toxins, pox virus envelope proteins, phosphodiesterases, and bacterial nucleases (45). PLD-family enzymes exhibit a conserved catalytic core and share a common catalytic mechanism; however, these enzymes encompass a very broad range of substrate specificities (45). Different enzymes in the PLD family may therefore have evolved through the fusion of a common catalytic core to separate domains for substrate recognition.

The C-terminal DNA-binding domain of BfiI (Fig. 3a) is structurally similar to the N-terminal DNA-binding domain of Mg$^{2+}$-dependent REases EcoRII (13). This finding suggests that different subtypes and lineages of REases might have evolved through the recombination of conserved structural blocks that became fused to different protein core domains.

Strikingly, DNA-binding domains of EcoRII/BfiI and B3 DNA-binding domain of Arabidopsis cold-responsive transcription factor RAV1 (38) share a similar structural fold (Fig. 3a). The B3-like DNA-binding domain was thought to be plant-specific and belongs to the family that currently includes ~300 proteins (InterPro IPR 003340) involved in the auxin-regulated and abscisic acid-regulated transcription (46). Identification of structural and functional homologues of the “plant-specific” B3 DNA-binding domain in the bacteria kingdom provokes speculations on the possible evolutionary relationships.

A PSI-BLAST search of the protein sequence database by using BfiI sequence as a query reveals significant similarities to the hypothetical protein MNBC0300766 of Mesorhizobium sp. BNC1 strain (see Fig. 9, which is published as supporting information on the PNAS web site). The degree of sequence similarity (60% of identical and 76% of similar amino acids) indicates that the unknown protein MNBC0300766 of Mesorhizobium sp. BNC1 strain is likely to be a restriction enzyme similar to BfiI. Moreover, typically for type IIS restriction modification system gene organization two ORFs possessing similarities to the DNA-methyltransferases are found in the vicinity of the gene of putative Mesorhizobium sp. BNC1 endonuclease (data not shown). Interestingly, the Mesorhizobium sp. BNC1 strain subjected for the genome sequencing has been isolated from an enrichment of industrial sewage receiving EDTA-containing wastewater effluents added to surface soil. This strain is able to use EDTA as a sole source of carbon and nitrogen, and thrives in an EDTA-rich environment (47). In such a metal-ion hostile environment, the presence of EDTA-resistant REase in bacteria might be an advantage.

Both the N- and C-terminal domains of BfiI (Fig. 1) represent independent structural and functional modules. The isolated DNA-binding domain binds specifically to the 5'-ACTGGG sequence characteristic for BfiI, whereas the N-terminal domain possesses a nonspecific nuclease activity (44). Mixing of the isolated N- and C-terminal domains, however, does not constitute the ability of BfiI to cleave phosphodiester bonds at specific sites (M. Zaremba and V.S., unpublished data). Thus, to function as a restriction enzyme, BfiI had to elaborate a nuclease activity control mechanism to limit double-strand breaks at the specific recognition sequence. The crystal structure of BfiI suggests how interdomain communications can lead to novel protein functions, not possessed by either domain alone.

BfiI, like many orthodox Mg$^{2+}$-dependent REases, is a dimer in solution (18). In contrast to the dimeric enzymes that recognize a single copy of their DNA, the dimeric form of BfiI binds two copies of its recognition sequence (18). Yet, when bound to two copies of its recognition site, the BfiI dimer cleaves only one phosphodiester bond at a time (48). The crystal structure reveals that BfiI, like Nuc from S. typhimurium, has a single active site located at the interface between the N-terminal domains (Fig. 2a and b). Similarly to Nuc, an isolated N-terminal domain of BfiI degrades DNA containing or lacking BfiI recognition sequence by introducing random cuts (44). Structural superposition between BfiI and Nuc reveals that in the wild-type BfiI the linker region interconnecting N- and C-terminal domains crosses the putative DNA-binding surface of the catalytic domain (Fig. 1) and sterically interferes with DNA binding. Distribution of positively charged residues in the interdomain linker suggests that it might mimic the DNA backbone (Fig. 4) and autoinhibit the DNA cleavage. The BfiI conformation seen in the crystal may therefore represent a latent form. The active site of BfiI, however, remains available for small artificial substrates like bis-p-nitrophenyl phosphate, which is hydrolyzed by BfiI (16). Thus, the negatively charged linker in BfiI does not disrupt the active site but rather creates a steric block for DNA binding.

To cleave DNA, the binding surface at the N-terminal domain must become available for DNA binding. Proteolytic cleavage of BfiI opens an access for DNA binding at the active site located at the catalytic domain, but phosphodiester bond hydrolysis then occurs at random rather than at specific sites (44). Thus, to focus cleavage at particular DNA sites, the opening of the catalytic domain in the wild-type BfiI has to be coupled to the specific DNA binding (see Fig. 10, which is published as supporting information on the PNAS web site). We propose that specific DNA binding to the C-terminal domain of BfiI may trigger a conformational rearrangement that leads to repositioning of the linker and unveils a surface for DNA binding at the catalytic N-terminal domain (Fig. 10). Such a cognate DNA-binding-induced conformational switch juxtaposes the catalytic domain to a proper configuration to enable BfiI to cleave a phosphodiester bond (48). The single active site in the dimeric catalytic domain then acts sequentially on the four target phosphodiester bonds. It has yet to be determined how the four target bonds, two from each recognition site, are placed in turn into the single active site of the enzyme.

We thank V. Stonyte for help with purification of the BfiI protein and S. Halford of the University of Bristol (Bristol, UK) and M. Zaremba, G. Sasnauskas, and C. Venclovas of the Institute of Biotechnology (Vilnius) for discussions and critical reading of the manuscript. Data sets were collected on the ID13 beamline of the European Synchrotron Radiation Facility (Grenoble), the European Molecular Biology Laboratory X13 beamline of the Deutsches Elektronen Synchrotron (Hamburg), and the International Institute of Molecular and Cell Biology (Warsaw). Measurements at the European Molecular Biology Laborato-
tory outstation were supported by the European Community Access to Research Infrastructure Action of the Improving Human Potential Program. V.S. is a recipient of Lithuanian State Fellowship. The work at the Institute of Biotechnology (Vilnius) was supported in part by the Howard Hughes Medical Institute, the Wellcome Trust, the Max Planck Society, and the Lithuania State Science and Studies Foundation.