Structure *14*, 1389–1400, September 2006 ©2006 Elsevier Ltd All rights reserved DOI 10.1016/j.str.2006.07.002

# The Crystal Structure of the Rare-Cutting Restriction Enzyme SdaI Reveals Unexpected Domain Architecture

**Giedre Tamulaitiene,[1] Arturas Jakubauskas,[1] Claus Urbanke,[2] Robert Huber,[3] Saulius Grazulis,[1] and Virginijus Siksnys[1,*]**

[1] Institute of Biotechnology
Graiciuno 8
LT-02241 Vilnius
Lithuania

[2] Abteilung Strukturanalyse
Medizinische Hochschule Hannover
Carl Neuberg Strasse 1
D-30632 Hannover
Germany

[3] Abteilung Strukturforschung
Max-Planck Institut für Biochemie
Am Klopferspitz 18a
D-82152 Martinsried
Germany

## Summary

**Rare-cutting restriction enzymes are important tools in genome analysis. We report here the crystal structure of SdaI restriction endonuclease, which is specific for the 8 bp sequence CCTGCA/GG ("/" designates the cleavage site). Unlike orthodox Type IIP enzymes, which are single domain proteins, the SdaI monomer is composed of two structural domains. The N domain contains a classical winged helix-turn-helix (wHTH) DNA binding motif, while the C domain shows a typical restriction endonuclease fold. The active site of SdaI is located within the C domain and represents a variant of the canonical PD-(D/E)XK motif. SdaI determinants of sequence specificity are clustered on the recognition helix of the wHTH motif at the N domain. The modular architecture of SdaI, wherein one domain mediates DNA binding while the other domain is predicted to catalyze hydrolysis, distinguishes SdaI from previously characterized restriction enzymes interacting with symmetric recognition sequences.**

## Introduction

Type II restriction endonucleases (REases) recognize short, 4–8 bp, nucleotide sequences and cleave phosphodiester bonds in DNA within or close to their recognition sites (Pingoud et al., 2005). They form a large and divergent protein family comprising more than 3600 functionally characterized enzymes that recognize more than 250 different nucleotide sequences (Roberts et al., 2005). Type II REases are subdivided into several subtypes according to their recognition site symmetry, structural organization, cofactor requirement, etc. (Roberts et al., 2003).

Orthodox Type IIP enzymes like EcoRI (G/AATTC) or EcoRV (GAT/ATC) recognize symmetric nucleotide sequences (shown in parentheses) and cleave within their recognition sites (designated by "/"). They share both a common structural core comprising the five-stranded mixed β sheet flanked by α helices (Venclovas et al., 1994; Aggarwal, 1995; Kovall and Matthews, 1998) and a weakly conserved active site structure (Pingoud et al., 2005). The DNA binding sites of Type IIP enzymes, however, are highly diverse and usually form a patch on the protein surface composed of amino acid residues located on the different structural elements (α helices, β strands, loops, etc.) (Jones et al., 1999; Pingoud and Jeltsch, 2001). Orthodox Type IIP REases interact with DNA as homodimers, and each subunit contributes to the recognition of half of the palindromic sequence (Pingoud et al., 2005). To achieve recognition of the interspaced 8 bp pseudopalindromic sequence GGCCNNNN/NGGCC, SfiI uses a monomer core similar to that of the orthodox restriction enzyme EcoRV, but it dimerizes in a very different way (Vanamee et al., 2005). Monomeric forms of Type IIP REases have been recently reported (Xu et al., 2004; Horton et al., 2006).

Type IIE REases like NaeI (GCC/GGC) recognize palindromic nucleotide sequences in a manner similar to the Type IIP enzymes and cleave DNA within the boundaries of their recognition sites; however, they possess a separate DNA binding domain to perform allosteric function. Crystal structures of the Type IIE enzymes NaeI and EcoRII indeed reveal two separate functional modules: an effector DNA binding domain and a catalytic domain (Huai et al., 2000; Zhou et al., 2004). The structures of the effector DNA binding domains differ in NaeI and EcoRII; however, their catalytic domains are similar to those of the orthodox Type IIP enzymes. Moreover, removal of the N-terminal effector DNA binding domain of EcoRII by proteolytic digestion generates an orthodox Type IIP restriction enzyme, demonstrating that the C-terminal domain of EcoRII contains all structural determinants required for DNA sequence recognition and catalysis (Mucke et al., 2002; Tamulaitis et al., 2006a, 2006b).

Restriction enzymes that belong to the Type IIS subtype recognize asymmetric nucleotide sequences and cut outside of their recognition sites. For example, the archetypal Type IIS enzyme FokI recognizes the GGATG site and makes a cut 9 and 13 nt from the recognition site on the top and bottom DNA strands, respectively. Crystal structure analysis of FokI reveals that it is composed of a specific DNA binding module fused to the cleavage domain that possesses a conserved REase catalytic core but cuts DNA in a nonspecific manner (Wah et al., 1997). Indeed, the isolated catalytic domain of FokI shows weak, nonspecific nuclease activity (Li et al., 1992). Modular architecture is also characteristic for another Type IIS enzyme, BfiI, which is composed of two DNA binding domains fused to the dimeric catalytic core similar to the nonspecific nuclease belonging to the phospholipase D family (Grazulis et al., 2005; Zaremba et al., 2004). Therefore, it is currently thought that separate modules for DNA recognition and cleavage are a distinctive feature of the Type IIS restriction

*Correspondence: siksnys@ibt.lt

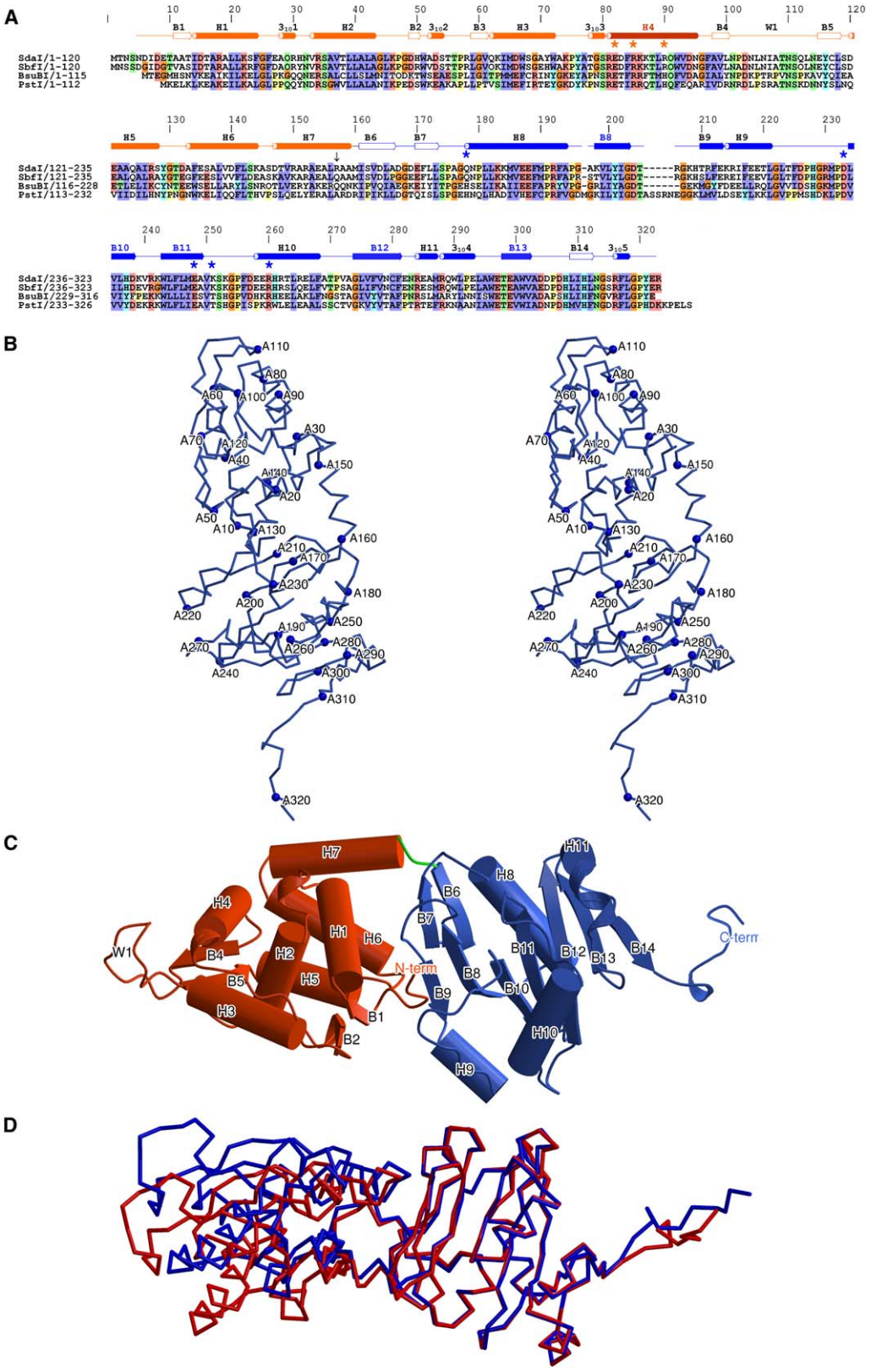


Figure 1. Structure of the SdaI Monomer

(A) Multiple alignment of SdaI, SbfI, BsuBI, and PstI restriction endonucleases. Colors are according to the ClustalX scheme in Jalview. Secondary structure elements of SdaI are shown above the alignment: $\alpha$ helices and $3_{10}$ helices are shown as cylinders, and $\beta$ strands are shown as arrows. The N domain secondary structure elements are colored in orange, and the C domain is shown in blue (according to domain colors in [C]). The putative recognition H4 helix is colored in red, and filled-in arrows correspond to the conserved endonuclease core. Putative active site and DNA binding residues of SdaI subjected to mutagenesis are designated, respectively, by blue and orange asterisks above the alignment.

enzymes that recognize asymmetric nucleotide sequences.

The Type IIP REase SdaI from *Streptomyces diastaticus* Ng7-324 recognizes the palindromic octanucleotide sequence CCTGCAGG and cleaves it after the A base to produce 4 nt 3′ overhangs. In this paper, we present the crystal structure of SdaI solved at 2.0 Å resolution. Structure analysis and biochemical data provided here demonstrate an unexpected modular arrangement of SdaI that is characteristic of Type IIS enzymes and suggest a novel, to our knowledge, mechanism for palindromic nucleotide sequence recognition and cleavage by REase.

## Results

### SdaI Is a Member of the BsuBI/PstI Restriction Endonuclease Family

The SdaI protein sequence reveals significant similarities to members of the BsuBI/PstI protein family, which includes the C terminus of nine proteins (Pfam accession number: PF06616) similar to the Type IIP REases BsuBI and PstI, specific for nucleotide sequence CTGCAG (Xu et al., 1992). The recognition sequence of BsuBI/PstI overlaps with the central part (underlined) of the SdaI target site, C<u>CTGCAG</u>G. Both SdaI and BsuBI/PstI cleave their target sites after an A base to produce 4 nt 3′ overhangs. Protein sequence alignment (Figure 1A) between SdaI and seeding Pfam family members BsuBI and PstI reveals ~30% identical and ~50% similar amino acids. The SbfI REase (EMBL accession number: DQ026300) from *Streptomyces sp*. included in the alignment (Figure 1A) is an isoschizomer of SdaI and shares 83% identical and 91% similar amino acid residues.

The PSI-BLAST search provides 23 protein sequences that are significantly similar to SdaI (Figure S1; see the Supplemental Data available with this article online). However, only BsuBI/PstI and SbfI are known to be functional restriction enzymes. Other proteins are or may be annotated as putative REases due to their sequence similarity to BsuBI/PstI. Interestingly, YenI demonstrates similarities (Figure S1) both to BsuBI/PstI REases and BsuBI/PstI methyltransferases (Antonenko et al., 2003), suggesting that YenI may be a bifunctional enzyme similar to Eco57I (Janulaitis et al., 1992a, 1992b).

### Overall Structure of the SdaI Monomer

The crystal structure of SdaI reveals two protein chains in the asymmetric unit. Despite the presence of the cognate 10 bp oligonucleotide in the crystallization mixture, no electron density for DNA is observed in the crystal. The SdaI monomer has an elongated shape and approximate dimensions of 101 × 41 × 37 Å. Each chain of the SdaI monomer is comprised of two domains (Figures 1B and 1C): the N-terminal domain (N domain), encompassing residues 5–159, and the C-terminal domain (C domain), extending from residue 161 to residue 323. The relative orientation of the N and C domains is slightly different in the individual SdaI monomers in the asymmetric unit (Figure 1D). Superposition of the backbone atoms of individual protein chains reveals a 2.4 Å rmsd for the full-length monomers, a 1.0 Å rmsd for the N domains, and a 0.96 Å rmsd for the C domain subfragments (residues 161–312). Superposition of the complete C domains (residues 161–323) increases the rmsd to 2.1 Å due to the differences in the conformation of the C-terminal tails (Figure 1D), which are presumably affected by the crystal packing interactions.

The surface buried at the N and C domain interface is 1400 $Å^2$, which is similar to the average protein-protein complex interface of 1600 ± 400 $Å^2$ (Lo Conte et al., 1999). The set of interdomain contacts involves five hydrogen bonds and a number of van der Waals and hydrophobic interactions provided mainly by the side chain of the Leu7 residue that is inserted into the hydrophobic pocket formed by the side chains of the Tyr201, Phe212, Phe225, and Met231 residues located at the C domain. The free space between the domains in the crystal is occupied by a HEPES molecule, presumably from the crystallization buffer.

### The C Domain: Structural Similarities to the Type IIP REases

The C domain of SdaI extends from the Ile161 residue to the Arg323 residue (Figure 1C). It is composed of nine β strands (B6–B14) and four α helices (H8–H11) that are arranged into an α/β architecture. The central β sheet is formed of seven strands (B8–B14), flanked by α helices H8, H9, H10, and H11. A DALI (Holm and Sander, 1993) database search for similar three-dimensional structures revealed significant similarities to REases FokI (PDB ID: 2FOK; Z score of 5.1) and NgoMIV (PDB ID: 1FIU; Z score of 4.3), TnsA endonuclease (PDB ID: 1F1Z; Z score of 4.1), and archaeal holiday junction resolvase (PDB ID: 1GEF; Z score of 4.0). Figure 2A shows the structural superposition of the SdaI C domain on the catalytic domain of FokI REase. Strands B8 and B10–B13 at the SdaI C domain overlap with the central β sheet at the FokI core and exhibit the same topology, indicating that the C domain of SdaI has the familiar REase α/β core.

### Putative Active Site Residues of SdaI Are Located at the C Domain

Most of the Type II enzymes with known structures belong to the PD-(D/E)XK superfamily of nucleases, containing the active site consensus (P)**D**$X_n$**(D/E)XK** (highly conserved residues are shown in bold type) (Aggarwal, 1995; Aravind et al., 2000; Bujnicki, 2004). The conserved active site residues are positioned at one end of the conserved β meander (Venclovas et al., 1994; Aggarwal, 1995; Pingoud and Jeltsch, 2001). The position of the second acidic residue, however, is less conserved

The trypsin cleavage site is indicated by the arrow above the SdaI sequence. The numbering above the alignment corresponds to the residue numbers of SdaI.

(B) Stereo $C_\alpha$ trace of the SdaI monomer. Every tenth residue is marked with a closed circle and is labeled.

(C) Cartoon diagram of the SdaI monomer: the N domain is depicted in orange, the C domain is depicted in blue, and the linker is depicted in green. α helices and β strands are numbered; W1 designates the wing in the wHTH motif.

(D) $C_\alpha$ trace for the superposition of two monomers (blue and red) of SdaI in the asymmetric unit. The transformational matrix of the C domain (residues 161–312) was used for the superposition of the entire molecule.

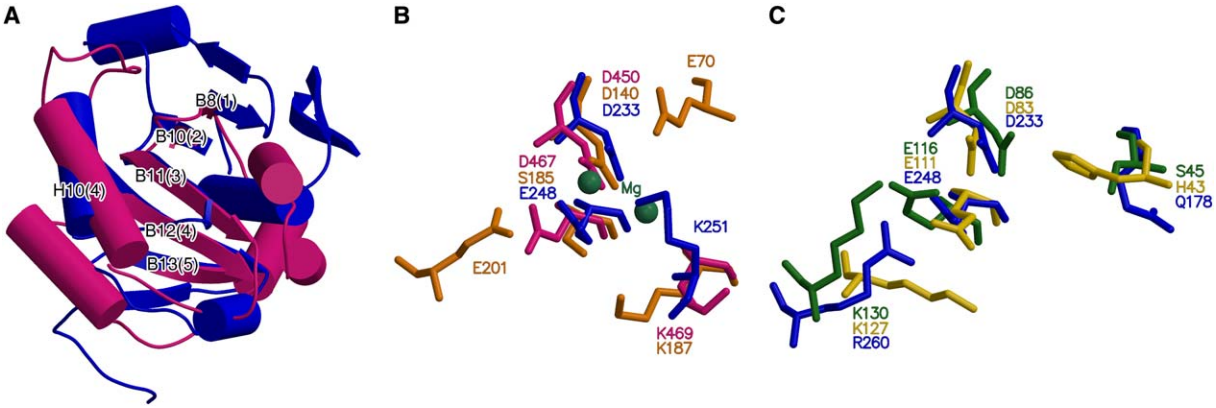


Figure 2. The C Domain of SdaI and the Putative Active Site

(A) Superposition of the SdaI C domain (blue) with the FokI catalytic domain (pink). Strands B8 and B10–B13 of SdaI correspond to the central β sheet (matching FokI strands are labeled in parentheses). Structurally equivalent α helices H10 of SdaI and α4 of FokI are also labeled.

(B) The active site of SdaI superimposed with those of FokI and NgoMIV. Putative SdaI active site residues are depicted in blue, FokI residues are depicted in pink, and NgoMIV residues are depicted in orange. Mg atoms from the NgoMIV structure (1FIU) are shown in green.

(C) The active site of SdaI superimposed with those of all3650 and Tt1808 proteins of the COG4636+ family. Putative SdaI active site residues are depicted in blue, all3650 residues are depicted in gold, and Tt1808 residues are depicted in green.

in the protein sequence. Indeed, in NgoMIV-like enzymes, it migrates to the helix outside the β meander (Deibert et al., 2000; Siksnys et al., 2004). Moreover, in MspI and HinP1I restriction enzymes, the second acidic residue is replaced by Asn117 and Gln81, respectively (Xu et al., 2004; Yang et al., 2005).

Structural superposition revealed (Figure 2B) that amino acid residues Asp233 and Glu248 of SdaI spatially overlap with the Asp450 and Asp467 residues at the active site of FokI, indicating that SdaI follows a canonical rather than a NgoMIV-like arrangement of acidic active site residues (Pingoud and Jeltsch, 2001). Strikingly, the Lys residue that typically follows the second acidic residue in the conserved active site motif (P)$DX_n$(D/E)XK is missing in SdaI. In fact, Lys251 is 2 residues away from the second conserved acidic residue, suggesting that the $PD^{233}X_{14}E^{248}X_2K^{251}$ motif may correspond to the active site of SdaI. However, Lys251 shows no spatial overlap with the conserved lysine at the active sites of canonical PD-(D/E)XK restriction enzymes (Figure 2B). Furthermore, it is not conserved in the homologous BsuBI/PstI REases, which contain threonine in the equivalent position (Figure 1A).

Recently, an active site in the COG4636+ family of putative nucleases remotely related to Holliday junction resolvases has been identified by using the fold recognition technique (Feder and Bujnicki, 2005). Most members of the COG4636+ family exhibit a new, unusual variation of the (P)$DX_n$(D/E)XK motif. In this motif, two acidic residues are conserved, while the catalytic Lys residue migrates in the sequence. The crystal structure of the COG4636+ family protein Tt1808 from *Thermus thermophilus Hb8* solved recently (PDB ID: 1WDJ) confirmed the active site structure predicted by modeling (Feder and Bujnicki, 2005). Structural comparison reveals that the SdaI putative active site residues Asp233 and Glu248 superimpose with acidic residues at the active sites of COG4636+ family proteins Tt1808 from *T. thermophilus Hb8* and all3650 from *Nostoc sp.* (Figure 2C). Interestingly, Gln178 of SdaI is spatially close to the Ser45 and His43 residues of the Tt1808 and all3650 proteins, respectively. Moreover, in most of the BsuBI/PstI family proteins, His and Ser correspond to the Gln178 residue in the SdaI sequence (see Figure S1). Furthermore, Arg260 of SdaI, which is absolutely conserved between BsuBI/PstI family members (see Figure S1), is spatially close to the Lys127 and Lys130 residues (Figure 2C) conserved in all3650, Tt1808, and other COG4636+ family proteins.

### Mutational Analysis of Putative Active Site Residues

In order to test the functional importance of the putative SdaI active site residues (Figure 2C), we constructed a set of alanine mutants and assayed the DNA cleavage and binding ability of each mutational variant (Table 1). The catalytic activity of mutant proteins was examined by using phage λ DNA. Alanine replacement of Asp233, Glu248, and Gln178 completely abolished the DNA cleavage activity of mutant proteins (Table 1). Arg260 and Lys251 mutations were less deleterious since mutant proteins still exhibited 1% and 5%, respectively, of the wild-type (wt) SdaI cleavage activity. Replacement of Lys251 to Thr, present in the equivalent position of BsuBI/PstI REases (Figure 1A), yielded a mutant that possessed ~10% of the wt activity. DNA binding affinity of mutant proteins was assayed by gel shift analysis with 30 bp duplex oligonucleotides that either possessed or lacked the SdaI recognition sequence. Mutation of Asp233 (Figure 3A), Glu248, Gln178, Lys251, and Arg260 (data not shown) did not significantly affect DNA binding, as the amount of mutant protein required to shift half of the labeled cognate 30 bp oligonucleotide varied less than 2- or 3-fold from the wt (Table 1).

### N Domain of SdaI: Structural Similarities to the Helix-Turn-Helix Motif

The N domain (residues 5–159) of SdaI is mainly α-helical and is composed of seven α helices, H1–H7; three $3_{10}$ helices; and five short β strands, B1–B5 (Figure 1C).

Table 1. DNA Cleavage and Binding Activities of SdaI Mutants

| Mutation | Location | Activity, %[a] | Effect on Specific DNA Binding Affinity[b] |
|---|---|---|---|
| Q178A | C domain | 0 | 2× enhanced |
| D233A | C domain | 0 | No effect |
| E248A | C domain | 0 | 2× enhanced |
| K251A | C domain | 5 | 2× enhanced |
| K251T | C domain | 10 | 3× reduced |
| R260A | C domain | 1 | 2× reduced |
| E82A | N domain wHTH | 0.01 | 30× reduced |
| R85A | N domain wHTH | 0 | No binding |
| R90A | N domain wHTH | 0.3 | 100× reduced |

[a] λ DNA cleavage activity compared to wt SdaI.
[b] The protein concentration at which it binds 50% DNA was estimated from DNA binding experiments as described in Experimental Procedures and compared to wt SdaI.

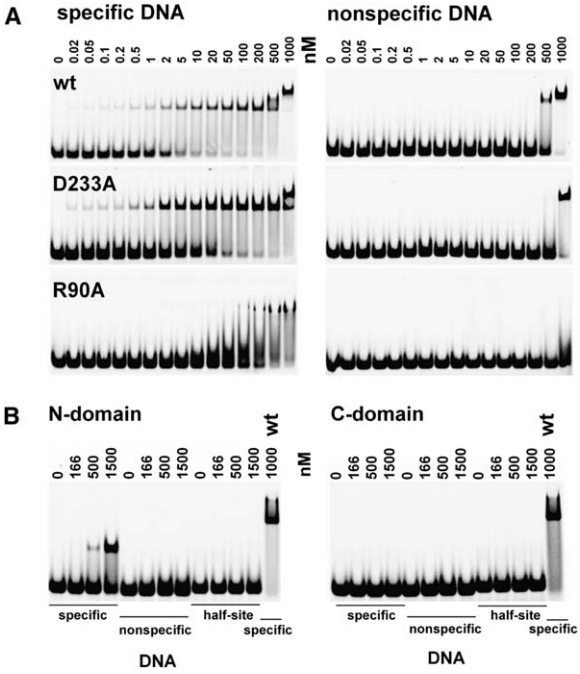


Figure 3. Gel Shift Analysis of DNA Binding

(A) Wild-type SdaI and mutant DNA binding. Only selected representatives of catalytic (D233A) and DNA binding (R90A) mutants are shown.

(B) N and C domain DNA binding. Increasing amounts of protein (concentrations are indicated above the lanes in terms of dimer for wt SdaI and the C domain and monomer for the N domain) were incubated with $^{33}$P-labeled 30 bp-specific and nonspecific oligodeoxynucleotides. "Half-site" oligonucleotides used in domain binding studies were 33 bp in length and contained only half of the SdaI recognition sequence. Protein-DNA complexes were analyzed in polyacrylamide gel under nondenaturing conditions as described in Experimental Procedures.

Structural comparisons of the N domain with unique domains present in the SCOP database revealed significant similarity to the DNA/RNA binding, three-helical bundle proteins (SCOP fold 46688 cf a.4), including SmtB repressor (PDB ID: 1SMT), DNA binding protein MJ223 (PDB ID: 1KU9), transcription factor TFIIE-β (PDB ID: 1D8J), the AFX DNA binding domain (PDB ID: 1E17), and transcription regulatory protein MotA (PDB ID: 1BJA). SmtB repressor showed the most extended structural similarities to SdaI (Figure 4): 58 $C_\alpha$ atoms of SmtB and the N domain of SdaI superimpose with an rmsd value of 1.6 Å.

In the N domain of SdaI, the α helices H2, H3, and H4; β strands B3, B4, and B5; and the loop connecting residues 100–115 form a winged HTH (wHTH) motif (Figure 4). The wHTH motif usually consists of two wings (W1 and W2) and alternating α helices and β strands forming a pattern of H1-B1-H2-H3-B2-W1-B3-W2 (Gajiwala and Burley, 2000). In SdaI, the corresponding secondary structure elements are arranged as H2-(B2-$3_{10}$2)-B3-H3-($3_{10}$3)-H4-B4-W1-B5, indicating that the second wing, W2, is missing. The wHTH motif is often found in DNA binding proteins, including both prokaryotic and eukaryotic transcription factors (Gajiwala and Burley, 2000). In FokI REase (1FOK), however, the wHTH domain is involved in protein-protein interactions (Wah et al., 1997). Interestingly, the SdaI N domain shows significant structural resemblance to the wHTH motif in the D3 subdomain of FokI (47 overlapping residues with an rmsd value of 1.6 Å). Since the crystal structure of SdaI was solved without DNA, it is not clear a priori if the wHTH motif of SdaI is involved in DNA binding.

### Identification of the DNA Recognition Determinants at the N Domain

Transcription factors containing the wHTH motif usually interact with DNA in the major groove either through residues located on helix H3 (the equivalent to H4 in SdaI) or wing W1, depending on which protein surface is more basic (Gajiwala and Burley, 2000). Helix H4 of SdaI (Figure 1A) contains three arginine and two lysine residues, while no basic residues are present in the W1 wing. Moreover, basic residues located on the H4 helix are conserved between SdaI and BsuBI/PstI, while residues positioned on the W1 wing are diverse (Figure 1A). Comparative structural analysis reveals that amino acid residues 1, 2, 5, 6, and 9 in the recognition helix of the canonical HTH motif are most often involved in the sequence-specific contacts (Suzuki et al., 1995; Pabo and Nekludova, 2000). The structurally equivalent residues located on the H4 helix of SdaI are Arg81, Glu82, Arg85, Lys86, Lys87, and Arg90. Therefore, we suggested that helix H4 contributes to the DNA binding interface of SdaI REase.

To determine the functional importance of the SdaI residues located on the H4 helix, Glu82, Arg85, and Arg90 residues (one from each predicted DNA binding patch on the H4 helix [Suzuki et al., 1995; Pabo and Nekludova, 2000]) were replaced by alanine, and the DNA cleavage and binding properties of mutant proteins were analyzed. Alanine replacement of Arg85 completely abolished both the DNA binding and cleavage ability of SdaI (Table 1). Mutation of the Glu82 and Arg90 residues significantly decreased DNA binding affinity and nearly completely abrogated the cleavage activity of SdaI (Figure 3A and Table 1), indicating that residues located on the H4 helix of the wHTH motif of SdaI are involved in sequence-specific contacts within the recognition site. Conservation of the SdaI residues involved in the sequence recognition in BsuBI/PstI REases (Figure 1A) suggests that the overlapping specificities of SdaI and

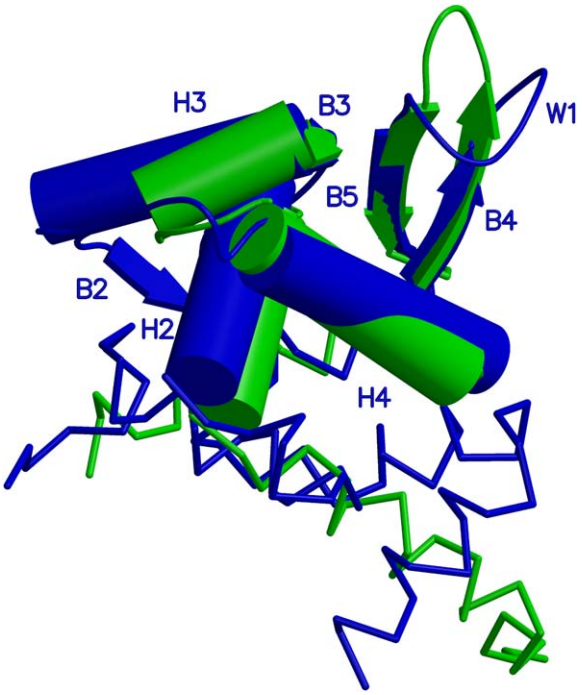


Figure 4. Superposition of the SdaI N Domain with the SmtB Repressor

The SdaI N domain is depicted in blue, and the SmtB repressor is depicted in green. wHTH motifs are shown in cartoon representation.

BsuBI/PstI most likely evolved on the same structural scaffold.

## Domain Separation by Proteolysis and Functional Analysis

Structural analysis supported by mutational studies suggests that SdaI is composed of two domains that may perform DNA binding and cleavage functions, respectively. In order to test this hypothesis and to examine the putative domain structure of SdaI, the full-length protein was subjected to limited proteolysis with trypsin. SDS-PAGE analysis yielded two distinct protein fragments with apparent masses of ~20 and ~16 kDa (Figure 5A). The sum of the apparent molecular masses of these fragments is close to the molecular mass of the full-length protein (36.5 kDa). In order to locate the proteolytic fragments in the amino acid sequence of SdaI, the N-terminal amino acid sequences of the trypsin fragments were determined by microsequencing. The N-terminal sequence of the ~20 kDa fragment matched exactly the sequence starting at Ala158 in the intact protein. Hence, the ~20 kDa fragment must have been liberated by a single trypsin cleavage at Arg157 and corresponds to the C domain of SdaI (Figure 5B). The N-terminal sequence analysis of the ~16 kDa fragment yielded two sets of sequence, corresponding to the N terminus of the intact SdaI protein with and without the first Met (Figure 5B).

Apparent mass values for the N and C domains of SdaI were estimated from their elution volumes during size-exclusion chromatography. The C domain, which had an apparent mass of ~20 kDa in SDS-PAGE, eluted from the gel-filtration column as a single peak, at a volume corresponding to an apparent mass of 40 kDa (Figures 5C and 5D). This value is close to that expected for a dimer of the C-terminal fragment. The gel-filtration analysis of the N-terminal fragment of SdaI gave an apparent mass of ~18 kDa, which is close to the value for the N domain monomer (Figures 5C and 5D). According to the crystal structure of SdaI, the linker region connecting the N and C domains contains only one amino acid residue, Met160. The trypsin cleavage site at

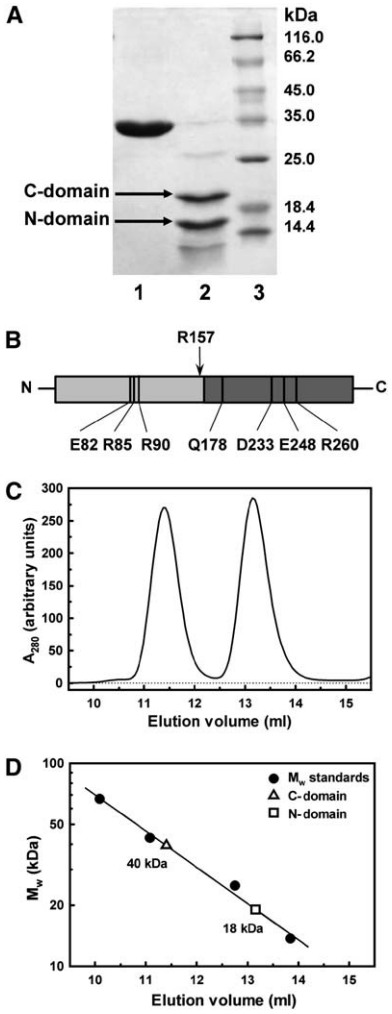


Figure 5. SdaI Domain Analysis

(A) Limited proteolysis of SdaI analyzed by SDS-PAGE. SdaI was digested with trypsin as described in Experimental Procedures. Lane 1, undigested SdaI; lane 2, SdaI digested with trypsin; lane 3, protein standards.

(B) Domain organization of SdaI. Trypsin cleaves SdaI at the position marked by the arrow to yield the N domain (residues 1–157, light gray) and the C domain (residues 158–323, dark gray). The residues important in DNA binding (Glu82, Arg85, and Arg90) are located in the N domain, and the catalytic amino acid residues Gln178, Asp233, Glu248, and Arg260 are located in the C domain.

(C) Elution profiles of the N-and C-terminal domains of SdaI on a Superdex75 HR column. Gel filtration was performed as described in Experimental Procedures.

(D) Molecular masses of SdaI domains. The apparent mass values of the SdaI N domain (open square) and C domain (open triangle) were calculated by interpolation from the standard curve (closed circle) obtained by using a set of proteins of known molecular mass.

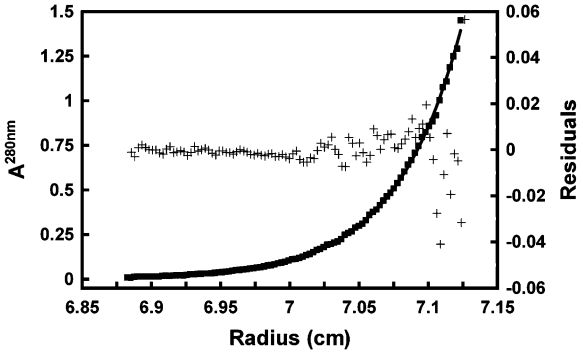


Figure 6. Sedimentation Equilibrium Analysis of SdaI

Sedimentation equilibrium of 5.7 μM SdaI (loading concentration) at 19,000 rpm and 25°C in 10 mM Tris (pH 7.4), 200 mM KCl, 0.1 mM EDTA, and 0.1 mM DTT. The solid line represents the theoretical concentration profile for a molar mass of 72.8 kg/mole; "+" is the difference between calculated and experimental data (closed square) (note the different ordinate for the residuals).

Arg157, located on α helix H7, is in the close vicinity of the proposed linker region.

Isolated N and C domains of SdaI were assayed for their ability to cleave and/or bind a specific DNA sequence. DNA cleavage analysis with phage λ DNA (five SdaI recognition sites) or pUC18 plasmid DNA (one SdaI site) revealed no DNA cleavage by either the N domain or the C domain (data not shown). Gel shift analysis of DNA binding revealed that the C domain shows no binding to the 30 bp oligonucleotide duplex possessing or lacking the SdaI recognition sequence (Figure 3B). The N domain, however, bound cognate DNA, albeit with much lower affinity compared to the wt enzyme (Figure 3B).

### SdaI Dimer Structure

In analytical ultracentrifugation, SdaI sedimented with $s_{20°C,W}$ = 4.44 S, yielding a frictional ratio of 1.4 for a dimer and 0.9 for a monomer. Thus, the measured sedimentation rate is incompatible with the protein being a monomer and indicates a dimeric state. Sedimentation equilibrium experiments (Figure 6) gave a molar mass of 73 ± 5 kDa at all concentrations and speeds used. This value shows the protein to be a dimer, and there are no indications that monomers or tetramers are present in significant amounts. The asymmetric unit in the SdaI crystal also contains a dimer. The distance between the active sites in the SdaI dimer, however, exceeds the distance between the scissile DNA phosphates by more than two times; therefore, the dimer present in the crystal, presumably, is not biologically relevant. In looking for the putative functional dimer, we have analyzed all possible A and B protomer contacts in the crystal (see Supplemental Results). However, in all cases, the distances between the active sites were not compatible with the distances between scissile phosphates; therefore, we think that neither dimer present in the crystal lattice is functionally significant. Furthermore, biochemical and crystallographic studies are required to identify the interface of the functional SdaI dimer.

## Discussion

Most of the Type II REases recognizing 4–6 bp target DNA are routinely used in molecular biology applications, while rare-cutting enzymes specific for continuous palindromic 8 bp sequences are often employed in genome analysis. Restriction enzymes specific for 8 bp sequences, however, are rare in nature; therefore, the expansion of their repertoire through protein engineering is a challenging task (Samuelson et al., 2006). The rational approaches for specificity engineering of the rare-cutting REases, however, are hindered by a lack of structural data.

We have solved the crystal structure of the SdaI REase at 2.0 Å resolution and have provided the first, to our knowledge, structure of an enzyme that recognizes the continuous palindromic 8 bp sequence CCTGCAGG and cleaves it, leaving 4 nt 3′ overhangs. According to the current nomenclature (Roberts et al., 2003), it is assigned to the Type IIP subtype. Similar to many orthodox Type IIP enzymes, SdaI cuts double-stranded DNA

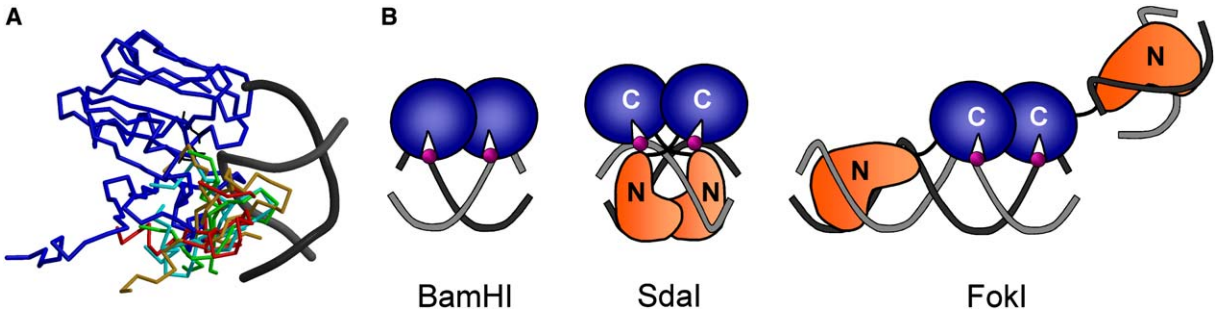


Figure 7. Model of SdaI Interaction with DNA

(A) Putative C domain contacts with the DNA minor groove. Structural superposition of the SdaI active site residues Asp233 and Glu248 (shown in black) with the structural equivalents in EcoRV, NaeI, BglI, and PvuII results in the overlay of the conserved β sheet core and positions the C-terminal domain of SdaI (blue) in the DNA minor groove. The conserved β sheet core of EcoRV, NaeI, BglI, and PvuII is omitted for clarity, and only the structural elements involved in sequence-specific contacts in the major groove (Huai et al., 2000) are shown: EcoRV (green, residues 101–113 and 175–195), NaeI (cyan, residues 101–113 and 135–158), PvuII (red, residues 75–88 and 129–152), BglI (gold, residues 149–162 and 256–279). The C domain of SdaI lacks such structural elements. The DNA is from EcoRV-DNA structure 1AZ0.

(B) Cartoon representation of BamHI, SdaI, and FokI complexes with DNA. The BamHI dimer is depicted in blue, the N-terminal DNA binding domains of SdaI and FokI are depicted in orange, and the catalytic C-terminal domains of SdaI and FokI are depicted in blue. DNA is shown in gray, and scissile phosphates are shown in magenta.

in a concerted manner via independent reactions at individual recognition sites (Bilcock et al., 1999).

In contrast to other Type IIP enzymes, which are single-domain proteins, the SdaI monomer is composed of two structural domains. The C domain of SdaI shows a typical REase fold characteristic of orthodox Type IIP/IIS enzymes (Figure 2A) and harbors amino acid residues Gln178, Asp233, and Glu248, which are crucial for catalysis (Figure 3A and Table 1). The structural arrangement of the putative catalytic/metal binding residues of SdaI is similar to the active sites of putative nucleases of the COG4636+ family (Feder and Bujnicki, 2005) that are remotely related to the Holliday junction resolvases/nucleases (Figures 2B and 2C) rather than to the canonical PD-(D/E)XK family REases. Moreover, the SdaI-like active site structure is likely to be conserved among members of the BsuBI/PstI family (Figure S1).

The cleavage position within the recognition site assumes that SdaI should approach DNA from the minor groove side (Anderson, 1993). Indeed, EcoRV, NaeI, and PvuII enzymes (Winkler et al., 1993; Cheng et al., 1994; Huai et al., 2001), producing blunt ends, or BglI (Newman et al., 1998), generating 3′ overhangs, position their catalytic residues at the scissile phosphates in the minor groove and interact with the target sequence in the major groove via a β ribbon protruding out of the conserved catalytic core (Figure 7A). While the SdaI catalytic domain has a typical REase fold, it makes a compact globule, which has no loops or other structural elements that might be involved in DNA sequence recognition (Figure 7A). Therefore, SdaI must use a different strategy for DNA sequence recognition.

Structural analysis and biochemical data indicate that SdaI residues involved in target site recognition are clustered within the separate N domain, which contains the wHTH motif common among DNA binding proteins. According to the mutational data, the Glu82, Arg85, and Arg90 residues located on the putative recognition helix of the wHTH motif are crucial for the DNA binding specificity of SdaI (Figure 3A and Table 1). Proteins containing the wHTH motif are usually dimers, and they interact with DNA by inserting recognition helices into the major groove. Each wHTH motif within the dimer usually interacts with half of the symmetric recognition site. Therefore, dimerization of the N domains of SdaI may be required in order to achieve recognition of its palindromic 8 bp site. Size-exclusion chromatography demonstrates that the isolated N domain of SdaI is a monomer (Figures 5C and 5D). It shows no cognate DNA binding at low protein concentrations; however, it specifically binds DNA at elevated concentrations (Figure 3B). The low cognate DNA binding affinity of the isolated N domain may be due to the decreased $K_A$ for the monomer-dimer association required for DNA binding. Therefore, high concentrations of the N domain may be necessary to achieve dimerization and trigger specific DNA binding (Figure 3B). Neither SdaI nor the isolated N domain binds to the oligonucleotide containing the half-site CCTG of the SdaI recognition sequence (Figure 3B).

Hence, SdaI presumably evolved through the fusion of the Holliday junction resolvase-like nuclease domain and the wHTH DNA binding domain. DNA sequence recognition is achieved by SdaI through the N domain, possessing the wHTH motif, while phosphodiester bond cleavage is performed by the C domain. The SdaI structure differs radically from that of the orthodox Type IIP restriction enzymes like BamHI, which has DNA binding and cleavage elements within a single domain (Figure 7B). The SdaI structure is also distinct from that of the SfiI REase (Vanamee et al., 2005) that recognizes the interrupted 8 bp sequence GGCCNNNN/NGGCC. Modular organization of SdaI is similar to that of the Type IIS enzyme FokI rather than the orthodox Type IIP REases (Figure 7B).

SdaI, similar to FokI, exhibits a modular architecture and possesses separate domains for sequence recognition and cleavage. However, unlike FokI, it recognizes the symmetric nucleotide sequence CCTGCAGG and cleaves it within the recognition site. In contrast to FokI, which is a monomer and dimerizes only in the presence of DNA, apo-SdaI in solution is a dimer. Moreover, while in the FokI-DNA complex dimerization is achieved through the catalytic C-terminal domains, both the N and C domains presumably contribute to the SdaI dimer interface in the SdaI-DNA complex. Indeed, the isolated C domains of SdaI form a dimer in solution (Figures 5C and 5D), while dimerization of the N domains that are monomers in solution seems to be necessary for cognate DNA binding (see above). The modular architecture of SdaI could be reconciled with its recognition site symmetry and the cleavage pattern if one assumes that SdaI binds specific DNA sequences through its N domain dimer by inserting recognition helices of the wHTH motif into the major groove and positioning two catalytic C domains at the scissile phosphates in the minor groove to generate a double-strand break (Figure 7B).

In conclusion, the rare-cutting SdaI restriction enzyme appears to be the first, to our knowledge, Type IIP REase that has separate structural and functional modules for sequence recognition and catalysis. It challenges the paradigm that modular architecture is a characteristic feature of Type IIS enzymes. Other Type IIP REases and hypothetical proteins belonging to the BsuBI/PstI family (Figure S1) presumably exhibit a similar modular structure.

## Experimental Procedures

### Strains, Plasmids, and DNA

Cloning of the SdaI restriction-modification system from *Streptomyces diastaticus* Ng7-324 will be described elsewhere. Recombinant wt and mutant genes of SdaI were cloned and expressed in the *Escherichia coli* ER2267 strain (*F′traD36proA⁺B⁺lacI^q Δ(lacZ)M15/e14⁻(McrA⁻)Δ(lac-proAB)endA1gyrA96(Nal^r)thi-1hsdR17($r_k^-m_k^+$)glnV44relA1recA1*). Strain ER2267 containing plasmid p184-SdaIM ($Cm^R$) bearing the *sdaIM* gene was used as a host for transformation of pAL-SdaIR plasmid ($Ap^R$) containing the *sdaIR* gene.

All mutants were obtained by the two-step "megaprimer" method (Barik, 1995). The *sdaIR* gene region was sequenced to confirm that only the designed mutations had been introduced.

All oligonucleotides, except the one used in crystallization, were purchased from Metabion. Phage λ DNA, pUC18, and the cognate 10 bp oligonucleotide were purchased from Fermentas.

### Protein Expression and Purification

The SdaI protein was purified by chromatography on Heparin-Sepharose (Amersham Biosciences), Phosphocellulose P11 (Whatman), Blue-Sepharose, and Q-Sepharose (Amersham Biosciences) columns. Fractions containing SdaI were pooled and dialyzed against the SdaI storage buffer (20 mM Tris-HCl [pH 7.5 at 25°C],

Table 2. Data Collection, Phasing, and Refinement Statistics

| Crystal | Native | $HgCl_2$ | |
|---|---|---|---|
| | | λ1 | λ2 |
| **Data Collection Statistics** | | | |
| Wavelength (Å) | 0.9755 | 1.00 | 1.0098 |
| Space group | $C222_1$ | $C222_1$ | $C222_1$ |
| Unit cell (Å) | a = 66.9, b = 238.0, c = 111.4 | a = 66.8, b = 235.9, c = 111.2 | a = 66.8, b = 235.9, c = 111.2 |
| Max. resolution (Å) | 2.0 | 2.52 | 2.54 |
| Number of unique reflections | 60,523 | 29,858 | 29,309 |
| Redundancy | 7.4 | 4.3 | 8.0 |
| $R_{merge}$[a] (last shell) | 0.097 (0.393) | 0.089 (0.281) | 0.081 (0.258) |
| Completeness (%) (last shell) | 100 (100) | 98.8 (94.7) | 99.1 (93.8) |
| I/σ(I) (last shell) | 5.6 (1.4) | 3.0 (2.4) | 8.2 (2.9) |
| **Phasing Statistics** | | | |
| $R_{iso}$[b] | | 0.257 | 0.247 |
| Phasing power[c] | | | |
| $Iso_{centric}$ | | 1.21 | 1.18 |
| $Iso_{acentric}$ | | 1.85 | 1.85 |
| $R_{Cullis}$[d], centric | | 0.79 | 0.79 |
| $R_{Cullis}$[d], acentric | | 0.65 | 0.66 |
| FOM[e] | | | |
| Centric (%) | 70.30 | | |
| Acentric (%) | 52.34 | | |
| All (%) | 54.14 | | |
| **Refinement Statistics** | | | |
| Resolution range (Å) | 30.0–2.0 | | |
| $R_{cryst}$[f] | 0.18 | | |
| $R_{free}$[f] | 0.21 | | |
| Number of unique reflections used in refinement | 54,406 | | |
| Test set size | 10% reflections, randomly selected | | |
| Rms bond length deviation | 0.011 Å | | |
| Rms angle deviation | 1.4° | | |
| Number of atoms | 5,713 | | |
| Number of solvent molecules | 470 | | |

[a] $R_{merge} = \Sigma|I_{hi} - \langle I_h \rangle|/\Sigma I_h$, where $I_{hi}$ is the intensity value of the $i^{th}$ measurement of reflection h, and $\langle I_h \rangle$ is the average measured intensity of reflection h.
[b] $R_{iso} = \Sigma|F_p - F_{ph}|/\Sigma|F_p|$, where $F_p$ and $F_{ph}$ are native and derivative structure factors, respectively.
[c] Phasing power = $\langle|F_{h(o)}|\rangle$/rmsd(ε), where $\langle|F_{h(o)}|\rangle$ is the mean calculated amplitude for the heavy atom model, and rmsd(ε) is the root-mean-square lack of closure error for the isomorphous differences.
[d] $R_{Cullis} = \Sigma|E|/\Sigma_h|F_p - F_{ph}|$, where E is lack of closure, and $F_p$ and $F_{ph}$ are native and derivative structure factors, respectively.
[e] FOM, figure of merit.
[f] $R_{cryst}/R_{free} = \Sigma_h|F_{h(o)} - F_{h(c)}|/\Sigma_h F_{h(o)}|$. $R_{free}$ was calculated with 10% of the data excluded from refinement.

100 mM NaCl, 1 mM EDTA, 2 mM DTT, 50% [v/v] glycerol) and stored at −20°C. The protein preparation was ~99% homogeneous by SDS-PAGE analysis. Mutant SdaI variants were expressed by following the procedures used for the wt protein and were purified to >90% homogeneity by chromatography on Heparin-Sepharose, Blue-Sepharose, and Phosphocellulose P11 columns.

### Crystallization and Data Collection

Prior to crystallization, small aliquots of SdaI were dialyzed against the crystallization buffer (10 mM Tris-HCl [pH 7.5 at 25°C], 100 mM KCl, 0.1 mM EDTA, 1 mM DTT, 0.02% $NaN_3$) and concentrated to ~12 mg $ml^{-1}$. Protein-DNA complexes were prepared by mixing 10 bp duplex 5′-T**CCTGCAGG**A-3′ (SdaI recognition site is underlined) with wt SdaI to the final protein concentration of 9.0 mg $ml^{-1}$ and a 2.2-fold molar ratio of DNA duplex to protein dimer. Crystals of SdaI were grown by hanging or sitting vapor diffusion over a reservoir of 1.7–2.0 M $(NH_4)_2SO_4$, 0.1 M Na-HEPES (pH 7.5 at 25°C). Drops were formed by mixing equal amounts of the SdaI-DNA complex and reservoir solution, equilibrated over the reservoir at 4°C for 3–5 days, and stored at 16°C. Crystals grew in 1–2 weeks.

The data set of native SdaI crystals was collected at the ESRF ID13 beamline (Grenoble, France) at 100 K. Data collection and refinement statistics are shown in Table 2. The data sets for the Hg derivative (7 month soak in 10 mM $HgCl_2$) were collected at the EMBL/DESY X31 beamline, Hamburg, Germany. Data were collected at two wavelengths: Hg inflection point and high-energy remote. The oscillation images were processed with MOSFLM (Leslie, 1992), SCALA, and TRUNCATE (CCP4, 1994).

### Structure Determination and Refinement

The structure of SdaI was solved by single isomorphous replacement with anomalous scattering. First, two Hg atom positions were identified after the inspection of isomorphous and anomalous Patterson maps, and an additional two were identified by calculation of crossphased difference Fourier's by using phases derived from the first Hg atom positions. All heavy atom sites were confirmed with crossdifference Fourier maps. The MLPHARE program from the CCP4 suite (CCP4, 1994) was used to produce initial phases, which gave an interpretable map after solvent flattening with DM (CCP4, 1994). More than 80% of the protein main chain was built automatically with the ARP/wARP program (Perrakis et al., 1999). The final model was built with the O molecular modeling program (Jones et al., 1991). Refinement was conducted with CNS (Brunger et al., 1998), and the final steps of refinement were made with REFMAC (Murshudov et al., 1997). The structure was refined to an $R_{cryst}$ of 17.7% and an $R_{free}$ of 21.2% with good geometry, except

for Asn313. The C-terminal part of the protein (residues 314–323, Figure 1D) adopts different conformations in both monomers, and Asn313, probably due to crystal packing forces, has bad geometry (the Asn313 in monomer A belongs to the generously allowed region, and in monomer B, it belongs to the additional allowed regions of the Ramachandran plot) despite the fact that its electron density for Asn313 is very good. The crystal structure of SdaI reveals 2 protein chains, 2 HEPES and 2 Tris molecules, 9 sulfate ions, and 470 water molecules in the asymmetric unit. Despite the presence of the cognate 10 bp oligonucleotide in the crystallization mixture, no electron density for DNA is observed in the crystal.

### Structure Analysis

Analysis of the stereochemical quality of the SdaI model was conducted with PROCHECK (Laskowski et al., 1993), and secondary structure was assigned with STRIDE (Frishman and Argos, 1995). TOP3D (Lu, 2000) was used for the structure's superpositions. Solvent-accessible surface areas were calculated with NACCESS (Hubbard and Thornton, 1993). Protein-protein contacts were analyzed with DIMPLOT (Wallace et al., 1995). Figures were prepared with MOLSCRIPT (Kraulis, 1991) and RASTER3D (Merritt and Bacon, 1997).

### DNA Cleavage Assay

The DNA cleavage activity of SdaI and mutant proteins was determined by incubating different amounts of protein in 50 μl reaction buffer (37 mM Tris-acetate [pH 7.0 at 37°C], 15 mM magnesium acetate, 150 mM potassium acetate, 0.1 mg $ml^{-1}$ BSA) (Fermentas) containing 1 μg phage λ DNA. The catalytic activity of the SdaI domains was assessed by adding N and C domains (final concentrations: 0.7 μM monomer and 1.6 μM dimer, respectively) to phage λ DNA (0.6 μg) or plasmid pUC18 (0.15 μg) in 30 μl reaction buffer SdaI. The reaction was carried out for 1 hr at 37°C, and products were analyzed by agarose gel electrophoresis. One unit of enzyme activity is the amount of protein required to completely hydrolyze 1 μg DNA in 60 min in a total reaction volume of 50 μl. Wild-type SdaI had a specific activity of $5 \times 10^6$ U $mg^{-1}$.

### Gel Shift Assay

DNA binding by wt SdaI, mutant proteins, and N and C domains was analyzed by the gel shift assay with 30 bp specific or nonspecific or 33 bp "half-site" oligonucleotide duplexes. The following oligonucleotides were used:

(1) specific 30 bp duplex containing the SdaI recognition sequence (underlined): 5′-AACCTACTACT**CCTGCAGG**TCCTA ATCGTC-3′ and 3′-TTGGATGATGA**GGACGTCC**AGGATTAGC AG-5′;
(2) nonspecific 30 bp duplex that lacks the recognition site for SdaI: 5′-AACCTACTACTCGTCGTGCTCCTAATCGTC-3′ and 3′-TTGGATGATGAGCAGCACGAGGATTAGCAG-5′;
(3) "half-site" 33 bp duplex containing half of the SdaI recognition sequence (underlined): 5′-AATGGGCTCGCACG**CCTG**GTA TTATCGATTGTA-3′ and 3′-TTACCCGAGCGTGC**GGAC**CATAAT AGCTAACAT-5′.

For the gel shift DNA binding experiments, one strand of the duplexes was 5′ end-labeled by [$\gamma$-$^{33}$P]ATP (Amersham Biosciences) by following standard procedures (Sambrook et al., 1989). In the SdaI binding experiments, the DNA concentration in the binding mixture was kept at 1 nM. In the case of isolated N- and C-terminal fragments, 0.5 μM unlabeled oligonucleotide plus 1 nM $^{33}$P-labeled oligonucleotide were present in the reaction mixture. Different amounts of proteins (wt SdaI, mutant, or proteolytic fragments) were incubated with cognate or noncognate oligonucleotide duplexes for 10 min at room temperature in 20 μl binding buffer (10 mM Tris-HCl [pH 7.5 at 25°C], 100 mM KCl, 0.1 mg $ml^{-1}$ BSA, 10% glycerol). Free DNA and protein-DNA complexes were separated by a nondenaturing 6% (w/v) polyacrylamide gel as described (Zaremba et al., 2004). The amount of protein required to shift half of the labeled cognate 30 bp oligonucleotide was evaluated.

### Limited Proteolysis and Isolation of SdaI Domains

Trypsin (Sigma) digestions were performed at 25°C by adding 1/10 (w/w) protease to 37.4 μg SdaI in 60 μl reaction buffer (10 mM Tris-HCl [pH 7.5 at 25°C], 0.1 mM EDTA, 1 mM DTT, 200 mM NaCl, 20% glycerol). The reaction was stopped by adding PMSF (phenylmethylsulphonylfluoride) to a final concentration of 2 mM. The reaction products were resolved by SDS-PAGE by using 15% polyacrylamide gels. Gels were stained with Coomassie brilliant blue-G250. The N terminus of the resulting fragments was determined in the University of Bristol Proteomics Facility.

For preparative isolation of N- and C-terminal fragments of SdaI, 0.88 mg SdaI protein was cleaved with trypsin under the above-mentioned conditions. The reaction was stopped by adding PMSF to a final concentration of 2 mM, and the resulting fragments were separated by gel filtration with a Superdex75 HR column (Amersham Biosciences) equilibrated with buffer containing 20 mM Tris-HCl (pH 7.5 at 25°C), 0.5 M KCl, 0.02% $NaN_3$. Fractions containing 16 kDa and 20 kDa fragments of SdaI, as determined by SDS-PAGE, were pooled separately, dialyzed against the SdaI storage buffer, and stored at −20°C.

### Analytical Centrifugation

Analytical ultracentrifugation was done with a Beckman/Coulter XLA analytical ultracentrifuge equipped with UV scanner optics by using the 8 hole An-Ti analytical rotor. For sedimentation equilibrium experiments, six-channel, double-sector centerpieces were used. Samples were spun for at least 30 hr, and it was assumed that equilibrium had been reached when the measured concentration profiles did not change measurably over a period of at least 12 hr. For signal smoothing, all profiles measured during this 12 hr period were averaged. Apparent molar masses were calculated from these averaged profiles as described earlier (Siksnys et al., 1999), by using a partial specific volume of $7.324 \times 10^{-4}$ $m^3$ $kg^{-1}$ calculated from the amino acid composition determined with the program SENTERP (Laue et al., 1992) (the program is available at http://www.jphilo.mailway.com/). Experiments were done at 12,000 and 19,000 rpm with loading concentrations between 1.9 and 5.7 μM.

Sedimentation velocity analysis was done in double-sector centerpieces made of charcoal-filled epon at a speed of 50,000 rpm. Sedimentation coefficients were evaluated by fitting ideal sedimentation boundaries calculated with a numerical integration of Lamm's differential equation (Kindler, 1997; Lamm, 1929) and were corrected for buffer viscosity and density by using the values given by Lide (1999).

### Supplemental Data

Supplemental Data include one figure and additional results and are available at http://www.structure.org/cgi/content/full/14/9/1389/DC1/.

### Acknowledgments

The authors thank W.J. Mawby at Bristol University for the N-terminal sequence analysis, G. Tamulaitis for advice on preparation and characterization of SdaI mutants, and M. Zaremba for valuable discussions. The authors also acknowledge UAB "Fermentas" for making the SdaI clone available and A. Janulaitis for the critical comments on the manuscript. Data sets were collected on the ID13 beamline of the European Synchrotron Radiation Facility (Grenoble) and the European Molecular Biology Laboratory (EMBL) X31 beamline of the Deutsches Elektronen Synchrotron (DESY) (Hamburg). Measurements at the EMBL outstation were supported by the European Community Access to Research Infrastructure Action of the Improving Human Potential Program. V.S. is a recipient of a Lithuanian State Fellowship. The work at the Institute of Biotechnology (Vilnius) was supported in part by the Howard Hughes Medical Institute and the Max Planck Society and by a Lithuanian Science Foundation doctoral fellowship to G.T.

Received: April 24, 2006
Revised: July 4, 2006
Accepted: July 5, 2006
Published: September 12, 2006

### References

Aggarwal, A.K. (1995). Structure and function of restriction endonucleases. Curr. Opin. Struct. Biol. *5*, 11–19.

Anderson, J. (1993). Restriction endonucleases and modification methylases. Curr. Opin. Struct. Biol. *3*, 24–30.

Antonenko, V., Pawlow, V., Heesemann, J., and Rakin, A. (2003). Characterization of a novel unique restriction-modification system from *Yersinia enterocolitica* O:8 1B. FEMS Microbiol. Lett. *219*, 249–252.

Aravind, L., Makarova, K.S., and Koonin, E.V. (2000). SURVEY AND SUMMARY: holliday junction resolvases and related nucleases: identification of new families, phyletic distribution and evolutionary trajectories. Nucleic Acids Res. *28*, 3417–3432.

Barik, S. (1995). Site-directed mutagenesis by double polymerase chain reaction. Mol. Biotechnol. *3*, 1–7.

Bilcock, D.T., Daniels, L.E., Bath, A.J., and Halford, S.E. (1999). Reactions of type II restriction endonucleases with 8-base pair recognition sites. J. Biol. Chem. *274*, 36379–36386.

Brunger, A.T., Adams, P.D., Clore, G.M., DeLano, W.L., Gros, P., Grosse-Kunstleve, R.W., Jiang, J.S., Kuszewski, J., Nilges, M., Pannu, N.S., et al. (1998). Crystallography & NMR system: a new software suite for macromolecular structure determination. Acta Crystallogr. D Biol. Crystallogr. *54*, 905–921.

Bujnicki, J.M. (2004). Molecular phylogenetics of restriction endonucleases. In Restriction Endonucleases, A. Pingoud, ed. (Berlin: Springer), pp. 63–93.

CCP4 (Collaborative Computational Project, Number 4) (1994). The CCP4 suite: programs for protein crystallography. Acta Crystallogr. D Biol. Crystallogr. *50*, 760–763.

Cheng, X., Balendiran, K., Schildkraut, I., and Anderson, J.E. (1994). Structure of PvuII endonuclease with cognate DNA. EMBO J. *13*, 3927–3935.

Deibert, M., Grazulis, S., Sasnauskas, G., Siksnys, V., and Huber, R. (2000). Structure of the tetrameric restriction endonuclease NgoMIV in complex with cleaved DNA. Nat. Struct. Biol. *7*, 792–799.

Feder, M., and Bujnicki, J.M. (2005). Identification of a new family of putative PD-(D/E)XK nucleases with unusual phylogenomic distribution and a new type of the active site. BMC Genomics *6*, 21.

Frishman, D., and Argos, P. (1995). Knowledge-based protein secondary structure assignment. Proteins *23*, 566–579.

Gajiwala, K.S., and Burley, S.K. (2000). Winged helix proteins. Curr. Opin. Struct. Biol. *10*, 110–116.

Grazulis, S., Manakova, E., Roessle, M., Bochtler, M., Tamulaitiene, G., Huber, R., and Siksnys, V. (2005). Structure of the metal-independent restriction enzyme BfiI reveals fusion of a specific DNA-binding domain with a nonspecific nuclease. Proc. Natl. Acad. Sci. USA *102*, 15797–15802.

Holm, L., and Sander, C. (1993). Protein structure comparison by alignment of distance matrices. J. Mol. Biol. *233*, 123–138.

Horton, J.R., Zhang, X., Maunus, R., Yang, Z., Wilson, G.G., Roberts, R.J., and Cheng, X. (2006). DNA nicking by HinP1I endonuclease: bending, base flipping and minor groove expansion. Nucleic Acids Res. *34*, 939–948.

Huai, Q., Colandene, J.D., Chen, Y., Luo, F., Zhao, Y., Topal, M.D., and Ke, H. (2000). Crystal structure of NaeI-an evolutionary bridge between DNA endonuclease and topoisomerase. EMBO J. *19*, 3110–3118.

Huai, Q., Colandene, J.D., Topal, M.D., and Ke, H. (2001). Structure of NaeI-DNA complex reveals dual-mode DNA recognition and complete dimer rearrangement. Nat. Struct. Biol. *8*, 665–669.

Hubbard, S.J., and Thornton, J.M. (1993). NACCESS—Atomic Solvent Accessible Area Calculations (http://wolf.bms.umist.ac.uk/naccess/).

Janulaitis, A., Petrusyte, M., Maneliene, Z., Klimasauskas, S., and Butkus, V. (1992a). Purification and properties of the Eco57I restriction endonuclease and methylase–prototypes of a new class (type IV). Nucleic Acids Res. *20*, 6043–6049.

Janulaitis, A., Vaisvila, R., Timinskas, A., Klimasauskas, S., and Butkus, V. (1992b). Cloning and sequence analysis of the genes coding for Eco57I type IV restriction-modification enzymes. Nucleic Acids Res. *20*, 6051–6056.

Jones, S., van Heyningen, P., Berman, H.M., and Thornton, J.M. (1999). Protein-DNA interactions: a structural analysis. J. Mol. Biol. *287*, 877–896.

Jones, T.A., Zou, J.Y., and Cowan, S.W. (1991). Improved methods for building protein models in electron density maps and the location of errors in these models. Acta Crystallogr. A *47*, 110–119.

Kindler, B. (1997). Akkuprog: Auswertung von Messungen chemischer Reaktionsgeschwindigkeit und Analyse von Biopolymeren in der Ultrazentrifuge. PhD thesis, Universität Hannover, Hannover, Germany.

Kovall, R.A., and Matthews, B.W. (1998). Structural, functional, and evolutionary relationships between lambda-exonuclease and the type II restriction endonucleases. Proc. Natl. Acad. Sci. USA *95*, 7893–7897.

Kraulis, P. (1991). Molscript - a program to produce both detailed and schematic plots of protein structures. J. Appl. Crystallogr. *24*, 946–950.

Lamm, O. (1929). Die Differentialgleichung der Ultrazentrifugierung. Ark. Mat. Astr. Fys. *21B*, 1–4.

Laskowski, R.A., Moss, D.S., and Thornton, J.M. (1993). Main-chain bond lengths and bond angles in protein structures. J. Mol. Biol. *231*, 1049–1067.

Laue, T.M., Shah, B.D., Ridgeway, T.M., and Pelletier, S.L. (1992). Computer-aided interpretation of analytical sedimentation data for proteins. In Analytical Ultracentrifugation in Biochemistry and Polymer Science, S.E. Harding, A.J. Rowe, and J.C. Horton, eds. (Cambridge, UK: The Royal Society of Chemistry), pp. 90–125.

Leslie, A. (1992). Recent changes to the MOSFLM package for processing film and image plate data. Joint CCP4 + ESF-EAMCB Newsletter on Protein Crystallography *26*.

Li, L., Wu, L.P., and Chandrasegaran, S. (1992). Functional domains in Fok I restriction endonuclease. Proc. Natl. Acad. Sci. USA *89*, 4275–4279.

Lide, D. (1999). CRC Handbook of Chemistry and Physics (Boca Raton, FL: CRC Press).

Lo Conte, L., Chothia, C., and Janin, J. (1999). The atomic structure of protein-protein recognition sites. J. Mol. Biol. *285*, 2177–2198.

Lu, G. (2000). TOP: a new method for protein structure comparisons and similarity searches. J. Appl. Crystallogr. *33*, 176–183.

Merritt, E.A., and Bacon, D.J. (1997). Raster3D: photorealistic molecular graphics. Methods Enzymol. *277*, 505–524.

Mucke, M., Grelle, G., Behlke, J., Kraft, R., Kruger, D.H., and Reuter, M. (2002). EcoRII: a restriction enzyme evolving recombination functions? EMBO J. *21*, 5262–5268.

Murshudov, G.N., Vagin, A.A., and Dodson, E.J. (1997). Refinement of macromolecular structures by the maximum-likelihood method. Acta Crystallogr. D Biol. Crystallogr. *53*, 240–255.

Newman, M., Lunnen, K., Wilson, G., Greci, J., Schildkraut, I., and Phillips, S.E. (1998). Crystal structure of restriction endonuclease BglI bound to its interrupted DNA recognition sequence. EMBO J. *17*, 5466–5476.

Pabo, C.O., and Nekludova, L. (2000). Geometric analysis and comparison of protein-DNA interfaces: why is there no simple code for recognition? J. Mol. Biol. *301*, 597–624.

Perrakis, A., Morris, R., and Lamzin, V.S. (1999). Automated protein model building combined with iterative structure refinement. Nat. Struct. Biol. *6*, 458–463.

Pingoud, A., and Jeltsch, A. (2001). Structure and function of type II restriction endonucleases. Nucleic Acids Res. *29*, 3705–3727.

Pingoud, A., Fuxreiter, M., Pingoud, V., and Wende, W. (2005). Type II restriction endonucleases: structure and mechanism. Cell. Mol. Life Sci. *62*, 685–707.

Roberts, R.J., Belfort, M., Bestor, T., Bhagwat, A.S., Bickle, T.A., Bitinaite, J., Blumenthal, R.M., Degtyarev, S.K., Dryden, D.T.F., Dybvig, K., et al. (2003). A nomenclature for restriction enzymes, DNA methyltransferases, homing endonucleases and their genes. Nucleic Acids Res. *31*, 1805–1812.

Roberts, R.J., Vincze, T., Posfai, J., and Macelis, D. (2005). REBASE—restriction enzymes and DNA methyltransferases. Nucleic Acids Res. *33*, D230–D232.

Sambrook, J., Fritsch, E.F., and Maniatis, T. (1989). Molecular Cloning: A Laboratory Manual (Cold Spring Harbor, NY: Cold Spring Harbor Laboratory Press).

Samuelson, J.C., Morgan, R.D., Benner, J.S., Claus, T.E., Packard, S.L., and Xu, S. (2006). Engineering a rare-cutting restriction enzyme: genetic screening and selection of NotI variants. Nucleic Acids Res. *34*, 796–805.

Siksnys, V., Skirgaila, R., Sasnauskas, G., Urbanke, C., Cherny, D., Grazulis, S., and Huber, R. (1999). The Cfr10I restriction enzyme is functional as a tetramer. J. Mol. Biol. *291*, 1105–1118.

Siksnys, V., Grazulis, S., and Huber, R. (2004). Structure and function of tetrameric restriction enzymes. In Restriction Endonucleases, A. Pingoud, ed. (Berlin: Springer), pp. 237–259.

Suzuki, M., Yagi, N., and Gerstein, M. (1995). DNA recognition and superstructure formation by helix-turn-helix proteins. Protein Eng. *8*, 329–338.

Tamulaitis, G., Mucke, M., and Siksnys, V. (2006a). Biochemical and mutational analysis of EcoRII functional domains reveals evolutionary links between restriction enzymes. FEBS Lett. *580*, 1665–1671.

Tamulaitis, G., Sasnauskas, G., Mucke, M., and Siksnys, V. (2006b). Simultaneous binding of three recognition sites is necessary for a concerted plasmid DNA cleavage by EcoRII restriction endonuclease. J. Mol. Biol. *358*, 406–419.

Vanamee, E.S., Viadiu, H., Kucera, R., Dorner, L., Picone, S., Schildkraut, I., and Aggarwal, A.K. (2005). A view of consecutive binding events from structures of tetrameric endonuclease SfiI bound to DNA. EMBO J. *24*, 4198–4208.

Venclovas, C., Timinskas, A., and Siksnys, V. (1994). Five-stranded $\beta$-sheet sandwiched with two $\alpha$-helices: a structural link between restriction endonucleases EcoRI and EcoRV. Proteins *20*, 279–282.

Wah, D.A., Hirsch, J.A., Dorner, L.F., Schildkraut, I., and Aggarwal, A.K. (1997). Structure of the multimodular endonuclease FokI bound to DNA. Nature *388*, 97–100.

Wallace, A.C., Laskowski, R.A., and Thornton, J.M. (1995). LIGPLOT: a program to generate schematic diagrams of protein-ligand interactions. Protein Eng. *8*, 127–134.

Winkler, F.K., Banner, D.W., Oefner, C., Tsernoglou, D., Brown, R.S., Heathman, S.P., Bryan, R.K., Martin, P.D., Petratos, K., and Wilson, K.S. (1993). The crystal structure of EcoRV endonuclease and of its complexes with cognate and non-cognate DNA fragments. EMBO J. *12*, 1781–1795.

Xu, G.L., Kapfer, W., Walter, J., and Trautner, T.A. (1992). BsuBI–an isospecific restriction and modification system of PstI: characterization of the BsuBI genes and enzymes. Nucleic Acids Res. *20*, 6517–6523.

Xu, Q.S., Kucera, R.B., Roberts, R.J., and Guo, H. (2004). An asymmetric complex of restriction endonuclease MspI on its palindromic DNA recognition site. Structure *12*, 1741–1747.

Yang, Z., Horton, J.R., Maunus, R., Wilson, G.G., Roberts, R.J., and Cheng, X. (2005). Structure of HinP1I endonuclease reveals a striking similarity to the monomeric restriction enzyme MspI. Nucleic Acids Res. *33*, 1892–1901.

Zaremba, M., Urbanke, C., Halford, S.E., and Siksnys, V. (2004). Generation of the BfiI restriction endonuclease from the fusion of a DNA recognition domain to a non-specific nuclease from the phospholipase D superfamily. J. Mol. Biol. *336*, 81–92.

Zhou, X.E., Wang, Y., Reuter, M., Mucke, M., Kruger, D.H., Meehan, E.J., and Chen, L. (2004). Crystal structure of type IIE restriction endonuclease EcoRII reveals an autoinhibition mechanism by a novel effector-binding fold. J. Mol. Biol. *335*, 307–319.

## Accession Numbers

The SdaI coordinates and structure factors have been deposited in the Protein Data Bank under accession code 2IXS.