

<https://doi.org/10.1038/s42003-025-07612-z>

Crystal structures of monomeric BsmI restriction endonuclease reveal coordinated sequential cleavage of two DNA strands



Rémi Sieskind¹ , Sophia Missouri¹, Clément Madru¹, Isciane Commenge¹, Germain Niogret², Marcel Hollenstein² , Yannick Rondelez³ , Ludovic Sauguet¹, Ahmed Haouz⁴ , Pierre Legrand⁵ & Marc Delarue¹

BsmI, a thermophilic Type IIS restriction endonuclease from *Bacillus stearothermophilus*, presents a unique structural composition, housing two distinct active sites within a single monomer. Recognition of the non-symmetrical 5'-GAATGC-3' sequence enables precise cleavage of the top and bottom DNA strands. Synthetic biology interventions have led to the transformation of BsmI into Nb.BsmI, a nicking endonuclease. Here we introduce Nt*.BsmI, tailored for top-strand cleavage, which is inactive on standard double-stranded DNA, but active on bottom-strand nicked DNA, suggesting a sequential cleavage mechanism. Crystallographic structures of pre- and post-reactive complexes with cognate DNA show one major conformational change, a retractable loop possibly governing sequential active site accessibility. The x-ray structures reveal the position of the divalent metal ions in the active sites and the DNA:protein interactions, while the models predicted by AlphaFold3 are incorrect. This comprehensive structural and functional study lays a foundation for rational enzyme redesign and potential applications in biotechnology.

Most of the bacteria and archaea possess one or several restriction/modification (R/M) systems, which consist of two distinct enzymes: a site-specific restriction endonuclease (RE) on the one hand and a corresponding methyltransferase on the other hand. The methyltransferase recognizes the same DNA sites as the coupled RE in the host's genetic material but adds a methyl moiety to adenine (N6-methyl-adenine) or to cytosine (N4-methyl-cytosine or 5-methyl-cytosine) to protect the endogenous DNA against RE cleavage. Upon phage infection, the non-modified exogenous DNA is rapidly digested by the RE. These systems enable bacteria to discriminate between self and non-self DNA, and they can be considered as a primitive immune system¹.

Among the wide variety of REs, the type II represents the most extensively studied family, with over 3500 different enzymes discovered and nearly 80 structures resolved by X-ray crystallography up to date². These REs recognize double-stranded DNA sequences ranging from 4 to 8 base pairs in length and require only Mg(II) as a cofactor to cleave both strands either at

or near their specific recognition sites³. Despite this common function, multiple sequence alignments within this family of enzymes reveal significant diversity, suggesting variations in their structures and mechanisms⁴. Most type II restriction endonucleases (REs) contain a PD-Xn-(D/E)XK catalytic motif that is responsible for the hydrolysis of phosphodiester bonds; however, some belong to other families, such as the H-N-H or GIY-YIG families of homing endonucleases⁵. Among those classified within the PD-Xn-(D/E)XK superfamily, a core structural invariant—a mixed beta-sheet flanked by two alpha-helices—can be consistently observed across all three-dimensional structures of type II restriction endonucleases available in the Protein Data Bank⁶. This scaffold carries the canonical catalytic motif and ensures that the important residues are maintained in the correct conformation⁶.

Type II RE act generally as homodimers or homotetramers and are subdivided into P, A and S subclasses based on the structure of the DNA motifs they recognize. Each monomer of type IIP RE recognizes a

¹Institut Pasteur, Université Paris Cité, CNRS UMR 3528, Unit of Architecture and Dynamics of Biological Macromolecules, 75724 Paris, France. ²Institut Pasteur, Université Paris Cité, CNRS UMR 3523, Unit of Bioorganic Chemistry of Nucleic Acids, 75724 Paris, France. ³Laboratoire Gulliver, UMR7083 CNRS/ESPCI Paris-PSL Research University, 10 rue Vauquelin, 75005 Paris, France. ⁴Institut Pasteur, Université Paris Cité, CNRS UMR 3528, Plateforme de cristallographie-C2RT, 75724 Paris, France. ⁵Synchrotron SOLEIL, 91192 Gif-sur-Yvette, France. ✉e-mail: delarue@pasteur.fr

palindromic sequence and cleaves one of the two strands symmetrically within its recognition site. Type IIA and IIS RE can bind to non-symmetrical sequences but type IIA enzymes cleave both strands within the recognition site, whereas type IIS enzymes can cleave at least one of the two strands at a defined distance from the recognition site³. However, even inside each of these classes, the quaternary structure of REs in the absence or presence of the DNA substrate remains highly variable. Some REs are monomeric, such as AgeI⁷, while others are oligomeric, like BamHI and EcoRI, in the apo form^{8,9}, but they form homodimers in the presence of DNA and cleave both strands simultaneously in a single binding event¹⁰. Others, such as MspI, BclI or HhaI act on DNA as monomers with one catalytic site and cleave the two strands in two distinct steps^{11–13}.

Here, we report the crystallographic structure of a type IIS RE, called BsmI, which is part of the R/M system of *B. stearothermophilus*. BsmI is notable for having two catalytic sites within a single monomer. This characteristic is shared by other enzymes, such as BtsI, BsrI, BtsCI or Mva1269I¹⁴, which are considered as fusions of ancestral heterodimers¹⁰. Like its isoschizomer Mva1269I, BsmI recognizes the sequence 5'-GAATGC-3', cleaving the top strand one base downstream of its recognition site and the bottom strand between the G and the C of the site. BsmI features two distinct catalytic motifs: one canonical (527–PD-X₁₇-E-X-K—548) and the other much less common (90–PD-X₁₇-E-X₁₂-QR—123). Interestingly, it cleaves the two strands in a single binding event but not simultaneously^{15,16}. The presence of two catalytic sites in a single polypeptide led synthetic biologists to inactivate one of the active sites, creating the nicking endonuclease called Nb.BsmI. This enzyme can recognize the same sequence and cleave only the bottom strand¹⁷. In the following, we describe functional and structural studies of BsmI that uncover its peculiar enzymatic mechanism, which will facilitate its further rational engineering.

Results

Design of the inactive mutant N0.BsmI and structure of the complex with its cognate DNA

Based on the previous work of Zhou et al.¹⁸, we co-expressed different mutants of the BsmI RE together with the two corresponding methylases found in the genome of the *B. stearothermophilus* NUB36 strain in a Mrr-, McrA-, McrBC-deficient *E. coli* strain¹⁸. Reducing the growth temperature from 37 °C to 30 °C improved the growth rate and allowed us to achieve an appropriate yield of soluble protein after purification. A quality control of the protein samples was conducted in accordance with the ARBRE-MOBIEU guidelines¹⁹. Considering the mutations described by Xu et al. to create Nb.BsmI¹⁷ and the sequence alignment of BsmI with a few other endonucleases carrying two catalytic sites in a single polypeptide (like BtsI, BsrI, Mva1269I, BtsCI, known as class II among the type IIS REs)^{14,20}, we designed a bottom-strand inactivated version called Nt^{*}.BsmI and a fully inactivated version of BsmI, hereafter called N0.BsmI. For each domain, we mutated the glutamic acid of the characteristic catalytic motif PD-X_n-(D/E)-X-K (E109V, for the bottom-strand cleaving domain, and E546V, for the top-strand one).

We first obtained crystals with the fully inactivated N0.BsmI in complex with its cognate DNA in the presence of Mg²⁺ ions. Attempts to resolve the structure by molecular replacement were unsuccessful. The structure was solved through experimental phasing of Eu³⁺-soaked crystals, utilizing the anomalous signal from the phosphorus atoms in the DNA backbone and from the six bound europium ions, of which five were located at the external surface of the complex and one near the top-strand nicking active site. The initial electron density map had a resolution close to 2.2 Å, allowing for the construction of a complete model. The asymmetric unit contained a single monomer of BsmI and one DNA molecule tightly bound inside a central cleft, suggesting that BsmI might be monomeric in solution, similar to its isoschizomer Mva1269I¹⁵. This hypothesis was confirmed by ultracentrifugation experiments, which indicated that, in the absence of DNA, BsmI sediments as a protein of 76.8 kDa (4.6 S), while in presence of its cognate DNA, it sediments as a DNA:protein complex of 86.5 kDa (5.3 S) (Figure S1). Although traces of higher-order oligomeric states could be

detected (less than 1% of the total mass), it is clear that BsmI is predominantly monomeric in solution, regardless of the presence of DNA. The theoretical sedimentation coefficients calculated from the obtained crystal structure were found to be 4.8 S in the apo-form and 5.3 S for the complex, which are close to the experimental values. These results show that the enzyme is in monomeric state both in the crystal and in solution.

Unique combination of remotely related folds for the three different subdomains

Structure analysis revealed, in accordance with previous sequence-based findings on its closely related isoschizomer Mva1269I¹⁵, that BsmI can be subdivided into three domains: (i) the N-terminal domain (in green in Fig. 1 and later) presenting a 8-strand mixed beta sheet flanked by alpha helices, (ii) a central domain (in blue) containing a smaller beta sheet, and (iii) the C-terminal domain (in orange) harboring a 10-strands mixed beta sheet flanked by alpha helices. The overall topology of the N- and C-terminal domains is similar to the invariant common structural core present in all members of the type II RE family (Figure S2). The N-terminal domain contains a catalytic motif that points toward the bottom strand, while the catalytic motif within the C-terminal domain is positioned to interact with the top strand at the level of its cleavage site. The central domain, a priori not directly involved in catalysis, will be called "Central domain" hereafter.

A global (whole) protein structure comparison search on the PDB using the DALI web-server²¹ did not find any high-confidence hits for the entire protein. However, applying the same process to the separated domains yielded interesting results for the top- and bottom-strand nicking domains. Among the proteins with a Z-score greater than 5, we identified the type II RE AgeI (PDB ID 5dwb), BsaWI (PDB ID 4zsf), NgoMIV (PDB ID 4abt) and EcoRI (PDB ID 1ckq) for the bottom-strand nicking domain; and SdaI (PDB ID 2ixs), PfoI (PDB ID 6eko), N.BsrD6I (PDB ID 5liq) for the top-strand nicking domain. Additionally, structural homologies of the top-strand nicking domain with FokI (PDB ID 1fok) and of the bottom-strand nicking domain with EcoRI (PDB ID 1ckq) are clearly visible on the views of the Fig. 2, as predicted by Armalyte et al.¹⁵.

Structure of the fully active wt.BsmI and the position of the divalent metal ions

Crystals of fully active BsmI, called hereafter wt.BsmI, with its cognate DNA could only be grown in the presence of Ca²⁺ and exhibited interesting differences compared to the N0.BsmI structure. First, the DNA is more deeply buried in the DNA-binding cleft and the top-strand nicking domain is closer to the DNA. This conformational change resulted in the loss of one symmetry axis in the crystallographic space group, transitioning from P2₁2₁2 (observed in all other obtained crystals) to P2₁. Secondly, a notable segment (residues from 484 to 503), which includes a loop and a β-strand, is shifted so that it is threaded in the middle of the large beta sheet of the bottom-strand nicking domain. In the structure of the wt.BsmI, the loop is positioned close to the DNA minor groove, insulating the two cleavage sites from one another; in the following, we refer to it as the *insulator loop*. In the structures of the inactive mutants, this loop retracts to form a short beta strand, while its N-terminal part forms a loop which is mainly disordered. The sequence of insulator loop is highly conserved, as shown by the multi-alignment of the top 100 sequences obtained from a BlastP search (see below, Figure S3). Intriguingly, among the mutations introduced to create the Nb.BsmI nicking enzyme, two of them are located near the insulator loop (Gly507Asp and Arg509Val). We illustrate this conformational change with a movie (Supplementary Movie SM1), with the caveat that it should not necessarily be taken as an actual movement during the catalytic cycle, but as a dynamic illustration of the unusual type of change of conformation of this long peptide fragment.

Importantly, the wt.BsmI structure reveals the presence of three calcium ions in the two active sites. Two of these ions are located in the bottom-strand nicking active site, coordinated by Asp9, Asp91 and Glu109, while one is found in the top-strand nicking active site, coordinated by Glu470 and Asp528. These ions occupy the same positions predicted by superimposing

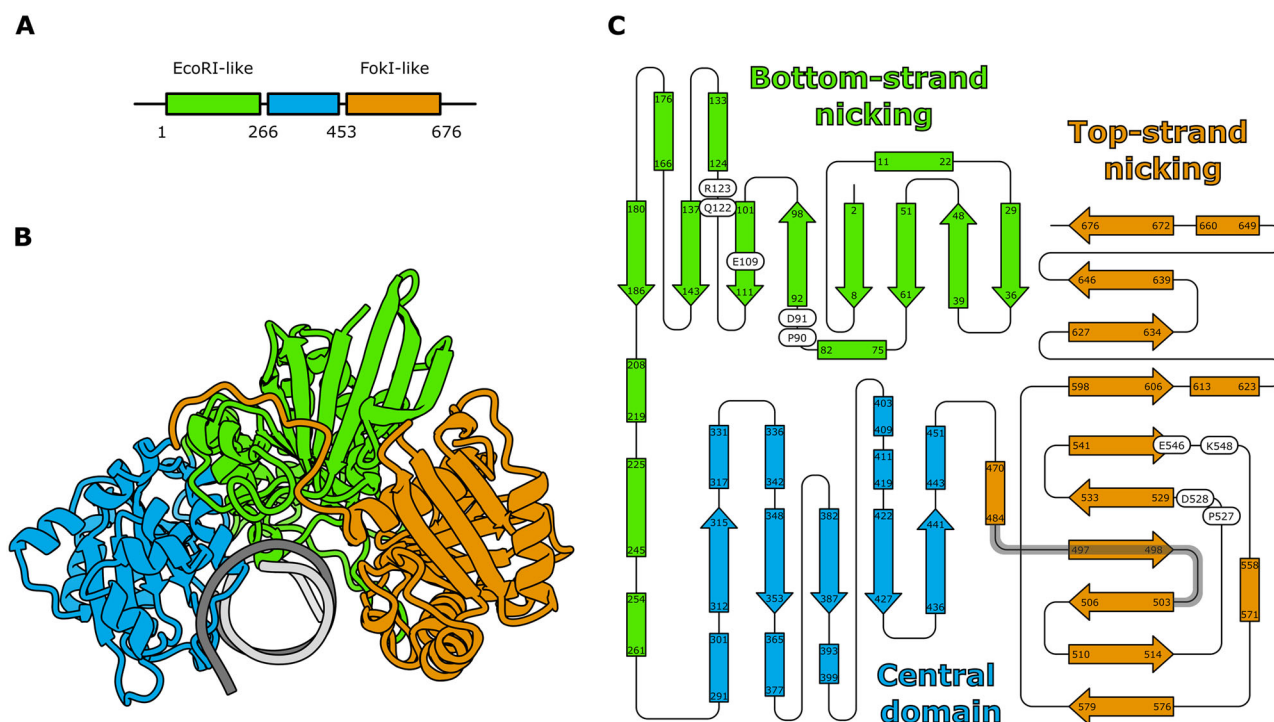


Fig. 1 | Structure of BsmI. **A** Tertiary structure of the enzyme in complex with its cognate DNA. The N-terminal contains the first mixed β -sheet flanked by α -helices and carrying the bottom-strand nicking catalytic residues. The primary and tertiary structure presents similarities with the EcoRI catalytic domain. The C-terminal contains the other mixed β -sheet flanked by α -helices and carrying the top-strand nicking catalytic residues. The primary and tertiary structure presents similarities with the FokI catalytic domain. The central domain contains a smaller mixed β -sheet surrounded by smaller α -helices. The tertiary structure presents no detectable global similarity with known structures in the PDB. **B** Delimitation of the three domains.

The green domain corresponds to the EcoRI-like domain, also called bottom-strand nicking domain, the blue domain corresponds to the central domain and the orange domain corresponds to the FokI-like domain also called top-strand nicking domain. The dark grey chain represents the DNA bottom strand, the light grey chain the DNA top strand. **C** Secondary structures of the enzyme. The residue numbers delimit the β -strands and the α -helices. The catalytic motifs residues are indicated with white rounded rectangles. The grey band on the fourth β -strand of the top-strand nicking domain signals the position of the retractable “insulator loop”.

the BamHI structure (which contains two divalent cations) onto each of the BsmI active sites (Fig. 3). Crystals of N0.BsmI could also be grown in the presence of Ca^{2+} ions and showed virtually no differences compared to crystals grown with Mg^{2+} , except that the electron density more clearly showed the two divalent cations, which correspond to those seen in wt.BsmI (one in each active site).

Structures of Nt*.BsmI and Nb.BsmI mutants give an insight on the cleavage mechanism

We also solved the structure of both Nt*.BsmI and Nb.BsmI mutants in complex with their cognate double-stranded DNA substrate in the presence of calcium or magnesium ions, respectively. Nb.BsmI in the presence of Ca^{2+} ions did not crystallize, but the presence of Mg^{2+} allows to obtain crystals with DNA. However, the substrate appears to be cleaved in the Nb.BsmI crystal. In contrast, Nt*.BsmI crystals obtained with Ca^{2+} and subsequently soaked in Mg^{2+} solutions were stable and did not show any cleavage of the DNA, consistent with an ordered reaction in which the nick in the bottom strand is a necessary step before proceeding toward the second cut. We do not observe any divalent ion in the bottom-strand nicking active sites of Nb.BsmI and Nt*.BsmI, and we see only one divalent ion (Calcium) in the top-strand nicking active site of Nt*.BsmI but none present in the Nb.BsmI site.

The structure of Nb.BsmI can be described as a snapshot of the post-reactive complex, following the cleavage of the phosphodiester bond in the bottom strand. Interestingly, despite the cut in the DNA, we do not see any conformational change compared to the inactive N0.BsmI. We observe the expected 5'-Phosphate and the 3'-OH moieties after the reaction, but the limited resolution (2.94 Å) does not provide direct insight into the presence and coordination of magnesium ions. The well-defined structure in the

electron density of the Nb.BsmI:DNA complex after cleavage of the bottom strand in the crystal and its high similarity to the other structures obtained with uncleaved DNA suggests that this intermediate step in the reaction is at least partially stable and not a transient species.

In all the resolved structures of the BsmI mutants, including the Nt*.BsmI, the Nb.BsmI and the N0.BsmI presented in this article, the Insulator loop is in a retracted position, away from the DNA. Figure 4A provides a detailed comparison of the relative positions of this loop in the Nt*.BsmI and Wt.BsmI structures. In Wt.BsmI the closure of the Insulator loop brings it in contact with the DNA minor groove. A more general view is presented in **Movie SM2**.

DNA recognition mechanism

Comparison of the cognate DNA with standard B-DNA of the same sequence reveals that the cleavage sites are located in a slightly distorted DNA helix (Fig. 4). This distortion is characterized by an opening of the minor groove, which can be directly attributed to the closure of the Insulator loop. The two nicking domains together with the Central domain form a cleft that presents a large contact surface with DNA. Their respective interaction surface with the DNA is 1053.9, 1112.6 and 445.4 Å². This suggests that the Central domain may have an important role in the sequence recognition specificity. As depicted in the Fig. 5A, the cleft surface facing the DNA is mainly positively charged, which is expected for a Protein:DNA binding interface. A thorough study of the refined models, including ions and water molecules, allowed us to enumerate all possible direct and water-mediated residues:DNA interactions. In Figure S4, we present a schematic diagram of these interactions, listing all the residues involved in one or more bond(s) with the DNA or with a DNA-interacting water molecule. The absence of interactions from residues in the top-strand

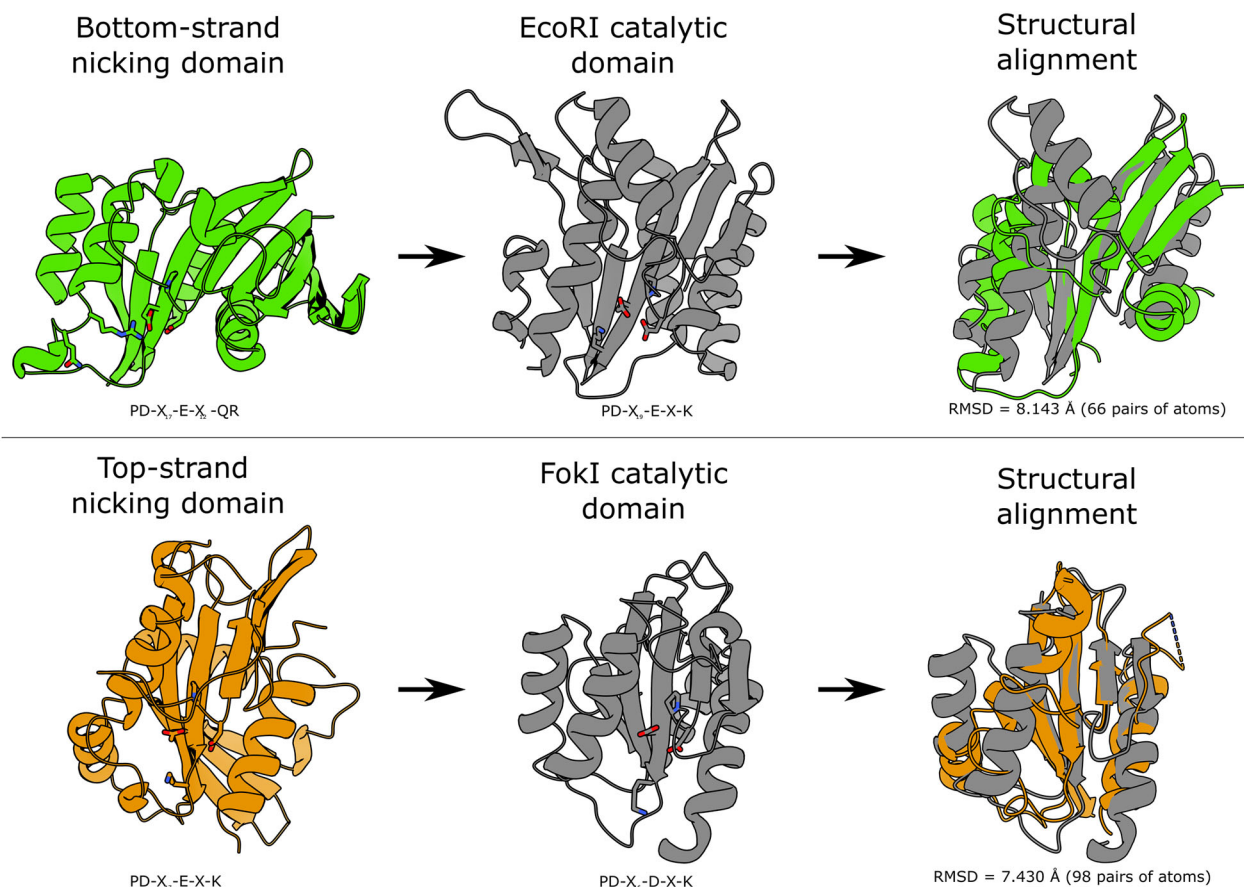


Fig. 2 | Structural similarities of the top- and bottom-strand nicking domains of BsmI (left) with FokI and EcoRI catalytic domains (right). The catalytic residues are shown in stick representation. The sequence motifs of the catalytic site are shown at the bottom. For clarity, the structural alignments were only performed between the residues [1-25, 74-142, 164-185] of BsmI (bottom strand nicking domain) and

the residues [35-114, 141-170, 201-210, 235-239] of EcoRI on the one hand, and between the residues [469-499, 520-571, 597-633] of BsmI (top strand nicking domain) and the residues [417-496, 514-547] of FokI on the other hand. The RMSD was obtained with the MatchMaker tool of ChimeraX⁴⁹.

nickering domain suggests that the complex is in a conformation compatible with bottom strand cleavage, but not with top strand cleavage. This is in line with the functional tests presented below, which indicate that cleavage of the bottom strand occurs first and is an obligatory step prior to cleavage of the top strand. An important aspect of the interactions involves the phosphate group and, occasionally the ribose moiety of the nucleotides. While these interactions do not confer specificity for the canonical recognition site, they do help stabilize the DNA interactions.

All the nucleobases accommodated in the double-stranded recognition site participate in binding, sometimes redundantly, with residue side chains or backbone atoms important for the high site-specificity of BsmI. Interestingly, the first two nucleobases of the recognition site (G and A) appear to have fewer contacts than the other nucleobases, interacting only through water molecules in the Nt^{*}.BsmI enzyme, whereas more contacts are observed in the wild-type complex. This suggests that the constraints on these two nucleobases could be more easily modified than the others, in future protein engineering experiments. Despite the important change in DNA position inside the cleft, the comparative map of Protein:DNA interactions presented on Figures S6A and S6B shows that in both conformations, the interaction is supported by the bottom-strand nicking domain and the central domain. A notable increase in the number of residues contacting the recognition site is observed in the wt.BsmI map. The loss of the water-mediated interactions can be explained by the lower resolution of the wt.BsmI electron density map, which did not allow for the placement of as many water molecules as the Nt^{*}.BsmI structure. Non crystallographic symmetry constraints were applied during refinement, and only a few

differences between the models of the two molecules in the asymmetric unit can be observed (Figures S6B and S6C).

To explain how BsmI might be impaired by methylation of the DNA, we examined close contacts between adenine (N6) or cytosine (N4) methylation sites on the DNA and the side chains of the protein. We found that the 2 central adenines (A6 on top strand and A7 on bottom strand) are recognized by the side chains of residues N118, Q119 and Q122 (Figure S4B-C), which are part of a highly conserved patch of residues (GNQAWQR in Figure S3).

Without divalent ions, AlphaFold3²², the last version of AlphaFold which can run predictions for the structures of complexes with DNA, ions and ligands, could not predict any complex between BsmI and its cognate DNA sequence. When divalent ions were included, the ions were correctly predicted but the DNA was positioned in the wrong orientation and shifted compared to the experimental model described here, suggesting that AlphaFold3 fails to accurately capture the essence of the interaction specificity between BsmI and its substrate (Figure S5).

Conserved motifs along the sequence and functional role of key residues

A BlastP search using the BsmI protein sequence as template allowed us to identify 184 closely related homologues with a coverage greater than 80%. These sequences were aligned using the MAFFT webserver. The multi-alignment (Figure S6) was processed with the Consurf program to highlight the pattern of conserved residues in a three-dimensional context^{23,24}. The output is presented on Fig. 5, with the protein surface colored according to

Fig. 3 | Nt.BsmI vs BsmI active sites comparison and active sites alignment of BsmI and BamHI.

The Nt.BsmI structure was solved at a resolution of 1.9 Å, whereas the BsmI structure was solved at a resolution of 2.85 Å. The calcium ions are represented as pink balls. The electron density was contoured at 1 sigma. **A, B** Details of the bottom-strand nicking catalytic domain of Nt.BsmI and BsmI crystallized in presence of calcium. Two calcium ions are present in the active site of the BsmI active enzyme whereas none are found in the active site of the inactive Nt.BsmI. **C, D** Details of the top-strand nicking catalytic domain. Only one calcium ion is present in both structures. **E** BsmI Bottom-strand nicking catalytic domain (green) structurally aligned with BamHI catalytic domain (light brown). **F** BsmI Top-strand nicking catalytic domain (orange) structurally aligned with the same BamHI catalytic domain. The blue spheres represent the calcium ions from BamHI pre-reactive structure (PDB id. 2BAM). The pink spheres represent the calcium ions found in wt.BsmI structure (PDB id. 9F38). The DNA represented in grey belongs to the BamHI structure.

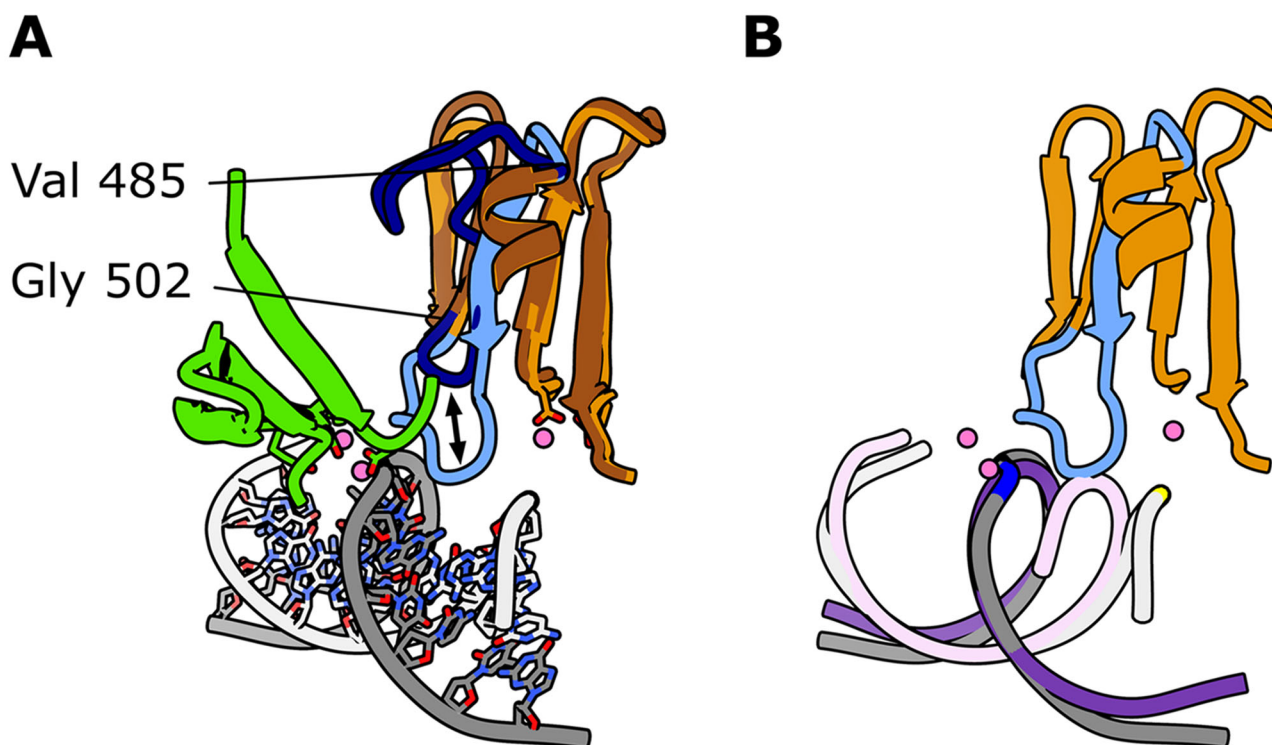
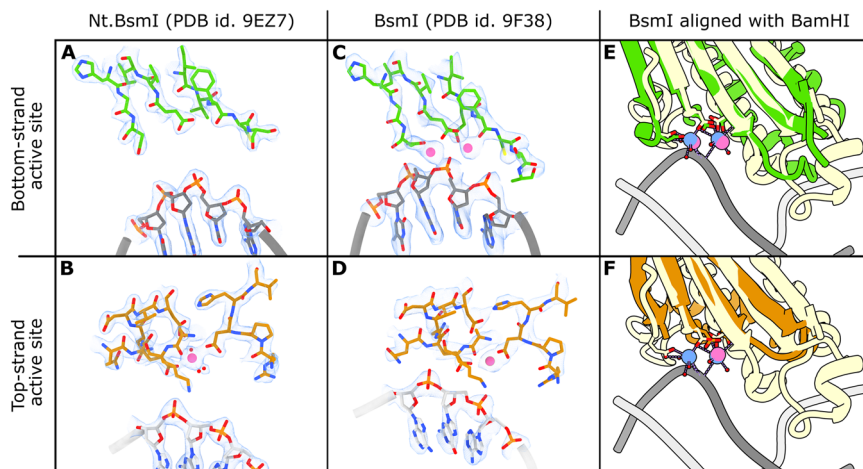


Fig. 4 | Insulator loop and DNA recognition. **A** Comparison of the different positions of the Insulator loop. The wt.BsmI structure is represented in light colors, while the Nt*.BsmI one (similar to all the other mutants) is represented in darker colors. The Insulator loop is depicted in blue and its retraction movement away from the DNA is materialized by a double arrow in black. The pink spheres correspond to the calcium ions found in the wt.BsmI structure and their known glutamate ligands are in ball-and-stick representation. The DNA backbone is in grey. **B** Comparison of the

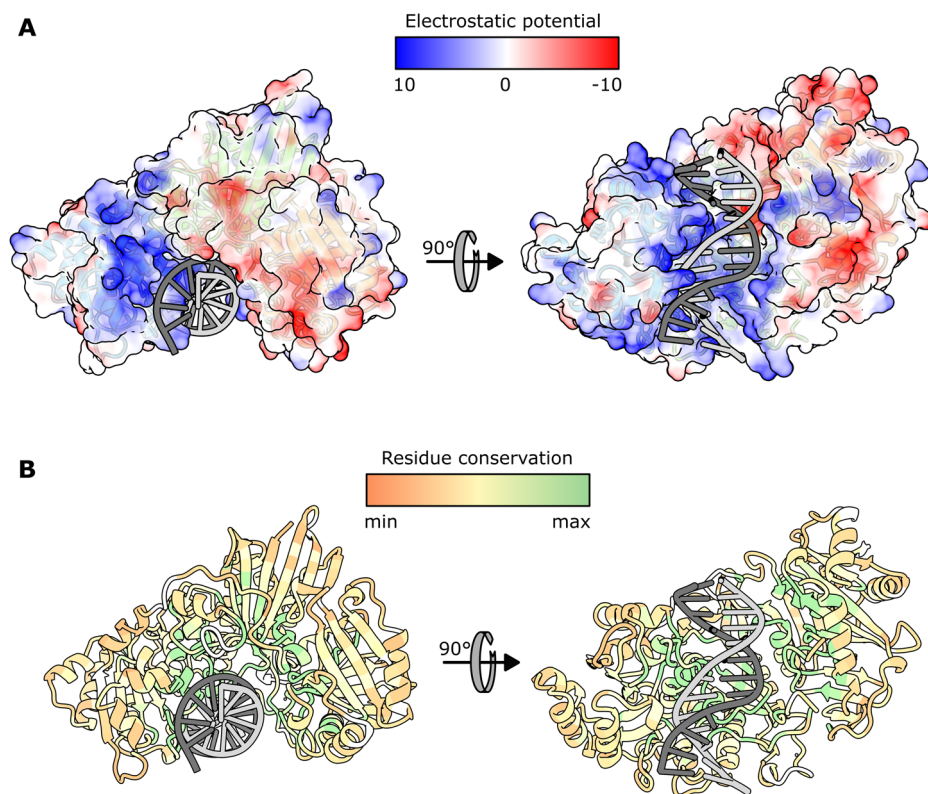
DNA conformation with B-DNA. The Insulator loop of the wt.BsmI is represented in light blue in the context of the beta-sheet of the top-strand nicking domain (in light brown). The blue and yellow patches on the DNA backbone illustrate the position of the scissile phosphodiester bonds of the bottom and the top strand respectively. The light and dark purple DNA backbone corresponds respectively to the top and bottom strands of the cognate DNA in the ideal B-form. The Insulator loop opens up the minor groove of the DNA to bring into position the second nicking site.

the calculated conservation levels from the multi-alignment. The multi-alignment itself reveals patches of highly conserved residues all along the BsmI reference sequence, while this representation shows how the conserved residues cluster at the surface of the cleft. All the residues involved in

the DNA sequence recognition, as depicted in Figure S4, as well as those in the catalytic motifs, are highly conserved, as expected.

Figure S3 presents a Logo representation of the conserved motifs in the MSA, highlighting highly conserved regions in the alignment. We selected

Fig. 5 | Electrostatic potential and residue conservation. **A** Coulomb electrostatic potential (ESP) mapped to the molecular surface of BsmI. The color coding is shown on top, in units of kT/e. The blue surfaces have a positive electrostatic potential and the red ones have a negative electrostatic potential. On the left, the view of the same orientation as in Fig. 1B. On the right, the view of the DNA-binding cleft. The electrostatic potential was calculated with the 'coulombic' function of ChimeraX (v1.5)⁴⁹ using default options (dielectric constant $\epsilon = 4.0$ inside the protein, $\epsilon = 78.0$ outside the protein temperature $T = 298$ K). **B** Residue conservation across the 100 first results of BlastP using the Consurf web server^{23,24}. The color scale is shown on top.



the eight most representative motifs and used the structure of BsmI to infer their function. Among 8 conserved motifs, two correspond to the expected catalytic motifs. While it is likely that this subfamily of monomeric RE comes from the fusion of two nucleases¹⁰, it is not known if the marginal PD-Xn-E-Xm-QR catalytic motif, which is rather unusual in the type II RE family compared to the PD-E-X-K motif, predates the fusion or not. Two other motifs correspond to an α -helix carrying residues coordinating metal ions in each active site, participating to the correct positioning of the catalytic ions inside the active site. Three motifs correspond to loops contacting either the major or the minor groove of the DNA, and are thus suspected to confer its DNA sequence specificity to the enzyme. The last region is the "Insulator loop" that appears to be mobile and to intercalate between the two catalytic sites in the active version of the enzyme. Its putative role will be commented in the discussion section.

Mva1269I, the isoschizomer of BsmI, (GenBankTM accession number AAY97906.1) ranks #181 (in term of sequence identity) among the 184 results of a BlastP run against the BsmI sequence with a coverage larger than 80% (see Table S1). The functional properties of Mva1269I have been described in ref. 15. A comparison of the two restriction enzymes sequences revealed that they share only 30.2% sequence identity, despite their functional equivalence and probable similarities in terms of mechanism (both are monomeric and carry both two catalytic sites in a single polypeptide). AlphaFold2 (<https://alphafold.ebi.ac.uk/>) was used to generate a structural model for the protein alone with a very high confidence score all along the chain. A structural alignment by ChimeraX with BsmI revealed an overall RMSD of 3.86 Å with 644 residues (Ca aligned), and 1.275 Å for 388 pruned residues (Figure S7).

Activity tests on BsmI mutants reveal a sequential and concerted cleavage mechanism

To study the mechanism leading to the cleavage of the two DNA strands, we mutated a conserved glutamate in the top-strand nicking site (E546V – Fig. 6A), or in the bottom-strand nicking site (E109V – Fig. 6B), or in both, to independently inactivate the cleavage of each of the two DNA strands. We

call these mutants Nb.BsmI (bottom strand nicking version), the Nt*.BsmI (bottom-cleavage-conditioned top strand nicking version) and the N0.BsmI (fully inactivated version).

The activity of the different versions of BsmI was assessed using a doubly labeled fluorescent DNA probe containing the BsmI recognition site. As depicted in Fig. 6C, the top strand was 5'-FAM-labeled and the bottom strand was 5'-TAMRA-labeled. When cleaved, each labeled strand is shortened, which can be visualized by electrophoresis on a polyacrylamide gel.

As expected, wt.BsmI cleaves the two strands, Nb.BsmI cleaves only the bottom strand, and N0.BsmI leaves the two strands uncut. However, Nt*.BsmI is also completely inactive (Fig. 6D). These results support the hypothesis that the two strands are cleaved in an ordered fashion. Indeed, it appears that the bottom strand must be cut first for the enzyme to be able to cleave the top strand. To verify this hypothesis, we tested the activity of the four versions of the enzyme with probes containing a nick on either the top or bottom strand (Fig. 6E, F). Indeed, Nt*.BsmI was able to cleave the top strand of the bottom-nicked probe, consistent with the proposed two-step mechanism. Interestingly, this experiment also shows that both nicked probes can be cleaved by BsmI and that the top-nicked probe can be cleaved by Nb.BsmI.

The presence of extra bands on the gel prompted us to explore the cleavage mechanism and the migration process in more detail. These bands could result from either star activity of the enzymes or electrophoresis artifacts. Consequently, we tested the different mutants in commercial buffers where BsmI and Nb.BsmI are known to be highly specific (rCuts-mart and r3.1 buffer from New England Biolabs, respectively) (Figure S8A). Since the results in the r2.1 buffer were completely similar to those obtained with the recommended buffers, we examined the migration profile of an oligonucleotide mixture mimicking the double-strand cleavage of the probe. We observed the exact same extra bands as for BsmI cleavage (Figure S8B-C), showing that the presence of the complementary strand influences the migration despite the harsh denaturing conditions in the loading dye and in the gel.

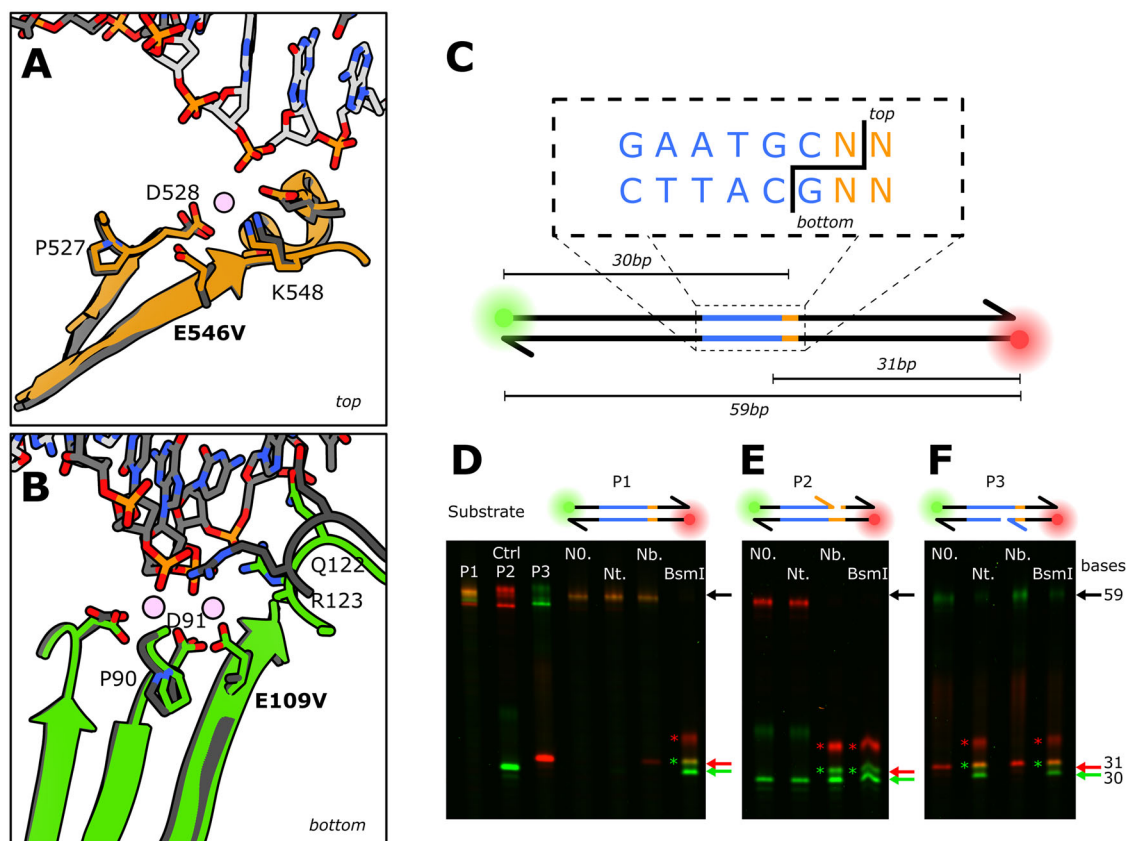


Fig. 6 | Effect of each catalytic sites inactivation on cleavage activity. **A** The top-strand nicking catalytic motif of the active version of BsmI is represented in orange while the one of the superimposed inactivated BsmI is in dark grey. The pink sphere represents a calcium ion. **B** The bottom-strand nicking catalytic motif of the active version of BsmI is represented in green while the one of the inactivated BsmI is superimposed in dark grey. **C** Design of the probe used to test enzyme activity. The strand presenting the top-strand recognition sequence is 5'-labeled with a FAM fluorophore, the bottom-strand one with a ROX fluorophore. When uncut, each strand migrates like a 59bp-long single-stranded DNA oligonucleotide. Once cut, the top strand is 30bp-long and the bottom strand is 31-bp long. **D** Cleavage test on a classical probe (no initial cut). The P1 control corresponds to the probe in absence of

enzyme. The P2 control is a mixture of the 59bp-long bottom strand and of the 30bp-long top strand. The P3 control is a mixture of the 59bp-long top strand and of the 31bp-long bottom strand. The probe was incubated 10 min at 65 °C in presence of 15 µg/mL of N0.BsmI, Nt*.BsmI, Nb.BsmI or wt.BsmI, respectively indicated by the staggered “N0.”, “Nt.”, “Nb.” and “BsmI” annotations at the top of the lanes. The green arrow symbolizes the expected position of the cut top strand and the red one indicates the expected position of the cut bottom strand. **E, F** The probes presenting already a nick on the bottom or on the top strand were treated similarly to the classical one. The green and red asterisks on the left side of the bands mark the position of extra-bands due to electrophoretic artifacts. The denatured strands were migrated by electrophoresis through a 16% polyacrylamide-8M urea gel.

In all structures, the bottom-strand nicking domain is closer to its DNA substrate than the top-strand domain, and its active site is well positioned for the first reaction to occur. When divalent ions are visible in the electron density map (with Ca^{2+} ions), the catalytic residues in the bottom-strand nicking site coordinate a divalent cation that is well positioned to trigger the phosphodiester bond hydrolysis. In contrast, in the top-strand nicking site, the divalent cation appears to be too far from the DNA.

Stereospecificity for the phosphorothioate bond cleavage

In order to shed light into the cleavage mechanism, we substituted the scissile phosphodiester bond with a phosphorothioate bond (Fig. 7A, C) in either one of its two diastereoisomers^{25–28}. We separated the two enantiomers of a 13-base-long oligonucleotide by FPLC using a pellicular anion exchange column and a NaCl gradient (Fig. 7B). The control single-stranded DNA was eluted in a single peak, while the phosphorothioated oligonucleotide produced two distinct peaks corresponding to the R_p and the S_p forms of the bond. It is known that the first peak corresponds to the R_p and the second to the S_p enantiomer²⁹.

The results of the test of the resistance to cleavage of the separated oligonucleotides containing a phosphorothioate bond at the scissile sites were analyzed after migration on a polyacrylamide gel. To reconstitute the same type of probes as previously, we treated the two previous strands

independently by annealing and then ligating the separated oligonucleotides with a 5'-labeled strand and a downstream strand to obtain the four desired probes (top-strand R_p and S_p , bottom-strand R_p and S_p). As shown in Fig. 7D, E, BsmI cleaves the control strands and the R_p enantiomer but cannot cut either the top-strand or the bottom-strand S_p . Interestingly, these results confirm the proposed two-step mechanism. When the bottom-strand contains the S_p phosphorothioate, both strands remain uncut. When the protection is on the top-strand, the bottom strand is cleaved while the top strand is not, consistent with the previous conclusion that the bottom strand must be cleaved before the top strand can be cut.

Normal mode analysis and the top-strand cleavage mechanism

A crystal structure does not necessarily account for the entirety of the conformations adopted by the protein along its catalytic cycle. Here, the observed conformation of the top-strand nicking active site is not compatible with a catalytic act, indicating that a conformational change is necessary for the enzyme to effectively cleave this bond. To visualize the collective degrees of freedom that allow the protein to rearrange its domains, we performed normal mode analysis using NOLB software³⁰. The lowest-frequency mode, which is considered the most relevant at a given temperature (i.e., with the highest amplitude), clearly shows the possibility for the top-strand nicking domain to move closer to the DNA (Supplementary

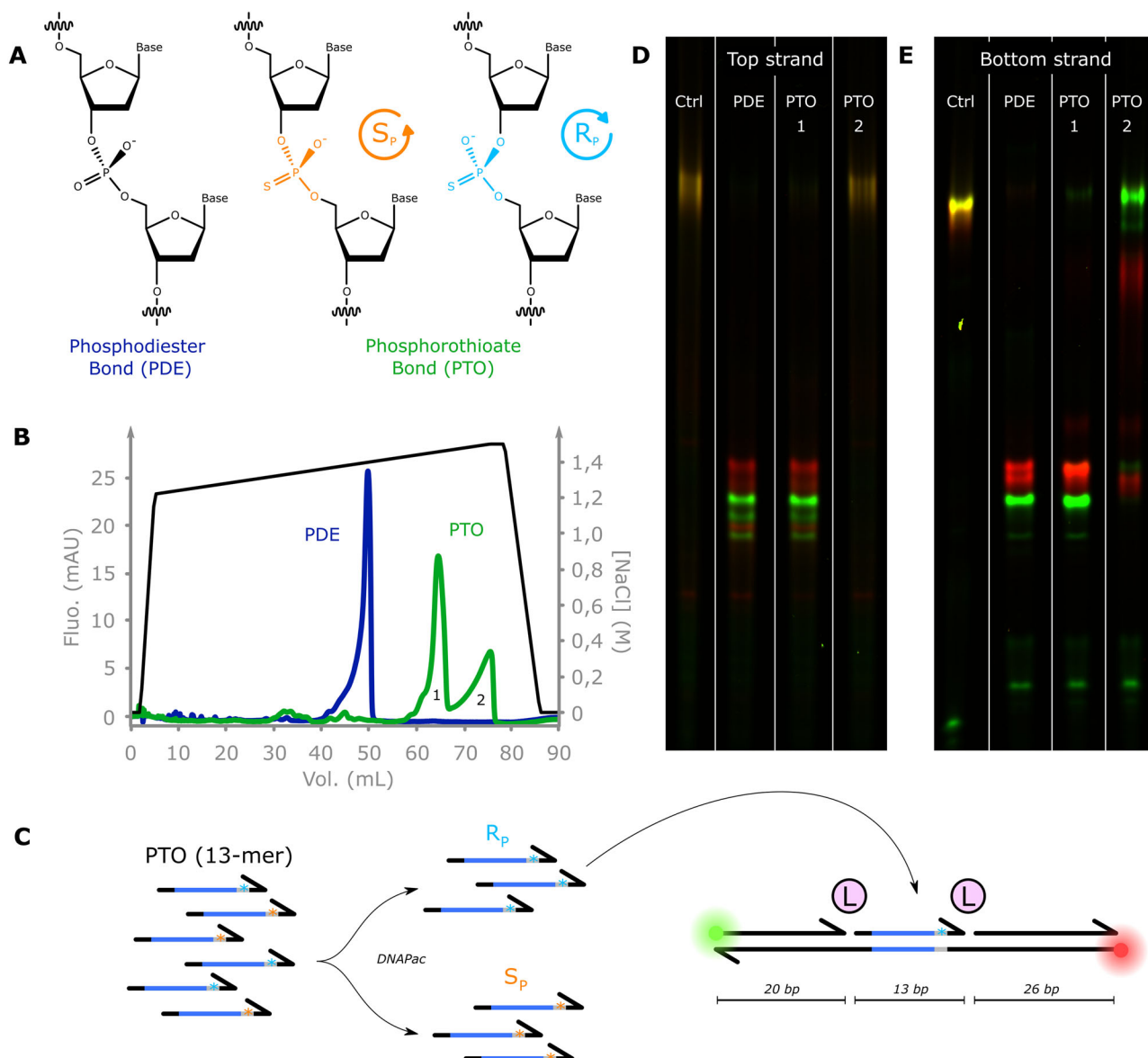


Fig. 7 | Effect of a phosphorothioate bond on cleavage activity. **A** Conformation of the phosphodiester bond and of the two diastereoisomers of the phosphorothioate bonds. **B** Chromatograms of the DNA Pac FPLC of the top strand 13-mer with phosphodiester bonds (PDE) only (in blue) and of the 13-mer presenting a phosphorothioate bond (PTO) instead of the scissile PDE bond (in green). The black curve represents the NaCl gradient applied during the enantiomers separation. The first peak is the R_P and the second the S_P enantiomer. **C** Scheme of the enantiopure

probe preparation. The oligonucleotides carrying a phosphorothioate bond, purchased from an industrial manufacturer, are a mixture of the two enantiomers. They were separated by HPLC, desalted, concentrated and then ligated to other oligonucleotides as indicated on the right to form the probe. **D**, **E** The control (Ctrl) represents the probe incubated without any enzyme. The probes obtained by ligation of the top- and bottom-strand 13-mers (PDE, PTO- R_P , PTO- S_P) were incubated with BsmI at 15 µg/mL at 65 °C for 10 min.

Movie SM2). Figure 8 illustrates the enzyme undergoing a large conformational change along the first mode, compared with the crystallographic structure. Beyond the global movement of the domain, which now aligns the top-strand scissile phosphodiester bond in the two models, the catalytic residues get closer to the DNA, suggesting that this movement could account for the top-strand cleavage mechanism. On the other hand, the normal mode analysis reveals significant DNA deformation that may be necessary to enable the movement of the protein domains. Indeed, this deformation could occur only if the bottom strand is cleaved, which explains the enzyme's inability to cut the top strand independently of the bottom strand's state.

Discussion

Here, we present the first high-resolution structure of a monomeric type IIS restriction enzyme (RE) with two active sites. To obtain the structure of

wt.BsmI bound to its cognate DNA, Ca^{2+} ions were used instead of Mg^{2+} , as they are catalytically inert but effectively mimic Mg^{2+} ions, at least in terms of their specific octahedral coordination geometry. This strategy of replacing Magnesium by Calcium ions has frequently been employed in structural studies of nucleic acids processing enzymes, such as DNA polymerases, for instance from the RB69 phage³¹. When possible, this is supplemented by structural studies of catalytically inert mutants in the presence of Mn^{2+} or Mg^{2+} ³². This approach was also adopted here, and we must be cautious in interpreting the separate effects of introduced mutations in relation to the effects of the change of divalent ions. See also^{33–36}. Here, we can assess whether the Ca^{2+} binding sites serve as a suitable model for Mg^{2+} ions by comparing them with known structures of the same family. BsmI, as a monomeric enzyme carrying two distinct active sites, belongs to a class of RE that can be engineered to yield nicking endonuclease(s) in each of the two DNA strands, by mutating glutamate residues in their catalytic sites^{17,20}.

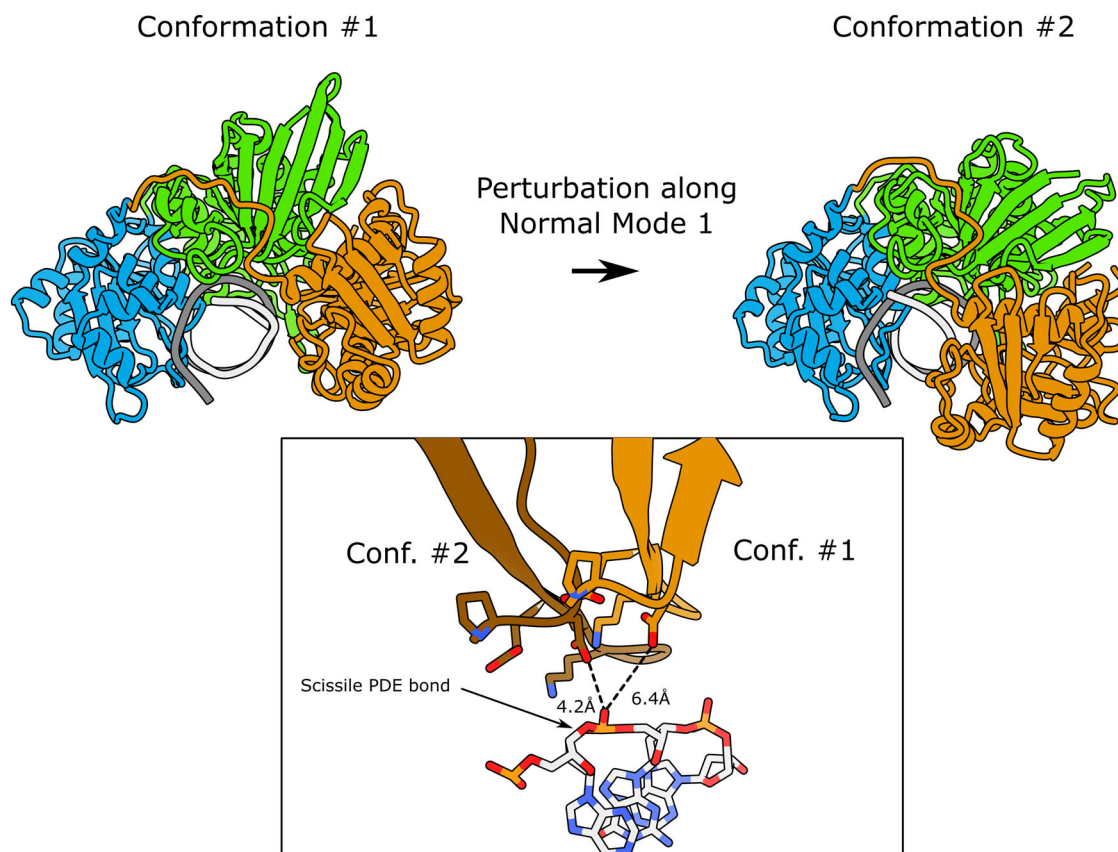


Fig. 8 | Effect of Normal mode perturbation on BsmI in complex with DNA. The upper panel shows the initial (#1) and final (#2) positions adopted by the domains following a deformation along the first normal mode with default value amplitude. The lower panel represents an alignment along the DNA molecules of the two models. The lighter chain corresponds to the initial conformation (the one observed

in the crystal structure), the darker chain corresponds to the positions of the catalytic motif residues after deformation along the first normal mode. Catalytic residues are shown in ball-and-stick representation. The distance between the catalytic glutamate and the scissile phosphodiester bond decreases from 6.4 to 4.2 Å. Normal Modes were calculated with NOLB³⁰.

BsmI has actually a commercially available bottom nicking version known as Nb.BsmI, obtained by inactivating the top-strand nicking active site. Following the same principle, we demonstrated the inactivation of the bottom-strand nicking active site and produced mutants with one or both the catalytic sites inactivated. They were crystallized to structurally study the catalytic mechanism.

In the structure of wt.BsmI in presence of calcium, we observe two metal ions in the bottom-strand cleaving active site, and only one in the top-strand cleaving active site (Fig. 3C, D). The alignment of the two beta-strands carrying the catalytic residues with the ones of BamHI for the bottom-strand cleaving domain shows striking similarities in the positions of the metal ions and their ligands (Fig. 3E). In contrast, the alignment for the top-strand cleaving domain shows that one of the two sites found in BamHI is occupied by a Ca^{2+} ion, while the other remains empty (Fig. 3F). Remarkably, the missing metal ion corresponds to the one that is lost after cleavage in BamHI³⁷. Since we know that the bottom strand must be cleaved to enable top-strand cleaving, we can confidently hypothesize that the cleavage of the bottom strand relaxes conformational constraints, allowing the top-strand cleaving domain-DNA complex to adopt an active conformation and to acquire the missing divalent cation. The post-reactive loss of metal ion in BamHI may indicate that the corresponding unoccupied site in BsmI becomes populated only during catalysis.

AlphaFold 3 succeeded in predicting the structure of the apo-enzyme accurately and in finding the correct position of the catalytic ions in one of the 5 models generated with the default parameters but could not place correctly the cognate DNA, in fact it is modeled upside-down. It is perhaps

appropriate here to alert the community of molecular biologists interested in structures or protein engineers interested in the RE specificity for DNA that the predictions of AlphaFold3 are not necessarily correct, even if its quality indicators all have a green light. It appears that crystal structures are not really dispensable at this stage, if one is interested in the atomic details of RE-DNA interactions.

Incidentally, the crystals of wt.BsmI, if optimized, could also serve as a valuable source for future detailed kinetic studies by time-resolved crystallography. Indeed, we know that the initial and ending points of the reaction can be accommodated within the same crystal packing environment. If the Ca^{2+} ions are carefully exchanged for Mg^{2+} or even Mn^{2+} ions, it might be possible to identify the different steps of at least the first nicking reaction at the atomic scale.

For the recognition of the cognate DNA sequence, we find that the side chains of highly conserved residues are involved in recognition of the 3 central adenines at atomic positions where steric clashes would occur upon N6-methylation, as expected. In general, there is a very good correspondence between highly conserved motifs and functional predictions based on structural analysis (See Figure S3). BsmI, like Mva1269I, cuts the bottom strand first before cleaving the top strand, most likely in a sequential and concerted mechanism during a single-binding event. However, it is also possible to observe top-strand cleavage by Nt*.BsmI and Wt.BsmI when provided with an artificially bottom-cleaved substrate. In the Fig. 5A, we clearly observe that the central and bottom-strand nicking domains of the DNA-binding cleft exhibit a positive electrostatic potential. In contrast, part of the top-strand nicking domain in this cleft is negatively charged. We can therefore hypothesize that the DNA is initially more attracted toward the bottom-strand nicking active site. The normal mode analysis performed on

the complex suggests that the enzyme is able, after the cut, to dynamically bring the top-strand-nicking active site closer to the DNA. However, this movement also requires a significant conformational change in the DNA itself. It seems reasonable to propose that nicking the bottom strand relaxes the conformational constraints in the DNA, thereby allowing a productive positioning of the top-strand scissile phosphodiester bond for cleavage by the top-strand nicking active site.

Monomeric endonucleases face a specific structural and functional challenge: they must maintain two active sites composed of three glutamates or aspartates each, located only a few angstroms apart to sequentially cleave both DNA strands. BsmI appears to have evolved to achieve cleavage in an ordered manner, with bottom-strand cleavage serving as a prerequisite for top-strand cleavage. This implies that the top-strand active site is not fully positioned with respect to DNA, and the divalent ions not fully in place, until

the bottom strand is cleaved. This presents a significant electrostatic challenge, as the glutamates in the active sites are only partially counterbalanced by the divalent ions. The structure suggests that a solution may lie in an essential loop that we refer to as the “insulator loop,” which effectively isolates one active site from the other. However, this loop may also play additional and yet unknown roles, such as facilitating the translocation of the nicked DNA to the second active site. Consequently, engineering this enzyme should consider this feature and give special attention to the Insulator loop.

Rational engineering of restriction endonuclease specificity remains a challenging exercise, especially due to the high redundancy of contacts between protein residues and the DNA bases, as well as the limited structural and detailed mechanistic knowledge of the different steps leading to DNA cleavage. This high redundancy imposes a quasi-exhaustive study of

Table 1 | X-ray diffraction data collection and refinement statistics

	N0.BsmI (Mg) (PDB id. 9EZ5)	Nt.BsmI (Ca) (PDB id. 9EZ7)	Nb.BsmI (Mg) (PDB id. 9EZO)	wt.BsmI (Ca) (PDB id. 9F38)
Data Collection				
Space group	<i>P</i> 2 ₁ 2 ₁ 2	<i>P</i> 2 ₁ 2 ₁ 2	<i>P</i> 2 ₁ 2 ₁ 2	<i>P</i> 2 ₁
Cell dimensions				
<i>a</i> , <i>b</i> , <i>c</i> (Å)	98.48, 132.58, 63.74	98.55, 132.99, 63.28	98.21, 130.42, 62.54	65.75, 129.90, 100.75
α , β , γ (°)	90.0, 90.0, 90.0	90.0, 90.0, 90.0	90.0, 90.0, 90.0	90.0, 95.5, 90.0
Wavelength (Å)	0.97857	0.97856	0.97856	0.97856
Resolution (Å)	50.0–1.923 (1.956–1.923)	50.0–2.031 (2.197–2.031)	50.0–2.944 (2.995–2.944)	48.7–2.85 (2.92–2.85)
Estimated Resolution limit by Staraniso (Å)*	1.87, 1.7, 3.28	2.04, 2.13, 2.49	2.62, 2.97, 4.22	4.49, 2.82, 2.69
<i>R</i> _{pim}	0.072 (2.173)	0.104 (1.437)	0.245 (2.199)	0.224 (1.497)
<i>R</i> _{merge}	0.267 (7.488)	0.380 (5.202)	0.778 (6.670)	0.481 (3.108)
<i>I</i> / σ <i>I</i>	5.8 (0.3)	6.5 (0.7)	4.0 (0.6)	2.9 (0.5)
<i>CC</i> _{1/2}	0.997 (0.336)	0.992 (0.344)	0.979 (0.315)	0.957 (0.194)
Completeness (%)	99.7	100	100	99.8
Redundancy	14.3	14.4	11.0	5.5
<i>R</i> _{merge} *	0.186 (2.061)	0.297 (2.389)	0.520 (2.439)	0.345 (1.292)
<i>I</i> / σ <i>I</i> *	9.6 (1.5)	7.9 (1.5)	5.5 (1.5)	4.5 (1.7)
<i>CC</i> _{1/2} *	0.997 (0.626)	0.993 (0.564)	0.985 (0.433)	0.975 (0.531)
Completeness (%)*	92.1 (51.1)	93.1 (69.5)	91.1 (65.7)	61.8 (15.9)
Refinement				
Resolution (Å)	34.08–1.85	29.45–2.03	78.45–2.63	48.70–2.85
Unique reflections	38701	39796	12735	24310
<i>R</i> _{work} / <i>R</i> _{free}	0.190 / 0.230	0.188 / 0.221	0.208 / 0.239	0.279 / 0.298
Number of atoms				
- Protein	5458	5414	5422	11042
- DNA	527	527	526	1054
- Ligand/ion	13	15	14	19
- Water	439	477	51	175
B-factors (Å²)				
- Protein	29.591	25.343	40.134	47.889
- DNA	33.615	28.622	54.782	47.572
- Ligand/ion	33.444	26.539	73.0395	62.593
- Water	34.340	32.410	18.488	28.620
R.m.s. deviations				
- Bond lengths (Å)	0.009	0.008	0.007	0.006
- Bond angles (°)	0.95	0.94	0.87	0.80

Values in parentheses are for highest resolution shell. Dataset from single crystal used per structure.

* Values calculated after truncation by STARANISO. Estimated resolution limits along the three crystallographic directions *a**, *b**, *c**.

epistasis for engineering the enzyme, with a potential explosion of the space to be explored that could be difficult to manage in wet experiments. Despite numerous structures of EcoRI, EcoRV, and BamHI at various stages of the cleavage mechanism, the specific transient contacts made during cleavage remain unclear.

In this study, we provide new insights on DNA cleavage by monomeric restriction enzymes carrying two active sites. Knowing the structure in several conditions, a first interesting and attainable goal could be to rationally create an active top-strand nicking enzyme that does not require prior nicking of the bottom strand. This could be explored by stabilizing the interactions between the top-strand nicking domain with DNA: (i) by relaxing some of the interactions between the bottom-strand nicking domain while maintaining the enzyme specificity or (ii) by engineering point mutations in the Insulator loop.

Methods

Protein purification

The methyltransferases operon BsmIM was codon optimized, ordered from Geneart (Thermo Fisher) and cloned in a pSB3K3 vector under the control of a Lac promoter. The plasmid was transformed into T7 Express LysY/Iq (New England Biolabs) and the cells were treated to become electro-competent. The different mutants were obtained by site-directed mutagenesis (KLD enzyme mix, New England Biolabs) cloned into a pIVEX2.3 d vector (Roche Applied Science) and transformed in the aforementioned electrocompetent cells (Table S2). The different cultures were grown in LB at 30 °C until they reached OD 0.4–0.6 at 600 nm, they were then induced with 400 μ M of IPTG and transferred at 20 °C for an overnight growth. Cell lysates were obtained using a cell disruptor (CellD) and centrifuged at 20,000 g for 30 min. The proteins purified by affinity (HisTrap, Cytiva), cation-exchange (Heparin, Cytiva) and size exclusion (Superdex 200 Increase, Cytiva) chromatography steps using the buffers described in Table S3.

Analytical Ultra-Centrifugation

Sedimentation velocity experiments were performed at 20 °C using an Optima AUC analytical ultracentrifuge (Beckman Coulter) equipped with double-UV and Rayleigh interference detection. The samples were used immediately after size-exclusion chromatography in 10 mM Tris-HCl pH 7.5, 100 mM NaCl, 5 mM MgSO₄. The protein and protein-DNA complex at a concentration of 2.3 μ M were spun at 42,000 rpm using an AN60-Ti rotor and 12 mm thick Epon double sector centerpieces. Absorbance and interference profiles were recorded every 4 min. Detection of concentrations as a function of radial position and time was performed by optical density measurements at wavelengths of 260 and 280 nm for proteins and protein-DNA complexes, respectively.

Buffer viscosity (η = 1.0307 cP) and density (ρ = 1.00453 g/mL) at 20 °C were estimated with Sednterp 3 (<http://www.jphilo.mailway.com/download.htm>). Partial specific volumes of each construct at 20 °C were estimated based on amino acid sequences using Sednterp 3. For BsmI/DNA complexes, partial specific volumes were estimated by weight averaging of the partial specific volume of the amino acid sequences (0.739 mL/g) and the nucleotide sequences (0.57 mL/g). Data were analyzed with Sedfit 16.1 using the continuous size distribution c(S) model³⁸.

Crystallogenesis

The proteins were mixed with the 13 bp dsDNA fragment (5'-GAG-GAATGCAGAC-3') and the concentrations were adjusted to reach 128 μ M (10 mg/mL) for the protein and 154 μ M for the DNA. Crystallization screening trials were carried out by the vapor diffusion method using a Mosquito TM nanodispensing system (STPLabtech, Melbourn, UK) following established protocols³⁹. Briefly, we set up crystallization sitting drops of 400 nl containing a 1:1 mixture of protein sample in complex with DNA and crystallization solutions (672 different commercially available conditions) equilibrated against 150 μ l of reservoir solution in multiwell plates (Greiner Bio-One). The crystallization plates were stored at 18 °C in a

RockImager (Formulatrix) automated imaging system to monitor crystal growth. Manual optimization was performed in Linbro plates with the hanging-drop method by mixing 1 μ l of protein samples with 1 μ l of reservoir solution. The best crystals were obtained with the crystallization condition containing 100 mM MES pH 6.5, 100 mM MgCl₂ or CaCl₂, 30% w/v PEG 400, 250 mM Guanidine HCl. The crystals were flash-cooled in liquid nitrogen for data collection using the crystallization solution as cryoprotectant.

X-ray data collection, structure solution and model refinement

Crystallographic data was collected at the synchrotron SOLEIL (France, Saint-Aubin) on beamlines PROXIMA-1⁴⁰ and processed by XDS⁴¹ with the XDSME⁴² or the AutoPROC⁴³ pipeline. The first structure was solved by experimental SAD phasing using crystals soaked in a Europium Chloride solution at the Europium LIII X-ray absorption edge (1.7757 Å wavelength). A single highly redundant datasets (multiplicity of 120) was assembled by merging data collected on 5 different Eu³⁺ derived crystals using XSCALE⁴¹. It turned out that most of the anomalous signal was contributed by phosphorus and sulphur atoms at this wavelength. Initial substructure sites were found using the SHELX C/D/E⁴⁴ pipeline and refined using Phaser⁴⁵. Phases were improved by density modification using PARROT⁴⁶. All the other structures were solved using molecular replacement with this model and refined with Buster⁴⁷ using Coot for manual reconstruction⁴⁸. Data collection and model refinement statistics are given in Table 1.

Coordinates and structure factors have been deposited in the Protein Data Bank under the accession codes (PDB id. 9EZ5) for N0.BsmI (with Mg²⁺); (PDB id. 9EZ7) for Nt*.BsmI (with Ca²⁺); (PDB id. 9EZD) for Nb.BsmI (with Mg²⁺); and (PDB id. 9F38) for wt.BsmI (with Ca²⁺). Figures showing the crystallographic models were generated with ChimeraX⁴⁹.

Activity assays

The oligonucleotides were purchased from Eurofins Genomics, annealed together in ultrapure water and, when needed, assembled using the T4 DNA ligase (New England Biolabs). The reactions contained 1 μ M of DNA probe, 1.5 μ g/mL (20 nM) of enzyme and, except when otherwise stated, the r2.1 buffer (New England Biolabs) in which both BsmI and Nb.BsmI are known to be active. The reactions were incubated at 65 °C for 20 min, heat-inactivated at 80 °C for another 20 min before the addition of the loading solution containing 95% v/v deionized formamide and 50 mM EDTA supplemented with bromophenol blue. The products were resolved by electrophoresis on a 16% w/v polyacrylamide gel containing 8 M urea. The results were imaged using a Typhoon FLA9500 laser-scanner (Cytiva).

PTO enantiomers separation

5'-phosphorylated 13-base-long oligonucleotides containing one phosphorothioate bond were resuspended in ultrapure water. The DNA PAC PA100 pellicular anion-exchange chromatography column (Thermo Fischer Scientific) was equilibrated with a solution containing 25 mM Tris-HCl pH 6.75. After injection of 5–20 nmol of oligonucleotide, the concentration in NaCl was gradually increased from 0 to 268 mM in 2 min and then from 268 mM to 400 mM at a flowrate of 1.5 mL/min. The recovered fractions corresponding to the peaks were dialyzed against ultrapure water (ReadyLyzer 10, MWCO 1 kDa, Serva) and concentrated in a speed vacuum concentrator (Eppendorf).

Normal mode analysis

Normal Modes were calculated by the Non-Linear Rigid-Block method (NOLB) using default values for the cut-off and the amplitude, and asking for the first 10 lowest frequency normal modes³⁰. Only the protein atoms were used in the model. 10 frames were included in the trajectory.

Statistics and reproducibility

After a broad initial screening, the crystals were optimized and reproduced twice. All the datasets presented in this article were collected on at least two

different crystals for each condition. All the enzymatic activity tests were performed at least three times.

Reporting summary

Further information on research design is available in the Nature Portfolio Reporting Summary linked to this article.

Data availability

The data underlying this article are available in the PDB at www.rcsb.org and can be accessed with identifiers 9EZ5, 9EZ7, 9EZD and 9F38. The raw data of the gel electrophoresis presented in this article are available as Supplementary Figs. (Figures S9, S10, S11 and S12). The supplementary data 1 file contains the ultracentrifuge raw data. The supplementary data 2 and 3 contains respectively the raw data of the chromatograms of the DNAPac separation of the oligonucleotides containing only phosphodiester (PDE) bonds and a phosphorothioate (PTO) bond at the scissile position presented in Fig. 7. The supplementary data 4 file contains the pdb files and the PDB validation reports. All data are available from the corresponding author on reasonable request.

Received: 24 July 2024; Accepted: 28 January 2025;

Published online: 07 March 2025

References

- Arber, W. & Linn, S. DNA Modification and Restriction. *Annu. Rev. Biochem.* **38**, 467–500 (1969).
- Roberts, R. J., Vincze, T., Posfai, J. & Macelis, D. REBASE—a database for DNA restriction and modification: enzymes, genes and genomes. *Nucleic Acids Res.* **43**, D298–D299 (2015).
- Roberts, R. J. et al. A nomenclature for restriction enzymes, DNA methyltransferases, homing endonucleases and their genes. *Nucleic Acids Res.* **31**, 1805–1812 (2003).
- Bujnicki, J. M. Crystallographic and Bioinformatic Studies on Restriction Endonucleases: Inference of Evolutionary Relationships in the “Midnight Zone” of Homology. *Curr. Protein Pept. Sci.* **4**, 327–337 (2003).
- Pingoud, A., Fuxreiter, M., Pingoud, V. & Wende, W. Type II restriction endonucleases: structure and mechanism. *Cell. Mol. Life Sci.* **62**, 685 (2005).
- Pingoud, A. & Jeltsch, A. Structure and function of type II restriction endonucleases. *Nucleic Acids Res.* **29**, 3705–3727 (2001).
- Tamulaitiene, G. et al. Restriction endonuclease AgeI is a monomer which dimerizes to cleave DNA. *Nucleic Acids Res.* **45**, 3547–3558 (2017).
- Newman, M., Strzelecka, T., Dorner, L., Schildkraut, I. & Aggarwal, A. Structure of restriction endonuclease BamHI phased at 1.95 Å resolution by MAD analysis. *Structure* **2**, 439–452 (1994).
- Modrich, P. & Zabel, D. EcoRI endonuclease. Physical and catalytic properties of the homogenous enzyme. *J. Biol. Chem.* **251**, 5866–5874 (1976).
- Pingoud, A., Wilson, G. G. & Wende, W. Type II restriction endonucleases—a historical perspective and more. *Nucleic Acids Res.* **42**, 7489–7527 (2014).
- Horton, J. R. et al. Structure of HhaI endonuclease with cognate DNA at an atomic resolution of 1.0 Å. *Nucleic Acids Res.* **48**, 1466–1478 (2020).
- Xu, Q. S., Roberts, R. J. & Guo, H.-C. Two crystal forms of the restriction enzyme MspI–DNA complex show the same novel structure. *Protein Sci.* **14**, 2590–2600 (2005).
- Sokolowska, M. et al. Monomeric Restriction Endonuclease BcnI in the Apo Form and in an Asymmetric Complex with Target DNA. *J. Mol. Biol.* **369**, 722–734 (2007).
- Too, P. H.-M., Zhu, Z., Chan, S.-H. & Xu, S. Engineering Nt.BtsCI and Nb.BtsCI nicking enzymes and applications in generating long overhangs. *Nucleic Acids Res.* **38**, 1294–1303 (2010).
- Armalyte, E. et al. Mva1269I: A Monomeric Type IIS Restriction Endonuclease from *Micrococcus Varians* with Two EcoRI- and FokI-like Catalytic Domains. *J. Biol. Chem.* **280**, 41584–41594 (2005).
- Bath, A. J., Milsom, S. E., Gormley, N. A. & Halford, S. E. Many Type IIS Restriction Endonucleases Interact with Two Recognition Sites before Cleaving DNA. *J. Biol. Chem.* **277**, 4024–4033 (2002).
- Xu, S.-Y., Zhu, Z. & Meixsell, T. Nicking endonuclease methods and compositions. Patent WO2006012593A1 (2012).
- Zhou, J., Zhu, Z. & Xu, S.-Y. Method for cloning and expression of BSMI Restriction-Modification system in *E. coli*. European Patent EP1199366B1 (2007).
- de Marco, A. et al. Quality control of protein reagents for the improvement of research data reproducibility. *Nat. Commun.* **12**, 2795 (2021).
- Chan, S.-H., Stoddard, B. L. & Xu, S. Natural and engineered nicking endonucleases—from cleavage mechanism to engineering of strand-specificity. *Nucleic Acids Res.* **39**, 1–18 (2011).
- Holm, L. Dali server: structural unification of protein families. *Nucleic Acids Res.* **50**, W210–W215 (2022).
- Abramson, J. et al. Accurate structure prediction of biomolecular interactions with AlphaFold 3. *Nature* **630**, 493–500 (2024).
- Landau, M. et al. ConSurf 2005: the projection of evolutionary conservation scores of residues on protein structures. *Nucleic Acids Res.* **33**, W299–W302 (2005).
- Ashkenazy, H. et al. ConSurf 2016: an improved methodology to estimate and visualize evolutionary conservation in macromolecules. *Nucleic Acids Res.* **44**, W344–W350 (2016).
- Eckstein, F. Phosphorothioate Analogues of Nucleotides—Tools for the Investigation of Biochemical Processes. *Angew. Chem. Int. Ed. Engl.* **22**, 423–439 (1983).
- Thorogood, H., Grasby, J. A. & Connolly, B. A. Influence of the Phosphate Backbone on the Recognition and Hydrolysis of DNA by the EcoRV Restriction Endonuclease. *J. Biol. Chem.* **271**, 8855–8862 (1996).
- Jahns, H. et al. Chirality matters: stereo-defined phosphorothioate linkages at the termini of small interfering RNAs improve pharmacology in vivo. *Nucleic Acids Res.* **50**, 1221–1240 (2022).
- Lan, W. et al. Structural investigation into physiological DNA phosphorothioate modification. *Sci. Rep.* **6**, 25737 (2016).
- Frederiksen, J. K. & Piccirilli, J. A. Chapter 14 - Separation of RNA Phosphorothioate Oligonucleotides by HPLC. In *Methods in Enzymology*, Biophysical, Chemical, and Functional Probes of RNA Structure, Interactions and Folding: Part A. Academic Press, Vol. 468, pp. 289–309 (2009).
- Hoffmann, A. & Grudin, S. NOLB: Nonlinear Rigid Block Normal-Mode Analysis Method. *J. Chem. Theory Comput.* **13**, 2123–2134 (2017).
- Xia, S., Wang, M., Blaha, G., Konigsberg, W. H. & Wang, J. Structural Insights into Complete Metal Ion Coordination from Ternary Complexes of B Family RB69 DNA Polymerase. *Biochemistry* **50**, 9114–9124 (2011).
- Gouge, J., Rosario, S., Romain, F., Beguin, P. & Delarue, M. Structures of Intermediates along the Catalytic Cycle of Terminal Deoxynucleotidyltransferase: Dynamical Aspects of the Two-Metal Ion Mechanism. *J. Mol. Biol.* **425**, 4334–4352 (2013).
- Cowan, J. A. Transition Metals as Probes of Metal Cofactors in Nucleic Acid Biochemistry. *Comments Inorg. Chem.* **13**, 293–312 (1992).
- Cowan, J. A. Magnesium activation of nuclease enzymes—the importance of water. *Inorganica Chim. Acta* **275–276**, 24–27 (1998).
- Kim, S. & Cowan, J. A. Inert cobalt complexes as mechanistic probes of the biochemistry of magnesium cofactors. Application to topoisomerase I. *Inorg. Chem.* **31**, 3495–3496 (1992).
- Cowan, J. A. Role of Metal Ions in Promoting DNA Binding and Cleavage by Restriction Endonucleases. In Pingoud, A. M. (ed), *Restriction Endonucleases*. Springer, Berlin, Heidelberg, pp. 339–360 (2004).

37. Viadiu, H. & Aggarwal, A. K. The role of metals in catalysis by the restriction endonuclease Bam HI. *Nat. Struct. Biol.* **5**, 910–916 (1998).
 38. Schuck, P. Size-Distribution Analysis of Macromolecules by Sedimentation Velocity Ultracentrifugation and Lamm Equation Modeling. *Biophys. J.* **78**, 1606–1619 (2000).
 39. Weber, P. et al. High-Throughput Crystallization Pipeline at the Crystallography Core Facility of the Institut Pasteur. *Molecules* **24**, 4451 (2019).
 40. Chavas, L. M. G. et al. PROXIMA-1 beamline for macromolecular crystallography measurements at Synchrotron SOLEIL. *J. Synchrotron Radiat.* **28**, 970–976 (2021).
 41. Kabsch, W. XDS. *Acta Crystallogr. D Biol. Crystallogr.* **66**, 125–132 (2010).
 42. Legrand, P. XDSME: XDS Made Easier GitHub repository (2017).
 43. Vonrhein, C. et al. Data processing and analysis with the autoPROC toolbox. *Acta Crystallogr. D Biol. Crystallogr.* **67**, 293–302 (2011).
 44. Sheldrick, G. M. Experimental phasing with SHELXC/D/E: combining chain tracing with density modification. *Acta Crystallogr. D Biol. Crystallogr.* **66**, 479–485 (2010).
 45. McCoy, A. J. et al. Phaser crystallographic software. *J. Appl. Crystallogr.* **40**, 658–674 (2007).
 46. Cowtan, K. Recent developments in classical density modification. *Acta Crystallogr. D Biol. Crystallogr.* **66**, 470–478 (2010).
 47. Bricogne, G. & Irwin, J. Refinement of Incomplete Structures with Maximum Likelihood in BUSTER-TNT. *Acta Crystallogr. D Biol. Crystallogr.* **60**, 2210–2221 (2004).
 48. Emsley, P., Lohkamp, B., Scott, W. G. & Cowtan, K. Features and development of Coot. *Acta Crystallogr. D Biol. Crystallogr.* **66**, 486–501 (2010).
 49. Goddard, T. D. et al. UCSF ChimeraX: Meeting modern challenges in visualization and analysis. *Protein Sci.* **27**, 14–25 (2018).
- the mutants. R.S., A.H., (S.M., C.M., and L.S.) performed (helped with) the crystallography experiments. R.S., A.H., I.C. and P.L. collected X-ray diffraction datasets. P.L. solved the structure by low-energy SAD method. R.S. and P.L. refined the structures, R.S., S.M., P.L. and M.D. analyzed the structure. M.D. performed normal mode analysis and supervised the study. R.S., M.H. and G.N. implemented the PTO enantiomers separation. All authors reviewed the final version of the manuscript.

Competing interests

The authors declare no competing interests.

Additional information

Supplementary information The online version contains supplementary material available at <https://doi.org/10.1038/s42003-025-07612-z>.

Correspondence and requests for materials should be addressed to Marc Delarue.

Peer review information *Communications Biology* thanks Albert Jeltsch and the other, anonymous, reviewer(s) for their contribution to the peer review of this work. Primary Handling Editors: Laura Rodríguez Pérez and David Favero.

Reprints and permissions information is available at <http://www.nature.com/reprints>

Publisher's note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Open Access This article is licensed under a Creative Commons Attribution-NonCommercial-NoDerivatives 4.0 International License, which permits any non-commercial use, sharing, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if you modified the licensed material. You do not have permission under this licence to share adapted material derived from this article or parts of it. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by-nc-nd/4.0/>.

© The Author(s) 2025

Acknowledgements

We thank the staff of the Crystallography core facility at the Institut Pasteur for carrying out robot-driven crystallization screenings and Biophysical core facility at the Institut Pasteur is acknowledged for carrying out AUC analysis and quality controls of our proteins samples. We acknowledge SOLEIL for provision of synchrotron radiation facilities and we would like to thank beamline staff for assistance in using beamline PROXIMA-1. Y.R. acknowledges financial support of the European Research Council (ERC, Consolidator Grant No. 647275 ProFF) and of the CNRS (PEPR MoleculeArxiv Grant No. ANR-22-PEXM-0002). M.D.'s laboratory is supported by I. Pasteur and CNRS.

Author contributions

R.S. and Y.R. designed and cloned the different mutant genes. R.S. and M.D. designed the study and wrote the manuscript. R.S. expressed and purified