Structure of the tetrameric restriction endonuclease *Ngo*MIV in complex with cleaved DNA

Markus Deibert[1], Saulius Grazulis[1,2], Giedrius Sasnauskas[2], Virginijus Siksnys[2] and Robert Huber[1]

The crystal structure of the *Ngo*MIV restriction endonuclease in complex with cleaved DNA has been determined at 1.6 Å resolution. The crystallographic asymmetric unit contains a protein tetramer and two DNA molecules cleaved at their recognition sites. This is the first structure of a tetrameric restriction enzyme–DNA complex. In the tetramer, two primary dimers are arranged back to back with two oligonucleotides bound in clefts on opposite sides of the tetramer. The DNA molecules retain a B-type conformation and have an enclosed angle between their helical axes of 60°. Sequence-specific interactions occur in both the major and minor grooves. Two Mg^{2+} ions are located close to the cleaved phosphate at the active site of *Ngo*MIV. Biochemical experiments show that interactions between the recognition sites within the tetramer greatly increase DNA cleavage efficiency.

Almost 3,000 type II restriction enzymes representing nearly 200 different specificities have now been discovered[1]. They usually recognize DNA sequences that vary between four to eight base pairs (bp) in length and cleave the phosphodiester bonds in the presence of Mg^{2+} within or close to the recognized sequence. It has become a paradigm that restriction endonucleases are arranged as homodimers reflecting the two-fold symmetry of their palindromic recognition sequence. Contrary to the current model, however, biochemical studies of *Sfi*I[2] and *Cfr*10I[3] restriction enzymes propose that their functionally active forms are homotetramers. It was demonstrated that tetramers of *Sfi*I and *Cfr*10I are able to interact with two copies of the DNA simultaneously and such interactions highly increase the efficiency of DNA cleavage. Recent kinetic studies of the *Sgr*AI restriction enzyme indicate that this enzyme also needs two recognition sites for optimal activity, suggesting that it has a possible tetrameric organization[4]. The type IIe restriction enzymes *Nae*I[5] and *Eco*RII[6], which are both reported to be homodimers, have also been shown to interact with two recognition sequences.

The *Ngo*MIV restriction enzyme from *Neisseria gonorrhoeae* strain M recognizes the palindromic hexanucleotide sequence 5′-GCCGGC-3′ and cleaves after the 5′-G on each strand to produce 4 bp 5′ staggered ends[7]. We have determined the crystal structure of *Ngo*MIV bound to the cognate 11 bp DNA fragment in the enzyme–product state in the presence of Mg^{2+} ions. The crystal structure revealed that *Ngo*MIV is arranged as a tetramer that interacts with two oligonucleotides cleaved at their recognition sites. As far as we know this is the first structure of a tetrameric restriction enzyme in complex with DNA.

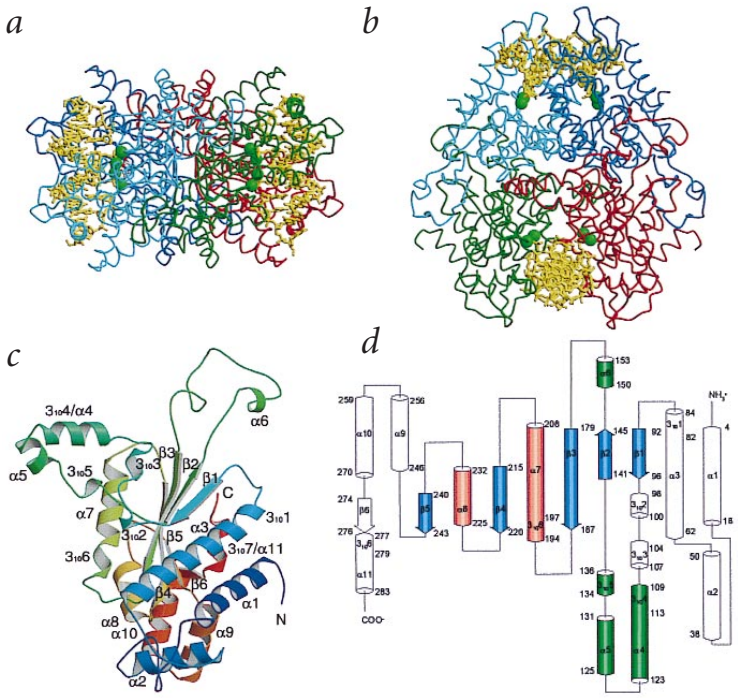

Fig. 1 Overall structure of the *Ngo*MIV–DNA complex. ***a***, Side view of the *Ngo*MIV–DNA complex. Monomers are colored blue (A), red (B), green (C) and gray (D). DNA is shown in yellow and Mg^{2+} ions as magenta spheres. ***b***, View of the *Ngo*MIV–DNA complex along the axis of one DNA duplex. ***c***, *Ngo*MIV monomer with labeled α-helices and 3_{10}-helices and β-strands. The chain runs from blue (N-terminus, N) to red (C-terminus, C). ***d***, Topology diagram of *Ngo*MIV. The central β-sheet is shown in blue, protruding arms in green, and recognition helices in red.

General architecture of the complex

The structure was solved using the technique of multiple isomorphous replacement with anomalous scattering. An experimental electron density map calculated at 2.5 Å was of extremely high quality and enabled most of the protein and all the DNA to be built. The structure has been refined to 1.6 Å resolution with excellent statistics and stereochemistry (Table 1). A total of

[1]Max-Planck-Institut für Biochemie, D-82152 Planegg-Martinsried, Germany. [2]Institute of Biotechnology, Lt-2028 Vilnius, Lithuania.

Correspondence should be addressed to M.D. *email:* markus.deibert@cii.de or V.S. *email: siksnys@ibt.lt*

1,279 solvent molecules were included in the final refined model.

The crystallographic asymmetric unit contains a *Ngo*MIV tetramer and two double-stranded oligonucleotides (Fig. 1*a*,*b*). The protein part shows local 222 symmetry, which is also followed by the mercury heavy atom positions. The *Ngo*MIV tetramer forms a compact globule with dimensions of 60 Å × 70 Å × 80 Å, which is built up as a dimer of dimers arranged back to back. Two DNA duplexes are bound in the large clefts formed between monomers in the primary dimer and face opposite directions. Therefore, DNA molecules are separated by the protein and positioned ~55 Å apart from each other (Fig. 1*a*,*b*). In the *Ngo*MIV tetramer the concavely shaped backsides of the dimers are slightly rotated around their local two-fold axis, such that the two bound DNA duplexes are positioned in an X-like fashion with their helical axes enclosing an angle of 60°.

Structure of the monomer and arrangement of dimer

The *Ngo*MIV monomer exhibits an α/β structure (Fig. 1*c*,*d*). The central region comprises a mixed six-stranded β-sheet (β1 to β6) with α-helices (α1, α3/3_{10}1, α7, α8, 3_{10}6/α11) on each side. The core structure of *Ngo*MIV is similar to that of other restriction enzymes[8]. A long 'arm' protrudes 25 Å out of the core of each monomer between strands β2 and β3 and is mainly involved in the tetramerization interface (Fig. 1*c*). The loop region between strands β1 and β2, including the helices 3_{10}4/α4, α5 and 3_{10}5 contribute both to the tetramer and dimer interfaces.

The tetramer of *Ngo*MIV in the asymmetric unit is built up of two primary dimers (comprising subunits A and B, and C and D, respectively). The primary dimers of *Ngo*MIV are structurally similar to that of dimeric restriction enzymes that produce 4 bp 5′ staggered ends after DNA cleavage. Therefore, the structural elements of *Ngo*MIV used for the dimer formation are similar to those of *Eco*RI, *Bam*HI and *Mun*I[9–11]. Dimerization of *Ngo*MIV occurs primarily by contacts between helices 3_{10}6/α7 from subunits A and B in one primary dimer (and from subunits C and D in another primary dimer), which cross over their counterpart of the related subunit at the Arg 194 residue in a V-like fashion. The N-termini of helices 3_{10}6/α7 point into the DNA major groove similarly to structurally equivalent helices α4 in *Eco*RI[9] and *Bam*HI[10] and helix 3_{10}4 in *Mun*I[11]. However, in contrast to the dimeric enzymes, the C-termini of the 3_{10}6/α7 helices of *Ngo*MIV contribute to the tetramerization interface. The helix α8 in *Ngo*MIV is the structural counterpart of helix α5 in *Eco*RI and *Mun*I and α6 in *Bam*HI, but it is positioned too far away to cross helix 3_{10}6/α7 to build up a four-helix bundle. Instead, the N-terminus of helix α8 is involved in cross contacts with a neighboring monomer while its C-terminal part and downstream region contact helix α5 of the symmetry related dimer. The loop region between strands β1-β2 of *Ngo*MIV is topologically similar to the dimerization arm (residues 59–75) of *Mun*I, which protrudes out of the core and wraps around the symmetry related monomer. The similar region connecting the first and the second strand of the β-sheet is also present in *Eco*RI (residues 60–88) and *Bam*HI (residues 77–94) and contributes to the dimerization interface. The length and conformation of these structural elements, however, differ between restriction enzymes.

Tetramer assembly

The tetrameric assembly of *Ngo*MIV is fixed by side-by-side contacts (between subunits A/D and B/C) and cross contacts (between subunits A × C and B × D) between primary dimers (Fig. 1*a*,*b*). Cross contacts are made exclusively by the helices 3_{10}4/α4 and α5/3_{10}5, which also contribute to the dimerization interface. The most extensive contacts in the *Ngo*MIV tetramer, between subunits A/D and B/C, are made by the tetramerization loop (residues 147–176, including helix α6) that spans across the neighboring monomer. The residues located on the α6 helix contact the C-terminus of the α3 helix while other residues positioned on the loop make extensive contacts with the N-terminal and C-terminal residues of *Ngo*MIV. Additional contacts at the A/D and B/C interface come from amino acid residues located on the C-terminal ends of helices α7 and α8 and downstream regions. In total, the contacts between subunits at the tetramerization interface comprise eight salt bridges and numerous hydrophobic interactions.

Accessible surface area calculations indicate that a large monomer surface area of ~4,200 Å^2 is buried upon tetramerization. The contribution of individual intersubunit contacts to the totally buried surface area differs notably. The cross contacts between monomers A × C contribute only ~400 Å^2 per monomer. The contacts between subunits AB of the primary dimers hide ~1,000 Å^2 per monomer while the largest surface area of ~2,800 Å^2 per monomer is buried in side-by-side contacts between subunits A/D. Thus, the surface area of each monomer shielded during tetramerization exceeds the surface area buried at the dimer interface by almost three times.

Of note is the fact that the loop connecting strands β2-β3 (Fig. 1*c*,*d*), which contributes major contacts to the tetramerization interface of *Ngo*MIV, is not present in the dimeric restriction enzymes *Eco*RI, *Bam*HI and *Mun*I, in which the two structurally equivalent β-strands are linked *via* a short loop. Interestingly, the *Cfr*10I restriction enzyme, which is functionally active as a tetramer[3], possesses a topologically similar loop (residue 140–181). The monomers in the *Cfr*10I crystal form an exact D2 symmetric tetramer with a similar back to back arrangement of dimers[3]. However, in contrast to *Ngo*MIV, the topologically similar loop connecting the β-strands of *Cfr*10I is not involved in tetramerization but rather contacts the N-terminal subdomain of the same subunit. The total contact surface area between primary dimers in the *Cfr*10I tetramer (2,300 Å) is, however, smaller than in *Ngo*MIV (4,200 Å). Notably, the disruption of the assumed dimer–dimer interface of *Cfr*10I by a W220A mutation renders the enzyme nearly inactive, suggesting that the loop might not be involved in the interdimer contacts[3].

Recognition of the specific DNA sequence

The *Ngo*MIV restriction enzyme binds to DNA from the major groove side, similar to *Eco*RI[12], *Bam*HI[10] and *Mun*I[11] but opposite to *Eco*RV[13], *Pvu*II[14] and *Bgl*I[15], which bind from the minor groove side. All four DNA strands in the *Ngo*MIV complex structure are cleaved and remain bound to the protein. The cleaved DNA retains a B-DNA-like conformation with no major kinks or bends as seen in the complexes of *Eco*RI[12], *Mun*I[11] and *Eco*RV[13] with specific DNA. Notably, there are local deviations from the standard B-DNA conformation near the active site in the *Ngo*MIV–DNA complex.

The recognition of the specific sequence is achieved through the number of direct contacts between protein side chains and DNA bases (Fig. 2*a*). Direct read out takes place primarily in the DNA major groove with one interaction occurring from the minor groove side. A single subunit of *Ngo*MIV makes hydrogen bonds to the bases of the half site GCC in the major groove, while a minor groove contact to the C base of the outer GC pair comes

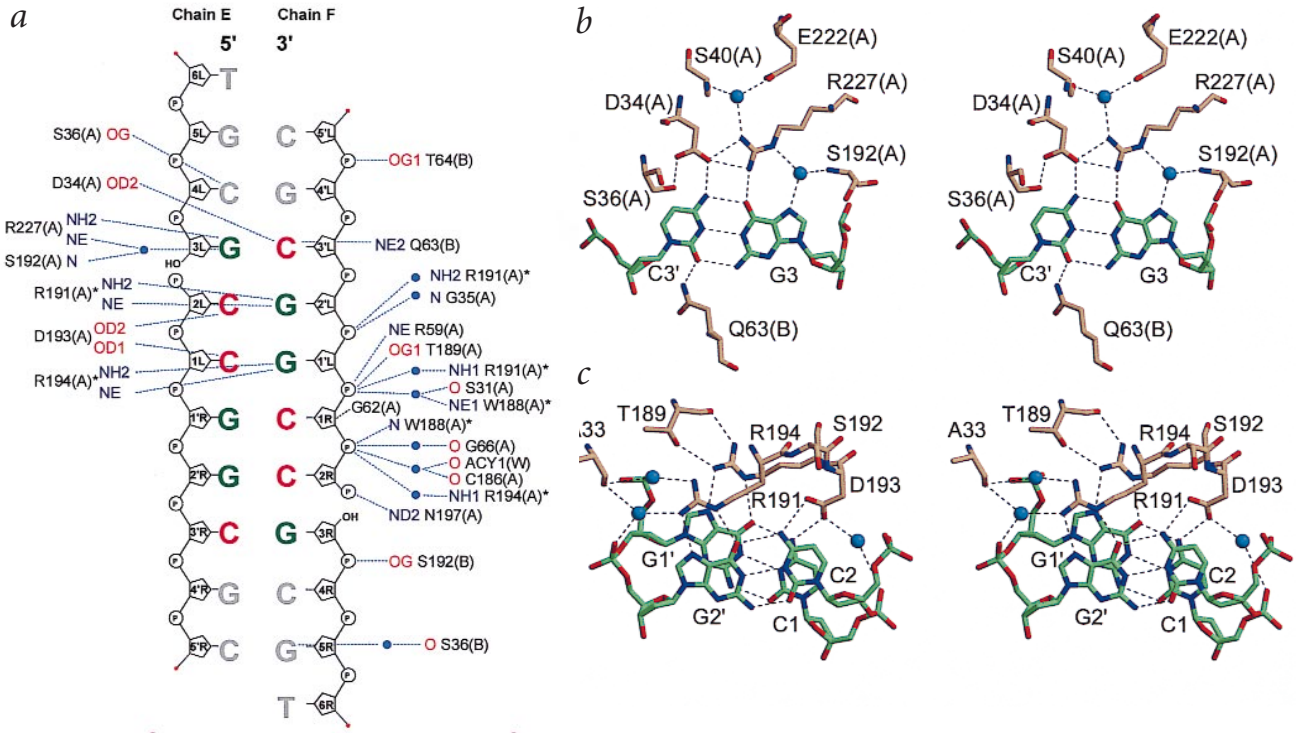

Fig. 2 *Ngo*MIV–DNA interactions. ***a***, NUCPLOT[37] sketch illustrating contacts between *Ngo*MIV and DNA1 (strands E and F). Only contacts from subunits A and B to one half site (designated L) of the recognition sequence are shown. Contacts to the sugar phosphate backbone are shown for the F chain only. Bases in the recognition sequence are colored according to their type (Gua, green; Cyt, red) and water molecules are blue dots. Asterisks represent residues that appear on the plot more than once. ***b***, Stereo view of the outer GC base pair (G3L C3′L) recognition. Water molecules are shown as blue dots. Asp 34 makes a hydrogen bond to Cyt 3′L. The NH2 atom of Arg 227 makes a direct hydrogen bond to the O6 oxygen of Gua 3L while the Nε of Arg 227 in companion with the main chain nitrogen of Ser 192 sandwiches a water molecule that is hydrogen bonded to the N7 nitrogen of Gua^{3L}. The amino group of Gln 63 from the neighboring subunit donates a hydrogen bond to the oxygen atom of Cyt 3′L in the minor groove. ***c***, Stereo view of the inner 5′-CCGG sequence recognition. NH2 and Nε atoms of Arg 191 and Arg 194 make bidentate hydrogen bonds to Gua 2′ and Gua 1′, respectively. The carboxylate oxygens of Asp 193 bridge the amino groups of the neighboring Cyt bases.

from the neighboring subunit. The major groove contacts are formed by residues located in the loop preceding helix α2, the N-terminus of helix $3_{10}6$/α7 and a region upstream of it and helix α8. The helices α7 and α8 are dominant at the recognition interface. Other structural elements involved in recognition are located on both sides of these central helices upstream of the DNA center and follow the surface curvature of the double-stranded oligonucleotide. Of note is the fact that the recognition helix $3_{10}6$/α7 also contributes both to the dimerization and tetramerization interfaces and might be involved in the crosstalking interactions between individual monomers in the *Ngo*MIV tetramer.

As all specific contacts are almost identical between the two DNA half sites in both DNA molecules we therefore describe them only for half site L of DNA1 (Fig. 2*a*). The major groove contacts to the outer G3L-C3′L base pair (Fig. 2*b*) arise from the residues located on two separate structural elements, namely on loops preceding helix α2 and helix α8. Asp 34 makes a hydrogen bond to the exocyclic amino group of Cyt 3′L [N4–Oδ2 (3.0 Å)]. The side chain of Asp 34 is also buttressed by a salt bridge to Arg 227 [Oδ2–NH1 (2.9 Å), Oδ2–NH2 (3.0 Å)] and a hydrogen bond to Ser 36 [Oδ1–Oγ (3.0 Å)]. The NH2 atom of Arg 227 makes a direct hydrogen bond to the O6 oxygen of Gua 3L [NH2–O6 (2.8 Å)] while the Nε of Arg 227 [Nε–O (2.9 Å)] in companion with the main chain nitrogen of Ser 192 [N–O (3.0 Å)] sandwiches a water molecule, which is hydrogen bonded to the N7 nitrogen of Gua 3L [O–N7 (2.8 Å)]. Arg 227 through the water molecule is involved in buttressing interactions with Ser 40 and Glu 222. The amino group of Gln 63, located at the N-terminus of helix α3 of the neighboring subunit, donates a hydrogen bond to the oxygen atom of Cyt 3′L [Nε2–O2 (2.8 Å)] in the minor groove. The side chain conformation of Gln 63 is probably fixed through the interaction of its carbonyl oxygen Oε1 atom with backbone nitrogen atoms of Gln 63 and Gly 62. Both subunits of a primary dimer thus contribute to the recognition of the outer GC pair of one half site.

The central 5′-CCGG part of the *Ngo*MIV recognition sequence is contacted by residues located on a short stretch of amino acids with the sequence 191RSDR just upstream of helix $3_{10}6$/α7. NH2 and Nε atoms of Arg 191 and Arg 194 (Fig. 2*c*) make bidentate hydrogen bonds to the Gua 2′ [Nε–O6 (3.0 Å), NH2–N7 (2.8 Å)] and Gua 1′ [Nε–O6 (3.1 Å), NH2–N7 (3.0 Å)], respectively. The carboxylate oxygens of Asp 193 bridges exocyclic amino groups of the neighboring Cyt 1L [Oδ1–N4 (2.9 Å)] and Cyt 2L [N4–Oδ2 (2.9 Å)] bases. In addition, Arg 191 and Arg 194 residues are involved in numerous buttressing interactions. Two highly localized water molecules are sandwiched between the Arg 191 and O1P oxygens of Gua 1′L and Gua 2′L [NH1–O (3.0 Å)–O1P (2.8 Å), NH2–O (2.9 Å)–O1P (2.9 Å)] and main chain carbonyl of Ala 33 [O–O (3.2 Å), O–O (3.6 Å)]. The side chain of Arg 194 is buttressed to the side chain and main chain oxygens of Thr 189. Thus, three neighboring amino acid residues, Arg 191, Asp 193 and Arg 194, completely satisfy the hydrogen bonding potential of two adjacent CG base pairs and unambiguously specify these bases.

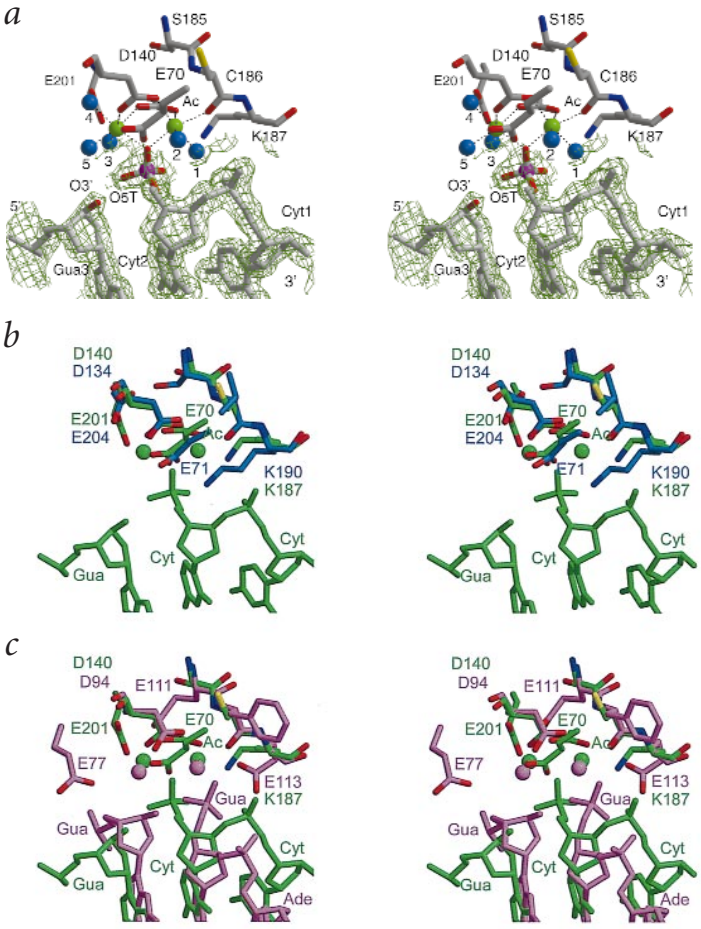

Fig. 3 The active site of *Ngo*MIV. ***a***, Stereo view of the coordination geometry of Mg^{2+} ions at the active site of *Ngo*MIV in the enzyme–product complex. Two Mg^{2+} ions are shown as green spheres. Both Mg^{2+} exhibit octahedral coordination. The O2P of the 5′ phosphate, the carboxylate of Asp 140 and the acetate molecule contribute to the coordination of both Mg^{2+} ions. The remaining three ligands of Mg^{2+} ion A are the backbone oxygen of Cys 186 and water molecules 1 and 2 (shown as blue spheres). Water molecules 3–5 complete the octahedral coordination of Mg^{2+} ion B. All Mg coordinated waters are fixed by either protein and/or DNA residues. The final $2F_o$ - F_c electron density for the DNA is shown contoured at 1.5 σ. ***b***, Stereo view of the superimposed active sites of *Ngo*MIV and *Cfr*10I. The *Cfr*10I residues are shown in blue, the *Ngo*MIV residues in green, with the two Mg^{2+} ions shown as green spheres. ***c***, Stereo view of the superimposed active sites of *Bam*HI and *Ngo*MIV. The *Bam*HI residues are shown in magenta, the *Ngo*MIV residues in green with the two metal ions as magenta (Mn^{2+} in *Bam*HI) and green (Mg^{2+} in *Ngo*MIV) spheres. Note that four acidic residues (Glu 77, Asp 94, Glu 111 and Glu 113) are located in the vicinity of the Mn^{2+} ions at the active site of *Bam*HI, while only three (Glu 70, Asp 140, Glu 201) are present at the active site of *Ngo*MIV. Lys 187 of *Ngo*MIV is structurally equivalent to Glu 113 of *Bam*HI.

The sequence region 194RPDR is present at the structurally equivalent positions of the *Cfr*10I restriction enzyme, which recognize the Pu/CCGGPy sequence[16]. Thus, *Ngo*MIV and *Cfr*10I probably utilize similar residues and structural mechanisms to interact with central CCGG bases. Of note is that the recognition of the AATT tetranucleotide common to *Eco*RI (G/AATTC) and *Mun*I (C/AATTG) restriction enzymes is achieved by a similar structural mechanism[11]. In these cases a short stretch of conserved residues (GNA(I/H)ER) located on the α-helices of *Eco*RI and *Mun*I is structurally equivalent to the helix $3_{10}6/\alpha7$ of *Ngo*MIV and is involved in numerous hydrogen bonding and van der Waals interactions with AATT bases. Thus, it is possible that restriction enzymes that share common inner tetranucleotides in their recognition sequences utilize conserved structural mechanisms to interact with a conserved part of the target DNA site. However, this may not be a general rule. The comparison of structural mechanisms of sequence recognition by *Bgl*II (A/GATCT) and *Bam*HI (G/GATCC) published recently[17] revealed that both proteins display considerably different protein–DNA contacts at common target base pairs.

In addition to the 16 direct *Ngo*MIV–DNA and two water mediated hydrogen bonds that satisfy the hydrogen bonding potential of bases located in the major groove, numerous interactions with the sugar-phosphate backbone, including six direct and eight water-mediated contacts to DNA phosphates, complete the intricate network of interactions within the recognition site of *Ngo*MIV (Fig. 2*a*). It is noteworthy that most of the contacts to the backbone phosphates of one half site come from the neighboring subunit.

Interestingly, there is one direct protein–base contact outside the 6 bp recognition sequence, which may explain the possible CG preference for the bases flanking the *Ngo*MIV recognition sequence. Ser 36 of each subunit A, B and C makes a single hydrogen bond to Cyt 4 of the corresponding DNA strands E, F, and H. The conformations of Ser 36, however, differ slightly between subunits. Ser 36 of monomer D thus makes a water mediated contact to the backbone phosphate of the complementary Cyt 3′ in strand H instead of direct contacts to the C base. Furthermore, the backbone oxygens of Ser 36 in subunits B and C are involved in water mediated contacts to the outermost guanine.

Active site architecture

Structural studies revealed conserved features of the divalent metal ion binding sites at the catalytic centers of *Eco*RI, *Eco*RV, *Pvu*II, *Bgl*I and *Mun*I. Three charged residues (two acidic and a Lys) are located at the ends of the β-sheet in the vicinity of the scissile phosphate. A sequence motif ($PDX_{9–18}$(E/D)YK; where X is any residue and Y is a hydrophobic one) corresponds to the structurally conserved active sites, although its conservation is weak; *Bam*HI has a Glu residue instead of Lys at the structurally equivalent position[10]. In *Cfr*10I[17] a Ser is present instead of the second acidic residue in the $PDX_{9–18}$(E/D)YK motif. Experimental data indicate that despite its different location in the linear sequence, Glu 204 of *Cfr*10I spatially substitutes for the second acidic residue in the PDX_{9-18}(E/D)YK motif[18]. Thus, in *Cfr*10I the sequence motif $133PDX_{55}KX_{13}E$ corresponds to the active site, which is structurally similar to those of other restriction enzymes, suggesting the importance of spatial, rather than sequence, conservation. Of note is an additional acidic residue that is presumably involved in the coordination of the second metal ion, and that has been identified at the active sites of *Eco*RV, *Bam*HI, *Cfr*10I and *Bgl*I restriction enzymes; however, it is not conserved in *Eco*RI and *Mun*I. Another variation of the canonical active site motif of restriction enzymes different from that described for *Ngo*MIV/*Cfr*10I has been reported recently for the *Bgl*II endonuclease[16].

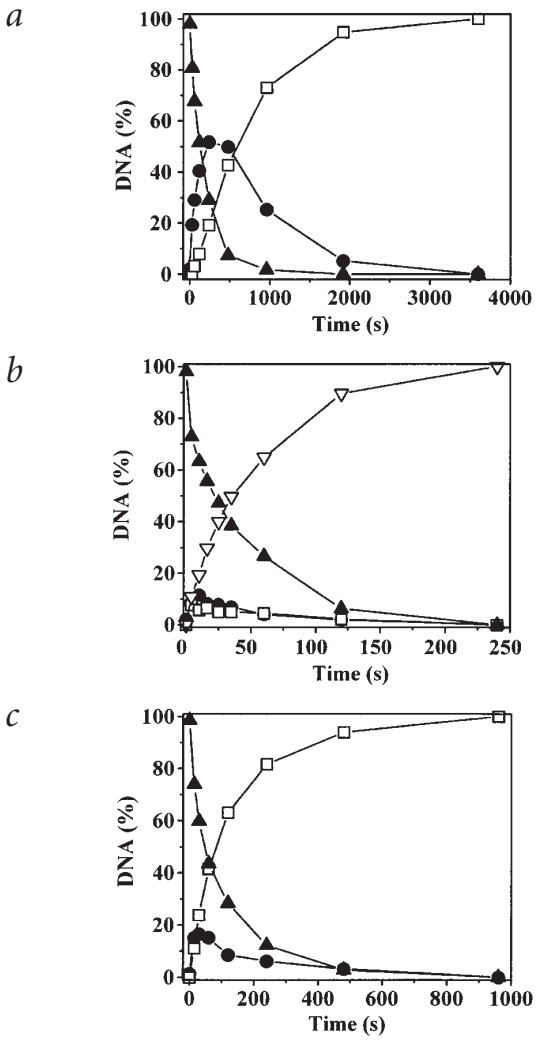

Fig. 4 Cleavage of plasmid DNA by *Ngo*MIV. ***a***, Cleavage of supercoiled plasmid pUCGC1 bearing a single recognition site of *Ngo*MIV. Reaction mixtures contained 2.3 nM pUCGC1, 50 nM *Ngo*MIV (as a tetramer), 10 mM $(CH_3COO)_2Mg$, 33 mM Tris-acetate (pH 8.0, 25 °C) and 66 mM CH_3COOK at 25 °C. The amounts of supercoiled (triangles), open circle (circles) and linear (squares) forms of pUCGC1 are shown. ***b***, Cleavage of supercoiled plasmid pUCGC2 bearing two recognition sites of *Ngo*MIV. Reaction mixtures contained 1.15 nM pUCGC2 (2.3 nM of target sequence), 50 nM *Ngo*MIV (as a tetramer) and other components as in (*a*). Amounts of supercoiled (filled triangles), open circle (circles), linear with a single double-stranded break (squares) and linear with two double stranded breaks (empty triangles) forms of pUCGC2 are shown. ***c***, Cleavage of supercoiled plasmid pUCGC1 bearing a single recognition site of *Ngo*MIV in the presence of cognate oligodeoxynucleotide. Reaction mixtures contained 2.3 nM pUCGC1, 200 nM specific self-complementary 12 bp duplex 5′-GAGGCCGGCCTC-3′, 50 nM *Ngo*MIV (as a tetramer), 10 mM $(CH_3COO)_2Mg$, 33 mM Tris-acetate (pH 8.0, 25 °C) and 66 mM CH_3COOK at 25 °C. Amounts of supercoiled (triangles), open circle (circles) and linear (squares) forms of pUCGC1 are shown.

The presence of two Mg^{2+} ions near the cleaved phosphate in the enzyme–product complex allowed us to unambiguously identify the catalytic/metal binding site of *Ngo*MIV (Fig. 3*a*). There are two Mg^{2+} ions in the active center, denoted A (low B-factor, 11.0–15.0 Å^2) and B (high B-factor, 22.0–31.0 Å^2). Both Mg^{2+} ions exhibit octahedral coordination. The O2P of the 5′ phosphate, the carboxylate of Asp 140 and the acetate molecule (from the reservoir of the crystallization buffer) contribute to the coordination of both Mg^{2+} ions. The acetate molecule, although not a prerequisite for catalysis (G.S. and V.S., unpublished results), probably stabilizes the enzyme product complex by keeping both metal ions together and therefore preventing their dissociation. The remaining three ligands of Mg^{2+} ion A are the backbone oxygen of Cys 186 and water molecules 1 and 2. Water molecules 3–5 complete the octahedral coordination of Mg^{2+} ion B. The ligand oxygen atoms are within 2.2–2.3 Å of Mg^{2+} ion A, and within 2.1–2.4 Å of Mg^{2+} ion B. Mg^{2+} ion B was substituted by Gd^{3+} in the heavy atom derivative search. All Mg-coordinating waters are fixed by either protein and/or DNA residues. Waters 1 and 2 are within hydrogen bonding distance of Nε of Lys 187, O1P of the adjacent Cyt 1L and the carboxylate of Glu 70. Water molecules 3 and 4 are hydrogen bonded to the carboxylate group of Glu 201 while water 5 is in contact with released the O3′ hydroxyl group of Gua 3 (3.3 Å), the carboxylate of Asp 140 (3.0 Å) and the O5T oxygen of the 5′ phosphate (3.2 Å).

The spatial locations of the active site residues of *Ngo*MIV are identical to those of *Cfr*10I (Fig. 3*b*). Interestingly, Glu 201 of *Ngo*MIV superimposes with Glu 204 of *Cfr*10I. Both Glu residues are located on α-helices while equivalent acidic residues at the active sites of other restriction enzymes are located at the ends of β-strands. Consequently, the sequence motif $139PDX_{46}KX_{15}E$ corresponds to the *Ngo*MIV active site. Thus, both *Ngo*MIV and *Cfr*10I show a structurally similar organization of their catalytic/metal binding sites, which differs from those of other restriction enzymes in linear sequence position of the active site residues. The metal chelating residues Glu 201 of *Ngo*MIV and Glu 204 of *Cfr*10I are located on the structurally equivalent helices that also contribute to the tetramerization interface and might be involved in the crosstalking interactions between dimers in the tetrameric restriction enzymes.

Type II restriction enzymes require only Mg^{2+} as a cofactor to catalyze the hydrolysis of DNA phosphodiester bonds to leave free 5′ phosphate and 3′ hydroxyl groups. Several mechanisms of catalysis, which differ in the number of metal ions involved and the moiety that activates the attacking water molecule, have been proposed[19–21]. In the *Ngo*MIV–product complex two Mg^{2+} ions separated by 3.7 Å are present at the active site (Fig. 3*a*). They are located symmetrically on both sides of the 5′ phosphate in the direction parallel to the former scissile bond. The location of the metal ions of *Ngo*MIV resembles the location of the metal ions at the active site of the *Bam*HI restriction enzyme[22] and the 3′-5′ exonuclease domain of DNA polymerase I[23]. Indeed, Mg^{2+} ions at the active site of *Ngo*MIV and Mn^{2+} ions at the active site of *Bam*HI can be superimposed in their enzyme–product complexes (Fig. 3*c*). Of note is the fact that four acidic residues (Glu 77, Asp 94, Glu 111 and Glu 113) are located in the vicinity of Mn^{2+} ions at the active site of *Bam*HI, while only three (Glu 70, Asp 140, Glu 201) are present at the active site of *Ngo*MIV. The Asp 140 of *Ngo*MIV and Asp 94 in *Bam*HI superimpose and both are involved in bridging interactions with metal ions. The Glu 201 of *Ngo*MIV and Glu 111 of *Bam*HI seem to be spatially equivalent and coordinate to the metal ions through the water molecules despite their different structural locations. The same is true for Glu 70 of *Ngo*MIV and Glu 77 of *Bam*HI.

The two metal ion mechanism of catalysis was proposed for the *Bam*HI restriction enzyme[21]. The structural properties of *Ngo*MIV are consistent with the two metal ion mechanism, with differences in some details. In the case of *Bam*HI, the attacking

water molecule located in the coordination sphere of the metal ion becomes activated by the Glu 113 carboxylate. In *Ngo*MIV, however, Lys 187 ocupies the spatially equivalent position. In our structure of the complex of *Ngo*MIV with cleaved DNA, water molecule 1 (Fig. 3*a*) is coordinated to the amino group of Lys 187 and would be in a geometrically favorable position for the in-line attack of the phosphorous atom at the scissile phosphodiester bond. This position might be occupied by the attacking water molecule in a *Ngo*MIV–substrate complex. However, one should keep in mind that interactions observed in the *Ngo*MIV–product complex might not be maintained in the complex with substrate DNA. A detailed analysis of the reaction mechanism for *Ngo*MIV should thus wait at least for the elucidation of the crystal structure of *Ngo*MIV–substrate complex. Nevertheless, a Lys residue structurally similar to Lys 187 of *Ngo*MIV is conserved in the active sites of all restriction enzymes with known crystal structures except for *Bam*HI and *Bgl*II[16]. In the recent *Bgl*II structure, a Gln residue substitutes for the conserved active site Lys. Thus, basic, acidic and neutral side chains are located at the structurally equivalent position in the active sites of restriction enzymes, suggesting a possible mechanistic divergence. All these structurally conserved residues interact with a metal bound water molecule and their roles in the activation/positioning of the attacking nucleophile still needs to be elucidated.

Structure-function relationships

Equilibrium sedimentation analysis indicates that *Ngo*MIV is a tetramer in solution (C. Urbanke, pers. comm.). The crystallographic evidence presented above reveals that the *Ngo*MIV tetramer is composed of two dimers with a separate DNA binding site in each (Fig. 1*a*,*b*). Processing at these sites in principle may be independent or cooperative. Cleavage patterns of plasmids containing a single or two recognition sites provide a general test for whether the restriction enzyme acts independently on the individual site or acts simultaneously at two copies of the recognition site[4]. We therefore studied the *Ngo*MIV cleavage of supercoiled plasmids pUCGC1 and pUCGC2, containing a single or two copies of the recognition sequence 5′ GCCGGC, respectively.

To eliminate possible effects of substrate binding or product release on the hydrolysis rates, *Ngo*MIV cleavage of pUCGC1 was studied under single turnover reaction conditions at a saturating enzyme concentration (2.3 nM substrate, 50 nM enzyme). Under these conditions the cleavage of pUCGC1 by *Ngo*MIV (Fig. 4*a*) follows a sequential reaction pathway; supercoiled plasmid DNA is converted into the linear product *via* an open circle DNA intermediate. The reaction pattern of the *Ngo*MIV cleavage of the plasmid containing a single recognition site (Fig. 4*a*) is similar to those reported for dimeric restriction enzymes[24,25] except that the cleavage of plasmid DNA by *Ngo*MIV is very slow.

According to the crystal structure, the *Ngo*MIV tetramer is composed of two dimers with a separate DNA binding site in each and is, therefore, able to interact with two copies of the recognition sequence. The observed low level of activity of *Ngo*MIV with the pUCGC1 plasmid might have resulted either from the cleavage of the single copy pUCGC1 plasmid by one of the dimers forming the tetramer or from the small fraction of tetramers interacting with two separate copies of pUCGC1. The data presented in Fig. 4*a* do not allow us to strictly discriminate between those possibilities. Under conditions of substrate deficiency ([E] >> [S]) binding of two plasmids by the same *Ngo*MIV tetramer seems to be less likely, although one cannot exclude the possibility that the *Ngo*MIV tetramer, similarly to *Sfi*I[26], binds two copies of the recognition sites in a cooperative manner.

The cleavage profile of pUCGC2 (Fig. 4*b*) containing two recognition sites located in *cis* differs significantly from that of pUCGC1. First, at the same enzyme concentration the supercoiled form of pUCGC2 is cleaved more than 10-fold faster than pUCGC1. Second, most of the supercoiled pUCGC2 is directly converted into the final reaction product (linear DNA with two double-stranded breaks) and only a small fraction of open circle DNA and linear DNA cut at one site is formed. Similar increases in the cleavage rate of plasmids containing two recognition sites have been reported for the tetrameric restriction enzymes *Sfi*I[27] and *Cfr*10I[3]. It has been demonstrated that both *Sfi*I[28] and *Cfr*10I[3] interact simultaneously with both recognition sites located in *cis* through the DNA looping. We suggest that a similar model could be applied to *Ngo*MIV: in the case of pUCGC2, *Ngo*MIV simultaneously binds two copies of its recognition sequence and rapidly cleaves all four phoshodiester bonds at both sites during the lifetime of the enzyme–DNA complex.

The requirement of two recognition sites for effective DNA hydrolysis by *Ngo*MIV was further tested by cleavage experiments with the pUCGC1 plasmid in the presence of a short oligodeoxynucleotide containing or lacking the *Ngo*MIV recognition site (Fig. 4*c*). Indeed, addition of 200 nM of the 12 bp oligodeoxynucleotide containing the recognition sequence of *Ngo*MIV increased the cleavage rate of the supercoiled form of pUCGC1 and significantly decreased the amount of the open circle DNA intermediate (compare Fig. 4*a* and *c*). In contrast, the oligodeoxynucleotide lacking the recognition sequence of *Ngo*MIV had no effect on the cleavage rate of pUCGC1 (data not shown). The activation of pUCGC1 cleavage by its cognate oligodeoxynucleotide is consistent with *Ngo*MIV having a tetrameric organization (Fig. 1). We suggest that, in the presence of the specific oligodeoxynucleotide, one of the dimers of the *Ngo*MIV tetramer interacts with the recognition site on the plasmid while the other one interacts with the specific oligodeoxynucleotide and such interactions in *trans* increase the cleavage rate of pUCGC1. Activation in *trans* by cognate oligonucleotides of plasmid DNA cleavage has been reported for tetrameric restriction enzymes *Sfi*I[29] and *Cfr*10I[3].

Cleavage studies of plasmids containing a single and two recognition sites demonstrate that two recognition sites are required for effective DNA cleavage by *Ngo*MIV and are consistent with *Ngo*MIV having a tetrameric architecture. Further mechanistic studies of *Ngo*MIV and its putative dimeric mutants are necessary to shed light on the mechanism of crosstalking interactions between dimers.

Methods

Crystallization. The protein was dissolved to a final concentration of 0.2 mM in buffer containing 20 mM Tris/HCl, pH 7.5, 200 mM NaCl, 0.02% (w/v) NaN_3. The oligonucleotide was dissolved in a 50 mM NaCl solution and annealed (heated to 95 °C and then cooled down overnight to room temperature). The protein and the oligonucleotide solutions were mixed to produce a stoichiometric complex solution (the molar protein to oligonucleotide ratio was 1:1). After a short centrifugation at 12,000 g to remove dust particles, 2 ml of the complex solution was mixed with 1 ml of the reservoir solution (25% (v/v) 2-methyl-2,4-pentanediol (MPD), 200 mM $MgCl_2$, 0.1 M 2-(N-morpholino) ethane sulphonate (MES) pH 6.5) in the depressions of the Cryschem® sitting drop crystallization plates. The crystals grew overnight at room temperature up to the size $0.5 \times 0.2 \times 0.1$ mm^3.

Table 1 Crystallographic data collection and refinement statistics

Crystal	NAT I	NAT II	Hg	Gd	I	Pt
Maximum resolution (Å)	2.8	1.6	3.3	3.4	3.0	3.0
Total number of reflections	126,655	2,553,379	113,063	100,647	153,143	139,677
Unique reflections[1]	27,705	160,612	19,707	17,869	25,816	24,955
Completeness (last shell) (%)	88.4 (91.7)	98.6 (88.7)	98.7 (99.8)	98.4 (99.9)	98.4 (99.1)	95.7 (94.3)
R_{merge} (%)[2]	9.4	4.7	19.3	20.4	13.3	13.6
R_{iso} (%)[3]			25.5	18.4	26.5	18.9
R_{cullis} ce / iso[4]			0.84	0.85	0.89	0.91
R_{cullis} (ac/iso) / ano[4]			0.84 / 0.95	0.81 / 0.91	0.87 / 0.91	0.90 / 0.91
Phasing power ce / iso[5]			0.58	2.29	1.13	1.08
Phasing power (ac/iso) / ano[5]			1.13 / 0.69	1.60 / 0.71	1.12 / 0.70	0.87 / 0.35
Refinement statistics (500–1.6 Å)						
Reflections ($F > 2\sigma(F)$)	158,154					
R_{cryst} (R_{free}) (%)[6]	17.4 (20.4)					
R.m.s. deviations						
Bond lengths (Å)	0.010					
Bond angles (°)	1.567					

[1]Only reflections with $I / \sigma(I) > 2.0$ were used for statistics.
[2]$R_{merge} = \sum_h \sum_i | I_{(h,i)} - \langle I_{(h)} \rangle | / \sum_h \sum_i I$, where $I_{(h,i)}$ is the intensity value of the ith measurement of h and $\langle I(h) \rangle$ is the corresponding value of h for all i measurements; the summation is over all measurements.
[3]$R_{iso} = \sum | F_P - F_{PH} | / \sum F_P$, where F_P and F_{PH} are the derivative and the native structure–factor amplitudes, respectively.
[4]$R_{cullis} = \sum |E| / \sum | F_{ph} - F_p |$, where E is lack of closure, F_{ph} and F_p are observed derivative and protein structure factors, respectively.
[5]Phasing power = $\langle | F_h | \rangle$ / r.m.s.(ε), where $\langle | F_h | \rangle$ is the mean calculated amplitude for the heavy atom model and r.m.s.(ε) is the root mean square lack of closure error for the isomorphous differences.
[6]$R_{cryst} = \sum | F_{obs} - F_{calc} | / \sum | F_{obs} |$, where F_{obs} and F_{calc} are the observed and calculated structure factor amplitudes, respectively; R_{free} was calculated using a random 5% of the reflection data that was omitted from all stages of the refinement.

X-ray data collection, phasing and refinement. The cocrystals belong to space group $P2_12_12_1$, with unit cell parameters a = 90.4 Å, b = 91.1 Å, c = 149.6 Å, and contain one complex in the crystallographic asymmetric unit. Data set NAT I and all the derivative data sets were collected at room temperature on a MAR research™ image plate mounted on a rotating anode generator RU200 at the CuKα wavelength of 1.5418 Å. After flash freezing, data set NAT II was collected at DESY, Hamburg (beamline BW6, λ = 1.0697 Å). Data were processed with the HKL package[30], and scaling and reduction was performed with the CCP4 package[31]. Phasing was done with SHARP[32] to a figure of merit (FOM) of 0.36 for the centric reflections, including all derivatives, and after solvent flattening with SOLOMON[33] ($FOM_{solomon} = 0.73$). Model building was performed with O[34] and refinement carried out with CNS[35] including bulk solvent correction. In the final model, all residues are in the allowed regions of the Ramachandran plot (90.3% in the most favored, 9.7% in the additional allowed regions) as calculated by the PROCHECK[36] program.

Cleavage of pUCGC1 and pUCGC2 plasmids by *Ngo*MIV. Plasmids pUCGC1 and pUCGC2 contain single and two recognition sites of *Ngo*MIV, respectively. The flanking sequences of *Ngo*MIV sites in the pUCGC1 and pUCGC2 plasmids were the same[3]. For pUCGC1, cleavage reaction mixtures contained 2.3 nM pUCGC1, 50 nM *Ngo*MIV (as a tetramer), 10 mM $(CH_3COO)_2Mg$, 33 mM Tris-acetate (pH 8.0, 25 °C) and 66 mM CH_3COOK at 25 °C. In the trans-activation studies of pUCGC1 cleavage, 200 nM of the specific self-complementary 12 bp duplex 5′-GAGGCCGGCCTC-3′ or the non-specific 12 bp duplex obtained by annealing the two 12 bp oligonucleotides 5′-GAGACCGGACTC-3′ and 3′-CTCTGGCCTGAG-5′ were added to the reaction mixture. For pUCGC2, cleavage reaction mixtures contained 1.15 nM pUCGC2 (2.3 nM of target sequence) and 50 nM *Ngo*MIV (as a tetramer) with the other components the same as used for pUCGC1 cleavage. Aliquots were removed after fixed time intervals, the reaction being stopped by adding a quarter volume of loading dye solution containing EDTA (50 mM EDTA, pH 8.0, 0.1% SDS, 50% glycerol, 0.01% bromphenol blue), and analyzed by agarose electrophoresis. Subsequent densitometric analysis of ethidium bromide stained gels using software provided by Ultra-Lum allowed us to quantify the amounts of the different DNA forms.

Coordinates. Coordinates have been deposited to the Protein Data Bank (accession code 1FIU).

Acknowledgments
We thank D.Stein for providing genes of *Ngo*MIV restriction-modification enzymes and R. Skirgaila for the purification of proteins. We also thank S. Halford for communicating results before publication and J. Richardson for critical reading of the manuscript. We thank G. Bourenkov and H. Bartunik at DESY, Hamburg, for assistance in making the high resolution measurements. The research in part was supported by NATO and the Lithuanian State Program for Biotechnology.

Received 4 April, 2000; accepted 31 July, 2000.

1. Roberts, R.J. & Macelis, D. REBASE: restriction enzymes and methylases. *Nucleic Acids Res.* **28**, 306–307 (2000).
2. Wentzell, L.M., Nobbs, T.J. & Halford, S.E. The *Sfi*I restriction endonuclease makes a four-strand DNA break at two copies of its recognition sequence. *J. Mol. Biol.* **248**, 581–595 (1995).
3. Siksnys, V., *et al.* The *Cfr*10I restriction enzyme is functional in a tetrameric form. *J. Mol. Biol.* **291**, 1105–1118 (1999).
4. Bilcock, D.T., Daniels, L.E., Bath, A.J. & Halford, S.E. Reactions of type II restriction endonucleases with 8-base pair recognition sites. *J. Biol. Chem.* **274** 36379–36386 (1999).
5. Topal, M.D., Thresher, R.J., Conrad, M. & Griffith, J. *Nae*I endonuclease binding to pBR322 DNA induces looping. *Biochemistry* **30**, 2006–2010 (1991).
6. Krüger, D.H., Barcak, G.J., Reuter, M. & Smith, H.O. *Eco*RII can be activated to cleave refractory DNA recognition sites. *Nucleic Acids Res.* **16**, 3997–4008 (1988).
7. Stein, D.C., Chien, R. & Seifert, H.S. Construction of a *Neisseria gonorrhoeae* MS11 derivative deficient in *Ngo*MI restriction and modification. *J. Bacteriol.* **174**, 4899–4906 (1992).
8. Aggarwal, A.K. Structure and function of restriction endonucleases. *Curr. Opin. Struct. Biol.* **5**, 11–19 (1995).
9. McClarin, J.A., *et al.* Structure of the DNA-*Eco*RI endonuclease recognition complex at 3 Å resolution. *Science* **234**, 1526–1541 (1986).
10. Newman, M., Strzelecka, T., Dorner, L.F., Schildkraut, I. & Aggarwal, A.K. Structure of *Bam* HI endonuclease bound to DNA: partial folding and unfolding on DNA binding. *Science* **269**, 656–663 (1995).
11. Deibert, M., Grazulis, S., Janulaitis, A., Siksnys, V. & Huber, R. Crystal structure of *Mun*I restriction endonuclease in complex with cognate DNA at 1.7 Å resolution. *EMBO J.* **18**, 5805–5816 (1999).
12. Kim, Y.C., Grable, J.C., Love, R., Greene, P.J. & Rosenberg, J.M. Refinement of *Eco*RI endonuclease crystal structure: a revised protein chain tracing. *Science* **249**, 1307–1309 (1990).
13. Winkler, F.K., *et al.* The crystal structure of *Eco*RV endonuclease and of its complexes with cognate and non-cognate DNA fragments. *EMBO J.* **12**, 1781–1795 (1993).
14. Cheng, X., Balendiran, K., Schildkraut, I. & Anderson, J.E. Structure of *Pvu*II endonuclease with cognate DNA. *EMBO J.* **13**, 3927–3935 (1994).
15. Newman, M., *et al.* Crystal structure of restriction endonuclease *Bgl*I bound to its interrupted DNA recognition sequence. *EMBO J.* **17**, 5466–5476 (1998).
16. Bozic, D., Grazulis, S., Siksnys, V. & Huber, R. Crystal structure of *Citrobacter freundii* restriction endonuclease *Cfr*10I at 2.15 Å resolution. *J. Mol. Biol.* **255**, 176–186 (1996).
17. Lucas, C.M., Kucera, R., Schildkraut, I. & Aggarwal, A. Understanding the immutability of restriction enzymes: crystal structure of *Bgl*II and its DNA substrate at 1.5 Å resolution. *Nature Struct. Biol.* **7**, 134–140 (2000).
18. Skirgaila, R., Grazulis, S., Bozic, D., Huber, R. & Siksnys, V. Structure-based redesign of the catalytic/metal binding site of *Cfr*10I restriction endonuclease reveals the importance of spatial rather than sequence conservation of active centre residues. *J. Mol. Biol.* **279**, 473–481 (1998).
19. Jeltsch, A., Alves, J., Wolfes, H., Maass, G. & Pingoud, A. Substrate-assisted catalysis in the cleavage of DNA by the *Eco*RI and *Eco*RV restriction enzymes. *Proc. Natl. Acad. Sci. USA* **90**, 8499–8503 (1993).
20. Kostrewa, D. & Winkler, F.K. Mg^{2+} binding to the active site of *Eco*RV endonuclease: a crystallographic study of complexes with substrate and product DNA at 2 Å resolution. *Biochemistry* **34**, 683–696 (1995).
21. Vipond, I.B., Baldwin, G.S. & Halford, S.E. Divalent metal ions at the active sites of the *Eco*RV and *Eco*RI restriction endonucleases. *Biochemistry* **34**, 697–704 (1995).
22. Viadiu, H. & Aggarwal, A.K. The role of metals in catalysis by the restriction endonuclease *Bam*HI. *Nature Struct. Biol.* **5**, 910–916 (1998).
23. Beese, L.S. & Steitz, T.A. Structural basis for the 3′-5′ exonuclease activity of *Escherichia coli* DNA polymerase I: a two metal ion mechanism. *EMBO J.* **10**, 25–33 (1991).
24. Erskine, S.G., Baldwin, G.S. & Halford, S.E. Rapid-reaction analysis of plasmid DNA cleavage by the *Eco*RV restriction endonuclease. *Biochemistry* **36**, 7567–7576 (1997).
25. Sasnauskas, G., Jeltsch, A., Pingoud, A. & Siksnys, V. Plasmid DNA cleavage by *Mun*I restriction enzyme: single-turnover and steady-state kinetic analysis. *Biochemistry* **38**, 4028–4036 (1999).
26. Embleton, M.L., Williams, S.A., Watson, M.A. & Halford, S.E. Specificity from the synapsis of DNA elements by the *Sfi*I endonuclease. *J. Mol. Biol.* **289**, 785–797 (1999).
27. Nobbs, T.J. & Halford, S.E. DNA cleavage at two recognition sites by the SfiI restriction endonuclease: salt dependence of cis and trans interactions between distant DNA sites. *J. Mol. Biol.* **252**, 399–411 (1995).
28. Wentzell, L.M. & Halford, S.E. DNA looping by the *Sfi*I restriction endonuclease. *J. Mol. Biol.* **281**, 433–444 (1998).
29. Wentzell, L.M., Nobbs, T.J. & Halford, S.E. The *Sfi*I restriction endonuclease makes a four-strand DNA break at two copies of its recognition sequence. *J. Mol. Biol.* **248**, 581–595 (1995).
30. Otwinowski, Z. & Minor, W. Processing of X-ray diffraction data collected in oscillation mode. *Methods Enzymol.* **276**, 307–326 (1996).
31. Collaborative Computational Project, Number4. The CCP4 Suite: programs for protein crystallography. *Acta Crystallogr. D* **50**, 760–763 (1994).
32. La Fortelle, E.d., Irwin, J.J. & Bricogne, G. SHARP: a maximum-likelihood heavy-atom parameter refinement and phasing program for the MIR and MAD methods. *Crystallogr. Computing* **7**, 1–9 (1997).
33. Abrahams, J.P. & Leslie, A.G.W. Methods used in the structure determination of the bovine mitochondrial F1 ATPase. *Acta Crystallogr. D* **52**, 30–42 (1996).
34. Jones, T.A., Zou, J.Y., Cowan, S.W. & Kjeldgaard, M. Improved methods for building protein models in electron density maps and the location of errors in these models. *Acta Crystallogr. A* **47**, 110–119 (1991).
35. Brünger, A.T., *et al.* Crystallography & NMR system: a new software suite for macromolecular structure determination. *Acta Crystallogr. D* **54**, 905–921 (1998).
36. Laskowski, R.A., McArthur, M.W., Moss, D.S. & Thornton, J.M. PROCHECK: a program to check the stereochemical quality of protein structures *J. Appl. Crystallogr.* **26**, 283–291 (1993)
37. Luscombe, N.M., Laskowski, R.A. & Thornton, J.M. NUCPLOT: a program to generate schematic diagrams of protein-nucleic acid interactions. *Nucleic Acids Res.* **25**, 4940–4945 (1997).