The EMBO Journal Vol.18 No.21 pp.5805–5816, 1999

Crystal structure of *Mun*I restriction endonuclease in complex with cognate DNA at 1.7 Å resolution

Markus Deibert[1,2], Saulius Grazulis[1,3], Arvydas Janulaitis[3], Virginijus Siksnys[2,3] and Robert Huber[1]

[1]Max-Planck-Institut für Biochemie, D-82152 Planegg-Martinsried, Germany and [3]Institute of Biotechnology, Lt-2028 Vilnius, Lithuania

[2]Corresponding authors
e-mail: deibert@biochem.mpg.de or Siksnys@ibt.lt

The *Mun*I restriction enzyme recognizes the palindromic hexanucleotide sequence C/AATTG (the '/' indicates the cleavage site). The crystal structure of its active site mutant D83A bound to cognate DNA has been determined at 1.7 Å resolution. Base-specific contacts between *Mun*I and DNA occur exclusively in the major groove. While DNA-binding sites of most other restriction enzymes are comprised of discontinuous sequence segments, *Mun*I combines all residues involved in the base-specific contacts within one short stretch (residues R115–R121) located at the N-terminal region of the $3_{10}4$ helix. The outer CG base pair of the recognition sequence is recognized solely by R115 through hydrogen bonds made by backbone and side chain atoms to both bases. The mechanism of recognition of the central AATT nucleotides by *Mun*I is similar to that of *Eco*RI, which recognizes the G/AATTC sequence. The local conformation of AATT deviates from the typical B-DNA form and is remarkably similar to *Eco*RI–DNA. It appears to be essential for specific hydrogen bonding and recognition by *Mun*I and *Eco*RI.

Keywords: crystal structure/*Mun*I/protein–DNA complex/restriction endonuclease

Introduction

Type II restriction enzymes recognize short, usually palindromic DNA sequences of 4–8 bp in length and cleave phosphodiester bonds in the presence of Mg^{2+} ions within or close to the recognized sequence. In response to the continuous demand for enzymes with new specificities, >3000 restriction endonucleases from a variety of bacteria were characterized (Roberts and Macelis, 1998). Due to these screening efforts, type II restriction enzymes became the largest family of functionally related enzymes, but less is known about how they achieve their specificity. More than 200 genes of restriction enzymes have been cloned and sequenced (Roberts and Macelis, 1998). Comparison of the deduced protein sequences revealed few or no similarities, suggesting that type II restriction enzymes might be structurally and functionally diverse. Crystallographic analysis of typical type II restriction enzymes *Eco*RI (McClarin *et al*., 1986; Kim *et al*., 1990), *Eco*RV (Winkler *et al*., 1993), *Pvu*II (Athanasiadis *et al*., 1994), *Bam*HI (Newman *et al*., 1995), *Cfr*10I (Bozic *et al*., 1996) and *Bgl*I (Newman *et al*., 1998), however, indicated considerable three-dimensional similarity despite the lack of sequence homologies. The common core motif consisting of a mixed five-stranded β-sheet flanked by two α-helices appeared to be characteristic for type II restriction enzymes (Aggarwal, 1995). Structural comparisons also revealed that restriction enzymes exhibiting a similar cleavage pattern, e.g. sticky end cutters *Eco*RI and *Bam*HI or blunt end cutters *Eco*RV and *Pvu*II, share extensive structural homologies that extend beyond the common core motif. Thus, type II restriction enzymes can be subdivided into structural subclasses in accordance with the DNA cleavage pattern. The detailed mechanism of DNA recognition by restriction enzymes within a single subclass seems to be different. However, the structural data currently available for the restriction enzymes are limited, and systematic studies are required to understand the possible diversity of the mechanisms of sequence recognition employed by type II restriction enzymes. In order to address this question, we have focused on the elucidation of the structural and molecular mechanisms of DNA sequence recognition by type II restriction enzymes recognizing partially overlapping nucleotide sequences within the same subclass of restriction enzymes.

The *Mun*I restriction enzyme from *Mycoplasma unidentified* recognizes the palindromic hexanucleotide sequence C/AATTG and cleaves it as indicated by the / (Stakenas *et al*., 1992). The recognition sequence of *Mun*I partially overlaps with that of *Eco*RI (G/AATTC). *Eco*RI was the first restriction enzyme characterized by X-ray crystallography that provided us with a molecular model of sequence discrimination (McClarin *et al*., 1986; Kim *et al*., 1990). Comparison of the protein sequences of *Mun*I and *Eco*RI revealed local similarities that mapped to the structural elements of *Eco*RI involved in catalysis and recognition of the AATT tetranucleotide (Siksnys *et al*., 1994). Solution of the crystal structure of *Mun*I and direct comparison of the molecular mechanism of sequence discrimination employed by *Mun*I and *Eco*RI therefore became an intriguing challenge.

Here we present the crystal structure of the D83A mutant of *Mun*I in complex with a DNA decamer containing its recognition sequence determined at 1.7 Å resolution. Crystallographic analysis revealed that both *Mun*I and *Eco*RI share a similar mechanism of the recognition of the central AATT tetranucleotide, while the mechanisms of the external base pair discrimination differ. Interestingly, the DNA in the complex with *Mun*I was significantly distorted at the central base pair, suggesting the importance of local DNA conformation in the mechanism of DNA recognition.

Results and discussion

Structure determination

Our efforts to obtain *Mun*I–DNA co-crystals mostly yielded crystals that diffracted just below 3 Å resolution and were not suitable for the structure solution. Biochemical studies revealed that the ability of *Mun*I to form stable protein–DNA complexes depends on the reaction conditions (pH, Ca^{2+} ions) (Lagunavicius *et al*., 1997). It was supposed that the ionization state of one of the putative active site carboxylate residues D83 or E98 was critical for the cognate DNA binding. Indeed, replacement of the D83 residue of *Mun*I by alanine generated a catalytically inactive mutant that formed tight sequence-specific complexes independently of the pH or the presence of metal ions (Lagunavicius and Siksnys, 1997). The co-crystallization trials of the D83A mutant with the decameric DNA (Figure 1) yielded highly ordered crystals.

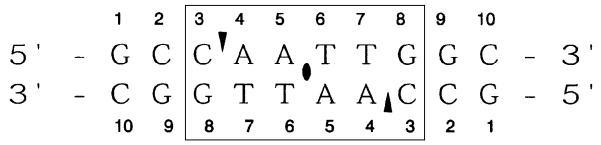

Fig. 1. DNA sequence of the decameric oligonucleotide used in the *Mun*I D83A–DNA complex. The recognition sequence is boxed. The cleavage sites are indicated by arrows. The position of the crystallographic 2-fold axis is shown.

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 of the DNA to be built. The structure has been refined to 1.7 Å resolution with excellent statistics and stereochemistry (Table I). A total of 646 solvent molecules were included in the final refined model.

Protein monomer structure

The *Mun*I restriction endonuclease consists of 202 amino acids, and with core dimensions of ~25 × 44 × 50 Å is one of the smallest proteins among the structurally characterized restriction enzymes (Figure 2A). The *Mun*I crystal contains a dimer and a single copy of double-stranded decamer in the asymmetric unit (Figure 2B).

The experimental electron density map was of excellent quality which allowed us to build nearly the whole model including side chains, except for the N-termini of both monomers which are flexible and could not be traced (residues 6–202 in monomer A and residues 3–202 in monomer B are visible). The monomer of *Mun*I displays an α/β architecture. A mixed five-stranded β-sheet is markedly curved and exhibits a left-handed twist of ~80°. On the concave side of the sheet, two N-terminal helices

Table I. Crystallographic data collection and refinement statistics

Data set	NAT I	NAT II	Iod I	Iod II	Pt	Os	Ta	Gd
Space group	$P2_12_12_1$	$P2_12_12_1$	$P2_12_12_1$	$P2_12_12_1$	$P2_12_12_1$	$P2_12_12_1$	$P2_12_12_1$	$P2_12_12_1$
Cell dimensions	*a* = 58.4 Å, *b* = 72.9 Å, c = 122.5 Å	*a* = 58.4 Å, *b* = 72.5 Å, *c* = 121.5 Å						
Resolution (Å)	2.3	1.7	3.0	3.2	2.5	2.5	2.4	3.0
Observed reflections[a]	80 635	177 834	30 193	31 181	59 529	55 354	54 170	31 048
Unique reflections	23 128	54 135	10 017	9105	18 607	18 452	20 157	10 640
Completeness (%)	97.0	93.6	92.4	99.7	97.5	98.6	94.7	97.5
overall (final shell)	(95.4)	(84.0)	(84.2)	(100)	(96.1)	(97.0)	(95.9)	(99.5)
R_{merge} (%)[b]	11.2	5.2	13.3	17.1	7.0	5.1	9.1	11.4
R_{iso} (%)[c]			22.7	21.2	19.6	14.4	16.5	15.9
Binding sites			2	2	2	4	1	1
R_{cullis} (centric/acentric)[d]			0.80/0.79	0.78/0.77	0.86/0.88	0.95/0.96	0.98/0.98	1.00/0.99
Phasing power (centric/acentric)[e]			0.85/1.30	0.93/1.40	0.68/0.91	0.39/0.54	0.24/0.37	0.15/0.24
FOM[f]: MIRAS/solvent flattened	0.42/0.72							
Refinement range	8.0–1.7 Å							
R_{cryst} (R_{free}) (%)[g]	17.97 (21.31)							
No. of solvent molecules	646							
R.m.s. bond length (Å)	0.009							
R.m.s. bond angles (°)	1.49							
Average *B*-factors (Å^2)								
Main chain	12.7							
Side chain	15.1							
Solvent	28.9							
DNA	12.9							

[a]Only reflections with $I/\sigma(I) > 2.0$ were used.
[b]$R_{merge} = (\sigma_h\sigma_i [|I(h, i) - \langle I(h)\rangle|]/\sigma_h\sigma_i I$, where $I(h, i)$ is the intensity value of the *i*th measurement of *h* and $\langle I(h)\rangle$ is the corresponding value of *h* for all *i* measurements, the summation is over all measurements.
[c]$R_{iso} = \sigma|F_P - F_{PH}|/(\sigma F_P$, where F_P and F_{PH} are the derivative and the native structure factor amplitudes, respectively.
[d]$R_{cullis} = \sigma_h|F_{h(obs)} - F_{h(calc)}|/\sigma_h F_{h(obs)}$
[e]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.
[f]FOM = figure of merit.
[g]$R_{cryst} = \sigma|F_{obs} - F_{calc}|/\sigma|F_{obs}|$, where F_{obs} and F_{calc} are the observed and calculated structure factor amplitudes, respectively; R_{free} was calculated using a random 5% of reflection data that were omitted from all stages of the refinement.

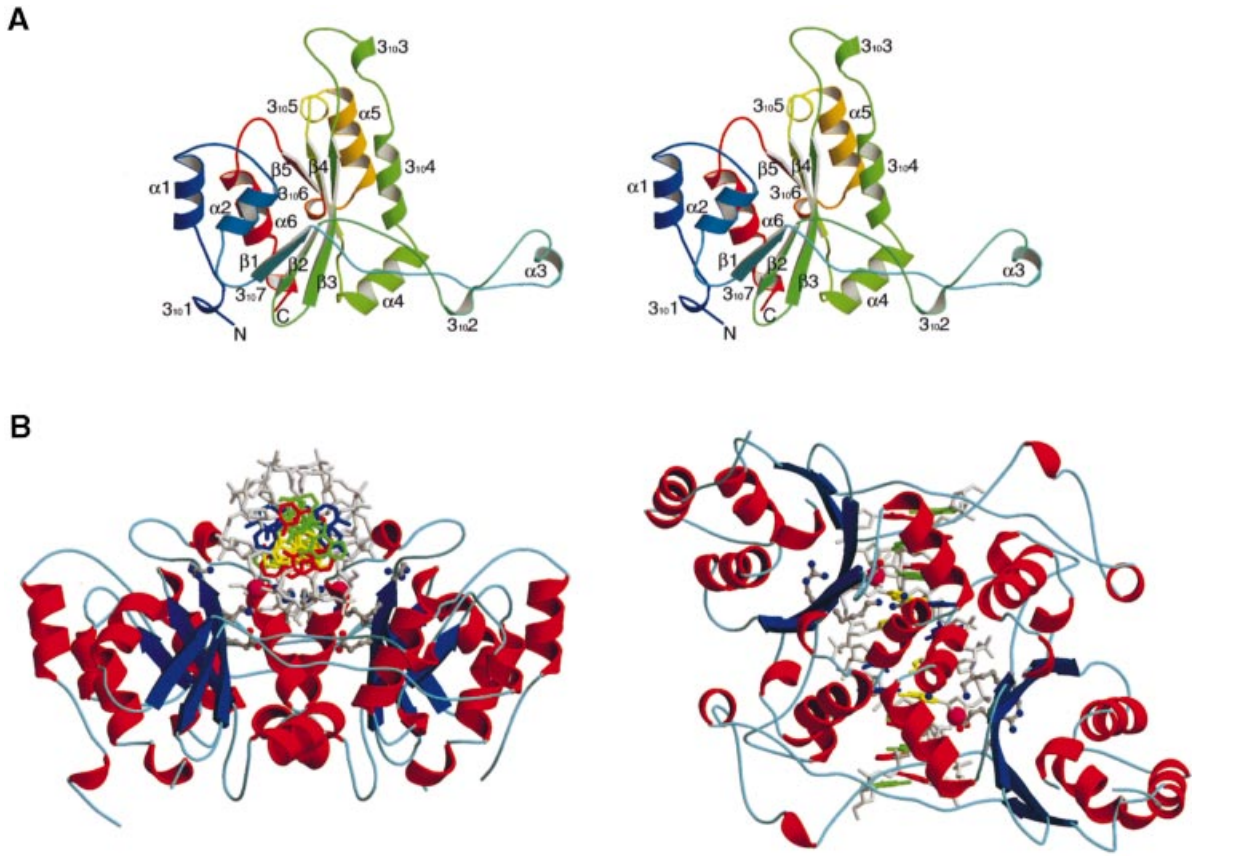

Fig. 2. (**A**) Stereo view of the *Mun*I monomer with labelled secondary structure elements. The chain runs from blue (N-terminus, N) to red (C-terminus, C). In the centre is the curved five-stranded β-sheet which spatially separates the three-helix bundle on the concave side from the DNA recognition elements on the convex side. The dimerization arm containing helices $3_{10}2$ and α3 protrudes away from the monomer. (**B**) Structure of the *Mun*I D83A dimer bound to cognate DNA viewed along the DNA axis (left), with the local 2-fold axis running vertically through the centre of the complex. α-helices and 3_{10} helices are shown in red, β-strands in blue, DNA backbone in grey, and the bases are coloured according to their type, i.e. G in green, C in red, T in blue and A in yellow; active site residues are shown in ball-and-stick representation, and scissile phosphate in magenta. View from the back with DNA behind the protein (right). Produced using MOLSCRIPT (Kraulis, 1991) and RASTER3D (Merrit and Murphy, 1994).

α1 and α2 pack against the C-terminal helix α6 forming a three-helix bundle. All other helices except for two at the N- and C-termini (helix $3_{10}7$ in monomer A and helices $3_{10}1$ and $3_{10}7$ in monomer B, respectively) are located at the convex side of the sheet. Helices $3_{10}4$ and α5 act as crossovers between β3 and β4 strands and β4 and β5 strands, respectively. The N-terminus of the $3_{10}4$ helix is directed towards the major groove of the DNA. Amino acid residues located at the N-terminus of the $3_{10}4$ helix or just upstream of it make contacts with the DNA bases within the *Mun*I recognition sequence. Helices α3 and $3_{10}2$ are part of the loop which protrudes ~25 Å away from the core of the protein and is involved in dimerization. It is noteworthy that all α-helical elements located on the convex side of the β-sheet are involved in either DNA recognition or dimerization.

Dimerization mode

The dimerization mode of restriction enzymes determines the position of the active sites within the dimer and thus the DNA cleavage pattern. Crystallographic analyses revealed structural similarities between restriction enzymes related by the DNA cleavage pattern and allowed them to be classified into three subfamilies. *Eco*RI, *Bam*HI and *Cfr*10I (McClarin *et al*., 1986; Kim *et al*., 1990; Newman *et al*., 1995; Bozic *et al*., 1996) belong to the family of restriction enzymes which produce DNA fragments with 5′ overhangs, whereas *Eco*RV and *Pvu*II (Winkler *et al*., 1993; Cheng *et al*., 1994) are members of the family of blunt end cutters. Only a single representative structure for *Bgl*I (Newman *et al*., 1998), which produces 3′ overhanging DNA fragments, is known to date. All these subfamilies differ with respect to the dimerization mode and structural elements involved in the intersubunit contacts.

*Mun*I produces fragments with 5′ overhangs after DNA cleavage. Three different regions of *Mun*I are involved in contacts between the monomers. Helices $3_{10}4$ and α5 are located at the subunit interface of the *Mun*I dimer. Two $3_{10}4$ helices coming from different subunits cross over one another at the R121 residue at an angle of ~30°. Additionally, the N-terminus of helix α4 of one monomer comes close to the C-terminus of helix α5 of the 2-fold symmetry-related monomer, and vice versa, forming a head-to-tail contact between the two helices. The loop connecting strands β1 and β2 (residues 50–79) of *Mun*I protrudes ~25 Å away from the core of the protein and spans across the neighbouring subunit. The α3 helix located on this loop is placed in parallel to the α5 helix of the symmetry-related subunit in a clamp-like fashion. The amino acid residues located on the structural elements involved in dimerization make a number of van der Waals

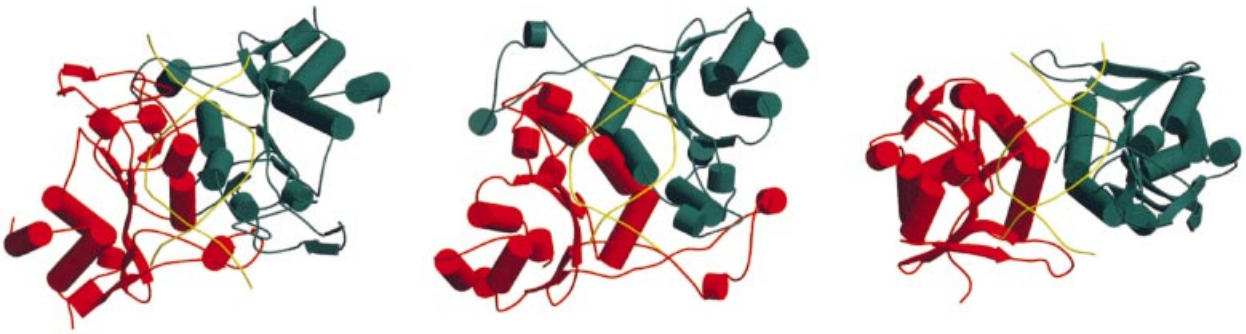

Fig. 3. Comparison of the dimerization modes of *Eco*RI (left), *Mun*I (middle) and *Bam*HI (right) in their complexes with cognate DNA in front of the dimers facing the viewer. Monomer subunits are displayed in orange and blue, helices as rods and strands as broad arrows. The DNA is shown in backbone representation as two yellow rods.

contacts, hydrogen bonds and salt bridges that contribute to the dimer stability. A surface of nearly 2300 Å^2 per monomer is buried at the interface between the two monomers of *Mun*I.

The dimerization mode of *Mun*I is strikingly similar to that of *Eco*RI and *Bam*HI (Figure 3). Four helices dominate at the dimer interface of *Mun*I, *Eco*RI and *Bam*HI. The $3_{10}4$ and α5 helices of *Mun*I located at the dimer interface are structurally equivalent to the α4 and α5 helices of *Eco*RI and α4 and α6 of *Bam*HI, respectively. Moreover, the $3_{10}4$ helices of *Mun*I, the α4 helices of *Eco*RI and the α4 helices of *Bam*HI of symmetry-related subunits cross over one another in an X-like fashion at a similar angle.

Although there is structural similarity between the four central helices of these three enzymes, their arrangement at the dimer interface differs: whereas in *Eco*RI and *Bam*HI they are organized as a four-helix bundle, in *Mun*I helix α5 does not cross helix $3_{10}4$ of the symmetry-related subunit but forms a head-to-tail contact to helix α4 of the other monomer. In addition to the four central helices, a loop located between structurally conserved β-strands is involved in the intersubunit contacts in *Mun*I, *Eco*RI and *Bam*HI dimers. The length, orientation and conformation of the loop differ between individual proteins; however, its topological position with respect to the β-sheet is conserved.

Comparison with other restriction endonucleases

Although there is little similarity at the primary sequence level, the structures of type II restriction endonucleases seem to be related to each other (Figure 4). A substructure consisting of a five-stranded β-sheet sandwiched between two helices was found to be common to those restriction enzymes with known structures (Aggarwal, 1995).

Spatial alignment reveals that the mixed five-stranded β-sheet (β1–β5) and two helices ($3_{10}4$ and α5) of *Mun*I overlap with structurally equivalent elements of *Eco*RI (β1–β5, and α4 and α5), *Bam*HI (β3–β7, and α4 and α6) and *Cfr*10I (β3–β7, and α7 and α8) (McClarin *et al*., 1986; Kim *et al*., 1990; Newman *et al*., 1995; Bozic *et al*., 1996) with an r.m.s. deviation of 3.9 Å (44 C_α), 2.1 Å (43 C_α) and 2.5 Å (34 C_α), respectively (Figure 5A–C).

The large r.m.s. difference between common core motifs of *Mun*I and *Eco*RI is due to the mixed β-sheet being more curved in *Mun*I. As well as the common core motif, *Mun*I shares with a *Eco*RI a topologically similar three-helix bundle (α1, α2 and α6 in both enzymes) which keeps the N- and C-termini in close proximity. In *Bam*HI, helices α2 and α3 are structurally equivalent to helices α3 and α6 of *Mun*I. In *Cfr*10I, only the C-terminal end of helix α3 is structurally equivalent to α2 in *Mun*I.

*Mun*I and *Eco*RI recognize related hexanucleotide sequences that differ only in the external base pair. Despite the overall fold similarities, both enzymes exhibit marked structural differences. In the *Eco*RI–DNA complex, two long loops (inner and outer arms) extend from the core of the protein and surround the DNA (McClarin *et al*., 1986; Kim *et al*., 1990). The residues located at the C-terminal ends of these loops are involved in the direct contacts with DNA bases in the recognition sequence of *Eco*RI. In the case of the *Mun*I–DNA complex, the outer arm is missing while the inner arm is considerably shorter and has a different conformation. The loop of *Mun*I (residues 50–79) involved in dimerization spans across the symmetry-related subunit and has a different orientation compared with the corresponding loop of *Eco*RI.

Despite overall structural differences, restriction enzymes producing 5′ overhangs upon DNA cleavage such as *Eco*RI and blunt end cutters such as *Eco*RV share a structurally similar substructure (Venclovas *et al*., 1994). Comparison of the *Mun*I and *Eco*RV structures (Winkler *et al*., 1993) revealed that 51 C_α atoms can be superimposed with an r.m.s. deviation of 2.7 Å (Figure 5D). This includes the five-stranded β-sheet (βc–βe, βg and βh in *Eco*RV and β1–β5 in *Mun*I), an α-helix (the C-terminus of αB in *Eco*RV and α2 in *Mun*I) and two loop regions which connect the α-helix with the first β-strand of the sheet (αB–βc in *Eco*RV and α2–β1 in *Mun*I) and the second to the third strand of the β-sheet (βd –βe in *Eco*RV and β2–β3 in *Mun*I). Furthermore, helices α6 of *Mun*I and αA of *Eco*RV are structurally equivalent.

Active site architecture

Mapping of the conserved sequence regions in *Mun*I and *Eco*RI restriction enzymes to the known X-ray structure of *Eco*RI allowed us to identify a sequence motif 82PDX$_{14}$EXK as a putative catalytic/Mg ion-binding site of *Mun*I (Siksnys *et al*., 1994). Subsequent mutational analysis supported the importance of residues D83, E98 and K100 of *Mun*I in catalysis (Lagunavicius and Siksnys, 1997). Crystal structure analysis of the D83A mutant of *Mun*I reveals that the side chains of glutamate E98 and lysine K100 residues are indeed located in close vicinity to the scissile phosphate (Figure 6).

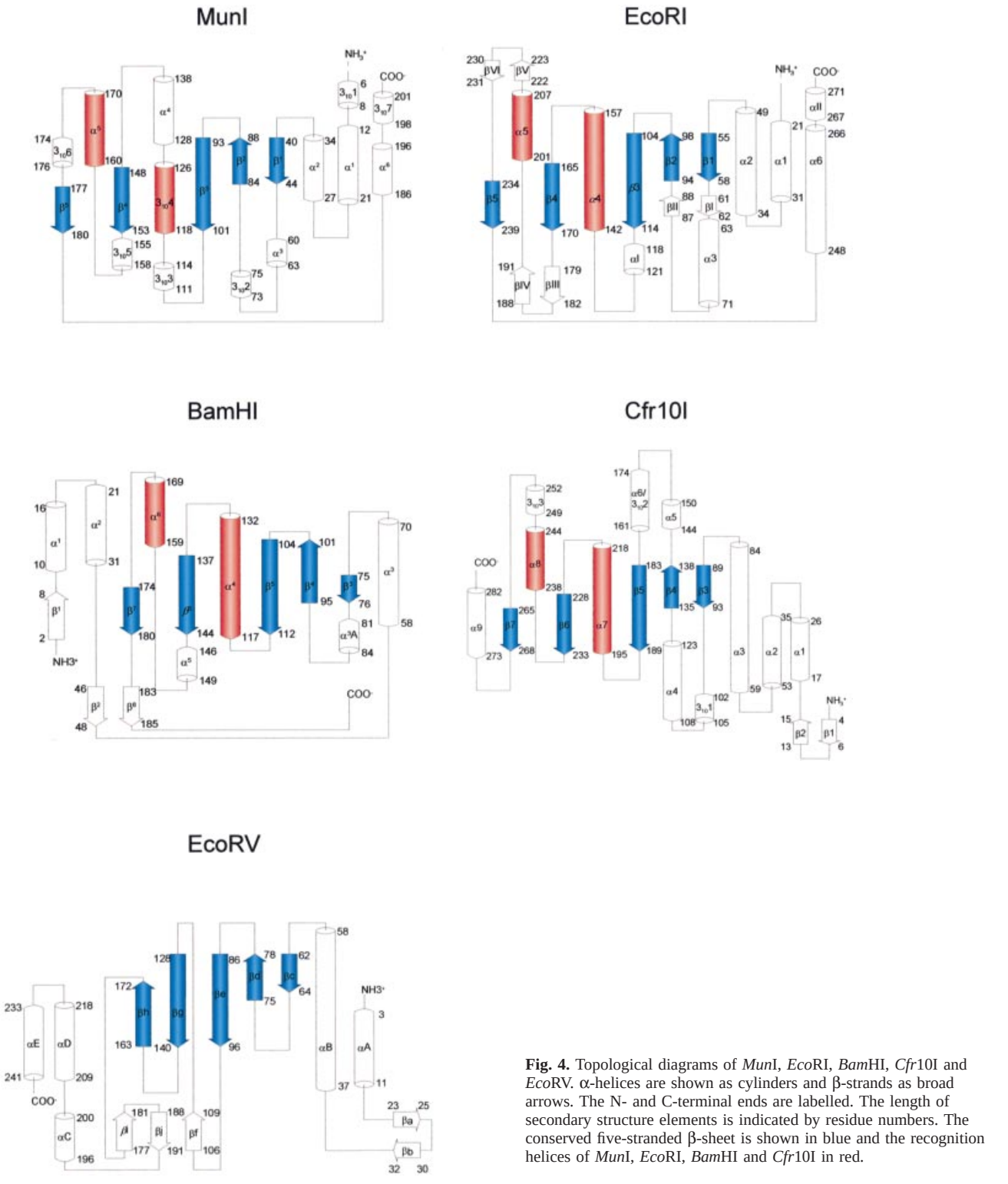

Fig. 4. Topological diagrams of *Mun*I, *Eco*RI, *Bam*HI, *Cfr*10I and *Eco*RV. α-helices are shown as cylinders and β-strands as broad arrows. The N- and C-terminal ends are labelled. The length of secondary structure elements is indicated by residue numbers. The conserved five-stranded β-sheet is shown in blue and the recognition helices of *Mun*I, *Eco*RI, *Bam*HI and *Cfr*10I in red.

Not surprisingly, residues E98 and K100 of *Mun*I superimpose with residues E111 and K113 at the catalytic/metal-binding site of *Eco*RI. The crystals of the D83A mutant of *Mun*I were grown in the presence of 50 mM $CaCl_2$; however, a Ca^{2+} ion was not present at the active site. Replacement of aspartate D83 by alanine probably decreases the metal-binding affinity and suggests that at least two acidic residues are required for the chelation of metal ion at the active site of *Mun*I. The putative Mg ion position in *Mun*I is occupied by a tightly bound water molecule that is hydrogen bonded to the OE1 of E98, the pro-R-oxygen of the scissile phosphate, the backbone NH of I99 and NZ of K100. It is interesting to note that the NZ atom of K100 and the NH1 and NH2 atoms of R121 make hydrogen bonds with pro-S-oxygen at the scissile phosphate. The side chains of residues K100 and R121 can adopt slightly different conformations in different monomers of *Mun*I, as indicated by higher *B*-factor values

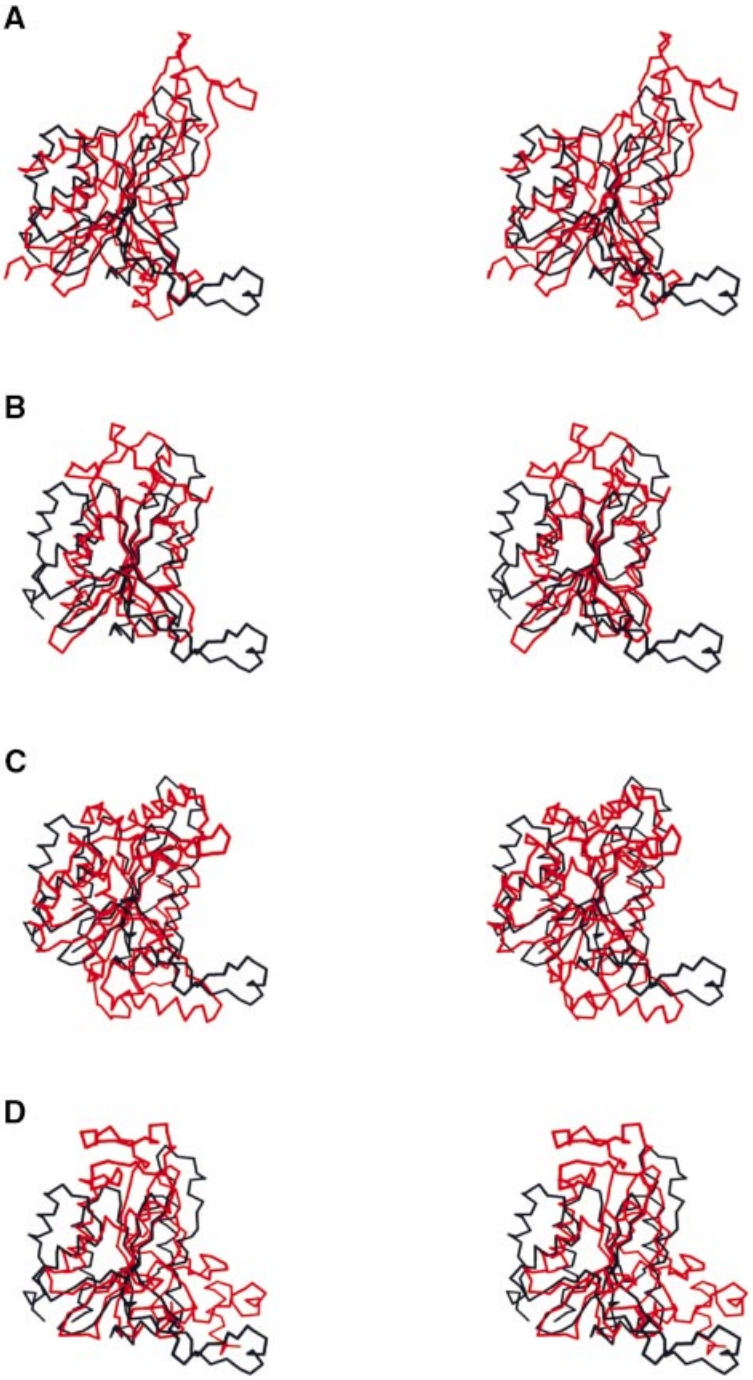

Fig. 5. Stereo views showing superpositions of C_α chains of *Mun*I on (**A**) *Eco*RI, (**B**) *Bam*HI, (**C**) *Cfr*10I and (**D**) *Eco*RV. *Mun*I is shown in black.

compared with neighbouring residues. In monomer A, the NZ atom of K100 is close to the water molecule which occupies the putative Mg-binding site, while in the second monomer it points away from the structurally equivalent water. In both monomers, K100 hydrogen-bonds the pro-S-oxygen of the scissile phosphate. The differences in the side chain conformations of R121 in different monomers did not change contacts to the pro-S-oxygen of the scissile phosphate, but in one orientation the single hydrogen bond to the E120 residue of the symmetry-related monomer has been lost.

The pro-R-oxygen is not contacted by protein residues and may become coordinated to the metal ion in the wild-type *Mun*I–DNA complex. The interactions of the non-bridging oxygens with electrophilic residues should polarize the P–O1P bond and facilitate the nucleophilic attack of a catalytic water molecule and/or stabilize the pentacoordinated transition state. It is worth noting that the arginine R121 residue is buttressed through hydrogen bonds to the pro-S-oxygen of the scissile phosphate, the N7 nitrogen of the A5 base and the OE1 carboxylate of glutamate E120 of the neighbouring subunit. Thus, at the active site of *Mun*I, R121 might be a central organizing residue that is involved in phosphodiester bond cleavage, base recognition and dimerization.

A two-metal ion mechanism of phosphodiester bond

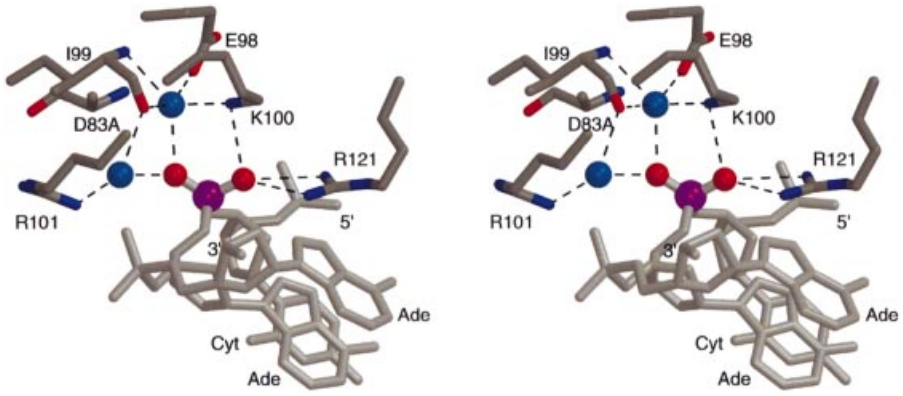

Fig. 6. Stereo view of *Mun*I active site arrangement. Active site residues are labelled. Water molecules are depicted as blue spheres, with hydrogen bonds shown as dashed black lines. One DNA strand is shown in stick representation, with the scissile phosphate shown as purple and O1P and O2P as red spheres.

Table II. DNA helix parameters

Base pair	Intra-base pair			Inter-base pair			
	Buckle (°)	Propel (°)	Open (°)	Rise (Å)	Tilt (°)	Roll (°)	Twist (°)
G1:C10	–9.43	12.77	–3.80	3.50	–3.71	7.92	28.23
C2:G9	–2.71	6.41	–2.69	3.58	4.06	–1.22	33.24
C3:G8	0.31	–1.85	–1.68	3.55	0.97	–0.28	40.27
A4:T7	–9.68	–1.92	0.84	3.63	–3.02	21.22	23.41
A5:T6	–0.98	4.04	–0.92	4.37	0.11	–53.32	40.77
T6:A5	–5.12	6.50	–1.57	3.49	3.51	21.22	21.91
T7:A4	8.87	–0.23	0.86	3.54	2.12	0.74	40.72
G8:C3	1.09	–0.18	–2.55	3.70	–4.11	–0.15	38.19
G9:C2	–3.63	4.81	1.34	3.30	3.20	4.25	27.35
C10:C1	5.91	0.03	–2.36				
Average	–1.34	3.04	–1.25	3.63	0.13	0.04	32.68
B-DNA[a]	0.0	–13.3	0.0	3.4	0.0	0.0	36.0
A-DNA[a]	0.0	–7.5	0.0	2.6	0.0	0.0	32.7

Intra- and inter-base pair parameters of the DNA oligonucleotide containing the *Mun*I recognition sequence, as determined by CURVES (Lavery and Sklenar, 1988, 1989).
[a]Average values from Hartmann and Lavery (1996).

cleavage was proposed for *Eco*RV (Baldwin *et al*., 1995; Kostrewa and Winkler, 1995; Vipond *et al*., 1995) and *Bam*HI (Newman *et al*., 1995). According to this mechanism, three and four acidic residues coordinate two metal ions at the active sites of *Eco*RV and *Bam*HI, respectively. In addition a substrate-assisted mechanism involving only one metal ion and a three metal ion-mediated substrate-assisted mechanism have been proposed for *Eco*RV (Jeltsch *et al*., 1993; Horton *et al*., 1998). However, in *Eco*RI (Kim *et al*., 1990), only a single metal-binding site was identified that structurally coincides with that of *Mun*I. The only other acidic residue of *Mun*I besides D83 and E98 that is within <10 Å distance from the scissile phosphate is E120 located on the $3_{10}4$ helix. Replacement of E120 by alanine (A.Lagunavicius and V.Siksnys, unpublished data) led to complete loss of the DNA cleavage activity of *Mun*I. This experimental observation does not necessarily imply the metal-binding role of this E120 residue since it presumably plays an important role in maintaining the dimer stability. Thus, both *Mun*I and *Eco*RI may utilize a single metal ion for catalysis; however, the crystal structures of *Mun*I and *Eco*RI in the presence of metal ions need to be determined in order to resolve this question.

DNA structure

The DNA molecules make end-to-end contacts in the crystal lattice forming a twisted quasi pseudo-continuous helix along the crystallographic *y*-axis. The DNA in the complex with *Mun*I has a distorted B-like conformation. The main feature of the DNA is a central kink. The extraordinarily large roll and rise parameters are characteristic for the inner AT/TA base pairs (Table II). The DNA in the *Mun*I complex is unwound and the DNA backbone exhibits a 'kink' at the adjacent adenine residues. The major and minor grooves of the DNA are widened by 3.5 and 5 Å, respectively, with a shallower major and a deeper minor groove in comparison with an uncomplexed oligonucleotide with the *Mun*I recognition sequence (Spink *et al*., 1995). The distorted DNA conformation is remarkably similar in *Mun*I– and *Eco*RI–DNA complexes. All atoms of the central AATT tetranucleotides of the two DNAs in the *Mun*I and *Eco*RI complexes can be superimposed with an r.m.s. deviation of 0.8 Å. However, there are also differences in the local conformational parameters of DNA in *Mun*I– and *Eco*RI–DNA complexes. In *Eco*RI–DNA, the overall parallel base stacking is maintained throughout the whole oligonucleotide sequence except for the central base pairs. *Mun*I–DNA appears to be comprised

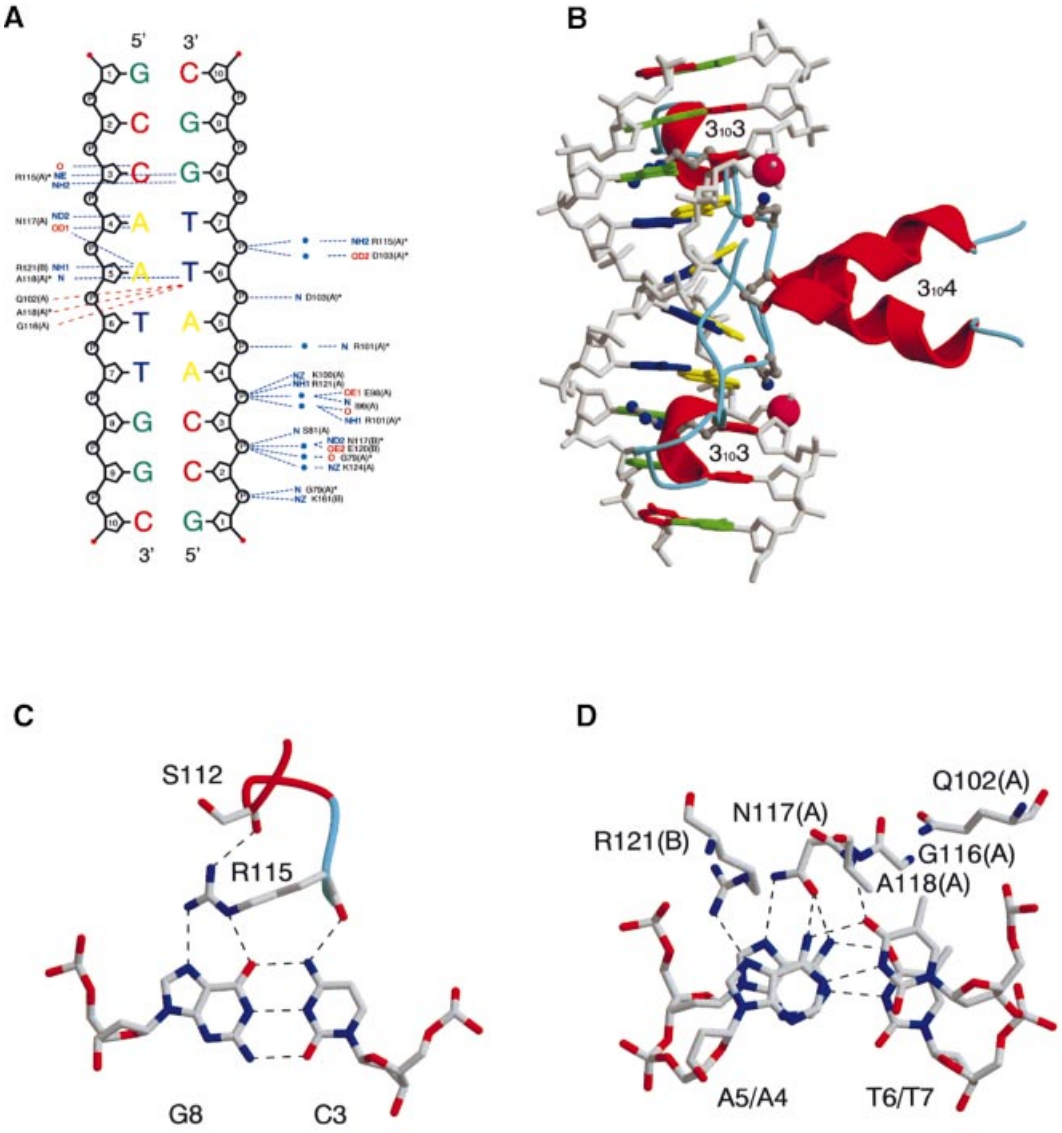

Fig. 7. Protein–DNA interactions. (**A**) NUCPLOT (Luscombe *et al*., 1997) sketch summarizing contacts between *Mun*I and DNA with the recognition sequence (C/AATTG) and the backbone. Bases are coloured according to their type, G in green, C in red, T in blue and A in yellow; the backbone is shown in black. Hydrogen bonds are displayed as dotted blue lines, van der Waals contacts as dashed red lines, and water molecules as blue dots; * indicates a residue on the plot more than once, amino acids are in one letter code and the monomer is in parentheses. The contacts to one recognition half-site and one DNA backbone strand are shown for clarity. The symmetry-related contacts are identical. (**B**) Side view of the recognition area of *Mun*I between helix $3_{10}3$ and $3_{10}4$. (**C** and **D**) Top view of the outer CG and the inner AATT base pair recognition, with hydrogen bonds indicated as dotted lines.

of two independent 5 bp DNA fragments connected via a kink in the centre. Calculation of the global axis curvature of *Mun*I–DNA assuming two 5 bp fragments yielded an overall bend of nearly 20°. This finding is consistent with the negative inclination angles of the bases to an overall axis of *Mun*I–DNA with an average value of –6.1°. There is a negative local roll of –5.5° between base C2 and C3 in each strand at the edge to the *Mun*I recognition sequence, breaking the rule of alternating signs of roll angles in the *Eco*RI–DNA structure. The base pairs C2:G9 and G9:C2 adjacent to the recognition sequence of *Mun*I exhibit positive propeller angles of 6° and 5°, respectively. The latter parameters for the base pairs G5:C10 and C10:G5 of *Eco*RI–DNA are negative (–4°). The conformation of the sugar pucker in *Mun*I DNA obeys C2′-endo except for the central kinked T6 where it is C3′-endo.

DNA recognition by MunI

In the *Mun*I–DNA complex, the protein faces the DNA major groove while the minor grove is exposed to the solvent. Therefore, it is not surprising that *Mun*I makes direct contacts to bases exclusively from the major groove side (Figure 7A and B). All amino acid residues involved in sequence-specific interactions lie within a single short region (residues 115–121) located between helices $3_{10}3$ and $3_{10}4$ and at the N-termini of the $3_{10}4$ helix of *Mun*I (Figure 7B).

The external CG base pair of *Mun*I-specific DNA sequence (C3:G8) is recognized solely by the R115 residue situated just after helix $3_{10}4$ (Figure 7C). The side chain of R115 is extended in the plane of the CG base pair. The main chain carbonyl oxygen of R115 makes a hydrogen bond with the exocyclic amino group of the C3 base. The

NE and NH2 atoms of the side chain guanidinium group of R115 donate the bidentate hydrogen bonds to the G8 base. The arginine–guanine interactions are the most common and predictable interactions in protein–DNA complexes; however, NH1 and NH2 atoms are most often involved in hydrogen bond interactions with the guanine base. In the case of *Mun*I, the hydrogen bond between the NH1 atom of residue R115 and the backbone oxygen of S112 probably fixes the stretched conformation of arginine and enables hydrogen bonding between NE and NH2 atom and the G base. The recognition of the CG base pair by main chain and side chain hydrogen bonds by a single arginine residue is unique as far as we know. It is interesting to note that the middle CG pair of the recognition sequence of *Bgl*I is recognized by the single K266 residue (Newman *et al*., 1998) by a mode that is very similar to the recognition of the CG base pair by R115 of *Mun*I.

The central AATT of the *Mun*I recognition sequence is bound specifically through van der Waals contacts and numerous hydrogen bond interactions (Figure 7D). The side chain methyl group of A118 makes a van der Waals contact with the 5C methyl group of the inner T base (T6), while the main chain NH group of the same A118 residue is at a distance of 3.1 Å from the O4 oxygen of the same T residue; however, the geometry is not optimal for this hydrogen bond. The CA atom of the main chain of G116 and the CG atom of the methylene group of Q102 are close enough to make van der Waals contacts with the methyl group of the inner T base. The methyl group of the outer T base (T7) is at a close enough distance (3.98 Å) from the CZ atom of R115 that is involved in the recognition of the CG base pair. The O4 oxygen atom of the outer T base is the only atom of potential donors and acceptors in the major groove that is not involved in the interactions with the protein. The OD oxygen of the side chain of the N117 residue is bridged between the exocyclic amino groups of adjacent adenine residues (A4 and A5) and makes three-centred bifurcated hydrogen bonds. The amino group of the side chain of the same N117 residue donates a hydrogen bond to the N7 nitrogen atom of the external adenine (A4). The N7 atom of the inner adenine (A5) residue is at hydrogen bonding distance from the NH1 atom of the R121 side chain of the symmetry-related monomer. Thus, the N117 residue makes three of the five hydrogen bonds between the AATT tetranucleotide and the *Mun*I protein. In total, there are 16 direct hydrogen bonds and six van der Waals contacts between the *Mun*I dimer and the recognition site.

The *Eco*RI restriction enzyme recognizes the hexanucleotide sequence G/AATTC that partially overlaps with that of *Mun*I. The crystal structure of *Eco*RI has provided us with the first structural mechanism of sequence recognition by type II restriction enzymes (McClarin *et al*., 1986; Kim *et al*., 1990). The current structure of the *Mun*I–DNA complex allows us to compare structural and molecular mechanisms of sequence discrimination by restriction enzymes recognizing related nucleotide sequences.

Although the recognition sequences of *Mun*I and *Eco*RI differ only in the external base pair, different structural elements are involved in the base-specific contacts. While *Eco*RI uses amino acid residues located on the extended chain motif, helices α4 and α5, to achieve the specificity for its recognition sequence, *Mun*I combines all residues involved in the base-specific contacts within one short segment (residues R115–R121) that includes the $3_{10}4$ helix.

The recognition of the middle A4:T7 and inner A5:T6 base pairs in *Mun*I and *Eco*RI is similar and accomplished mainly by residues located on the structurally equivalent α4 and $3_{10}4$ helices of *Eco*RI and *Mun*I and by a few residues just upstream of these helices. The conserved sequence motif GNAXER (Siksnys *et al*., 1994) corresponds to the structurally conserved N-termini of the α4 and $3_{10}4$ helices of *Eco*RI and *Mun*I. The amino acid residues located in the latter motif starting with G116 in *Mun*I and G140 in *Eco*RI cover 10 out of 12 major groove donor–acceptor contacts to the central four base pairs of the recognition sequence. An additional six van der Waals contacts to the methyl groups of the inner thymidines (T6), including the glutamine Q102 in *Mun*I and Q115 in *Eco*RI, complete this intricate network.

The discrimination of the external base pairs by *Mun*I and *Eco*RI is achieved by different mechanisms. The single R115 residue of *Mun*I recognizes the CG base pair (C3:G9) through the hydrogen bonds made by backbone oxygen to the amino group of cytosine and by the side chain guanidinium group to the acceptor atoms of the G base. Amino acid residues involved in the recognition of the external GC base pair by *Eco*RI are located on the separate structural elements of *Eco*RI. Arginines 200 and 203 positioned on the N-terminal part of the α5 helix sandwich a water molecule that makes hydrogen bonds with the acceptor sites of the G base. The backbone oxygen of residue A138 located on the extended chain motif of *Eco*RI makes a hydrogen bond to the amino group of the C base. The side chain of the Met137 residue located on the extended chain motif makes additional van der Waals contact with the C base. Thus, *Mun*I recognizes the CG base pair via a single R115 residue, whilst in the case of *Eco*RI three residues located on two separate structural elements are involved in the recognition of the external GC base pair.

In both cases, a number of buttressing interactions help to fix the side chain conformation of the protein residues involved in the hydrogen bond interactions with the base edges. R200 of *Eco*RI not only helps to sandwich a water molecule that donates hydrogen bonds to the G base, but also makes a hydrogen bond to the backbone oxygen of residue A139. Such a buttressing interaction might be important for both fixation of the side chain of R200 and stabilization of the conformation of the extended chain motif, suggesting a cooperative nature of the recognition network.

Similarly, the R115 residue of *Mun*I is buttressed to S112 through the hydrogen bond of the NH1 atom to the backbone oxygen of S112 located on the $3_{10}3$ helix. Such an interaction might be important in maintaining the correct side chain orientation of R115 that makes a direct interaction with the G base. R162 located on the α5 helix of *Mun*I makes a hydrogen bond to the backbone oxygen of the G114 residue and might also contribute to the stability of the $3_{10}3$ helix The replacement of the R115 residue by alanine, asparagine or glutamine led to the loss of *Mun*I DNA cleavage and binding properties. The conservative arginine to lysine replacement produced an enzyme that retained only 0.1% of the cleavage activity

of the wild-type enzyme (A.Lagunavicius and V.Siksnys, unpublished data). Such deleterious effects of mutations of a single residue cannot be explained simply by a loss of a few hydrogen bonds to a G base (Jen-Jacobson, 1995).

Protein–phosphate interactions

*Mun*I makes an extensive set of contacts with the sugar–phosphate backbone of the DNA. Protein–phosphate interactions occur throughout the second to the seventh nucleotide of each strand of the DNA (Figure 7A). In total, there are six direct and eight water-mediated contacts between both side chain and main chain atoms of *Mun*I and the phosphate oxygens. The amino acid residues of *Mun*I involved in the protein–phosphate contacts come from several regions including the loop connecting helix $3_{10}2$ and strand β2, and strand β3 and helices α5 and $3_{10}4$. It is noteworthy that amino acids residues G79, S81, I99, R101 and D103 involved in the protein–phosphate contacts are located respectively just upstream and downstream of the catalytic/metal-binding site of *Mun*I and probably help to position and fix active site residues at the scissile phosphate. Protein–phosphate contacts of amino acid residues N117, E120, R121 and K124 located on the $3_{10}4$ helix similarly position the $3_{10}4$ recognition helix in the DNA major groove. R121 and N117 are buttressed by hydrogen bonds between backbone phosphates and bases, while residue E120 also contributes to the interface between the monomers. Thus, an intricate network of hydrogen bonds connects *Mun*I residues involved in base recognition, DNA backbone contacts and in dimer interface formation. It is noteworthy that R121 of *Mun*I is structurally similar to residues R145 and R122 of *Eco*RI and *Bam*HI, respectively. While in *Bam*HI this residue participates in the direct readout of the inner guanosine of the recognition sequence, R145 of *Eco*RI and R121of *Mun*I contact the inner adenine. It is interesting to note that R145 of *Eco*RI and R121 of *Mun*I are close to the scissile phosphate and probably polarize it, which might facilitate cleavage.

In the *Eco*RI–DNA interface, contacts to six phosphates appear to act as 'clamps' to position base recognition elements and stabilize distorted DNA conformation (Jen-Jacobson *et al*., 1996). These phosphates are recognized by the protein with extremely high geometric precision. It was suggested that these 'clamp' phosphates are positioned so that the protein can exert a torsional strain to 'kink' the DNA. Thus, contacts between *Eco*RI and these phosphates presumably contribute to the DNA-binding specificity of *Eco*RI along with the direct contacts to DNA bases. It is noteworthy that a comparison of the protein–DNA backbone contacts between *Eco*RI and *Mun*I DNA complexes reveals that the contacts with the 'clamp' phosphates come from the same parts of the proteins and are highly conserved. All 'clamp' phosphates revealed by analysis of the *Eco*RI–DNA complex are also present in the *Mun*I–DNA complex. It was suggested that contacts to the 'clamp' phosphates stabilize the distorted DNA conformation in the *Eco*RI–DNA complex (Jen-Jacobson *et al*., 1996). Since the DNAs in *Mun*I and *Eco*RI complexes are distorted in the same way, it is not very surprising that these 'clamp' contacts are conserved between *Mun*I and *Eco*RI.

Role of direct and indirect 'readout' in the mechanism of specific sequence recognition by MunI

The direct 'readout' model of the recognition assumes that the discrimination between different DNA sequences can be achieved by direct hydrogen bonding of the protein residues to the bases. For a given nucleotide sequence, the donor and acceptor groups located on the base edges in the major groove of DNA make a unique pattern that might be recognized by a specific combination of amino acids situated on the protein surface. Indeed, amino acid side chain and polypeptide backbone atoms of *Mun*I make a number of direct hydrogen bonds and van der Waals contacts to the bases in the major groove. Amino acid residues of *Mun*I that make sequence-specific contacts with bases are positioned within a short region (residues 115–121) located between helices $3_{10}3$ and $3_{10}4$ and at the N-termini of the $3_{10}4$ helix. In contrast to *Mun*I, other restriction enzymes with known X-ray structures utilize discontinuous sequence segments located on different structural elements for the site-specific interaction with DNA. In total, the *Mun*I dimer makes 12 hydrogen bonds and two non-bonded interactions with the recognition sequence. All potential hydrogen bond donor and acceptor groups (except for the O4 oxygen of the T7 base) located on the base edges in the major groove are involved in direct contacts with corresponding amino acid residues of *Mun*I. Such a highly saturated network of hydrogen bonds is characteristic for the DNA–protein interfaces of restriction enzymes.

Analysis of the direct contacts between *Mun*I and the DNA bases, however, suggests that some of the hydrogen bonding interactions might be dependent on the local conformational features of the DNA. Indeed, bridging interactions between the side chain oxygen of residue N117 and exocyclic amino groups of adjacent adenine residues would be impossible in canonical B-DNA since amino groups are too far apart. The conformation of the DNA, however, is distorted in the complex with *Mun*I. The base pairs at the central AT/TA steps exhibit unusually high rise and roll values (Table II). Due to these local conformational changes, the distance between N6 exocyclic amino groups of adjacent adenine residues (A4 and A5) is reduced and therefore enables the bridging interactions with the N117 residue. Thus, the direct recognition by N117 through hydrogen bonding seems to be coupled to the indirect 'readout' of the local conformational features of the DNA. The recognition sequences of *Mun*I (CAATTG) and *Eco*RI (GAATTC) have a common AATT sequence. It is noteworthy that the local conformation of the central AATT tetranucleotide in the *Mun*I–DNA complex is very similar to that in the *Eco*RI–DNA complex (McClarin *et al*., 1986; Kim *et al*., 1990) which overlap with a 0.8 Å r.m.s. difference. The slightly higher rise and roll values are characteristic for the AA/TT step in the *Eco*RI–DNA complex. However, the distances between corresponding donor–acceptor groups of adjacent adenine bases are the same in *Eco*RI and *Mun*I complexes since the lower rise at the AT/TA step in *Mun*I is coupled to the lower roll. Consequently, asparagines N141 of *Eco*RI and N117 of *Mun*I make similar bridging interactions with the exocyclic amino groups of the adjacent adenine residues. The unusual DNA conformation of the A/T step

in *Mun*I and *Eco*RI complexes is probably fixed by numerous contacts with the oxygen atoms of the sugar–phosphate backbone of the DNA. These contacts presumably stabilize the distorted conformation of the AATT sequence. It is noteworthy that protein–phosphate contacts are well conserved between *Mun*I and *Eco*RI and might contribute to the DNA-binding specificity (see above).

The conformation of the central AATT tetranucleotide is very similar in the *Mun*I and *Eco*RI complexes despite the different external bases. The local conformation of the CAATTG and GAATTC nucleotides seems to be determined primarily by the AATT sequence. Thus, the *Mun*I and *Eco*RI restriction enzymes use the increased flexibility of the AATT in the same way to position the base and protein functional groups precisely and ensure a tight hydrogen bond network at the recognition interface. The direct and indirect 'readouts' are probably highly cooperative since the distortion of the DNA facilitates formation of direct contacts between base edges and protein. An indirect 'readout' of the sequence-dependent conformational properties of DNA might be utilized by *Mun*I and *Eco*RI to discriminate against other DNA sequences. The test for DNA distortability by docking a structurally similar four-helix bundle in the DNA major groove should allow discrimination against hexanucleotide sequences lacking an AATT tetranucleotide. The recognition of the external base pairs flanking the AATT sequence is achieved by *Mun*I and *Eco*RI through direct 'readout'. Thus, both the direct and indirect 'readouts' seem to be coupled effectively and both contribute to the mechanism of sequence discrimination by *Mun*I and *Eco*RI. We suppose that other restriction enzymes such as *Apo*I that recognize the Pu/AATTPy sequence that has a common central tetranucleotide with *Mun*I and *Eco*RI might use a similar mechanism of sequence discrimination.

Materials and methods

Crystallization

*Mun*I D83A was expressed and purified as described previously (Lagunavicius and Siksnys, 1997). The iodinated DNA oligonucleotide used in crystallization trials was synthesized in-house and purified on a Mono Q® column from Pharmacia using a NaCl gradient. After desalting, precipitating and evaporating, it was dissolved in water. The other DNA oligonucleotides were purchased from MWG in the lyophilized HPLC grade state and also dissolved in water. Annealing was performed by heating the DNA solution to 95°C for 5 min and slow cooling overnight. Crystals suitable for X-ray diffraction were obtained by mixing 2 μl of complex solution (protein:DNA ratio 1:1, concentration 0.2 mM) with 1 μl of reservoir solution (14% PEG 8000, 200 mM NaCl, 50 mM $CaCl_2$, 0.1 M MES pH 6.0) and equilibrating the mixture against 0.5 ml of the latter at 20°C in cryschem plates using the sitting drop vapour diffusion method. Crystals grew within 4 days in space group $P2_12_12_1$ with a = 58.44 Å, b = 72.89 Å, c = 122.50 Å, $\alpha = \beta = \gamma = 90°$, and contained one complex in the asymmetric unit with a solvent content of 54%. They could be flash-frozen after transfer to reservoir solution containing an additional 15% glycerol.

Data collection

All data sets (except dataset Native II) were collected on a MARresearch™ image plate mounted on a rotating anode generator RU200 at a wavelength of λ = 1.5418 Å at room temperature. The data set Native II crystal was frozen and measured at DESY-Hamburg (Beamline BW6) λ = 1.006 Å. The X-ray intensities were evaluated with the MOSFLM package (Leslie, 1991); data scaling and reduction was performed with the CCP4 package (Collaborative Computational Project Number 4, 1994).

Phasing and refinement

The *Mun*I D83A–DNA complex was solved using the technique of multiple isomorphous replacement with anomalous scattering. Heavy atom positions were identified either by inspection of isomorphous difference Patterson maps or by calculation of cross-phased difference Fourier maps using phases derived from the iodine heavy atom positions. The CCP4 program MLPHARE (Collaborative Computational Project Number 4, 1994) was used to produce initial phases to 2.5 Å resolution (Table I), which were improved by solvent flattening using SOLOMON (Abrahams and Leslie, 1996). The experimental electron density map was of high quality and enabled all the DNA and most of the protein residues with side chains to be built using the interactive graphics program O (Jones *et al*., 1991). Additional single crystal averaging with AVE (Collaborative Computational Project Number 4, 1994) exploiting the 2-fold local symmetry clearly revealed the extended dimerization arm. The structure was refined with the program X-PLOR version 3.851 (Brünger, 1992) and in the final steps with the program CNS (Brünger *et al*., 1998) using the parameters derived by Engh and Huber (1991) and Parkinson *et al*. (1996). Several cycles of refinement and manual rebuilding were performed, resulting in 646 solvent molecules being included in the final model, with excellent statistics and stereochemistry (Table I). The DNA structure was analysed using the program CURVES (Lavery and Sklenar, 1988, 1989) (Table II).

The atomic coordinates of the *Mun*I-D83A–DNA complex have been deposited in the Protein Data Bank with entry code 1D02.

Acknowledgements

We thank A.Lagunavicius for protein purification and J.Richardson for reading the manuscript. This work was supported by Volkswagenstiftung and NATO linkage grant No. 960928.

References

Abrahams,J.P. and Leslie,A.G.W. (1996) Methods used in the structure determination of the bovine mitochondrial F1 ATPase. *Acta Crystallogr. D*, **52**, 30–42.

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

Athanasiadis,A., Vlassi,M., Kotsifaki,D., Tucker,P.A., Wilson,K.S. and Kokkinidis,M. (1994) Crystal structure of *Pvu*II endonuclease reveals extensive structural homologies to *Eco*RV. *Nature Struct. Biol.*, **1**, 469–475.

Baldwin,G.S., Vipond,I.B. and Halford,S.E. (1995) Rapid reaction analysis of the catalytic cycle of the *Eco*RV restriction endonuclease. *Biochemistry*, **34**, 705–714.

Bozic,D., Grazulis,S., Siksnys,V. and Huber,R. (1996) Crystal structure of *Citrobacter freundii* restriction endonuclease *Cfr*10I at 2.15 Å resolution. *J. Mol. Biol.*, **255**, 176–186.

Brünger,A.T. (1992) *X-PLOR Version 3.1. A System for Crystallography and NMR*. Yale University Press, New Haven, CT.

Brünger,A.T. *et al*. (1998) Crystallography and NMR system: a new software suite for macromolecular structure determination. *Acta Crystallogr. D*, **54**, 905–921.

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

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

Engh,R.A. and Huber,R. (1991) Accurate bond and angle parameters for X-ray protein structure refinement. *Acta Crystallogr. D*, **4**, 392–400.

Horton,N.C., Newberry,K.J. and Perona,J.J. (1998) Metal ion-mediated substrate-assisted catalysis in type II restriction endonucleases. *Proc. Natl Acad. Sci. USA*, **95**, 13489–13494.

Jeltsch,A., Alves,J., Wolfes,H., Maass,G. and Pingoud,A. (1993) 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.

Jen-Jacobson,L. (1995) Structural-perturbation approaches to thermodynamics of site-specific protein–DNA interactions. *Methods Enzymol.*, **259**, 305–344.

Jen-Jacobson,L., Engler,L.E., Lesser,D.R., Kurpiewski,M.R., Yee,C. and McVerry,B. (1996) Structural adaptations in the interaction of *Eco*RI endonuclease with methylated GAATTC sites. *EMBO J.*, **15**, 2870–2882.

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

Kim,Y.C., Grable,J.C., Love,R., Greene,P.J. and Rosenberg,J.M. (1990) Refinement of *Eco*RI endonuclease crystal structure: a revised protein chain tracing. *Science*, **249**, 1307–1309.

Kostrewa,D. and Winkler,F.K. (1995) 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.

Kraulis,P. (1991) MOLSCRIPT: a program to produce both detailed and schematic plots of proteins. *J. Appl. Crystallogr.*, **24**, 946–950.

Lagunavicius,A. and Siksnys,V. (1997) Site-directed mutagenesis of putative active site residues of *Mun*I restriction endonuclease: replacement of catalytically essential carboxylate residues triggers DNA binding specificity. *Biochemistry*, **36**, 11086–11092.

Lagunavicius,A., Grazulis,S., Balciunaite,E., Vainius,D. and Siksnys,V. (1997) DNA binding specificity of *Mun*I restriction endonuclease is controlled by pH and calcium ions: involvement of active site carboxylate residues. *Biochemistry*, **36**, 11093–11099.

Lavery,R. and Sklenar,H. (1988) The definition of generalized helicoidal parameters and of axis curvature for irregular nucleic acids. *J. Biomol. Struct. Dynam.*, **6**, 63–91.

Lavery,R. and Sklenar,H. (1989) Defining the structure of irregular nucleic acids: conventions and principles. *J. Biomol. Struct. Dynam.*, **6**, 655–667.

Leslie,A.G.W. (1991) Crystallographic computing 5. In Moras,D., Podjarni,A.D. and Thierry,J.C. (eds), *Crystallographic Computing 5*. Oxford University Press, Oxford, UK, Vol. 5, pp. 50–61.

Luscombe,N.H., Laskowski,R.A. and Thornton,J.M. (1997) NUCPLOT: a program to generate schematic diagrams of protein–nucleic acid interactions. *Nucleic Acids Res.*, **25**, 4940–4945.

McClarin,J.A., Frederick,C.A., Wang,B.C., Greene,P., Boyer,H.W., Grable,J. and Rosenberg,J.M. (1986) Structure of the DNA–*Eco*RI endonuclease recognition complex at 3 Å resolution. *Science*, **234**, 1526–1541.

Merrit,E.A. and Murphy,M.E.P. (1994) Raster3D Version 2.0. A program for photorealistic molecular graphics. *Acta Crystallogr. D*, **50**, 869–873.

Newman,M., Strzelecka,T., Dorner,L.F., Schildkraut,I. and Aggarwal, A.K. (1995) Structure of *Bam*HI endonuclease bound to DNA: partial folding and unfolding on DNA binding. *Science*, **269**, 656–663.

Newman,M., Lunnen,K., Wilson,G., Greci,J., Schildkraut,I. and Phillips, S.E. (1998) Crystal structure of restriction endonuclease *Bgl*I bound to its interrupted DNA recognition sequence. *EMBO J.*, **17**, 5466–6476.

Parkinson,G., Vojtechovsky,J., Clowney,L., Brunger,A.T. and Berman, H.M. (1996) New parameters for the refinement of nucleic acid-containing structures. *Acta Crystallogr. D*, **52**, 57–64.

Roberts,R.J. and Macelis,D. (1998) REBASE—restriction enzymes and methylases. *Nucleic Acids Res.*, **26**, 338–350.

Siksnys,V., Zareckaja,N., Vaisvila,R., Timinskas,A., Stakenas,P., Butkus,V. and Janulaitis,A. (1994) CAATTG-specific restriction-modification *munI* genes from *Mycoplasma*: sequence similarities between R.*Mun*I and R.*Eco*RI. *Gene*, **142**, 1–8.

Spink,N., Nunn,C.M., Vojtechovsky,J., Berman,H.M. and Neidle,S. (1995) Crystal structure of a DNA decamer showing a novel pseudo four-way helix–helix junction. *Proc. Natl Acad. Sci. USA*, **92**, 10767–10771.

Stakenas,P.S., Zaretskaia,N.M., Manelene,Z.P., Mauritsas,M.M., Butkus, V.V. and Ianulaitis,A.A. (1992) *Mycoplasma* restriction–modification system *Mun*I and its possible role in pathogenesis processes. *Mol. Biol.*, **26**, 546–557.

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

Vipond,I.B., Baldwin,G.S. and Halford,S.E. (1995) Divalent metal ions at the active sites of the *Eco*RV and *Eco*RI restriction endonucleases. *Biochemistry*, **34**, 697–704.

Winkler,F.K. *et al.* (1993) The crystal structure of *Eco*RV endonuclease and of its complexes with cognate and non-cognate DNA fragments. *EMBO J.*, **12**, 1781–1795.

Received July 27, 1999; revised September 10, 1999; accepted September 13, 1999