The EMBO Journal Vol.17 No.18 pp.5466–5476, 1998

Crystal structure of restriction endonuclease *Bgl*I bound to its interrupted DNA recognition sequence

Matthew Newman[1,2], Keith Lunnen[3], Geoffrey Wilson[3], John Greci[3], Ira Schildkraut[3] and Simon E.V.Phillips[2]

School of Biochemistry and Molecular Biology, and North of England Structural Biology Centre, University of Leeds, Leeds LS2 9JT, UK and [3]New England Biolabs, 32 Tozer Road, Beverly, MA 01915, USA

[1]Present address: Imperial Cancer Research Fund, 44 Lincoln's Inn Fields, London, WC2A 3PX, UK

[2]Corresponding authors
e-mail: M.Newman@icrf.icnet.uk and SEVP@bmb.leeds.ac.uk

The crystal structure of the type II restriction endonuclease *Bgl*I bound to DNA containing its specific recognition sequence has been determined at 2.2 Å resolution. This is the first structure of a restriction endonuclease that recognizes and cleaves an interrupted DNA sequence, producing 3′ overhanging ends. *Bgl*I is a homodimer that binds its specific DNA sequence with the minor groove facing the protein. Parts of the enzyme reach into both the major and minor grooves to contact the edges of the bases within the recognition half-sites. The arrangement of active site residues is strikingly similar to other restriction endonucleases, but the coordination of two calcium ions at the active site gives new insight into the catalytic mechanism. Surprisingly, the core of a *Bgl*I subunit displays a striking similarity to subunits of *Eco*RV and *Pvu*II, but the dimer structure is dramatically different. The *Bgl*I–DNA complex demonstrates, for the first time, that a conserved subunit fold can dimerize in more than one way, resulting in different DNA cleavage patterns.

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

Introduction

Type II restriction endonucleases comprise one of the major families of endonucleases (Roberts and Halford, 1993; Pingoud and Jeltsch, 1997). They usually recognize a short palindromic DNA sequence between 4 and 8 base pairs in length, and in the presence of Mg^{2+}, specifically catalyse the hydrolysis of phosphodiester bonds at precise positions within or close to this sequence. Although type II restriction endonucleases are ubiquitous within prokaryotes, there is generally no sequence similarity among them.

Despite this lack of sequence similarity, crystallographic analyses of several restriction endonucleases have revealed considerable three-dimensional similarity, correlating well with the type of cleavage pattern produced by these enzymes. *Eco*RV (Winkler *et al.*, 1993) and *Pvu*II (Athanasiadis *et al*., 1994; Cheng *et al*., 1994) have conserved subunit cores, consisting of a central five-stranded β-sheet flanked by α-helices, which interact to form structurally similar homodimers. They bind their uninterrupted 6-bp recognition sequences with the minor groove facing the protein and cleave the DNA to produce blunt-ended fragments. *Bam*HI (Newman *et al*., 1995) and *Eco*RI (McClarin *et al*., 1986) also have conserved subunit cores, consisting of five β-strands and two α-helices, but they are topologically distinct from the *Eco*RV-like enzymes and form very different dimer structures. They bind their 6-bp recognition sequences with the major groove facing the protein and cleave to produce four base 5′ overhangs. *Cfr*10I (Bozic *et al*., 1996) and the catalytic domain of *Fok*I (Wah *et al*., 1997) have a similar subunit fold to the *Bam*HI-like enzymes, and also produce DNA products with four base 5′ overhangs. Despite the structural differences between the *Eco*RV- and *Bam*HI-like enzymes, their active sites are well conserved, consisting of a triad of charged amino acid residues (Aggarwal, 1995; Pingoud and Jeltsch, 1997).

*Bgl*I, the type II restriction endonuclease from *Bacillus globigii* (Wilson and Young, 1976; Lee and Chirikjian, 1979), produces a different cleavage pattern from those of known structures. It recognizes the interrupted DNA sequence GCCNNNNNGGC and cleaves between the fourth and fifth unspecified base pair to produce 3′ overhanging ends (Bickle and Ineichen, 1980; Lautenberger *et al*., 1980; Van Heuverswyn and Fiers, 1980). A subunit of *Bgl*I has a molecular mass of 35 kDa, and consists of 299 amino acid residues. It displays no significant sequence homology to any other restriction endonucleases. Owing to the nature of its cleavage pattern, it was anticipated that *Bgl*I would represent a new structural class of restriction endonucleases.

To understand further the molecular basis of DNA recognition and hydrolysis by type II restriction endonucleases, the structure of *Bgl*I bound to a DNA oligonucleotide containing its specific recognition sequence was determined to a resolution of 2.2 Å. We have compared the structure of *Bgl*I with those of *Eco*RV and *Pvu*II. Surprisingly, despite the fact *Bgl*I recognizes an interrupted DNA sequence and produces 3′ overhanging ends, whereas *Eco*RV recognizes a contiguous 6-bp sequence and produces blunt ends, a subunit of *Bgl*I shows significant structural similarity to subunits of the *Eco*RV-like enzymes. It appears that dramatically different cleavage patterns are achieved by *Bgl*I and *Eco*RV using conserved subunit folds, combined with alternative modes of dimerization. In addition, we propose a two-metal-ion catalytic mechanism based on the presence of two calcium ions that are coordinated at the active site of *Bgl*I.

Results and discussion

Structure determination

The gene coding for the *Bgl*I restriction endonuclease was cloned from *B.globigii* genomic DNA and overexpressed

in an *Escherichia coli* strain that co-expressed the *Bgl*I methyltransferase. Purified *Bgl*I endonuclease had a specific activity of ~1 250 000 U/mg. Crystals were obtained of *Bgl*I bound to a 17 bp DNA oligonucleotide containing its specific recognition sequence. In order to obey the symmetry imposed by the crystal lattice, the DNA oligonucleotide that was used in the final crystallization experiments incorporated an A:A mismatch at the central base pair (Figure 1). The structure was solved using the technique of multiple isomorphous replacement with anomalous scattering. An experimental electron density map calculated at 2.5 Å resolution was of extremely high quality and enabled most of the protein and all the DNA to be built. The structure has been refined to 2.2 Å resolution with good statistics and stereochemistry (Table I). Two hundred and fifty three solvent atoms were included in the final refined model. All protein main chain torsion angles are located in the energetically allowed regions for L-amino acids.

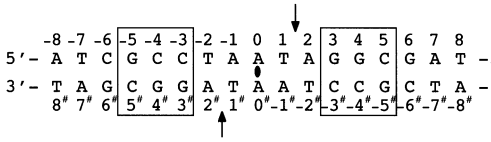

Fig. 1. Sequence of the 17 base pair DNA fragment used in the *Bgl*I–DNA complex. Recognition half-sites are boxed. The sites of cleavage are indicated by arrows. The position of the crystallographic two-fold is shown and the symmetry related strand is indicated by #. To avoid crystallographic disorder, an A:A mismatch was incorporated at the centre of the oligonucleotide.

Overall structure

The structure of the *Bgl*I bound to a DNA oligonucleotide containing its specific recognition sequence is shown in Figure 2A. There is only one protein subunit plus one DNA strand in the asymmetric unit of the crystal structure, and thus the model of the biologically active homodimer is generated by the application of a crystallographic two-fold operation. The *Bgl*I dimer measures ~75×55×55 Å. There is a large DNA binding cleft, ~45 Å in length, which is long enough to accommodate 14 base pairs of a DNA duplex. The DNA is bound with the minor groove at the centre of the 17 bp oligonucleotide facing the protein, as in the *Eco*RV and *Pvu*II protein–DNA complexes (Winkler *et al*., 1993; Cheng *et al*., 1994). Thus, it appears that enzymes that cleave their DNA to produce blunt-ended or single-stranded 3′ overhanging ends approach the DNA from the minor groove side, so that the active site is oriented correctly with respect to the scissile phosphodiester bond (Cheng *et al*., 1994).

A subunit of *Bgl*I has an α/β structure with a large central six-stranded β-sheet flanked by α-helices (Figure 2B). The central sheet contains strands β^1, β^2, β^3, β^7, β^8, β^9 and β^{12} (β^7 and β^9 form a single pseudo-continuous β-strand). β^1, β^2 and β^3 are anti-parallel and form a β-meander which contains the active site residues Asp116, Asp142 and Lys144. α-helix α^4 packs against the concave surface of the central β-sheet, and contributes residues to both the dimer interface and the active site (Glu87). α^4 has an inserted residue (Ala89) that is accommodated through the formation of a bulge that protrudes from the

Table I. Data collection and refinement statistics

	Native	Br1	Br2	Br3	Hg	Os
Data collection statistics						
Wavelength (Å)	0.919	0.919	0.919	0.919	0.919	1.140
Resolution (Å)	2.2	2.3	2.3	2.3	2.3	2.8
R_{merge} [a] (%)	5.5 (7.8)	3.4 (6.5)	3.8 (8.1)	3.3 (6.7)	3.9 (10.7)	3.1 (5.4)
I>3σ(I) (%)	94.4 (81.1)	95.2 (84.6)	90.8 (74.6)	93.7 (83.1)	86.3 (63.4)	95.7 (90.6)
Completeness (%)	90.8 (57.1)	94.7 (62.2)	95.7 (66.6)	93.7 (56.9)	86.5 (56.5)	90.8 (53.2)
Phasing statistics						
R_{iso} [b] (%)		5.5	18.5	18.8	27.2	17.0
Phasing power						
$iso_{centric}$ [c]		1.1	0.7	0.7	0.9	0.4
$iso_{acentric}$ [c]		1.9	0.9	0.9	1.1	0.4
ano[d]		1.8	1.7	2.0	1.6	1.1
FOM[e] centric (%)	47.9					
acentric (%)	57.9					
sol. flat. (%)	91.3					
Refinement statistics						
Resolution (Å)	10–2.2					
No. refls. |F|>2σ(|F|)	17359					
R-factor[f] (%)	18.1					
R-free[g] (%)	24.4					
RMSD[h] bonds (Å)	0.009					
angles (°)	1.4					

Values in parentheses are for data in the highest resolution shell.
[a]$R_{merge} = \Sigma|I_{obs}-<I>|/\Sigma I_{obs}$
[b]$R_{iso} = \Sigma|F_p-F_{ph}|/\Sigma F_p$, where F_p and F_{ph} are the structure factor amplitudes of the native and derivative data, respectively.
[c]Isomorphous phasing power is defined as $<|F_h|>$/r.m.s. (ε), where $<|F_h|>$ 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.
[d]Anomalous phasing power is defined as $<|\Delta F_{calc}|>$/r.m.s. (ε), where $<|\Delta F_{calc}|>$ is the mean calculated Bijvoet difference from the heavy-atom model and r.m.s. (ε) is the root mean square lack of closure for the anomalous differences.
[e]Figure of merit. [f] R-factor = $\Sigma|F_{obs}-F_{calc}|/\Sigma|F_{obs}|$, where F_{obs} and F_{calc} are the observed and calculated structure factor amplitudes, respectively.
[g]The free *R*-value (Brünger, 1992) was calculated using a random 10% of reflection data that was omitted from all stages of the refinement. [h]Root mean square deviation from ideal stereochemical parameters (Engh and Huber, 1991; Parkinson *et al*., 1996).

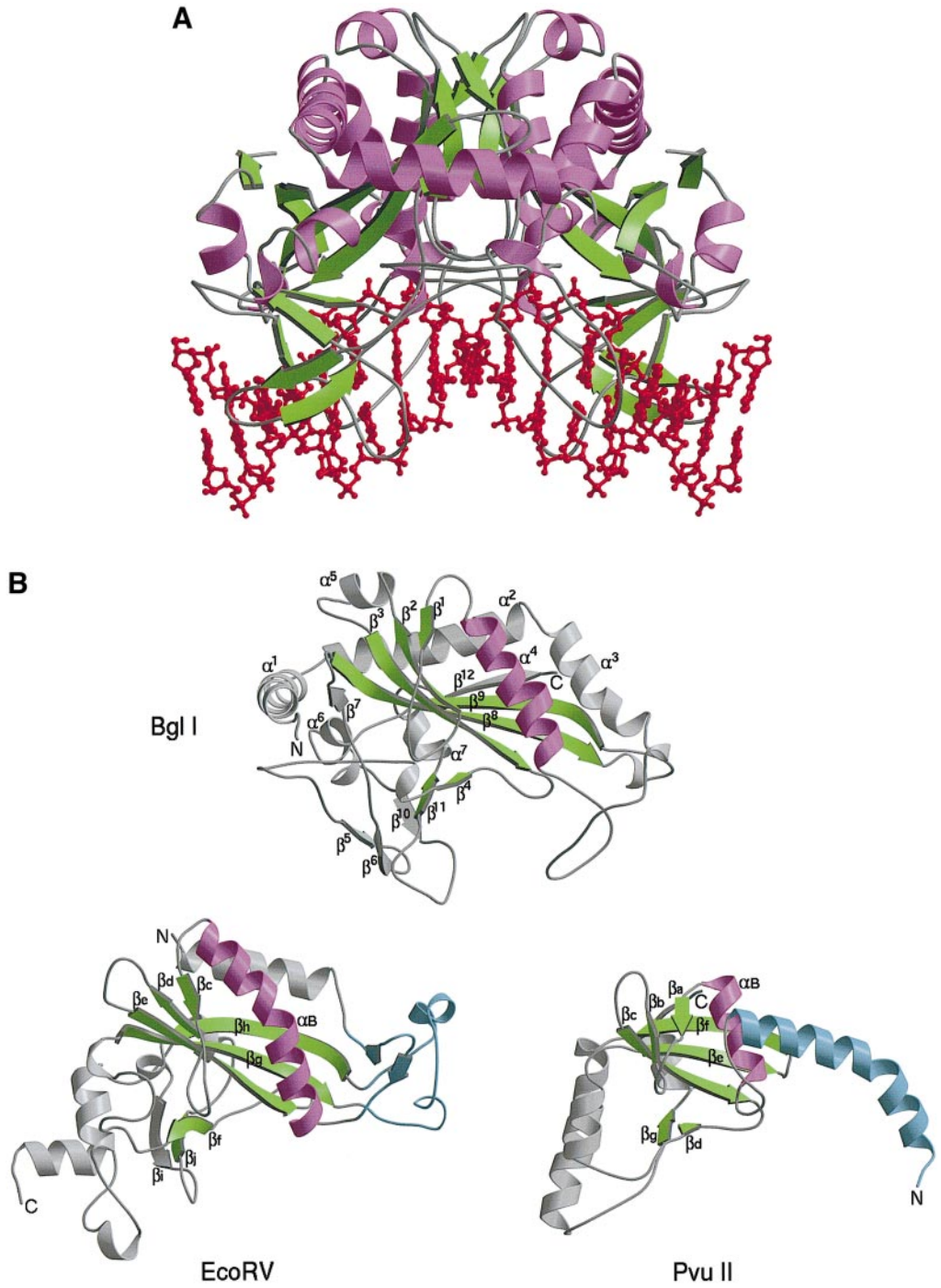

Fig. 2. (**A**) Structure of the *Bgl*I dimer bound to DNA containing its recognition sequence, with a crystallographic two-fold running vertically through the centre of the complex. The DNA helical axis is horizontal. α-helices and 3_{10}-helices are shown in purple, β-strands in green and the DNA in red. (**B**) Secondary structural assignment of a *Bgl*I subunit and comparison with *Eco*RV and *Pvu*II (PDB codes 1rva and 1pvi, respectively). Secondary structural elements were defined using the Kabsch and Sander (1983) algorithm implemented in PROCHECK (Laskowski *et al*., 1993), but were modified in some regions according to the hydrogen-bonding criteria of Baker and Hubbard (1984). 3_{10}-helices are shown, but are not labelled. The N- and C-terminal ends are labelled. Only relevant labels are given for *Eco*RV, as in Winkler *et al*. (1993), and *Pvu*II, as in Athanasiadis *et al*. (1994). Common elements of secondary structure between *Bgl*I, *Eco*RV and *Pvu*II are shown in purple for α-helices and green for β-strands. The dimerization sub-domains are coloured blue for *Eco*RV and *Pvu*II. Produced using MOLSCRIPT (Kraulis, 1991) and RASTER3D (Merritt and Murphy, 1994).

side of the helix (Kavanaugh *et al*., 1993; Harrison *et al*., 1994). The helix does not incorporate any proline residues, and does not contain any other distortion. A smaller three-stranded anti-parallel β-sheet (β^4, β^{10} and β^{11}) is situated at the bottom of the larger β-sheet. This smaller sheet contains most of the residues involved in specific DNA recognition.

The dimer interface is extremely large and is formed by residues comprising one side of the *Bgl*I subunit. The interface area calculated using the method of Lee and

Richards (1971) is ~3100 Å^2, which is significantly larger than that for proteins of a similar molecular weight (Janin *et al*., 1988). The equivalent areas for *Eco*RV and *Bam*HI were ~2200 and 1600 Å^2, respectively. The dimer interface is formed by several segments of the polypeptide chain: the N-terminus, the C-terminal region of helix α^3, helix α^4, the loop between β^1 and β^2, just before strand β^5, and the loops between β^7 and β^8, and between β^8 and β^9 (Figure 3). The dimer interface contains 40 hydrogen bonds, including 12 salt bridges. Additionally, 22 water molecules are involved in water-mediated hydrogen bonds between the two subunits.

Similarity to EcoRV and PvuII

Surprisingly, despite the fact that *Bgl*I is the first structure of a type II restriction endonuclease that binds and interrupted recognition site and hydrolyses its phosphodiester bonds to produce 3′ overhanging ends, analysis of the subunit fold reveals extensive similarities to *Eco*RV (Winkler *et al*., 1993) and *Pvu*II (Athanasiadis *et al*., 1994; Cheng *et al*., 1994). A region consisting of five β-strands of the central β-sheet (incorporating the β-meander), an α-helix and two β-strands of the smaller β-sheet is shared by the three enzymes (β^1, β^2, β^3, β^8, β^9, α^4, β^4 and β^{11} in *Bgl*I; βc, βd, βe, βg, βh, αB, βf and βj in *Eco*RV; and βa, βb, βc, βe, βf, αB, βd and βg in *Pvu*II; Figures 2B and 3). This common region includes residues that comprise the active sites and DNA recognition elements. A pair-wise least-squares superposition of *Bgl*I and *Eco*RV gives a root mean square (r.m.s.) difference of 2.1 Å from 91 C_α pairs that can be aligned closer than 3.8 Å. *Bgl*I and *Pvu*II are less similar (*Pvu*II being significantly smaller), giving an r.m.s. difference of 1.5 Å for 41 C_α pairs that can be aligned closer than 3.8 Å. Outside the common region, however, *Bgl*I shows significant differences to *Eco*RV and *Pvu*II, many of which occur in parts of the protein structure involved in dimerization (Figure 2B).

DNA structure

The DNA is primarily B-form but has a slight curvature, resulting in a 20° bend away from the protein (Figure 2A). There are no major bends or kinks of the type seen in *Eco*RI or *Eco*RV (specific) DNA complexes (McClarin *et al*., 1986; Winkler *et al*., 1993). The average DNA helical parameters for the *Bgl*I–DNA complex are 32° for twist and 3.5 Å for rise (Table II), with most of the sugars in the standard 2′-endo conformation for B-DNA. The A:A mismatch at the centre of the oligonucleotide does not distort the B-form structure significantly, with both bases remaining in the anti conformation, but results in a localized shift of the adenine bases within the DNA helix. The DNA helical parameters at the mismatch have similar values to those obtained from an identical mismatch in the NF–κB p50 homodimer complex (Mueller *et al*., 1995).

Some of the largest deviations from B-form DNA occur within, or adjacent to, the recognition half-site. There is a 16° roll between base pairs 3 and 2 opening up the minor groove [other roll angles are generally small ($<|8°|$)]. There are positive propeller angles at base pairs 3 and 2 (7° and 13°, respectively, where the mean value for B-DNA is –13°). Base step 3–2 also has the smallest twist of 26°. Surprisingly, there are large tilts within the recognition half-site, associated with base steps 5–4 and 4–3. These

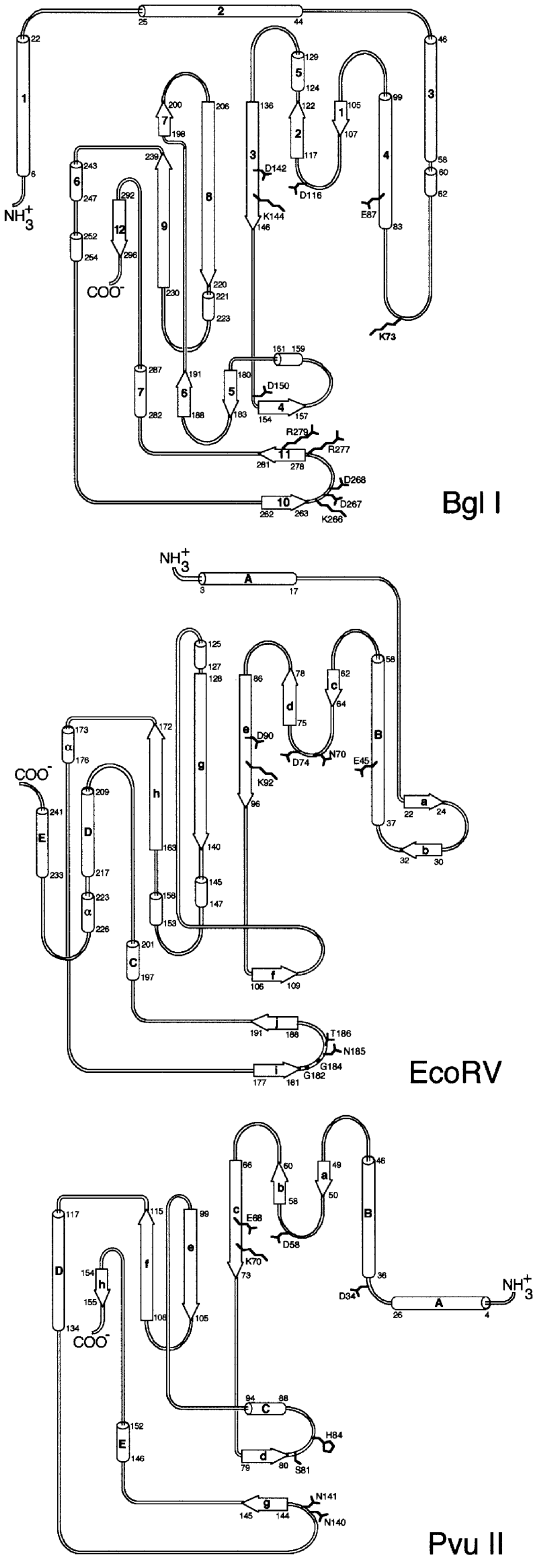

Fig. 3. Schematic diagram showing the secondary structural topology of *Bgl*I, *Eco*RV and *Pvu*II. α-helices and 3_{10}-helices are shown as cylinders and β-strands as broad arrows. α-helices and β-strands are labelled according to Figure 2B. 3_{10}-helices are unlabelled. The N- and C-terminal ends are labelled. The extent of the secondary structural element is indicated by residue numbers. Residues thought to be involved in either the active site or specific DNA recognition are shown. For *Eco*RV, the recognition R-loop is between βi and βj, and the Q-loop is between βc and βd.

Table II. DNA helical parameters

	Base pair	Intra-base pair			Inter-base pair			
		Buckle (°0)	Propell (°)	Open (°)	Rise (Å)	Tilt (°)	Roll (°)	Twist (°)
8	$A^{-8\#}$:T^{8}	−12.4	−4.6	2.0				
					3.1	−3.4	5.4	30.9
7	$T^{-7\#}$:A^{7}	−2.3	−16.2	8.2				
					3.3	2.4	6.9	34.0
6	$C^{-6\#}$:G^{6}	0.0	−4.5	−0.8				
					3.5	−0.1	−3.2	36.2
5	$G^{-5\#}$:C^{5}	−0.2	1.1	−0.2				
					3.1	−10.5	−7.9	33.9
4	$C^{-4\#}$:G^{4}	3.7	−21.4	−2.4				
					3.3	9.2	−3.2	33.1
3	$C^{-3\#}$:G^{3}	9.3	7.2	−4.1				
					4.1	−6.5	16.0	25.5
2	$T^{-2\#}$:A^{2}	15.7	12.5	9.7				
					4.0	−0.5	2.5	34.8
1	$A^{-1\#}$:T^{1}	7.2	−9.9	−0.6				
					3.9	6.8	7.3	29.5
0	$A^{0\#}$:A^{0}	0.0	−18.2	24.3				
Average[a]		0.0	−5.3	2.8	3.5	0.0	3.0	32.2
B-DNA[b]		0.0	−13.3	0.0	3.4	0.0	0.0	36.0
A-DNA[b]		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 *Bgl*I recognition sequence, as determined by CURVES (Lavery and Sklenar, 1988, 1989). Values for only half the oligonucleotide are shown, as values for the other half are related due to the two-fold crystallographic symmetry.
[a]Average values are calculated over entire 17 bp oligonucleotide.
[b]Average values from Hartmann and Lavery (1996).

are of similar magnitude but of different sign, thus the resultant effect on the DNA helix is negligible. Tilt angles about the short axis of a base pair are rarely seen as they are unfavourable energetically. The major groove width varies from 10.6 Å at the centre of the oligonucleotide, to 13.9 Å in the recognition half-site, and the minor groove width varies from 4.6 Å at the ends of the duplex to 9.4 Å at the recognition half-site. The average values for B-form DNA are 11.7 and 5.7 Å for the major and minor groove widths, respectively (Saenger, 1984). The widening of both the major and minor grooves probably provides improved access for the recognition elements of the protein.

DNA recognition

A *Bgl*I subunit sits entirely on one DNA half-site, and cleaves within that half of the DNA oligonucleotide (Figure 4A). There is no cross-over mode of binding of the type seen in *Bam*HI, where a protein subunit forms contacts to both DNA half-sites (Newman *et al*., 1995). It has been proposed that the cross-over mode of binding is important as it guarantees that specific recognition leads to the correct formation of both active sites, followed by concerted double-strand cleavage (Pingoud and Jeltsch, 1997). Although there is no cross-over binding in *Bgl*I, there are extensive subunit–subunit contacts (see above), which include an intricate intersubunit hydrogen-bonding network that involves the active site region. The β^{1}–β^{2} loop of one subunit, which contains the active site residue Asp116, interacts with conserved helix α^{4} of the other subunit. It is possible that these intersubunit contacts could fulfil the same function in *Bgl*I as the cross-over mode of binding does in *Bam*HI, i.e. non-specific binding in one half-site leads to inactivated catalytic centres in both protein subunits.

Although *Bgl*I contacts bases within both the major and minor grooves, it does not surround the DNA like *Eco*RV. This is reflected by a smaller *Bgl*I dimer–DNA interface than that for *Eco*RV (1671 versus 2354 $Å^2$, respectively). The only direct base contacts made by *Bgl*I are within the recognition sequence, with most occurring in the major groove. There are no direct contacts to the unspecified five base pairs between the two recognition half-sites, although there are contacts to the sugar-phosphate backbone within this region.

Major groove. The major groove contacts involve residues situated on, or near to, the small three stranded recognition β-sheet, comprising strands β^{4}, β^{10} and β^{11} (Figures 3 and 4A). Strands β^{4} and β^{11} have counterparts in both *Eco*RV (Winkler *et al*., 1993) and *Pvu*II (Athanasiadis *et al*., 1994; Cheng *et al*., 1994), forming structurally conserved DNA recognition regions. The details of specific base contacts vary, however: *Eco*RV recognizes DNA via the R-loop between strands βi and βj (Winkler *et al*., 1993), whereas *Pvu*II uses anti-parallel strands βd and βg (Cheng *et al*., 1994). Within the recognition half-site of *Bgl*I there are eight direct protein–DNA hydrogen bonds and one water-mediated hydrogen bond (Figure 4B). Thus, the hydrogen-bonding potential with the major groove is satisfied totally. In addition to these hydrogen bonds, there are numerous buttressing interactions which help to correctly orient and immobilize protein groups involved in DNA recognition.

The outer G:C (-5#:5) base pair is contacted by Arg279 from strand β^{11} and Asp150 from just before strand β^{4}. The guanidinium group of Arg279 lies in the plane of Gua-5#, donating two hydrogen bonds [$N^{\eta 1}$...N7 (3.4 Å), $N^{\eta 2}$...O6 (2.8 Å)]. Arginine–guanine interactions are probably the most commonly observed interactions in

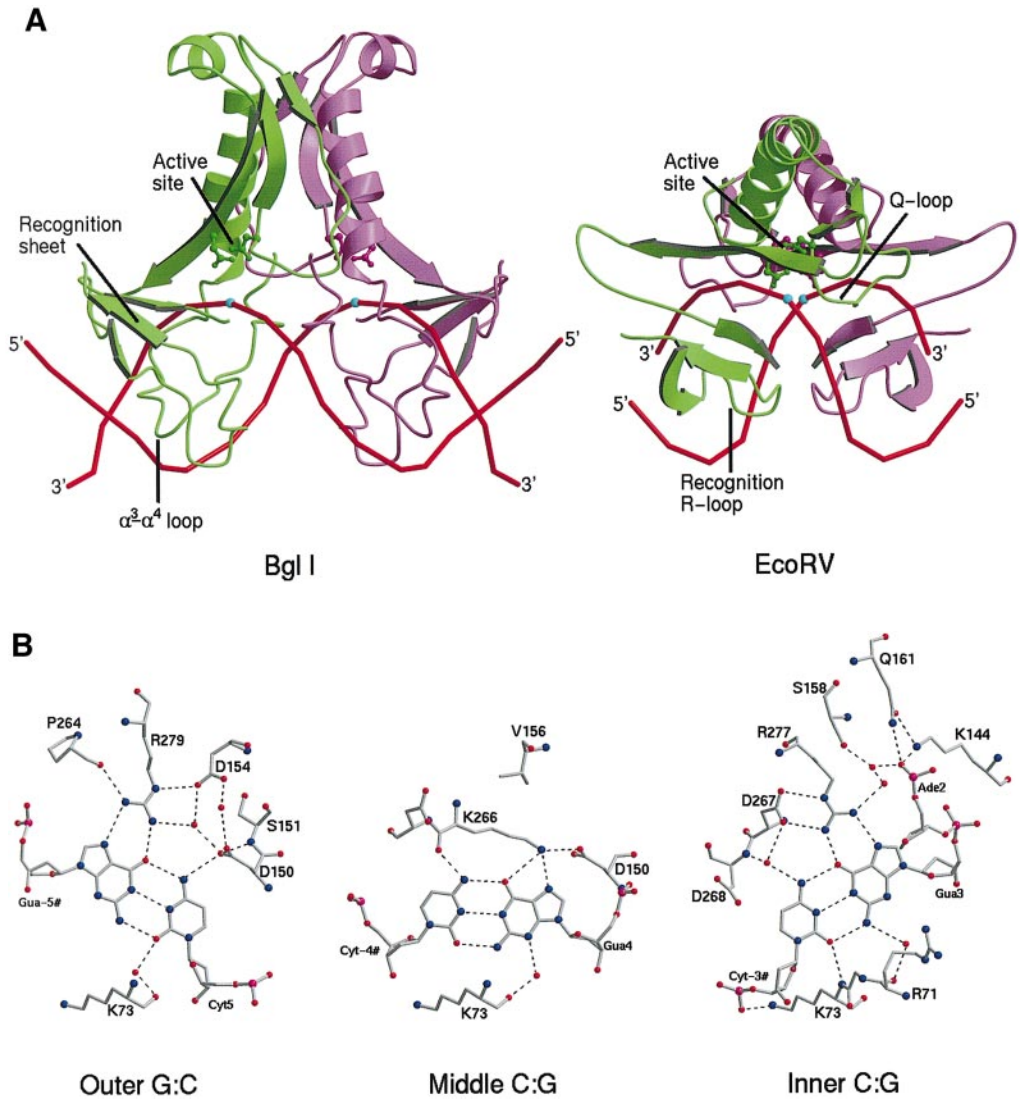

Fig. 4. (**A**) A simplified view of *Bgl*I and *Eco*RV showing the DNA recognition and active site elements. Only residues 62–159 and 262–282 are shown for *Bgl*I and only residues 34–109 and 177–193 are shown for *Eco*RV. The intersubunit 2-fold is oriented vertically for both enzymes. One subunit is coloured green and the other is purple. Active site residues (Asp116 and Asp142 for *Bgl*I, and Asp74 and Asp90 for *Eco*RV) are shown as a ball-and-stick representation. The DNA is shown in red by bonds drawn through the O5′ positions for simplicity. Cyan spheres indicate the position of the active site phosphates. (**B**) A view of the interactions between the outer (G:C), middle (C:G) and inner (C:G) base pairs of the recognition sequence of *Bgl*I. Dashed lines indicate hydrogen bonds. Selected water molecules are shown as red spheres. Interactions are shown for both the major and minor groove edges of the base pairs. Buttressing interactions, which serve to orient and immobilize recognition groups, are also shown. For the inner base pair, part of the hydrogen-bonding network with the scissile phosphodiester bond is shown. The calcium ions plus their liganded waters (including Wat4) have been omitted for clarity, as has the base of Ade2. Produced using MOLSCRIPT (Kraulis, 1991) and RASTER3D (Merritt and Murphy, 1994).

protein–DNA complexes (Pabo and Sauer, 1992). The outer cytosine base, Cyt5, is involved in a single hydrogen bond to Asp150 [N4...OD1 (3.2 Å)]. Contacts to the middle C:G (-4#:4) base pair arise solely from Lys266, situated on the loop between strands β^{10} and β^{11}. Lys266 main chain torsion angles are in the α_L conformation and the side chain is extended in the plane of the C:G base pair. Cyt-4# is involved in a single hydrogen bond to the main chain carbonyl group of Lys266 [N4...O (2.8 Å)]. The side-chain amino group of Lys266 is involved in a bifurcated hydrogen bond to Gua4 [NZ...N7 (2.8 Å), NZ...O6 (3.4 Å)]. Main chain and side-chain hydrogen bonds from a single lysine to both the guanine and cytosine bases in a G:C base pair are uncommon, but have been observed previously in λ repressor–operator complex (Beamer and Pabo, 1992). Only the guanine base of the inner C:G (-3#:3) base pair is involved in direct protein contacts. The guanidinium group of Arg277 (situated just before strand β^{11}) lies in the plane of Gua3 and donates two hydrogen bonds to the N7 and O6 atoms [NH1...N7 (3.0 Å), NH2...O6 (2.6 Å)]. Cyt-3# is hydrogen bonded to a highly ordered water molecule [N4...OH2 (3.0 Å), B = 9 $Å^2$]. This water molecule is further hydrogen bonded to residues Asp267 [OD2...OH2 (3.1 Å)] and Asp268 [N...OH2 (2.8 Å)], which are situated on the loop between strands β^{10} and β^{11}. In addition, an intricate hydrogen-bonding network, involving two highly ordered water molecules, links Arg277 to the active site phosphate. This feature, coupling DNA-recognition elements to the active site phosphate, has also been observed within

the recognition interface of the *Bam*HI–DNA complex (Newman *et al*., 1995).

Minor groove. *Bgl*I contacts bases within the minor groove of the recognition half-site via the large loop (residues 59–82) between helices α^3 and α^4 (Figures 3 and 4A). PvuII forms specific minor groove base contacts using a similar part of its structure, between helices αA and αB (Asp34, Figure 3), but contacts a different half-site from *Bgl*I (Nastri *et al*., 1997). *Eco*RV also forms minor groove base contacts, but uses a different part of its structure to achieve this, known as the Q-loop (Winkler *et al*., 1993; Figures 3 and 4A). *Bgl*I forms a direct protein–DNA hydrogen bond between the main-chain amide of Lys73 and Cyt-3# [N...O2 (3.2 Å)], and three water-mediated hydrogen bonds to Gua3, Gua4 and Cyt5 (Figure 4B). Although these hydrogen bonds confer no DNA specificity, the GCC sequence may be recognized by a process of indirect read-out owing to its preference for a wide minor groove (Goodsell *et al*., 1993). Other DNA sequences without such a wide minor groove may give restricted access to the α^3–α^4 loop, disrupting hydrogen bonds to bases and nearby phosphate groups.

Phosphate contacts. There are extensive contacts between the enzyme and the sugar-phosphate backbone of the DNA. The *Bgl*I dimer contacts both DNA strands at all phosphate groups from –7 to 6, except for nucleotides –2 and –1. Each subunit makes 17 direct hydrogen bonds to the phosphate groups of the DNA duplex, with 21 more being mediated by water molecules. Of the direct phosphate contacts, 10 involve side-chain groups (five of which are from arginine or lysine residues) and seven are from main chain NH groups. The residues involved in direct or water-mediated hydrogen bonds with the DNA backbone come from several segments of the enzyme: the loop between α^3 and α^4, the loop between β^1 and β^2, the C-terminal end of β^3, residues near the small three-stranded recognition β-sheet (involving strands β^4, β^{10} and β^{11}), the β-hairpin between β^5 and β^6 and the C-terminal end of β^8 (Figure 3).

Active site

Type II restriction endonucleases require only Mg^{2+} as a cofactor to catalyse the hydrolysis of a phosphodiester bond, producing 5′ phosphate and 3′ hydroxyl groups. Hydrolysis proceeds with stereochemical inversion of configuration at the phosphorous atom (Connolly *et al*., 1984; Grasby and Connolly, 1992). This possibly involves a direct in-line attack on the phosphate by an activated water molecule, leading to a pentavalent transition state stabilized by Mg^{2+} ions. Although several catalytic mechanisms have been proposed (Jeltsch *et al*., 1993; Kostrewa and Winkler, 1995; Vipond *et al*., 1995), there is currently no detailed understanding of the steps involved in phosphodiester hydrolysis. For example, it is not clear whether one or two Mg^{2+} ions are involved at the active site (Vipond *et al*., 1995; Groll *et al.*, 1997). Also, it is not known which group activates the attacking water molecule, or which group is responsible for protonation of the leaving group.

*Bgl*I represents the first structure of a type II restriction endonuclease that cleaves its DNA substrate producing 3′ overhanging ends, and analysis of the three-dimensional structure reveals that it too possesses the conserved triad of charged amino acid residues adjacent to the scissile phosphodiester bond. Residues Asp116, Asp142 and Lys144 in *Bgl*I can be aligned spatially with active site residues Asp74, Asp90 and Lys92 in *Eco*RV (Winkler *et al*., 1993), with only small differences in the positions of side-chain groups (Figure 5A). These residues are situated on the edge of a β-meander in the central β-sheet (comprising strands β^1, β^2 and β^3 in *Bgl*I). Residues Asp74 and Asp90 are essential for the catalytic function of *Eco*RV, and are probably involved in co-ordinating Mg^{2+}-cofactor ions. Substitution of either of these residues by alanine renders *Eco*RV inactive (Selent *et al*., 1992; Groll *et al*., 1997). Importantly, inspection of the experimental electron density map (calculated using MIR phases), and subsequent $2|F_o|-|F_c|$ maps (calculated using model phases), identified two Ca^{2+} ions at the active site of *Bgl*I (Figure 5B, Table III). Calcium 1 has octahedral co-ordination geometry, involving ligands from the carboxylate groups of Asp116 and Asp142, the carbonyl of Ile143 and a non-bridging oxygen O2P of the active site phosphate Ade2. The remaining two ligands are water molecules. Calcium 2 has pentagonal bipyramidal co-ordination geometry that is typical of a Ca^{2+} ion, but not of Mg^{2+}. Ligands are from the carboxylate of Asp116, a non-bridging oxygen O2P of the active site phosphate Ade2, O3′ of Thy1, and four water molecules. The ligand distances lie within the ranges 2.3–2.7 Å and 2.3–2.8 Å, for calcium ions 1 and 2, respectively. No DNA cleavage product was observed in the *Bgl*I–DNA crystal structure, as Ca^{2+} ions do not generally support hydrolysis of DNA by restriction endonucleases (Vipond and Halford, 1995).

The Ca^{2+} positions are very similar to the divalent cation positions in the 3′–5′ exonuclease domain of DNA polymerase I (Beese and Steitz, 1991; Figure 5C). They are also similar to the proposed position of two ions in the *Bam*HI–DNA complex, identified from two highly ordered water molecules (Newman *et al*., 1995). However, the positions of the Ca^{2+} ions observed at the active site of *Bgl*I differ from the divalent cation positions identified in the *Eco*RV–DNA complex (Kostrewa and Winkler, 1995), although this complex was non-productive, as soaking the crystals in solutions containing Mg^{2+} did not lead to DNA cleavage. For *Bgl*I, the metal ions lie along a direction parallel to the scissile phosphodiester bond, whereas for *Eco*RV, they are positioned on a line approximately perpendicular to it, with the Mg^{2+} ions co-ordinated by Asp74 and Asp90, and Glu45 and Asp74, respectively (Figure 5A). The *Bgl*I–DNA structure reveals that Glu87 superimposes with Glu45 of *Eco*RV (Figure 5A), both residues being situated on a conserved α-helix (α^4 in *Bgl*I and αB in *Eco*RV). In the *Bgl*I active site, however, Glu87 does not participate in Ca^{2+} binding directly, but instead forms a hydrogen bond to the calcium-1-bound water molecule Wat3 (Figure 5B). In *Eco*RV, Glu45 has been shown to be important for catalysis, but unlike Asp74 and Asp90, substitution of this residue by an alanine does not abolish activity entirely (Selent *et al*., 1992; Groll *et al*., 1997). *Pvu*II does not have an acidic residue at a structurally equivalent position on conserved helix αB (Athanasiadis *et al*., 1994; Cheng *et al*., 1994).

Wat4 provides a possible candidate for the nucleophilic water (Figure 5B), and is in a very similar position to a

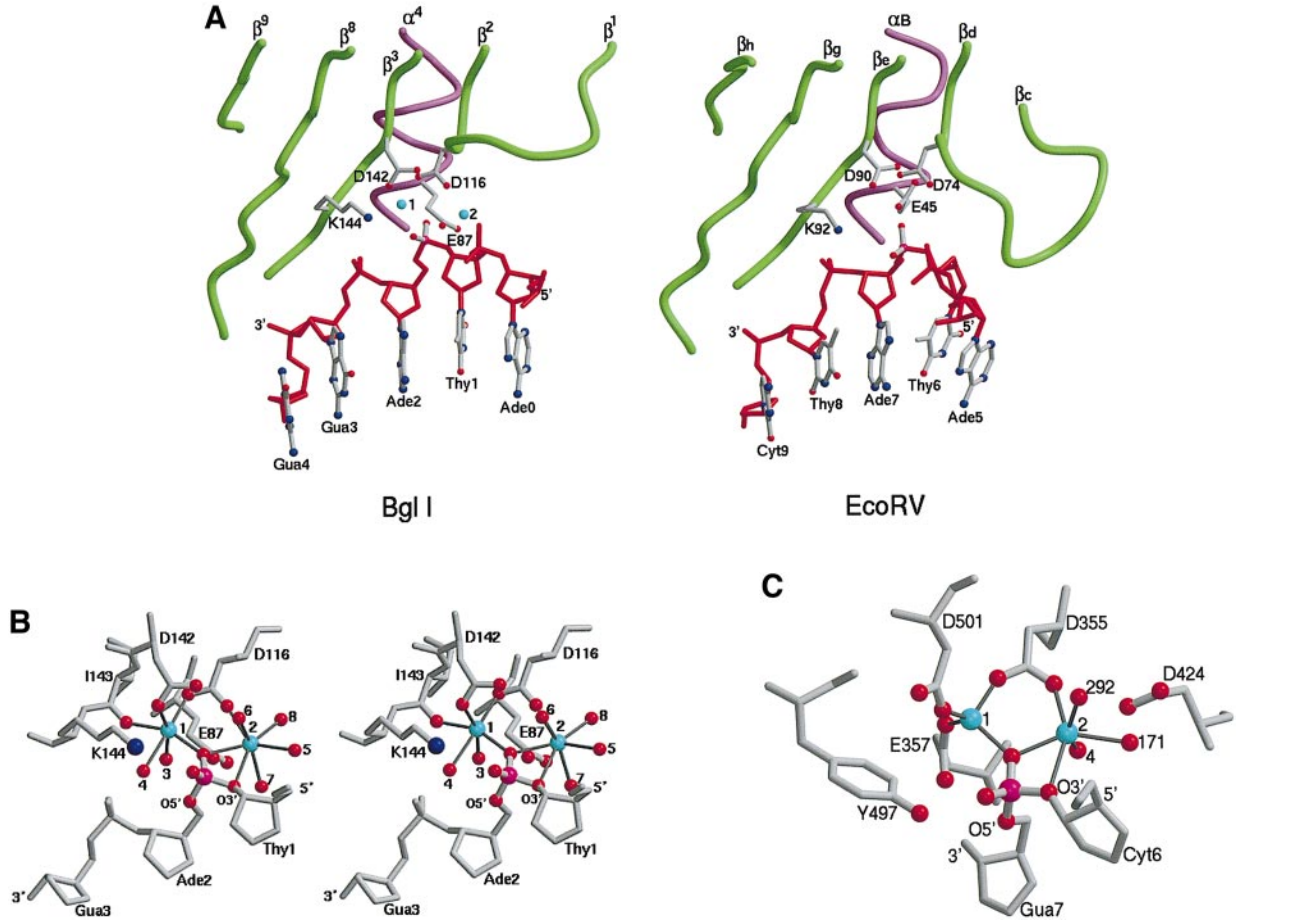

Fig. 5. The active site of *Bgl*I. (**A**) Comparison of the active sites of *Bgl*I and *Eco*RV. Both views are in a similar orientation and show similar regions of the structures. The central β-sheet is green with an α-helix (α^4 in *Bgl*I and αB in *Eco*RV) is purple. Only five bases of a single strand of DNA are shown for clarity, with the sugar–phosphate backbone in red. Active site residues and the active site phosphate are shown as a ball-and-stick representation. Calcium ion positions for the *Bgl*I–DNA complex are shown as cyan spheres. (**B**) Stereo view of the co-ordination geometry of the Ca^{2+} ions in the active site of *Bgl*I. Active site residues are shown as a ball-and-stick representation. Only a small segment of the DNA phosphate backbone on either side of the scissile phosphodiester bond is shown. DNA bases have been omitted for clarity. Calcium ions are shown as cyan spheres. Atoms co-ordinated to the calcium ions are shown connected by solid grey lines. (**C**) The active site of the 3′–5′ exonuclease domain of DNA polymerase I (PDB code 1kfs) shown as a ball-and-stick representation. Divalent cations, shown as cyan spheres, are Zn^{2+} at position 1 and Mg^{2+} at position 2. The attacking water molecule is not shown. Produced using MOLSCRIPT (Kraulis, 1991) and RASTER3D (Merritt and Murphy, 1994).

water observed in the crystal structure of the *Bam*HI–DNA complex (Newman *et al*., 1995). In addition to being co-ordinated by calcium 1, Wat4 is hydrogen bonded to Lys144 N^{ζ} (2.9 Å), Ade2 O5′ (2.8 Å) and Gua3 O1P (2.7 Å). It is well positioned for in-line attack opposite the O3′ leaving group, being 3.2 Å from the phosphorus atom of the scissile phosphodiester bond, with the Wat4–P-O3′ angle almost linear (164°). There is, however, no convenient acidic group that could act as a general base to deprotonate the attacking water molecule (Wat4). A substrate-assisted catalysis model has been proposed for restriction endonucleases, where the pro-R_p oxygen atom of the phosphate 3′ to the scissile phosphodiester bond deprotonates the attacking water molecule (Jeltsch *et al*., 1993). The main problem with this model is that although Gua3 O1P in *Bgl*I is well placed to deprotonate Wat4, the phosphate oxygen would be a poor proton acceptor owing to its unfavourably low pK_a (p$K_a \leqslant 2$).

Based on the active site structure of *Bgl*I, one possible mechanism is that a Mg^{2+} ion at site 1 could help to activate the attacking water molecule (Wat4), a Mg^{2+} at site 2 could help to stabilize the negative charge on the 3′ oxyanion leaving group, and both ions could be involved in stabilizing the pentavalent transition state. Although

Table III. Active site calcium ion interactions

Ion[a]	Ligand[a]	Distance (Å)
Calcium 1 (18)	Ade2 O2P (15)	2.4
	Asp116 OD2 (11)	2.3
	Asp142 OD1 (9)	2.5
	Ile143 O (14)	2.4
	Wat3 O (12)	2.4
	Wat4 O (13)	2.7
Calcium 2 (22)	Ade2 O2P (15)	2.5
	Asp116 OD1 (15)	2.3
	Thy1 O3′ (13)	2.8
	Wat5 O (21)	2.6
	Wat6 O (6)	2.5
	Wat7 O (23)	2.4
	Wat8 O (14)	2.4

[a]Atomic temperature factors (Å^2) are given in parentheses.

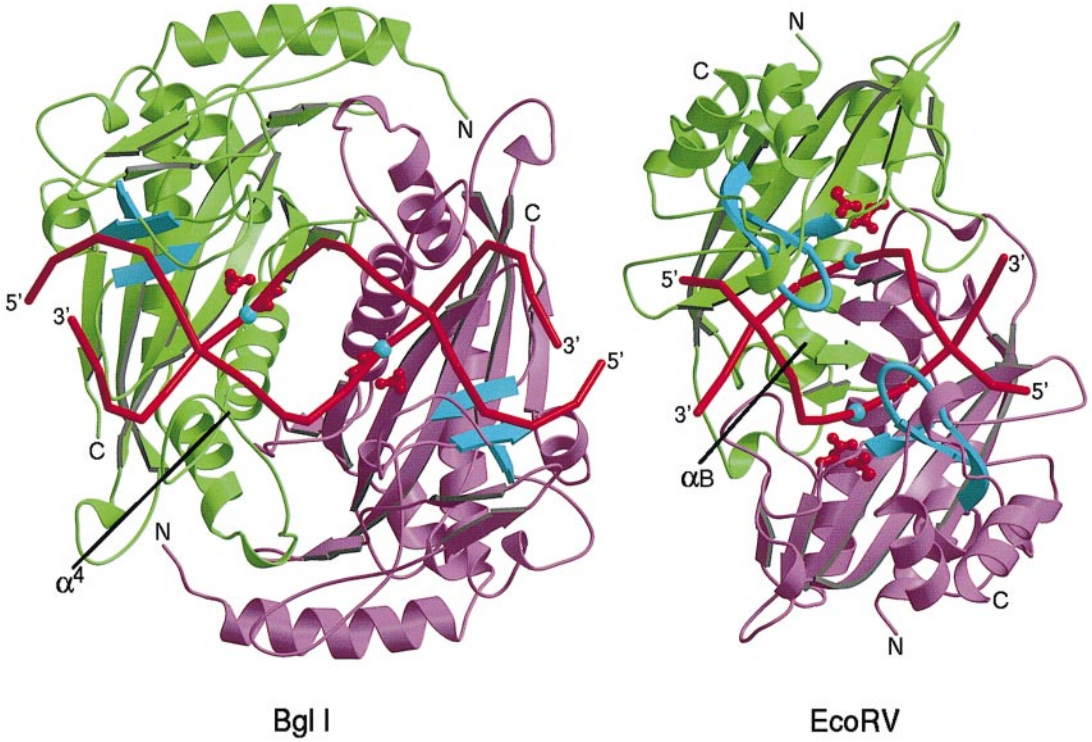

Fig. 6. Comparison of the dimer structure of *Bgl*I and *Eco*RV, viewed down the molecular two-fold axis, with the DNA above the protein. One subunit is coloured green and the other purple. The DNA is indicated in red by bonds drawn through the O5′ positions. Active site residues (Asp116 and Asp142 for *Bgl*I, Asp74 and Asp90 for *Eco*RV) are shown in red as a ball-and-stick representation. The position of the active site phosphate is shown as a cyan sphere. The DNA recognition sheet is coloured blue for *Bgl*I, as is the equivalent sheet, plus recognition R-loop, for *Eco*RV. The N- and C-terminal ends of the protein, and the 5′- and 3′-ends of the DNA, are labelled. Conserved helices α^4 in *Bgl*I and αB in *Eco*RV are labelled at their N-terminal ends. Produced using MOLSCRIPT (Kraulis, 1991) and RASTER3D (Merritt and Murphy, 1994).

Ca^{2+} appears to mimic the role of the active cofactor Mg^{2+} in specific DNA binding by *Eco*RV (Vipond and Halford, 1995), it must be borne in mind that the *Bgl*I–DNA structure is properly considered a non-productive complex, and that Ca^{2+} may be a poor model for the binding of Mg^{2+} (Engler *et al*., 1997). The co-ordination distance of Mg^{2+} is 2.0 Å, whereas Ca^{2+} has a longer co-ordination distance of 2.5 Å. Substitution of Ca^{2+} by Mg^{2+} would therefore require small positional adjustments of the groups involved in co-ordinating the divalent cations.

Cleavage pattern

Despite the fact that a subunit of *Bgl*I displays a significant degree of structural similarity to subunits of *Eco*RV and *Pvu*II, *Bgl*I has a remarkably different dimer structure. All three enzymes have a conserved α-helix at their dimer interface: α^4 in *Bgl*I, and αB in *Eco*RV and *Pvu*II (Figures 2B and 3). However, the remaining subunit–subunit interactions differ considerably between the three enzymes. In *Eco*RV, dimerization occurs primarily through residues situated on a β-hairpin (involving strands βa–βb) and a large loop between strands βg and βh. These two loops form the dimerization sub-domain (Winkler *et al*., 1993). This sub-domain interacts with the corresponding sub-domain from another protein subunit. In *Pvu*II, the dimerization sub-domain is replaced by α-helix αA and the loop between αA and αB (Athanasiadis *et al*., 1994; Cheng *et al*., 1994). In *Bgl*I, there is no equivalent of the dimerization sub-domain (Figure 2B), and as a result of the altered subunit–subunit interactions, one *Bgl*I subunit is rotated with respect to the other by 86° compared with *Eco*RV (101° compared with *Pvu*II). This creates a *Bgl*I dimer that is less 'U-shaped' than either the *Eco*RV or *Pvu*II dimers. The *Bgl*I subunits are also ~11 Å further apart along a direction almost parallel to the DNA helical axis, creating the water-filled cavity at the dimer interface (Figure 6). The resulting 'screw' transformation of the *Bgl*I subunits, correctly places the active sites and DNA recognition elements adjacent to the scissile phosphodiester bonds and DNA half-sites, respectively. The two active sites of *Bgl*I are separated by ~12 Å along a direction parallel to the DNA helical axis, which is the correct distance to produce DNA fragments with single-stranded three base extensions.

The structure of *Bgl*I–DNA complex shows that dimeric restriction endonucleases can use a conserved *Eco*RV-like subunit fold, combined with alternative modes of dimerization, to generate cleavage patterns with blunt or 3′ overhanging ends. The observation that proteins use conserved folds that are combined in alternative ways to recognize different DNA sequences, was previously noted in the structures of the MetJ (Somers and Phillips, 1992) and Arc (Raumann *et al*., 1994) repressor complexes. The repressor homodimers display significant structural similarity to each other, but cooperative contacts between pairs of homodimers involve dramatically different regions in the two structures, because the half-sites in the MetJ and Arc operator sequences that have different spacings. It remains to be seen whether restriction endonucleases

that generate 5′ overhanging ends of various lengths, can achieve these cleavage patterns by alternative dimerization of a conserved, *Bam*HI-like subunit fold.

Materials and methods

Cloning and purification

The *Bgl*I restriction/modification system was cloned from a 4.8 kb *Eco*RI fragment of *B.globigii* genomic DNA using methylase selection (Lunnen *et al*., 1988). Independent clones were isolated and shown by DNA sequencing to contain an open reading frame (orf) that coded for the N-terminus of purified *Bgl*I endonuclease. These clones lacked any R.*Bgl*I endonuclease activity which was attributed to various frameshift errors in the R.*Bgl*I orf. An overexpressing plasmid coding for the full-length R.*BglI* gene was constructed (DDBJ/EMBL/GenBank accession No. AF050216) from these clones and was introduced into an *E.coli* strain (NEB #815) which harboured a compatible plasmid expressing the *Bgl*I methylase gene. *Bgl*I was purified on Heparin Sepharose and Q-Sepharose columns. Purified *Bgl*I (protein concentration 15 mg/ml in 20mM Tris–HCl pH 7.5, 10 mM β-mercaptoethanol, 200 mM NaCl) had a specific activity of ~1 250 000 U/mg. SDS–PAGE analysis showed a single band at >95% purity.

Crystallization

DNA oligonucleotides for use in crystallization trials were synthesized using an in-house service and purified on a DIONEX anion exchange HPLC (final concentration 20 mg/ml in 50 mM NaCl, Tris–HCl pH 7.5, 1 mM EDTA). Diffraction quality crystals were obtained by co-crystallizing *Bgl*I with a self-complementary 17mer (protein:DNA ratio was 1:2, Figure 1) using the vapour diffusion method. Crystals measuring up to 600×300×100 μm were obtained from 7–12% PEG4000, 75–150 mM Li_2SO_4, 100 mM Tris–HCl pH 8.5 at room temperature. The crystals were subsequently transferred to a stabilizing solution containing 8–10% PEG8000, 125–150 mM $Ca(acetate)_2$ and 100 mM PIPES pH 6.5. The space group was C2221 and unit cell dimensions were a = 78.5 Å, b = 81.6 Å, c = 117.1 Å. Heavy-atom screening was performed on a Rigaku RU200 rotating anode generator with a Siemens Xentronics X100A area detector. Two heavy-atom derivatives were identified from crystals soaked in PCMBS and K_2OsCl_6. In addition, modified oligonucleotides with 5 bromo-dU (Br) were purchased from Cruachem: Br1 5′-ABrC-GCCBrAATAGGCGAT-3′; Br2 5′-ABrCGCCTAABrAGGCGAT-3′; and Br3 5′-ATCGCCBrAABrAGGCGAT-3′.

Data collection

Data were collected at BM14 ESRF using a CCD detector from crystals frozen at 100 K. X-ray fluorescence spectra for the bromine k-edge and osmium L_{111}-edge were measured from *Bgl*I–DNA derivatized crystals to select the appropriate wavelengths for the maximum anomalous signal. X-ray data were collected by the oscillation method ($\Delta\phi$ = 0.5°) and reduced to profile fitted intensities using the HKL suite (Otwinowski and Minor, 1997) (Table I). Data were placed on an approximately absolute scale using TRUNCATE, before derivative-to-native anisotropic scaling using SCALEIT (Collaborative Computational Project Number 4, 1994).

Phasing and refinement

The *Bgl*I–DNA complex was solved using the technique of multiple isomorphous replacement with anomalous scattering. Heavy-atom positions were identified by inspection of either isomorphous or anomalous difference Pattersons, or by calculation of cross phased difference Fouriers using phases derived from the PCMBS heavy-atom positions. The maximum-likelihood phase refinement program SHARP (De la Fortelle and Bricogne, 1997) was used to produce initial phases to 2.5 Å resolution (Table I), which was improved by solvent flattening using SOLOMON (Abrahams and Leslie, 1996). The experimental electron density map was of an extremely high quality and enabled all the DNA and 298 out of 299 amino acids to be built using the interactive graphics program O (Jones *et al*., 1991). The structure was refined using the program X-PLOR Vs. 3.86 (Brünger *et al*., 1987), employing a bulk solvent correction in the final stages. Several rounds of refinement and manual rebuilding were performed, resulting in 253 solvent molecules being included in the model, with excellent final statistics and stereochemistry (Table I). DNA structure was analysed using the program CURVES (Lavery and Sklenar, 1988, 1989).

Acknowledgements

We thank E.Fanchon, L.Smith, A.Thomson, E.Vinecombe and C.Wilmot for helping with data collection at ESRF and J.Jäger for useful discussions. We thank beam-line staff at both the SRS, Daresbury and ESRF, Grenoble. This work was supported in part by BBSRC, MRC and HHMI grants. Full co-ordinates and structure factors are being deposited in the Brookhaven Protein Data Bank.

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. 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. *Struct. Biol.*, **1**, 469–475.

Baker,E.N. and Hubbard,R.E. (1984) Hydrogen bonding in globular proteins. *Prog. Biophys. Mol. Biol.*, **44**, 97–179.

Beamer,L.J. and Pabo,C.O. (1992) Refined 1.8 Å crystal structure of the λ repressor–operator complex. *J. Mol. Biol.*, **227**, 177–196.

Beese,L.S. and Steitz,T.A. (1991) Structural bases for the 3′–5′ exonuclease activity of *Escherichia coli* DNA polymerase I: a two metal ion mechanism. *EMBO J.*, **10**, 25–33.

Bickle,T.A. and Ineichen,K. (1980) The DNA sequence recognized by *Bgl*I. *Gene*, **9**, 205–212.

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) Free *R* value: a novel statistical quantity for assessing the accuracy of crystal structures. *Nature*, **355**, 472–475.

Brünger,A.T., Kuriyan,J. and Karplus,M. (1987) Crystallographic R-factor refinement by molecular dynamics. *Science*, **235**, 458–460.

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.*, **D50**, 760–763.

Connolly,B.A., Eckstein,F. and Pingoud,A. (1984) The stereochemical course of the restriction endonuclease *Eco*RI-catalysed reaction. *J. Biol. Chem.*, **259**, 10760–10763.

De la Fortelle,E. and Bricogne,G. (1997) Maximum-likelihood heavy-atom parameters refinement in the MIR and MAD methods. *Methods Enzymol.*, **276**, 472–494.

Engh,R.A. and Huber,R. (1991) Accurate bond and angle parameters for X-ray protein-structure refinement. *Acta Crystallogr.*, **A47**, 392–400.

Engler,L.E., Welch,K.K. and Jen-Jacobson,L. (1997) Specific binding by *Eco*RV endonuclease to its DNA recognition site GATATC. *J. Mol. Biol.*, **269**, 82–101.

Goodsell,D.S., Kopka,M.L., Cascio,D. and Dickerson,R.E. (1993) Crystal structure of CATGGCCATG and its implications for A-tract bending models. *Proc. Natl Acad. Sci. USA*, **90**, 2930–2934.

Grasby,J.A. and Connolly,B.A. (1992) Stereochemical outcome of the hydrolysis reaction catalysed by the *Eco*RV restriction endonuclease. *Biochemistry*, **31**, 7855–7861.

Groll,D.H., Jeltsch,A., Selent,U. and Pingoud,A. (1997) Does the restriction endonuclease *Eco*RV employ a two-metal-ion mechanism for DNA cleavage? *Biochemistry*, **36**, 11389–11401.

Harrison,C.J., Bohm,A.A. and Nelson,H.C.M. (1994) Crystal structure of the DNA binding domain of the heat shock transcription factor. *Science*, **263**, 224–227.

Hartmann,B. and Lavery,R. (1996) DNA structural forms. *Quart. Rev. Biophys.*, **29**, 309–368.

Janin,J., Miller,S. and Chothia,C. (1988) Surface, subunit interfaces and interior of oligomeric proteins. *J. Mol. Biol.*, **204**, 155–164.

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.

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.*, **A47**, 110–119.

Kabsch,W. and Sander,C. (1983) Dictionary of protein secondary structure: pattern recognition of hydrogen-bonded and geometrical features. *Biopolymers*, **22**, 2577–2637.

Kavanaugh,J.S., Moo-Penn,W.F. and Arnone,A. (1993) Accommodation of insertions in helices: the mutation in haemoglobin catonsville (Pro37α-Glu-Thr38α) generates a $3_{10}\rightarrow\alpha$ bulge. *Biochemistry*, **32**, 2509–2513.

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.J. (1991) MOLSCRIPT: A program to produce both detailed and schematic plots of protein structures. *J. Appl. Crystallogr.*, **24**, 946–950.

Laskowski,R.A., MacArthur,M.W., Moss,D.S. and Thornton,J.M. (1993) PROCHECK: a program to check stereochemical quality of protein structures. *J. Appl. Crystallogr.*, **26**, 283–291.

Lautenberger,J.A., White,C.T., Haighwood,N.L., Edgell,M.H. and Hutchison,C.A. (1980) The recognition site of type II restriction enzyme *Bgl*I is interrupted. *Gene*, **9**, 213–231.

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

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

Lee,B.K. and Richards,F.M. (1971) The interpretation of protein structures: estimation of static accessibility. *J. Mol. Biol.*, **55**, 379–400.

Lee,Y.H. and Chirikjian,J.G. (1979) Sequence-specific endonuclease *Bgl*I. *J. Biol. Chem.*, **254**, 6838–6843.

Lunnen,K.D. *et al.* (1988) Cloning type-II restriction and modification genes. *Gene*, **74**, 25–32.

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.

Merritt,E.A. and Murphy,M.E.P. (1994) Raster 3D version 2.0: A program for photorealistic molecular graphics. *Acta Crystallogr.*, **D50**, 869–873.

Mueller,C.W., Rey,F.A., Sodeoka,M., Verdine,G.L. and Harrison,S.C. (1995) Structure of the NF–κB p50 homodimer bound to DNA. *Nature*, **373**, 311–317.

Nastri,H.G., Evans,P.D.,Walker,I.H. and Riggs,P.D. (1997) Catalytic and DNA binding properties of *Pvu*II restriction endonuclease mutants. *J. Biol. Chem.*, **272**, 25761–25767.

Newman,M., Strzelecka,T., Dorner,L.F., Schildkraut,I. and Aggarwal,A.K. (1995) Restriction endonuclease *Bam*HI partially folds and unfolds on DNA binding: crystal structure of *Bam*HI–DNA complex at 2.2 Å resolution. *Science*, **269**, 656–663.

Otwinowski,Z. and Minor,W. (1997) Processing of X-ray diffraction data collected in oscillation mode. *Methods Enzymol.*, **276**, 307–326.

Pabo,C.O. and Sauer,R.T. (1992) Transcription factors – structural families and principles of DNA recognition. *Annu. Rev. Biochem.*, **61**, 1053–1095.

Parkinson,G., Vojtechovsky,J., Clowney,L., Brünger,A.T. and Berman, H.M. (1996) New parameters for the refinement of nucleic acid containing structures. *Acta Crystallogr.*, **D52**, 57–64.

Pingoud,A. and Jeltsch,A. (1997) Recognition and cleavage of DNA by type-II restriction endonucleases. *Eur. J. Biochem.*, **246**, 1–22.

Raumann,B.E., Rould,M.A., Pabo,C.O. and Sauer,R.T. (1994) DNA recognition by β-sheets in the Arc repressor–operator crystal structure. *Nature*, **367**, 754–757.

Roberts,R.J. and Halford,S.E. (1993) Type II restriction endonucleases. In Linn,S.M., Lloyd,R.S. and Roberts,R.J. (eds), *Nucleases*, Second Edition. Cold Spring Harbor Laboratory Press, Cold Spring Harbor, NY, pp. 35–88.

Saenger,W. (1984) *Principles of Nucleic Acid Structure*. Springer, New York.

Selent,U. *et al.* (1992) A site-directed mutagenesis study to identify amino acid residues involved in the catalytic function of the restriction endonuclease *Eco*RV. *Biochemistry*, **31**, 4808–4815.

Somers,W.S. and Phillips,S.E.V. (1992) Crystal structure of the *met* repressor–operator complex at 2.8 Å resolution reveals DNA recognition by β-strands. *Nature*, **359**, 387–393.

Van Heuverswyn,H. and Fiers,W. (1980) Recognition sequence for the restriction endonuclease *Bgl*I from *Bacillus globigii*. *Gene*, **9**, 195–203.

Vipond,I.B. and Halford,S.E. (1995) Specific DNA recognition by *Eco*RV restriction endonuclease induced by calcium ions. *Biochemistry*, **34**, 1113–1119.

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.

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

Wilson,G.A. and Young,F.E. (1976) Restriction and modification in the *Bacillus subtilis* genuspecies. In Schlessinger,D. (ed.), *Microbiology*. American Society for Microbiology, Washington DC, pp. 350–357.

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 June 24, 1998; revised July 15, 1998; accepted July 16, 1998