The EMBO Journal vol.12 no.5 pp.1781–1795, 1993

The crystal structure of *Eco*RV endonuclease and of its complexes with cognate and non-cognate DNA fragments

Fritz K.Winkler, David W.Banner, Christian Oefner, Demetrius Tsernoglou[1], Raymond S.Brown[2], Stephen P.Heathman[3], Richard K.Bryan[4], Philip D.Martin[5], Kyriakos Petratos[6] and Keith S.Wilson[7]

Pharma Research-New Technologies, F.Hoffmann-La Roche Ltd, 4002 Basel, Switzerland

Present addresses: [1]European Molecular Biology Laboratory, Postfach 10.2209, 6900 Heidelberg, Germany, [2]Howard Hughes Medical Institute, Laboratory of Molecular Medicine, Children's Hospital/Enders 670, 320 Longwood Avenue, Boston, MA 02115, USA, [3]Institute for Transuranium Element, Postfach 2340, 7500 Karlsruhe 1, Germany, [4]Laboratory of Molecular Biophysics, University of Oxford, Rex Richards Building, South Parks Road, Oxford OX1 3QU, UK, [5]Biochemistry Department, Wayne State University, Detroit, MI 48201, USA, [6]Institute of Molecular Biology and Biotechnology, PO Box 1527, Heraklion 71110, Crete, Greece and [7]European Molecular Biology Laboratory (EMBL), c/o Deutsches Elektronen Synchrotron (DESY), 2000 Hamburg 52, Germany

Communicated by J.N.Jansonius

The crystal structure of *Eco*RV endonuclease has been determined at 2.5 Å resolution and that of its complexes with the cognate DNA decamer GGGATATCCC (recognition sequence underlined) and the non-cognate DNA octamer CGAGCTCG at 3.0 Å resolution. Two octamer duplexes of the non-cognate DNA, stacked end-to-end, are bound to the dimeric enzyme in B-DNA-like conformations. The protein–DNA interactions of this complex are prototypic for non-specific DNA binding. In contrast, only one cognate decamer duplex is bound and deviates considerably from canonical B-form DNA. Most notably, a kink of ~50° is observed at the central TA step with a concomitant compression of the major groove. Base-specific hydrogen bonds between the enzyme and the recognition base pairs occur exclusively in the major groove. These interactions appear highly co-operative as they are all made through one short surface loop comprising residues 182–186. Numerous contacts with the sugar phosphate backbone extending beyond the recognition sequence are observed in both types of complex. However, the total surface area buried on complex formation is >1800 Å^2 larger in the case of cognate DNA binding. Two acidic side chains, Asp74 and Asp90, are close to the reactive phosphodiester group in the cognate complex and most probably provide oxygen ligands for binding the essential cofactor Mg^{2+}. An important role is also indicated for Lys92, which together with the two acidic functions appears to be conserved in the otherwise unrelated structure of *Eco*RI endonuclease. The structural results give new insight into the physical basis of the remarkable sequence specificity of this enzyme.

Key words: DNA deformation/DNA recognition/protein–DNA interaction/restriction endonuclease

Introduction

The most remarkable property of type II restriction endonucleases is the high specificity with which they cleave double-stranded DNA at defined positions. For *Eco*RV, a change in just one base pair in its hexameric recognition sequence can reduce the ratio k_{cat}/K_m for DNA cleavage by a factor of $\geq 10^6$ (Taylor and Halford, 1989). Of the nearly 2100 type II restriction endonucleases identified from a wide variety of prokaryotes (Roberts and Macelis, 1992) only a few have been studied in detail by biochemical and biophysical techniques (Bennett and Halford, 1989). They all need Mg^{2+} as an essential cofactor for catalytic activity and hydrolyse phosphodiester bonds to leave 5′ phosphate and 3′ hydroxy groups. *In vitro*, their specificity can be relaxed under certain buffer conditions, known as 'star' conditions, such as low ionic strength, elevated pH, the presence of organic solvents or the substitution of Mg^{2+} by Mn^{2+} (reviewed in Bennett and Halford, 1989). The two biochemically best studied type II restriction endonucleases are *Eco*RI (recognition site G/AATTC, cleavage site indicated by /) and *Eco*RV (GAT/ATC) which both function as dimers. Comparison of their amino acid sequences does not reveal any significant sequence identity (Bougueleret *et al.*, 1984). Systematic studies with short synthetic DNA duplexes have shown that any single base pair substitution in the target sites of these two enzymes produces dramatic cleavage rate reductions (Lesser *et al.*, 1990; Thielking *et al.*, 1990; Alves,J. and Pingoud,A., personal communication). For *Eco*RI, large decreases occur both in substrate affinity (measured in the absence of Mg^{2+}) and in catalytic rate constant k_{cat} (Lesser *et al.*, 1990; Thielking *et al.*, 1990). For *Eco*RV, no preferential binding to target sites is observed in the absence of Mg^{2+} (Taylor *et al.*, 1991). Nevertheless similarly large rate decreases of $\geq 10^5$ are observed for all *Eco*RV star sites (sites with one base pair substitution in the recognition sequence) (Alves,J. and Pingoud,A., personal communication). The relaxation of specificity observed with many restriction endonucleases when using Mn^{2+} in place of Mg^{2+} has been investigated for the star site GTTATC of the *Eco*RV enzyme (Vermote and Halford, 1992). A dramatic decrease of specificity from a discrimination factor of 3×10^5 with Mg^{2+} to only 6 with Mn^{2+} is observed.

The first X-ray structure determination of a restriction endonuclease, that of a complex between *Eco*RI endonuclease and a cognate DNA fragment, revealed in detail its protein–DNA interactions and showed that the bound DNA deviates significantly from B-form DNA (Kim *et al.*, 1990; Rosenberg, 1991). Here we present the analysis of the structure of *Eco*RV endonuclease and of its complexes with cognate and non-cognate DNA fragments and show that considerable conformational changes occur in both protein and DNA on complex formation. The active site, which has local structural homology to that of *Eco*RI, is described in detail and candidate catalytic residues are identified. Site-directed mutagenesis experiments inspired by these structural results have already confirmed these suggestions for the *Eco*RV enzyme (Thielking *et al.*, 1991; Selent *et al.*, 1992). *Eco*RV has been shown to have a high affinity for Mg^{2+} when bound to its cognate site and a low affinity when bound to non-cognate sites (Taylor and Halford, 1989). Based on the detailed analysis of the protein–DNA interactions in the cognate and non-cognate complexes we suggest that the kinked DNA conformation of the cognate duplex is required to generate a high affinity Mg^{2+} binding site and thereby plays an important role for the discrimination against non-cognate sites. Some aspects of these structure determinations have been discussed in a preliminary fashion (Winkler, 1992).

```
        AAAAAAAAA           aaa    bbb    BBBBBBBBBBBBBB
               |         |         |         |         |
  1     SLRSDLINALYDENQKYDVCGIISAEGKIYPLGSDTKVLSTIFELFSRP

      BBBBBBBB   ccc          dddd       eeeeeeeeeee
               |         |         |         |         |
 51   IINKIAEKHGYIVEEPKQQNHYPDFTLYKPSEPNKKIAIDIKTTYTNKEN

           ffff                  ggggggggggggg
               |         |         |         |         |
101   EKIKFTLGGYTSFIRNNTKNIVYPFDQYIAHWIIGYVYTRVATRKSSLKT

                  hhhhhhhhhh    iiiii      jjjj    CCCCC
               |         |         |         |         |
151   YNINELNEIPKPYKGVKVFLQDKWVIAGDLAGSGNTTNIGSIHAHYKDFV

      C       DDDDDDDDDD              EEEEEEEEE
               |         |         |         |
201   EGKGIFDSEDEFLDYWRNYERTSQLRNDKYNNISEYRNWIYRGRK
```

Fig. 1. Amino acid sequence (single letter code) of *Eco*RV endonuclease (Bougueleret *et al.*, 1984). The mature protein starts at amino acid 2 as the N-terminal methionine of the coding sequence is cleaved off. The secondary structure assignments as derived by the algorithm of Kabsch and Sanders (1983) are indicated above the sequence with capital letters for α-helices and lower case letters for β-strands.

Results

Crystallography

The three crystal structures determined are referred to as 1RVE for the free enzyme, 2RVE for the complex with non-cognate DNA and 3RVE for that with cognate DNA. The asymmetric units of 1RVE and 2RVE contain one non-crystallographic dimer, consisting of protein chains A and B, and, in the case of 2RVE, of four octamer DNA strands (C, D, E and F) paired in two duplexes (CD and EF). In the asymmetric unit of 3RVE there are three protein subunits (chains A, B and C) and three decamer DNA strands (D, E and F). They are paired in a non-crystallographic dimeric complex (AB-DE) and in a crystallographic one (CC′-FF′) where C′ and F′ are related to C and F by a crystallographic twofold axis of symmetry along the *b* axis.

Earlier attempts to interpret a 2.5 Å electron density map of the crystal structure of the uncomplexed enzyme (1RVE) calculated with phases derived from the isomorphous and anomalous differences of a lead acetate derivative and refined by twofold molecular averaging and solvent flattening, were not successful (Winkler *et al.*, 1987). After finding a second, mercury, derivative it was realized that the non-crystallographic twofold axis was not sufficiently exact to be used for twofold averaging. However, a new map based on phases derived from the two derivatives and improved by solvent flattening still proved very difficult to interpret. Significant progress was only achieved when all data, hitherto collected on film with 5–8 crystals needed per data set, were recollected using an electronic area detector. With these data, measured from only one or two crystals per data set, much improved heavy atom refinement and phasing statistics and an easily interpretable electron density map were obtained. Subsequent to this initial structure determination, the structures of the complexes with DNA (2RVE and 3RVE) were solved by molecular replacement (MR). Experimental

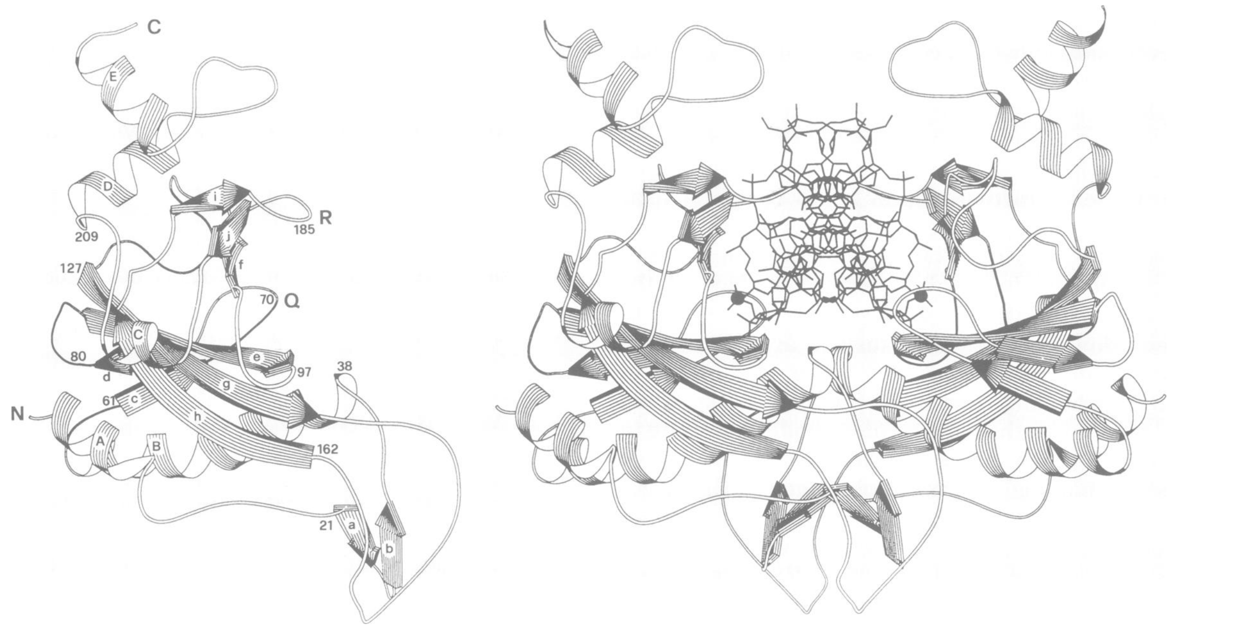

Fig. 2. *Eco*RV ribbon diagrams (Priestle, 1988). (**Left**) Monomer structure (subunit A of 1RVE) with α-helices and β-strands labelled according to Figure 1. N and C mark the N- and C-termini of the protein, Q and R the location of the Q-turn (residues 68–71) and recognition loop (residues 182–187) respectively, both important for DNA interaction. In addition the approximate positions of selected residues are indicated with residue numbers. (**Right**) Dimer structure (crystallographic dimer of 3RVE) with a stick model of the bound cognate DNA fragment. The two symmetrically disposed scissile phosphodiester groups are emphasized by black circles.

details and statistical information on the structure determinations and refinements are given in Materials and methods.

Tertiary and quaternary protein structure

*Eco*RV endonuclease has 244 amino acids (numbered 2-245, Figure 1) and is functional as a dimer. Its structure determination has revealed a mixed α/β architecture. The secondary structure assignments (Figure 1) are essentially identical for all seven crystallographically independent chains of *Eco*RV indicating that no secondary structure changes accompany DNA binding. The core of the asymmetrically shaped subunit (Figure 2) consists of a mixed parallel/antiparallel sheet formed by three long β-strands (βe, βg and βh) and this sheet is extended at one end of βe by two shorter antiparallel strands (βd and βc). The 60 N-terminal residues containing helices αA and αB and one long loop (residues 141 – 161, connecting the ends of the antiparallel strands βg and βh) are located on the bottom face of this central sheet. On the opposite face, a short triple-stranded antiparallel sheet (βi, βj and βf) and three short α-helices (αC, αD and αE) form the top side of the molecule. The U-shaped dimer (Figure 2) has dimensions of approximately $\sim 60 \times 60 \times 40$ Å^3.

*Eco*RV has a number of poorly ordered chain segments which could not be modelled (indicated in Figure 3, listed in Table IV). Three exposed chain segments, 13 – 18, 141 – 149 and 221 – 229, are disordered to a large extent in all seven monomers. The conformation of segment 34 – 37 is poorly defined in most of the seven monomers. Segments 68 – 70 and 183 – 186 show very weak density in the free enzyme structure, are better defined in the complex with non-cognate DNA and are well-defined in the complex with cognate DNA. On the other hand, the C-terminal four residues which are ordered in the free enzyme have become disordered in both complexes with DNA, indicating some intrinsic flexibility of the C-terminus.

There are significant differences in the tertiary structures of the seven monomers. Superimpositions with different subsets of α-carbon atoms show that there are two structurally conserved subdomains which are linked in different relative orientations. This structural organization is illustrated in Figure 3. The first, rather small, subdomain consists of segments 19 – 32 and 150 – 160. It forms most of the dimer interface and we term it the 'dimerization subdomain'. The other, much larger conserved substructure comprises the N-terminal helix αA and, except for segment 141 – 166, all residues from the start of helix αB to the C-terminus. All residues involved in interactions with DNA belong to this subdomain which we term 'DNA binding subdomain'. The remaining segments (13 – 18, 33 – 37, 141 – 149 and 161 – 165) link these two subdomains in a flexible fashion. The N-terminal two turns of helix αB, formally assigned to the 'DNA binding subdomain', also contribute to this flexible joint by bending away from the C-terminal part of the helix in somewhat different directions. Interestingly, residues in these two helical turns contribute directly to the dimer interface, to DNA binding and possibly to catalysis.

The three independent protein chains A, B and C in 3RVE show only modest variation in the relative orientation of their two subdomains. Much larger variations occur between the four monomers present in 1RVE and 2RVE which, in their respective crystals, form dimers with considerable deviations from twofold symmetry. Looking at the superposition of the seven monomers along the axis of helix αB (Figure 3) reveals that a major component of the interdomain flexibility correlates with different windings and orientations of the N-terminal turns of αB. The variable subdomain orientations produce differences in the relative position and orientation of the conformationally invariant residues at the two ends of each link. These are apparently accommodated through the high conformational flexibility of the three linker segments, 13 – 18, 33 – 37 and 141 – 149. For the fourth linking segment, residues 161 – 165, the required structural adaptations are smaller and the same extended conformation is observed in all seven monomers. Small concerted changes in this segment's main chain torsion angles appear to be sufficient to cope with the different interdomain orientations.

Dimerization produces a four-stranded antiparallel sheet involving strands βa and βb of both subunits (Figure 2). This pairing, including contacts between a number of buried hydrophobic side chains, is structurally very well conserved in all dimers. Given this conserved dimer interface and the observed interdomain flexibility, quite dramatic changes result at the quaternary structure level. In other words, the relative position and orientation of the two DNA binding subdomains is highly variable. The N-terminal two turns of αB also contribute to the dimer interface but as they are more part of the 'DNA binding subdomain' and follow its motion relative to the 'dimerization subdomain' their interaction is not invariant. To first approximation, two states of the αB – αB interface are observed, one in the free enzyme, the other when cognate or non-cognate DNA is bound. In the free enzyme dimer, the N-terminal parts of the antiparallel αB helices interact with each other over one additional turn of the helix, that is, each of the two helices has slid past the other by half a turn as compared with the state in which DNA is bound.

DNA conformation

The DNA conformation observed in the two types of complex is very different (Figure 4). In the non-cognate complex, the two octamer duplexes show the characteristic shape of B-form DNA with the base pair planes being essentially parallel to each other and normal to the helix axis. Seen in the same view, the cognate DNA fragment appears compressed along the helix axis and the major groove has become narrower and deeper.

The DNA conformations were analysed with the program CURVES (Lavery and Sklenar, 1989) which is designed for the analysis of irregular helices. In the non-cognate complex, the parameters for the two octamer duplexes are within the range typically observed for B form DNA fragments (Kennard and Hunter, 1989). To first approximation the two octamer duplexes appear to form a pseudo-continuous hexadecamer duplex. However, their helical axes as calculated by the CURVE algorithm do not join up smoothly. They are displaced by ~4.5 Å in the long direction of the base pairs near the twofold axis and inclined by nearly 20° with respect to each other (Figures 4 and 5). Nevertheless, the two central base pair planes are essentially parallel, 3.4 Å apart and partly stacked. The corresponding inter base-pair twist angle, calculated as if the DNA were continuous, is essentially zero.

The most remarkable conformational feature of the bound cognate DNA fragment is a central kink of ~50° (Figure 5).

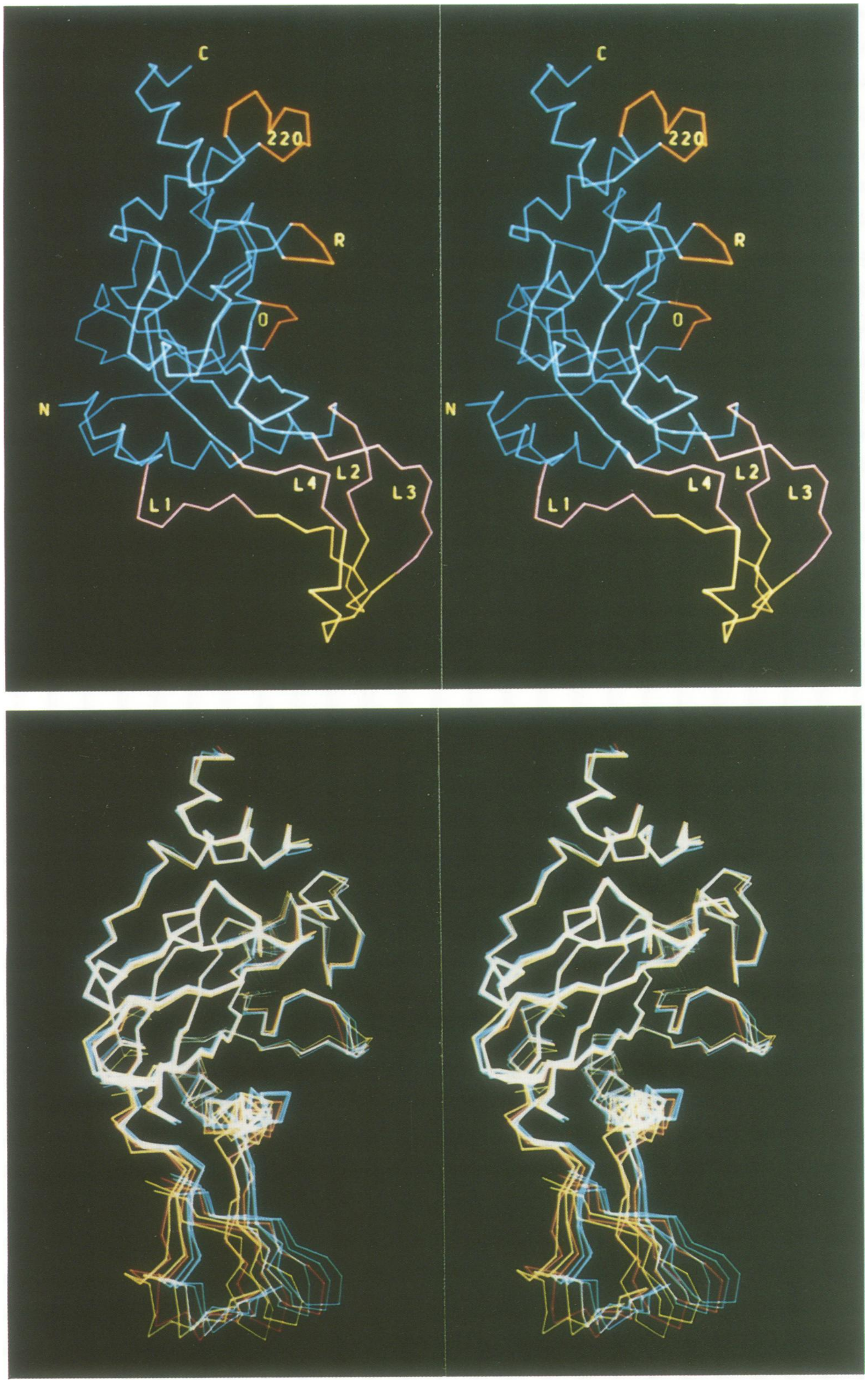

Fig. 3. *Eco*RV subdomain structure. (**Top**) Stereo diagram showing the α-carbon tracing of one *Eco*RV monomer (1RVE, subunit A, similar view as Figure 2). The dimerization subdomain is shown in yellow, the DNA binding subdomain in blue, the linking chain segments in magenta and those chain segments of the DNA binding subdomain that are poorly ordered in some or all structures (see Table IV) in orange. The four linking segments are labelled L1 to L4 and comprise residues 13–18, 33–37, 141–149 and 161–165 respectively. Other labels are defined in Figure 2. (**Bottom**) Superimposition of the seven independent monomers from 1RVE (yellow), 2RVE (red) and 3RVE (blue) after optimal matching of the DNA binding subdomains. The disordered or poorly ordered chain segments have been omitted.

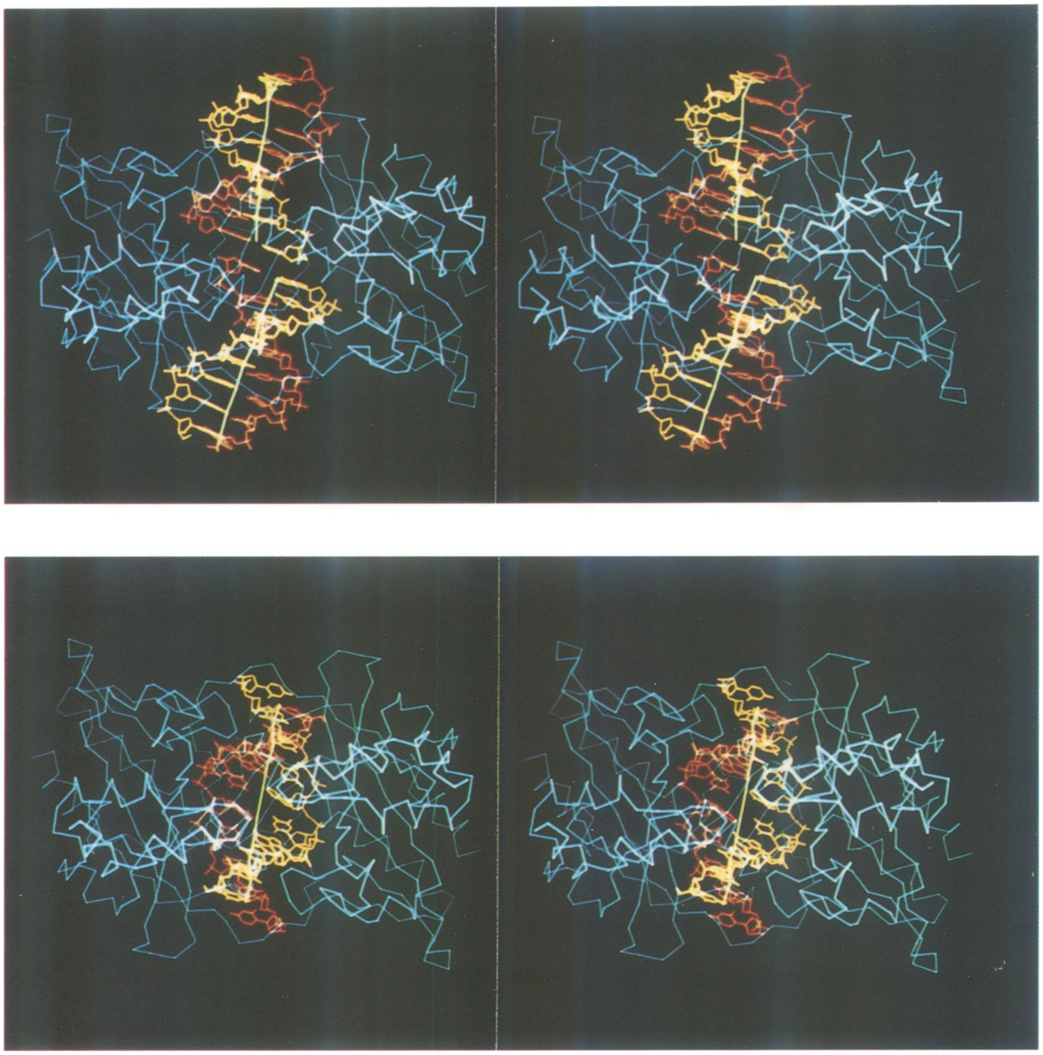

Fig. 4. Complexes of *Eco*RV with non-cognate and cognate DNA. Stereo diagram of the non-cognate complex (2RVE) seen along the approximate molecular twofold axis (**top**) and of the cognate complex (3RVE) seen along the exact crystallographic twofold axis (**bottom**). The α-carbon tracings of the protein subunits are represented in blue, the atomic models of the DNA strands in yellow and red respectively. The paths of the helical axes of the two octamer duplexes and of the decamer duplex as determined by the CURVE algorithm (Lavery and Sklenar, 1989) are indicated by yellow lines.

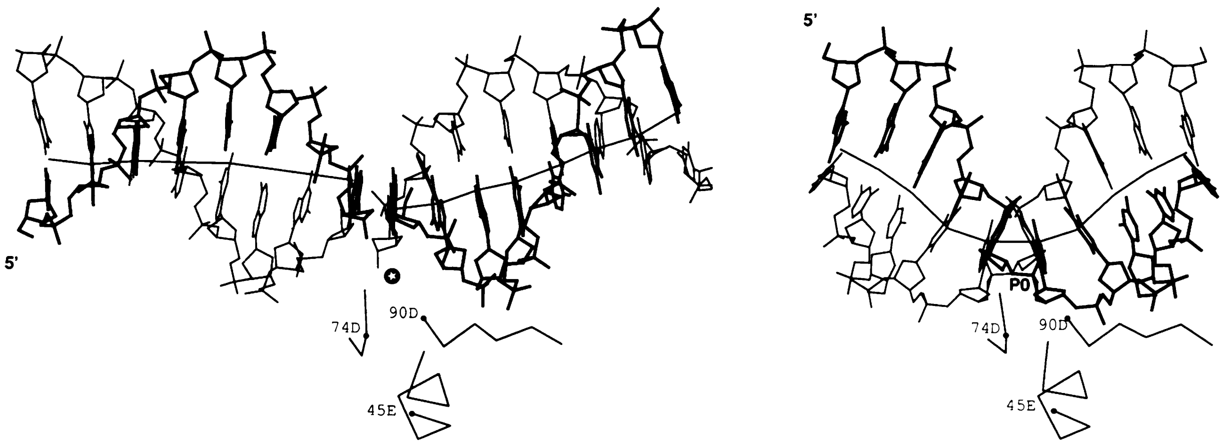

Fig. 5. Non-cognate and cognate DNA conformations. Side view of non-cognate (**left**) and cognate (**right**) DNA in their respective complexes, seen perpendicular to the quasi-twofold or twofold axis. The α-carbon positions of segments 38–45 and 73–76 and 90–95 (subunit A in 2RVE, subunit C in 3RVE) serve as common reference frames (r.m.s. error in superimposition = 0.73 Å). The DNA strands interacting predominantly with the represented protein subunits (E1–S1 interaction, Table II) are drawn in thicker lines. The scissile phosphodiester group of the cognate DNA is marked P0 just underneath the scissile O3'-P bond. The corresponding position with respect to the protein subunit of the non-cognate complex is indicated by a white star. The lines marking the paths of the helical axis are drawn to illustrate the small and large central kinking in the two complexes.

Table I. Selected helical and conformational parameters of the bound cognate DNA fragment

		Twist (°)		Rise (Å)		Roll (°)		P_i-P_i distance (Å)			Sugar pucker P (°)			Sugar torsion angle δ (°)			
		DE	FF′	DE	FF′	DE	FF′	D	E	F	D	E	F	D	E	F	
5′ dG	−5										130	194	149	117	148	145	−5
p		32.5	40.1	2.84	3.34	−1.3	6.2										
dG	−4							6.85	6.92	6.73	165	146	179	136	127	150	−4
p		33.3	32.9	3.70	3.51	−3.7	0.9										
dG	−3							6.84	6.84	6.50	147	150	156	139	139	140	−3
p		35.4	36.6	4.10	3.81	−2.4	−3.9										
dA	−2							6.48	6.59	6.89	$\underline{41}$	$\underline{26}$	$\underline{55}$	$\underline{74}$	$\underline{75}$	$\underline{95}$	−2
p		19.3	19.8	3.22	3.28	15.1	11.2										
dT	−1							5.48	6.15	5.59	$\underline{22}$	$\underline{15}$	$\underline{0}$	$\underline{76}$	$\underline{82}$	$\underline{93}$	−1
p		23.0	16.0	4.18	4.39	49.0	51.0										
dA	0							6.08	5.67	6.21	166	$\underline{43}$	$\underline{27}$	140	$\underline{81}$	$\underline{95}$	0
p		19.9	19.8	3.26	3.28	4.7	11.2										
dT	1							6.76	6.43	6.60	150	163	145	132	128	139	1
p		37.4	36.6	3.96	3.81	−5.0	3.2										
dC	2							6.44	6.60	6.46	150	138	119	137	127	123	2
p		40.4	32.9	3.22	3.51	0.4	0.9										
dC	3							6.81	6.57	6.40	148	156	165	133	140	142	3
p		29.6	40.1	3.23	3.34	−3.8	6.2										
dC	4										172	154	178	154	143	136	4

C3′-*endo*-like sugar conformations are underlined.

Selected conformational and helical parameters of the three decamer strands D, E and F of 3RVE are listed in Table I. The torsion angles of the sugar–phosphate backbone (not listed) are all within the range typically observed in the crystal structures of small DNA fragments. The kink can apparently be achieved by concerted adjustments of the backbone torsion angles within the ranges observed for A- or B-DNA fragments (Kennard and Hunter, 1989). Of the 10 furanose sugars per strand, two (in strand D) and three (in strands E and F) fitted the density better in C3′-*endo*-like rather than C2′-*endo*-like conformations. The three DNA strands D, E and F in 3RVE superimpose much better (0.54 Å r.m.s. difference for all atoms) than the four strands C, D, E and F present in 2RVE (1.08 Å). This is related to the fact that the twofold symmetry is obeyed much better in the non-crystallographic dimer of the cognate complex. It also indicates that some of the protein–DNA interactions are weak in the non-cognate complex such that crystal packing effects more easily disturb the inherent twofold symmetry. Within experimental error, strands E and F in 3RVE have essentially identical structures while the conformation of strand D appears somewhat different at and around the scissile phosphodiester group. In addition to the central kink, the cognate DNA duplexes DE and FF′ are unwound by ~45° between the central four base pairs. This is similar to the untwisting observed in the non-cognate complex between the two base pairs stacked end-to-end at the local twofold axis. Apparently, the *Eco*RV dimer contacts the DNA backbone in both cognate and non-cognate binding mode such that the central part of the DNA tends to be unwound.

The large positive roll angle in the centre of the cognate DNA makes the major groove narrower and deeper while the minor groove becomes shallower. The shape and width of the two grooves become remarkably similar to that present in short A-form DNA fragments. Despite some A-DNA characteristics of the central part of the bound decamer, the C3′-*endo* sugar conformations occur in this part as well, we prefer to describe this duplex as consisting of two 5 bp segments of B-form-like DNA joined with a kink of ~50° (Figure 5).

Two symmetrically disposed kinks with a positive roll of similar magnitude at TG (CA) steps have been observed in a complex between the catabolite activator protein (CAP) and a 30 bp DNA fragment (Schultz *et al.*, 1991). The DNA fragment in the *Eco*RI complex (Kim *et al.*, 1990; entry 1RIE in the Protein Data Bank, Bernstein *et al.*, 1977) is also essentially unstacked at the central step (AT in this case). However, the base pair roll is in the opposite direction such that the major groove opens up. In addition, the overall path of the helix remains almost linear as the negative roll between the central two base pairs is compensated at the adjacent AA steps on each side by a positive roll. It appears that with *Eco*RI the unstacking is primarily used to improve the accessibility of the major groove for direct protein–base interactions.

Protein – DNA interactions

Overall structures. In both complexes the DNA is similarly embedded in the bowl of the U-shaped protein dimer (Figures 2 and 4). The floor of the bowl is formed by the N-terminal turns of the two symmetry related, antiparallel αB helices. Regarding the two octamer duplexes in 2RVE as a continuous hexadecamer duplex, the minor groove side of the DNA is seen to face the floor of the binding site in both types of complex. This contrasts with the situation in the *Eco*RI–DNA complex (Kim *et al.*, 1990) where it is the major groove which is orientated towards the protein. A large fraction of the contacts between *Eco*RV and the cognate DNA fragment are mediated by two short loops. The loop between residues 182 and 187 embraces the central part of the DNA by reaching from both sides into the major groove. As it makes all the base-specific interactions we refer to it as the recognition loop. It is poorly ordered in both the free enzyme and the non-cognate complex. The other loop, forming a β-turn, interacts predominantly with the

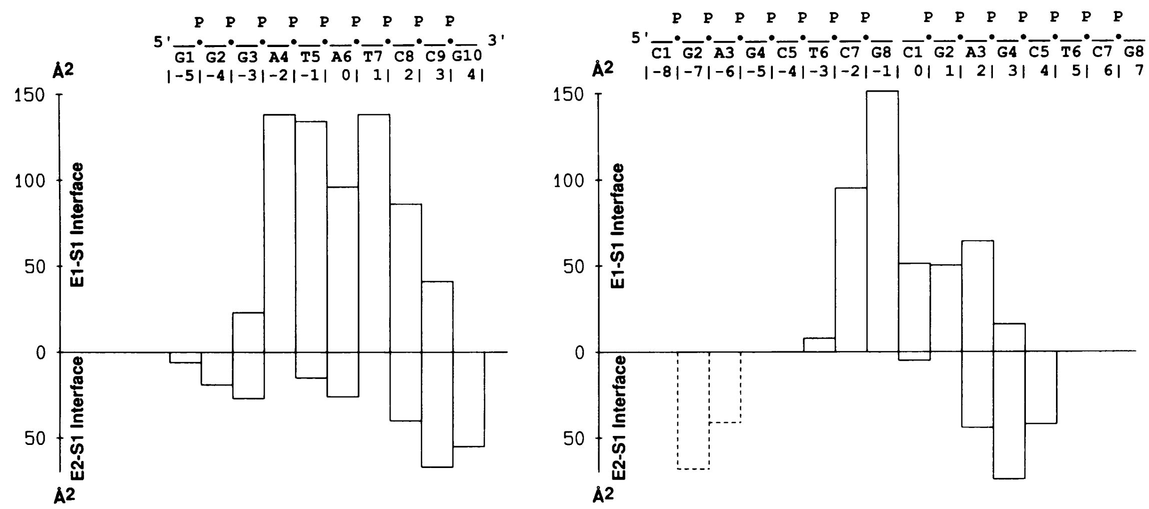

Fig. 6. Protein contact areas of DNA residues in the cognate (**left**) and non-cognate (**right**) complexes. Contact surface areas for the residues of a bound DNA strand (taken as S1) with the two protein subunits E1 and E2 of the *Eco*RV dimer. The S1–E1 interface values are plotted upwards, those of the S1–E2 interface downwards. Average values for the equivalent strands (D, E and F in 3RVE and F,C and D,E in 2RVE respectively) are plotted. The contacts at the 5′ end of the non-cognate complex (residues −7 and −6; dashed lines) occur only in the case of the F,C strand and are not averaged. The areas were calculated as described in Table II.

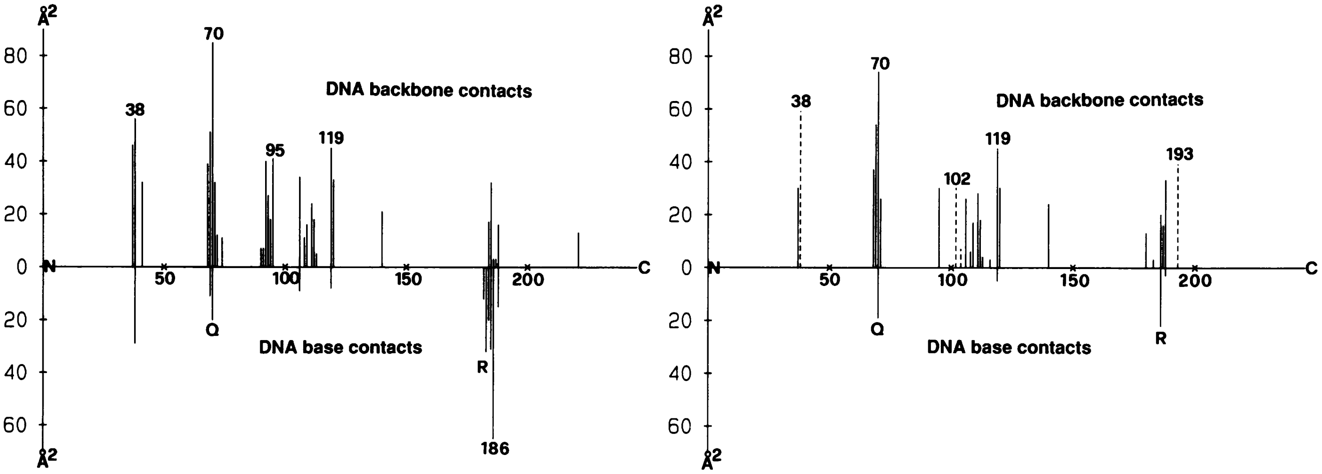

Fig. 7. DNA contact areas of the protein residues in the cognate (**left**) and non-cognate (**right**) complexes. Contact surface areas for the residues of a protein subunit (taken as E1) with the bound DNA duplex (E1–S1S2 interfaces). Contacts with the sugar–phosphate backbone are plotted upwards, those with the base pairs downwards. Average values for the equivalent subunits (A, B and C in 3RVE and A, B in 2RVE respectively) are plotted. Q and R mark the minor and major groove contacts of the Q turn and recognition loop respectively. Residue numbers are indicated for some peak contacts. The dashed lines indicate contacts that occur only in one subunit and have not been averaged.

sugar–phosphate backbone in both types of complex and contacts the bases in the minor groove at both ends of the recognition sequence. It comprises residues Gln68, Gln69, Asn70 and His71 and we have termed it the Q turn because of its two glutamine residues. It has been fitted as a type I β-turn (Wilmot and Thornton, 1988) in the three subunits present in 3RVE and as a type II β-turn in those of 2RVE. However, the differences in the fit to the electron density for the two alternative turn conformations are barely significant at the available resolution of 3.0 Å. In the free enzyme the Q loop is poorly ordered indicating that its conformation is variable in the absence of DNA.

In order to gain a more quantitative picture of the importance of various intermolecular interactions within the dimeric complexes, we have calculated the corresponding buried surface areas (Table II). In this analysis, we refer to the two protein subunits of a dimer as E1 and E2 and to the two bound DNA molecules as strands S1 and S2. Note that in the case of the non-cognate complex, S1 and S2 refer to the two pseudo-hexadecamer strands. As they are composed of two octamer strands lacking 5′ phosphate groups there is no phosphate group in the non-cognate complex equivalent to the scissile phosphodiester group.

The protein–protein dimer interface (E1–E2) buries ~2240 Å^2 of surface area in both the free enzyme and in the non-cognate complex and 2470 Å^2 in the cognate complex. The increase results almost entirely from the additional contacts formed between the two recognition loops across the major groove. About two-thirds of the E1–E2 contacts are provided by non-polar atoms. In each complex, the protein–DNA interface (E1E2–S1S2) buries a considerable amount of surface area, but in this case more polar than non-polar atoms are buried in the contact areas (60% versus 40%). Including the additional protein–protein contacts formed in the cognate complex, >1800 Å^2 more of total surface area is buried more when DNA is bound in

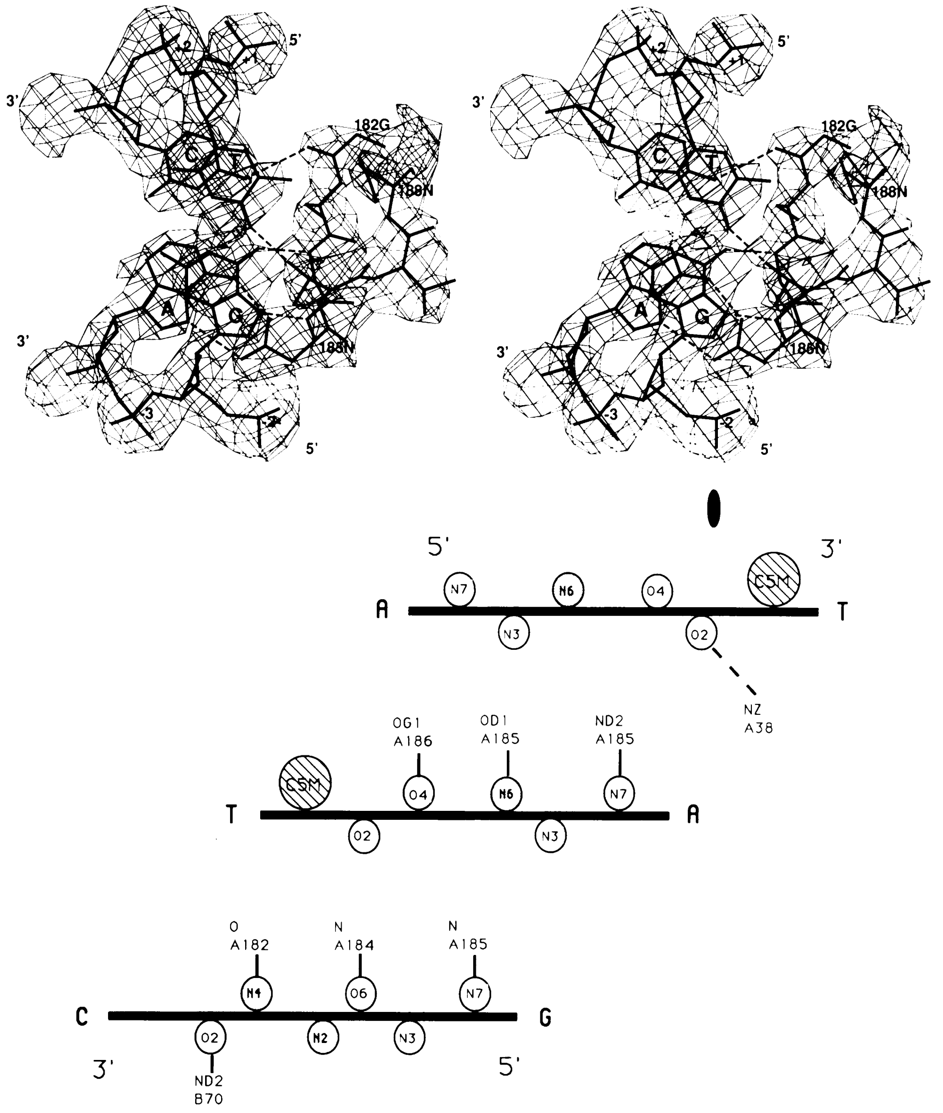

Fig. 8. Cognate DNA base recognition. (**Top**) Stereo diagram with electron density (coefficients $2F_{\text{obs}} - F_{\text{calc}}$, contoured at 1.2 × r.m.s. density) showing the hydrogen bonding interactions (dashed lines) between the recognition loop of subunit C (3RVE) and the GC and AT base pairs of the corresponding GAT half site. (**Bottom**) Schematic diagram illustrating the hydrogen bonds to the canonical base pairs (drawn as thick horizontal bars) in the major and minor grooves of the bound cognate decamer GGGATATCCC. Only one half of the palindromic recognition site is shown with the twofold symmetry axis indicated by a black oval. With one exception all hydrogen bonds (drawn as connecting vertical lines) with this half site are made by one protein subunit (A). The hydrogen bond donor and acceptor atoms (shaded and blank circles respectively) of the bases in the major and minor groove are shown above and below the base pair bars respectively. The long hydrogen bond (4 Å) between Lys38 and O2 of the central thymine is shown as a dashed line. The thymine methyl groups (large hatched circles) are both located in loose hydrophobic pockets.

the cognate than in the non-cognate binding mode. About one-third of this increase results from the protein–base-pair contacts in the major groove of the target site.

Protein–DNA backbone interactions. Calculated contact areas between DNA and protein in the two kinds of complex are plotted for DNA residues in Figure 6 and for amino acid residues in Figure 7. The DNA residues, carrying a 5′ phosphate group, are assigned numbers increasing from the 5′ end, with residue 0 defined as containing the scissile phosphate group. Strikingly similar contact patterns are observed for the two kinds of complex. The *Eco*RV dimer can thus interact with B-form DNA and with the kinked cognate DNA in an overall similar fashion. In the crystal structure of the cognate complex, the three crystallographically independent, equivalent sets of protein–DNA interactions are very similar; the average values are plotted in Figures 6 and 7. Extended DNA-backbone–protein contacts, exceeding 80 Å^2 contact area per DNA residue, occur in the central part of the decamer strand between residues −2 and +2. They are made predominantly with one of the two subunits and we define this as the S1–E1 interface (Table II). As will be shown, E1 carries the catalytic site that cleaves S1. In this interface there are 10 hydrogen bonds between the protein and the DNA phosphates. Six of these are with donors from the four contiguous residues 92–95 belonging to strand βe. The ammonium group of Lys92 lies between the two adjacent phosphates of residues 0 and +1.

Despite the poor twofold symmetry of the non-cognate complex, the major contacts with the protein are conserved in both equivalent interfaces. Averaged values are shown in Figure 6. Some clear differences, however, exist, the most significant being the additional contacts formed by residues

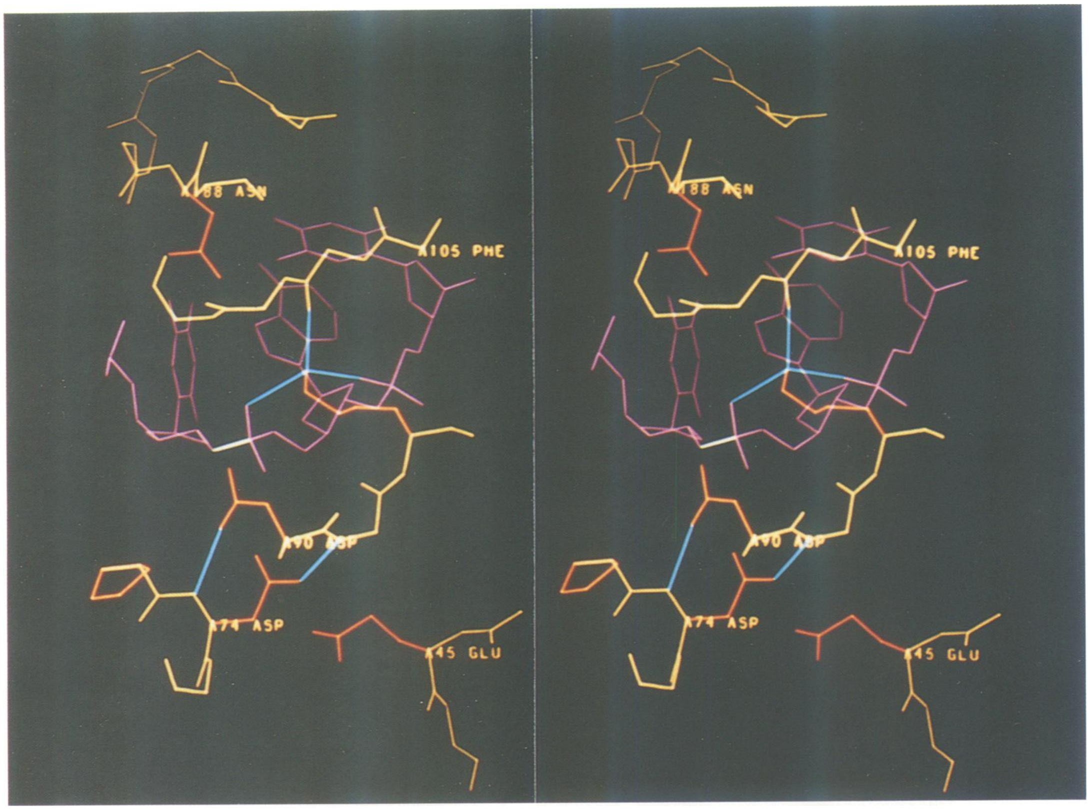

Fig. 9. Stereo diagram of the active site region in the cognate complex (subunit A, 3RVE). Main chain protein segments (44–46, 73–75, 90–92, 105–108 and 182–189) are shown in yellow. Selected side chains are shown in red and their hydrogen bonding contacts in blue. The DNA residues –1, 0 and +1 (TAT) are shown in magenta and the scissile O3′-P bond is emphasized in white.

–7 and –6 of the F strand. These contacts involve Lys102 and His193 and are part of the S1–E2 interface. Equivalent contacts might occur in cognate complexes with longer DNA fragments. Extension of the model of the cognate decamer duplex with B-DNA brings the side chains of Lys102 and His193 close to the DNA backbone at residues corresponding to positions –7 and –6. At residues –2 and –1 the contact areas in the non-cognate complex are similar in size to those in the cognate complex. At the level of detailed interactions, however, the corresponding contacts are very different in the two types of complex. As best seen in Figure 5, the sugar–phosphate backbones of the cognate and non-cognate S1 strands are displaced by nearly 3 Å along the helical path after optimal superimposition of their associated E1 DNA binding domains. 3′ to the missing scissile phosphate, the non-cognate S1–E1 interface is rather tenuous and this is the other place where there are clear differences between the two equivalent parts of this complex. Unlike in the cognate complex, the backbone of residues 0 to +2 does not run close and parallel to strand βe of E1. Only the long side chains of Tyr95 and Arg140 of the E1 subunit still reach the S1 backbone at phosphates +2 and +3 respectively. In addition, Thr37, which contacts the +1 sugar moiety in the cognate complex, can now make a hydrogen bond to the +2 phosphate. On the other hand, the contact area involving residues +2 to +4 of S1 and the Q turn of E2 is again very comparable to that observed in the cognate complex. 31% of the total protein–DNA interface in the non-cognate complex is provided by the contacts with the Q turn as against 22% in the case of the cognate complex.

Major and minor groove contacts. Figure 7 shows the amino acid residues which contact the bases in the two grooves of the bound DNA. In the cognate complex, all sequence-specific hydrogen bonds are made through the major groove and they all involve the recognition loop comprising residues 182–187 (Figure 8). Each loop forms six hydrogen bonds with the outer two base pairs of one GAT half site while no hydrogen bond contacts are made to the inner TA base pair. The three hydrogen bonds to the GC base pair are formed with one main chain oxygen and two main chain nitrogen atoms. The hydrogen bond partners to the adjacent AT base pair are provided by the side chains of Asn185 and Thr186. An additional protein–protein contact is formed across the twofold axis and involves the two symmetry-related methyl groups of Thr186. The two interacting recognition loops cover the major groove side of the six recognition base pairs. The participation of main chain atoms in hydrogen bond formation, the small size of the recognition loop residues (Gly182-Ser183-Gly184-Asn185-Thr186-Thr187) and the presence of interactions within and between symmetry-related recognition loops generates a highly co-operative set of interactions. The two central base pairs are shielded from solvent access on their major groove side by their outer base pair neighbours and by the two recognition loops. They make no hydrogen bonds to the protein. Direct readout of all six base pairs appears sterically difficult if not impossible in this case because of the compression of the major groove through the large central kink.

In the minor groove, the only possible protein–DNA hydrogen bonds are between ND2 of Asn70 and O2 of the cytosine of the GC base pair (in each half site) and/or N3 of the adenine of the adjacent AT base pair. Other hydrogen bonding contacts may be mediated through water molecules. The only other close contact between protein and cognate DNA in the widened minor groove involves Lys38. The two dyad-related lysine side chains run antiparallel underneath the wedge-shaped opening between the two unstacked base pairs. Each ammonium group appears to form an intramolecular salt bridge with Asp36. While these minor groove contacts may contribute some favourable binding energy in the cognate complex, they cannot provide much discrimination against other DNA sequences which could form very similar contacts.

The active site

Like many nucleases (Saenger, 1991), *Eco*RI and *Eco*RV endonucleases cleave with inversion at the phosphorus atom (Connolly *et al.*, 1984; Grasby and Connolly, 1992). The most simple and common mechanism compatible with this finding is an SN2 type reaction, i.e. attack of an activated water molecule in line with the 3′-OH leaving group. Divalent metal ions, ligated by one of the non-esterified phosphodiester oxygens, are thought to play an essential role in this mechanism through the polarization of the reactive phosphate group and/or the stabilization of the pentavalent, doubly charged transition state. In most cases, acidic residues are needed to form a high affinity metal ion binding site and are therefore essential for catalysis. In addition, one may expect a general base and a general acid, suitably positioned, to deprotonate the attacking water and to protonate the leaving group respectively.

As shown in Figure 9, the side chains of Asp74, Asp90, Lys92 and Asn188 are found close to the reactive phosphate group in the complex with cognate DNA. The three charged residues and the β strand with Asp90 and Lys92 appear structurally and functionally conserved in the otherwise unrelated *Eco*RI structure (Rosenberg, 1991; Winkler, 1992). Inspired by these findings, Selent *et al.* (1992), have mutated the *Eco*RV residues; the observed dramatic reductions of the nucleolytic activity confirm their importance for catalysis. Furthermore, we have suggested that the two carboxylates and the reactive phosphate group participate in the formation of the Mg^{2+} binding site (Selent *et al.*, 1992; Winkler, 1992). In each of the seven independent monomers present in the three different crystal structures this region of the protein is well determined and all structures superimpose very well. Asp74 and Asp90 make intramolecular hydrogen bonds with the main chain nitrogens of residues Ile91 and Asp74 respectively. The ammonium group of Lys92 bridges between phosphates 0 and +1 in the cognate complex and appears to make a third good hydrogen bond to the main chain oxygen of residue Thr106. It also is in van der Waals contact with the carboxylate of Asp90 but in this case the hydrogen bonding geometry is poor. Close to the carboxylate of Asp74 there is a third acidic residue, Glu45, whose side chain appears partly mobile as indicated by relatively high thermal parameters. Asp74 and Glu45 form the binding site for the Pb^{2+} metal ion used as a heavy atom derivative. In addition, the large rate reduction of 10^4 observed for the E45A mutant with both Mg^{2+} and Mn^{2+} is indicative of a catalytic residue (Selent *et al.*, 1992). The presence of two adjacent metal binding sites is not uncommon in metal-dependent enzymes, examples being the 3′−5′ exonuclease of DNA polymerase I (Beese and Steitz, 1991) or xylose isomerase (Jenkins *et al.*, 1992). An important function of Glu45 is indicated, but its precise role remains to be determined. In the structure of the *Eco*RI−DNA complex, there is no counterpart for Glu45.

Given the conserved rigid structure of the active site region in all seven monomers we consider it unlikely that large structural rearrangements will take place upon binding of the essential Mg^{2+} ion. Mg^{2+} is known to have a strong preference for regular octahedral coordination, and bidentate ligation of carboxylates to Mg^{2+}, apart from being incompatible with the above requirement, does not occur in Mg^{2+} complexes determined at very high resolution (Glusker, 1991). We think that the scissile phosphodiester group, Asp74 and Asp90 each provide one oxygen ligand for the Mg^{2+} ion.

No obvious candidate for the general base is seen close to the position from where a water molecule would have to be positioned for in-line attack. All the acidic side chains appear to be too far away and Lys92, although ideally positioned, is in direct contact with three negatively charged groups rendering the required substantial lowering of its pK unlikely. It has also been suggested that the adjacent (+1) phosphodiester group could activate the water (Jeltsch *et al.*, 1992) but we see no obvious reason why its pK should be much elevated from its normal value below 2.

Discussion

Conformational flexibility and DNA binding modes

The way in which *Eco*RV endonuclease interacts with DNA through two β-turns and other structural components represents a novel arrangement of DNA binding elements and does not resemble any of the other known DNA recognition motifs such as the helix−turn−helix, zinc finger, basic region−leucine-zipper or β-ribbon (Steitz, 1990; Freemont *et al.*, 1991; Harrison, 1991).

The structures of free and DNA-bound *Eco*RV endonuclease show that this dimeric enzyme can undergo large changes in its quaternary structure through motions between two flexibly linked functional subdomains. Two of the four linking segments are highly disordered in all seven crystallographically independent monomers. The large deviations from twofold symmetry observed in some of the non-crystallographic dimers are most likely caused by crystal packing forces and indicate that these quarternary structure transitions do not cost much energy.

There is experimental evidence that *Eco*RV locates its target sequence through facilitated diffusion along DNA (Taylor *et al.*, 1991). The observed structural flexibility may serve several purposes in this process. The enzyme can readily respond to changes in the local DNA structure and might translocate from one non-cognate binding site to the next by moving one of its two DNA binding subdomains at a time. Furthermore, the flexible DNA recognition loop can probe the local DNA sequence during this process and a smooth transition to the cognate binding mode can easily be envisaged on encounter of the target sequence. In most known protein−DNA complexes (Steitz, 1990; Freemont *et al.*, 1991) it is possible to dock the DNA into its pre-formed binding site without any great problems. This is not the case for the observed *Eco*RV−DNA complexes. The flexibility of its quaternary structure and, in particular,

Table II. Protein–protein and protein–DNA contact surfaces (Å^2)

Interface	1RVE	2RVE		3RVE		
	AB	AB		AB	CC′	
E1–E2	1115 (66)	1125 (69)		1244 (67)	1227 (68)	
		A-FC	B-DE	AE	BD	CF
E1–S1		451 (40)	423 (38)	764 (41)	729 (38)	757 (36)
		A-DE	B-FC	AD	BE	CF′
E2–S1		187 (32)	309 (37)	335 (44)	346 (46)	336 (46)
		AB-FCDE		AB-DE	CC′-FF′	
E1E2–S1S2		1370 (37)		2173 (41)	2186 (39)	
E1E2–B1B2		141 (11)		471 (35)	444 (31)	

A dimeric complex is defined as consisting of the protein subunits E1, E2 and the DNA strands S1,S2 (B1,B2 base pairs only). The actual protein chains or DNA strands involved in each interface are indicated by the respective chain identifiers A–F (see Table IV). Contact areas were determined by calculating the molecular surface area occluded from contact with a water probe (radius 1.4 Å) by the bound respective molecule or molecular fragment. Each value is the average of rolling the water probe over the two molecular surfaces involved in each contact. Values in parenthesis give the percentage of apolar atoms involved in each contact. The disordered chain segments (Table IV) were omitted from these calculations. Buried surface areas are twice the values listed.

of the recognition loop are certainly important for a sufficient opening of the DNA binding cleft during the initial association of protein and DNA.

In addition to the recognition loop, two other surface segments of the DNA binding subdomain show conformational variability in some or all of the seven independently determined monomers. One is the Q turn which contacts the DNA across the minor groove in both types of complex. The variable turn conformation with its two mobile glutamine side chains appears ideally suited to optimize the interactions with a variety of local DNA structures. The other, almost completely disordered segment comprises residues 222–229. At its N-terminus, it lies close to the 3′ end of the sugar–phosphate backbone of the decameric cognate DNA. Contacts between this segment and a longer DNA fragment appear possible and may induce it to assume a preferred conformation. Indeed, the change of Arg221 to Gln results in a 30-fold increase in K_m for DNA cleavage at the recognition site but no real change in k_{cat} [0.6 min^{-1} instead of 0.9 min^{-1} (Vermote, 1991)].

The structure of the cognate complex, although observed in crystals grown in the absence of Mg^{2+}, undoubtedly represents a state very close to that with productively bound substrate. The strongly kinked DNA conformation and the observed protein–DNA interactions, in particular the base-specific hydrogen bonds, characterize the cognate DNA binding mode. On the other hand, it can be questioned whether the non-cognate complex with two octamer duplexes is representative of a non-cognate complex with continuous DNA strands. The 4 Å dislocation between the helical axes of the two end-to-end stacked octamer duplexes is structurally impossible with continuous DNA. However, the small untwisting and kinking observed in this complex, both in the direction of those present in the cognate complex, make functional sense and may be expected to be found in a continuous non-cognate duplex as well. More importantly, though, we think that the strong contacts with the sugar–phosphate backbone of residues −2 and −1 (Figure 6) are likely to be structurally preserved in a true non-cognate complex. Modelling the 3′ extension of a strand fixed in this way shows that the phosphodiester group of residue 0 then falls 2.5–3 Å short of the position it occupies in the cognate complex. It is indeed possible to construct a symmetric complex with continuous B-form DNA (F.K.Winkler, unpublished results). The majority of the protein–DNA interactions as well as the small kinking and central unwinding are maintained but the lateral 4 Å dislocation is removed. We therefore think that the salient features observed in the 2RVE complex are representative of the true non-cognate DNA binding mode.

Alternatively, the structure of the 2RVE complex could be regarded as a model for a product–enzyme complex with TCG-CGA in place of GAT-ATC and lacking the central 5′ phosphate groups. If so, this would imply that after cleavage the complex switches back into the non-cognate DNA binding mode before product dissociation.

Comparison of the observed DNA–protein interface with biochemical data on binding and recognition

*Eco*RV binds to all DNA sequences, cognate and non-cognate, with equal affinity in the absence of Mg^{2+} (Taylor *et al.*, 1991). This means that the excess binding energy resulting from the considerably larger interaction interface in the cognate complex (Table II) is compensated by similarly large, unfavourable energy terms. Major contributions must come from the DNA deformation and in particular from the unstacking between the two central base pairs. Stacking enthalpies for the 10 possible dinucleotide steps in DNA duplexes derived from melting curves and from theoretical calculations differ appreciably (Delcourt and Blake, 1991, and references cited therein). Taking the experimentally derived ΔH values, unstacking at the TpA step may cost 6–8 kcal/mol. Another unfavourable local conformation is indicated at the preceding ApT step. An anomalously low twist angle of ~22° (see Table I) is observed at all three crystallographically independent examples of this step. The associated energy cost is, however, difficult to estimate. The fact that six potential hydrogen bonds, those to the central two base pairs in the major groove, are not realized may represent another unfavourable term. In the non-cognate binding mode these sites are solvent accessible. Again, the associated energy penalty is difficult to estimate and relevant experimental studies are lacking.

The *Eco*RV–DNA interface and in particular the recognition contacts have been probed with modified DNA substrates (Fliess *et al.*, 1988; Mazzarelli *et al.*, 1989; Newman *et al.*, 1990) and with mutant proteins (Thielking *et al.*, 1991; Vermote *et al.*, 1992). There is very satisfactory agreement in the sense that modifications which directly perturb the observed interactions produce large reductions in the cleavage rates. In some cases, however, such large decreases are observed in the absence of severe stereochemical conflicts. An important finding in the most detailed of such studies (Newman *et al.*, 1990) was that the large decreases in k_{cat}/K_m were almost entirely the result of a largely decreased catalytic rate constant. This was a first indication that the specificity of *Eco*RV manifests itself only in the steps after substrate binding, that is upon Mg^{2+} binding and in the subsequent steps of phosphodiester hydrolysis.

Biologically, the most relevant substrate modification studies are those where each base pair of the target site has been individually replaced by one of the three other possible base pairs (Alves,J. and Pingoud,A., personal communication). With each single base pair substitution large cleavage rate reductions of 10^4-10^6 were observed. For the outer two base pairs these substitutions disrupt the perfect match of hydrogen bonding and van der Waals contacts with the recognition loop. The tight packing of this interface and the strong coupling between the individual interactions within this short loop may explain why the effects are so large. However, the similarly large effects with substitutions of the third, innermost base pair of the GAT half site are more difficult to rationalize. Thermodynamic, biochemical, structural and theoretical studies indicate that the TpA step is the least stable of the 10 possible dinucleotide steps in double-stranded DNA (Saenger, 1984; Travers and Klug, 1990). It would appear natural for *Eco*RV endonuclease to use this in the discrimination against other central dinucleotide sequences. Based on the experimentally derived stacking enthalpies (Delcourt and Blake, 1991), the resulting discrimination ranges from 0.3 kcal/mol against a TpT (ApA) step to at most 2 kcal/mol against a GpC step. This is considerably less than the discrimination actually observed (Alves,J. and Pingoud,A., personal communication) and additional factors must contribute. The large positive roll at this base pair step brings the O4 and N6 atoms of the two adjacent T and A bases into close proximity. This close contact is electrostatically favourable only for opposite partial atomic charges and thus would certainly discriminate against CpA or TpG steps. Further discrimination may result from differences in the stacking energy with the preceding AT base pair, in particular because the low twist angle, of ~20° indicates an unfavourable local conformation.

The structural basis of specificity

The remarkable specificity of *Eco*RV endonuclease has been demonstrated by measuring cleavage rates at cognate and non-cognate sites located on plasmid DNA (Taylor and Halford, 1989) and on short synthetic fragments (Alves,J. and Pingoud,A., personal communication). Unlike the *Eco*RI enzyme (Lesser *et al.*, 1990; Thielking *et al.*, 1990), *Eco*RV shows no specific binding in the absence of Mg^{2+} (Taylor *et al.*, 1991). On the other hand, the apparent affinity for Mg^{2+} is high for cleavage at cognate site ($K_{d,Mg}$ < 1 mM) but very low at non-cognate sites ($K_{d,Mg}$ >> 10 mM) (Taylor and Halford, 1989). Halford *et al.* (1993) suggest that the intrinsic activity of *Eco*RV at cognate and non-cognate sites could be very similar and that the observed rates are essentially determined by the fractional saturation of the relevant enzyme–DNA complex with Mg^{2+}. This could not be directly verified experimentally as Mg^{2+} concentrations much above 10 mM introduce adverse effects (Record *et al.*, 1977; Halford *et al.*, 1993). An alternative to higher Mg^{2+} concentrations is to use Mn^{2+}. Although catalytically ~20 times less effective than Mg^{2+}, Mn^{2+} binds with higher affinity as evident from the cleavage rates measured with Mg^{2+}/Mn^{2+} mixtures (Vermote and Halford, 1992). Indeed, it was observed that manganese-supported cleavage at the non-cognate site GTTATC is nearly as fast as that at the cognate site. Furthermore, $K_{d,Mn}$ for cleavage at GTTATC is such that the metal binding site is essentially saturated at the concentration (10 mM) used in the experiment. These results seem to confirm that recognition is primarily coupled to Mg^{2+} affinity and that, once Mg^{2+} and DNA are productively bound, the recognition interactions have no further effect on the energetics of the subsequent bond-making and bond-breaking steps. Structurally, productive binding is achieved when a phosphodiester group of the DNA substrate is positioned in the active site such that it provides a non-esterified oxygen ligand to complete the Mg^{2+} binding site together with the carboxylates of Asp74 and Asp90. At cognate sites, the energetically unfavourable deformation of the DNA is compensated by the many additional interactions in the cognate binding mode. At any non-cognate site, however, enforcing the productive binding mode costs energy because the tightness of the recognition interface and the co-operative nature of its interactions make it impossible to relax such perturbations locally at a modest energy cost. The question arises as to what extent the dramatic relaxation of specificity reported for *Eco*RV with Mn^{2+} in place of Mg^{2+} (Vermote and Halford, 1992) is only apparent (in the sense that the stronger intrinsic binding of Mn^{2+} simply results in a higher concentration of productive complex for the same metal ion concentration). The apparent Mg^{2+} and Mn^{2+} dissociation constants with cognate DNA substrates are both significantly lower than 1 mM (Taylor and Halford, 1989) but have not been determined more precisely. It appears unlikely, however, that they differ by more than a factor of 10–100 as found for example for xylose isomerase (van Tilbeurgh *et al.*, 1992). The corresponding amount of stabilization through such tighter Mn^{2+} binding is hardly sufficient to account for the large rate difference at the non-cognate site GTTATC observed with Mg^{2+} and Mn^{2+} respectively. Any other effect enhancing the fractional saturation of productively bound substrate, such as, for example, the presence of an additional Mn^{2+} (but not Mg^{2+}) binding site in the cognate binding mode, could account for the apparent discrepancy. Alternatively, there could be true relaxation of specificity with Mn^{2+} in the sense that the energy difference between productive binding at cognate and non-cognate sites is less than with Mg^{2+}. However, this is difficult to rationalize structurally. Based on the difference in ionic radius, a productively bound phosphodiester group would be some 0.3 Å further away from the essential aspartates of *Eco*RV in the case of Mn^{2+} as compared with Mg^{2+}. In addition, the less crowded coordination shell of Mn^{2+} may be more tolerant of violations of octahedral geometry. Perhaps this permits relaxation of some of the strain introduced by a star substitution or by a mutation in the recognition loop without reducing the metal affinity as drastically as in the case of Mg^{2+}.

In conclusion, the structural and functional studies on *Eco*RV endonuclease have given us considerable insight into the origin of the high specificity of this enzyme. More work is needed, though, to elucidate the details of the catalytic

Table III. Data collection and MIR statistics

1RVE	Apoenzyme	$P2_12_12_1$	a = 58.2 Å b = 71.7 Å c = 130.6 Å	
2RVE	Complex with non-cognate DNA (CGAGCTCG)	$P2_1$	a = 68.4 Å b = 79.6 Å c = 66.4 Å	β = 104.6°
3RVE	Complex with cognate DNA (GGGATATCCC)	$C222_1$	a = 60.2 Å b = 78.4 Å c = 371.3 Å	

Diffraction data

Stucture code		$d_{\min}$ (Å)	No. of crystals/ measurements	No. of unique reflections	Completeness of data (%)	R_{merge}[a]	Comments[d]
1RVE	Native	2.50	4/86 437	16 509	84 (95 to 2.70 Å)	0.069	RT/X100/XENGEN
	$PbAc_2$	2.50	1/43 206	15 957	81 (92 to 2.70 Å)	0.073	RT/X100/XENGEN
	Chlormerodrin	2.50	2/41 395	16 654	85 (95 to 2.70 Å)	0.069	RT/X100/XENGEN
2RVE	Native	2.80	1/32 155	17 027	99	0.047	RT/X100/XDS
	$PbAc_2$	3.16	1/17 403	8974	76	0.057	3°C/X100/XDS
	Chlormerodrin	2.83	1/32 495	15 422	92	0.068	RT/X100/XDS
3RVE	Native	3.00	1/57 567	15 988	90	0.073	RT/IP/MOSFLM

MIR analysis

Structure code		resolution range (Å)	mean fractional isomorphous difference	R_{Cullis}[b] (no. of centric reflections)	No. of sites major/minor	r.m.s. F_h/E_{iso}[c]
1RVE	$PbAc_2$	25.0–2.78	0.16	0.495 (1932)	2	2.4
	Chlormerodrin	25.0–2.78	0.21	0.485 (1946)	3	2.3
	Mean figure of merit: 0.70 for 13 782 reflections (25.0–2.78 Å)					
2RVE	$PbAc_2$	25.0–3.16	0.25	0.601 (337)	2/1	1.9
	Chlormerodrin	25.0–3.00	0.22	0.563 (496)	2	1.7
	Mean figure of merit: 0.47 for 13 383 reflections (25.0–3.00 Å)					

[a] $R_{\text{merge}} = \Sigma_h\Sigma_i|I(h)_i - <I(h)_i>|/\Sigma_h\Sigma_i I_{hi}$ where $I(h)_i$ is the ith measured intensity of reflection, *h*, and $<I(h)>$ its mean intensity.
[b] $R_{\text{Cullis}} = \Sigma_h|F_{h(\text{obs})} - F_{h(\text{calc})}|/\Sigma_h \mathbf{F}_{h(\text{obs})}$ (centric reflections only)
[c] F_h, heavy atom structure factor; E_{iso}, residual lack of closure
[d] Data measured at (RT = room temperature; °C), collected with device (X100 = Siemens-Nicolet X100 area detector, IP = image plate) and processed with software Xengen (Howard *et al.*, 1987), XDS (Kabsch, 1988) or MOSFLM (Leslie *et al.*, 1986).

mechanism and to rationalize more quantitatively the wealth of kinetic data obtained with different substrates and mutant enzymes. The local active-site similarity to *Eco*RI contrasts with the major structural and mechanistic differences in the way these two enzymes recognize their target sequences (Heitman, 1992; Halford *et al.*, 1993). Nevertheless, the requirement for DNA deformation in order to generate a high affinity Mg^{2+} binding site may be shared by both enzymes, although it appears less important for discrimination in the case of *Eco*RI. Like *Eco*RV, the enzyme *Taq*I (Zebala *et al.*, 1992) also shows little specificity in binding to its recognition site and *Eco*RV may well turn out to be an archetype for a number of type II restriction endonucleases.

Materials and methods

Crystallization and data collection

Purification and crystallization of *Eco*RV endonuclease have been described by D'Arcy *et al.* (1985) and the crystallization of its complexes with DNA fragments by Winkler *et al.* (1991). Data for the crystals of the uncomplexed enzyme and of the complex with the non-cognate octamer were collected with a Nicolet-Siemens X100 area detector using CuKα radiation produced by an Elliott GX21 rotating anode generator equipped with a graphite monochromator. Data frames of 0.1° or 0.2° with exposure times between 60 and 120 s were measured. The area detector data were processed using the Xengen (Howard *et al.*, 1987) or XDS (Kabsch, 1988) software packages. The data of the cognate complex crystals were collected with the image plate system at the EMBL outstation using synchrotron radiation (beam-line X31 of the DORIS storage ring at DESY, Hamburg). Integration of these intensities was carried out using the MOSFLM program system. Statistical information on data collection and processing is given in Table III.

Structure determinations

EcoRV endonuclease (1RVE). Native crystals were soaked in a solution containing 15% PEG 4000, 100 mM NaCl, 0.5 mM EDTA and 20 mM HEPES buffer at pH 6.8 (1RVE storage buffer) before mounting. Two heavy atom derivatives were used for phase determination. Crystals of the lead derivative were soaked for 4 days in storage buffer with 1 mg/ml lead acetate, those of the mercury derivative for 6 days in storage buffer at pH 7.4 with 1 mg/ml of chlormerodrin (3-chloromercuri-2-methoxypropyl urea). The two sites of the lead derivative were already known from the analysis of a difference Patterson synthesis calculated with 6 Å resolution diffractometer data (Winkler *et al.*, 1987). The correct enantiomorph had been determined using the lead anomalous differences for phasing a difference Fourier synthesis of a gold derivative which was not used for higher resolution phasing. The mercury sites were obtained from a lead derivative phased difference Fourier map. The final heavy atom parameters used for phase determination were obtained by a Dickerson type refinement (Dickerson *et al.*, 1968) as implemented in the REFINE program from the CCP4 suite of programs (Daresbury Laboratories, UK). Anomalous differences were included in the phasing. After further improvement through five cycles of solvent flattening (Bricogne, 1976; Wang, 1985) the resulting 2.8 Å electron density map showed well-defined side chain densities and tracing of the polypeptide chain was straightforward with the exception of a few poorly

Table IV. Refinement statistics

Stage			1RVE		2RVE		3RVE		
0	Start	R factor (resolution range in Å)	0.375 (10−2.5)		0.426 (12−3)		0.394 (8−3)		
1	TNT[a]	R factor (resolution range)	0.240 (10−2.5)		0.274 (12−3)				
2	X-PLOR[a]	R factor (resolution range)	0.221 (8−2.5)		0.218 (6−3)		0.193 (8−3)		
3	TNT	R factor (resolution range)	0.185 (8−2.5)		0.192 (8−3)		0.176 (8−3)		
		No. of reflections	15 963		12 143		15 231		
		σ cut-off	–		$\sigma(F) > \lvert F \rvert$		–		
		Data completeness (%)	84		89		89		
		Geometry:							
		r.m.s. bonds (Å)	0.017		0.019		0.017		
		r.m.s. bond angles (°)	2.07		2.06		2.26		
		B factors							
		B_{av} (Å^2)	27.8		43.0		28.0		
		r.m.s. B (bonded atoms)	3.0		1.9		3.1		
		No. of H_2O	49		33		23		
Final model		Protein chains	A,B		A,B		A,B,C		
		DNA strands			C−D, E−F		D−E,F		
		Completeness of chain[b]	A	B	A	B	A	B	C
		No. of atoms with unit occupancy	1828	1815	1800	1808	1785	1782	1783
		Disordered chain segments[c]	13−17	13−19	13−17	13−17	13−17	13−18	13−18
			142−148	142−148	142−148	142−149	141−149	141−148	141−149
			183−187	183−186	184−185	183−185			
			221−228	221−228	221−228	222−228	222−228	222−228	222−228
					241−245	242−245	242−245	242−245	242−245
		Disordered single residues[c]	–	–	–	–	83,99,100	–	–
		Disordered side chains[c]	–	–	–	–	57,98	67,70,85, 98,99,101, 102	57,85 98,99,100

[a] Refinement carried out with the software packages TNT (Tronrud *et al.*, 1987) and X-PLOR (Brünger *et al.*, 1990) respectively.

[b] One complete *Eco*RV chain (residues 2−245) contains 2023 non-hydrogen atoms.

[c] Disordered parts of the protein chains are present in the models in a tentative conformation (except for the missing C-termini in 2RVE and 3RVE), but have been assigned zero occupancies.

defined segments (see Table IV). An initial, 82% complete model was built using FRODO (Jones, 1985).

Non-cognate complex with CGAGCTCG (2RVE). One enzyme dimer complexed with two octamer duplexes was expected to be present in the asymmetric unit of the monoclinic unit cell (Winkler *et al.*, 1991). Examination of the cross rotation function peaks obtained with data in the resolution range 8.0−4.0 Å and using the monomeric or dimeric structure as a search model indicated a change in the quaternary structure as the monomer model yielded clearly better results. The two highest peaks were related by an approximate twofold rotation axis corresponding to the top peak of the self rotation function. Determination of the translation parameters from the results of translation function calculations, again using data in the resolution range 8.0−4.0 Å, was straightforward. A $2F_{obs} - F_{calc}$ electron density map (resolution range 12.0−3.0 Å) calculated with phases of the molecular replacement solution, showed strong but poorly continuous density in the region where DNA was expected to be located. The protein density was of variable quality with indications of larger displacements of some chain segments. Therefore, data were also collected for crystals derivatized with the same heavy atom compounds used to solve the 1RVE structure. Difference Fourier maps calculated with the molecular replacement phases readily revealed two strong sites for the lead and the mercury derivative at the expected positions. The poor heavy atom refinement and phasing statistics (Table III) indicated substantial non-isomorphism and the corresponding MIR density map looked rather uninterpretable. However, combination of the MR and MIR phases followed by solvent flattening produced a map which showed well-defined density for both DNA duplexes. They were fitted into the density with sugar conformations constrained to C2′-*endo*. Substantial corrections to the protein model had to be made in a few places, mainly in rigid body displacements of the chain segments of the dimerization subdomain accompanied by local adjustments at the junctions.

Cognate complex with GGGATATCCC (3RVE). A 1:1 stoichiometry (one enzyme dimer/decamer duplex) was expected from the biochemical analysis of the crystals (Winkler *et al.*, 1991). The unit cell volume suggested that there could be 1.5 dimeric complexes in the asymmetric unit (calculated V_m = 2.23 Å^3/dalton). Analysis of the cross rotation function, using data in the resolution range 8.0−4.5 Å and one protein subunit of the 2RVE structure as a model, suggested that in addition to two subunits related by a non-crystallographic twofold axis, there was a third protein subunit oriented such that its associated twofold axis was parallel to a crystallographic twofold axis along *y*. After optimization of the rotation parameters of these three solutions, translation function searches were carried out using data in the resolution range 8.0−4.0 Å. All 12 independent, intermolecular translation vectors (nine Harker and three cross vectors) ranked at the top in the respective 2- or 3-dimensional searches. After R factor minimization as implemented in MERLOT the MR protein model produced an R factor of 0.445 with data between 10.0 and 3.0 Å resolution. The three decamer DNA strands appeared well defined in a $3F_{obs} - 2F_{calc}$ electron density map. An initial DNA model was built using A- and B-form DNA fragments positioned such that their backbones fitted well into the backbone density of the decamer duplex. A-form DNA was needed to give a satisfactory fit in the central part of the decamer. The base pairs were then optimally positioned in the density and, where necessary, small corrections were made to the backbone structure. The DNA model produced in this way fitted the density very well but contained a number of poor local geometries which were only corrected during refinement. No corrections were indicated at this point for the protein model except that, for the first time, the density for residues 183−186 was clearly defined and a satisfactory model could be built for this loop.

Structure refinement

Refinement of the starting models proceeded similarly for all three structures. Initial rounds of restrained geometry least squares refinement using the TNT software (Tronrud *et al.*, 1987) with frequent manual rebuilding using FRODO (Jones, 1985) were carried out in the case of the 1RVE and 2RVE structures. Refinement was continued (or, for 3RVE, started), with simulated annealing following the slow cooling protocol of X-PLOR (Brünger *et al.*, 1990) with additional cycles of positional least squares refinement. Simulated annealing with heating to 4000 K was partly carried out to explore whether some of the poorly defined, tentatively modelled chain segments would become clearer. Although large structural changes took place in these regions the resulting models remained unsatisfactory and subsequent omit and difference density maps still showed rather poor and fragmented density. It was therefore concluded that these segments are severely disordered and all atoms belonging to them were assigned zero occupancies in the subsequent refinement cycles. The DNA of the cognate complex, initially built with serious local strain, was carefully examined after X-PLOR refinement. In a number of places the backbone conformations were highly unusual and indeed they could all be manually corrected to more standard conformations without deterioration of the fit to the density. The final rounds of refinement of all three structures were carried out using the TNT software again. Manual rebuilding during this stage mainly consisted in correcting protein side chain conformations and in further improving the DNA backbone conformation. Individual isotropic temperature factors were refined with tight constraints between those of bonded atoms. Given the modest resolution of the three structures only a few bound water molecules were added to the models and refined. They were selected from the highest peaks (>3 times r.m.s. density) of a final difference density map under the constraint of being within hydrogen bonding distance to two hydrogen bond acceptor or donor atoms of the protein or DNA respectively. Information of the completeness and quality of the refined structures is given in Table IV. Final difference density maps including low resolution terms from 15 Å were inspected in the regions of the chain segments or residues assigned zero occupancies as described above and listed in Table IV. Although improved continuous density was seen in some of the regions, convincing, single-conformation models could not be built. They all appear disordered although to a variable extent in the seven different subunits.

The refined coordinates have been deposited with the Brookhaven Protein Data Bank as 1RVE (apoenzyme), 2RVE (non-cognate DNA complex) and 4RVE (cognate DNA complex).

Acknowledgements

We would like to thank S.Halford (University of Bristol, UK) and A.Pingoud (Medizinische Hochschule Hannover, Germany) and their collaborators for the continuous exchange of ideas and results prior to publication. We are particularly grateful to S.Halford for providing some of the enzyme used in this work and for critical reading of the manuscript. We also thank C.Broger for his help in the calculation and representation of contact surface areas. This project was initiated at the European Molecular Biology Laboratory (Heidelberg, Germany) and a substantial amount of the work described was carried out while the authors were working at that laboratory.

References

Beese,L.S. and Steitz,T.A. (1991) *EMBO J.*, **10**, 25–33.

Bennett,S.P. and Halford,S.E. (1989) *Curr. Top. Cell Regul.*, **30**, 57–104.

Bernstein,F.C., Koetzle,T.F., Williams,G.J.B., Meyer,E.F.Jr, Brice,M.D., Rodgers,J.D., Kennard,O., Shimanouchi,T. and Tasumi,M. (1977) *J. Mol. Biol.*, **112**, 535–542.

Bougueleret,L., Schwarzstein,M., Tsugita,A. and Zabeau,M. (1984) *Nucleic Acids Res.*, **12**, 3659–3676.

Bricogne,G. (1976) *Acta Crystallogr.*, **A32**, 832–847.

Brünger,A.T., Krukowsi,A. and Erickson,J.W. (1990) *Acta Crystallogr.*, **A46**, 585–593.

Connolly,B.A., Eckstein,F. and Pingoud,A. (1984) *J. Biol. Chem.*, **259**, 10760–10763.

D'Arcy,A., Brown,R.S., Zabeau,M., vanResandt,R.W. and Winkler,F.K. (1985) *J. Biol. Chem.*, **252**, 1987–1990.

Delcourt,S.G. and Blake,R.D. (1991) *J. Biol. Chem.*, **266**, 15160–15169.

Dickerson,R.E., Weinzierl,J.E. and Palmer,R.A. (1968) *Acta Crystallogr.*, **B24**, 997–1001.

Fliess,A., Wolfes,H., Seela,F. and Pingoud,A. (1988) *Nucleic Acids Res.*, **16**, 11781–11793.

Freemont,P.S., Lane,A.N. and Sanderson,M.S. (1991) *Biochem. J.*, **278**, 1–23.

Glusker,J. (1991) *Adv. Protein Chem.*, **42**, 1–76.

Grasby,J.A. and Connolly,B.A. (1992) *Biochemistry*, **31**, 7855–7861.

Halford,S.E., Taylor,J.D., Vermote,C.L.M. and Vipond,I.B. (1993) In Eckstein,F. and Lilley,D.M.J. (eds), *Nucleic Acids and Molecular Biology*. Springer Verlag, Berlin, Vol. 7.

Harrison,S.C. (1991) *Nature*, **353**, 715–719.

Heitman,J. (1992) *BioEssays*, **14**, 445–454.

Howard,A.J., Gilliland,G.L., Finzel,B.C., Poulos,T.L., Ohlendorf,D.H. and Salemme,F.R. (1987) *J. Appl. Crystallogr.*, **20**, 383–387.

Jeltsch,A., Alves,J., Maass,G. and Pingoud,A. (1992) *FEBS Lett.*, **304**, 4–8.

Jenkins,J. *et al.* (1992) *Biochemistry*, **31**, 5449–5458.

Jones,T.A. (1985) *Methods Enzymol.*, **115**, 157–171.

Kabsch,W. (1988) *J. Appl. Crystallogr.*, **21**, 916–924.

Kabsch,W. and Sanders,C. (1983) *Biopolymers*, **22**, 2577–2637.

Kennard,O. and Hunter,W.N. (1989) *Q. Rev. Biophys.*, **22**, 327–379.

Kim,Y., Grable,J.C., Love,R., Greene,P.J. and Rosenberg,J. (1990) *Science*, **249**, 1307–1309.

Lavery,R. and Sklenar,H. (1989) *J. Biomol. Struct. Dyn.*, **7**, 655–667.

Leslie,A.G.W., Brick,P., and Wonacott,A.J. (1986) *CCP4 Newsletter*, **18**, 33–39.

Lesser,D.R., Kurpiewski,M.R. and Jen-Jacobson,L. (1990) *Science*, **250**, 776--778.

Mazzarelli,J., Scholtissek,S. and McLaughlin,L.W. (1989) *Biochemistry*, **28**, 4616–4622.

Newman,P.C., Williams,D.M., Cosstick,R., Seela,F. and Connolly,B.A. (1990) *Biochemistry*, **29**, 9902–9910.

Priestle,J. (1988) *J. Appl. Crystallogr.*, **21**, 572–576.

Record,M.T.,Jr, deHaseth,P.L. and Lohman,T.M. (1977) *Biochemistry*, **16**, 4791–4906.

Roberts,R.J. and Macelis,D. (1992) *Nucleic Acids Res.*, **20**, 2167–2180.

Rosenberg,J.M. (1991) *Curr. Opin. Struct. Biol.*, **1**, 104–113.

Saenger,W. (1984) In Cantor,C.R. (ed.), *Principles of Nucleic Acid Structure*, Springer-Verlag, Berlin, p. 185.

Saenger,W. (1991) *Curr. Opin. Struct. Biol.*, **1**, 130–138.

Schultz,S.C., Shields,G.S. and Steitz,T.A. (1991) *Science*, **253**, 1001–1007.

Selent,U., Rüter,T., Köhler,E., Liedtke,M., Thielking,V., Alves,J., Oelgeschläger,T., Wolfes,H., Peters,F. and Pingoud,A. (1992) *Biochemistry*, **31**, 4808–4815.

Steitz,T.A. (1990) *Q. Rev. Biophys.*, **23**, 205–280.

Taylor,J.D. and Halford,S.E. (1989) *Biochemistry*, **28**, 6198–6207.

Taylor,J.D.,Badcoe,I.G., Clarke,A.R. and Halford,S.E. (1991) *Biochemistry*, **30**, 8743–8753.

Thielking,V., Alves,J., Fliess,A., Maass,G. and Pingoud,A. (1990) *Biochemistry*, **29**, 4682–4691.

Thielking,V., Selent,U., Köhler,E., Wolfes,H., Pieper,U., Geiger,R., Urbanke,C., Winkler,F.K. and Pingoud,A. (1991) *Biochemistry*, **30**, 6416–6422.

Thielking,V., Selent,U., Köhler,E., Landgraf,A., Wolfes,H., Alves,J. and Pingoud,A. (1992) *Biochemistry*, **31**, 3727–3732.

Travers,A.A. and Klug,A. (1990) In Cozzarelli,N.R. and Wang,J.C. (eds), *DNA Topology and its Biological Effects*. Cold Spring Harbor Laboratory Press, Cold Spring Harbor, NY, pp. 57–106.

Tronrud,D.E., Ten Eyck,L.F. and Matthews,B.W. (1987) *Acta Crystallogr.*, **A43**, 489–501.

van Tilbeurgh,H., Jenkins,J., Chiadmi,M., Janin,J., Wodak,S.J., Mrabet,N.T. and Lambeir,A. (1992) *Biochemistry*, **31**, 5467–5471.'

Vermote,C.L.M. (1991) Ph.D. Thesis, University of Bristol.

Vermote,C.L.M. and Halford,S.E. (1992) *Biochemistry*, **31**, 6082–6089.

Vermote,C.L.M., Vipond,I.B. and Halford,S.E. (1992) *Biochemistry*, **31**, 6089–6097.

Wang,B.C. (1985) *Methods Enzymol.*, **115**, 90–112.

Wilmot,C.M. and Thornton,J.M. (1988) *J. Mol. Biol.*, **203**, 221–232.

Winkler,F.K. (1992) *Curr. Opin. Struct. Biol.*, **2**, 93–99.

Winkler,F.K., Brown,R.S., Leonard,K., and Berriman,J. (1987) In Moras,D., Drenth,J., Strandberg,B., Suck,D. and Wilson,K. (eds), *Crystallography in Molecular Biology*. Plenum Press, London, pp. 345–352.

Winkler,F.K., D'Arcy,A., Blöcker,H., Frank,R. and van Boom,J.H. (1991) *J. Mol. Biol.*, **217**, 235–238.

Zebala,J.A, Choi,J. and Barany,F. (1992) *J. Biol. Chem.*, **267**, 8097–8105.

Received on December 29, 1992