3438–3445 *Nucleic Acids Research, 1999, Vol. 27, No. 17*

Structural analysis of a mutational hot-spot in the *Eco*RV restriction endonuclease: a catalytic role for a main chain carbonyl group

Mark P. Thomas, R. Leo Brady, Stephen E. Halford*, Richard B. Sessions and Geoffrey S. Baldwin

Department of Biochemistry, School of Medical Sciences, University of Bristol, University Walk, Bristol BS8 1TD, UK

Received May 26, 1999; Revised and Accepted June 25, 1999

PDB accession nos R1B94SF–R1B97SF

ABSTRACT

Following random mutagenesis of the *Eco*RV endonuclease, a high proportion of the null mutants carry substitutions at Gln69. Such mutants display reduced rates for the DNA cleavage step in the reaction pathway, yet the crystal structures of wild-type *Eco*RV fail to explain why Gln69 is crucial for activity. In this study, crystal structures were determined for two mutants of *Eco*RV, with Leu or Glu at residue 69, bound to specific DNA. The structures of the mutants are similar to the native protein and no function can be ascribed to the side chain of the amino acid at this locus. Instead, the structures of the mutant proteins suggest that the catalytic defect is due to the positioning of the main chain carbonyl group. In the enzyme–substrate complex for *Eco*RV, the main chain carbonyl of Gln69 makes no interactions with catalytic functions but, in the enzyme–product complex, it coordinates a metal ion bound to the newly liberated 5'-phosphate. This re-positioning may be hindered in the mutant proteins. Molecular dynamics calculations indicate that the metal on the phosphoryl oxygen interacts with the carbonyl group upon forming the pentavalent intermediate during phosphodiester hydrolysis. A main chain carbonyl may thus play a role in catalysis by *Eco*RV.

INTRODUCTION

The *Eco*RV endonuclease is a type II restriction enzyme that cleaves DNA specifically at its recognition sequence, GAT↓ATC (where ↓ marks the point of cleavage), in a reaction that requires only Mg^{2+} as a cofactor (1,2). It exists as a homodimer (2,3) and binds to the recognition sequence in a symmetrical fashion, so that its two active sites are juxtaposed against the two scissile bonds in the DNA, one in each strand (4). During the reaction at the recognition site, the DNA cleavage step is faster than the subsequent (rate-limiting) dissociation of the cleaved product (5). Consequently, *Eco*RV usually cleaves both DNA strands within one DNA-binding event (6). Under optimal reaction conditions, it cleaves its recognition sequence over 10^6 times faster than any other sequence (7,8). Yet under the same conditions lacking Mg^{2+}, *Eco*RV binds all sequences with equal affinity and shows no preference for the recognition site (9–12). Ca^{2+} ions fail to support DNA cleavage by *Eco*RV but can mimic Mg^{2+} in promoting sequence-specific binding (13,14). The discrimination against DNA cleavages at sites other than the recognition site is thus achieved in the ternary *Eco*RV–DNA–metal ion complex rather than the binary *Eco*RV–DNA complex (13–15).

In the crystal structures of *Eco*RV bound to its recognition site in the absence of Mg^{2+} (4,16,17), the DNA lies in a cleft between the two protein subunits and is contacted primarily by two peptide loops from each subunit, though several other segments of the protein contact the DNA phosphates (18,19). The R (for recognition) loop, residues 182–188, lies in the major grove of the DNA and makes several hydrogen bonds to bases in the recognition sequence. The Q loop, so-called because it contains two glutamines, comprises residues 67–72. It interacts extensively with the sugar–phosphate backbone in the minor grove. When bound to *Eco*RV, the GATATC sequence is distorted from B-form DNA, with a marked bend towards the major grove (13,20), which places the scissile phosphodiester bond in the active site of the enzyme. In the crystal structure of *Eco*RV bound to non-specific DNA, the DNA is again located in the cleft between the subunits but it retains a B-form structure and fails to enter the active site (4). The R loop makes no contacts to the non-specific DNA but the Q loop still makes intimate contacts with the sugar–phosphate backbone. In the structure of the free protein without DNA, both R and Q loops are largely disordered (4). The Q loop is thus a conformationally mobile segment of polypeptide that can adjust to a variety of DNA structures.

The Q loop has been probed by random mutagenesis followed by selection for null mutants on the basis of viability in cells lacking the *Eco*RV methyltransferase (21). While many of the residues in the Q loop could be mutated to yield the null phenotype, mutations at Gln69 occurred at much higher frequency than at other positions. Mutants in which Gln69 had been replaced by Glu or Leu (Q69E and Q69L, respectively) were analysed further (22). In the absence of divalent metals, Q69L binds DNA with the same affinity and lack of specificity as wild-type *Eco*RV while Q69E binds less strongly. With Ca^{2+} present, both mutants bind DNA specifically

*To whom correspondence should be addressed. Tel: +44 117 928 7429; Fax: +44 117 928 8274; Email: s.halford@bris.ac.uk

in the same manner as native *Eco*RV. For Q69E, the rate of the DNA cleavage step in the reaction pathway is ~10^4 times lower than wild-type *Eco*RV while the cleavage step with Q69L is 100-fold slower than wild-type. Nevertheless, the reduced rate of phosphodiester hydrolysis by Q69L, relative to wild-type, results in the steady-state turnover of the mutant being rate-limited by this step rather than by product dissociation. Thus, Q69L has a steady-state k_{cat} just 5-fold lower than wild-type, but its initial product is DNA cut in one strand, in contrast to the double strand breaks by wild-type *Eco*RV. In addition, Q69L requires higher concentrations of Mg^{2+} than wild-type and shows a sigmoidal dependence upon the Mg^{2+} concentration, indicating two metal ions per strand scission. Transient kinetics on Q69L also showed that its reaction involves a slow conformational change preceding DNA cleavage that has no equivalent with wild-type *Eco*RV (22).

However, the crystal structures of *Eco*RV do not explain why mutations at Gln69 should affect the DNA cleavage step in the reaction pathway. In the complexes of *Eco*RV bound to either specific or non-specific DNA (4), the side chain of Gln69 lies parallel to the sugar–phosphate backbone of the DNA and its methylene groups make van der Waals contacts to both the DNA and to other regions of the protein, but the carbonyl (CO) and amide groups at the end of the side chain do not make hydrogen bonds to the DNA. In addition, the main chain CO and amino groups of Gln69 do not seem to have any specific function in the enzyme–substrate complex. In contrast, the structure of *Eco*RV bound to the product from its DNA cleavage reaction shows the peptidyl CO of Gln69 coordinated to one of the two Mg^{2+} ions liganded to the 5'-phosphate of the cleaved DNA (16). The interaction between the peptidyl CO and the metal ion is observed only in the crystal structure of the enzyme–product complex and not in the structures obtained after adding divalent metal ions to crystals of *Eco*RV bound to its recognition site. Both the metal ions and the scissile phosphate are positioned differently in the *Eco*RV–substrate–metal complex compared to the *Eco*RV–product–metal complex (see below; Fig. 4).

The objective of this study was to identify a role for Gln69 in catalysis by *Eco*RV. The Q69L and Q69E mutants were crystallised as complexes with specific DNA and their structures determined by X-ray crystallography. Since the diffraction data for the mutants were collected at 100K, valid comparisons between the mutant and wild-type structures also required the acquisition of structures for the wild-type enzyme bound to specific DNA at 100K. Further insights into possible roles for Gln69 were obtained by molecular dynamics (MD) modelling of the pentavalent intermediate during phosphodiester hydrolysis by *Eco*RV.

MATERIALS AND METHODS

Materials

Wild-type *Eco*RV and the Q69L and Q69E mutants (21) were purified as described previously (3,22). The oligodeoxynucleotide AAAGATATCTT was synthesised on a Millipore Expedite system with reagents from Cruachem (Glasgow, UK), purified by HPLC and annealed to form the duplex (23).

Methods

All crystallisations were by the sitting drop method. The protein (2 µl, 15 mg/ml, ~260 µM dimer) in 10 mM phosphate, pH 7.5, 250 mM NaCl, 1 mM EDTA, 0.1 mM dithiothreitol and 0.1% (w/v) sodium azide was mixed with the duplex form of AAAGATATCTT (2 µl, 3 mg/ml, ~460 µM) in 10 mM cacodylate, pH 6.0, and, in some instances, with $CaCl_2$ at a final concentration of 5 mM. Solutions of PEG 4000 (0.7 µl) were added to yield final concentrations of PEG of 0.8–1.4% (w/v). Crystals of wild-type (±Ca^{2+}) and mutant proteins grew in 5–6 days at 18°C. All of the crystals grew isomorphously in space group *P*1 with one enzyme–substrate complex in the asymmetric unit. The cryoprotectant solution used during data collection was the same as the crystallisation solution except that the concentration of PEG 4000 was 8.8% (w/v) and that glycerol was present at 23.4% (v/v). Diffraction data were collected at 100K at station PX7.2 of the CLRC synchrotron radiation source (Daresbury, Cheshire, UK) on a MAR30 image plate or in Bristol using Cu*K*α radiation from a rotating anode and a 20 cm Mac-Science image plate and processed through Denzo/Scalepack (24). Cell constants and data statistics are given in Table 1. Data for Q69E were collected from two crystals: data for the other structures were from single crystals.

The structures were solved by molecular replacement using AMoRe (25) with the PDB (Protein Data Bank, Brookhaven; 26) entry 1RVA (16) as the search molecule. Structures were refined with the CCP4 suite of programs (27), chiefly with Refmac (28). Electron density interpretation and model building used the program O (29). Waters were added through ARP (30) in three cycles with σ cut-offs set at 3.5, 3.25 and 3.0, with cycles of refinement through Refmac between each cycle through ARP. Further refinement and rebuilding led to the removal of several water molecules due to high B-factors, lack of electron density and/or poor hydrogen bonding contacts. In the case of crystals obtained in the presence of Ca^{2+}, the areas of density picked by ARP as being waters were changed to Ca^{2+} ions if all of the following criteria were met: larger than expected for a water molecule; visible at greater than 5σ in an $F_o - F_c$ map; lower than expected B-factors for waters in those positions; hexa-coordination to other ligands. Distances between the Ca^{2+} ions and coordinating waters were restrained to 2.56 Å and to protein side chain oxygens and DNA phosphate oxygens to 2.74 Å (31). This was followed by further refinement through Refmac. Refinement statistics are given in Table 1.

Superpositions were calculated using lsqkab (32) in the CCP4 suite (27). MD simulations of a sphere of radius 25 Å, centred on the active site of the B subunit, were carried out in Discover 2.97 under the Consistent Valence Force Field, as described previously (15).

RESULTS

Crystal structures

The following structures of *Eco*RV proteins bound to the *Eco*RV recognition sequence were determined from X-ray diffraction data collected at 100K: wild-type *Eco*RV (noted as ED); wild-type *Eco*RV in the presence of Ca^{2+} ions (EDC); a mutant *Eco*RV with Leu in place of Gln69 (Q69L); a mutant *Eco*RV with Glu in place of Gln69 (Q69E). In all four cases,

Table 1. Data and refinement statistics

	*Eco*RV with DNA	*Eco*RV with DNA + Ca^{2+}	Q69L with DNA	Q69E with DNA
Space group	*P*1	*P*1	*P*1	*P*1
Cell constants				
a, b, c (Å)	48.5, 48.6, 63.9	48.6, 49.4, 63.9	48.7, 49.2, 64.0	48.4, 48.8, 63.6
α, β, γ (°)	97.0, 108.9, 107.0	96.9, 108.7, 107.5	96.7, 108.9, 107.5	96.6, 108.9, 107.4
Observed reflections	73 981	158 169	113 353	148 523
Unique reflections	29 850	37 718	38 240	21 703
R_{merge}	2.3%	5.0%	5.2%	9.9%
Completeness (overall)	95.1% (30.0–2.05 Å)	91.9% (20.0–1.90 Å)	92.7% (15.0–1.90 Å)	96.5% (12.5–2.30 Å)
Completeness (outer shell)	86.6% (2.21–2.05 Å)	82.7% (2.05–1.90 Å)	87.5% (1.99–1.90 Å)	94.8% (2.38–2.30 Å)
R_{cryst}	20.6%	20.6%	21.9%	18.9%
R_{free}	28.2%	27.0%	27.9%	28.2%
r.m.s.d. from ideal bond lengths (Å)	0.012	0.010	0.013	0.011
r.m.s.d. from ideal bond angles (°)	1.6	1.7	1.3	2.3
Number of water molecules	357	350	358	257
Other atoms	–	2 Ca^{2+}	–	–

the duplex oligodeoxynucleotide with the recognition sequence was the undecamer used by Kostrewa and Winkler (16), AAAGATATCTT. Though Q69E binds DNA weakly in the absence of Ca^{2+} (22), it still crystallised as a DNA–protein complex, presumably due to the high concentrations of both DNA and protein in the crystallisation drop. However, no crystals were obtained for the mutant proteins bound to DNA in the presence of either Ca^{2+} or Mg^{2+} ions.

The four structures reported in this paper have been deposited at the PDB, with the following entry codes (and structure factors): for ED, 1B95 (R1B95SF); for EDC, 1B94 (R1B94SF); for Q69L, 1B97 (R1B97SF); for Q69E, 1B96 (R1B96SF). A brief account of the overall architecture of the new structures and their similarities to previous structures for *Eco*RV from room temperature data (4,15,16) is given here. The two protein chains of the dimeric enzyme are labelled A and B and are numbered 2–245 from N- to C-termini (the N-terminal methionine in the gene sequence is not present in the mature protein; 33). The two DNA chains are labelled C and D and are numbered 1–11 from 5'- to 3'-termini, with the scissile bond between nucleotides 6 and 7 on both strands. The catalytic residues E45, D74 and D90 of the A chain act on the C strand and those of the B chain on the D strand. The two subunits are related by a non-crystallographic dyad and have numerous differences in conformational details.

The four structures have good geometries (Table 1) and are very similar to each other, with the protein Cα atoms superimposing closely (Fig. 1). The side chains and the DNA, except around the relatively disordered regions (residues 15–19, 97–101, 143–153 and nucleotide 1), also overlay well. The structures of ED and EDC are at resolutions equal to or greater than the equivalent structures reported before, with greater completeness in the data, lower values of R_{merge} and comparable values of R_{cryst} and (where available) R_{free}. There are thus more water molecules in these structures, which results in more complete hydrogen-bonding networks on the surface of the protein and the DNA. Nonetheless, the current structure of wild-type *Eco*RV complexed with specific DNA in the absence of Ca^{2+} (ED) is essentially identical to that from Kostrewa and Winkler (1RVA; 16). An overlay of all 488 Cα atoms in ED and 1RVA

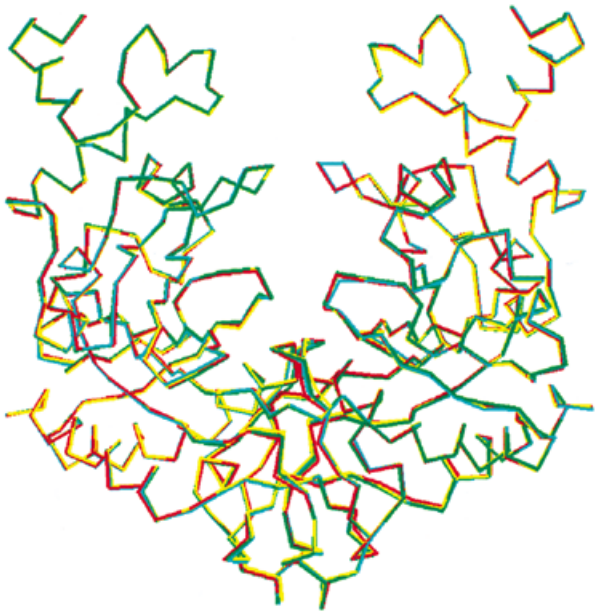

Figure 1. An overlay of the Cα atoms in both peptide chains from the following structures of *Eco*RV proteins bound to specific DNA: wild-type *Eco*RV (ED), in red; wild-type *Eco*RV + Ca^{2+} (EDC), in green; the Q69L mutant, in blue; the Q69E mutant, in yellow. For clarity, the DNA has been omitted in all four cases.

yields a root mean square displacement (r.m.s.d.) of 0.29 Å. The structures differ only in the rotation of some of the amino acid side chains, the position of some atoms in the Q loop and in the position and number of water molecules. This comparison shows that the structure is not affected by the temperature of data collection.

EDC has two Ca^{2+} ions: one (visible at 7σ in an $F_o - F_c$ omit map) in the B subunit, octahedrally coordinated to Asp74, Asp90, the scissile phosphate and three water molecules, at a locus called the 90/74 site; the other (refined at 50% occupancy and visible at 5.5σ in an $F_o - F_c$ omit map) at the 90/74 site of the A subunit. In the previous structure determinations of *Eco*RV with DNA and Ca^{2+}, the Ca^{2+} ions were either soaked into pre-existing crystals (16) or, as in this work, present during the crystallisation (17). The previous studies used higher concentrations of $CaCl_2$ (30–40 mM) than that used here (5 mM), the latter being similar to the concentration used in DNA-binding studies by gel-retardation (13). This accounts for why EDC, unlike the earlier structures, shows incomplete occupancy of the 90/74 site in one subunit. In one model for the *Eco*RV–DNA–Ca^{2+} complex (17), several amino acids were excluded, due to high B-factors and/or lack of electron density. Nonetheless, the 466 Cα atoms in this model superimpose on the equivalent atoms in EDC with a r.m.s.d. of 0.33 Å. There is also good agreement in the positions of most side chains.

A superposition of the Cα atoms of ED and EDC (Fig. 1) yields a r.m.s.d. of 0.25 Å. The binding of Ca^{2+} to the complex of *Eco*RV with its recognition site thus has virtually no effect on the conformation of the protein or the DNA (not shown). The lack of divalent metal ions in the co-crystals with the mutant proteins is thus unlikely to have a significant effect on their structures.

The overall structures of the two mutant *Eco*RV proteins bound to specific DNA are also very similar to wild-type (Fig. 1). The Cα atoms in ED overlay those in Q69E and Q69L with r.m.s.d. values of 0.25 and 0.24 Å, respectively. The similarity includes the DNA structure and the positions of most of the backbone and side chain atoms of the Q loop. In Q69E, the configuration of the Glu side chain at residue 69 follows that for the Gln in wild-type *Eco*RV. In both Q69E and ED, the terminal atoms of the side chain of residue 69 are within hydrogen bonding distance of two residues in the opposite subunit of the dimer: Thr37, both backbone nitrogen and side chain oxygen; Arg140, one or more of the side chain nitrogens. In Q69L, Leu cannot make these hydrogen bonds, but this has no detectable effect on the orientation of the side chain at position 69 or on either side chain or backbone positions of Thr37 and Arg140. However, either Glu or Leu at position 69 result in slightly closer van der Waals contacts between the methylene side chain and the deoxyribose backbone of the DNA than is the case with the wild-type Gln (not shown).

Main chain carbonyl of Gln69

Since the only discernible function for Gln69 is the interaction of its peptidyl CO with one of the two Mg^{2+} ions bound to the 5'-phosphate in the enzyme–product complex (16), the enzyme–substrate complexes for both wild-type *Eco*RV and the two mutants were examined to see if the positioning of the main chain CO at residue 69 might account for the effect of the mutations. In one of the two subunits of the enzyme–substrate

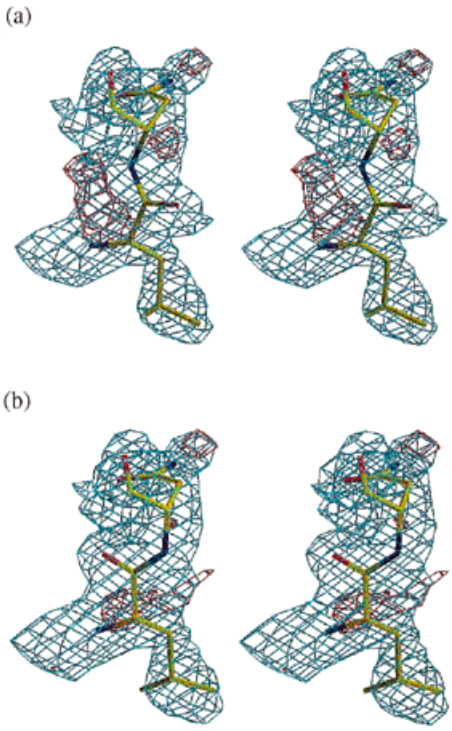

Figure 2. Stereo diagrams of the electron density maps for Leu69 and Asn70 in the B subunit of the Q69L mutant of *Eco*RV, showing the two alternative orientations for the main chain CO at residue 69. In each case, the $2F_o - F_c$ map (shown in blue) is calculated with coefficients from the accompanying coordinates. However, the corresponding $F_o - F_c$ map (2σ contour shown in red) demonstrates that the alternative conformation for the main chain CO is also present in the crystal. (**a**) The model with the main chain CO in the OUT orientation and (**b**) the model with the main chain CO in the IN orientation. The active site region is located on the left-hand side of these pictures.

complexes for wild-type *Eco*RV, with and without Ca^{2+} (ED and EDC), and for Q69L, the electron density maps (Fig. 2) are consistent with equal proportions of two alternative configurations at residue 69: an 'OUT' orientation where the main chain CO points away from the active site (Fig. 3a); an 'IN' orientation where the CO is flipped through ~150° so that it points into the active site (Fig. 3b). In each of these three structures, the electron density for the opposite subunit of the dimer is consistent only with the OUT orientation for the main chain CO at residue 69, as in Figure 3a (Table 2). However, the enzyme–substrate complex for Q69E has the OUT orientation for the main chain CO in both subunits (Table 2). In contrast, the enzyme–product complex for wild-type *Eco*RV displays the IN orientation in both subunits (Fig. 3c). Indeed, the main chain CO at residue 69 must have the IN orientation if it is to interact with a Mg^{2+} bound to the 5'-phosphate.

The overall configuration of the peptide backbone in the Q loop is largely unaffected by the flip in the orientation of the main chain CO at residue 69 (Fig. 2). Nevertheless, the two alternative configurations for the peptidyl CO at 69 are generally accompanied by an alternation in the position of the side chain of the adjacent residue, Asn70 (Table 2). In most cases where

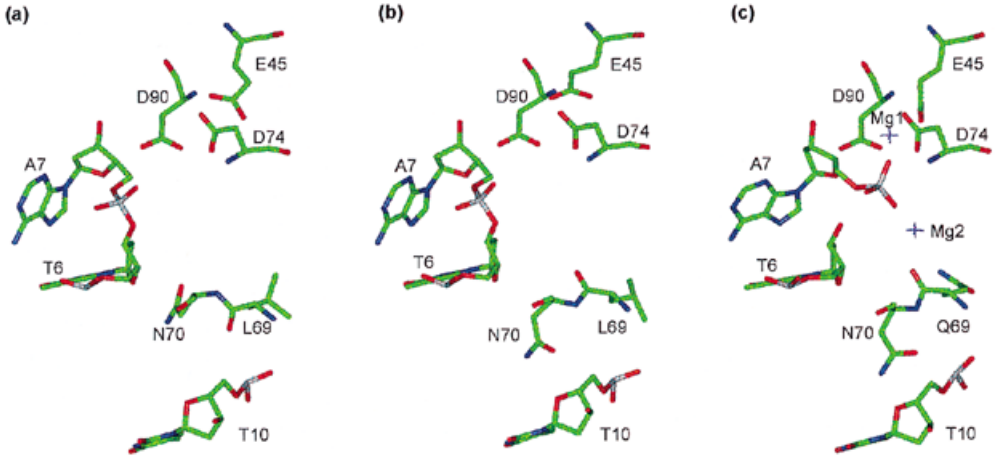

Figure 3. Alternative conformations at the active site of *Eco*RV. (**a** and **b**) The two possible orientations for the main chain carbonyl of L69 and the side chain of N70, in the B subunit of Q69L bound to the duplex form of AAAGATATCTT. In (a), the main chain CO of L69 points away from the active site and the side chain of N70 points towards the recognition sequence (cf. Fig. 2a). In (b), the main chain CO of L69 points into the active site and the side chain of N70 points away from the recognition sequence (cf. Fig. 2b). (**c**) The enzyme–product complex for wild-type *Eco*RV (1RVC; 16), with the main chain CO of Q69 pointing into the active site and the side chain of N70 pointing away from the recognition sequence. All three panels show, in addition to residues 69 and 70: the active site residues, D90, E45 and D74; the TpA step (T6 and A7) cleaved by *Eco*RV; the thymidine immediately following the recognition sequence (T10). Also shown in (c) are two Mg^{2+} ions (purple crosses).

Table 2. Orientation of the main chain carbonyl at residue 69 and the side chain of Asn70

	Main chain CO of Residue 69		Side chain of Asn70	
	Subunit A	Subunit B	Subunit A	Subunit B
ED	IN/OUT	OUT	IN	OUT
EDC	OUT	IN/OUT	IN	IN
Q69L	OUT	IN/OUT	IN	OUT
Q69E	OUT	OUT	IN	IN
1RVC	IN	IN	OUT	OUT

ED, Q69L and Q69E refer to wild-type and mutant *Eco*RV proteins bound to the *Eco*RV recognition sequence. EDC is the wild-type protein–DNA complex with Ca^{2+}. 1RVC is the PDB entry for wild-type *Eco*RV bound to the products of its DNA cleavage reaction (16). The terms IN and OUT mark, respectively, whether the relevant moieties point into or out of the active site region: IN/OUT denotes cases where the electron density indicates that both orientations are populated.

the main chain CO at 69 points away from the active site, such as in both subunits of Q69E, the side chain of Asn70 enters the DNA minor groove (Fig. 3a) and makes a sequence-specific hydrogen bond with the O2 carbonyl group of the cytosine at the 3'-end of the recognition sequence (4; not shown). Conversely, in most cases where the main chain CO at 69 points into the active site, such as in both subunits of the *Eco*RV–product complex, a rotation around the Cβ–Cγ bond of Asn70 leaves its side chain inappropriately positioned for any base-specific contact to the recognition sequence (Fig. 3b and c). Instead, the side chain of Asn70 then makes van der Waals interactions with the deoxyribose of the first nucleotide after the recognition sequence.

Molecular dynamics

Phosphodiester hydrolysis by *Eco*RV proceeds with inversion of configuration at the target phosphorus and thus involves the direct attack of a water molecule in line with the 3'-leaving group, presumably via a pentavalent intermediate (34). To investigate whether the position of the main chain CO at residue 69 affects catalysis by *Eco*RV, molecular mechanics were used to generate a model of the pentavalent intermediate (Fig. 4).

In the crystal structures of *Eco*RV bound to its DNA substrate in the presence of Ca^{2+}, one Ca^{2+} ion is found in each subunit, at the 90/74 site between Asp90 and Asp74 and the *pro*S oxygen of the scissile phosphate (16,17; this study, EDC). In contrast, both the *Bam*HI and *Bgl*I endonucleases incorporate two Ca^{2+} ions per active site (35,36). However, when both Ca^{2+} and Mn^{2+}, or just Co^{2+} ions, are added to *Eco*RV–DNA crystals (16), metal ions bind to both the 90/74 site and to an additional site between Asp74 and Glu45, the 74/45 site (Fig. 4a). Both wild-type *Eco*RV and Q69L need at least two metal ions per catalytic event (15,22,23,37), but the metal at the 74/54 site is ~5 Å away from the scissile phosphate. Yet no DNA cleavage occurs in the *Eco*RV crystals with two metal ions per active site (16), even though Co^{2+} gives almost the same rates as Mg^{2+} (15), while the combination of Ca^{2+} and Mn^{2+} gives higher activity than Mn^{2+} alone (37). The crystal structure of the *Eco*RV–substrate complex (Fig. 4a) is of an inactive conformation.

A model for the active conformation was derived previously by applying MD to the crystal structure of the *Eco*RV–DNA–Co^{2+} complex (15; the structure used as the start point for MD

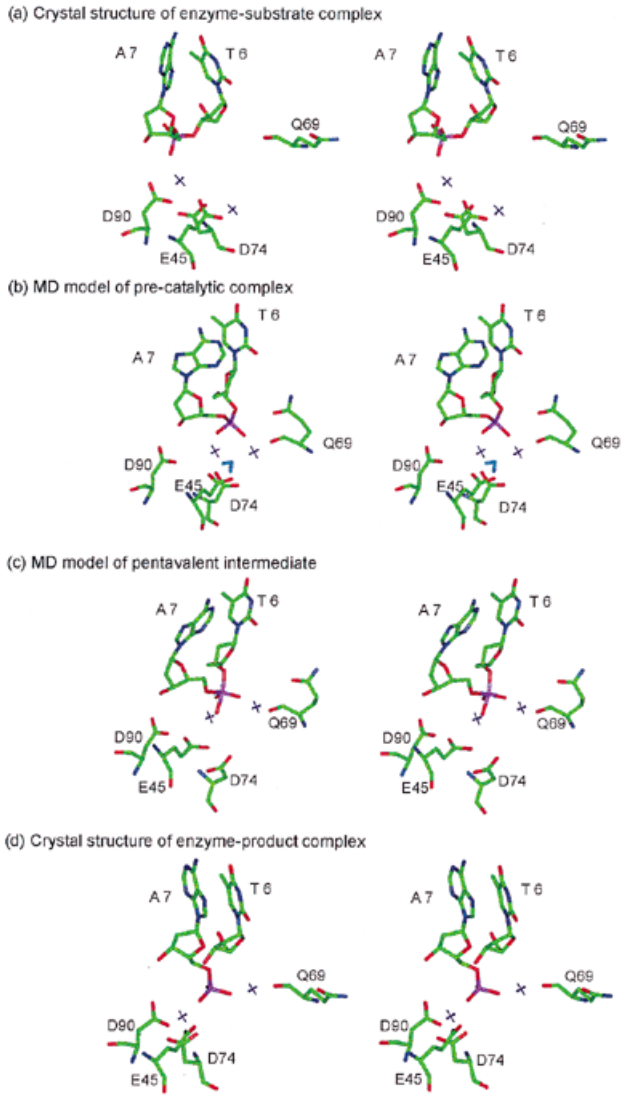

Figure 4. Structures for catalysis by *Eco*RV. (**a**) The active site in the B subunit from the crystal structure of *Eco*RV bound to its recognition sequence and to two Co^{2+} ions (D.Kostrewa and F.Winkler, personal communication). (**b**) The MD model for the active conformation of *Eco*RV, derived from the crystal structure in (a) (15; for MD, the Co^{2+} ions were converted to Mg^{2+}). (**c**) The MD model for the pentavalent intermediate during phosphodiester hydrolysis, derived from the scheme in (b) utilising the water noted in (b) between the right-hand Mg^{2+} and E45. (**d**) The active site in the B subunit from the crystal structure of *Eco*RV bound to the products of its DNA cleavage reaction and to two Mg^{2+} ions (1RVC; 16). All four stereo diagrams show: the TpA step (T6 and A7) cleaved by *Eco*RV; the active site residues, D90, D74 and E45; Q69; two metal ions (purple crosses).

already had the IN orientation for the main chain CO at 69). During the MD trajectory, a spontaneous transition occurred to a new state: the pucker of the deoxyribose 3' to the scissile bond flipped from C3'-*endo* to C2'-*endo* and the target phosphate moved 3.4 Å, to a position close to its location in the crystal structure of the enzyme–product complex (Fig. 4b). [The phosphate in the enzyme–product complex (Fig. 4d) is 4.5 Å away from its position in the enzyme–substrate complex (16).] After the transition, the metals at both the 90/74 and the 74/45 sites contact oxygens on the scissile phosphate: the *pro*R oxygen replaces the *pro*S moiety on the metal at the 90/74 site while the *pro*S oxygen moves to the metal at the 74/45 site, but the latter metal is pulled partly out of its site so that Glu45 becomes an outer shell ligand. However, the inner shell water between Glu45 and the 74/45 metal is placed to attack the phosphorus (15).

To convert the scheme for the active conformation (Fig. 4b) into a model for the pentavalent intermediate (Fig. 4c), the water between the 74/45 metal and Glu45 was reconfigured *in silico* as an OH and a H and these were attached to the phosphorus at the scissile bond and the carboxylate of Glu45, respectively. The resultant structure was subjected to energy minimisation followed by MD for 200 ps. In the minimisation, the aberrant bond lengths created by the reconfiguration reverted to standard lengths. In the subsequent MD, the coordination geometry of the phosphorus changed from tetrahedral to trigonal bipyramidal in the first 15 ps of the trajectory and essentially no further changes in the structure occurred during the remainder of the trajectory. During these adjustments, the phosphorus at the scissile bond remained in the same position but the oxygens surrounding the phosphorus took up new positions as the coordination geometry of the phosphorus changed. This in turn re-positioned the metal ions attached to the phosphoryl oxygens. In particular, the metal attached to the *pro*S oxygen, that had been near the 74/45 site, moved from an apical to an equatorial position that was ~6 Å away from Glu45, but which was now only 2 Å away from the main chain CO of Gln69, the correct distance for an inner shell ligand. The model for the pentavalent intermediate (Fig. 4c) thus has marked similarities to the crystal structure of the enzyme–product complex (Fig. 4d).

DISCUSSION

Random mutagenesis of *Eco*RV had identified Gln69 as a hotspot for the generation of null mutants, which indicates that Gln69 plays a crucial role in DNA recognition and/or catalysis (21). The biochemical analysis of *Eco*RV proteins mutated at residue 69 had shown that the defect could be isolated primarily to the DNA cleavage step in the reaction pathway (22). The replacement of Gln69 by Glu has a graver effect on catalysis than the substitution with Leu (22). Hence, it might have been anticipated that the overall structure of the Q69E protein would display a greater perturbation from the wild-type structure than Q69L. However, the high resolution crystal structures of the DNA–protein complexes for Q69E and Q69L show that this is not the case. The overall structures of the complexes with the mutant proteins are very similar to each other and to wild-type *Eco*RV (Fig. 1), with respect to both the protein and the DNA. In addition, while the amide and CO groups at the end of the side chain of Gln69 are within hydrogen bonding distance of Thr37 and Arg140 in the opposite subunit of the dimer, Leu at this position cannot duplicate these interactions, yet the Gln to Leu substitution has no effect on the orientation of the side chain at residue 69 or on the positioning of Thr37 and Arg140. Moreover, the severe consequences of mutations at Thr37 (18) seem not to be due to interactions across the subunit interface (38). The effect of mutations at Gln69 cannot therefore be assigned to the loss of the functional

groups at the end of the side chain or to any gross perturbation to either protein or DNA conformation.

Discrete perturbations were, however, observed among the structures analysed here in the orientation of the main chain CO at residue 69 (Fig. 3 and Table 2). As noted previously (16), the peptide backbone of the *Eco*RV protein has two alternate configurations at residue 69: either the OUT orientation, in which main chain CO points away from the active site (Fig. 3a), or the IN orientation, in which the main chain CO points into the active site (Fig. 3b). The two orientations for the main chain CO at residue 69 are usually accompanied by alternate conformations for the side chain of the adjacent residue, Asn70 (Table 2). In most cases where the main chain CO at 69 has the OUT orientation, the side chain of Asn70 enters the minor groove of the DNA and contacts the cytosine base at the 3'-end of the GATATC recognition sequence for *Eco*RV. However, this sequence-specific contact is usually missing when the main chain CO at 69 has the IN orientation: instead, the side chain of Asn70 makes van der Waals contacts with the deoxyribose of the nucleotide immediately after the 6 bp recognition sequence. DNA recognition and catalysis by *Eco*RV might thus entail the following sequence of events: Asn70 is perhaps involved first in the recognition of the specific DNA sequence; at a later stage in the reaction pathway, after the enzyme has identified its recognition sequence, a segment of the Q loop undergoes a flip in its conformation, so as to bring the main chain CO of Gln69 into the active site where it plays a role in catalysis; by this stage, the base-specific contact from Asn70 is no longer required.

For wild-type *Eco*RV and the moderately active Q69L mutant, the different configurations at residues 69 and 70 are both populated in the crystal structures, which suggests that the alternate conformations of these two proteins have similar free energies. The reaction pathway for Q69L features a conformational change preceding DNA cleavage that is much slower than any equivalent process with the wild-type enzyme (22). Hence, the difference between Q69L and wild-type *Eco*RV may be the energy barrier between the conformational states, rather than the free energy *per se*, perhaps as a consequence of the more intimate van der Waals contacts between the Q loop and the DNA backbone with Leu at position 69 instead of Gln. In contrast, only the OUT configuration for the main chain CO at residue 69 is populated in the crystal structure of Q69E (Table 2). Q69E has virtually no DNA cleavage activity (22), so this result is consistent with the view that catalysis by *Eco*RV requires the IN orientation for the main chain CO at residue 69.

A model for the pentavalent intermediate during phosphodiester hydrolysis by *Eco*RV was generated by two cycles of MD simulations (Fig. 4b and c), starting from the crystal structure of the enzyme–substrate complex for *Eco*RV. The resultant model (Fig. 4c) duplicates several of the distinctive features seen in the crystal structure of the enzyme–product complex for *Eco*RV (Fig. 4d), but which are absent in the crystal structure of the enzyme–substrate complex (Fig. 4a). MD thus appears to have yielded a plausible structure for this short-lived intermediate. If so, then the main chain CO of Gln69 is not just the final resting place for the metal ion attached to the *pro*S phosphoryl oxygen after the hydrolytic reaction, but is instead a coordination ligand for the key intermediate in the reaction. If the interaction between the main chain CO of Gln69 and the metal ion on the *pro*S phosphoryl oxygen is established upon the formation of the pentavalent intermediate, as suggested here, then this interaction would stabilise the transition state of the reaction and thus enhance the catalytic rate. Such an interaction demands the IN orientation for the main chain CO at position 69.

Other proposals for the mechanism of phosphodiester hydrolysis by *Eco*RV lack roles for either the main chain or the side chain functionalities of Gln69 (39,40). However, such proposals do not lead directly to a structure for the *Eco*RV–product complex akin to the crystal structure of the enzyme–product complex. In contrast, the scheme proposed here provides a direct route from the crystal structure of an inactive form of the enzyme–substrate complex, in which no DNA cleavage occurs, to the crystal structure of the enzyme–product complex (Fig. 4). A rearrangement at the active site allows the scissile phosphate to interact with metal ions at both the 90/74 and the 74/45 sites (15), while the subsequent formation of the pentavalent intermediate leads to the transfer of the second metal onto the main chain CO of Gln69 (Fig. 4c).

Many enzymes employ main chain amino or CO groups for stereospecific interactions with their substrates or cofactors. The *Eco*RV restriction enzyme provides one example: the R loop makes numerous hydrogen bonds to the bases in the recognition sequence, but several of these bonds emanate from peptidyl amino or carbonyl groups (4). Other examples are provided by Ca^{2+}-binding proteins, that often use main chain CO groups to coordinate Ca^{2+} ions (41). It is, however, substantially less common for peptidyl amino or CO groups to play direct roles in catalysis, as proposed here for the main chain CO of Gln69 in *Eco*RV. Nonetheless, there exist precedents for this behaviour. In both T4 endonuclease V and penicillin acylase, a key catalytic function is the N-terminal amino group of the polypeptide (42,43)

ACKNOWLEDGEMENTS

We thank John Barker, Mark Banfield, Louse Hancox and Neil Stanford for advice and discussions, Janet Baber for technical support, the staff of the Synchrotron Radiation Source at Daresbury for assistance and the BBSRC and the Wellcome Trust for financial support.

REFERENCES

1. Schildkraut,I., Banner,C.D.B., Rhodes,C.S. and Parekh,S. (1984) *Gene*, **27**, 327–329.
2. D'Arcy,A., Brown,R.S., Zabeau,M., van Resandt,R.W. and Winkler,F.K. (1985) *J. Biol. Chem.*, **260**, 1987–1990.
3. Luke,P.A., McCallum,S.A. and Halford,S.E. (1987) *Gene Amplif. Anal.*, **5**, 183–207.
4. Winkler,F.K., Banner,D.W., Oefner,C., Tsernoglou,D., Brown,R.S., Heathman,S.P., Bryan,R.K., Martin,P.D., Petratos,K. and Wilson,K.S. (1993) *EMBO J.*, **12**, 1781–1795.
5. Erskine,S.G., Baldwin,G.S. and Halford,S.E. (1997) *Biochemistry*, **36**, 7567–7576.
6. Halford,S.E. and Goodall,A.J. (1988) *Biochemistry*, **27**, 1771–1777.
7. Taylor,J.D. and Halford,S.E. (1989) *Biochemistry*, **28**, 6198–6207.
8. Waters,T.R. and Connolly,B.A. (1994) *Biochemistry*, **33**, 1812–1819.
9. Taylor,J.D., Badcoe,I.G., Clarke,A.R. and Halford,S.E. (1991) *Biochemistry*, **30**, 8743–8753.
10. Thielking,V., Selent,U., Köhler,E., Landraf,A., Wolfes,H., Alves,J. and Pingoud,A. (1992) *Biochemistry*, **31**, 3727–3732.
11. Szczelkun,M.D. and Connolly,B.A. (1995) *Biochemistry*, **34**, 10724–10733.

12. Erskine,S.G. and Halford,S.E. (1998) *J. Mol. Biol.*, **275**, 759–772.
13. Vipond,I.B. and Halford,S.E. (1995) *Biochemistry*, **34**, 1113–1119.
14. Martin,A.M., Horton,N.C., Lusetti,S., Reich,N.O. and Perona,J.J. (1999) *Biochemistry*, **38**, 8430–8439.
15. Baldwin,G.S., Sessions,R.B., Erskine,S.G. and Halford,S.E. (1999) *J. Mol. Biol.*, **288**, 87–104.
16. Kostrewa,D. and Winkler,F.K. (1995) *Biochemistry*, **34**, 683–696.
17. Perona,J.J. and Martin,A.M. (1997) *J. Mol. Biol.*, **273**, 207–225.
18. Wenz,C., Jeltsch,A. and Pingoud,A. (1996) *J. Biol. Chem.*, **271**, 5565–5573.
19. Thorogood,H., Grasby,J.A. and Connolly,B.A. (1996) *J. Biol. Chem.*, **271**, 8855–8862.
20. Stöver,T., Köhler,E., Fagin,U., Wende,W., Wolfes,H. and Pingoud,A. (1993) *J. Biol. Chem.*, **268**, 8645–8650.
21. Vipond,I.B. and Halford,S.E. (1996) *Biochemistry*, **35**, 1701–1711.
22. Hancox,E.L. and Halford,S.E. (1997) *Biochemistry*, **36**, 7577–7585.
23. Baldwin,G.S., Vipond,I.B. and Halford,S.E. (1995) *Biochemistry*, **34**, 705–714.
24. Otwinowski,Z. and Minor,W. (1996) *Methods Enzymol.*, **276**, 307–326.
25. Navaza,J. (1994) *Acta Crystallogr.*, **A50**, 157–163.
26. Abola,E.E., Sussman,J.L., Prilusky,J. and Manning,O. (1997) *Methods Enzymol.*, **277**, 556–571.
27. Collaborative Computational Project, Number 4 (1994) *Acta Crystallogr.*, **D50**, 760–763.
28. Murshudov,G.N., Vagin,A.A. and Dodson,E.J. (1997) *Acta Crystallogr.*, **D53**, 240–255.
29. Jones,T.A., Zou,J.-Y., Cowan,S.W. and Kjeldgaard,M. (1991) *Acta Crystallogr.*, **A47**, 110–119.
30. Lamzin,V.S. and Wilson,K.S. (1993) *Acta Crystallogr.*, **D49**, 129–147.
31. Boel,E., Brady,L., Brzozowski,A.M., Derewenda,Z., Dodson,G.G., Jensen,V.J., Petersen,S.B., Swift,H., Thim,L. and Woldike,H.F. (1990) *Biochemistry*, **29**, 6244–6249.
32. Kabsch,W. (1976) *Acta Crystallogr.*, **A32**, 922–923.
33. Bouguleret,L. Schwarzstein,M., Tsugita,A. and Zabeau,M. (1984) *Nucleic Acids Res.*, **12**, 3659–3676.
34. Grasby,J.A. and Connolly,B.A. (1992) *Biochemistry*, **31**, 7855–7861.
35. Viadiu,H. and Aggarwal,A.K. (1998) *Nature Struct. Biol.*, **5**, 910–916.
36. Newman,M., Lunnen,K., Wilson,G., Greci,J., Schildkraut,I. and Phillips,S.E.V. (1998) *EMBO J.*, **17**, 5466–5476.
37. Vipond,I.B., Baldwin,G.S. and Halford,S.E. (1995) *Biochemistry*, **34**, 697–704.
38. Stahl,F., Wende,W., Wenz,C., Jeltsch,A. and Pingoud,A. (1998) *Biochemistry*, **37**, 5682–5688.
39. Jeltsch,A., Alves,J., Wolfes,H., Maass,G. and Pingoud,A. (1993) *Proc. Natl Acad. Sci. USA*, **90**, 8499–8503.
40. Horton,N.C., Newberry,K.J. and Perona,J.J. (1998) *Proc. Natl Acad. Sci. USA*, **95**, 13489–13494.
41. Heizman,C.W. and Hunziker,W. (1991) *Trends Biochem. Sci.*, **16**, 98–103.
42. Vassylev,D.G., Kashiwagi,T., Mikami,Y., Ariyoshi,M., Iwai,S., Ohtsuka,E. and Morikawa,K. (1995) *Cell*, **83**, 773–782.
43. Duggleby,N.J., Tolley,S.P., Hill,C.P., Dodson,E.J., Dodson,G. and Moody,P.C.E. (1995) *Nature*, **373**, 264–268.