Metal-Dependent DNA Cleavage Mechanism of the I-*Cre*I LAGLIDADG Homing Endonuclease[†,‡]

Brett Chevalier,[§] Django Sussman,[§] Christian Otis,[||] Ann-Josée Noël,[||] Monique Turmel,[||] Claude Lemieux,[||] Kathy Stephens,[⊥] Raymond J. Monnat, Jr.,[⊥] and Barry L. Stoddard*[,§]

Graduate Program in Molecular and Cellular Biology, University of Washington and Fred Hutchinson Cancer Research Center, 1100 Fairview Avenue North A3-025, Seattle, Washington 98109, Departments of Genetics and Pathology, Box 357470, University of Washington, Seattle, Washington 98195, and Program in Evolutionary Biology, Department de biochimie, Laval Universite, Quebec City, Quebec G1K 7P4, Canada

Received May 20, 2004; Revised Manuscript Received August 13, 2004

ABSTRACT: The LAGLIDADG homing endonucleases include free-standing homodimers, pseudosymmetric monomers, and related enzyme domains embedded within inteins. DNA-bound structures of homodimeric I-*Cre*I and monomeric I-*Sce*I indicate that three catalytic divalent metal ions are distributed across a pair of overlapping active sites, with one shared metal participating in both strand cleavage reactions. These structures differ in the precise position and binding interactions of the metals. We have studied the metal dependence for the I-*Cre*I homodimer using site-directed mutagenesis of active site residues and assays of binding affinity and cleavage activity. We have also reassessed the binding of a nonactivating metal ion (calcium) in the wild-type enzyme−substrate complex, and determined the DNA-bound structure of two inactive enzyme mutants. The conclusion of these studies is that the catalytic mechanism of symmetric LAGLIDADG homing endonucleases, and probably many of their monomeric cousins, involves a canonical two-metal mechanism in each of two active sites, which are chemically and structurally tethered to one another by a shared metal ion. Failure to occupy the shared metal site, as observed in the presence of calcium or when the metal-binding side chain from the LAGLIDADG motif (Asp 20) is mutated to asparagine, prevents cleavage by the enzyme.

Homing is the process by which mobile intervening genetic sequences, either introns or inteins, are specifically duplicated into cognate recipient alleles that lack such a sequence (*1*−*5*). The process is induced by an endonuclease encoded by an ORF harbored within the intervening sequence (*6*). The endonuclease specifically recognizes a target sequence corresponding to the intron insertion site, generates a DNA double-strand break, and induces cellular mechanisms that repair the break. If the intron-containing allele is used as a template for repair, the endonuclease ORF is duplicated into the target site and the homing cycle is completed. Transfer of mobile introns can be extremely efficient, frequently occurring between different subcellular compartments of unrelated organisms (*7*), and sometimes allowing introns to overrun diverse lineages within entire biological families (*8*).

Homing endonucleases are widespread and found within introns and inteins in all biological superkingdoms. On the basis of primary sequence homology, four homing enzyme families have been identified: the LAGLIDADG, GIY-YIG, HNH, and His-Cys Box endonucleases (*5*). The largest family, LAGLIDADG, contains several hundred identified ORFs, many of which have been shown to encode functional endonucleases (*5*, *9*). A pair of conserved LAGLIDADG sequence motifs form the core of a structural interface between endonuclease domains or subunits, and contribute conserved acidic metal-binding residues to two overlapping enzyme active sites (*10*, *11*). Endonucleases containing a single motif per ORF form homodimers that recognize palindromic consensus target sequences; those with two motifs fold to form pseudosymmetric monomers capable of recognizing DNA sites with significant asymmetry (*5*). The structures of six LAGLIDADG enzymes bound to their DNA targets have been determined: two homodimers (I-*Cre*I[1] and I-*Mso*I) (*12*, *13*), two pseudosymmetric monomers [I-*Ani*I (*14*) and I-*Sce*I (*15*)], one artificially engineered chimera [H-*Dre*I, composed of a domain of I-*Dmo*I fused to a subunit of I-*Cre*I (*16*)], and an intein-associated endonuclease [PI-*Sce*I (*17*)]. The structures of four of these proteins (I-*Cre*I, I-*Mso*I, I-*Sce*I, and H-*Dre*I) were determined at relatively high resolutions (2.5−1.9 Å) and demonstrate the presence

[†] Funded by NIH grants to B.L.S. (GM4985 7), R.J.M. (CA88942), and B.C. (T32 CA80416), by the Natural Sciences and Engineering Research Council of Canada (Grant GP0002830 to C.L.), and by a fellowship from the American Cancer Society to D.S.

[‡] The I-*Cre*I D20N (PDB entry 1T9I) and Q47E (PDB entry 1T9J) structural models and corresponding diffraction data have been deposited in the Protein Data Bank.

* To whom correspondence should be addressed: Fred Hutchinson Cancer Research Center, 1100 Fairview Ave. N. A3-025, Seattle, WA 98109. E-mail: bstoddar@fhcrc.org. Phone: (206) 667-4031. Fax: (206) 667-6877.

[§] Graduate Program in Molecular and Cellular Biology, University of Washington and Fred Hutchinson Cancer Research Center.

[||] Departments of Genetics and Pathology, University of Washington.

[⊥] Laval Universite.

[1] Abbreviations: I-, intron-encoded (i.e., I-*Cre*I); BSA, bovine serum albumin; dATP, deoxyadenosine triphosphate; WT, wild type; TBE, Tris-borate EDTA.

10.1021/bi048970c CCC: $27.50
Published on Web 10/16/2004

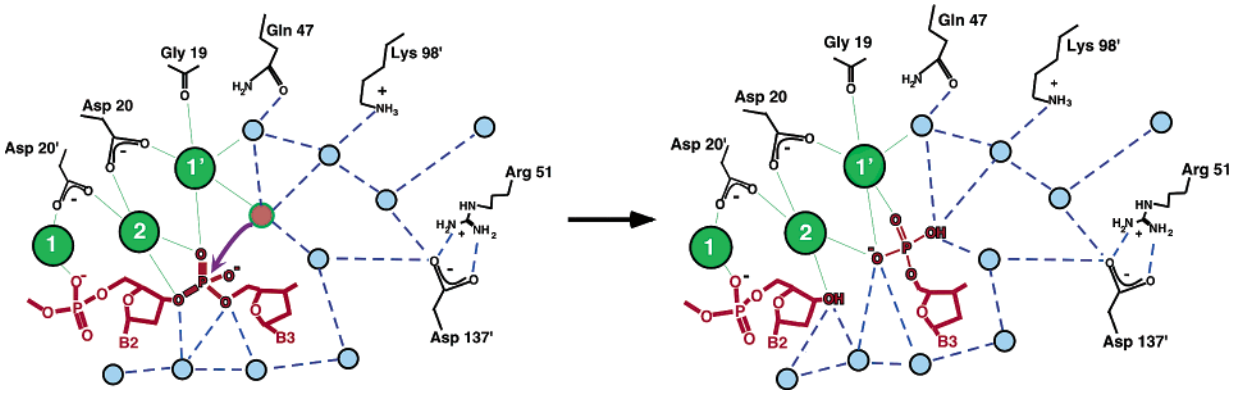

FIGURE 1: Schematic diagram of the mechanism of nucleophilic attack proposed for I-*Cre*I and homologous enzymes. Hydrolysis of the scissile phosphate follows a canonical two-metal mechanism. While a single metal ion (1/1′) and its associated water nucleophile are bound in each active site, the second metal ion (2) is shared between the active sites. Asp 20 and Asp 20′ are contributed by the LAGLIDADG motifs of the enzyme subunits and make the only direct contacts between the enzyme and the bound metal ions. A second residue mutated in this study (Gln 47) is within contact distance of a metal-associated, inner-shell water molecule, while the third and final residue that has been mutated (Lys 98) contacts a water belonging to the same metal ion's outer hydration shell. The active site pocket that contains additional ordered water molecules surrounding the scissile phosphate is also flanked by Arg 51 and Asp 137.

of three bound divalent metal ions distributed across a pair of overlapping active sites, with one central metal shared between the active sites. In contrast, the structure of one monomer (I-*Ani*I) and that of the intein-associated PI-*Sce*I, determined at a lower resolution, only demonstrate the presence of two bound metal ions; the central shared metal ion is not visible.

The I-*Cre*I active site and those of its relatives listed above appear to employ a canonical two-metal mechanism for phosphodiester hydrolysis (Figure 1). An unshared metal in each of the two overlapping active sites positions and activates a nucleophilic water, while the third metal ion stabilizes the transition-state phosphoanion and the 3′-hydroxylate leaving group for both strand cleavage events (*13*). The shared central metal is jointly coordinated by one conserved acidic residue from each LAGLIDA**D**G motif (Asp 20 in I-*Cre*I) and by oxygen atoms from scissile phosphates on each DNA strand. The unshared metals are individually coordinated by a single LAGLIDADG carboxylate oxygen, nonbridging DNA oxygen atoms, and a well-ordered coordination shell of water molecules. In I-*Cre*I, one of the metal-bound water molecules is in contact with a catalytically essential glutamine residue (Gln 47). A well-ordered network of water molecules is distributed in a large pocket surrounding the DNA scissile phosphate group; these solvent molecules extend from the metal-bound nucleophile to the leaving group 3′-oxygen. The walls of the active site solvent pocket are lined with several basic residues that contact both DNA and solvent atoms, including Arg 51 and Lys 98.

With the exception of the direct metal-binding residues from the LAGLIDADG motifs, the residues described above for the I-*Cre*I active site are only moderately conserved within the LAGLIDADG enzymes that have been crystallographically visualized (*18*) (Table 1). In addition, for the one monomeric structure shown to contain three bound catalytic metals (I-*Sce*I), the general scheme of sharing one metal between active sites is maintained, but the coordination and individual contacts with the bound metals differ significantly, both in comparison to those of I-*Cre*I and when its individual active sites are superimposed (*15*). Therefore, details of the structural mechanism of nucleophilic activation appear to differ significantly between enzyme subfamilies and even between active sites within individual asymmetric enzymes, probably contributing to differential rates of cleavage at the two scissile phosphates.

To directly characterize the role of bound divalent metal ions for the symmetric LAGLIDADG family member I-*Cre*I, we have conducted a series of biochemical and crystallographic studies. First, the relative activity and specificity of the wild-type enzyme were assayed for a panel of seven different divalent cations. Second, the binding sites of nonactivating calcium ions were redetermined by anomalous difference Fourier analysis (while the binding of activating manganese ions to all three sites had previously been demonstrated using anomalous difference mapping, the binding of calcium had been similarly modeled based on only $2F_o - F_c$ difference Fourier analyses, which are not as definitive). Third, three active site residues were mutagenized: the only residue involved in direct metal ligation (Asp 20 from the LAGLIDA**D**G motif), a second residue that contacts a metal-bound inner-shell water molecule (Gln 47), and a third residue that does not participate in metal binding but instead is found in a basic pocket near the scissile phosphate and nucleophilic water residue (Lys 98). The kinetic behavior and DNA binding properties of these mutants were assessed. Finally, the X-ray crystal structures of two point mutations at metal-associated residues, each differing from the wild type by a single atom (D20N and Q47E), were determined in an attempt to visualize the structural basis for their lack of catalytic activity.

MATERIALS AND METHODS

Cloning and Site-Directed Mutagenesis of the I-CreI Gene. The I-*Cre*I gene was PCR amplified from *Chlamydomonas reinhardtii* chloroplast DNA (SAG 11-32b) with primers 5′-TAAATAAAACATCCCATGGATACAAAATATAATAAAG-3′ and 5′-CACCGGATATGGACCGAACTGTCTC-3′. The underlined nucleotides in the former primer are those that were modified in the original I-*Cre*I sequence to introduce an *Nco*I site into the initiation codon. The addition of this site leads to a change within the second codon (N → D substitution). After cleavage with *Nco*I and *Xho*I, the PCR product was cloned into pET30a (Novagen, Madison, WI)

Table 1: Summary of Conserved Motif and Active Site Residues for LAGLIDADG Homing Endonuclease Structures

Enzyme	LAGLIDAD$_n$G	Metal Binding	Basic Pocket
I-*Cre*I	LAGFVDG**D**G	D_{20} Q_{47}	K_{98} R_{51}
I-*Mso*I	IAGFLDG**D**G	D_{21} Q_{49}	K_{104} K_{54}
I-*Dmo*I	LLGLIIG**D**G IKGLYVA**E**G	D_{21} Q_{42} E_{117} N_{129}	K_{120} K_{43} – K_{130}
I-*Ani*I	LVGLFEG**D**G LVGFIEA**E**G	D_{15} L_{36} E_{148} Q_{171}	K_{94} D_{40} K_{227} G_{174}
I-*Sce*I	GIGLILG**D**A LAYWFMD**D**G	D_{44} E_{61} D_{145} N_{192}	K_{122} – K_{223} –
PI-*Sce*I	LLGLWIG**D**G LAGLIDS**D**G	D_{218} D_{229} D_{326} T_{341}	K_{301} R_{231} K_{403} H_{343}
PI-*Pfu*I	LAGFIAG**D**G IAGLFDA**E**G	D_{149} D_{173} E_{250} M_{263}	L_{220} – K_{322} –

using *Escherichia coli* Max Efficiency Stabl2 cells (Invitrogen Life Technologies, Carlsbad, CA). Plasmid DNA was isolated with the Quiagen Plasmid MidiKit (Qiagen, Mississauga, ON). Mutations were introduced within codons 20, 47, and 98 of the I-*Cre*I gene using either Transformer Site-Directed Mutagenesis Kit version 2.0 (Clonetech, Palo Alto, CA) or the Quick Change Kit (Stratagene, La Jolla, CA). Three residue 20 point mutants (D20N, -L, and -A), four residue 47 mutants (Q47N, -A, -M, and -E), and two residue 98 mutants (K98R and -A) were prepared and their sequences verified.

Protein Expression and Purification. The coding sequences of variants Q47N, Q47A, K98R, and K98A were derived from a pET expression construct in which the wild-type sequence is cloned downstream of a region specifying an N-terminal histidine tag as described above; these two mutants were produced as fusion proteins and purified by affinity chromatography with subsequent removal of the tag as previously described (*19*). These I-*Cre*I constructs exhibit two single-amino acid differences relative to the wild-type protein sequence: an extra alanine residue at the N-terminus and an asparagine instead of an aspartate at position 2. The remaining mutants and additional preparations of the wild-type enzyme were derived from an untagged pET expression construct, which allowed their direct purification using cation exchange chromatography as unmodified proteins, also as previously described (*12*, *13*). The sequences of the latter variants differed from that of the wild type at only the mutagenized sites.

All protein preparations and DNA solutions were routinely treated with chelex resin prior to reconstitution with defined metal ion conditions in the experiments described below, using protocols previously optimized in our laboratories for homing endonuclease analyses (*20*).

Metal Dependence. The relative ability of the wild-type enzyme to catalyze the cleavage of its cognate homing site DNA target sequence, and variants of that sequence, was assayed as previously described (*21*−*23*) (Figure 2). The labeled DNA substrate (10 nM, 214 bp restriction fragment labeled with Klenow polymerase) was digested with 20 nM I-*Cre*I at 42 °C for 60 min in 100 mM Tris (pH 8.0), 100 mM KCl, 100 μg/mL BSA, and 2 mM cation. Reactions were quenched with loading buffer containing 0.5% SDS and proteinase K to effect product release. Samples were electrophoresed on 7% 0.5× TBE polyacrylamide gels containing 0.05% SDS at 100 V.

Determination of Affinity by Gel Shift Analyses. The affinity of enzyme constructs for the I-*Cre*I target site was determined in the presence of calcium, to prevent cleavage, by electrophoretic retardation (gel shift) assays (see Figure 3 for raw data and curve fitting and Table 2 for a summary). For each protein, the 232 bp end-labeled substrate was incubated for 20 min at 20 °C at an initial concentration of 0.1 or 1.0 pM in the presence of increasing concentrations of the protein ranging from 10 to 333 000 pM (333 nM), in a buffer containing 50 mM Tris-HCl (pH 8.5), 1.0 mM $CaCl_2$, 1.0 mM dithiothreitol, 1.0 μg/μL bovine serum albumin, 0.05 μg/μL poly(dI-dC), and 10% glycerol. A control assay without protein was also carried out. Bound and free DNA species were separated on a 7% polyacrylamide gel (Figure 3), containing 23 mM Tris-borate (pH 7.5) and 1.0 mM $CaCl_2$. Gels were dried, and the relative proportions of the bound and unbound DNA substrate were determined with the Fujix BAS1000 Phosphorimager and MacBas 2.5 (Fuji Photo System). Concentrations of the bound protein in each assay were inferred by assuming that two subunits of I-*Cre*I bind one DNA molecule. K_d values were estimated using the equation $[ES] = (S_tE_f)/(K_d + E_f)$, where [ES] is the concentration of the protein−DNA complex, S_t is the total concentration of the DNA substrate, and E_f is the concentration of the free protein.

The gel shift experiments for the wild type, Q47N, and Q47A were carried out in one group of experiments, with protein expressed and purified using His tag affinity chro-

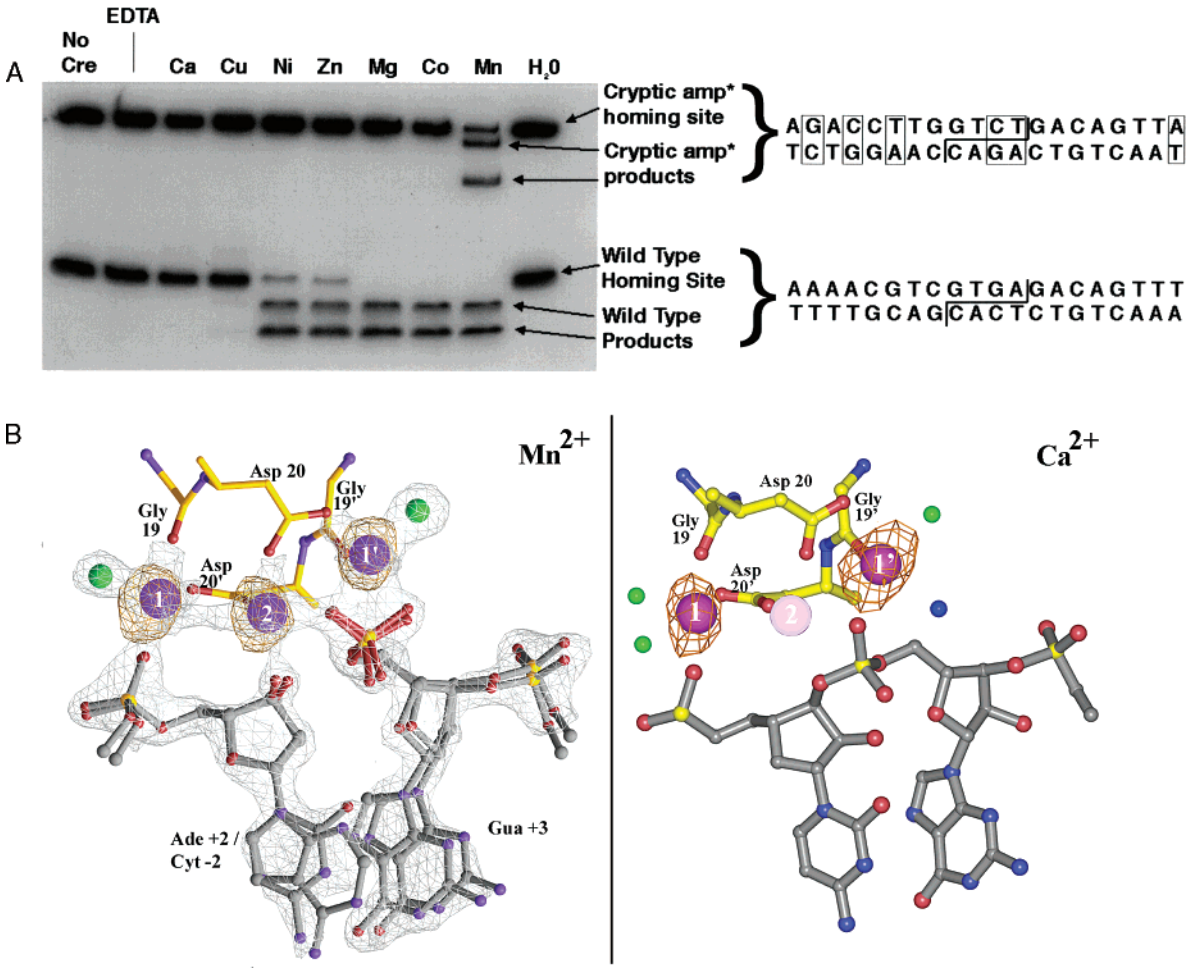

FIGURE 2: (A) Relative cleavage activities and substrate sequence specificity of I-*Cre*I as a function of metal ion identity. All cleavage reactions were carried out under optimal conditions of pH and salt concentration that yield complete cleavage of substrate, with magnesium as the cation, as described in Materials and Methods. Note the ability of the enzyme to cleave an alternate substrate, containing multiple sequence polymorphisms relative to the wild-type homing site, in the presence of manganese. (B) Anomalous difference maps demonstrate the binding of three manganese ions in an active I-*Cre*I−DNA complex [left; reprinted with permission from ref *13* (Copyright 2001 Nature Publishing Group)] as opposed to binding of calcium in an inactive I-*Cre*I complex (right). The position of the missing shared metal ion is indicated with the light sphere for comparison.

matography as described above. The remaining experiments, including an independently purified wild-type enzyme preparation, were carried out in a second, independent set of experiments using proteins expressed and purified as untagged constructs by cation exchange chromatography. The specific binding activities of tagged and untagged wild-type enzymes from these preparations were found to agree within the expected range of error for such experiments (data not shown). The total variance in the measured K_d values for the wild-type enzyme from both preparation methods, and independent preparations of each, is approximately 10-fold (from 0.1 to 1 nM), indicating that mutants displaying K_d values within or near that range should be interpreted as displaying minimal deficiencies in DNA binding affinity.

As presented in the Results and Discussion, the affinity of the enzyme is dependent upon the concentration of calcium ions (and presumably on the concentration of metal ions in general). Additional measurements of affinity at elevated concentrations of calcium indicate that for all the enzyme species reported in this paper, the affinity reaches a plateau at approximately 1 mM metal ion; therefore, affinities are quantitated as described above in the presence of 1 mM Ca^{2+}.

Endonuclease Assays. The 232 bp ^{33}P-labeled DNA substrate was generated by PCR amplification of *Chlamydomonas zebra* (SAG 10.83) chloroplast DNA using either end-labeled primers (*19*) or unlabeled primers in the presence of [α-^{33}P]dATP. The endonuclease activity of wild-type and variant I-*Cre*I was assayed as follows. The DNA substrate (0.5 nM) was incubated in the presence of I-*Cre*I (50 or 500 nM) at 37 °C in buffer A [20 mM Tris-HCl (pH 8.5), 1.0 mM dithiothreitol, 2.0 μg/mL bovine serum albumin, and 1.0 mM $MgCl_2$]. Aliquots (10 μL) of the incubation mixture were taken at 0 and 30 min, and reactions were stopped as described previously (*19*). The resulting samples were electrophoresed in a 6% polyacrylamide−7 M urea gel; the gel was then dried and exposed to an imaging plate (Fuji Photo System).

Determination of Kinetic Parameters. Kinetic parameters were determined under single-turnover conditions using a model developed previously for restriction endonucleases (*24*). These methods are used in place of steady-state kinetic assays because the rate-limiting reaction step during turnover by I-*Cre*I is product release. In these experiments, the first-order rate of disappearance of the substrate is measured at variable enzyme concentrations, against a single concentra-

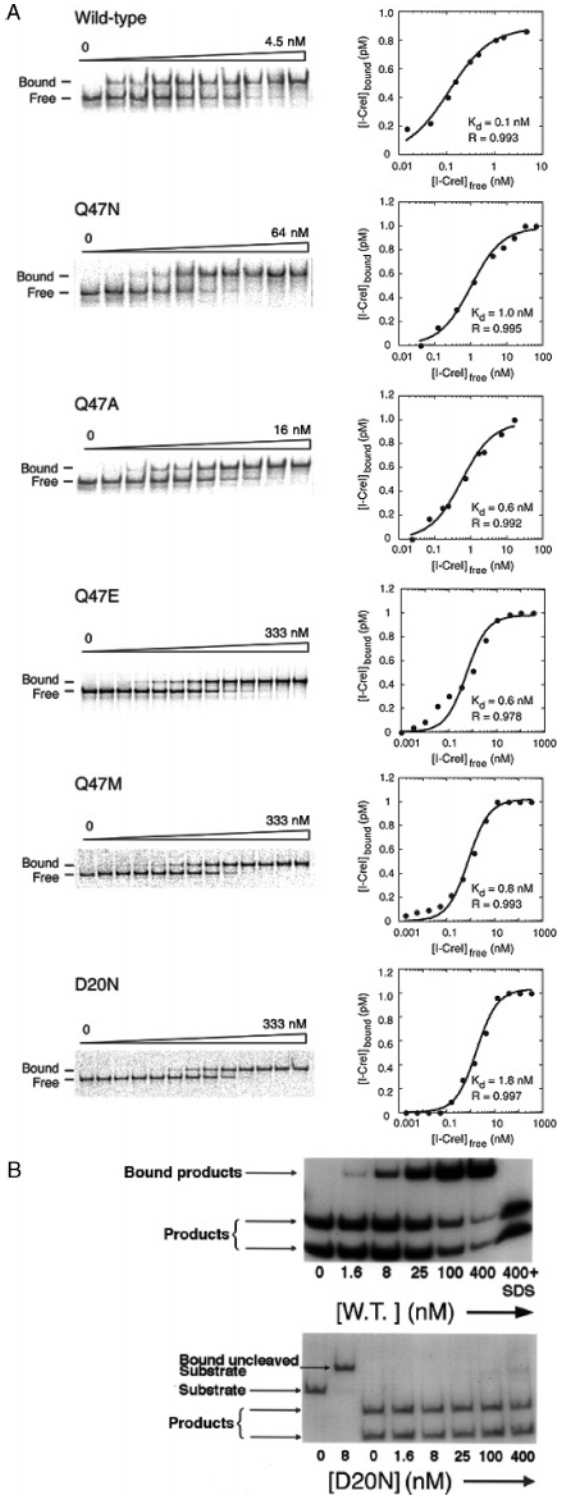

FIGURE 3: (A) DNA binding by the wild-type and variant I-*Cre*I protein. For each protein, the 232 bp end-labeled substrate was incubated at an initial concentration of 1.0 pM in the presence of 1.0 mM Ca^{2+} and increasing concentrations of the protein (as indicated above the electrophoretic profiles). A control assay without protein was also carried out. Bound and free DNA species were separated on a 7% polyacrylamide gel (left column), and concentrations of bound I-*Cre*I were plotted against the concentrations of free I-*Cre*I (right column). The correlation coefficients (R) of these experimental data with the equation describing the concentration of bound enzyme as a function of free enzyme are given. The estimated K_d values are also given. (B) Binding and gel shift of separate cleaved product half-sites by the wild type and D20N mutants. The point mutant efficiently binds the substrate but fails to end-hold cleaved products, as described in the Results and Discussion.

Table 2: Kinetic Parameters

I-*Cre*I	k^*_{max} (min^{-1})	K_m* (nM)	k^*_{max}/K_m* ($nM^{-1}\ min^{-1}$)	K_D (nM) (1 mM Ca^{2+})
WT	3×10^{-2}	0.0001	0.3	0.1
D20N[a,b]	–	–	–	1.8
D20A,L[b,c]	–	–	–	ND
Q47E[a,b]	–	–	–	0.6
Q47N	5×10^{-2}	350	1.4×10^{-4}	1.0
Q47M[b]	–	–	–	0.8
Q47A[b]	–	–	–	0.6
K98R[c]	5×10^{-2}	340	1.5×10^{-4}	ND
K98A[c]	3×10^{-3}	35	8.6×10^{-5}	N[c]

[a] These mutants were crystallized in complex with their DNA target sites as described in the text and figures. [b] These mutants did not display any detectable activity at the limits of the assay [corresponding to k^*_{max} values of less than $\sim 10^{-6}\ min^{-1}$, at time points extended as far as 48 h (see Materials and Methods)]. [c] These mutants, in gel shift affinity measurements, did not demonstrate a clean two-state shift from the unbound to the bound complex as a function of increasing enzyme concentration, and instead appear to exhibit behavior consistent with nonspecific association at elevated protein concentrations, and dissociation during electrophoretic analyses.

tion of substrate (10^{-14} M) far below the K_m of the reaction. The values of the pseudo-Michaelis constants for a single cleavage event (K_m* and k^*_{max}) are estimated by curve fitting the individual k* values obtained at various protein concentrations according to the equation $k^* = (k^*_{max}[\text{I-}Cre\text{I}]_0)/([\text{I-}Cre\text{I}]_0 + K_m^*)$, where $[\text{I-}Cre\text{I}]_0$ is the initial protein concentration. The value of k^*_{max} represents the magnitude of the limiting rate of the cleavage event leading to formation of fully cleaved products, while the value of K_m* reflects the affinity of the enzyme for the uncleaved substrate, by indicating the ratio of microscopic rate constants leading to loss versus formation of the initial ES complex. This constant is distinct from the more traditional steady-state value of K_m, which reflects an apparent dissociation constant averaged from all bound enzyme species (*25*).

In the assays with wild-type I-*Cre*I, the labeled DNA substrate (0.001 pM) was incubated with I-*Cre*I (0.005−2.0 pM) at 37 °C in buffer A. Identical conditions were used in the assays with the Q47N variant, except that the DNA substrate concentration was 250 pM and the protein concentration ranged from 5.0×10^4 to 8.0×10^5 pM. Aliquots of the reaction mixtures were taken at regular intervals (0−90 min), and the reactions were stopped as described previously (*26*). Samples were electrophoresed in 5% polyacrylamide−7 M urea gels, and the relative proportions of cleaved products were determined using the Fujix BAS1000 Phosphorimager and the MacBas 2.5 software (Fuji Photo System). The pseudo-first-order constant k* describing the disappearance of the DNA substrate at a given protein concentration was determined using the equation $[S] = [S]_0 e^{-k^* t}$, where $[S]_0$ and $[S]$ correspond to the substrate concentration at time 0 and time t, respectively. The parameters K_m* and k^*_{max} were estimated by curve fitting the k* values obtained at various protein concentrations according to the equation $k^* = (k^*_{max}[\text{I-}Cre\text{I}]_0)/([\text{I-}Cre\text{I}]_0 + K_m^*)$, where $[\text{I-}Cre\text{I}]_0$ is the initial protein concentration. All these parameters were calculated using KaleidaGraph 3.0 (Synergy Software, Reading, PA).

Those mutants that displayed no measurable activity after 90 min were then re-assayed after 24 and 48 h, and found to generate no measurable cleavage products above back-

ground. Subsequent experiments with the wild-type enzyme, in which the enzyme was incubated with nonspecific DNA at 37 and 42 °C and then assayed as described above, demonstrated significant loss of specific activity (>95%) after 24 h at 37 °C, and more rapid loss of activity ($t_{1/2} \sim 6$ h at 42 °C) (data not shown). Therefore, meaningful kinetic measurements using longer incubation time points are not possible.

Product Binding by the Wild Type and the D20N Mutant. The relative abilities of the two active constructs (WT and D20N) to retain bound products were visualized by gel shift analysis (Figure 3B). The labeled DNA substrate (10 nM, either intact or predigested with I-*Cre*I) was incubated with I-*Cre*I for 30 min at 0 °C in 100 mM Tris (pH 8.0), 10 mM $MgCl_2$, 100 mM KCl, 100 μg/mL BSA, and 25 ng/mL nonspecific competitor DNA [poly(dI-dC)]. Samples were electrophoresed on 7% 0.5× TBE polyacrylamide gels at 110 V and 4 °C. SDS (0.5%) and proteinase K treatment of the substrate−I-*Cre*I complex resulted in the release of product fragments.

Crystallography. The DNA for cocrystallization was purchased from Oligos Etc. (Wilsonville, OR) and consisted of two strands of sequence: 5′-GCAAAACGTCGTGAGA-CAGTTTCG-3′ and its complement, 5′-CGAAACTGTCT-CACGACGTTTTGC-3′. The construct forms a 24 bp blunt-ended pseudopalindromic duplex that differs at four positions between the two half-sites (Figure 1). Crystals were grown for wild-type, D20N, and Q47E I-*Cre*I bound to this DNA construct using a 2.7:1 (molar ratio) DNA/protein solution by hanging drop vapor diffusion against a reservoir containing 20 mM NaCl, 100 mM MES (pH ranging from 6.3 to 6.7), and PEG 400 (ranging from 20 to 35%, v/v). The final concentration of I-*Cre*I in the DNA/protein solution was 3.5 mg/mL. The crystallization drops also contained 10 mM calcium cations that inhibit cleavage, allowing identification of the bound metal ions by anomalous difference Fourier maps. For all three separate experiments, the crystal belongs to space group $P2_1$, with the following unit cell dimensions: $a \approx 43$ Å, $b \approx 68$ Å, and $c \approx 88$ Å (Table 1).

Crystals were removed directly from the drops in which they grew, suspended in a fiber loop, frozen in liquid nitrogen, and maintained at 100 K during data collection. Data from crystals of the D20N−DNA and Q47E−DNA complexes were collected at the Advanced Light Source (beamline 5.0.2) and in house, respectively; for comparison, a data set on the wild-type enzyme complex was collected in house. All data sets include a full 360° of oscillations to facilitate separation of redundant sets of Bijvoet pairs and calculation of anomalous difference Fourier maps and subsequent visualization of bound metal ions. Data were reduced using the DENZO/SCALEPACK crystallographic data reduction package (*27*) (Table 3).

The structures were determined via molecular replacement using EPMR (*28*) with the wild-type I-*Cre*I−DNA complex structure as an initial search model. All structures were modeled in XtalView (*29*) and O (*30*) and refined using CNS (*31*) with 5% of the data set aside for cross validation (*32*). The final refinement statistics (Table 1) for D20N were as follows: $d_{min} = 1.6$ Å, $R_{work} = 17.2$, and $R_{free} = 19.8$. Those for Q47E were as follows: $d_{min} = 2.0$ Å, $R_{work} = 21.8$, and $R_{free} = 25.4$. The refinement of a wild-type complex (but not the calculation of calcium anomalous difference Fourier maps, presented below) has been previously reported (*13*). Geometric analysis of the structures using PROCHECK (*33*) indicates no residues in any structure with generously allowed or unfavorable backbone dihedral angles.

Table 3: Data and Refinement Statistics

	D20N	Q47E
Data		
source	ALS 5.0.2	RAXIS-IV
resolution limit (Å)	1.6	2.0
wavelength (Å)	1.1	1.54
space group	*P*21	*P*21
unit cell parameters		
a, *b*, *c* (Å)	43.8, 67.3, 89.4	43.1, 67.9, 87.7
α, β, γ (deg)	90, 92.4, 90	90, 91.6, 90
no. of measured reflections	175670	81511
no. of unique reflections	67746	31039
R_{merge}[a]	0.035 (0.084)	0.042 (0.255)
completeness (%)[a]	98.9 (98.9)	90.6 (57.3)
Refinement		
R_{work} (%)	17.2	21.8
R_{free} (%)[b]	19.8	25.4
resolution (Å)	50−1.6	50−2.0
no. of atoms	5112	4608
no. of waters	656	189
rms deviations		
bond lengths (Å)	0.010	0.008
bond angles (deg)	1.460	1.272
Ramachandran plot		
favorable (%)	91.1	91.7
allowed (%)	8.9	8.3
generous (%)	0	0
unfavorable (%)	0	0
mean *B* value (Å^2)		
overall	19.7	38.8
protein	17.5	38.7
DNA	18.5	38.5
solvent	31.7	42.8
cations	18.1	39.1

[a] The numbers in parentheses are statistics from the highest-resolution shell. [b] R_{free} was calculated with 5% of the data withheld from the refinement.

RESULTS AND DISCUSSION

Metal Dependence. Seven divalent metal ions were assayed for cleavage activity with I-*Cre*I (Figure 2). Two of these metals (calcium and copper) fail to support cleavage; two (nickel and zinc) display reduced cleavage activity, and three (magnesium, cobalt, and manganese) display full activity under the conditions that were tested. The use of manganese in place of magnesium allows recognition and cleavage of DNA target sequences containing mismatches in a variety of positions relative to the target site, including the sequence shown in Figure 2 which contains five mismatches relative to the wild-type homing sequence and is found as a cryptic target site in the bacterial Amp^R sequence.

The binding of nonactivating calcium ions to the wild-type enzyme−DNA complex was visualized by collecting a new, highly redundant (360°) diffraction data set on a home X-ray source ($l = 1.54$ Å) and examining the resulting anomalous difference Fourier maps (Figure 2B). In contrast to a previous anomalous difference mapping study in the presence of manganese (*18*) (which demonstrated binding of three metal ions at equivalent occupancy), this equivalent analysis with calcium demonstrates that calcium fails to bind

in the central, "shared" site, possibly due to steric crowding imparted by its larger size. A density peak at this position in $2F_o - F_c$ maps, which was previously interpreted as a bound calcium ion (*18*), is now believed to represent a mixture of bound solvent and/or sodium ions. It is likely that the failure of calcium and other large divalent cations to facilitate cleavage may be due to their inability to bind in the sterically constrained central metal binding site, in which all six direct oxygen ligands are contributed by protein side chains and the DNA phosphate backbone.

Substrate Binding Affinity of the Wild-Type Enzyme and Point Mutants. The DNA binding affinities (in the presence of calcium) of wild-type I-*Cre*I and a series of active site point mutants were measured as described in Materials and Methods (see Figure 3 for raw data and curve fitting and Table 2 for a summary). For all mutants that display clean, two-state gel shift behavior (D20N and all Q47 variants), the measured dissociation constant varies by no more than 10-fold relative to that of the wild-type enzyme, demonstrating that these mutations induce relatively small effects on binding affinity as described in the text. In contrast, nonconservative mutations at D20 (D20A and -L) and both mutations at K98 (K98R and -A) appear to be significantly compromised in DNA binding, and fail to demonstrate clean, quantifiable gel shift behavior, possibly reflecting reduced affinities and elevated dissociation kinetics (Table 2).

The measured dissociation constant (K_d) for dissociation of I-*Cre*I from its DNA target is dependent on the concentration of the divalent cation. In the presence of 1 mM calcium, the lowest measured K_d for the wild-type enzyme is approximately 0.1 nM (Figure 3A). The affinity of the wild type (and mutant enzyme species described below) does not significantly increase at calcium ion concentrations of >1 mM. Decreasing the concentration of calcium to 0.1 mM increases the measured K_d for the same preparation of the wild-type enzyme to approximately 1.0 nM (data not shown). This effect is consistently reproducible for individual samples from the same wild-type enzyme preparations; however, as mentioned above, 10-fold variations in the value of K_d represent the limits of accuracy for this system and therefore should be considered to reflect relatively mild effects on affinity.

A series of mutations at Q47 uniformly display small, variable decreases in DNA binding affinity measured at 1 mM calcium, with K_d values ranging from approximately 0.6 to 1.0 nM (Figure 3A). Reduction of the calcium concentration to 0.1 mM causes a more significant reduction in affinity than that observed for the wild-type enzyme, with the K_d increasing by 100−500-fold in various experiments. Therefore, these mutants appear to exhibit compromised affinities for metal binding in the presence of DNA, but are still capable of binding DNA with near-wild-type affinities when sufficient concentrations of metal ion are present.

Mutations at residue D20 demonstrate greater heterogeneity in their effect on DNA affinity than those mutations at Q47 (Figure 3A). A conservative mutation at this position to asparagine (D20N) produces a clean shift from the unbound DNA substrate to the bound complex at 1 mM calcium, with a measured K_d slightly elevated relative to that of the wild-type enzyme. As observed for Q47 mutants, this mutant demonstrates an elevated sensitivity to reduced metal concentrations. This same mutation completely eliminates the ability of the enzyme to form a stable complex with DNA constructs corresponding to the cleaved products (Figure 3B). In contrast, the wild-type enzyme is strongly rate-limited by product release, and readily forms a stable complex with precleaved DNA products (Figure 3B).

Less conservative mutations at this position (D20A and D20L) fail to demonstrate a clean shift to the bound DNA complex at limiting concentrations of calcium (presumably due to an increased dissociation rate during the electrophoretic analysis), indicative of a more significant alteration of structure and binding interactions. Whereas the D20N mutation causes the loss of a single bound metal ion (see the results of structural analysis below), the D20L mutant would be expected to be compromised in the binding of metal ions at all three positions.

Kinetic Parameters of Wild-Type and Mutant Enzymes. Because the I-*Cre*I endonuclease is strongly rate-limited by product release, the kinetic behavior of the enzyme was studied by determining the pseudo-Michaelis rate constants for single cleavage events using the methods described by Halford and co-workers (*24*).

The measured kinetic rate constants (Figure 4 and Table 2), k^*_{max} and K_m^*, of the wild-type enzyme are 0.03 min^{-1} and 1.0×10^{-4} nM, respectively, giving a value for catalytic efficiency (k^*_{max}/K_m^*) of 0.3 nM^{-1} min^{-1} (Figure 4 and Table 2). With one exception (Q47N), point mutations at residues D20 and Q47 cause the complete loss of measurable cleavage activity, after incubation for up to 48 h (at which point the wild-type enzyme displays significant thermal inactivation as described in Materials and Methods). The Q47N variant must be present in a considerably greater amount than the wild-type protein for observation of the same rate of cleavage. It displays a value for k^*_{max} almost identical to that found for wild-type I-*Cre*I (3×10^{-2} min^{-1}); however, its K_m^* value (350 nM) is 6 orders of magnitude higher than that of wild-type I-*Cre*I. Unlike the wild-type enzyme, which retains the bound product after cleavage, Q47N releases its products after cleavage (Figure 4C). This behavior is similar to the product binding behavior of the catalytically inactive D20N mutant described above (Figure 3B). In the reactions catalyzed by wild-type I-*Cre*I, the cleavage products remain so tightly bound to the enzyme that the addition of a strong detergent (SDS) is necessary for dissociation of the complex before electrophoresis on polyacrylamide gels; the wild-type enzyme is even capable of shifting independent cleaved half-sites into a single bound complex (Figure 3B). In contrast, release of the Q47N cleavage products is not dependent upon the addition of a detergent; the mutant can bind substrate but cannot form a stable complex with cleaved products.

Whereas the dissociation constants (K_d) for wild-type and mutant enzymes, determined in the presence of 1 mM Ca^{2+}, range from 0.1 to 1 nM, the single-turnover kinetic measurements described above are performed using enzyme and DNA concentrations significantly lower than these K_d values. On the basis of the results of these studies, which indicate that the single-turnover cleavage rates are dependent on enzyme concentration (Figure 4B) and the fact that k_{cat}^* for the wild-type enzyme reaches a maximum at ~0.002 nM enzyme, it appears to be possible that the affinity of the enzyme for DNA is greater in the presence of Mg^{2+} than in the presence of Ca^{2+}. It is also possible, based on the data described above, that some of the mutations that do not

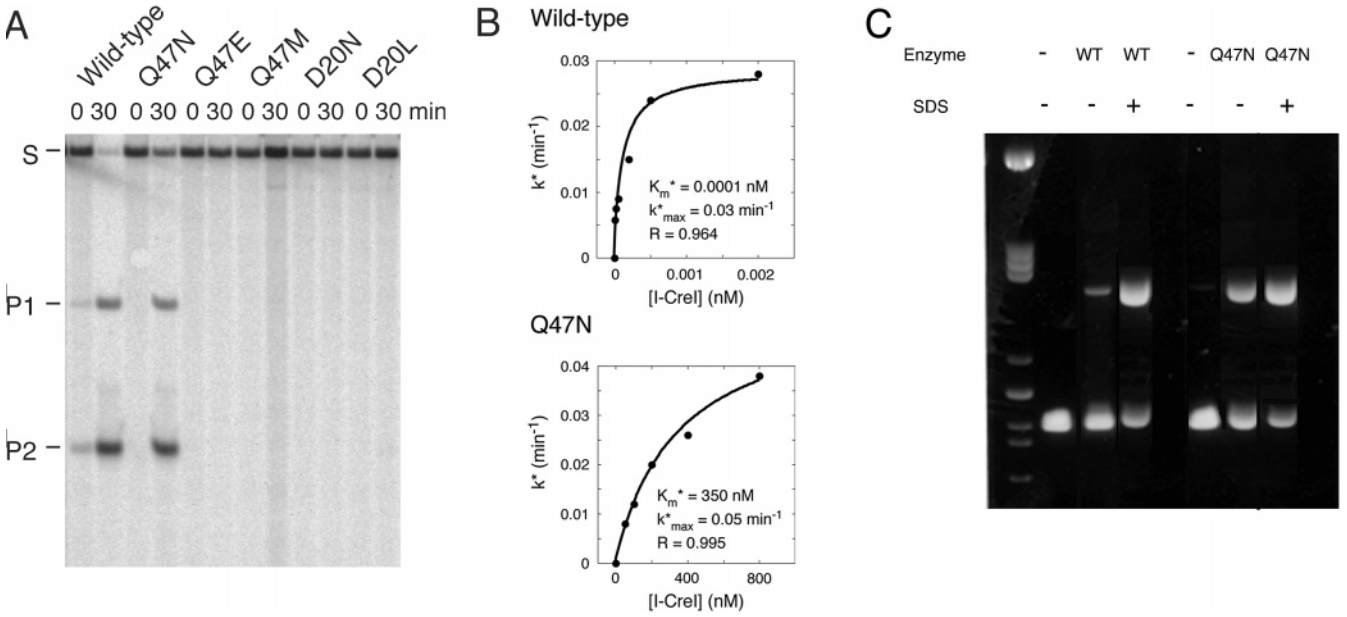

FIGURE 4: (A) Cleavage activity of wild-type and variant I-*Cre*I proteins. The 232 bp end-labeled substrate (S) was incubated at an initial concentration of 0.5 nM in the presence of each protein at 500 nM (panel B). Aliquots were removed from the reaction mixtures immediately after protein addition and 30 min later. The reaction products (P1 and P2) were separated on 6% polyacrylamide−7 M urea gels. (B) Cleavage kinetics of wild-type and Q47N I-*Cre*I under single-turnover conditions. For each protein, a set of cleavage reactions was carried out in which the variable concentration of the protein exceeded the initial concentration of the 232 bp substrate (0.001 and 250 pM in the assays with wild-type and Q47N I-*Cre*I, respectively). The first-order rate constants (k^*) obtained for these reactions are plotted against the protein concentrations. The correlation coefficients (R) of these experimental data with the equation describing k^* as a function of K_m^* and k^*_{max} are given. The estimated K_m^* and k^*_{max} values are also given. (C) Product release by the wild-type enzyme (left) is dependent on the presence of SDS; product release by Q47N (right) is not dependent on SDS.

significantly affect the affinity measured with Ca^{2+} (specifically, mutants at residue 47) do affect the value of K_m in the presence of Mg^{2+}.

Lysine 98, a Residue Not Involved in Metal Binding. As described above, the general effect of one conservative mutation at D20 (which directly binds catalytic metal ions) and most mutations at Q47 (which contacts a metal-bound water) is to only slightly reduce DNA binding affinity (in the presence of calcium), but to inactivate the enzyme's catalytic activity. As a point for comparison, individual mutations were introduced at a lysine residue (K98) that is not involved in metal binding, but is instead involved in indirect DNA contacts and formation of a solvent pocket near the scissile phosphate. The mutations at this position (K98R and K98A) are enzymatically active. The rate of cleavage is similar to that of the wild-type enzyme, but the value of K_m^* is increased significantly (Table 2). These mutants fail to demonstrate clean shifts to bound complexes in gel retardation analyses that permit unambiguous determination of K_d values.

Structures of Metal-Binding Residue Point Mutants. To better understand the structural and catalytic relationships between bound catalytic metals and coordinating residues D20 and Q47 (the latter through a metal-bound water), the structures of both D20N and Q47E mutants in the presence of DNA and Ca^{2+} ions were determined to high resolution (1.6 and 2.0 Å, respectively). Both mutations are essentially isosteric; only the charge, but not the size or shape, of the side chain is altered.

Since both terminal oxygens (OD1 and OD2) of D20 each make direct contact with an active site metal ion, the substitution of nitrogen for oxygen and the associated loss of a negative charge should alter the coordination of the metals. As expected, the terminal nitrogen of D20N no longer contacts the positively charged metals; instead, both D20N residues in the homodimer contort ~90° around their χ_1 axis (between Cα and Cβ) and each coordinate one of the two ions in the active sites (Figures 5A and 6A). These ions are positioned between the outer and central metals found in the productive protein−DNA complex. The twist of the D20N residues and the associated metal ions push the LAGLIDADG helices ~1 Å up and away from the DNA below; this further alters the shape of the solvent pocket and the positions of many of the water molecules within it. Finally, the scissile phosphates are no longer productively positioned within the minor groove for nucleophilic attack; rather, they adopt a B-form structure like the rest of the phosphates in the DNA duplex (Figure 6C).

Unlike D20, Q47 makes no direct contact with metal ions, the nucleophile, scissile phosphates, or leaving groups, nor is it strictly conserved across the LAGLIDADG endonuclease family. Since this residue contacts a metal-bound water and also is hydrogen bonded to D20, it was expected that a Q47E mutant would alter the position of the residue with respect to D20, the metal-bound water, and/or other elements of the active site. However, the high-resolution structure of Q47E is essentially identical to that of the wild-type protein; even the molecules within the extensive solvent network are similarly positioned and easily superimposed (Figures 5B and 6B). This implies that the exchange of the nitrogen for a second oxygen at each of the Q47 residues eliminates catalysis not by physically rearranging the active site but rather by changing the charge environment immediately surrounding this residue. Therefore, this mutation likely inhibits I-*Cre*I by drawing positive charge away from the metal ions at the expense of their ability to stabilize the

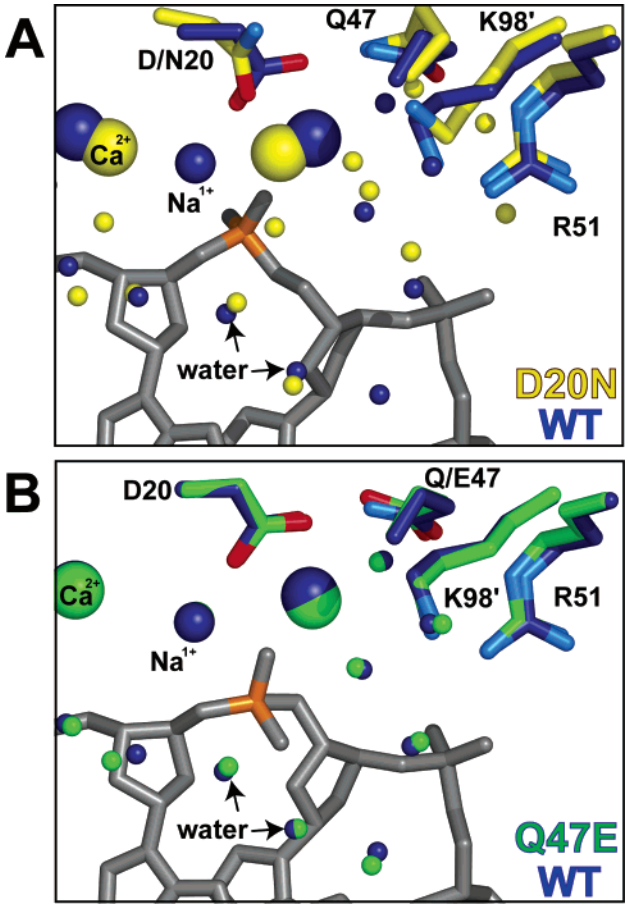

FIGURE 5: (A) Superposition of the D20N (yellow) and WT (blue) I-*Cre*I active sites viewed from the side of the scissile phosphate (stereoview). Residues 20, 47, and 98′ are clearly distorted; the solvent network is disrupted, and the two outer metals are pulled together at the exclusion of the central metal ion (modeled as Na^+ in the wild-type structure; see the text). DNA is from the D20N structure; the scissile phosphate is colored orange. (B) Superposition of the Q47E (green) and WT (blue) I-*Cre*I active sites viewed from the same orientation as panel A. Unlike D20N, the Q47E structure is essentially identical to the wild type. DNA is from the Q47E structure; the scissile phosphate is colored orange. The central sodium ion is present in both structures.

nucleophile, pentacoordinate phosphate intermediate and/or the 3′-oxygen leaving groups. Alternately, this change may reverse the polarity of a hydrogen bonding network within the solvent molecules in the pocket that extends from the nucleophile to the leaving group (Figure 1B).

At the central, shared metal binding site in the Q47E structure (within contact distance of D20 and D20′), density corresponding to an octahedrally coordinated ion is observed. As for the results observed for the wild-type I-*Cre*I−DNA−Ca^{2+} complex (described above), electron density at the position corresponding to the central bound metal ion could not be appropriately modeled by a calcium ion, and anomalous difference experiments again indicate that calcium fails to occupy that site. Given the lack of an anomalous signal, the octahedral coordination around the density, and the components of the mother liquor, a sodium ion is modeled at this position in the Q47E structure.

In contrast to wild-type I-*Cre*I and its isosteric point mutant Q47E, the structure of I-*Sce*I bound to uncleaved DNA contains all three bound calcium ions (*15*). In that structure, the distances between metal ligands at the center of the complex are slightly longer than those observed for I-*Cre*I, apparently allowing the calcium ion to bind at that site. Therefore, cleavage by these enzymes appears to be inhibited in the presence of calcium by different combinations of parameters: the failure to occupy a catalytically essential site (by I-*Cre*I) and by inappropriate physical characteristics for nuclease activity, such as the size-to-charge ratio of the cation and the pK_a of metal-bound solvent molecules (for both enzymes).

Variance of Active Site Structure and Cleavage Mechanisms across the LAGLIDADG Family. The most general conclusion of the work presented in this paper is that the catalytic mechanism of I-*Cre*I involves a canonical two-metal mechanism in each identical active site, both of which are chemically and structurally tethered to one another by a shared metal ion. This shared metal appears to participate in the cleavage of both DNA strands, by stabilizing each independent transition state and leaving group in the separate reactions.

The structures of the three additional LAGLIDADG endonuclease−DNA complexes that have been determined at relatively high resolution (I-*Mso*I, I-*Sce*I, and H-*Dre*I) all indicate the presence of three bound divalent metal ions coordinated by a pair of overlapping active sites. The structures of these enzyme−DNA complexes differ somewhat in the precise position and binding interactions of the metals but point to similar mechanisms where each strand is cleaved using a canonical two-metal mechanism for phosphodiester hydrolysis. Whether this unusual structural feature, a shared central divalent metal ion, imparts any particular kinetic order (or simultaneity) to the individual cleavage events is not known for the homodimeric enzymes. In contrast, the structure of the asymmetric I-*Sce*I−DNA complex (*15*) clearly demonstrates that DNA cleavage must involve sequential cleavage of coding and noncoding DNA strands, with a significant conformational rearrangement of the active sites relative to DNA occurring between the two reactions.

At physiological pH, phosphate ester bonds have large barriers to cleavage even though they are thermodynamically unstable (*34*). To efficiently catalyze the cleavage of phosphate esters, several chemical features are required, including a nucleophile, a basic moiety to activate and position the nucleophile, a general acid to protonate the leaving group, and the presence of one or more positively charged groups to stabilize the phosphoanion transition state (*35*). The diversity of chemical groups and metal ions available to proteins has made it possible for evolution to arrive at many diverse strategies that satisfy the above requirements. A common feature of many nuclease catalysts (and other phosphoryl transfer enzymes) is the use of bound metal ions as cofactors. Restriction endonucleases utilizing one, two, and even three bound metals per active site have been observed and described crystallographically (*36*). Metal ions can act as Lewis acids by lowering the pK_a of their directly coordinated water molecules. A resulting hydroxide may then serve as either a nucleophile or a general base. Alternatively, a metal-bound water molecule can efficiently protonate the leaving group. Perhaps most importantly, the positive charge of the divalent metal ion can stabilize the −2 charge that accumulates at the phosphoanion transition state relative to the −1 charge of the ground state.

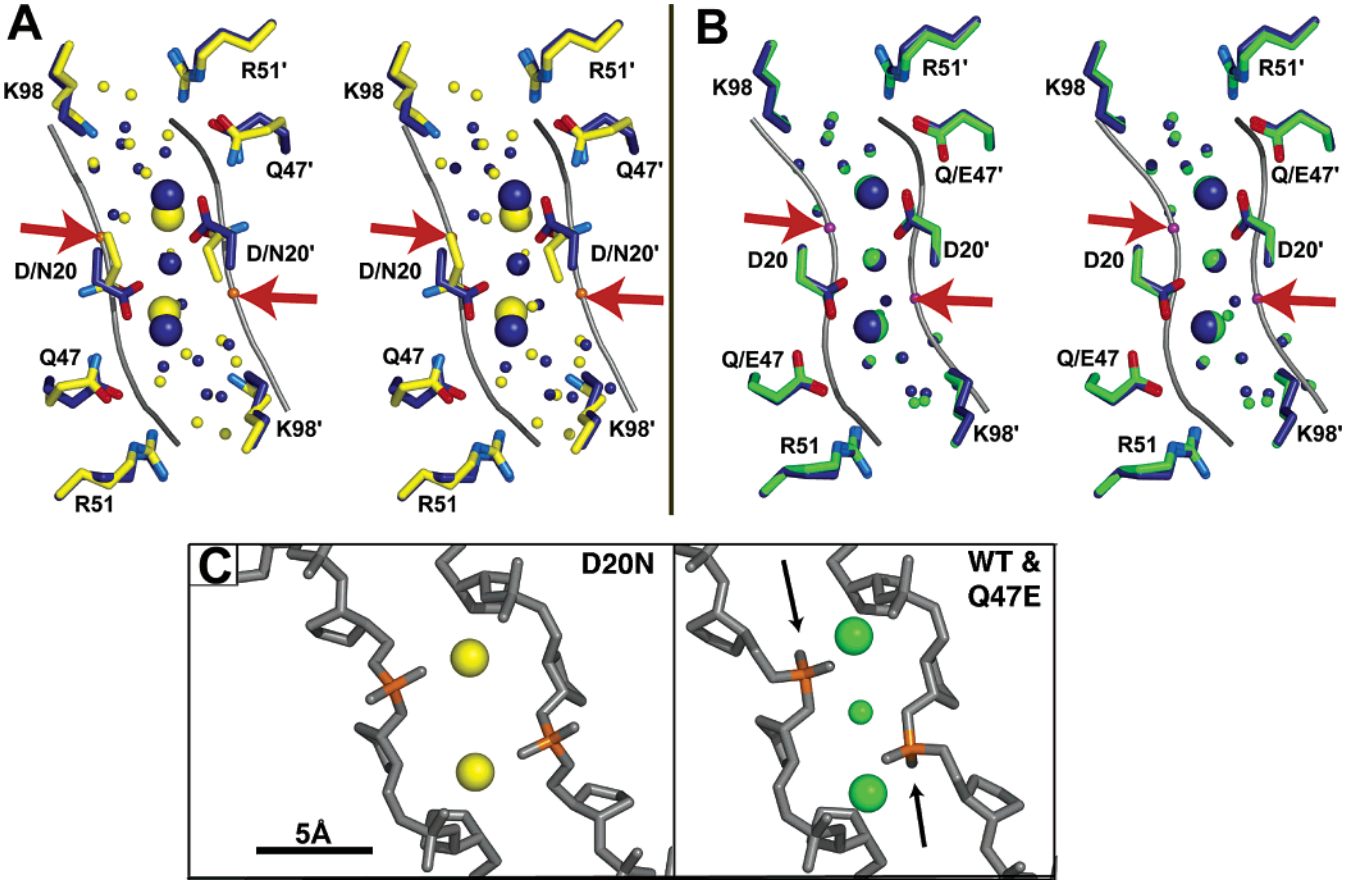

FIGURE 6: (A) Stereoview of superimposed active site residues from D20N (yellow carbons) and wild type (blue carbons) I-*Cre*I, both from structures in complex with their DNA target. The DNA backbone from the D20N cocrystal structure is represented with thin ribbons, and the positions of the scissile phosphates are denoted with red arrows. The distortion of the solvent network (small spheres) and residues 20, 20′, 47, 47′, 98, and 98′ is clearly visible. While the wild-type structure has three metals between the two active sites, the D20N structure clearly has only two (large spheres). The structure is viewed looking down through the of the enzyme, toward the bound DNA underneath. (B) Stereoview of superimposed active site residues from Q47E (green carbons) and wild-type I-*Cre*I, again from their respective DNA-bound complexes. Comparison of these two active sites reveals them to be essentially identical; even the positions of the water molecules within the solvent network are maintained. Arrows point to the scissile phosphates (DNA from the Q47E structure). The orientation is the same as in panel A. (C) Structure of the DNA backbone around the active sites in the D20N (left) and Q47E/WT (right) structures. In the Q47E and wild-type cocrystal structures, the scissile phosphates are flipped into a narrow minor groove ready for nucleophilic attack (arrows show the direction of attack). In contrast, the scissile phosphates are not productively positioned within the minor groove in the D20N structure, but are B-form. Similarly, the minor groove is ~2 Å wider in the D20N structure.

The general mechanistic features of DNA hydrolysis described above are clearly imparted in I-*Cre*I, at least in part, by the three residues that have been shown to be important for catalysis in this paper: aspartate 20, glutamine 47, and lysine 98. Their corresponding residues in the related enzyme I-*Ceu*I have also been shown to be essential for activity (*37*). However, outside of the I-*Cre*I/I-*Ceu*I enzyme branch, only the metal-binding residues from the LAGLIDADG motif (Asp 20 in I-*Cre*I) are well-conserved, whereas the remaining residues in the active site (including Q47 and K98) are remarkable for their chemical and structural diversity (*18*) (Table 1). In particular, those residues, such as Lys 98, involved in interactions with solvent molecules (including those in contact with the scissile phosphate) are poorly conserved, and in some cases absent. The only obvious common chemical feature of many of those residues is the capacity to either donate or accept one or more hydrogen bonds. It is possible that these peripheral active site residues are responsible for positioning and polarizing the solvent network in the active site to facilitate efficient proton transfer to and from nucleophiles and 3′ leaving groups. Each branch of closely related enzymes may have adopted a unique active site solvent packing arrangement that is highly specialized; furthermore, this rapidly diverging enzyme family may be broadly sampling and adopting significantly different combinations and configurations of chemical groups and associated water molecules to fulfill the catalytic roles described above.

The mechanism of DNA cleavage has also been extensively studied for several restriction endonucleases [including *Mun*I (*38*), *Cfr*10I (*39*), and *Eco*RV (*40*−*43*)], via site-directed mutagenesis of their active sites. Many of these studies focus on obvious candidates in various active sites for binding of divalent metal ions, and lead to relatively consistent conclusions about the role of those metal ions in positioning and activating water nucleophiles and proton donors, and in stabilizing phosphoanion transition states. Additionally, these studies indicate that metal binding and release can be a dynamic and transient component of the reaction pathway, with metal ions being acquired and/or lost at specific positions as the scissile phosphodiester bond is converted to a free 5′-terminal phosphate and 3′-hydroxyl. All of these results are consistent with the effect of similar mutations in the I-*Cre*I active site. Of particular interest to

the results reported here is a recent study of the role of long-range electrostatic contacts in the active site of *Eco*RV (*40*). That study indicated that mutations that alter the charge of polar residues located more than 7 Å from the scissile phosphate decrease cleavage rates by 3−5 orders of magnitude. Furthermore, second-site point mutations that restore wild-type charge balance in the active site restore a significant amount of catalytic efficiency. This observation indicates that "moderate-range electrostatic effects play a significant role in modulating the efficiency of phosphoryl transfer" (*40*). This same interpretation would appear to be an attractive way to summarize the effects of the Q47E mutation in I-*Cre*I (which alters the charge some distance from the scissile phosphate and bound metal, but not the structure of the active site). It also provides a reasonable explanation for the variable conservation patterns observed for "outer-shell" polar residues such as K98. As summarized above, these residues are highly variable in the LAGLIDADG enzyme family but collectively maintain similar charge states around the scissile phosphate pocket.

ACKNOWLEDGMENT

We thank the structural biology program at the Fred Hutchinson Cancer Research Center for assistance and advice on this project.

REFERENCES

1. Dujon, B. (1989) Group I introns as mobile genetic elements: facts and mechanistic speculations—a review, *Gene 82*, 91−114.
2. Lambowitz, A. M., and Belfort, M. (1993) Introns as mobile genetic elements, *Annu. Rev. Biochem. 62*, 587−622.
3. Belfort, M., and Perlman, P. S. (1995) Mechanisms of intron mobility, *J. Biol. Chem. 270*, 30237−30240.
4. Belfort, M., and Roberts, R. J. (1997) Homing endonucleases: keeping the house in order, *Nucleic Acids Res. 25*, 3379−3388.
5. Chevalier, B. S., and Stoddard, B. L. (2001) Homing endonucleases: structural and functional insight into the catalysts of intron/intein mobility, *Nucleic Acids Res. 29*, 3757−3774.
6. Jacquier, A., and Dujon, B. (1985) An intron-encoded protein is active in a gene conversion process that spreads an intron into a mitochondrial gene, *Cell 41*, 383−394.
7. Turmel, M., Cote, V., Otis, C., Mercier, J. P., Gray, M. W., Lonergan, K. M., and Lemieux, C. (1995) Evolutionary transfer of ORF-containing group I introns between different subcellular compartments (chloroplast and mitochondrion), *Mol. Biol. Evol. 12*, 533−545.
8. Cho, Y., Qiu, Y.-L., Kuhlman, P., and Palmer, J. D. (1998) Explosive invasion of plant mitochondria by a group I intron, *Proc. Natl. Acad. Sci. U.S.A. 95*, 14244−14249.
9. Dalgaard, J. Z., Klar, A. J., Moser, M. J., Holley, W. R., Chatterjee, A., and Mian, I. S. (1997) Statistical modeling and analysis of the LAGLIDADG family of site-specific endonucleases and identification of an intein that encodes a site-specific endonuclease of the HNH family, *Nucleic Acids Res. 25*, 4626−4638.
10. Heath, P. J., Stephens, K. M., Monnat, R. J., and Stoddard, B. L. (1997) The structure of I-*Cre*I, a group I intron-encoded homing endonuclease, *Nat. Struct. Biol. 4*, 468−476.
11. Duan, X., Gimble, F. S., and Quiocho, F. A. (1997) Crystal structure of PI-SceI, a homing endonuclease with protein splicing activity, *Cell 89*, 555−564.
12. Chevalier, B., Turmel, M., Lemieux, C., Monnat, R. J., and Stoddard, B. L. (2003) Flexible DNA target site recognition by divergent homing endonuclease isoschizomers I-CreI and I-MsoI, *J. Mol. Biol. 329*, 253−269.
13. Chevalier, B. S., Monnat, R. J., Jr., and Stoddard, B. L. (2001) The homing endonuclease I-CreI uses three metals, one of which is shared between the two active sites, *Nat. Struct. Biol. 8*, 312−316.
14. Bolduc, J. M., Spiegel, P. C., Chatterjee, P., Brady, K. L., Downing, M. E., Caprara, M. G., Waring, R. B., and Stoddard, B. L. (2003) Structural and biochemical analyses of DNA and RNA binding by a bifunctional homing endonuclease and group I intron splicing cofator, *Genes Dev. 17*, 2875−2888.
15. Moure, C. M., Gimble, F. S., and Quiocho, F. A. (2003) The crystal structure of the gene targeting homing endonuclease I-SceI reveals the origins of its target site specificity, *J. Mol. Biol. 334*, 685−696.
16. Chevalier, B. S., Kortemme, T., Chadsey, M. S., Baker, D., Monnat, R. J., Jr., and Stoddard, B. L. (2002) Design, activity and structure of a highly specific artificial endonuclease, *Mol. Cell 10*, 895−905.
17. Moure, C. M., Gimble, F. S., and Quiocho, F. A. (2002) Crystal structure of the intein homing endonuclease PI-SceI bound to its recognition sequence, *Nat. Struct. Biol. 9*, 764−770.
18. Chevalier, B. S., and Stoddard, B. L. (2001) Homing endonucleases: structural and functional insight into the catalysts of intron/intein mobility, *Nucleic Acids Res. 29*, 3757−3774.
19. Lucas, P., Otis, C., Mercier, J. P., Turmel, M., and Lemieux, C. (2001) Rapid evolution of the DNA-binding site in LAGLIDADG homing endonucleases, *Nucleic Acids Res. 29*, 960−969.
20. Galburt, E. A., Chevalier, B., Tang, W., Jurica, M. S., Flick, K. E., Monnat, R. J., and Stoddard, B. L. (1999) A novel endonuclease mechanism directly visualized for I-PpoI, *Nat. Struct. Biol. 6*, 1096−1099.
21. Durrenberger, F., and Rochaix, J.-D. (1993) Characterization of the cleavage site and the recognition sequence of the I-*Cre*I DNA endonuclease encoded by the chloroplast ribosomal intron of *Chlamydomonas reinhardtii*, *Mol. Gen. Genet. 236*, 409−414.
22. Thompson, A. J., Yuan, X., Kudlicki, W., and Herrin, D. L. (1992) Cleavage and recognition pattern of a double-strand-specific endonuclease (I-*Cre*I) encoded by the chloroplast 23S rRNA intron of *Chlamydomonas reinhardtii*, *Gene 119*, 247−251.
23. Wang, J., Kim, H.-H., Yuan, X., and Herrin, D. L. (1997) Purification, biochemical characterization and protein-DNA interactions of the I-*Cre*I endonuclease produced in *Escherichia coli*, *Nucleic Acids Res. 25*, 3767−3776.
24. Halford, S. E., Johnson, N. P., and Grinsted, J. (1980) The EcoRI restriction endonuclease with bacteriophage λ DNA, *Biochem. J. 191*, 581−592.
25. Fersht, A. (1985) in *Enzyme structure and mechanism*, pp 475, W. H. Freeman and Co., New York.
26. Drouin, M., Lucas, P., Otis, C., Lemieux, C., and Turmel, M. (2000) Biochemical characterization of I-CmoeI reveals that this H−N−H homing endonuclease shares functional similarities with H−N−H colicins, *Nucleic Acids Res. 28*, 4566−4572.
27. Otwinowski, Z., and Minor, W. (1997) Processing of X-ray diffraction data collected in oscillation mode, *Methods Enzymol. 276*, 307−326.
28. Kissinger, C. R., and Gehlhaar, D. K. (1997) *EPMR*, Agouron Pharmaceuticals, La Jolla, CA.
29. McRee, D. E. (1999) A Versatile Program for Manipulating Atomic Coordinates and Electron Density, *J. Struct. Biol. 125*, 156−165.
30. Jones, T. A., Zou, J.-Y., Cowan, S. W., and Kjeldgaard, M. (1991) Improved methods for building protein models in electron density maps and the location of errors in these models, *Acta Crystallogr. A47*, 110−119.
31. Brunger, A. T., Adams, P. D., Clore, G. M., DeLano, W. L., Gros, P., Grosse-Kunstleve, R. W., Jiang, J. S., Kuszewski, J., Nilges, M., Pannu, N. S., Read, R. J., Rice, L. M., Simonson, T., and Warren, G. L. (1998) Crystallography & NMR system: A new software suite for macromolecular structure determination, *Acta Crystallogr. D54*, 905−921.
32. Brunger, A. (1993) Assessment of phase accuracy by cross validation: the free *R* value. Methods and Applications, *Acta Crystallogr. D49*, 24−36.
33. Laskowski, R. J., Macarthur, M. W., Moss, D. S., and Thornton, J. M. (1993) PROCHECK: a program to check the stereochemical quality of protein structures, *J. Appl. Crystallogr. 26*, 283−291.
34. Westheimer, F. H. (1987) Why nature chose phosphates, *Science 235*, 1173−1178.
35. Galburt, E., and Stoddard, B. L. (2002) Catalytic mechanisms of restriction and homing endonucleases, *Biochemistry 41*, 13851−13860.
36. Pingoud, A., and Jeltsch, A. (2001) Structure and function of type II restriction endonucleases, *Nucleic Acids Res. 29*, 3705−3727.

37. Turmel, M., Otis, C., Cote, V., and Lemieux, C. (1997) Evolutionarily conserved and functionally important residues in the I-CeuI homing endonuclease, *Nucleic Acids Res. 25*, 2610−2619.
38. Lagunavicius, A., and Siksnys, V. (1997) Site-directed mutagenesis of putative active site residues of MunI restriction endonuclease: replacement of catalytically essential carboxylate residues triggers DNA binding specificity, *Biochemistry 36*, 11086−11092.
39. Skirgaila, R., Grazulis, S., Bozic, D., Huber, R., and Siksnys, V. (1998) Structure-Based Redesign of the Catalytic/Metal Binding Site of Cfr10I Restriction Endonuclease Reveals the Importance of Spatial Rather Than Sequence Conservation of Active Centre Residues, *J. Mol. Biol. 279*, 473−481.
40. Horton, N. C., Otey, C., Lusetti, S., Sam, M. D., Kohn, J., Martin, A. M., Ananthnarayan, V., and Perona, J. J. (2002) Electrostatic contributions to site specific DNA cleavage by EcoRV endonuclease, *Biochemistry 41*, 10754−10763.
41. Sam, M. D., and Perona, J. J. (1999) Catalytic roles of divalent metal ions in phosphoryl transfer by EcoRV endonuclease, *Biochemistry 38*, 6576−6586.
42. Thomas, M. P., Brady, R. L., Halford, S. E., Sessions, R. B., and Baldwin, G. S. (1999) Structural analysis of a mutational hotspot in the EcoRV restriction endonuclease: a catalytic role for a main chain carbonyl group, *Nucleic Acids Res. 27*, 3438−3445.
43. Vipond, I. B., and Halford, S. E. (1996) Random mutagenesis targeted to the active site of the EcoRV restriction endonuclease, *Biochemistry 35*, 1701−1711.

BI048970C