The LAGLIDADG Homing Endonuclease Family

Brett Chevalier, Raymond J. Monnat, Jr., Barry L. Stoddard

1 Introduction

The LAGLIDADG protein family includes the first identified and biochemically characterized intron-encoded proteins (Dujon 1980; Lazowska et al. 1980; Jacquier and Dujon 1985), as described in this volume by Dujon. It has been variously termed the 'DOD', 'dodecapeptide', 'dodecamer', and 'decapeptide' endonuclease family, based on the conservation of a ten-residue sequence motif (Dujon 1989; Dujon et al. 1989; Belfort et al. 1995; Belfort and Roberts 1997; Dalgaard et al. 1997; Chevalier and Stoddard 2001). The LAGLIDADG endonucleases are the most diverse of the homing endonuclease families. Their host range includes the genomes of plant and algal chloroplasts, fungal and protozoan mitochondria, bacteria and *Archaea* (Dalgaard et al. 1997). One reason for the wide phylogenetic distribution of LAGLIDADG genes appears to be their remarkable ability to invade unrelated types of intervening sequences, including group I introns, archaeal introns and inteins (Belfort and Roberts 1997; Chevalier and Stoddard 2001). Descendents of LAGLIDADG homing endonucleases also include the yeast HO mating type switch endonuclease (Jin et al. 1997), which is encoded by an independent reading frame rather than within an intron, but does carry remnants of an inactive intein domain (Haber and Wolfe, this Vol.), and maturases that assist in RNA splicing (Delahodde et al. 1989; Lazowska et al. 1989; Schafer et al. 1994; Geese and Waring 2001; Caprara and Waring, this Vol.).

B. Chevalier, B.L. Stoddard (e-mail: bstoddar@fhcrc.org)
Division of Basic Sciences, Fred Hutchinson Cancer Research Center, 1100 Fairview Ave. N. A3–025, Seattle, Washington 98109, USA

R.J. Monnat, Jr.
Departments of Pathology and Genome Sciences, Box 357470, University of Washington, Seattle, Washington 98195, USA

Nucleic Acids and Molecular Biology, Vol. 16
Marlene Belfort et al. (Eds.)
Homing Endonucleases and Inteins

Members of the LAGLIDADG family are segregated into groups that possess either one or two copies of the conserved LAGLIDADG motif. Enzymes that contain a single copy of this motif, such as I-CreI (Thompson et al. 1992; Wang et al. 1997) and I-CeuI (Turmel et al. 1997), act as homodimers and recognize consensus DNA target sites that are constrained to palindromic or near-palindromic symmetry. Enzymes that have two copies of the LAGLIDADG motif (such as I-SceI, the first LAGLIDADG enzyme to be discovered) act as monomers, possess a pair of structurally similar nuclease domains on a single peptide chain, and are not constrained to symmetric DNA targets (Agaard et al. 1997; Dalgaard et al. 1997; Lucas et al. 2001). In both subfamilies, the LAGLIDADG motif residues play both structural and catalytic roles (see below).

Free-standing LAGLIDADG endonucleases (i.e., those that are not covalently associated with intein domains) recognize DNA sites that typically range from 18 to 22 base pairs. They cleave both DNA strands across the minor groove, to generate mutually cohesive four base 3′ overhangs (Chevalier and Stoddard 2001). Like most, if not all nucleases, LAGLIDADG homing endonucleases require divalent cations for activity.

Upon invasion of a novel biological target site, homing endonucleases and their associated mobile introns or inteins can persist, diversify and spread to similar sites of related hosts. The evolution of related homing endonucleases subsequent to a founding intron invasion event has been elegantly described for at least one LAGLIDADG endonuclease branch, which contains the I-CreI enzyme (Lemieux et al. 1988; Turmel et al. 1995; Chevalier et al. 2003). (A similar study has more recently been reported for the intein-associated PI-SceI lineage; Posey et al. 2004.) I-CreI is encoded within a group I intron present in the chloroplast large subunit (LSU) rDNA of the green alga *Chlamydomonas reinhardtii*; the insertion site of this intron corresponds to position 2593 in the *Escherichia* coli 23S rDNA (Turmel et al. 1995). Sequence analysis of chloroplast and mitochondrial LSU rDNAs from numerous other green algae have disclosed 15 similar open-reading frames (ORFs) within identically positioned introns. Three of these genes were shown to encode active endonucleases that are isoschizomers of I-CreI, including I-MsoI from *Monomastix* (Lucas et al. 2001). Although the native target sites of I-CreI and I-MsoI differ at 2 out of 22 base pair positions, each endonuclease efficiently cleaves both target sites. Threading the I-MsoI sequence onto the I-CreI structure suggests significant protein sequence divergence, especially at residues involved in DNA binding (Lucas et al. 2001); this observation has been confirmed by an X-ray crystal structure of I-MsoI (Chevalier et al. 2003). The structure also implies that the I-MsoI enzyme might recognize a larger number of sites (i.e. is more promiscuous in its DNA recognition profile) than does I-CreI.

2 Structures of LAGLIDADG Homing Endonucleases

The structures of six LAGLIDADG enzymes bound to their DNA targets have been determined. These include two isoschizomeric homodimers (I-CreI: Heath et al. 1997; Jurica et al. 1998; Chevalier et al. 2001, 2003 and I-MsoI: Chevalier et al. 2003), which are both encoded within group I introns in the 23S rDNA of the green algae *Chlamydomonas reinhardtii* and *Monomastix*; two pseudosymmetric monomers (I-AniI: Bolduc et al. 2003 and I-SceI: Moure et al. 2003), which are encoded in mitochondrial introns of the fungi *Aspergillus nidulans* and *Saccharomyces cerevisiae*; one artificially engineered chimera (H-DreI: Chevalier et al. 2002, which is composed of a domain of the monomeric archaeal enzyme I-DmoI fused to a subunit of I-CreI); and an intein-associated endonuclease from yeast (PI-SceI: Moure et al. 2002). Structures of two additional enzymes have also been determined in the absence of DNA: the archaeal intron-encoded I-DmoI (encoded within an intron in the 23S rRNA gene of *Desulfurococcus mobilis*; Silva et al. 1999), and the archaeal intein-encoded PI-PfuI (found in the ribonucleotide reductase gene of *Pyrococcus furiosus*; Ichiyanagi et al. 2000). These crystallographic structures illustrate the structural and functional significance of the LAGLIDADG motif, the mechanism of DNA recognition and binding, and the structure and likely mechanism of their active sites.

LAGLIDADG enzyme domains form an elongated protein fold that consists of a core fold with mixed α/β topology (α-β-α-β-α). The overall shape of this domain is a half-cylindrical "saddle" that averages approximately 25×25×35 Å, with the longest dimension along a groove formed by the underside of the saddle. The surface of the groove is formed by an antiparallel, four-stranded β-sheet that presents a large number of exposed basic and polar residues for DNA contacts and binding. Each individual β-strand crosses the groove axis at an angle of ~45° and displays a continuous N- to C-terminal bend. The length of the core protein domain is often increased by extended loops connecting the β-strands at the periphery of the β-sheet structure. The β-sheets are stabilized by hydrophobic packing between the tops of the sheets and the α-helices of the core enzyme fold.

In the case of homodimeric enzymes, the full endonuclease structure is generated by a two-fold symmetry axis located at the N-termini of the individual subunits. For monomeric LAGLIDADG enzymes, a pseudo dyad symmetry axis at the same position arranges individual domains from a single peptide chain into similar relative positions (Fig. 1). For the monomeric enzymes the C- and N-terminal helices of the two related core domains (Dalgaard et al. 1997) are connected by flexible linker peptides with lengths between 3 residues to over 100 residues. In either enzyme subfamily, the complete DNA-binding surfaces of the full-length enzymes are 70–85 Å long, and thus can accommodate DNA targets of up to 24 base pairs.

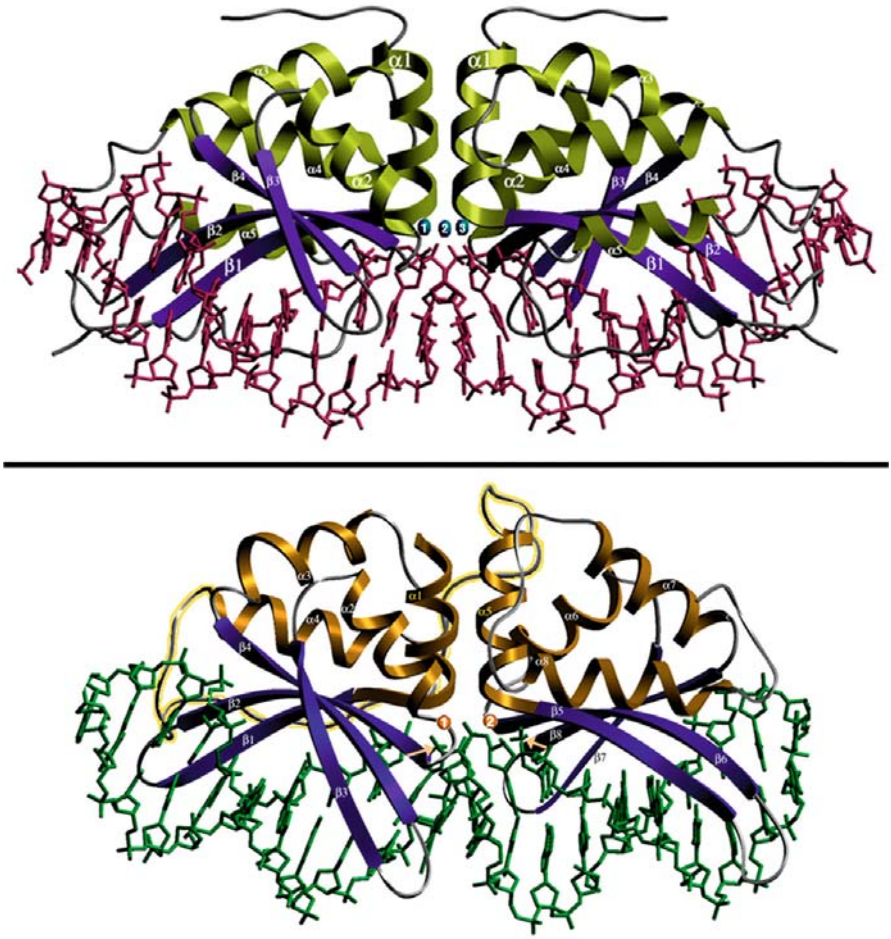

Fig. 1. Ribbon diagrams of homodimeric I-CreI (*top*) and asymmetric, monomeric I-AniI (*bottom*) endonucleases. In the latter, the core α-ββ-α-ββ-α domain fold is duplicated within the single polypeptide chain, and a long flexible linker (*highlighted in yellow*) connects their N- and C-termini, respectively. In the two structures, a perfect dyad symmetry axis, or a pseudo-symmetry axis, extends vertically in the plane of the page between the central two helices at the domain interface. The DNA target of I-CreI is 22 base pairs long and is a pseudo-palindrome; the DNA target of I-AniI is 19 base pairs long and asymmetric. Both enzymes use a four-stranded, antiparallel β-sheet to contact individual base pairs in the major groove of each DNA half-site. The DNA is in a very similar, slightly bent conformation in both structures. In I-CreI, there are three bound metal ions visible in the presence of manganese or magnesium. In the I-AniI structure, only two bound metal ions are visible, as discussed in the text.

The LAGLIDADG motif plays three distinct, but interrelated, roles in the structure and function of this enzyme family (Fig. 2). The first seven amino acid residues of each conserved motif form the last two turns of the N-terminal helices in each folded domain, which are packed against one another. Individual side chains from these helices participate either in core packing within individual domains or in contacts across the interdomain interface. The final three conserved residues (typically a Gly-Asp/Glu-Gly sequence) facilitate a tight turn from the N-terminal α-helix into the first β-strand of each DNA-binding surface. The conserved acidic residues of these sequences are positioned in the active sites and bind divalent cations that are essential for catalytic activity.

The structure and packing of the parallel, two-helix bundle in the domain interface of the LAGLIDADG enzymes are strongly conserved among the otherwise highly diverged members of this enzyme family. Helix packing at this interface is not mediated by a classic "ridges into grooves" strategy, but rather by small residues such as glycine and alanine that allow van der Waals contacts between backbone atoms along the helix–helix interface. The first two glycine and/or alanine residues in the LAGLIDADG motif participate directly in the dimer interface and allow tight packing of the helices. The close packing of the interface helices in these enzymes reflects the need to pack two symmetry-related endonuclease active sites less than 10 Å apart, to facilitate cleavage of homing site DNA across the narrow minor groove.

Despite little primary sequence homology among the LAGLIDADG homing endonucleases outside of the motif itself, the topologies of the endonuclease domains of the enzymes visualized to date, and the shape of their DNA-

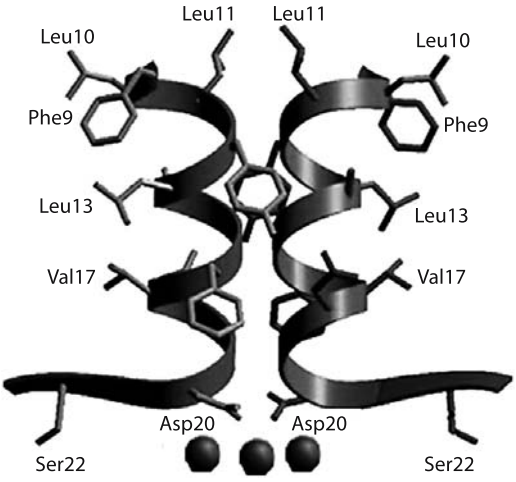

Fig. 2. The LAGLIDADG helices at the domain interface of I-CreI. Other enzymes in the family that have been visualized crystallographically have very similar motifs and structures. Note that a series of hydrophobic residues (Phe 9, Leu 10, 11 and 13, and Val 17) pack into the core of the individual enzyme domains, while other aromatic residues and small residues are involved in packing around and between the helices, respectively. The conserved acidic residues (Asp 20 and 20′) are contributed to the active sites where they participate in metal binding

bound β-sheets, are remarkably similar. A structural alignment of endonuclease domains and subunits in their DNA-bound conformation indicates that the structure of the central core of the β-sheets is well conserved (Bolduc et al. 2003). At least 12 Cα positions within these β-sheets are in close juxtaposition and have a Cα root-mean-square deviation (RMSD) of approximately 1 Å. These positions correspond to residues that make contacts to base pairs ±1 to 6 in each DNA half-site (see below). The conformations of the more distant ends of the β-strands and connecting turns are more poorly conserved, displaying RMSD values of over 3 Å for DNA-contacting residues. Similar alignments of intein-associated endonuclease domains indicate a more diverged structure of the β-sheet motifs.

In contrast to the LAGLIDADG enzymes, which contain a relatively compact structure in which DNA-binding and catalytic activities are intimately connected, the HNH and GIY-YIG homing endonuclease families have been shown by sequence analyses and by structural comparisons to display bipartite structures with separable catalytic and DNA-binding domains (Dalgaard et al. 1997; Derbyshire et al. 1997; VanRoey et al. 2001, 2002; Sitbon and Pietrokovski 2003; Shen et al. 2004). These enzymes often share common DNA-binding domain structures, which may indicate a common ancestral origin for a useful and reuseable binding domain. For example, both the GIY-YIG enzyme I-TevI and the HNH enzyme I-HmuI share a common helix-turn-helix motif at their C-termini that is critical for DNA recognition and binding (VanRoey et al. 2001; Shen et al. 2004). This pattern of swapping structural domains (which are usually part of tandemly arranged functional regions) is generally not observed for the LAGLIDADG family. However, recent analyses of homing endonuclease sequence alignments indicate that, in rare cases, the core fold of LAGLIDADG enzymes can be tethered to additional functional domains involved in DNA binding, usually termed NUMODS (nuclease-associated modular DNA-binding domains; Sitbon and Pietrokovski 2003). For example, a single copy of a canonical NUMOD1 region is found downstream (C-terminal) from the LAGLIDADG core of the intron-associated gene product of ORF Q0255 in yeast. This motif is similar to a conserved region of the bacterial sigma54-activator DNA-binding protein, and its C-terminal 15 amino acids are also similar to the N-terminal helix of typical helix-turn-helix (HTH) DNA-binding domains (Wintjens and Rooman 1996). In HTH domains, this helix is responsible for sequence-specific interactions with DNA.

3 Mechanisms of DNA Target Site Recognition and Specificity

As catalysts of the genetic mobility of introns and inteins, LAGLIDADG homing endonucleases (as well as other enzymes with the same biological func-

tion) must balance two somewhat contradictory requirements: they need to be highly sequence-specific, in order to promote precise intron transfer in their host genomes which are most often a chloroplast or mitochondrial genome, and yet must retain sufficient site recognition flexibility to allow successful lateral transfer in the face of sequence variation in genetically divergent hosts. LAGLIDADG homing endonucleases appear to solve these apparently contradictory problems by using a flexible homing site recognition strategy in which a well-defined, but limited, number of individual polymorphisms are tolerated by the enzyme without significant loss of binding affinity or cleavage efficiency. The biochemical basis of this flexible recognition strategy is to make phased, undersaturating DNA–protein contacts across long DNA target sites (Moure et al. 2002, 2003; Chevalier et al. 2003). The length of the interface provides overall high specificity, while formation of a broadly distributed set of phased, subsaturating contacts across the interface facilitates the recognition and accommodation of specific polymorphisms at individual target site positions (Fig. 3). The overall specificity of the LAGLIDADG endonucleases is not well established, but is generally thought to range from 1 in 10^8 to 10^9 random sequences for an average length of 20–22 base pairs (Chevalier et al. 2003).

In the protein–DNA interfaces visualized at high resolution (2.5–1.9 Å) for the LAGLIDADG family (I-CreI, I-MsoI, I-SceI and H-DreI), a set of four antiparallel β-strands in each enzyme domain provide direct and water-mediated contacts between residue side chains and nucleotide atoms in the major groove of each DNA half-site (Fig. 3). These contacts extend from base pairs ±3 to base pairs ±11 (the central four base pairs from –2 to +2, which are flanked by the scissile phosphate groups, are not in contact with the protein). Typically, strands β1 and β2 extend the entire length of this interface in each half-site, while strands β3 and β4 provide additional contacts to base pairs ±3, 4 and 5 in each complex. The LAGLIDADG endonucleases typically make contacts to approximately 65–75% of possible hydrogen-bond donors and acceptors of the base pairs in the major groove, make few or no additional contacts in the minor groove, and also contact approximately one-third of the backbone phosphate groups across the homing site sequence. These contacts are split evenly between direct and water-mediated interactions. A schematic of contacts formed by I-CreI to its pseudo-palindromic target site is shown in Fig. 3.

In the structures listed above, the DNA target is gradually bent around the endonuclease binding surface, giving an overall curvature across the entire length of the site of approximately 45°. In the homodimeric enzyme–DNA complexes with I-CreI and I-MsoI, the DNA is locally overwound between bases –3 to +3 (twist rising to ~50°), with a corresponding deformation in the base pair propeller twist and buckle angles for those same bases, leading to

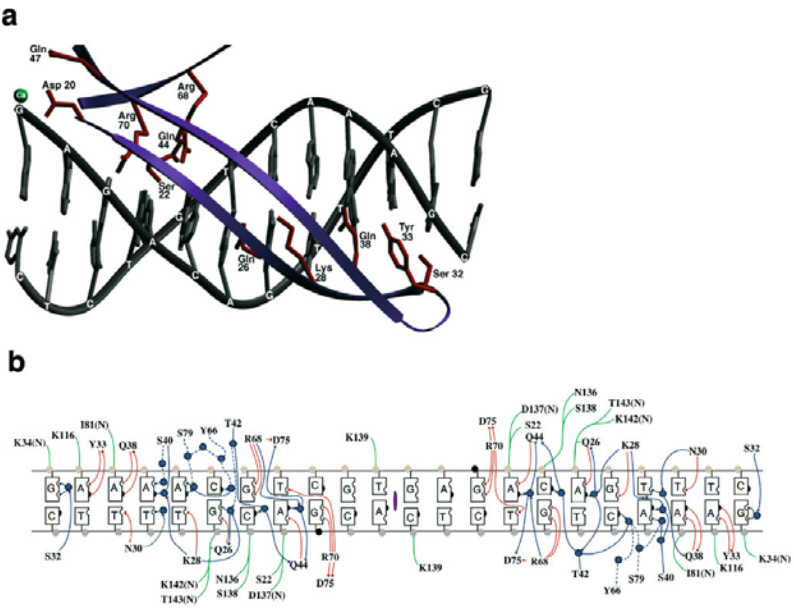

Fig. 3. Structural mechanism of DNA recognition by LAGLIDADG enzymes. A Structure of the β-sheet from a subunit of I-CreI in complex with its corresponding DNA target half-site. Note that every other side chain from a β-strand is pointed into the DNA major groove, and that the residues from adjacent β-strands are staggered in their positions to permit contact to several sequential bases. B Schematic of all observed contacts (both direct and water-mediated) between the I-CreI subunits and both DNA target half-sites, which differ in sequence at several positions (the full-length site is a pseudo-palindrome). The *blue circles* represent ordered water molecules. *Indentations* on bases represent H-bond acceptor groups; *bulges* on the bases represent H-bond donors. *Red lines* are direct contacts, *blue lines* are water-mediated, and *green lines* are contacts to backbone atoms of the DNA. *Dashed lines* represent 'double indirect' contacts to bases via two sequential bridging water molecules

narrowing of the minor groove at the site of DNA cleavage. The bending of the DNA is symmetric (Jurica et al. 1998). In the DNA complex with the monomeric enzymes, the central four base pairs of the cleavage sites generally display negative roll values, which translate into a similar narrowing of the minor groove. As a result, in all of these structures, the scissile phosphates are positioned approximately 5–8 Å apart and are located near bound metal ions in the active sites.

The distributions of related target site sequences that are recognized and cleaved by individual LAGLIDADG enzymes have been previously described using a variety of site preference screens (Argast et al. 1998; Gimble et al. 2003). In those experiments, target site variants that are recognized by the native enzyme are recovered from a randomized homing site library and se-

quenced. Using these data, the information content (specificity) at each base pair of the target site can be calculated using a computational method that accounts for the probability of each possible base being found at each position across the site (Schneider et al. 1986). The determination of crystallographic structures of the corresponding enzyme–DNA complexes, with the explicit visualization of direct and water-mediated contacts, facilitates an analysis of the correlation between the number and type of intermolecular contacts made to each base pair with the information content at each of these positions. Three general conclusions from these analyses are: (i) the specificity of base pair recognition to structurally unperturbed DNA sequence is proportional to the number of H-bond contacts to each base pair; (ii) the degree of specificity is not significantly attenuated by the use of solvent molecules as chemical bridges between nucleotide atoms and protein side chains; and (iii) information content is increased at individual base pairs, particularly near the center of the cleavage site, by indirect recognition of DNA conformational preferences.

4 Mechanism of DNA Cleavage

The kinetics and mechanism of catalysis have been particularly well studied for the I-CreI enzyme (Chevalier et al. 2004); many of these results appear to be generalizable to the LAGLIDADG family. The measured single-turnover kinetic rate constants, k^*_{max} and K_m^*, of the wild-type I-CreI 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 $nM^{-1}min^{-1}$. This enzyme and its relatives are all dependent on divalent cations for activity, similar to most if not all known endonucleases. A wide variety of divalent metal ions have been assayed for cleavage activity with I-CreI and display a wide range of effects (Chevalier et al. 2004). Two 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 tested. The use of manganese in place of magnesium allows recognition and cleavage of a broader repertoire of DNA target sequences than is observed with magnesium, as is seen for a variety of endonuclease catalysts such as restriction enzymes.

The structures of the four endonuclease–DNA complexes that have been solved at relatively high resolution (I-CreI, I-MsoI, I-SceI and H-DreI) all indicate the presence of three bound divalent metal ions coordinated by a pair of overlapping active sites, with one shared metal participating in both cleavage reactions by virtue of interacting with the scissile phosphates and 3′ hydroxyl leaving groups on both DNA strands. The structures of these four enzymes 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 (Fig. 4). Whether this

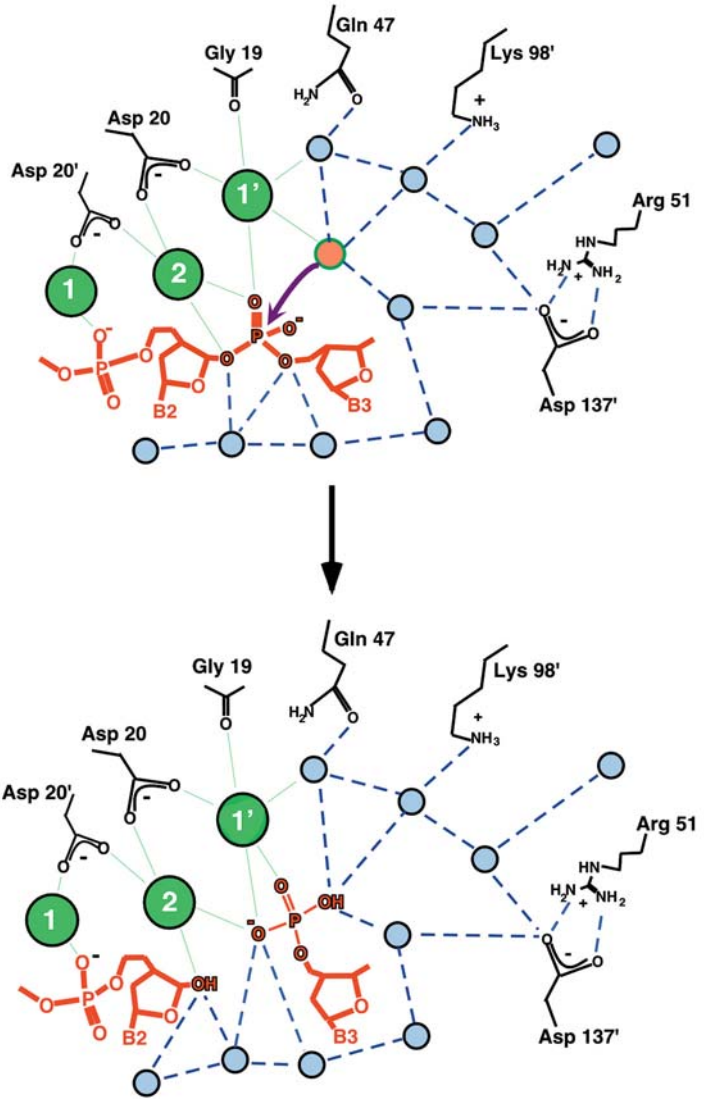

Fig. 4. Proposed mechanism of DNA hydrolysis for the I-CreI homing endonuclease. Other LAGLIDADG enzymes are also thought to follow a canonical two-metal phosphoryl hydrolysis pathway, but with significant variation in the positions and/or roles of basic residues and ordered water molecules. Those residues shown are all known to be essential or extremely important for DNA cleavage by I-CreI. Other LAGLIDADG enzymes display significant divergence at all positions except for the direct metal-binding residues (Asp 20 and 20′) from the LAGLIDADG motifs (Table 1)

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-SceI–DNA complex (Moure et al. 2003) clearly demonstrates that DNA cleavage must involve sequential cleavage of coding and non-coding DNA strands, with a significant conformational rearrangement of the active sites relative to DNA occurring between the two reactions.

In contrast, the structures of DNA complexes of one monomeric enzyme (I-AniI; Bolduc et al. 2003) and of the intein-associated PI-SceI, solved at lower resolution (~3 Å; Moure et al. 2002), have thus far revealed the presence of only two bound metal ions; a central, shared metal ion is not visible. It is unclear whether this reflects a significant difference in catalytic mechanism, reduced occupancy or poor structural ordering of the central metal ion, or simply a limitation of lower resolution crystallographic data.

In the high-resolution structures listed above, a single independently bound metal in each of the two endonuclease active sites coordinates a directly ligated water molecule, which is appropriately positioned for an in-line hydrolytic attack on a scissile phosphate group. The third, 'shared' central metal ion stabilizes the transition state phosphoanion and the 3′ hydroxylate leaving group for both strand cleavage events (Chevalier et al. 2001). In I-CreI, the central metal is jointly coordinated by one conserved acidic residue from each LAGLIDA**D**G motif and by oxygen atoms from the scissile phosphates of each DNA strand. The unshared metals in each individual active site are also coordinated by a single LAGLIDADG carboxylate oxygen, as well as a non-bridging DNA oxygen atom and a well-ordered coordination shell of water molecules. One of the metal-bound water molecules in the active site is often in contact with a catalytically essential glutamine or asparagine residue. In addition to the attacking water molecule a well-ordered network of water molecules is distributed in a large pocket surrounding the DNA scissile phosphate group. These ordered solvent molecules extend from the metal-bound nucleophile to the leaving group 3′ oxygen and are themselves positioned or coordinated by several basic residues that line the solvent pocket.

At physiological pH, phosphate ester bonds have large barriers to cleavage even though they are thermodynamically unstable (Westheimer 1987). To efficiently catalyze the cleavage of phosphate esters, several chemical features are required, including a nucleophile, a basic moiety to activate and position that 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 (Galburt and Stoddard 2002). 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 endonucleases (and other phosphoryl transfer enzymes) is the

use of bound metal ions as cofactors, and a basic residue (such as a lysine) that directly activates the water molecule for nucleophilic attack.

The metal-dependent features of DNA hydrolysis described above are clearly imparted in the LAGLIDADG endonucleases by the conserved acidic residues of their namesake sequence motif, which directly coordinate divalent cations. However, the remaining residues in the active site are remarkable for their chemical and structural diversity (Chevalier and Stoddard 2001; Table 1). In fact, no enzyme in this family has an essential residue that has been unambiguously identified as a general base for activation of a water nucleophile. Indeed, these enzymes are unique compared to other hydrolytic endonucleases in that the basic residues in their active sites are not generally found in contact distance with metal-bound waters. Catalytically important basic residues, such as Lys 98 in I-CreI, which are 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 reactions 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

Table 1. Summary of conserved motif and active site residues for LAGLIDADG homing endonuclease structures

Enzyme	LAGLIDADG	Metal Binding		Basic Pocket	
I-CreI	LAGFVDGDG	D20	Q47	K98	R51
I-MsoI	IAGFLDGDG	D21	Q49	K104	K54
I-DmoI	LLGLIIGDG	D21	Q42	K120	K43
	IKGLYVAEG	E117	N129	–	K130
I-AniI	LVGLFEGDG	D15	L36	K94	D40
	LVGFIEAEG	E148	Q171	K227	G174
I-SceI	GIGLILGDA	D44	E61	K122	–
	LAYWFMDDG	D145	N192	K223	–
PI-SceI	LLGLWIGDG	D218	D229	K301	R231
	LAGLIDSDG	D326	T341	K403	H343
PI-PfuI	LAGFIAGDG	D149	D173	L220	–
	IAGLFDAEG	E250	M263	K322	–

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.

References

Agaard C, Awayez MJ, Garrett RA (1997) Profile of the DNA recognition site of the archaeal homing endonuclease I-DmoI. Nucleic Acids Res 25:1523–1530

Argast GM, Stephens KM, Emond MJ, Monnat RJ (1998) I-PpoI and I-CreI homing site sequence degeneracy determined by random mutagenesis and sequential *in vitro* enrichment. J Mol Biol 280:345–353

Belfort M, Roberts RJ (1997) Homing endonucleases – keeping the house in order. Nucleic Acids Res 25:3379–3388

Belfort M, Reaban ME, Coetzee T, Dalgaard JZ (1995) Prokaryotic introns and inteins: a panoply of form and function. J Bacteriol 177:3897–3903

Bolduc JM, Spiegel PC, Chatterjee P, Brady KL, Downing ME, Caprara MG, Waring RB, Stoddard BL (2003) Structural and biochemical analyses of DNA and RNA binding by a bifunctional homing endonuclease and group I intron splicing cofactor. Genes Dev 17:2875–2888

Chevalier BS, Stoddard BL (2001) Homing endonucleases: structural and functional insight into the catalysts of intron/intein mobility (review; 144 refs). Nucleic Acids Res 29:3757–3774

Chevalier BS, Monnat RJ Jr, Stoddard BL (2001) The homing endonuclease I-CreI uses three metals, one of which is shared between the two active sites (see comments). Nature Struct Biol 8:312–316

Chevalier BS, Kortemme T, Chadsey MS, Baker D, Monnat RJ Jr, Stoddard BL (2002) Design, activity and structure of a highly specific artificial endonuclease. Mol Cell 10:895–905

Chevalier B, Turmel M, Lemieux C, Monnat RJ Jr, Stoddard BL (2003) Flexible DNA target site recognition by divergent homing endonuclease isoschizomers I-CreI and I-MsoI. J Mol Biol 329:253–269

Chevalier B, Sussman D, Otis C, Noel A-J, Turmel M, Lemieux C, Stephens K, Monnat RJ Jr, Stoddard BL (2004) Metal-dependent DNA cleavage mechanism of the I-CreI LAGLIDADG homing endonuclease. Biochemistry 43:14015–14026

Dalgaard JZ, Klar AJ, Moser MJ, Holley WR, Chatterjee A, Mian IS (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

Delahodde A, Goguel V, Becam AM, Creusot F, Perea J, Banroques J, Jacq C(1989) Site-specific DNA endonuclease and RNA maturase activities of two homologous intron-encoded proteins from yeast mitochondria. Cell 56:431–441

Derbyshire V, Kowalski JC, Dansereau JT, Hauer CR, Belfort M(1997) Two-domain structure of the td intron-encoded endonuclease I-TevI correlates with the two-domain configuration of the homing site. J Mol Biol 265:494–506

Dujon B (1980) Sequence of the intron and flanking exons of the mitochondrial 21S rRNA gene of yeast strains having different alleles at the omega and rib-1 loci. Cell 20:185–197

Dujon B (1989) Group I introns as mobile genetic elements: facts and mechanistic speculations – a review. Gene 82:91–114

Dujon B, Belfort M, Butow RA, Jacq C, Lemieux C, Perlman PS, Vogt VM (1989) Mobile introns: definition of terms and recommended nomenclature. Gene 82:115–118

Galburt E, Stoddard BL (2002) Catalytic mechanisms of restriction and homing endonucleases. Biochemistry 41:13851–13860

Geese WJ, Waring RB (2001) A comprehensive characterization of a group IB intron and its encoded maturase reveals that protein-assisted splicing requires an almost intact intron RNA. J Mol Biol 308:609–622

Gimble FS, Moure CM, Posey KL (2003) Assessing the plasticity of DNA target site recognition of the PI-SceI homing endonuclease using a bacterial two-hybrid selection system. J Mol Biol 334:993–1008

Heath PJ, Stephens KM, Monnat RJ, Stoddard BL (1997) The structure of I-CreI, a group I intron-encoded homing endonuclease. Nat Struct Biol 4:468–476

Ichiyanagi K, Ishino Y, Ariyoshi M, Komori K, Morikawa K (2000) Crystal structure of an archaeal intein-encoded homing endonuclease PI-PfuI. J Mol Biol 300:889–901

Jacquier A, 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

Jin Y, Binkowski G, Simon LD, Norris D (1997) Ho endonuclease cleaves MAT DNA in vitro by an inefficient stoichiometric reaction mechanism. J Biol Chem 272:7352–7359

Jurica MS, Monnat RJ Jr, Stoddard BL (1998) DNA recognition and cleavage by the LAGLIDADG homing endonuclease I-CreI. Mol Cell 2:469–476

Lazowska J, Jacq C, Slonimski PP (1980) Sequence of introns and flanking exons in wild-type and box3 mutants of cytochrome b reveals an interlaced splicing protein coded by an intron. Cell 22:333–348

Lazowska J, Claisse M, Gargouri A, Kotylak Z, Spyridakis A, Slonimski PP (1989) Protein encoded by the third intron of cytochrome b gene in *Saccharomyces cerevisiae* is an mRNA maturase. Analysis of mitochondrial mutants, RNA transcripts proteins and evolutionary relationships. J Mol Biol 205:275–289

Lemieux B, Turmel M, Lemieux C (1988) Unidirectional gene conversions in the chloroplast of *Chlamydomonas* inter-specific hybrids. Mol Gen Genet 212:48–55

Lucas P, Otis C, Mercier JP, Turmel M, Lemieux C (2001) Rapid evolution of the DNA-binding site in LAGLIDADG homing endonucleases. Nucleic Acids Res 29:960–969

Moure CM, Gimble FS, Quiocho FA (2002) Crystal structure of the intein homing endonuclease PI-SceI bound to its recognition sequence. Nature Struct Biol 9:764–770

Moure CM, Gimble FS, Quiocho FA (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

Posey KL, Koufopanou V, Burt A, Gimble FS (2004) Evolution of divergent DNA recognition specificities in VDE homing endonucleases from two yeast species. Nucleic Acids Res 32:3947–3956

Schafer B, Wilde B, Massardo DR, Manna F, del Giudice L, Wolf K (1994) A mitochondrial group-I intron in fission yeast encodes a maturase and is mobile in crosses. Curr Genet 25:336–341

Schneider TD, Stormo GD, Gold L, Ehrenfeucht A (1986) Information content of binding sites on nucleotide sequences. J Mol Biol 188:415–431

Shen BW, Landthaler M, Shub DA, Stoddard BL (2004) DNA binding and cleavage by the HNH homing endonuclease I-HmuI. J Mol Biol 342:43–56

Silva GH, Dalgaard JZ, Belfort M, Roey PV (1999) Crystal structure of the thermostable archaeal intron-encoded endonuclease I-DmoI. J Mol Biol 286:1123–1136

Sitbon E, Pietrokovski S (2003) New types of conserved sequence domains in DNA-binding regions of homing endonucleases. Trends Biochem Sci 28:473–477

Thompson AJ, Yuan X, Kudlicki W, Herrin DL (1992) Cleavage and recognition pattern of a double-strand-specific endonuclease (I-CreI) encoded by the chloroplast 23S rRNA intron of *Chlamydomonas reinhardtii*. Gene 119:247–251

Turmel M, Cote V, Otis C, Mercier JP, Gray MW, Lonergan KM, Lemieux C (1995) Evolutionary transfer of ORF-containing group I introns between different subcellular compartments (chloroplast and mitochondrion) Mol Biol Evol 12:533–545

Turmel M, Otis C, Cote V, Lemieux C (1997) Evolutionarily conserved and functionally important residues in the I-CeuI homing endonuclease. Nucleic Acids Res 25:2610–2619

VanRoey P, Waddling CA, Fox KM, Belfort M, Derbyshire V (2001) Intertwined structure of the DNA-binding domain of intron endonuclease I-TevI with its substrate. EMBO J 20:3631–3637

VanRoey P, Meehan L, Kowalski JC, Belfort M, Derbyshire V (2002) Catalytic domain structure and hypothesis for function of GIY-YIG intron endonuclease I-TevI. Nat Struct Biol 9:806–811

Wang J, Kim H-H, Yuan X, Herrin DL (1997) Purification, biochemical characterization and protein-DNA interactions of the I-CreI endonuclease produced in *Escherichia coli*. Nucleic Acids Res 25:3767–3776

Westheimer FH (1987) Why nature chose phosphates. Science 235:1173–1178

Wintjens R, Rooman M (1996) Structural classification of HTH DNA-binding domains and protein-DNA interaction modes. J Mol Biol 262:294–313