Molecular Cell, Vol. 2, 469–476, October, 1998, Copyright ©1998 by Cell Press

DNA Recognition and Cleavage by the LAGLIDADG Homing Endonuclease I-*Cre*I

Melissa S. Jurica,* Raymond J. Monnat, Jr.,† and Barry L. Stoddard*‡

*Fred Hutchinson Cancer Research Center and the Graduate Program in Molecular and Cell Biology
University of Washington
Seattle, Washington 98109
†Department of Pathology
University of Washington
Seattle, Washington 98195

Summary

The structure of the LAGLIDADG intron-encoded homing endonuclease I-*Cre*I bound to homing site DNA has been determined. The interface is formed by an extended, concave β sheet from each enzyme monomer that contacts each DNA half-site, resulting in direct side-chain contacts to 18 of the 24 base pairs across the full-length homing site. The structure indicates that I-*Cre*I is optimized to its role in genetic transposition by exhibiting long site-recognition while being able to cleave many closely related target sequences. DNA cleavage is mediated by a compact pair of active sites in the I-*Cre*I homodimer, each of which contains a separate bound divalent cation.

Introduction

The homing endonucleases (Lambowitz and Belfort, 1993; Mueller et al., 1993; Belfort and Perlman, 1995; Belfort et al., 1995; Belfort and Roberts, 1997) are a diverse family of proteins encoded by open reading frames in genetically mobile, self-splicing introns (Figure 1). Similar endonucleases have also been identified as optional, independently folded domains in self-splicing protein introns, termed "inteins" (Pietrokovski, 1994; Cooper and Stevens, 1995; Belfort and Roberts, 1997). Homing endonucleases have been identified in a diverse collection of unicellular eukaryotes, archea, and eubacteria, with the largest number identified in organellar genomes. These proteins display the ability to recognize and cleave long DNA target or homing sites of 15–40 bp in a homologous allele that lacks the intron or intein sequence, and they promote the lateral transfer of their encoding intron or intein to these sites by a targeted transposition mechanism termed "homing" (Belfort et al., 1995).

In the case of a group I intron and intein sequences, the mobile intervening sequence, including the endonuclease ORF, is subsequently transferred to the cleaved recipient allele by double-strand break repair/gene conversion. Mobile group II introns, in contrast, use a combination of endonucleolytic cleavage and a target DNA primed reverse transcription mechanism for lateral transfer (Yang et al., 1996). Group I introns self-splice at the RNA level by folding into a specific multidomain structure with an active site that catalyzes the cleavage ligation reaction. Open reading frames can be inserted into loop regions peripheral to this catalytic core. The protein I-*Cre*I is a homing endonuclease encoded by an ORF contained within a group I intron in the *Chlamydomonas reinardtii* chloroplast 23S rRNA gene. I-*Cre*I is a member of the largest family of homing endonucleases in which each member contains one or two copies of a conserved LAGLIDADG or dodecapeptide protein motif. Proteins containing this motif include mobile intron endonucleases, self-splicing inteins (Belfort and Perlman, 1995; Cooper and Stevens, 1995), mitochondrial RNA maturases (Waring et al., 1982; Hensgens et al., 1983), and free-standing nuclear-encoded proteins such as the yeast HO mating type switch endonuclease (Kostriken et al., 1983). All of the LAGLIDADG homing endonucleases cleave their homing site DNA to generate 3′ extended cohesive ends, usually four bases in length (Belfort and Roberts, 1997). No structures of this large family of proteins have been determined complexed to a nucleic acid target.

The homing site recognized and cleaved by I-*Cre*I is 22 bp in length and is a degenerate palindrome (Figure 2). Genetic analysis has indicated that any single nucleotide base pair in this site may be altered while preserving cleavage activity and that additional changes may also be accommodated, depending on the specific sites of mutation (Argast, 1998). The positions in the homing site that differ between half-sites (i.e., that are not palindromically conserved) are less sensitive to mutation in these screens, as expected for a homodimeric endonuclease. This pattern of site specificity appears to be quite beneficial for an enzyme that is responsible for a simple transposition event by maintaining high enough specificity to minimize nonspecific cleavage in the target genome while also allowing the enzyme to target closely related variants of its normal homing site. Such a strategy might minimize host toxicity while increasing intron mobility between host strains.

Recently determined X-ray crystal structures of the intron-encoded I-*Cre*I homing endonuclease from *Chlamydomonas reinhardtii* (Heath et al., 1997) and of the intein-encoded endonuclease PI-*Sce*I from yeast (Duan et al., 1997) indicate that the LAGLIDADG motif plays a role in protein folding, in dimerization or interdomain packing, and in catalysis. In the structure of I-*Cre*I, much of the LAGLIDADG motif forms an α helix that packs against its symmetry-related partner. This symmetric interface is conserved in the endonuclease domain of the intein protein PI-*Sce*I. This monomeric enzyme contains two copies of the LAGLIDADG motif that pack α helices together in the same manner as I-*Cre*I. Both enzymes exhibit a pronounced groove of dimensions that are complementary to double-stranded DNA. An analysis of these structures and their cleavage pattern leads to the prediction that the LAGLIDADG endonucleases would form most of their sequence-specific contacts in the major groove to either side of the center of the homing site and cleave homing site DNA across the minor groove.

‡To whom correspondence should be addressed (e-mail: bstoddar@fred.fhcrc.org).

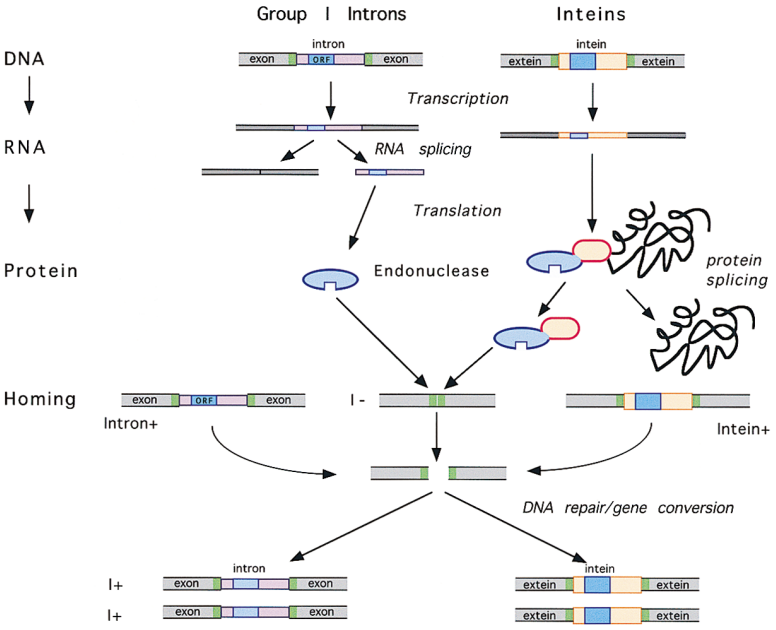

Figure 1. The Homing Endonuclease Homing Cycle

Intron mobility or homing is targeted and initiated by an endonuclease making a site-specific DNA double-strand break at a homing site (green region) in an intronless allele (I−) of the host gene. Double-strand break repair transfers the mobile intron from an intron-containing allele (I+), which includes the open reading frame (ORF) encoding the endonuclease activity. In the case of group I introns, the invading intron reconstitutes the intact host gene at the RNA level by self-splicing out the intron sequence. In contrast, inteins work at the protein level to splice out the invading protein and ligate the host protein components together.

Results

Overall Structure and Homing Site Recognition

The refined I-*Cre*I/DNA complex structure was determined to 3 Å resolution (Table 1). The structure (Figure 3) consists of a nearly full-length enzyme homodimer (ten C-terminal residues are disordered in each subunit), a 24–base pair DNA duplex corresponding to the sequence of the native homing site in *Chlamydomonas* chloroplast rDNA, and two bound calcium ions that support DNA binding but prevent cleavage (Wang et al., 1997). The entire complex is approximately 75 Å long and buries 3400 Å^2 of surface area each from the endonuclease and DNA in the molecular interface. The protein monomer of I-*Cre*I is a single domain of mixed α/β topology. Four α helices pack against one side of an extended β ribbon created by four antiparallel β strands. The other side of the β ribbon forms the DNA binding surface. The homing site DNA is slightly bent around this endonuclease binding surface. The LAGLIDADG sequence is positioned at the dimer interface. The first seven residues (13–19) form an α helix (α1) that is perpendicular to the β ribbon and packs against the symmetry-related helix (α1′). This forms a parallel two-helix bundle that uses van der Waals contacts between small side chains and protein backbone atoms for very close packing of the interface. This close packing is very important for positioning the conserved aspartate residues (Asp-20 and Asp-20′) at the DNA interface juxtaposed with the two scissile phosphates across the minor groove of the DNA. The final LAGLIDADG residues create a tight turn from the α1 helix into the first β strand of the β ribbon structure. Each enzyme monomer places this long, concave, antiparallel β ribbon in the major groove of the homing half-sites, with the two longest strands each contributing about 8 residues to the interface (Figures 3 and 4). These β ribbons are part of a saddle that is approximately 18 Å in diameter and displays sufficient helical twist to maintain direct protein–DNA contacts across 9 base pairs of DNA in each homing half-site. The β ribbon architecture facilitates the recognition of extended DNA sequence motifs such as homing sites by maintaining

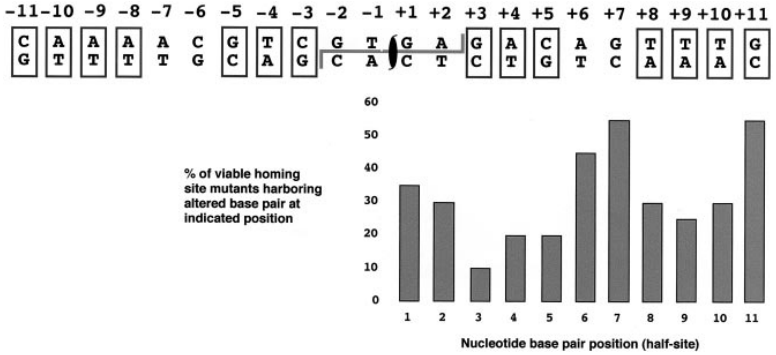

Figure 2. Homing Site Recognition by I-*Cre*I

The enzyme recognizes a degenerate palindrome 22 bp in length and cleaves across the minor groove to liberate cohesive 3′ ends of length 4 bases each. Top, the homing site found in the rDNA target gene. The positions that are conserved between half-sites are indicated by boxes. Bottom, frequency of mutations observed in homing site randomization studies that select active, cleavable DNA variants (data compiled from Argast, 1998). Individual mutations are tolerated at all positions; in many active mutants, 2–4 sites are altered. Note that mutational frequencies in these studies are generally lower at positions that are not conserved between half-sites in the rDNA gene, as expected for a homodimeric endonuclease. The sequence used for oligonucleotides in the protein/DNA crystallization (Experimental Procedures) is the wild-type homing site sequence but with a mutation at bp +11 that does not alter enzyme binding or activity.

Table 1. X-Ray Data and Refinement Statistics

Crystallographic Data	
Space group	$P2_12_12_1$
Cell parameters (Å)	a = 67.56, b = 89.26 c = 177.516
Crystalline form	Orthorhombic
Number of dimer/DNA complexes per AU	2
Resolution (Å)	3.0
Number of reflections measured	86360
Number of unique reflections	20643
Redundancy	4.2
Completeness (outer resolution bin 3.05–3.0)	96.6% (79.6%)
Average I/σ(I)	7.4
R_{sym} (outer resolution bin 3.05–3.0)	12.0% (29.7%)
Refinement	
Resolution range (Å)	100–3.0
R factor (%)	24.2
R-free (%)	28.0
Number of protein atoms	4948
Number of nucleic acid atoms	1960
Number of solvent molecules	8
Ramachandran plot (percent core, allowed, generous, disallowed)	87.1, 12.9, 0.0, 0.0
rms on bonds (Å)	0.008
rms on angles (°)	1.322
Average B (Å^2) (protein, DNA)	15.3, 10.7

a spacing of 6–7 Å between each accessible side chain on the protein surface. This closely parallels the spacing between alternating base pairs and phosphate groups in B-form DNA (Phillips, 1994). Flexibility of the β strands also allows contacts with nucleotide bases over a longer distance than could be achieved with a more rigid structural element such as an α helix. The large buried surface area created by the extensive molecular interface accounts for the high stability of the I-*Cre*I/DNA complex. This stability is illustrated by the fact that in vitro I-*Cre*I releases its DNA product only in the presence of SDS (Seligman et al., 1997; Wang et al., 1997).

Binding of the endonuclease induces small conformational changes to both the enzyme and the DNA substrate. Superposition of the bound and unbound enzyme structures reveals differences primarily in two loop regions. Residues 29–37, which connect strands β1 and β2 in the protein–DNA interface, participate in DNA binding and adopt a significantly altered conformation. This loop donates three side chains (Asn-30, Ser-32, and Tyr-33) to nucleotide contacts at the ends of the homing site and provides the distinctive "twist" of the β ribbon that allows it to maintain contact over a long DNA site. The last 16 visible residues in the protein–DNA complex (138–153) become ordered upon binding and make a number of nonspecific contacts to the DNA phosphate backbone. In addition, crystal contacts cause residues 113–123 (which are disordered in the unbound protein) to be clearly visible in the cocrystal structure. In contrast, residues of the N-terminal α helices at the dimer interface, including those in the LAGLIDADG motif, are virtually unchanged. The rms difference between the two enzyme structures is 2.1 Å. The bound DNA is slightly curved with a bend of less than 10° induced in the middle of each half-site near the position of the fifth base pair. These shallow perturbations of the homing site DNA

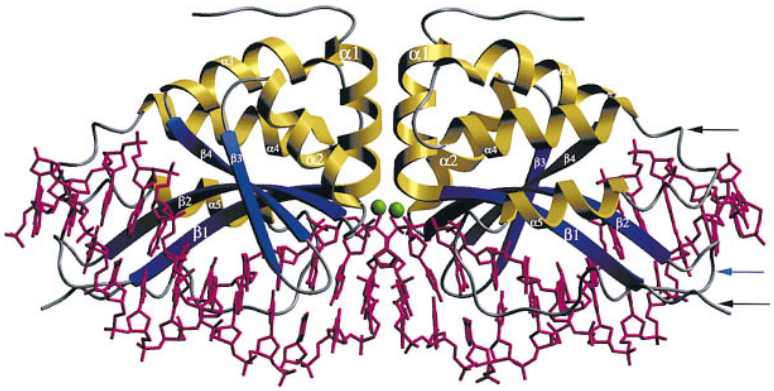

Figure 3. The Structure of I-*Cre*I Complex with a 24–Base Pair Synthetic DNA Homing Site

Each calcium ion (green spheres) is coordinated by a single aspartate residue (Asp-20) from the LAGLIDADG motif, by an oxygen from the scissile phosphate, and by another oxygen from the phosphate group directly across the minor groove from the cleavage site. The distance between the bound calcium ions is 7 Å. The loop connecting β1 and β2 in each subunit (blue arrow) displays the largest conformational change upon DNA binding. In addition, the loop connecting helices α4 and α5 (residues 113–123, top black arrow) and most of the C-terminal tail (residues 138–153, bottom black arrow) are ordered in this DNA complex structure. These residues were disordered in the structure of the apo enzyme. Figures 3 and 4 were made with the program SETOR (Evans, 1993).

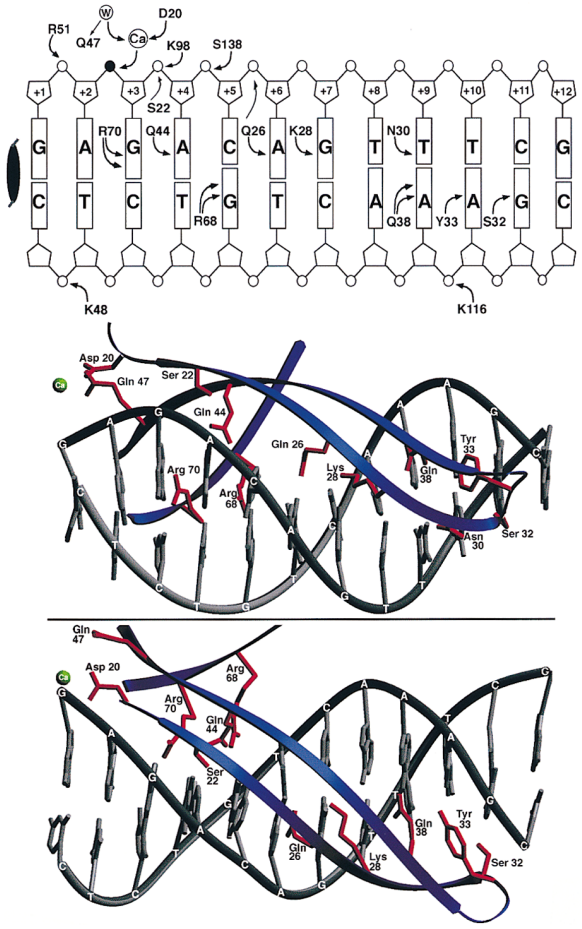

Figure 4. The Endonuclease–DNA Interface

Top, schematic diagram of direct contacts between the endonuclease and the homing site (contacts are shown for a protein monomer and DNA half-site). The two strands are numbered with + and − values as shown, and the scissile phosphate is shown in black. Water-mediated contacts are not shown. Bottom, two views of the contacts between nucleotide bases in a homing half-site and side chains from the antiparallel β ribbon formed by strands β1, β2, and β4 of a single protein subunit. The β3 strand also contributes side chains to the DNA interface, but none make direct contact with nucleotide bases, and it is therefore not shown to improve clarity.

appear to be driven by the interaction of the β groove of the enzyme and the complementary major groove of the DNA homing site. Watson-Crick base pairs are properly maintained throughout the homing site, although several exhibit measurable propeller twist. This is especially obvious near the cleavage site where the minor groove is narrowed as a result of bound enzyme and metal ions.

Residues from the DNA-binding surface of the protein contact nine consecutive base pairs (3 through 11) in each half-site (Figure 4). These contacts are made in the major groove and consist of 12 direct interactions to atoms of nucleotide bases and up to three additional water-mediated contacts. This is in general agreement with methylation protection analysis (Wang et al., 1997), although the contacts extend three base pairs beyond the region protected in those experiments. The protein contacts involve approximately 40% (15 of 36) of the potential hydrogen bond donors and acceptors in the major groove of each homing half-site. Three of the nucleotide base pairs (3, 5, and 9) exhibit more than one direct atomic interaction with the protein (Figure 4). These base pairs are among the more strongly conserved positions in the homing site in genetic randomization studies (Argast, 1998). The remaining homing site positions make single contacts to side-chain atoms, allowing flexibility at these positions in these same studies. This flexibility is also observed in methylation interference studies, where specific methylation of either of two bases contacted (−5 and +7) does not significantly interfere with protein binding (Wang et al., 1997). Each I-*Cre*I monomer also contacts six phosphate groups in each half-site. The base pairs in the center of the homing site (between the scissile phosphates) and the final base pair at each end of the homing site are not involved in protein contacts. This is in some contrast with the methylation interference data, which shows that methylation of the central +1 base reduces protein binding. There is no obvious structural explanation for this as there is room for a methyl group at the N7 position of this guanosine.

Endonuclease Active Site

In the cocrystal structure, two calcium ions (which substitute for activating cations such as Mg^{2+} or Mn^{2+} and act as inhibitors of the cleavage reaction [Wang et al., 1997]) are bound at the protein dimer interface. Each calcium ion is independently coordinated by a single aspartate residue (Asp-20 from the LAGLIDADG), a single oxygen from a scissile phosphate, and an additional oxygen atom from the phosphate located directly across the minor groove (Figures 4 and 5). The main-chain carbonyl of residue 19 from the opposing monomer and at least one water molecule complete the coordination sphere around the cations. This water molecule is loosely contacted by Gln-47, which is a catalytically important residue as shown by mutagenesis studies (Seligman et al., 1997). The direct oxygen contacts to the bound metal exhibit normal calcium bond distances of 2.2–2.5 Å, although the coordination is not strictly octahedral. There is no indication of a bound metal between the symmetry-related aspartate residues.

Two different side chains, Arg-51 and Lys-98, are located in the enzyme active site and are candidates to act as a Lewis acid (stabilizing the pentacoordinate transition state) or to activate a proton donor in the cleavage reaction. Mutations in each of these residues have been observed to sharply reduce I-*Cre*I endonucleolytic activity (Seligman et al., 1997). Arg-51 makes a strong contact to an adjacent phosphate in the cleavage site. Lys-98 is provided by the opposing protein monomer in the homodimer. Each of the side chains is within 6.5 Å of the scissile phosphate in this inactive complex. It is straightforward to surmise that one or more of these residues moves slightly, possibly in concert with nearby DNA backbone atoms, in the presence of an activating cation to contact the cleaved phosphate. The structural equivalent of the Lys-98 and -98′ side chains in the I-*Cre*I dimer is conserved in both domains of the monomeric PI-*Sce*I endonuclease (Lys-301 and Lys-403) (Duan et al.,

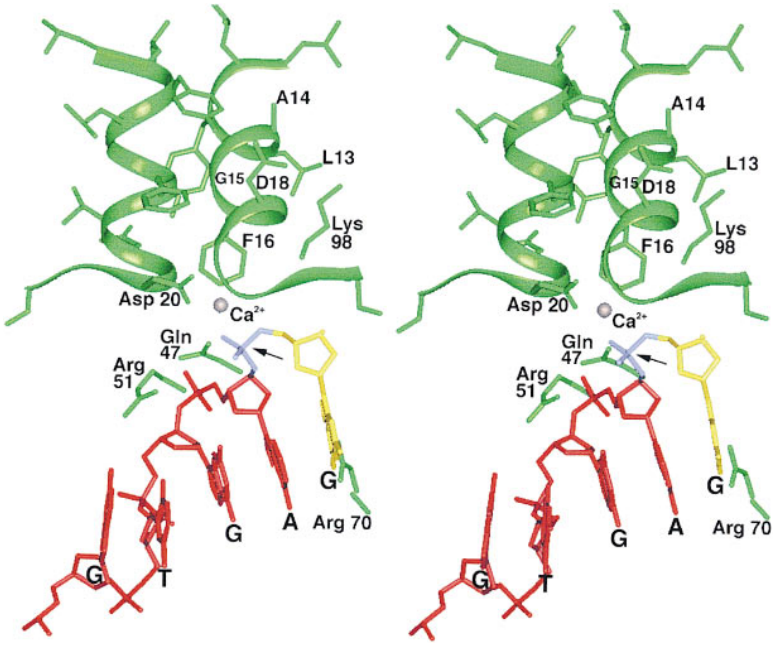

Figure 5. I-*Cre*I Homing Endonuclease Active Site

Atoms are shown for a single DNA strand in the half-site complex, the LAGLIDADG-containing protein helices, and the most likely catalytic side chains. The four nucleotide bases that form the single-stranded overhang after cleavage are colored red, the scissile phosphate group is blue, and the adjoining nucleotide base (Gua+3, which retains the cleaved phosphate) is yellow. The bound calcium ion is a gray sphere and is coordinated as described. The figure was made with the program QUANTA 96 (1996).

1997). Superposition of these two structures, using the LAGLIDADG helices as an anchor, positions these PI-*Sce*I lysine residues in virtually the same location as Lys-98 and -98′ in the I-*Cre*I/DNA complex.

Five additional residues, which when mutated abolish I-*Cre*I endonuclease activity (Seligman et al., 1997), are located in or near the enzyme active site. Arg-70 makes a strong contact to the guanosine nucleotide base directly 3′ of the cleaved phosphate and may play a role in generating binding energy for the reaction or destabilizing the substrate complex. Leu-39, Leu-91, and Asp-75 are near the active site and appear to play a role in protein packing and stabilization. Asp-75 is also adjacent to two side chains, Gln-44 and Arg-70, that contact bound DNA. Similar genetic and biochemical studies of I-*Ceu*I identified residues equivalent to Asp-20, Gln-47, and Arg-70 as sites of inactivating mutations (Turmel et al., 1997). Alanine scanning mutagenesis of PI-*Sce*I (He et al., 1998) identified Lys-301 (the equivalent of Lys-98 in I-*Cre*I) as a critical residue and implicated at least three other amino acids that are in or near the active site of I-*Cre*I after superposition: Asp-229, Arg-231, and His-343. In the superimposed structures, Asp-229 and Arg-231 of PI-*Sce*I correlate most closely to the positions of Arg-51 and Gln-50, respectively, in the I-*Cre*I structure. His-343 corresponds to the position of Arg-51′ in the I-*Cre*I symmetry-related subunit.

Discussion

Homing Site Recognition and Transposition

The structure of the I-*Cre*I enzyme–DNA complex is in excellent agreement with previous genetic and biochemical studies. These experiments indicate that homing site sequence of at least 20 base pairs is required to achieve a maximal binding affinity of 0.2 nM (Thompson et al., 1992; Durrenberger and Rochaix, 1993; Wang et al., 1997), that sequence-specific contacts are distributed across the entire length of the homing site, and that base pair substitutions can be tolerated at many different homing site positions without seriously disrupting homing site binding or cleavage (Argast, 1998). Approximately half of the base pairs are strongly conserved in functional homing site variants. The four central base pairs are comparatively insensitive to mutation in binding and cleavage assays.

In order to function most efficiently as a catalyst of lateral transposition, a homing endonuclease should clearly minimize its activity toward essential genes in the host genome, while maximizing its ability to recognize and cleave closely related variants of the host target site. Based on the length of the I-*Cre*I homing site and the overall pattern of sequence degeneracy across that site in genetic screens (Argast, 1998), the overall cleavage frequency across random sequence is estimated at 10^{-8} to 10^{-9}, which is the equivalent of approximately 9–10 consecutive base pairs of inviolate sequence. This is clearly specific enough to prevent nonspecific cleavage events in the chloroplast genome (the organelle to which I-*Cre*I is localized). At the same time, variation of any one nucleotide base pair in the homing site is insufficient to prevent cleavage and homing, and in fact, many different combinations of two or more sequence changes are still recognized efficiently by the enzyme. This might allow homing and transposition of mobile introns and inteins to remain efficient between rapidly diverging host strains in the immediate environment. Similar results and conclusions may be drawn from recently reported structural and genetic studies of I-*Ppo*I (Argast, 1998; Flick et al., 1998; Wittmayer et al., 1998), which also exhibits subsaturating protein contacts across its homing site, and degeneracy and flexibility at most base pairs across its target sequence, and is localized to the nucleolus of its *Physarum* host. However, that protein, which is a member of the “His-Cys box” homing

endonuclease family, exhibits a completely different sequence, protein fold, and active site architecture than I-CreI, as well as a unique endonuclease catalytic mechanism. A comparison of these structures provides an excellent example of the unique end results of separate evolutionary pathways for mobile genetic sequences and their associated protein cofactors.

Active Site and Catalysis

Homing endonucleases that cleave DNA across the minor groove must address specific steric requirements in order to hydrolyze two phosphate groups separated by only 10–12 Å in undistorted B-form DNA. They must position the catalytic machinery for a nuclease attack in close proximity to two closely opposed scissile phosphates. This problem is solved by the unrelated, nuclear homing endonuclease I-*Ppo*I by drastically widening the minor groove at the cleavage site (Flick et al., 1998) and positioning a unique active site near the labile phosphate group. In contrast, the LAGLIDADG motif of I-*Cre*I is used to generate a closely packed subunit interface, with equivalent active site residues positioned at the center of the cleavage site (Duan et al., 1997; Heath et al., 1997). The structure of I-*Cre*I bound to its DNA homing site supports the presence of two separate active sites within the endonuclease homodimer, with an architecture reminiscent of other endonucleases that cleave DNA in a site-specific manner. The best studied of these site-specific endonucleases are the type II restriction endonucleases (Roberts and Halford, 1993; Aggarwal, 1995; Pingoud and Jeltsch, 1997; Aggarwal and Wah, 1998). These restriction endonucleases contain a conserved PD...(D/E)X(K/R) motif in which the acidic side chains are positioned near the scissile phosphate to bind a divalent cation and water molecules in an activated complex, and the positively charged side chain is positioned to stabilize the pentacoordinate phosphoanion transition state. This strategy appears to have been loosely reproduced in the LAGLIDADG homing endonucleases, with an aspartate residue from the LAGLIDADG motif (Asp-20) and an essential glutamine residue (Gln-47) participating in metal binding, and a basic residue (probably Lys-98) interacting with the labile phosphate.

The precise mechanism of cleavage has not yet been established for any endonuclease, but models have been proposed that involve either substrate-assisted or metal-assisted catalysis (Pingoud and Jeltsch, 1997). For both mechanisms, DNA cleavage is presumed to occur through an attack of a water molecule on a phosphorus atom yielding 3′-OH and 5′-phosphate ends. Hydrolysis of the scissile phosphodiester bonds requires at least three functional entities: a general base to activate the attacking nucleophile, a Lewis acid to stabilize the negatively charged oxyanions in the pentacoordinate transition state, and acids that stabilize and protonate the hydroxyl leaving group. Based on the structure of I-*Cre*I bound to its homing site, we favor a mechanism involving single-metal ion assisted hydrolysis of the phosphodiester bond, with a separate active site responsible for each cleavage event. A water molecule coordinated to the bound metal ion would be the most appropriately positioned candidate to act as an in-line nucleophile in this reaction.

An unresolved question is whether other LAGLIDADG endonucleases form two completely separate but closely packed active sites, as suggested for I-*Cre*I (Heath et al., 1997), or a single active site (or two partially overlapping active sites) as suggested for PI-*Sce*I (Duan et al., 1997). The latter possibility could be achieved by coordinating a single divalent cation between the two equivalent subunits or domains using conserved aspartate residues from each LAGLIDADG helix and a different positively charged residue from each domain to contact a phosphate during cleavage. The inability to uncouple active sites in PI-*Sce*I, including mutations of individual aspartate residues that eliminate both cleavage and nicking activity, supports this model (Gimble and Stephens, 1995). A distinct possibility may be that these two enzymes, which have strongly diverged both in sequence (outside of the LAGLIDADG region) and in quaternary structure (through a gene duplication event) may have also diverged in their active site structure and function to the point that the two enzymes follow distinctly different structural mechanisms of DNA cleavage.

In summary, the structure and mechanism of homing endonucleases such as I-*Cre*I are dictated by several separate aspects of their function. First, the unusual position of their open reading frames inside self-splicing intervening sequences may limit the size of many of these enzymes. However, these intron-encoded endonucleases must also recognize and cleave long homing sites with sufficient specificity to minimize toxicity to the host while accommodating subtle changes in target site sequence in order to maximize their ability to induce genetic homing or transposition between divergent host strains. In order to achieve these somewhat disparate challenges, homing endonucleases have evolved unique strategies to economize their folded structures, to read long DNA sequences, and to exhibit controlled flexibility in site recognition.

Experimental Procedures

Crystallization

The I-*Cre*I endonuclease was subcloned, overexpressed, and purified as previously described (Heath et al., 1997). The DNA was purchased from Oligos Etc. (Wilsonville, OR) and consisted of two strands of sequence: 5′-GCAAAACGTCGTGAGACAGTTTCG-3′ and its complement 5′-CGAAACTGTCTCACGACGTTTTGC-3′. The construct forms a pseudopalindromic, 24 bp blunt-ended duplex. Crystals were grown using a 2:1 molar ratio solution of DNA:protein by hanging drop vapor diffusion against a reservoir containing 14% PEG 3350, 100 mM MES (pH 6.5) with 10 mM $CaCl_2$, 20 mM NaCl, and 0.2% 1–3 propanediol. The final concentration of I-*Cre*I in the DNA:protein complex solution was 4 mg/ml.

Data Collection

Crystals were transferred sequentially to aliquots of the crystallization reservoir with 15% and 30% (w/v) glucose added as a cryoprotectant. A single crystal was suspended in a fiber loop, frozen in liquid nitrogen, and maintained at 100 K during data collection. All data collection was performed on a Rigaku RAXIS IV area detector mounted on a Rigaku RU200 rotating anode X-ray generator equipped with mirror focusing optics (Molecular Structure Corporation). Data were reduced using the DENZO/SCALEPACK crystallographic data reduction package (Otwinowski and Minor, 1997). The statistics for data collection are in Table 1. The crystal belongs to

space group $P2_12_12_1$, with unit cell dimensions a = 67.56 Å, b = 89.26 Å, c = 177.52 Å.

Structure Refinement

The structure of the complex was solved by molecular replacement with the program EPMR (Kissinger and Gehlhaar, 1997), using a truncated model of the apoenzyme dimer as a search probe. There are two I-*Cre*I dimer/DNA homing site complexes in the asymmetric unit of the crystal. Noncrystallographic symmetry averaging was used to extend phases and improve maps for the incomplete model: an 18 bp double-stranded model of B-form DNA was docked to the concave surface of the protein expected to contact DNA, and a mask was generated around this model complex with the program MAMA (Keywegt and Jones, 1996). Two-fold and four-fold NCS averaging was performed within the mask with the program RAVE (Keywegt and Jones, 1996), and maps were calculated using the CCP4 package (1979). The subsequent maps were of excellent quality and allowed an unambiguous model for most of the DNA and truncated protein loops to be built in QUANTA 96 (1996). Subsequent expansion of the NCS mask and averaging to cover the full model revealed the remaining DNA and protein density. This initial model was refined using X-PLOR (Brünger, 1992) against the native data set with 6% (1202) of reflections set aside for R_{free}, which was used to monitor all stages of the refinement (Brünger, 1993). Four-fold or two-fold NCS restraints were used for the protein component throughout the refinement and a bulk-solvent correction was used. Restrained group B factors were refined once the R factors had dropped to reasonable values. For each residue, a single overall B factor for its main-chain atoms and a second separate overall B factor for its side-chain atoms were refined. The final refined model consists of residues 2–154 and 12 base pairs of B-form DNA for each of four protein subunits and DNA half-sites. The unique orientation of the pseudopalindromic DNA sequence (which differs at five positions in the 12 nucleotide half-site) cannot be determined from density maps at this resolution and has been modeled in a single orientation, based on the best value of R_{free} in the refinements. Geometric analysis of the structure (Laskowski et al., 1993) indicates that there are no residues with generously allowed or unfavorable backbone dihedral angles, and over 87% of the residues are in the core region of the Ramachandran plot.

Acknowledgments

This work was supported by NIH grant GM49857 (B. L. S.), NIH grant CA48022 (R. J. M.), and graduate training grant fellowships from the NSF and the NIH (M. S. J.). We acknowledge the efforts of Pat Heath, David Erho, Kathryn Stephens, and Gretchen Argast in the initial subcloning, purification, and crystallization studies, and Roland Strong, Kam Zhang, and Betty Shen for advice during the crystallographic analysis.

Received June 24, 1998; revised August 14, 1998.

References

Aggarwal, A.K. (1995). Structure and function of restriction endonucleases. Curr. Opin. Struct. Biol. *5*, 11–19.

Aggarwal, A.K., and Wah, D.A. (1998). Novel site-specific DNA endonucleases. Curr. Opin. Struct. Biol. *8*, 19–25.

Argast, G.M. (1998). I-*Ppo*I and I-*Cre*I homing site sequence degeneracy determined by random mutagenesis and sequential in vitro enrichment. J. Mol. Biol. *280*, 345–353.

Belfort, M., and Perlman, P.S. (1995). Mechanisms of intron mobility. J. Biol. Chem. *270*, 30237–30240.

Belfort, M., and Roberts, R. (1997). Homing endonucleases—keeping the house in order. Nucleic Acids Res. *25*, 3379–3388.

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

Brünger, A. (1992). XPLOR Version 3.1: A System for X-ray Crystallography and NMR (New Haven, CT: Yale University Press).

Brünger, A. (1993). Assessment of phase accuracy by cross validation: the free R value. Methods and Applications. Acta Crystallogr. D *49*, 24–36.

CCP4 (1979). The SERC (UK) collaborative computing project No. 4, a suite of programs for protein crystallography. (Warrington, UK: Daresbury Laboratory).

Cooper, A.A., and Stevens, T.H. (1995). Protein splicing: self-splicing of genetically mobile elements at the protein level. Trends Biochem. Sci. *20*, 351–356.

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.

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.

Evans, S. (1993). SETOR: hardware-lighted three-dimensional solid model representations of macromolecules. J. Mol. Graph. *11*, 134–138.

Flick, K.E., Jurica, M.S., and Monnat, R.J. (1998). DNA binding and cleavage by the nuclear intron-encoded homing endonuclease I-*Ppo*I. Nature *394*, 96–101.

Gimble, F.S., and Stephens, B.W. (1995). Substitutions in conserved dodecapeptide motifs that uncouple the DNA binding cleavage activities of PI-*Sce*I endonuclease. J. Biol. Chem. *270*, 5849–5856.

He, Z., Crist, M., Yen, H., Duan, X., Quiocho, F.A., and Gimble, F.S. (1998). Amino acid residues in both the protein splicing and endonuclease domains of the PI-*Sce*I intein mediate DNA binding. J. Biol. Chem. *273*, 4607–4615.

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.

Hensgens, L.A.M., Bonen, L., de Haan, M., van der Horst, G., and Grivell, L.A. (1983). Two intron sequences in yeast mitochondrial COX1 gene: homology among URF-containing introns and strain-dependent variation in flanking exons. Cell *32*, 379–389.

Keywegt, G.J., and Jones, T.A. (1996). Halloween...masks and bones. In From First Map to Final Model, S. Bailey, R. Hubbard, and D. Waller, eds. (Warrington, UK: SERC Daresbury Laboratory), pp. 59–66.

Kissinger, C.R., and Gehlhaar, D.K. (1997). EPMR: A Program for Crystallographic Molecular Replacement by Evolutionary Search (La Jolla, CA: Agouron Pharmaceuticals).

Kostriken, R., Strathern, J.N., Klar, A.J.S., Hicks, J.B., and Heffron, F. (1983). A site-specific endonuclease essential for mating-type switching in Saccharomyces cerevisiae. Cell *35*, 167–174.

Lambowitz, A.M., and Belfort, M. (1993). Introns as mobile genetic elements. Annu. Rev. Biochem. *62*, 587–622.

Laskowski, R.A., 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.

Mueller, J.E., Brysk, M., Loizos, N., and Belfort, M. (1993). Homing endonucleases. In Nucleases, S.M. Linn, R.S. Lloyd, and R.J. Roberts, eds. (Cold Spring Harbor, NY: Cold Spring Harbor Laboratory Press), pp. 111–143.

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

Phillips, S.E.V. (1994). The β-ribbon DNA recognition motif. Annu. Rev. Biophys. Biomol. Struct. *23*, 671–701.

Pietrokovski, S. (1994). Conserved sequence features of inteins (protein introns) and their use in identifying new inteins and related proteins. Protein Sci. *3*, 2340–2350.

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

QUANTA 96 (1996). X-Ray Structure Analysis User's Reference (San Diego, CA: Molecular Simulations, Inc).

Roberts, R.J., and Halford, S.E. (1993). Type II restriction endonucleases. In Nucleases, S.M. Linn, R.S. Lloyd, and R.J. Roberts, eds.

(Cold Spring Harbor, NY: Cold Spring Harbor Laboratory Press), pp. 35–88.

Seligman, L.M., Stephens, K.M., Savage, J.H., and Monnat, R.J. (1997). Genetic analysis of the *Chlamydomonas reinhardtii* I-*Cre*I mobile intron homing system in *Escherichia coli*. Genetics *147*, 1653–1664.

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.

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.

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

Waring, R.B., Davies, R.W., Scazzocchio, C., and Brown, T.A. (1982). Internal structure of a mitochondrial intron of *Aspergillus nidulans*. Proc. Natl. Acad. Sci. USA *79*, 6332–6336.

Wittmayer, P.K., Mckenzie, J.L., and Raines, R.T. (1998). Degenerate DNA recognition by I-PpoI endonuclease. Gene *206*, 11–21.

Yang, J., Zimmerly, S., Perlman, P.S., and Lambowitz, A.M. (1996). Efficient integration of an intron RNA into double-stranded DNA by reverse splicing. Nature *381*, 280–281.

Brookhaven Protein Data Bank Accession Number

The accession number for the atomic coordinates is 1bp7. Coordinates are also available immediately from B. L. S. (bstoddar@fred.fhcrc.org).