Quarterly Reviews of Biophysics, Page 1 of 47. © 2005 Cambridge University Press
doi:10.1017/S0033583505004063 Printed in the United Kingdom

Homing endonuclease structure and function

Barry L. Stoddard

Division of Basic Sciences, Fred Hutchinson Cancer Research Center, 1100 Fairview Ave N., A3-023, Seattle WA 98109, USA

Abstract. Homing endonucleases are encoded by open reading frames that are embedded within group I, group II and archael introns, as well as inteins (intervening sequences that are spliced and excised post-translationally). These enzymes initiate transfer of those elements (and themselves) by generating strand breaks in cognate alleles that lack the intervening sequence, as well as in additional ectopic sites that broaden the range of intron and intein mobility. Homing endonucleases can be divided into several unique families that are remarkable in several respects: they display extremely high DNA-binding specificities which arise from long DNA target sites (14–40 bp), they are tolerant of a variety of sequence variations in these sites, and they display disparate DNA cleavage mechanisms. A significant number of homing endonucleases also act as maturases (highly specific cofactors for the RNA splicing reactions of their cognate introns). Of the known homing group I endonuclease families, two (HNH and His-Cys box enzymes) appear to be diverged from a common ancestral nuclease. While crystal structures of several representatives of the LAGLIDADG endonuclease family have been determined, only structures of single members of the HNH (I-HmuI), His-Cys box (I-PpoI) and GIY-YIG (I-TevI) families have been elucidated. These studies provide an important source of information for structure–function relationships in those families, and are the centerpiece of this review. Finally, homing endonucleases are significant targets for redesign and selection experiments, in hopes of generating novel DNA binding and cutting reagents for a variety of genomic applications.

Tel.: 1-206-667-4031; Fax: 1-206-667-6877; E-mail: bstoddar@fred.fhcrc.org

1. Intron mobility and homing endonucleases: basic principles

1.1 Intron and intein homing

Homing is the transfer of an intervening sequence (either an intron or intein) to a homologous allele that lacks the sequence (Dujon, 1989; Dujon *et al.* 1989; Belfort & Perlman, 1995), leading to gene conversion and dominant transmission and inheritance of the mobile element (Fig. 1). Invasion of ectopic sites, including transfer across biological kingdoms and between different genomic compartments, has also been documented (Cho *et al.* 1998; Goddard & Burt, 1999). Homing is initiated by an endonuclease, that is encoded within the mobile intervening sequence, that recognizes a DNA target site and generates a single- or double-strand break.

Transfer of group I introns and inteins is initiated solely by the endonuclease activity, and is completed by cellular mechanisms that repair the strand breaks via homologous recombination, using the intron- or intein-containing allele as a template (Colleaux *et al.* 1986; Dujon, 1989; Belfort & Perlman, 1995). In contrast, homing of mobile group II introns is completed by more complex processes, each involving a combination of strand cleavage, a reverse splicing reaction that generates a DNA–RNA hybrid intermediate, and finally reverse transcription using the inserted RNA intron as a template (Moran *et al.* 1995; Zimmerly *et al.* 1995a, b; Curcio & Belfort, 1996; Mills *et al.* 1996; Shearman *et al.* 1996; Guo *et al.* 1997; Matsuura *et al.* 1997; Cousineau *et al.* 1998; Yang *et al.* 1998). These sequential activities are also encoded within the intron, on a single multifunctional polypeptide chain.

In any of these cases, the homing endonuclease gene (HEG) is duplicated into the target site (Chevalier & Stoddard, 2001). The endonuclease reading frames, which represent extremely specific transposable elements, avoid disrupting host gene function by embedding themselves

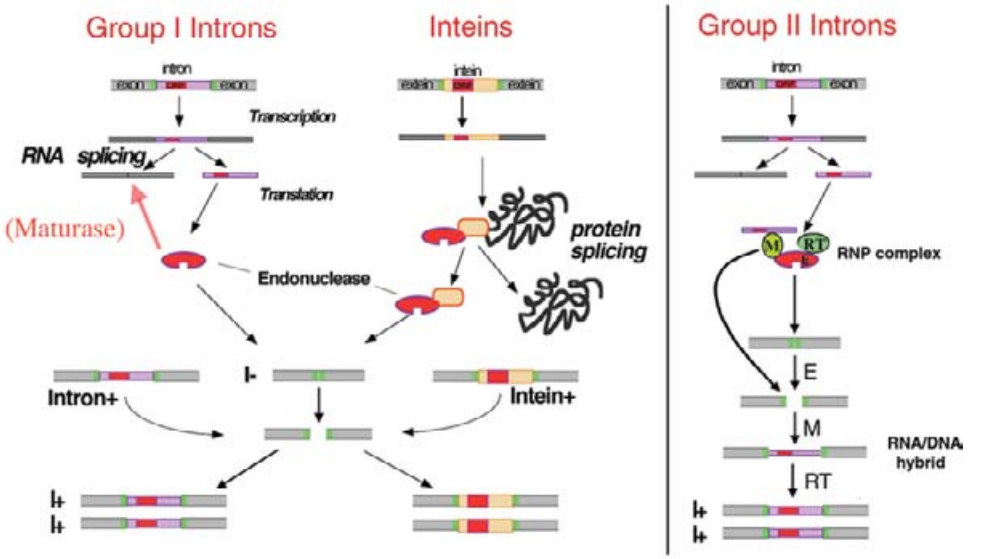


Fig. 1. Generalized homing mechanisms for mobile group I introns, inteins and group II introns. While the first two elements rely on cell-based repair of double-strand breaks, induced by the endonuclease, the latter element uses a more complex mechanism of reverse splicing and subsequent reverse transcription of the intron, as described in the text.

into sequences that are capable of self-splicing at the RNA (introns) or protein (inteins) level, as well as by occasionally transferring themselves, as stand-alone ORFs, into intergenic regions.

Homing endonucleases are highly specific, recognizing DNA target sites ranging from 14 to 40 base pairs (bp) in length, and have evolved to cleave target sequences within cognate alleles without being overly toxic to the organism (Belfort *et al.* 1995; Belfort & Roberts, 1997; Chevalier & Stoddard, 2001). In order to be evolutionarily successful, they must balance two somewhat contradictory requirements: they need to be highly sequence-specific, in order to promote precise intron or intein transfer in their host genomes while avoiding toxicity, and yet must retain sufficient flexibility of site recognition (i.e. reduced fidelity) to permit lateral transfer in the face of sequence variation in diverging targets and hosts.

Homing appears to be widespread. This genetic phenomenon was first described for a group I intron of budding yeast (Dujon, 1980; Jacquier & Dujon, 1985; Colleaux *et al.* 1986). In the 1970s, the genetic marker 'ω' in *Saccharomyces cerevisiae* was found to transfer to strains lacking the marker when crossed to ω^+ strains. This marker corresponded to a 1·1 kb group I intron found in the large ribosomal RNA (rRNA) gene of the mitochondrial genome. Subsequent analysis indicated that the gene duplication event required a double-strand break at the intron insertion site and the expression of an ORF within the intron itself (Jacquier & Dujon, 1985). This ORF was further shown to encode a site-specific endonuclease capable of recognizing and cleaving the intronless allele, and thereby initiating the homing event (Colleaux *et al.* 1986). This protein, now called I-SceI based on current nomenclature conventions, was the first of many hundreds of homing endonucleases to be identified.

Intervening sequences capable of homing are found in all branches of life (phage, Eubacteria, Archaea, and eukaryotes). Approximately one-quarter to one-third of group I introns, and a similar fraction of group II introns, are estimated to contain internal ORFs, and a significant

number of these appear to be mobile (Dujon, 1989; Belfort & Perlman, 1995; Curcio & Belfort, 1996; Lykke-Andersen *et al.* 1997a; Chevalier & Stoddard, 2001). Within eukaryotes, these elements are found within nuclear, mitochondrial and chloroplast genomes. In contrast, spliceosomal introns are not known to be mobile or to harbor homing endonucleases. However, structural similarities between group II introns and spliceosomal introns have lead to suggestions that some introns in the human genome might be relics of once-mobile group II introns (Mohr *et al.* 1993; Michel & Ferat, 1995; Lambowitz *et al.* 1999). Further evidence for this theory is provided by experiments in which engineered group II introns invade ectopic chromosomal sites (Guo *et al.* 2000).

The homing mechanism of mobile group I introns requires the endonuclease activity of the intron-encoded protein (Dujon, 1989; Loizos *et al.* 1994) (Fig. 1). Many of these enzymes (particularly those encoded within chloroplast or mitochondrial genomes) are specific solely for the intronless target site, recognizing and cleaving a homing-site sequence that corresponds to what will eventually become flanking exons after the intron invasion event. The DNA lesions generated by the homing enzymes are subsequently repaired by homologous recombination, using host cell repair machinery (Bell-Pedersen *et al.* 1989; Dujon, 1989; Belfort & Perlman, 1995; Mueller *et al.* 1996). In comparison, many mobile introns found within phage genomes display greater diversity in their homing mechanisms: some endonucleases generate single-strand breaks (nicks) that are converted to double-strand breaks during DNA replication, and some recognize both intron-deficient and intron-containing target sites (Belfort, 1989; Belfort & Perlman, 1995; Belfort *et al.* 1995; Belfort & Roberts, 1997; Edgell & Shub, 2001; Landthaler & Shub, 2003). This latter, more promiscuous activity can drive complex patterns of intron competition in mixed phage infections (Goodrich-Blair & Shub, 1996).

While the homing mechanism of inteins is similar to group I introns (involving a double-strand break repair pathway induced solely by the endonuclease), the system is remarkable for the structural and functional fusion of the endonuclease with its intein host (Wende *et al.* 1996; He *et al.* 1998; Moure *et al.* 2002) (Fig. 1). In these systems, the endonuclease reading frame is positioned in-frame with the surrounding intein sequence, at a location corresponding to a surface loop of the structural intein protein. The intein itself is also encoded in-frame with the surrounding sequence of its host protein, again at a location corresponding to an amenable position on the protein surface. The entire protein construct, which can be likened to a series of Russian nesting dolls, is translated as a single polypeptide chain, harboring activities of the homing endonuclease, the intein peptide ligase, and the host protein. Whereas the endonuclease is not involved in, and is in fact usually dispensable for the intein's ability to ligate and splice the flanking peptides of the host protein, significant portions of the intein's surface have often been co-opted for participation in DNA recognition and binding by the endonuclease. The structure–function relationships of fused homing endonuclease–intein domain systems have been particularly well described, using X-ray crystallography and biochemical analyses, for the PI-SceI endonuclease–intein system (Wende *et al.* 1996; Duan *et al.* 1997; He *et al.* 1998; Moure *et al.* 2002).

The homing mechanism employed by mobile group II introns, termed 'retrohoming' (Fig. 1) differs significantly from the mechanisms described above (Lazowska *et al.* 1994; Moran *et al.* 1995; Eskes *et al.* 1997; Matsuura *et al.* 1997). This mechanism is more complex and requires the use of a ribonucleoprotein (RNP) consisting of the intron-encoded protein and the spliced intron RNA (Curcio & Belfort, 1996; Guo *et al.* 1997). Individual domains of these proteins function as an endonuclease (Zimmerly *et al.* 1995a), maturase (Mohr *et al.* 1993) and reverse-transcriptase

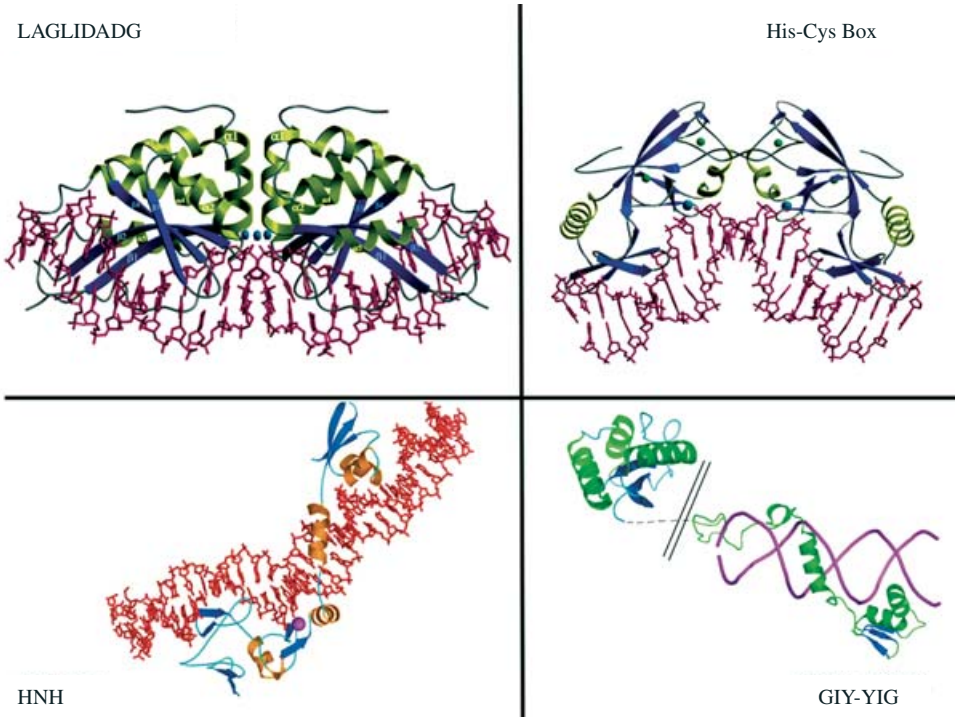


Fig. 2. Representative structures of the four families of homing endonucleases for which such studies have been reported. The enzymes shown are I-CreI (a LAGLIDADG homodimer; 22-bp DNA site), I-PpoI (a His-Cys box homodimer; 14-bp site), I-HmuI (a phage HNH enzyme; 26-bp site) and I-TevI (a phage GIY-YIG enzyme; 37-bp site).

(Kennell *et al.* 1993). Once the intron-encoded ORF has been translated, the protein product binds the RNA transcript and aids in proper splicing and excision of the RNA intron, forming a stable RNP in the process. The spliced intron product, in complex with the protein, recognizes the intronless allele using base-pairing between the protein-bound RNA and the DNA target site (Zimmerly *et al.* 1995a). The RNA lariat then *reverse* splices into the target site (Zimmerly *et al.* 1995b; Yang *et al.* 1998), and the complementary DNA strand is cleaved by the protein's endonuclease domain. Finally, the protein's reverse-transcriptase domain synthesizes DNA using the invading RNA as template. Cellular machinery completes the homing of the intervening sequence by replacing the invading RNA with DNA.

1.2 Homing endonuclease phylogeny and evolution

There are five known families of homing endonuclease catalysts, differing greatly in their conserved nuclease active-site core motifs and catalytic mechanisms, biological and genomic distributions, and wider relationships to non-homing nuclease systems (Fig. 2). These endonuclease families, and their structure–function relationships, are independently summarized in Sections 2–6 of this review. The largest known family, termed the 'LAGLIDADG' endonucleases (Section 2), is typically encoded in mitochondrial or chloroplast genomes in single-cell eukaryotes. These enzymes are translated within introns as free-standing enzymes (either homodimers, or pseudosymmetric monomers) or within inteins as in-frame fusion

proteins, which results in a bifunctional protein with endonuclease and protein ligation activities (Dalgaard *et al.* 1997; Chevalier *et al.* 2003). Multiple structures of the major variants and types of these endonucleases have been determined, as described later.

The second and third families, known as the 'HNH' and 'His-Cys box' endonucleases (Sections 3 and 4 respectively), appear to be derived from a common ancestral nuclease, and are built around a consensus nuclease active site architecture known as a '$\beta\beta\alpha$-metal' motif (Friedhoff *et al.* 1996; Kuhlmann *et al.* 1999). Although only one representative enzyme from each family has been visualized structurally (I-HmuI and I-PpoI respectively) (Flick *et al.* 1998; Shen *et al.* 2004), comparison of these two structures to those of non-specific bacterial colicins (Kleanthous *et al.* 1999; Ko *et al.* 1999; Pommer *et al.* 2001; Cheng *et al.* 2002; Sui *et al.* 2002; Hsia *et al.* 2004; Mate & Kleanthous, 2004) have provided a solid framework for understanding the mechanisms of these enzymes. Despite their common core folds and active site architectures, these two homing endonuclease families are sufficiently diverged in structure–function relationships that they are treated and described as separate families in most descriptions of homing endonucleases. The 'HNH' homing endonuclease family is generally found in phage introns (Mehta *et al.* 2004). Its prototypical members (such as I-HmuI) recognize long asymmetric DNA sites (24 bp and longer). These enzymes possess a highly extended, modular, monomeric structure, in which the relatively non-specific nuclease domain, located near the enzyme's N-terminus, is tethered to additional structural motifs that confer and restrict DNA-binding specificity (Shen *et al.* 2004). Many (but not all) of these enzymes nick a single strand of their asymmetric DNA target sequence. In contrast, the His-Cys box enzymes are generally encoded in nucleolar introns within rDNA host genes, have evolved compact homodimeric structures (Flick *et al.* 1998), recognize shorter symmetric DNA target sites with higher overall specificity than the HNH enzymes, and generate double-strand breaks that are resolved during homing in a manner similar to the LAGLIDADG systems.

The fourth enzyme family, known as the GIY-YIG endonucleases (Section 5) are also encoded in phage introns and possess similar types of modular structures to that described above for the HNH endonucleases. This enzyme family has clearly encountered the HNH enzyme family, resulting in exchanges or swaps of DNA-binding domains which are tethered to their unique catalytic domains (VanRoey *et al.* 2001; Shen *et al.* 2004). This process probably has occurred through frequent contact of these enzymes and their reading frames in mixed infections of common bacterial targets. The GIY-YIG catalytic domain, which contains the conserved sequence signatures that give this family its name, is quite non-specific in its inherent cleavage activity (even more so than the nuclease domain of the HNH enzymes), again being restricted to target sites that are dictated by the appended DNA-binding modules. Unlike the endonuclease families noted above, a GIY-YIG active site has not been visualized in complex with a bound DNA target, and, therefore, is less thoroughly understood mechanistically. However, comparisons of active site structures indicate that they may use a similar strategy for phosphate bond cleavage as that of the HNH and His-Cys box enzymes (Kowalski *et al.* 1999).

Finally, a small number of studies extending back to the mid-1990s have demonstrated that a fifth, thus far uncharacterized type of homing endonuclease (Section 6) is responsible for the presence and persistence of introns in cyanobacterial tRNA genes (Biniszkiewicz *et al.* 1994; Bonocora & Shub, 2001). The prototypical enzyme, found in *Synechocystis* and now named I-SspI, displays sequence identity to intergenic, free-standing reading frames that often appear to also encode active endonucleases. This enzyme displays limited homology to known nuclease superfamilies, and is represented by only a small handful of identified open reading frames.

1.3 Intron mobility and life cycles

At their most basic level, homing endonuclease genes are extremely efficient parasitic elements that take advantage of host DNA double-strand break repair mechanisms for propagation (Lambowitz & Belfort, 1993; Goddard & Burt, 1999). They are closely associated with intervening sequences, and use these sequences as a refuge that enables the replication of their ORF without deleterious effects to a host gene. The existence of multiple evolutionarily distinct families of homing endonucleases indicates several independent ancestral origins for these elements. Furthermore, the distribution of homing endonucleases among intervening sequences suggests the origin of homing endonucleases and intervening sequences to also be independent (Quirk *et al.* 1989; Loizos *et al.* 1994; Lykke-Andersen *et al.* 1997a; Goddard & Burt, 1999).

The relationship between homing endonuclease-containing mobile intervening sequences and their host genes is dynamic. The extra-Mendelian inheritance of mobile intervening sequences assures that they rapidly become fixed in a population. After this occurs, there is little selective pressure for the maintenance of a functional homing endonuclease, and this ORF is subsequently lost, followed eventually by the intron itself. Once the intron is lost, the site is ripe for re-invasion. This cycle has been illustrated in yeast by the ω element in yeast (Posey *et al.* 2004). A sampling of 20 species enabled the categorization of the ω element into functional, non-functional and absent states. Comparisons of the phylogenetic analyses of the intron, the ω endonuclease (I-SceI), and the host yeast species present evidence for multiple rounds of rapid invasion by horizontal transmission, slow degeneration and eventual loss of the intron with an invasion frequency of once about every 2 million years.

There has been speculation whether homing endonucleases and/or their mobile intervening sequences confer any benefit to their host gene or organism. The best evidence indicates that sequences capable of homing are usually nothing more than opportunistic selfish DNA. Evidence for this conjecture is provided by the cycle of invasion and elimination of the ω element as described above. If these invasive elements benefit their hosts, it is unlikely that they would be eliminated from the host genome at such a rapid rate. However, it does appear possible that a small fraction of homing endonuclease genes have evolved a more host-beneficial role. In so doing, these enzymes may have generated selective pressure to be retained and, therefore, broken free of the invasion–elimination cycle of most homing sequences. For example, a subset of homing endonucleases also aid in the splicing of their host intron (and are thus termed 'maturases') (Lambowitz *et al.* 1999; Bolduc *et al.* 2003). While the presence of the intron itself might not benefit the host *per se*, the requirement of the protein for efficient removal of the intervening sequence may have provided enough selective pressure to maintain a functional protein, which in turn aided in the survival of the intron and its homing endonuclease ORF. Furthermore, some homing endonucleases appear to have been adopted by the host as free-standing enzymes. One LAGLIDADG enzyme in particular, HO, is responsible for the mating switch in yeast (Jin *et al.* 1997). This protein is closely related to the endonuclease domain of the intein-encoded PI-SceI and may have arisen from a gene duplication event or from the remnants of a parallel invasion by the intervening sequence. Finally, in bacteriophage some homing endonucleases appear to confer advantage to their host by specifically cleaving the DNA of the heterologous phage during mixed infections (Goodrich-Blair & Shub, 1996). Whether potentially 'beneficial' homing endonucleases have an invasion–elimination cycle analogous to ω has yet to be determined, but it seems possible that the dynamics of intron recycling described above are altered for such systems.

Although the mechanism by which horizontal transfer of mobile intervening sequences between distantly related species is not well understood, a growing body of evidence suggests that it occurs frequently. Studies of more than 300 diverse land plants reveals a recent massive invasion by a homing endonuclease-containing group I intron in the mitochondrial cox1 gene (Cho *et al.* 1998; Goddard & Burt, 1999). Similarly, homologous group I introns and associated homing endonucleases also exist in identical positions of large subunit RNA genes in an algal chloroplast and an amoeboid protozoon mitochondria (Lemieux *et al.* 1988; Lucas *et al.* 2001). The wide distribution of LAGLIDADG-containing inteins, as found in eukaryotes, bacteria and Archaea, are also suggestive of horizontal transmission.

2. LAGLIDADG homing endonucleases

2.1 Distribution and diversity

The LAGLIDADG protein family includes the first identified and biochemically characterized intron-encoded proteins (Dujon, 1980; Lazowska *et al.* 1980; Jacquier & Dujon, 1985). It has also been variously termed the 'DOD', 'dodecapeptide', 'dodecamer', and 'decapeptide' endonuclease family, based on the conservation of a 10-residue sequence motif (Dujon, 1989; Dujon *et al.* 1989; Belfort *et al.* 1995; Belfort & Roberts, 1997; Dalgaard *et al.* 1997; Chevalier & Stoddard, 2001). The biological host range of the LAGLIDADG endonucleases is quite broad. This includes the genomes of plant and algal chloroplasts, fungal and protozoan mitochondria, and Archaea (Dalgaard *et al.* 1997). This enzyme family is found in several unrelated types of intervening sequences, including group I introns, archaeal introns and inteins (Belfort & Roberts, 1997; Chevalier & Stoddard, 2001). Descendants 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) and maturases that assist in RNA splicing (Delahodde *et al.* 1989; Lazowska *et al.* 1989; Schafer *et al.* 1994; Geese & Waring, 2001).

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 (Fig. 3*a*) (Thompson *et al.* 1992; Wang *et al.* 1997) and I-CeuI (Turmel *et al.* 1997), act as homodimers and recognize consensus DNA targets sites that are constrained to palindromic or near-palindromic symmetry. Enzymes that have two copies of the LAGLIDADG motif (such as I-AniI; Fig. 3*b*) 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, various residues within the LAGLIDADG motif residues play structural and/or catalytic roles (see below). Prior to structural analyses of multiple members of both branches of the LAGLIDADG family, a central question facing investigators was whether the two LAGLIDADG motifs found in the monomeric enzymes formed a single active site, and whether the answer to this question would be reflected in the structure and mechanism of their dimeric cousins (Gimble *et al.* 1996; Lykke-Andersen *et al.* 1996, 1997b; Wende *et al.* 1996). As discussed below, it is now known that both forms of these enzymes display two unique active sites, responsible for cleavage of each DNA strand. The active sites are found to be immediately adjacent to one another in the enzyme tertiary and/or quaternary structure; domain swapping of catalytic residues between the active sites is commonly observed, as is the use of shared metal-binding sites in the dimeric enzymes.

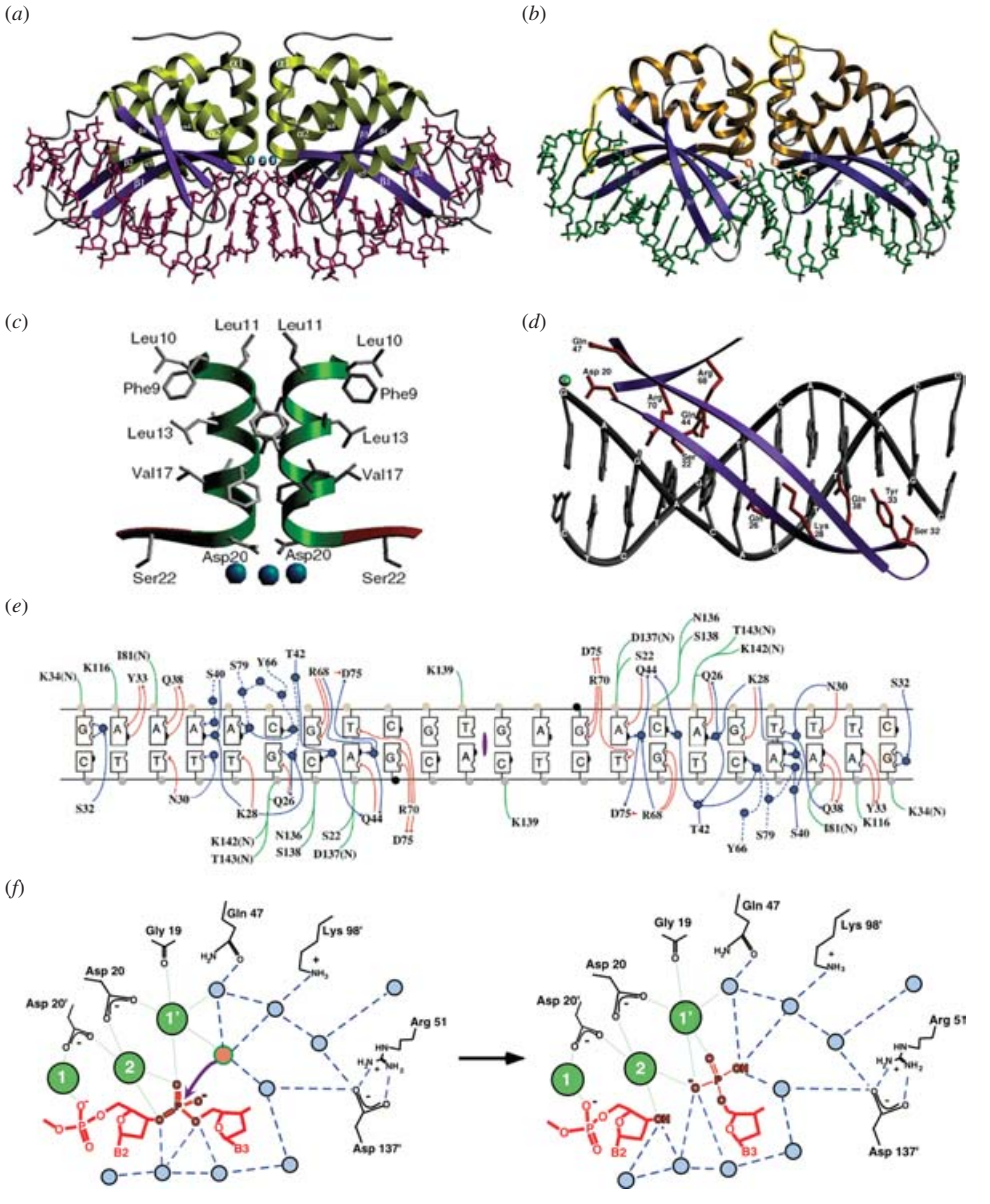


Fig. 3. (*a*) DNA bound complex of the I-CreI homodimer. (*b*) DNA bound complex of the I-AniI monomer. The peptide linker connecting the two related domains is outlined in yellow. (*c*) The LAGLIDADG motifs form the helices at the domain interface of the I-CreI structure and serve a similar role in all known LAGLIDADG enzymes. (*d*) Structure and orientation of the DNA-binding antiparallel β-sheets from I-CreI. (*e*) Summary of direct and water-mediated contacts between the I-CreI enzyme and the bases of its DNA target site. Direct bonds are shown as solid arrows (red to bases, green to backbone atoms); water-mediated contacts are shown as blue dashed and solid lines. The scissile phosphates are indicated with black closed circles. (*f*) Proposed catalytic mechanism for I-CreI, as described in the text. The nucleophilic water is red; surrounding ordered water molecules are blue.

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 bp (Durrenberger & Rochaix, 1993; Perrin *et al.* 1993; Dalgaard *et al.* 1994; Agaard *et al.* 1997). They cleave both DNA strands across the minor groove, to generate mutually cohesive four-base 3′ overhangs (Colleaux *et al.* 1988; Thompson *et al.* 1992; Durrenberger & Rochaix, 1993). Like most nucleases, LAGLIDADG homing endonucleases require divalent cations for activity.

In contrast, DNA binding by intein-associated LAGLIDADG enzymes (typified by the well-studied PI-SceI enzyme) displays many additional features that involves recruitment of adjacent protein domains. The PI-SceI endonuclease–intein combination binds a 31-bp site. Surprisingly, the majority of the energetic contribution to binding is derived from interactions with the intein peptide splicing domain; the endonuclease domain contains the active sites, but on its own exhibits relatively weak, non-specific DNA binding (Gimble & Wang, 1996; Wende *et al.* 1996; He *et al.* 1998). The full-length complex of PI-SceI with its target site involve a kinetically complex, two-step mechanism, in which the intein splicing domain forms tight interactions to the 5′ end of the target site (approximately base pairs 5–21). Bending of the DNA (to a total curvature of $\sim 65^{\circ}$) allows subsequent formation of the final ES complex, with the endonuclease domains contacting the 3′ end of the site. In this complex, the LAGLIDADG active sites are positioned near the scissile phosphate bonds, across the DNA minor groove (Christ *et al.* 1999; Moure *et al.* 2002).

2.2 Structural studies

The structures of several LAGLIDADG enzymes bound to their DNA targets have been determined. These include two homodimeric isoschizomers [I-CreI (Heath *et al.* 1997; Jurica *et al.* 1998; Chevalier *et al.* 2001, 2003) and I-MsoI (Chevalier *et al.* 2003)], that are both encoded within group I introns in the 23S rDNA of green algae *Chlamydomonas reinhardtii* and *Monomastix*; two pseudosymmetric monomers [I-AniI (Bolduc *et al.* 2003) and I-SceI (Moure *et al.* 2003)] that are encoded in mitochondrial introns of the fungi *Aspergillus nidulans* and *S. 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 have also been determined of two additional enzymes 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-β-α-β-β-α) (Heath *et al.* 1997) (Fig. 3*a*, *b*). The overall shape of this domain is a half-cylindrical 'saddle' that averages approximately 25 Å × 25 Å × 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 (Fig. 3*d*). Each individual β-strand crosses the groove axis at an angle of $\sim 45^{\circ}$ 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, and the C- and N-terminal helices of the core domains are connected by flexible linker peptides with lengths between 3 and >100 residues (Dalgaard *et al.* 1997). 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 bp.

The LAGLIDADG motif plays three distinct, but interrelated roles in the structure and function of this enzyme family (Fig. 3*c*). 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 or 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 is 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 <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-bound β-sheets, are remarkably similar. A structural alignment of several 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). These positions correspond to residues that make contacts to base pairs 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. Similar alignments of intein-associated endonuclease domains indicate a more diverged structure of the β-sheet motifs.

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 & 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 & Rooman, 1996). In HTH domains, this helix is responsible for sequence-specific interactions with DNA.

2.3 DNA recognition

LAGLIDADG homing endonucleases use a flexible homing-site recognition strategy in which 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 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 the contacts across the interface facilitate the recognition and accommodation of specific polymorphisms at individual target site positions. The overall specificity of the LAGLIDADG endonucleases is not well established, but is generally thought to range from 1 in 10^8–10^9 random sequences for an average length of 2022 bp (Chevalier *et al.* 2003). Recent *in vivo* studies using transfected cell lines, however, indicate that the repertoire of sites capable of being recognized and cleaved in the nucleus is more restricted than that which is observable using *in vitro* screens, implying higher specificity (i.e. reduced frequency of cutting) under conditions of nuclear chromatin packaging of DNA (B. L. Stoddard, unpublished observations).

In the protein–DNA interfaces visualized at high resolution (2·5–1·5 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*d*, *e*). 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 direct contact with protein side chains). Typically, strands $\beta 1$ and $\beta 2$ extend the entire length of this interface in each half site, while strands $\beta 3$ and $\beta 4$ 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.

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 $\sim 45^\circ$. 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 $\sim 50^\circ$), with a corresponding deformation in the base pair propeller twist and buckle angles for those same bases, leading to narrowing of the minor groove at the site of DNA cleavage. 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 information content (specificity) of recognition and cleavage by I-CreI, at each base pair of its DNA target site, has been correlated between the number and type of inter-molecular contacts made by the enzyme to each base pair. Three general conclusions from these analyses are: (i) the specificity of base-pair recognition to structurally unperturbed DNA sequence is proportional to the number hydrogen bond contacts to each base pair; (ii) the degree of specificity is not significantly reduced 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.

2.4 DNA cleavage

The mechanism of catalysis has been particularly well studied for the I-CreI homodimeric enzyme (Chevalier *et al.* 2004); many of these results appear to be generalizable to the LAGLIDADG family. As described below, however, the precise use of bound divalent metal ions for cleavage by related enzymes (particularly the monomeric endonucleases) is still unsettled, with reports of a canonical two-metal mechanism by the homodimeric enzymes not yet unambiguously supported by definitive structural and mutagenic experiments. Several lines of evidence indicate that members of the broader enzyme family may display significant variation in their catalytic mechanisms.

The measured single-turnover kinetic rate constants k^*_{max} and K^*_m, of the wild-type I-CreI enzyme are $0{\cdot}03\ min^{-1}$ and $1{\cdot}0 \times 10^{-4}$ nM, respectively, giving a value for catalytic efficiency (k^*_{max}/K^*_m) of $0{\cdot}3\ nM^{-1}\ min^{-1}$. This enzyme and its relatives are all dependent on divalent cations for activity, similar to most known endonucleases. A wide variety of divalent metal ions have been assayed for cleavage activity with I-CreI and display a 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 and other phosphoryl transfer enzymes.

The structures of four endonuclease–DNA complexes that have been solved at relatively high resolution of 2·5 Å or better (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 the resulting interpretation that one shared metal participates in both cleavage reactions by virtue of interacting with the scissile phosphates and 3′ hydroxyl leaving groups on both DNA strands, as well as conserved acidic residues from the LAGLIDADG motifs (Fig. 3*f*). The structures of these four enzymes differ in the precise position and binding interactions of the metals, but implicate similar possible mechanisms where each strand is cleaved using a canonical two-metal mechanism for phosphodiester hydrolysis (Fig. 3*f*). In I-CreI, the 'shared' metal site has been directly and unambiguously observed only via anomalous X-ray scattering experiments for a product complex in the presence of manganese. In structures of the uncleaved substrate complex (determined in the presence of non-activating calcium), this site is not occupied, either indicating that the site might be occupied only transiently after the cleavage event (Chevalier *et al.* 2004). However, recent structural studies of the homodimeric enzyme I-CeuI unambiguously demonstrate the presence of a bound metal ion, in an uncleaved DNA complex, at a position that bridges the scissile phosphates and the active sites (B. L. Stoddard, unpublished observations).

Complicating the interpretations described above, the structures of DNA complexes of one monomeric enzyme (I-AniI) (Bolduc *et al.* 2003) and of the intein-associated PI-SceI (Moure *et al.* 2002), solved at lower resolution ($\sim$2·6 and 3 Å respectively), have 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. For the homodimeric enzymes, the question of whether the structural feature of a shared divalent metal ion between two active sites would impart any particular kinetic constraints on the individual cleavage events remains to be answered. In contrast, regardless of their precise catalytic mechanism, the structures of the asymmetric monomeric enzymes in complex with DNA (Moure *et al.* 2003) clearly demonstrate that catalysis probably involves 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 the high-resolution structures of I-CreI, a metal-bound water molecule in each active site is appropriately positioned for an in-line hydrolytic attack on a scissile phosphate group. In addition to the attacking water molecule, a well-ordered network of solvent 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 and coordinated by several basic residues that line the solvent pocket. These residues are remarkable for their chemical and structural diversity (Chevalier *et al.* 2004). 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. Furthermore, 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 essential residues in I-CreI, such as Lys98 and Gln47, that 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. Thus, 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 required for DNA hydrolysis.

3. HNH homing endonucleases

3.1 Distribution and diversity

The HNH nuclease motif is the least restricted of those motifs that are found in the group I homing endonuclease families. In addition to group I, group II and intein-associated homing endonucleases, it has been identified in non-specific bacterial and fungal nucleases, (such as the colicin enzyme family) (Friedhoff *et al.* 1999b; Kuhlmann *et al.* 1999), and is also found in a broad group of DNA-acting enzymes including transposases, restriction endonucleases (such as KpnI), DNA packaging factors, and a bacterial factor involved in a developmentally controlled DNA rearrangement (Dalgaard *et al.* 1997; Mehta *et al.* 2004). The HNH family can be divided into at least eight sub-families (Mehta *et al.* 2004) based on the exact sequence of the core HNH signature and on the presence or absence of uniquely conserved flanking residues, such as cysteine pairs that might participate in binding additional metal ions (Mehta *et al.* 2004). Various members of the HNH subfamilies are encoded in-frame with additional protein domains of known structure and function, including reverse transcriptase, DNA repair enzyme AP2 domains, cloacin ribonucleases, and various DNA-binding motifs (Sitbon & Pietrokovski, 2003; Mehta *et al.* 2004). Whereas enzymes in the bacterial colicin family generally contain a canonical HNH motif near their C-terminus, the group I homing endonucleases display a related 'HNN' motif, located closer to the N-terminus. The various cleavage activities of these enzymes reflects their catalytic diversity, with colicins digesting double-stranded DNA targets non-specifically, and homing endonucleases inducing either double-strand breaks or single-strand nicks in their duplex DNA targets (Chevalier & Stoddard, 2001).

The HNH motif is the hallmark sequence signature of a broadly distributed nuclease fold (Fig. 4), which forms a non-specific catalytic center in many enzymes, now known as a '$\beta\beta\alpha$-metal'

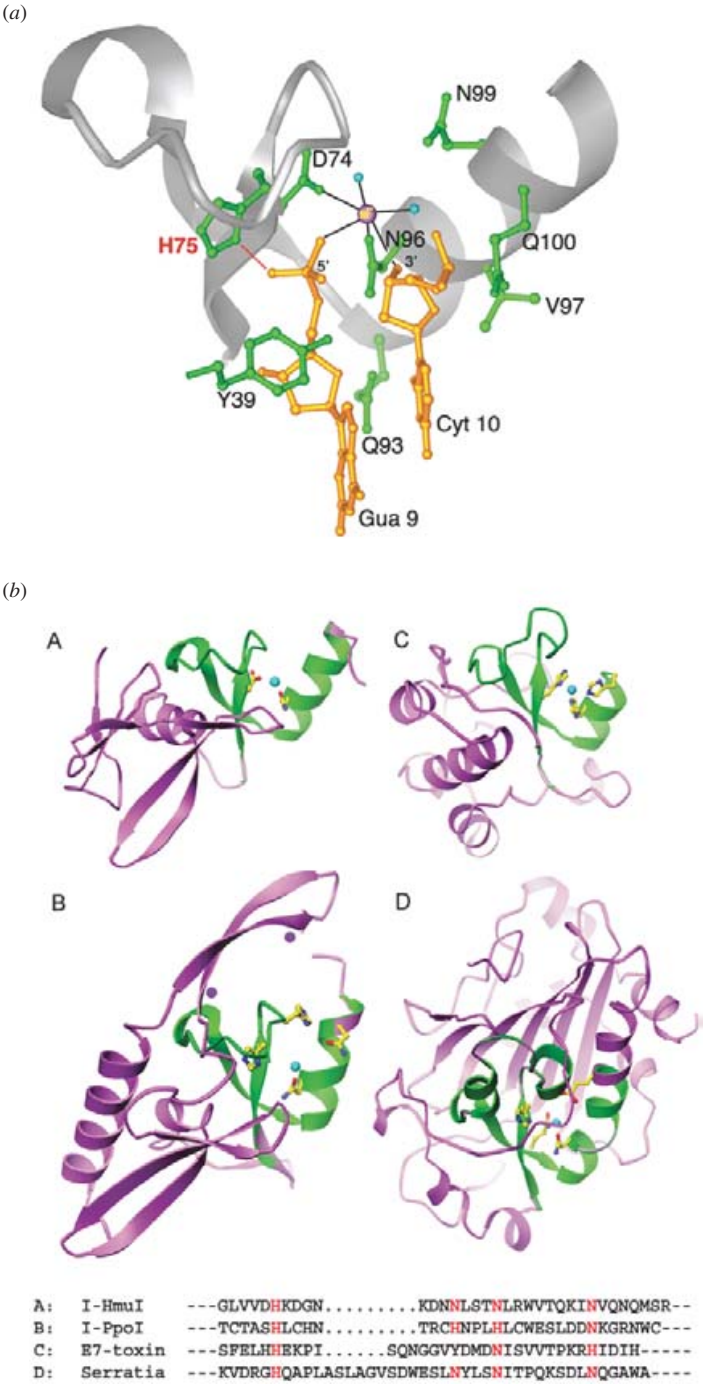


Fig. 4. (*a*) The $\beta\beta\alpha$-metal nuclease active site fold, containing highly conserved residues of the HNH motif. The active site shown here corresponds to the I-HmuI HNH homing endonuclease. Protein side-chains are green; DNA atoms for the nucleotides on either side of the scissile bond (Gua9 and Cyt10) are yellow. (*b*) Examples of the $\beta\beta\alpha$-metal fold in a variety of additional nuclease enzymes: two homing endonucleases (I-HmuI and I-PpoI, labeled 'A' and 'B'), one non-specific bacterial colicin (E7 from *E. coli*, labeled 'C') and a non-specific fungal nuclease (from *Serratia*, labeled 'D').

fold (Friedhoff *et al.* 1999b; Kuhlmann *et al.* 1999). The evolutionary distribution of this active site and its corresponding catalytic residues, and its ubiquitous involvement in a variety of nucleic acid-processing pathways, was postulated based on its presence in a variety of nucleases, including the colicins, the HNH homing endonucleases (this section), the His-Cys box homing endonucleases (described in Section 4), non-specific nucleases from *Serratia* and *Vibrio vulnificus* and T7 phage endonuclease. The crystal structures of nuclease domains of colicins E7 and E9, in complex with duplex DNA have been determined in the absence and presence of bound metal ions respectively (Hsia *et al.* 2004; Mate & Kleanthous, 2004). In those non-specific enzymes the HNH motif and its $\beta\beta\alpha$-metal active site is bound across the minor groove, primarily contacting DNA phosphate groups near the 3′ hydroxyl of the scissile phosphate. As described below and in Section 3.2, this active-site motif is similarly positioned across the minor groove in the HNH endonuclease I-HmuI and the His-Cys box endonuclease I-PpoI, but is augmented in both cases by additional DNA-contacting β-sheets that impart specific interactions to neighboring DNA base pairs.

Therefore, it appears that the 'HNH/N' fold appears to have diverged from a common ancestor to adopt a variety of diverse biological roles requiring the action of a nuclease. These enzymes range from structurally minimal non-specific bacterial colicins, to more elaborate endonucleases including HNH/N and His-Cys box homing endonucleases. The root mean square (rms) deviation of the HNH/N core structures of three of the enzymes mentioned above (E7 colicin, I-PpoI and I-HmuI) range from 1·0 to 1·5 Å for $\sim$30 superimposed residues (Shen *et al.* 2004). In contrast, the *V. vulnificus* (Vvn) (Li *et al.* 2003) and *Serratia* (Friedhoff *et al.* 1999b) nucleases, which also contain $\beta\beta\alpha$-metal nuclease active sites, display much greater rms deviation values for the same core residues ($\sim$5 Å) and very dissimilar folds and topologies that appears to indicate a separate line of convergent evolution to a similar $\beta\beta\alpha$-metal active site.

Members of the HNH family encoded within group I introns that have been relatively well studied include I-HmuI and I-HmuII from the SPO1 and SP83 introns respectively, of two closely related *B. subtilis* bacteriophage (Goodrich-Blair *et al.* 1990; Goodrich-Blair & Shub, 1994, 1996; Landthaler *et al.* 2004), I-BasI from the *Bastille* bacteriophage (Landthaler & Shub, 2003), phage T4 endonuclease I-TevIII (Eddy & Gold, 1991), the I-CmoeI endonuclease from the psbA gene of the *Chlamydomonas moewusii* chloroplast (Drouin *et al.* 2000), and a related ORF from the same gene in *C. reinhardtii* (Holloway *et al.* 1999).

These HNH homing endonucleases display significantly diverged biochemical properties outside of their defining motif. For example, I-HmuI cleaves only one strand of its DNA substrate (Landthaler *et al.* 2004), whereas I-TevIII and I-CmoeI generate double-strand breaks with 5′ and 3′ overhangs respectively (Eddy & Gold, 1991; Drouin *et al.* 2000). Furthermore, I-TevIII and I-CmoeI each appear to have additional domains found in other enzymes. The former contains two putative DNA-binding zinc-finger domains (Eddy & Gold, 1991), while the latter contains a degenerate GIY-YIG sequence (Drouin *et al.* 2000).

3.2 Structure–function studies: I-HmuI

The structure of the I-HmuI homing endonuclease has been determined in complex with its bound cognate DNA target site (Fig. 5) (Shen *et al.* 2004). The structure of the complex spans 25 bp. Approximately 2700 Å^2 of accessible surface area is buried for each molecule. The DNA construct which was synthesized and used in crystallization experiments as two intact oligonucleotides is cleaved on one strand, precisely at the observed site of nicking in biochemical

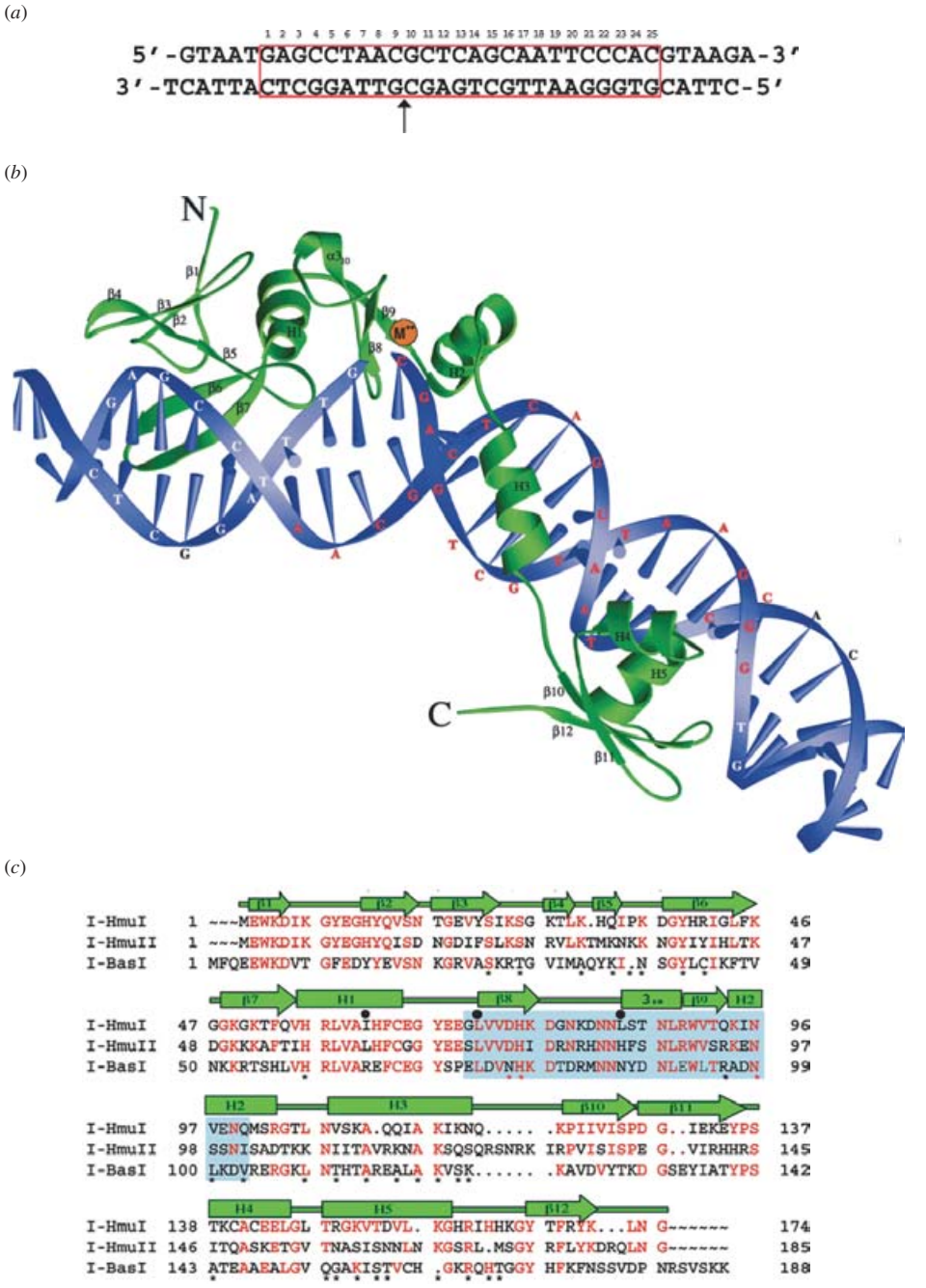


Fig. 5. Structure of the phage HNH homing endonuclease I-HmuI, in complex with its DNA target site. (*a*) The DNA construct used to grow crystals, with the base pairs on contact to the enzyme outlined. The site of cleavage (a single nick) is indicated with the arrow. (*b*) Structure of the full-length enzyme–DNA complex. (*c*) Sequence superposition of homologous enzymes from three closely related phage species, each of which is encoded in a similarly positioned group I intron in the phage DNA polymerase gene. The HNH motif is outlined in purple, residues that contact the DNA in I-HmuI are indicated with asterisks; residues mutated to methionine, in order to experimentally phase the I-HmuI X-ray data are indicated with dots. The secondary structure of I-HmuI is also shown.

assays (Landthaler & Shub, 2003). There is one bound catalytic metal ion associated with the free 5′ phosphate, the corresponding 3′ hydroxyl, and a pair of enzyme side-chains. The DNA displays a distinct bend of $\sim 40^{\circ}$ approximately 4–5 bp downstream to the site of cleavage, corresponding to a significant widening of the minor groove in that region.

The structure of the I-HmuI protein is extraordinarily elongated, with a series of distinct sequential structural domains and motifs distributed along the DNA target (Fig. 5*b*). The structure corresponds well with the length and approximate boundaries of the I-HmuI DNA recognition site as estimated by biochemical protection and interference studies. The protein engages in several different structural modes of interaction with the DNA. The target site is bound at one end within its major groove by an N-terminal antiparallel β-sheet, along its middle seven base pairs within the minor groove by a two α-helices, and at the opposite end within the major groove by a C-terminal HTH motif. In order to form this complex, the polypeptide straddles the phosphate backbone twice, once at the site of cleavage and again nearer to the 3′ end of the target site. The minor groove at the central seven base pairs of the site is significantly widened, in order to permit the deep penetration of side-chains to within direct contact distances to the corresponding base pairs. A single valine side-chain (Val108) from helix 3 is deeply intercalated into the DNA base pairs at the center of the bend. The protein DNA-interaction surface in this region is remarkably hydrophobic, with contacts made by Val97 and Val108, Leu105, and Ala111 and Ala115.

I-HmuI contains a $\beta\beta\alpha$-metal nuclease active site, as described above, interacting non-specifically with the DNA backbone at and near the site of cleavage (Fig. 4). The N-terminal 70 residues of this same domain form an antiparallel β-sheet structure that binds the DNA target site over a contiguous stretch of 9 bp via contacts in the major groove. Unlike the His-Cys box enzyme I-PpoI (Section 4), which is a homodimer that contains a pair of structural zinc ions buried in the interior of each subunit's fold and a bound magnesium ion in its active site, I-HmuI is a monomer and only contains a single bound metal ion in its active site. This metal ion, which is a manganese in the crystal structure, is coordinated by a conserved asparagine residue from the HNN motif, an additional aspartate, and a non-bridging oxygen atom from the scissile phosphate group of the DNA substrate.

While the position of the bound catalytic metal ion is similar between the related HNH/N enzymes visualized to date, the chemical identities and roles of their metal-binding residues differ significantly, with the metal ion variously coordinated by one Asn and one Asp residue (in the HNH endonuclease I-HmuI), a single Asn (in I-PpoI), or a pair of histidine residues (in the bacterial E9 colicin). The differences in metal-binding displayed by these structures may correlate with subtle differences in catalytic mechanisms and metal specificity profiles (recently reviewed by Galburt & Stoddard, 2002). For I-HmuI and the other HNH enzymes, the structure of the active site consistently supports a mechanism where an activated hydroxyl acts as a nucleophile for an in-line displacement, with the metal appropriately positioned to stabilize the phosphoanion transition state and the leaving group (Fig. 6), similar to that described in detail for I-PpoI and the *Serratia* nuclease. Conserved histidine residues in the HNH motif (H75 in I-HmuI; H98 in I-PpoI; His103 in E9 colicin nuclease) appear to play similar roles in each of these enzymes as a general base.

The C-terminal DNA-binding domain of I-HmuI contains two sequential α-helices and a final HTH structural motif. This region of the enzyme collectively resembles the C-terminal region of the GIY-YIG homing endonuclease I-TevI from phage T4 (Fig. 7; also described in Section 5), a similarity that was previously predicted, based on weak sequence homology (Landthaler & Shub,

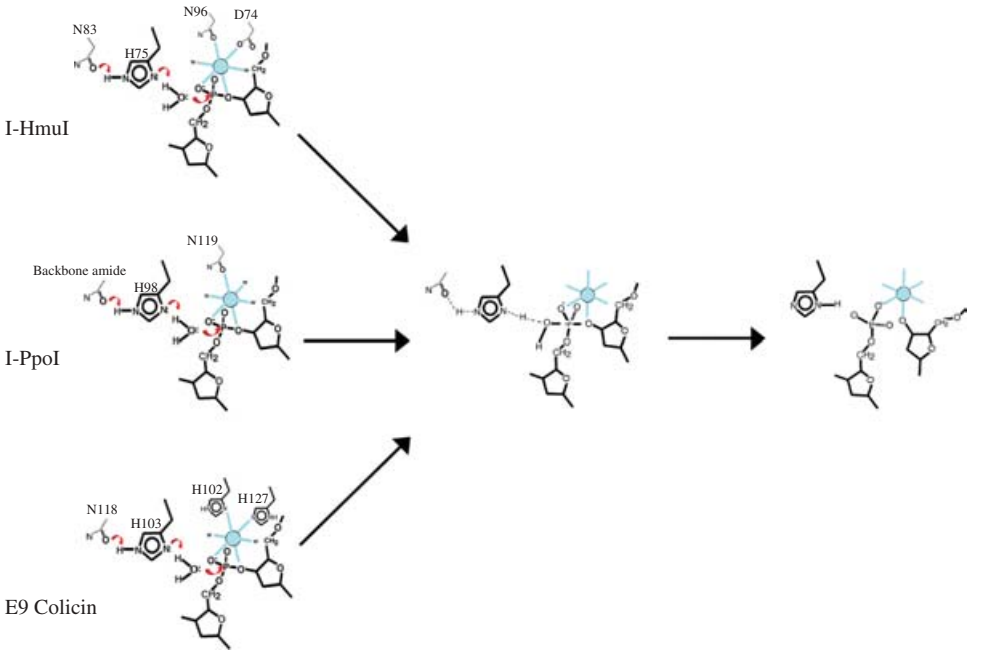
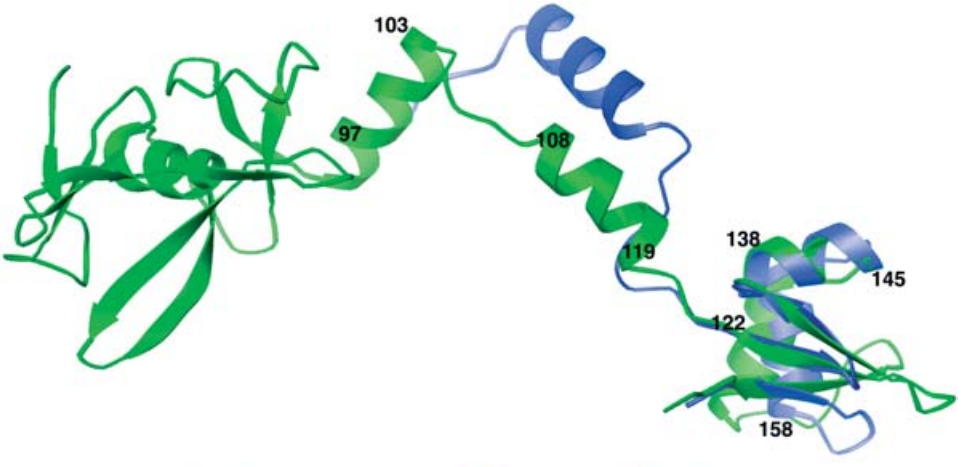


Fig. 6. Consensus mechanism of DNA cleavage for the HNH/His-Cys box/colicin nucleases. Differences in metal-binding contacts in the active site are shown for I-HmuI, I-PpoI and the E9 colicin.




```
I-HmuI    97 VENQMSRGTL NVSKA.QQIA KIKNQ..... .KPIIVISPD G..IEKEYPS  137
I-TevI   184 ~~~~~~~~~~ ~DITKSKISE KMKGKKPSN. IKKISCD... ....GVIFDC  214

I-HmuI   138 TKCACEELGL TRGKVTDVL. KGHRIHHKGY TFRYK...LN G~~~~~~      174
I-TevI   215 AADAARHFKI SSGLVTYRV. KSDKWNW.....FYINA~~~~ ~~~~~~~      243
```

Fig. 7. Structural and sequence alignment of the C-terminal DNA-binding region of the HNH enzyme I-HmuI (green) and the GIY-YIG enzyme I-TevI (purple). Differences in the types of contacts made by these homologs are described in the text.

2003). The extended pair of helices in this region are deeply inserted along the minor groove at the center of the target site, from base pairs 10–16, while the HTH motif interacts with bases in the major groove at the 3′ end of the DNA target site. These motifs are also termed 'NUMOD3' (Nuclease-associated Module, type 3) and 'IENR1' (Intron-Encoded Nuclease Repeat) domains respectively, according to the nomenclature developed for homing endonucleases and their functional relatives (Sitbon & Pietrokovski, 2003). The DNA-binding region at the N-terminus of I-HmuI, similar to that used by I-PpoI as described above, is similarly termed a 'NUMOD4' domain using this nomenclature.

The primary difference between the C-terminal DNA-binding regions of I-HmuI and I-TevI is that whereas for the first enzyme a pair of helices are intercalated deeply into a long stretch of minor groove, in the latter GIY-YIG enzyme the single helix makes primarily non-specific contacts to the phosphate backbone of the DNA duplex, which is relatively unperturbed from a B-form structure. In that enzyme, an additional zinc-finger domain serves as a structural linker between the GIY-YIG catalytic nuclease domain and the C-terminal DNA-binding region, but does not itself participate in DNA recognition (Dean *et al.* 2002). That structural element is replaced in I-HmuI with the additional α-helix, which actually forms part of the HNH/N $\beta\beta\alpha$-metal motif of the nuclease active site. Therefore, the precise site of domain swapping (or sharing) between the HNH and GIY-YIG endonucleases appears to be a tetrapeptide sequence between the two helices (H2 and H3) in I-HmuI that sequentially associate with the minor groove at the center of the DNA target site.

It has been previously reported (Belle *et al.* 2002) that the segF endonuclease displays similarities in primary sequence with the GIY-YIG family at its N-terminus. This enzyme is related to other seg genes in T4, but is most closely related to an HNH ORF (mobD, also in T4 phage) at its C-terminus. In light of the structural analysis of I-HmuI, it appears as if that protein might also be the result of a discrete set of one or more domain swaps. Additionally, the structural motifs present in the I-TevI DNA-binding region appear to differ significantly among related GIY-YIG enzymes (Edgell & Shub, 2001), particularly in the presence or absence of a zinc-finger structural domain, indicating that domain swapping between endonuclease families is being accompanied by rapid divergence and further structural swapping within those same families. This observation appears to reflect a generalizable feature of many nuclease families, which display a high propensity to swap and recombine functional and structural domains in different systems. For example, MutH, lambda exonuclease and the Tn7 transposase share a common ancestor with type II restriction enzymes (Bujnicki, 2000).

Previous quantitative site-recognition studies on LAGLIDADG (Section 3.1) and His-Cys box enzymes (Section 4) indicate that specificity, or information content, at individual base pairs increases linearly as the number of protein hydrogen-bond contacts to those bases increases (Argast *et al.* 1998; Chevalier *et al.* 2003; Gimble *et al.* 2003). The I-HmuI endonuclease (and by analogy, most similar HNH endonucleases) is predicted to exhibit significantly lower overall specificity, in spite of its longer target site, because of the low number and density of sequence-specific contacts distributed over its site. The enzyme makes visible contacts to 14 out of 25 bp (seven via contacts in the minor groove to the central portion of the site; the remaining seven via contacts in the major groove at the more distal ends). Two of these positions, flanking the site of strand cleavage (base pairs 9 and 10) are primarily in contact with enzyme side-chains via a pair of structurally ordered, bridging water molecules, while an adjacent base pair (position 11) exhibits a mixture of direct and water-mediated contacts. In addition, contacts are observed to 12 of the 50 phosphate groups of the duplex backbone. Overall, 52 observed contacts are evenly distributed

between individual atoms of bases and atoms of the DNA backbone, corresponding to two contacts per base pair. Four of the contacts to bases are in the form of intercalated, van der Waals interfaces between aliphatic side-chains (V108 and A111) and base pairs 13 and 14 (the primary site of the bend in the DNA substrate). The large number of contacts in the minor groove of the DNA target site, between base pairs 10 and 16, should also reduce total information content and overall specificity for the complex, due to the fewer differences in chemical properties between base pairs in the minor groove.

The I-HmuI homing endonuclease displays the interesting property of cleaving a target site that contains the inserted mobile intron, in addition to the intronless target site visualized in this study. The insertion of the intron between base pairs 6 and 7 therefore has the effect of altering the sequence of five of the first six base pairs in the target site. However, the contacts formed by the enzyme within this region of the target are almost entirely to the DNA backbone; indicating a mechanism by which I-HmuI can tolerate intron insertion.

Finally, the eight codons of the homing target site correspond to highly conserved residues in the host DNA polymerase amino-acid sequence: E_{595}**PN**A**QQ**F**P**$_{602}$. Interestingly, twice as many of the first and second positions of these codons (6 out of 8 for both positions) are engaged in visible direct interactions with protein side-chains as are the third, more variable 'wobble' bases from the same codons (3 out of 8). It is possible that a homing endonuclease such as I-HmuI, specific for DNA encoding a highly conserved amino-acid sequence, could achieve its balance of selectivity and flexibility by preferentially targeting the first and second positions of codons, while avoiding recognition of bases that are more subject to silent mutations. Experimental demonstrations of the importance of highly conserved coding sequences at homing endonuclease binding and cleavage sites have been made previously: for example, I-BmoI cleavage (but not binding) is dependent on a base pair that is always present as the second position of a universally conserved arginine codon in thymidylate synthases (Edgell *et al.* 2003).

4. His-Cys box homing endonucleases

4.1 Distribution and diversity

As mentioned in Section 3.2, the His-Cys box homing endonucleases are related to the HNH family as a result of divergence from an ancestral nuclease. For all His-Cys box family members identified thus far, the host genes are nuclear rDNA loci from several species of protists. This family is characterized by a series of conserved histidine and cysteine residues distributed over an ~100 amino-acid stretch of the protein (Johansen *et al.* 1993; Muscarella & Vogt, 1993); these residues encompass a diverged variation of the HNH motif. The family includes six enzymes that have demonstrated activity, and 17 putative members encoded by full-length ORFs. The enzymes with demonstrated activity are I-PpoI from *Physarum polycephalum* (Muscarella *et al.* 1990; Wittmayer & Raines, 1996; Wittmayer *et al.* 1998), I-DirI from *Didymium iridis* (Johansen *et al.* 1997) and a set of four highly related proteins (I-NjaI, I-NanI, I-NitI and I-NgrI) from several *Naegleria* species (Elde *et al.* 1999, 2000; Decatur *et al.* 2000). Twenty His-Cys box pseudogenes, which appear to have degenerated from formerly active endonucleases, have also been identified in group I introns found at the same position of rDNA sequences as those encoding active endonucleases (Cho *et al.* 1998; Goddard & Burt, 1999; Haugen *et al.* 1999; Foley *et al.* 2000; Muller *et al.* 2001; Bhattacharya *et al.* 2002; Nozaki *et al.* 2002).

As nicely reviewed in a recent review chapter on this enzyme family (Galburt & Jurica, 2005), the position of the ORFs encoding the His-Cys box homing endonucleases in nuclear eukaryotic rDNA genes, rather than organellar or prokaryotic genomes, raises several questions regarding their expression. In the nuclei of eukaryotes, RNAs destined to be translated into protein are transcribed by RNA polymerase II (Pol II) and subjected to extensive processing (5′ capping, splicing and polyadenylation), whereas rDNA is transcribed by RNA Pol I in the nucleolus and is subjected to unique cleavage and modification events prior to incorporation in ribosomal subunits. Therefore, for many members of this family it is unclear which polymerase generates the translated message and what pre-mRNA processing is necessary for endonuclease expression. This situation is in contrast to those homing endonucleases encoded in organellar introns, where transcription and translation take place in the same compartment, pre-mRNA processing is minimal, and a single RNA polymerase appears responsible for transcription.

Studies of I-PpoI expression from the PpLSU3 intron integrated into the *S. cerevisiae* genomic rDNA showed that I-PpoI is probably expressed from an RNA Pol I transcript, and that expression was not dependent on internal processing of the intron (Lin & Vogt, 1998). It appears that I-PpoI is translated from the full-length excised intron; however, the mechanism of expression for other His-Cys box endonucleases might be more complicated (Vader *et al.* 1999; Decatur *et al.* 2000).

Prior to the structure determination of I-PpoI (Section 4.2), a number of biochemical studies provided details of its basic catalytic features. In agreement with its pattern of conserved histidine and cysteine residues, spectroscopic studies indicated there are one or more bound zinc atoms per protein subunit (Flick *et al.* 1997). The DNA-binding properties of I-PpoI have been well-characterized (Ellison & Vogt, 1993; Wittmayer & Raines, 1996; Wittmayer *et al.* 1998). It binds and cleaves an asymmetric pseudo-palindromic homing site and induces a significant DNA bend. The specific homing site for I-PpoI is long (14 bp) compared to restriction endonucleases, but shorter than the sites recognized by most LAGLIDADG endonucleases. I-PpoI can be activated *in vitro* by several divalent metal ions, including Mg^{2+}, Mn^{2+}, Ca^{2+}, Co^{2+} and Zn^{2+} (Lowery *et al.* 1992). Homing-site binding and cleavage are salt-dependent, with K_d's ranging from 1 to 100 nM in the presence of 10–275 mM NaCl. Cleavage is particularly efficient, with a k_{cat}/K_m of $10^8\ M^{-1}\ s^{-1}$ for I-*Ppo*I. Cleavage occurs at the center of the homing site to generate 4 nt, 3′ overhangs.

4.2 Structure–function studies: I-PpoI

In addition to being the most thoroughly characterized His-Cys box endonuclease using traditional biochemical methods, I-PpoI is the only member of the His-Cys box family of homing endonucleases with a known structure. This small enzyme (163 residues) forms a stable homodimer; X-ray structures have been solved of the apoenzyme and of several complexes bound to DNA (Flick *et al.* 1998; Galburt *et al.* 1999, 2000). I-PpoI displays a fold of mixed α/β topology (Fig. 2) with the dimensions of $25 \times 35 \times 80$ Å. The extended length of the dimer allows the protein to fully interact with its DNA homing site. Each monomer of I-PpoI contains a three-stranded, antiparallel β-sheet flanked by two long α-helices and a long carboxy-terminal tail. The folded structure is stabilized by two buried structural zinc ions that are 15 Å apart in the monomer. The central interface of the enzyme dimer is small and highly solvated, with subunit contacts that bury only 700 $Å^2$ of surface area. The C-terminal tails (residues 146–163) are domain-swapped and extend 34 Å across opposite monomers, burying an additional 900 $Å^2$ per subunit.

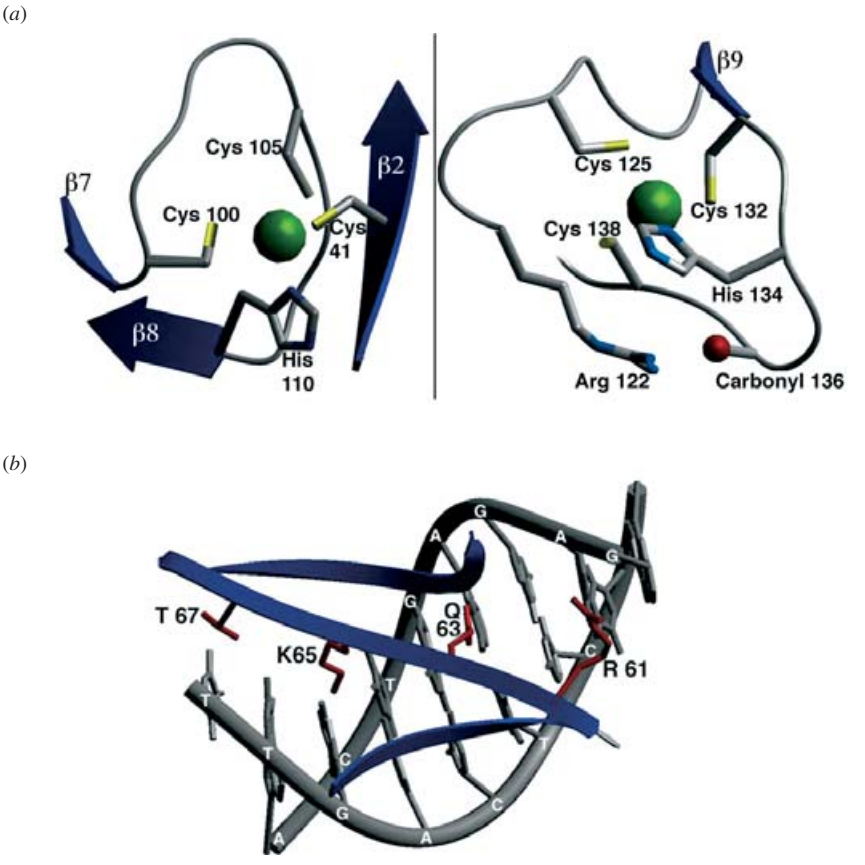


Fig. 8. (*a*) Structural zinc ions in the core of the His-Cys box endonuclease I-PpoI, bound by conserved residues of the enzymes' namesake motif. This structural feature is unique to the His-Cys box branch of the HNH superfamily of enzymes. (*b*) DNA binding and contacts by the I-PpoI antiparallel β-sheet platform. The pattern and strategy of DNA recognition is similar to, but more limited than the use of β-sheets by the LAGLIDADG enzymes.

Eight of the conserved cysteine and histidine residues that give this enzyme family its name are involved in zinc coordination (Fig. 8*a*). In contrast to other zinc-binding motifs in DNA-binding proteins that are primarily associated with DNA recognition, the I-PpoI zinc-binding motifs play a central role in stabilizing the folded structure of the enzyme, localizing the bound metal ions to the protein core.

The remaining conserved residues of the His-Cys box homing endonucleases are positioned in the active site of the enzyme, forming a canonical $\beta\beta\alpha$-metal nuclease motif related to the HNH endonuclease structure described in Section 4.1 (Figs 4*b*, 6). A conserved asparagine coordinates a divalent metal ion (which appears to be magnesium, again similar to the HNH enzyme family), which in turn contacts the 3′ hydroxyl of the cleaved DNA and four bound water molecules. The metal is positioned to interact with the scissile phosphate. A conserved histidine residue (His98) is positioned to activate a water molecule, quite similar to the active site of the HNH endonuclease I-HmuI (where His75 acts as the critical base). Overall, the metal-dependent mechanism of DNA cleavage for the HNH and His-Cys box enzymes appears to be similar.

Outside of the sequence elements noted above, the His-Cys box family is highly divergent. Many of the other His-Cys box enzymes contain an additional long N-terminal extension that make them nearly 50% larger than I-PpoI, and the C-terminal tail of I-PpoI that serves as a dimerization motif is not conserved. As all these enzymes cleave nearly symmetric homing sites (see below for details), it is likely they also dimerize but may have evolved alternate means for achieving a dimeric interface.

4.3 DNA recognition

The characterized His-Cys box endonucleases recognize extended palindromic and pseudopalindromic homing sites (up to 20 bp). Members of the family cleave to generate cohesive 3′ overhangs (similar to the LAGLIDADG enzyme family) indicating cleavage across the minor groove. Also similar to other homing endonuclease systems, their DNA target sites may vary in sequence at many of their individual nucleotide positions while still being recognized and cleaved. Site preference for I-PpoI has been explored by an *in vitro* cleavable site selection strategy (Argast *et al.* 1998), demonstrating that I-PpoI tolerates base-pair substitutions at several positions within the homing site with some positions being more stringently recognized.

Similar to the structure of the nuclease domain of I-HmuI described in Section 4.1 (as well as the LAGLIDADG enzymes), the antiparallel β-sheet in each subunit of I-PpoI forms a folded structure complementary to the major groove of B-form DNA (Fig. 8*b*), and makes a series of side-chain contacts to the DNA bases that represent both direct and water-mediated contacts. However, this motif is more compact than the corresponding DNA-binding surfaces of the LAGLIDADG enzymes, with a single β-strand providing the majority of these contacts. From this strand, alternating side-chains contact the DNA bases of the target site: Arg61 contacts the phosphate backbone near the scissile phosphate, Gln63 makes direct bipartite contact to Ade+6, Lys65 contacts Ade-8 and Gua-9 through water-mediated interactions, and Thr67 contacts the 5′-terminus of the DNA oligonucleotide. A small number of additional contacts are made in this region by the two additional β-strands that flank β4. In addition, one base-specific contact in each half-site is made in the minor groove, between the scissile phosphates.

Endonucleases that cleave across the minor groove of DNA (liberating 3′ overhangs) must cleave two phosphate groups separated by only 10–12 Å in an undistorted B-form double helix. As described above, the LAGLIDADG homing endonucleases have evolved an elegant structural solution to this problem: the conserved LAGLIDADG motif is used to generate a closely packed subunit or domain interface that appropriately positions active-site residues at the center of the active site with minimal separation between them. I-PpoI uses a different strategy: it severely bends the homing-site DNA, making the scissile phosphates more accessible for the two separate enzyme active sites.

At the point of the bend in each half-site, Leu116 points into the minor groove and contacts Ade±2, unstacking these bases from their immediate neighbors (Gua±3). The 55° bend across the central 4 bp of the homing site increases the average width of the minor groove in this region to 10 Å. The major groove across these same base pairs narrows to less that 5 Å, and the scissile phosphates are separated by 22 Å.

4.4 DNA cleavage

Enzymatic cleavage by I-PpoI has been studied both biochemically and structurally. I-PpoI cleaves its target site with a k_{cat}/K_m of 10^8 M^{-1} s^{-1} and is activated by many divalent metal ions

(in order of activity: $Mg^{2+} > Mn^{2+} > Ca^{2+} = Co^{2+} > Ni^{2+} > Zn^{2+}$) (Lowery *et al.* 1992; Ellison & Vogt, 1993; Wittmayer & Raines, 1996). To directly visualize and characterize the mechanism of bond cleavage exhibited by I-PpoI, its structure has been determined in five distinct intermediates on the reaction pathway: one of apoenzyme (Galburt *et al.* 2000), one in a bound substrate complex prior to full DNA bending (Galburt *et al.* 2000), two in an uncleaved precursor state immediately prior to bond cleavage (Galburt & Stoddard, 2002) and a final structure as a bound product complex (Flick *et al.* 1998). These studies allow a quite thorough description of cleavage mechanism (Fig. 6) that is relevant to the broader collection of HNH endonucleases, from which I-PpoI and the other His-Cys box enzymes are diverged.

A single divalent (Mg^{2+}) cation is bound to the protein–DNA complex in each active site of the protein dimer. The bound magnesium ion is coordinated in a six-fold geometry by the side-chain oxygen of Agn119, the bridging 3′ oxygen and a single non-bridging oxygen atom of the scissile phosphate, and three well-ordered water molecules. The average metal-oxygen ligand bond distance is 2·5 Å. This distance is slightly longer (by ~0·2 Å) than typical magnesium bond distances observed in a more highly charged metal-binding site, and is indicative of the relatively neutral character of the metal ligands in I-PpoI.

The single bound metal ion in I-PpoI appears to serve three distinct roles in catalysis. Direct interaction of the bound metal ion with the scissile phosphate indicates that magnesium stabilizes the phosphoanion intermediate and the 3′ hydroxylate leaving group. Second, a water molecule in the inner coordination sphere of the metal is appropriately positioned to donate a proton to the 3′ hydroxylate leaving group. The metal ion decreases the pK_a of this water molecule and accelerates proton transfer. Finally, the bound metal forms a geometrically strained octahedral complex with surrounding protein, DNA and solvent atoms that is relaxed after DNA bond cleavage. The basic details of this mechanism are in good agreement with biochemical studies of cleavage by I-PpoI from several independent laboratories (Friedhoff *et al.* 1999a; Mannino *et al.* 1999); these studies also indicate that this mechanism is more broadly exploited by additional microbial nuclease families (Friedhoff *et al.* 1999a; Miller *et al.* 1999).

In the uncleaved enzyme–substrate complexes a water molecule is positioned for in-line attack on the scissile phosphate, ~3·9 Å from the phosphorous atom. The δN of His98 appears to be directly hydrogen-bonded to the water molecule, may to act as a general base and may also participate in stabilization of the phosphoanion transition-state.

Because the observed nucleophilic water molecule is not associated with a bound cation or any other electrophilic group, its pK_a is likely to be higher than a metal-bound water nucleophile. Because the pK_a of an uncharged histidine residue is only ~6, it would seem likely that a side-chain acting as an activating base must be itself activated through an interaction with a hydrogen-bond acceptor. In I-PpoI the backbone carboxylate oxygen of Cys105 is 2·8 Å from the His 98 εN, and is positioned to form a linear hydrogen bond. In the structure of the HNH endonuclease I-HmuI described previously (Shen *et al.* 2004), the histidine base is similarly coordinated by the side-chain amide oxygen of Asn83, and in the *Serratia* nuclease the putative general base (His89) displays a similar interaction between its εN and Asn106 (Miller *et al.* 1999).

A series of conformational changes are observed in the active site as a result of DNA bond cleavage. The free 5′-phosphate moves by over 2·5 Å from its position in the substrate complex and forms a 2·8 Å electrostatic bond with a guanido-nitrogen of Arg61, which moves by ~0·5 Å. The movement of the 5′-phosphate disrupts the interaction between its non-bridging oxygen and the bound metal ion. A fourth well-resolved water molecule is added to the inner metal coordination sphere, which assumes a more ideal octahedral geometry. The previously

strained metal bond angles relax by $\sim 20^{\circ}$, and the average orthogonal coordination angle for the ligands is $89{\cdot}6^{\circ}$ ($\sigma = 5{\cdot}3$). The metal ion does not move significantly upon cleavage, and maintains interactions with Asn119 and the 3′ oxygen leaving group of the cleaved phosphodiester bond.

These structures indicate that the phosphoanion transition state is stabilized through contacts with the bound metal ion and the imidazole ring of His98. This contact exists in the E-S complex as a polar interaction with the hydrolytic water molecule and is maintained in the free 5′-PO_3 group of the E-P complex. Arg61 does not appear to play a role in transition state stabilization, because the distance from this side-chain to the scissile phosphate prior to bond cleavage is too long, at 5·5 Å. Arg61 does, however, appear to stabilize the final product complex, and thus may help to drive the equilibrium of the reaction forward by inhibiting re-ligation.

5. GIY-YIG homing endonucleases

5.1 Distribution and diversity

This family of endonucleases is characterized by a conserved GIY-($X_{10\text{-}11}$)-YIG motif (Kowalski *et al.* 1999). GIY-YIG endonucleases have been found in the T4 bacteriophage both as free-standing enzymes (F-TevI, F-TevII) (Sharma *et al.* 1992) and within mobile group I introns (I-TevI, I-TevII) (Bell-Pedersen *et al.* 1990). I-TevI, which is encoded within a mobile intron embedded in the thymidylate synthase gene of bacteriophage T4 (Bell-Pedersen *et al.* 1991), has been extensively characterized by NMR and crystallographic analyses, and is described in detail below.

GIY-YIG ORFs have also been reported in introns of fungal mitochondria (Tian *et al.* 1991; Paquin *et al.* 1994; Saguez *et al.* 2000), algal mitochondria (Kroymann & Zetsche, 1997; Denovan-Wright *et al.* 1998) and algal chloroplasts (Paquin *et al.* 1995; Holloway *et al.* 1999). Beyond their namesake signature, several additional conserved motifs are also involved in the folding and activity of their N-terminal catalytic domains (Fig. 9*a*). This nuclease core is tethered to a C-terminal DNA-binding region; the architecture and domain organization of this portion of the enzyme is highly variable across the GIY-YIG superfamily (Derbyshire *et al.* 1997; Kowalski *et al.* 1999). Beyond the known homing endonucleases and their closely related free-standing homologs, the GIY-YIG signature has also been found in a small number of additional DNA-acting enzymes, including restriction endonucleases and repair systems (Kowalski *et al.* 1999).

Biochemical studies of the 28 kDa I-TevI and the 30 kDa I-TevII reveal these monomeric enzymes to recognize long homing sites (37 bp and 31 bp respectively) and cleave their DNA many bases away from the intron insertion site (Derbyshire *et al.* 1997). Both enzymes bind primarily across the minor groove and phosphate backbone and cleave their DNA substrates to leave 2-bp 3′ overhangs. Furthermore, I-TevI is extremely tolerant of base-pair changes in its homing site. No specific bases are essential for activity and a variety of insertions and/or deletions between the cleavage and insertion sites are permitted.

Limited proteolysis and footprinting experiments have shown I-TevI to be a bipartite enzyme with distinct N-terminal nuclease and C-terminal DNA binding domains separated by a long flexible linker (Fig. 9*b*). The linker region contains a unique zinc-finger domain that is important for relative organization and orientation of the separate enzyme domains on the DNA molecule. The site of cleavage, which is engaged by the N-terminal nuclease domain in the complex, is

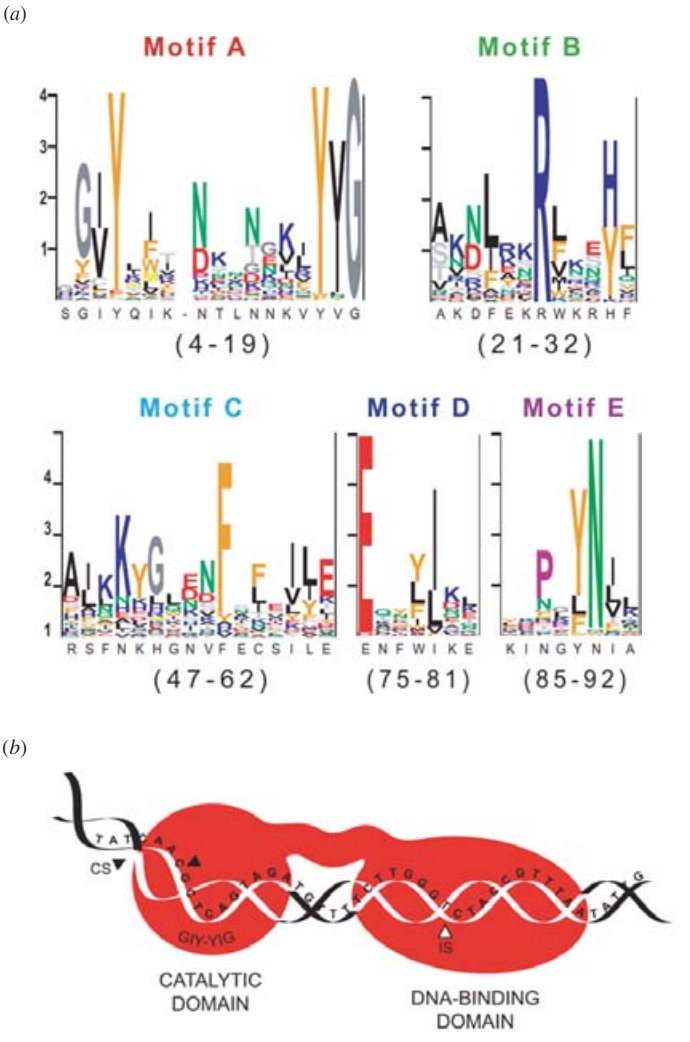


Fig. 9. (*a*) Conserved sequence motifs across the GIY-YIG endonuclease catalytic domain. (*b*) Schematic of the bipartite DNA binding strategy employed by the I-TevI endonuclease. [Figure reproduced from *Homing Endonucleases and Inteins* (Springer-Verlag) with permission of the author (V. Derbyshire).]

located near the 5′ end of the 37-bp site. The position of the intron insertion site is nearer to the 3′ end of the site, within the region of DNA sequence that is bound by the C-terminal DNA-binding domain. The overall DNA-binding specificity of the enzyme is dictated by this region of the enzyme, as is its binding affinity (the isolated C-terminal region of the enzyme binds its DNA target with the same affinity as the full-length enzyme to the corresponding full-length DNA homing site). In contrast, the nuclease domain does not independently recognize its cleavage target, and does not significantly contribute to overall binding affinity in the complex of the full-length enzyme with its target site. The nuclease domain does, however, impart some limited target sequence preferences near the cleavage site, that are observed in site cleavage selection experiments conducted with the full-length enzyme (Bryk *et al.* 1995; Edgell *et al.* 2004b).

The enzyme can bind and cleave substrates with a limited range of insertions and deletions between the 3′ binding site and the 5′ cleavage site. The presence of the zinc-finger domain appears to restrict the ability of the enzyme to tolerate even larger variations in this region of the target site, presumably by acting as a structural cassette that limits free rotation and extension of the nuclease domain in the full-length enzyme (Dean *et al.* 2002). This feature appears to be somewhat similar to interdomain contacts in the bipartite restriction endonuclease FokI (Wah *et al.* 1997), which limit the conformational and catalytic freedom of the nuclease domain in the full-length enzyme.

It appears that there is very little sequence similarity outside of the GIY-YIG module among the members of the family, except for proteins highly related to I-TevI, such as I-BmoI, an intron-encoded enzyme from *Bacillus mojaversis* (Edgell & Shub, 2001). The latter two enzymes can be aligned through both their GIY-YIG domains and through parts of their DNA-binding domains. I-BmoI has a HTH domain analogous to that of I-TevI, but does not have a zinc-finger domain. Instead, it appears to have three copies of the minor-groove binding α-helix present in I-TevI (Sitbon & Pietrokovski, 2003).

In vivo experiments have shown that I-TevI can repress its own expression (Gott *et al.* 1988). *In vitro* experiments have demonstrated that the DNA-binding domain of I-TevI binds to this operator sequence with the same affinity as to the homing site, even though there are six substitutions in the enzyme's primary recognition sequence. However, I-TevI cleaves its operator site very poorly, relative to its homing site target. The crystal structure of the I-TevI DNA-binding domain bound to the operator sequence demonstrated that the DNA structure is largely unchanged in this complex, relative to the homing site, except for small differences in the regions where the DNA interacts with the extended regions between the subdomains (Edgell *et al.* 2004a).

5.2 Structure–function studies: I-TevI

While the structure of a full-length GIY-YIG homing endonuclease has not been determined, extensive structural studies have been reported of both the catalytic domain in the absence of DNA (Kowalski *et al.* 1999; VanRoey *et al.* 2002) and of the C-terminal binding region in complex with its DNA homing target (VanRoey *et al.* 2001) and its closely related operator sequence, as mentioned above (Edgell *et al.* 2004a).

NMR studies of the catalytic domain of I-TevI (Kowalski *et al.* 1999) initially revealed this structure to have a mixed α/β topology with the GIY-YIG residues located in a three-stranded β-sheet. Two highly conserved residues (Arg27 and Glu75) reside in the α-helices and are required for catalytic activity. The subsequent crystal structure of this domain (VanRoey *et al.* 2002), conducted using constructs containing inactive point mutations at residues 27 and 75, greatly expanded the resolution of these studies. These analyses provided a clear depiction of a novel nuclease fold, consisting of a twisted three-stranded antiparallel β-sheet flanked by two α-helices on one side, and a third α-helix on the other (Fig. 10). The GIY and YIG sequences are located on the first two β-strands of the core fold respectively. The GIY residues are located in the enzyme core and appear to serve a structural role, while the YIG residues (Y_{18}-V_{19}-G_{20} in this enzyme) appear to be involved both in structural stabilization and in catalysis. The side-chain of Val18 is buried in the core of the domain, while Tyr17 and Gly19 are exposed on a surface of the domain that also encompasses residues from the conserved motifs. This surface is thought to be the site of DNA interaction and cleavage.

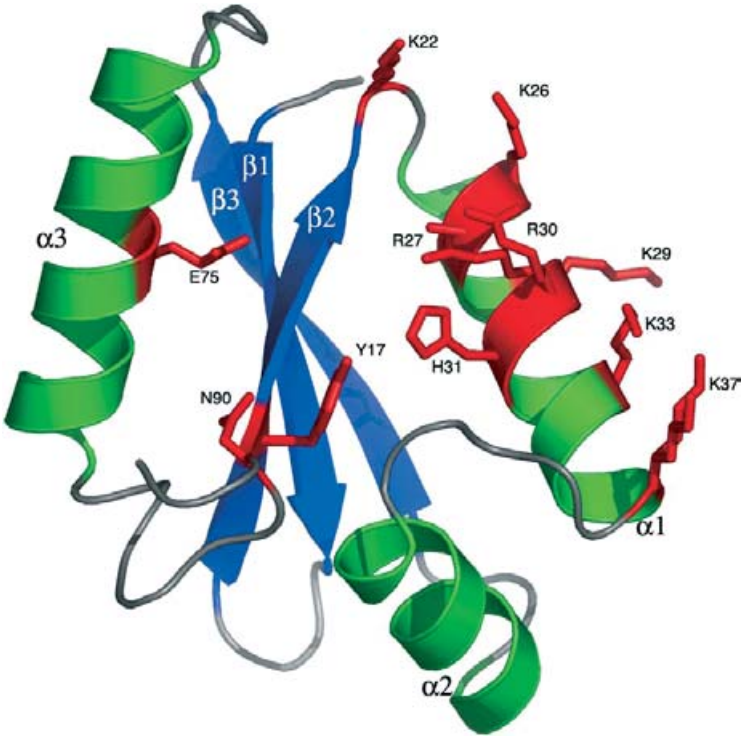


Fig. 10. Structure of the GIY-YIG catalytic domain, with side-chains shown that are conserved and/or known to participate in DNA binding and catalysis. The catalytic triad of E75, R27 and Y17 are similarly positioned to the residues N119, R55 and H98 in the I-PpoI endonuclease, as discussed in the text.

The relative positions of Arg27 and Glu75 (the two conserved residues that are the sites of inactivating point mutations mentioned above) and Tyr17 from the 'YIG' signature are quite similar to conserved active-site residues of the His-Cys box enzyme I-PpoI (corresponding to Arg61, Asn119 and His98 respectively). In addition, the position of bound Mn^{2+} in I-TevI is nearly identical to the site of the same bound metal ion in I-PpoI ternary complexes. This similarity implies that the GIY-YIG nuclease domain might utilize a similar DNA strand cleavage mechanism as the His-Cys box enzymes, as well as the broader collection of HNH nucleases. However, while there is coincidence of key catalytic residues between I-PpoI and I-TevI, the local fold of I-TevI does not correspond to the $\beta\beta\alpha$-Me structural motif that is shared by the His-Cys box and HNH homing endonucleases, consistent with the GIY-YIG enzymes forming an independent fold family, which has possibly converged on a similar catalytic mechanism.

Crystallographic studies of the C-terminal DNA-binding region of I-TevI bound to its 20-bp DNA site (VanRoey *et al.* 2001) demonstrate it to consist of three separate DNA-binding sub-domains: a zinc finger, followed by an α-helix, and then a HTH motif (Fig. 2). The latter two elements are similar to the same elements in the C-terminal region of the I-HmuI endonuclease, as described earlier in Section 3 (Fig. 7). These domains adopt an extended structure that is wound along the DNA molecule, making a combination of contacts to the phosphate backbone and limited hydrophobic contacts to individual bases. A limited number of additional base contacts are made with residues in the regions of peptide that serve as linkers between the I-TevI

subdomains. With the exception of the HTH domain, which places a helix in the major groove, the protein occupies the minor groove and does not greatly alter the DNA conformation (in contrast, the DNA target of I-HmuI is significantly bent in response to the intercalation of a protein helix in the DNA minor groove).

As mentioned above, the catalytic domain and the DNA-binding region are connected by an extensive linker and a zinc-finger domain; the latter element makes a small number of contacts to the DNA backbone. This portion of the molecule acts as structural organizing domain that bridges the nuclease and C-terminal domains (Dean *et al.* 2002). Deletion of the zinc finger causes the full-length enzyme to become more tolerant to changes in distance between the cleavage site and the 3′ DNA binding site. This implies the zinc finger might act as a structural spacer or organizer, helping to structure the adjoining linker region, orient the catalytic domain on its cognate site, and influence the distance between the primary DNA binding site and the position of phosphate cleavage.

6. Cyanobacterial introns and their homing endonucleases

In addition to initiating intron homing (movement into a cognate intronless gene) and transposition (transfer into a new, ectopic site), homing endonucleases are responsible for the persistence of intervening sequences at locations that would otherwise result in their rapid elimination. This is most readily observed in prokaryotic genomes, which are thought to respond to pressure to remain small (and thus minimize the time required for genomic replication) by streamlining their coding content. Thus, the conservation of introns in phage genomes appears to be caused by the presence of active homing endonuclease genes within a high percentage of the members of individual lineages of those introns. As described in Section 1 of this review, this can serve to ensure continual replacement of non-mobile genes with invasive cousins during phage infections and life cycles.

In addition to their presence in bacteriophage, genomic and phylogenetic analyses demonstrate the presence, persistence, and horizontal transmission of mobile group I introns in cyanobacterial tRNA genes, generally at positions corresponding to their anticodon loops. The best studied of these systems is an ORF, encoding an endonuclease, embedded in a group I intron which is found in the $tRNA^{fMet}$ genes of two closely related cyanobacterial isolates, *Synechocystis* PCC 6803 (now named ‘I-SspI’, according to the accepted nomenclature) and PCC 6906 (Biniszkiewicz *et al.* 1994). The reading frames of these endonucleases correspond to small polypeptide chains of ∼150 codons; their 5′ and 3′ ends comprise elements of the P1 stem and P2 stem-loop sequences of the intron respectively. An additional, homologous gene in an intergenic region in T7 phage (termed ‘gene 5.3′) is clearly related to these reading frames. Although no meaningful clues as to structure–function relationships for these intron-encoded proteins can be elucidated from sequence-based alignments with annotated proteins, recent analyses indicate that these enzymes might be distantly related to a broad family of archael Holliday junction resolvase enzymes (B. L. Stoddard, unpublished data).

The physiological DNA homing site for I-SspI is a pseudopalindromic sequence (5′-CGTCGGGCTCATAACCCGAAGG-3′) of 20–22 bp, cleaved to generate complementary 3-base, 3′ cohesive overhangs of sequence 5′-CAT-3′ and 5′-ATG-3′ (Bonocora & Shub, 2001). The site of cleavage of the coding strand of the host gene corresponds precisely to the resulting intron insertion site. Within the DNA target sequence, 14 of the 22 bp share palindromic

symmetry with their counterparts in the opposing DNA half-site, strongly implying that the enzyme acts as a homodimer and displays flexible (reduced fidelity) recognition of several base pairs in the target. These features of recognition are similar to those described for the LAGLIDADG and His-Cys box enzymes described earlier.

7. Intron splicing (maturase) activity by homing endonucleases

7.1 Distribution and diversity

A variety of *in vitro* studies have demonstrated that many group I introns fold inefficiently (Treiber & Williamson, 2001), and only self-splice under non-physiological conditions such very high concentrations of salt or divalent cations (Coetzee *et al.* 1994; Ho *et al.* 1997). *In vivo*, many if not all of these introns are associated with tightly bound proteins that function to promote a stable and active RNA conformation (Lambowitz & Perlman, 1990; Lambowitz *et al.* 1999). Such splicing cofactors include various *nuclear*-encoded protein that sometimes have an additional RNA-related activity (such as a tRNA synthetase that moonlights as a splicing factor), and *maturase* proteins that are encoded within the introns themselves. These proteins do not cleave the RNA directly; instead they aid in folding the RNA into a conformation that favors RNA-catalyzed splicing.

Two nuclear-encoded group I intron splicing factors have been extensively characterized. The yeast CBP2 protein specifically promotes splicing of the fifth intron (bi5) in the mitochondrial cytochrome *b* (COB) gene (McGraw & Tzagoloff, 1983). In contrast, the *Neurospora* mitochondrial tyrosyl tRNA synthetase (termed CYT-18) promotes splicing of a wide variety of group I introns (Mannella *et al.* 1979; Walweber *et al.* 1997). These two unrelated proteins use quite different mechanisms to bind their RNA target and facilitate splicing. CBP2 promotes splicing by a 'tertiary capture' mechanism, where an unstable but significantly folded RNA structure is captured and stabilized (McGraw & Tzagoloff, 1983; Weeks & Cech, 1995a, b, 1996). CYT-18 uses a different mechanism where the protein binds to an early intermediate in the intron folding pathway (via a stable P4-P6 RNA subdomain) and then acts as a scaffold for the nucleation and assembly of the complete folded intron structure (Saldanha *et al.* 1995, 1996; Caprara *et al.* 1996a, b, 2001).

The fact that intron-encoded proteins could be critical for post-transcriptional splicing was first demonstrated through genetic and sequencing analysis of three group I introns in the *S. cerevisiae* cytochrome *b* (COB) gene. Based on a variety of analyses, it seems likely that RNA maturase activity is a secondary adaptation of LAGLIDADG homing endonucleases (for further discussion see Lambowitz & Belfort, 1993).

The precise relationship of DNA- and RNA-binding activities and surfaces used by these proteins has not been determined. The majority of known group I maturases are members of the LAGLIDADG protein family, many of which are bifunctional (acting both as homing endonucleases and as RNA-splicing maturases) (Belfort & Roberts, 1997; Chevalier & Stoddard, 2001). These proteins promote splicing of the cognate intron encoding them, and in one case a very close intron relative (Lambowitz *et al.* 1999). Introns that require maturases for splicing do not share obvious similarities in sequence or predicted secondary structure, suggesting that maturases recognize a variety of diverse features of their own unspliced RNA targets. In addition, at least one intron has recently been shown to require *both* a nuclear-encoded protein and an intron-encoded maturase to be spliced *in vitro* (Bassi *et al.* 2002).

7.2 I-Ani and the bI3 maturase: model systems

The I-AniI protein is a monomeric 250-residue LAGLIDADG protein encoded within a group I intron in the *A. nidulans* mitochondrial apocytochrome *b* gene; the protein displays both endonuclease and RNA maturase activity (Ho *et al.* 1997). In its role as a splicing cofactor, the protein operates independently and does not require additional protein factors to facilitate intron splicing. In contrast, the third intron of the *S. cerevisiae* cytochrome *b* gene (the bI3 group I intron) requires two separate proteins to fold into a stable, splicing competent conformation: a LAGLIDADG maturase (termed the bI3 maturase) and the nuclear-encoded Mrs1 protein (Bassi *et al.* 2002; Bassi & Weeks, 2003). The maturase component of this more complex system is no longer a functional DNA endonuclease.

7.2.1 I-AniI

The cognate intron containing the gene that encodes I-AniI does not exhibit self-splicing activity under physiological conditions. RNase protection assays indicate that the intron is incapable of forming a stable, fully reactive folded conformation in the absence of a protein cofactor (Waring *et al.* 1982). This is consistent with the apparent lack of a full complement of RNA sequence signatures indicative of stabilizing tertiary interactions between the P4-P6 domain and the intron core domain, such as the tetraloop and its receptor site and the P5abc three-way junction observed in the tetrahymena group I intron (Ho & Waring, 1999). I-AniI binds its cognate intron with a picomolar dissociation constant, similar to intron-binding affinities measured for the CBP2 and CYT-18 proteins (Solem *et al.* 2002). Binding of I-AniI reduces dependence on magnesium for splicing and causes a 30-fold rate enhancement of splicing (to 2·1 min^{-1}) relative to optimal self-splicing conditions (Ho & Waring, 1999; Solem *et al.* 2002).

I-AniI binds its cognate intron with a 1:1 stoichiometry and forms a very stable RNP complex (Ho & Waring, 1999; Solem *et al.* 2002). Association kinetics of protein–RNA complex formation and splicing revealed that I-AniI binding is a multi-step process that includes an intermediate 'encounter' complex that is relatively slowly resolved into a native, splicing competent complex (Solem *et al.* 2002). Deletion analyses indicate that I-AniI preassociates in a 1:1 stoichiometric ratio with an unfolded COB intron and facilitates correct folding of the intron catalytic core (Solem *et al.* 2002). Tight RNA-binding requires most of the intact intron (Geese & Waring, 2001), while deletion of sequences and secondary structure near the 5′ and 3′ splice sites had minimal effects on binding (Solem *et al.* 2002). These results suggest that I-AniI binds to regions of the COB pre-RNA that are far apart in primary sequence but brought together in close proximity in the folded core (Geese & Waring, 2001; Solem *et al.* 2002).

Crystallographic analysis has shown that the self-splicing Tetrahymena LSU intron (TtLSU), lacking regions P1-P2, P9.1 and P9.2, has major dimensions of 110 and 65 Å (Golden *et al.* 1998). Electron microscopy of the entire intron revealed a globular molecule with a diameter of ∼116 Å (Wang *et al.* 1994). In contrast, the core of the I-AniI cognate intron has ∼25% fewer nucleotides than TtLSU. Deletion analysis argues against recognition of any specific subdomain of the RNA; however very little of the intron can be deleted without significantly impairing binding and splicing (Geese & Waring, 2001; Solem *et al.* 2002). This indicates that the problem of stabilizing the intron has been solved predominantly by recognition of a region or regions of mostly RNA tertiary structure. The question then arises as to whether the RNA has adapted structurally to the protein, vice versa or a combination of both.

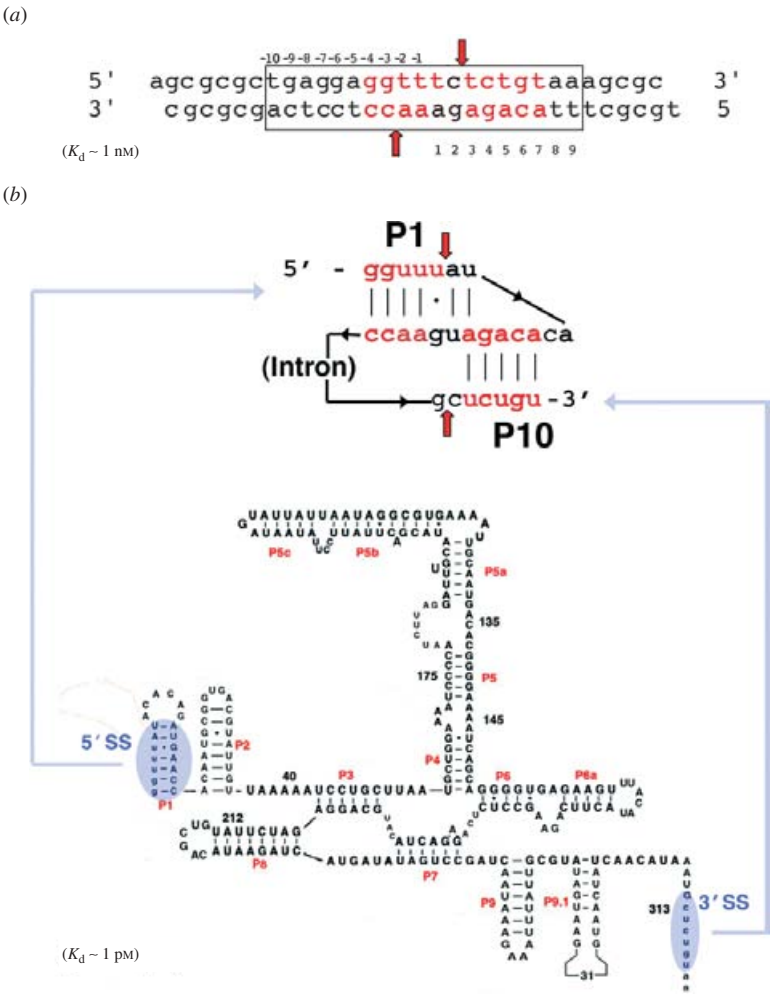


Fig. 11. Sequence and secondary structure of the DNA (*a*) and RNA (*b*) binding targets of the I-AniI homing endonuclease/maturase. The enzyme binds these constructs with nanomolar and picomolar dissociation constants, as shown. Although the sequence of the DNA half-sites bears significant resemblance to the P1/P10 splice site boundaries in the RNA intron active site, experimental evidence indicates that the protein uses distinct surfaces to achieve target recognition and binding, as discussed in the text.

The similarity between respective DNA-binding sites and the RNA structure near the 5′ and 3′ splice sites of the intron (Fig. 11) suggested that the same protein surface, using analogous sets of interactions, is used to perform both functions. However, because competition experiments indicate that I-AniI binds either the wild-type COB intron RNA, or a low-affinity intron variant in the presence of saturating concentrations of its DNA target site substrate, separate binding sites for DNA and RNA are suggested (Chatterjee *et al.* 2003; Geese *et al.* 2003). In addition, two different DNA-binding mutants of I-AniI (in which the conserved LAGLIDADG acidic residues were individually changed to alanine) have little effect on the protein's RNA maturation activity (Chatterjee *et al.* 2003).

A co-crystal of I-AniI endonuclease/maturase complexed with its DNA substrate (Fig. 3*b*) has been solved (Bolduc *et al.* 2003). Overall, the structure of I-AniI is similar to those of other

LAGLIDADG endonucleases. Subsequent mutational analyses of I-AniI, based on this structure, have provided clues to the location of the intron RNA-binding site on that protein. Alignment of I-AniI with two of its closest maturase relatives (encoded by the *S. cerevisiae* bI3 and *V. inaequalis* bI5 introns) revealed conserved basic residues that fall on one side of I-AniI not in contact with the DNA (Bolduc *et al.* 2003). Subsequently, a single arginine to glutamine substitution on this surface of the C-terminus of I-AniI was shown to reduce intron binding and splicing ~10-fold while having no effects on DNA substrate hydrolysis or affinity (Bolduc *et al.* 2003). Furthermore, a C-terminal fragment of I-AniI that begins at the second LAGLIDADG motif does not cleave substrate DNA, as expected, but does facilitate splicing of COB intron almost as efficiently as the full-length protein (Downing *et al.* 2005). The corresponding N-terminal fragment, beginning at the first and ending at the second motif, has no splicing activity.

7.2.2 The bI3 maturase

Many important features of RNA recognition and protein-facilitated intron splicing for the bI3 intron are different from those described above for the COB intron of *Aspergillus*. As previously mentioned the intron requires both an internally encoded LAGLIDADG maturase protein and an auxiliary, nuclear-encoded Mrs1 protein (a divergent member of the RNAse H superfamily of DNA junction-resolving enzymes) to form a splicing-competent fold. The resulting RNP particle, with a mass of 420 kDa, includes the intron RNA, a single maturase monomer, and two dimers of Mrs1. The assembly pathway of this complex, and the kinetics of individual binding events and subsequent splicing reactions, have been thoroughly characterized (Bassi & Weeks, 2003).

The crystal structure of the bI3 maturase has been determined in the absence of bound nucleic acid targets (Longo *et al.* 2005), and the resulting model compared to biochemical data that define the RNA- and DNA-binding surfaces of the protein. The RNA surface that is recognized by the enzyme has been further defined by chemical probing experiments. The protein, which is extremely similar to previously visualized members of the LAGLIDADG family, binds at a distal, non-conserved structure in the peripheral P5-P4-P6 domain of the intron, where it appears to interact primarily with bases in RNA minor grooves, as well as with the intron phosphate backbone. Cross-linking experiments with the maturase indicate that the β-saddle surface of the N-terminal domain of the protein (which is also involved in DNA binding) is extensively involved in RNA binding. These results are in contrast with the I-AniI enzyme, which appears to primarily associate with regions of its cognate group I intron that are near the RNA active site and splice junctions, using a surface on the C-terminal domain that is distant from the DNA-binding β-saddle surface. The striking divergence of RNA-binding mechanisms for these two systems, which use highly similar protein scaffolds, again reinforces the general impression that RNA-splicing activities tend to arise subsequent to DNA cleavage and homing activities.

8. Engineering and directed evolution of homing endonucleases

8.1 Applications and considerations

A variety of groups have postulated that engineered, gene-specific DNA-modification enzymes, such as homing endonucleases or tandem zinc-finger nuclease (ZFN) chimeras, might be used

to drive targeted recombination and gene repair events (reviewed in Uil *et al.* 2003; Wickelgren, 2003). Such reagents might potentially be used, in place of retroviral integrases, as the catalytic machinery that drives highly specific gene therapy applications, including repair of endogenous gene function, as opposed to non-specific 'gene augmentation'. A proof of principle demonstration of this idea, using the ZFN strategy, has been described (Urnov *et al.* 2005). In that study, the gene encoding the 'common gamma chain' (IL2rγ), which is mutated in X-linked severe combined immunodeficiency disorder (XSCID), was efficiently repaired in cultured cells by transiently introducing an engineered ZFN and a corrective allele of the mutant gene.

Such a therapeutic strategy appears to be an attractive concept. Replacement of a mutated endogenous gene might ensure the correct stoichiometric expression of the protein and its various splice variants. Perhaps more importantly, corrective endogenous gene targeting, unlike approaches that introduce additional gene copies, might avoid dysregulation due to lack of endogenous promoter control, loss of function due to epigenetic modifications such as methylation, and oncogenic cell transformation due to non-specific integration or bystander gene activation. These problems have all plagued gene therapy trials (Nabel, 2004).

In addition to gene therapy applications, Burt and colleagues (Burt, 2003) have proposed that engineered homing endonucleases might be used to manipulate natural populations, by driving a gene conversion event leading to incorporation and spread of a desirable phenotype in a non-Mendelian dominant manner. In modeling experiments, complete gene replacement throughout a population can be imposed in less than 20 generations. A proof of principle for this type of experiment has been reported, through a study in which a naturally occurring homing endonuclease gene from yeast (I-SceI) was modified and introduced into *Drosophila*, and then shown to be active and able to cut chromosomal DNA at the appropriate target sequence (Gong & Golic, 2003). As an example of a possible application, a variety of groups are engaged in efforts to control the *Anopheles gambiae* mosquito vector (responsible for spread of the malaria pathogen, *Plasmodium falciparum*) by genetically engineering insects that are refractory to infection or reproductively disabled (Hahn & Nuzhdin, 2004). Attempts to use well-established genetic conversion mechanisms such as transposons to progressively knock out or reduce gene functions in model insect populations have encountered difficulties due to low efficiency of genetic transfer, a lack of suitable target genes, and the generally low specificity of inter- and intragenic transfer of these elements (Ito *et al.* 2002). However, the recent identification of high quality *Anopheles* targets that are essential for *Plasmodium* propagation (Osta *et al.* 2004) offers the possibility of exploiting two selective pressures that favor spread of an inactivating homing endonuclease gene into those loci: the inherent mobility of the endonuclease itself, and the superior fitness of insects that are refractory to infection (Hahn & Nuzhdin, 2004).

8.2 Alternative systems

As described above, non-specific nuclease domains have been tethered to sequence-specific DNA binding modules such as zinc fingers (Smith *et al.* 1999, 2000) and used *in vivo* to stimulate homologous recombination (Bibikova *et al.* 2001, 2003) and more recently to drive sequence correction of a disease-causing allele associated with a severe genetic disorder (Urnov *et al.* 2005). Thus, such constructs show enormous potential for the creation of gene-specific reagents. Despite the relative ease of designing such highly specific ZFN reagents, however, comparison of several of their properties to those of homing endonucleases indicate that both are worthy of

continued development and assessment for gene-specific applications. For example, the nuclease domains of ZFN constructs, which usually are allosterically regulated by inhibitory contacts with their DNA-binding domains, appear to display significant non-specific DNA nicking and cleaving activity in the engineered chimeras. In addition, these constructs can generate multiple adjacent phosphate cleavage events within a single bound DNA target site, which may enhance non-conservative break repair outcomes (Smith *et al.* 1999). In contrast, the cleavage activity of the LAGLIDADG homing endonucleases are tightly coupled to cognate site binding, and the action of these enzymes, by virtue of their tight product binding properties, appear to strongly enhance the ratio of homologous recombination events relative to undesirable, non-conservative double-strand break repair events such as non-homologous end-joining and single-strand annealing.

Additionally, current ZFN chimeras require expression of two separate chains to generate double-strand breaks (Smith *et al.* 2000; Bibikova *et al.* 2003), and more total coding sequence to generate the active enzyme, than do group I homing endonucleases in order to target DNA sites of similar length. This latter issue may be an issue for those requirements where packaging of the endonuclease and the corrective allele into a delivery vehicle must be accomplished.

A variety of efforts have been described in the literature to increase the inherent specificity of type II restriction endonucleases for similar applications. However, efforts to engineer these enzymes have generally been unsuccessful (Lanio *et al.* 2000), due to the tight interdependence of enzyme structure, substrate recognition and catalysis, and in any case the gap in specificity between these enzymes (which recognize cognate sites of 4–8 bp in length) and what is required for gene-specific activity (at least 12–14 bp) would seem to be too wide to effectively bridge.

Finally, mobile group II homing endonuclease systems are promising for targeted gene disruptions, as they are easily engineered for novel specificities by altering their cognate intron sequences (their DNA specificity is dictated by base pairing with the RNA component of the intron–protein complex, rather than only by protein contacts to DNA) (Guo *et al.* 1997, 2000). Using such systems, high-throughput knockouts of entire gene families have been reported (Perutka *et al.* 2004). However, these systems are more appropriate for gene disruptions (via insertion of the mobile element) than for gene conversion. In addition, mobile group II systems require the presence and packaging of significant amounts of genetic information, including the reading frame for a large, multifunctional protein reading frame (RT, endonuclease and maturase) and the cognate intron sequence for the generation of a reactive RNP for reverse splicing and gene insertion.

8.3 Homing endonuclease engineering and selection efforts

The past 20 years of homing endonuclease research, including recent analyses of their evolution and divergence that have been facilitated by high-throughput sequencing programs, have conclusively demonstrated that these enzyme's reading frames are prone to rapid divergence, structural shuffling and recombination, continuous adaptation to and invasion of ectopic target insertion sites, rapid expansion throughout novel target lineages, and cyclical acquisition and loss. As a result, it is widely believed that the actual and potential site recognition repertoires of homing endonucleases are extremely broad. If the primary mechanisms by which evolution and selection have driven specificity changes in homing endonucleases could be duplicated and harnessed in the laboratory setting, it is possible that a wide variety of such enzymes could be generated for the applications noted above.

8.3.1 Domain fusions and shuffling

One early strategy to alter homing endonuclease specificity, practiced for the intein-associated enzymes (which use localized regions of the intein to make sequence-specific DNA contacts, in addition to the endonuclease itself) has been to exchange entire intein domains (or portions of them) with domains from other inteins or unrelated DNA-binding proteins. An advantage of working with inteins is that in some cases the DNA-recognition region (DRR) of the protein splicing domain can be uncoupled from the endonuclease domain that cleaves the DNA. Experiments of this types have demonstrated that the PI-SceI protein splicing domain can be used as a site-specific DNA-binding module in chimeric protein constructs: domain swapping between the PI-SceI and a homolog from *Candida tropicalis* (PI-CtrIP) was conducted to design altered specificity proteins (Steuer *et al.* 2004).

Several independent studies have demonstrated that domains from unrelated free-standing LAGLIDADG enzymes can be structurally fused to create fully active, chimeric homing endonucleases that recognize corresponding chimeric target sites (Chevalier *et al.* 2002; Epinat *et al.* 2003; Steuer *et al.* 2004). This technology requires extensive repacking of the domain interface, but allows the creation of new protein scaffolds with novel specificities. Of particular note, two separate groups generated an artificial highly specific endonuclease by fusing domains of homing endonucleases I-DmoI and I-CreI; the novel protein was termed H-DreI and DmoCre respectively. In the first of these reported studies, protein engineering to generate the chimeric endonuclease (Fig. 12) was accomplished by combining computational redesign and an *in vivo* protein folding screen (Chevalier *et al.* 2002). The resulting enzyme binds a long chimeric DNA target site with nanomolar affinity, cleaving it precisely at a rate equivalent to its natural parents. The structure of the engineered protein in complex with its DNA target demonstrated the accuracy of the protein interface redesign algorithm and reveals how catalytic function is maintained during the creation of the new endonuclease.

A related experiment has demonstrated that a single chain, monomeric endonuclease can be generated from a homodimer predecessor, by generating a fusion of genes that encoded each subunit connected with an artificial linker (Epinat *et al.* 2003). In these experiments, a linker derived from I-DmoI was used to join two copies of the I-CreI gene to generate a pseudo-symmetric single-chain enzyme, which cleaves DNA with the same specificity as native I-CreI *in vitro*. Furthermore, this construct was shown to initiate homologous recombination in both yeast and mammalian cells. Finally, the role and mutability of interfacial residues between LAGLIDADG helices has been examined by grafting side-chains from the homodimeric I-CreI into the corresponding positions in the monomeric I-DmoI enzyme, resulting in enzymes with novel nicking activities and oligomeric properties (Silva & Belfort, 2004).

8.3.2 Base-pair specificity changes: selections, screens and redesigns

Several methods that have been used and described to alter homing endonuclease specificity, for the most part at the level of individual base-pair alterations in the cognate target site. These strategies can be broadly divided into those that select or screen for DNA-binding activity and those that select or screen for cleavage (Fig. 13).

An adaptation of a bacterial two-hybrid strategy (Joung *et al.* 2000) selected altered specificity variants of the intein-encoded PI-SceI homing endonuclease (Gimble *et al.* 2003). The DNA-binding specificities of the selected variants ranged from being relaxed (i.e. able to cleave the

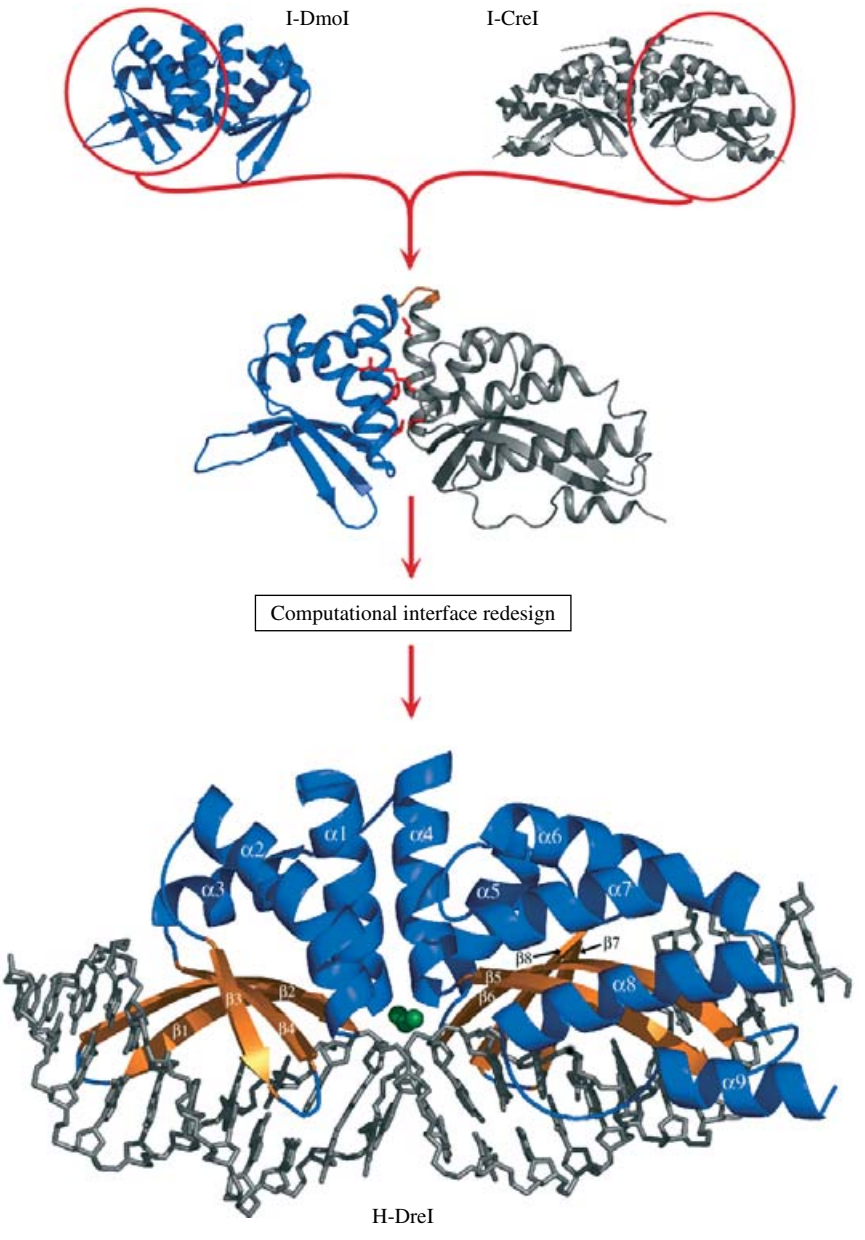


Fig. 12. Generation of an artificial, chimeric homing endonuclease from individual domains of the I-DmoI and I-CreI parental enzymes, as described in the text.

wild-type and mutant targets equally) to being dramatically shifted to preferring the selection targets. However, none of the variants displayed the same degree of specificity as wild-type PI-SceI.

A strategy for isolating I-CreI derivatives with increased affinities for altered target sites has been described (Seligman *et al.* 2002; Sussman *et al.* 2004). Endonuclease mutants with single amino-acid substitutions at positions predicted to make base-specific DNA contacts were assayed against appropriate DNA target-site mutants in an *E. coli* based system. In this system, cleavage of target sites results in cells being converted from lac^+ to lac^-. Additionally,

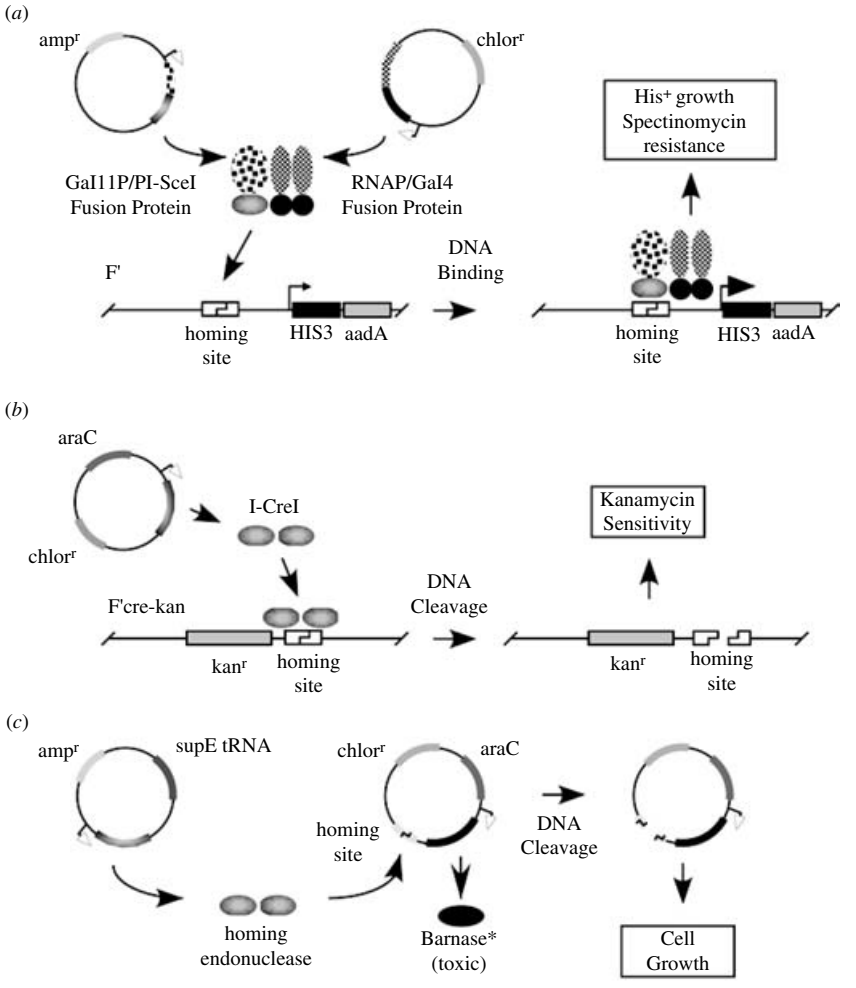


Fig. 13. Bacterial selections and screens for altered specificity homing endonucleases. (*a*) Bacterial two-hybrid binding selection (Gimble *et al.* 2003). Gal11P/PI-SceI fusion variants expressed from a plasmid library associate with a RNAPα/Gal4 fusion protein encoded by a separate plasmid. If the PI-SceI variant binds to a homing site located on an F', it recruits the RNAPα subunit proximal to the weak P*lac* promoter, leading to an increase in *HIS3* and *aadA* expression and growth on histidine-selective and spectinomycin-selective media. (*b*) Screen for I-CreI cleavage activity (Seligman *et al.* 2002). Plasmid-encoded I-CreI derivatives that bind and cleave an I-CreI homing site located on an F' lead to its elimination. Concomitant loss of an adjacent kanamycin antibiotic marker yields kanamycin-sensitive cells. (*c*) Selection for homing endonuclease activity (Gruen *et al.* 2002). Co-existence of two plasmids kills bacterial cells because one expresses an amber nonsense allele of a toxic gene product, barnase, and the other expresses an amber tRNA suppressor. However, when the tRNA suppressor plasmid also expresses a homing enzyme that cleaves the homing site on the barnase expression plasmid, the cells survive due to elimination of the barnase gene. [Figure and caption reproduced from *Homing Endonucleases and Inteins* (Springer-Verlag) with permission of the author (F. Gimble).]

undesirable activity (for example, cleavage of the original wild-type site) can be suppressed through a secondary 'negative' screen for elimination of an essential reporter (such as an antibiotic resistance marker). Using such methods, enzyme variants with shifted, rather than completely altered specificity proteins were obtained. A similar genetic selection strategy has also

been described in which a toxic gene product, barnase, expressed by a reporter plasmid, kills all of the bacteria except those that also encode a homing endonuclease variant of I-SceI that cleaves a homing site within the barnase plasmid (Gruen *et al.* 2002).

Finally, a research group at a company dedicated to the engineering and application of homing endonucleases has recently described a novel assay system designed to report on the generation of double-strand break-induced homologous recombination in eukaryotic cells (Perez *et al.* 2005). The use of recombination-based assays, as opposed to those that report on either DNA binding or DNA cleavage, promises to add significant contributions to the field of homing endonuclease engineering and selection.

9. References

Agaard, C., Awayez, M. J. & Garrett, R. A. (1997). Profile of the DNA recognition site of the archaeal homing endonuclease I-DmoI. *Nucleic Acids Research* **25**, 1523–1530.

Argast, G. M., Stephens, K. M., Emond, M. J. & Monnat, R. J. (1998). I-*Ppo*I and I-*Cre*I homing site sequence degeneracy determined by random mutagenesis and sequential *in vitro* enrichment. *Journal of Molecular Biology* **280**, 345–353.

Bassi, G. S., Oliveira, D. M. D., White, M. F. & Weeks, K. M. (2002). Recruitment of intron-encoded and co-opted proteins in splicing of the bI3 group I intron RNA. *Proceedings of the National Academy of Sciences USA* **99**, 128–133.

Bassi, G. S. & Weeks, K. M. (2003). Kinetic and thermodynamic framework for assembly of the six-component bI3 group I intron ribonucleoprotein catalyst. *Biochemistry* **42**, 9980–9988.

Belfort, M. (1989). Bacteriophage introns: parasites within parasites? *Trends in Genetics* **5**, 209–213.

Belfort, M. & Perlman, P. S. (1995). Mechanisms of intron mobility. *Journal of Biological Chemistry* **270**, 30237–30240.

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

Belfort, M. & Roberts, R. J. (1997). Homing endonucleases – keeping the house in order. *Nucleic Acids Research* **25**, 3379–3388.

Bell-Pedersen, D., Quirk, S., Clyman, J. & Belfort, M. (1990). Intron mobility in phage T4 is dependent upon a distinctive class of endonulceases and independent DNA sequences encoding the intron core: mechanistic and evolutionary implications. *Nucleic Acids Research* **18**, 3763–3770.

Bell-Pedersen, D., Quirk, S. M., Aubrey, M. & Belfort, M. (1989). A site-specific endonuclease and co-conversion of flanking exons associated with the mobile td intron of phage T4. *Gene* **82**, 119–126.

Bell-Pedersen, D., Quirk, S. M., Bryk, M. & Belfort, M. (1991). I-TevI, the endonuclease encoded by the mobile td intron, recognizes binding and cleavage domains on its DNA target. *Proceedings of the National Academy of Sciences USA* **88**, 7719–7723.

Belle, A., Landthaler, M. & Shub, D. A. (2002). Intronless homing: site-specific endonuclease SegF of bacteriophage T4 mediates localized marker exclusion analogous to homing endonucleases of group I introns. *Genes and Development* **16**, 351–362.

Bhattacharya, D., Friedl, T. & Helms, G. (2002). Vertical evolution and intragenic spread of lichen-fungal group I introns. *Journal of Molecular Evolution* **55**, 74–84.

Bibikova, M., Beumer, K., Trautman, J. K. & Carroll, D. (2003). Enhancing gene targeting with designed zinc finger nucleases. *Science* **300**, 764.

Bibikova, M., Carroll, D., Segal, D. J., Trautman, J. K., Smith, J., Kim, Y. G. & Chandrasegaran, S. (2001). Stimulation of homologous recombination through targeted cleavage by chimeric nucleases. *Molecular and Cellular Biology* **21**, 289–297.

Biniszkiewicz, D., Cesnaviciene, E. & Shub, D. A. (1994). Self-splicing group I intron in cyanobacterial initiator methionine tRNA: evidence for lateral transfer of introns in bacteria. *EMBO Journal* **13**, 4629–4635.

Bolduc, J. M., Spiegel, P. C., Chatterjee, P., Brady, K. L., Downing, M. E., Caprara, M. G., Waring, R. B. & 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 and Development* **17**, 2875–2888.

Bonocora, R. & Shub, D. A. (2001). A novel group I intron-encoded endonuclease specific for the anticodon region of tRNA-fMet genes. *Molecular Microbiology* **39**, 1299–1306.

Bryk, M., Belisle, M., Mueller, J. E. & Belfort, M. (1995). Selection of a remote cleavage site by I-*Tev*I, the td intron-encoded endonuclease. *Journal of Molecular Biology* **247**, 197–210.

Bujnicki, J. M. (2000). Phylogeny of the restriction endonuclease-like superfamily inferred from comparison of

protein structures. *Journal of Molecular Evolution* **50**, 39–44.

Burt, A. (2003). Site-specific selfish genes as tools for the control and genetic engineering of natural populations. *Proceedings of the Royal Society of London* (*B*): *Biological Sciences* **270**, 921–928.

Caprara, M. G., Lehnert, V., Lambowitz, A. & Westhof, E. (1996a). A tyrosyl-tRNA synthetase recognizes a conserved tRNA-like structural motif in the group I intron catalytic core. *Cell* **87**, 1135–1145.

Caprara, M. G., Mohr, G. & Lambowitz, A. (1996b). A tyrosyl-tRNA synthetase protein induces tertiary folding of the group I intron catalytic core. *Journal of Molecular Biology* **257**, 512–531.

Caprara, M. G., Myers, C. A. & Lambowitz, A. (2001). Interaction of the Neurospora crassa mitochondrial tyrosyl-tRNA synthetase (CYT-18 protein) with the group I intron P4-P6 domain. Thermodynamic analysis and the role of metal ions. *Journal of Molecular Biology* **308**, 165–190.

Chatterjee, P., Brady, K. L., Solem, A., Ho, Y. & Caprara, M. G. (2003). Functionally distinct nucleic acid binding sites for a group I intron-encoded RNA maturase/DNA homing endonuclease. *Journal of Molecular Biology* **329**, 239–251.

Cheng, Y.-S., Hsia, K.-C., Doudeva, L. G., Chak, K.-F. & Yuan, H. S. (2002). The crystal structure of the nuclease domain of colicin E7 suggests a mechanism for binding to double-stranded DNA by the HNH endonucleases. *Journal of Molecular Biology* **324**, 227–236.

Chevalier, B., Sussman, D., Otis, C., Noel, A. J., Turmel, M., Lemieux, C., Stephens, K., Monnat Jr., R. J. & Stoddard, B. L. (2004). Metal-dependent DNA cleavage mechanism of the I-CreI LAGLIDADG homing endonuclease. *Biochemistry* **43**, 4015–4026.

Chevalier, B., Turmel, M., Lemieux, C., Monnat, R. J. & Stoddard, B. L. (2003). Flexible DNA target site recognition by divergent homing endonuclease isoschizomers I-CreI and I-MsoI. *Journal of Molecular Biology* **329**, 253–269.

Chevalier, B. S., Kortemme, T., Chadsey, M. S., Baker, D., R. J. Monnat, J. & Stoddard, B. L. (2002). Design, activity and structure of a highly specific artificial endonuclease. *Molecular Cell* **10**, 895–905.

Chevalier, B. S., Monnat, Jr., R. J. & Stoddard, B. L. (2001). The homing endonuclease I-CreI uses three metals, one of which is shared between the two active sites. *Nature Structural Biology* **8**, 312–316.

Chevalier, B. S. & Stoddard, B. L. (2001). Homing endonucleases: structural and functional insight into the catalysts of intron/intein mobility. *Nucleic Acids Research* **29**, 3757–3774.

Cho, Y., Qiu, Y.-L., Kuhlman, P. & Palmer, J. D. (1998). Explosive invasion of plant mitochondria by a group I intron. *Proceedings of the National Academy of Sciences USA* **95**, 14244–14249.

Christ, F., Schoettler, S., Wende, W., Steuer, S., Pingoud, A. & Pingoud, V. (1999). The monomeric homing endonuclease PI-SceI has two catalytic centres for cleavage of the two strands of its DNA substrate. *EMBO Journal* **18**, 6908–6916.

Coetzee, T., Herschlag, D. & Belfort, M. (1994). Escherichia coli proteins, including ribosomal protein S12, facilitate in vitro splicing of phage T4 introns by acting as RNA chaperones. *Genes and Development* **8**, 1575–1588.

Colleaux, L., D'Auriol, L., Betermier, M., Cottarel, G., Jacquier, A., Galibert, F. & Dujon, B. (1986). Universal code equivalent of a yeast mitochondrial intron reading frame is expressed into E. coli as a specific double strand endonuclease. *Cell* **44**, 521–533.

Colleaux, L., D'Auriol, L., Galibert, F. & Dujon, B. (1988). Recognition and cleavage site of the intron-encoded omega transposase. *Proceedings of the National Academy of Sciences USA* **85**, 6022–6026.

Cousineau, B., Smith, D., Cavanagh, S. L., Mueller, J. E., Yang, J., Mills, D., Manias, D., Dunny, G., Lambowitz, A. M. & Belfort, M. (1998). Retrohoming of a bacterial group II intron: mobility via complete reverse splicing, independent of homologous DNA recombination. *Cell* **94**, 451–462.

Curcio, M. J. & Belfort, M. (1996). Retrohoming: cDNA-mediated mobility of group II introns requires a catalytic RNA. *Cell* **84**, 9–12.

Dalgaard, J. Z., Garrett, R. A. & Belfort, M. (1994). Purification and characterization of two forms of I-*Dmo*I, a thermophilic site-specific endonuclease encoded by an archaeal intron. *Journal of Biological Chemistry* **269**, 28885–28892.

Dalgaard, J. Z., Klar, A. J., Moser, M. J., Holley, W. R., Chatterjee, A. & 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 Research* **25**, 4626–4638.

Dean, A. B., Stanger, M. J., Dansereau, J. T., Vanroey, P., Derbyshire, V. & Belfort, M. (2002). Zinc finger as distance determinant in the flexible linker of intron endonuclease I-TevI. *Proceedings of the National Academy of Sciences USA* **99**, 8554–8561.

Decatur, W. A., Johansen, S. & Vogt, V. M. (2000). Expression of the Naegleria intron endonuclease is dependent on a functional group I self-cleaving ribozyme. *RNA* **6**, 616–627.

Delahodde, A., Goguel, V., Becam, A. M., 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.

Denovan-Wright, E. M., Nedelcu, A. M. & Lee, R. W. (1998). Complete sequence of the mitochondrial DNA

of Chlamydomonas eugametos. *Plant Molecular Biology* **36**, 285–295.

Derbyshire, V., Kowalski, J. C., Dansereau, J. T., Hauer, C. R. & Belfort, M. (1997). Two-domain structure of the td intron-encoded endonuclease I-TevI correlates with the two-domain configuration of the homing site. *Journal of Molecular Biology* **265**, 494–506.

Downing, M. E., Brady, K. L. & Caprara, M. G. (2005). A C-terminal fragment of an intron-encoded maturase is sufficient for promoting group I intron splicing. *RNA* **11**, 437–446.

Drouin, M., Lucas, P., Otis, C., Lemieux, C. & 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 Research* **28**, 4566–4572.

Duan, X., Gimble, F. S. & Quiocho, F. A. (1997). Crystal structure of PI-SceI, a homing endonuclease with protein splicing activity. *Cell* **89**, 555–564.

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, R. A., Jacq, C., Lemieux, C., Perlman, P. S. & Vogt, V. M. (1989). Mobile introns: definition of terms and recommended nomenclature. *Gene* **82**, 115–118.

Durrenberger, F. & 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*. *Molecular and General Genetics* **236**, 409–414.

Eddy, S. R. & Gold, L. (1991). The phage T4 nrdB intron: a deletion mutant of a version found in the wild. *Genes Development* **5**, 1032–1041.

Edgell, D. R., Derbyshire, V., Vanroey, P., Labonne, S., Stanger, M. J., Li, Z., Boyd, T. M., Shub, D. A. & Belfort, M. (2004a). Intron-encoded homing endonuclease I-TevI also functions as a transcriptional autorepressor. *Nature Structural Molecular Biology* **11**, 936–944.

Edgell, D. R. & Shub, D. A. (2001). Related homing endonucleases I-BmoI and I-TevI use different strategies to cleave homologous recognition sites. *Proceedings of the National Academy of Sciences USA* **98**, 7898–7903.

Edgell, D. R., Stanger, M. J. & Belfort, M. (2003). Importance of a single base pair for discrimination between intron-containing and intronless alleles by endonuclease I-BmoI. *Current Biology* **13**, 973–978.

Edgell, D. R., Stanger, M. J. & Belfort, B. (2004b). Coincidence of cleavage sites of intron endonuclease I-TevI and critical sequences of the host thymidylate synthase gene. *Journal of Molecular Biology* **343**, 1231–1241.

Elde, M., Haugen, P., Willassen, N. P. & Johansen, S. (1999). I-NjaI, a nuclear intron-encoded homing endonuclease from Naegleria, generates a pentanucleotide 3′ cleavage-overhang within a 19 base-pair partially symmetric DNA recognition site. *European Journal of Biochemistry* **259**, 281–288.

Elde, M., Willassen, N. P. & Johansen, S. (2000). Functional characterization of isoschizomeric His-Cys box homing endonucleases from Naegleria. *European Journal of Biochemistry* **267**, 7257–7266.

Ellison, E. L. & Vogt, V. M. (1993). Interaction of the intron-encoded mobility endonuclease I-*Ppo*I with its target site. *Molecular and Cellular Biology* **13**, 7531–7539.

Epinat, J. C., Arnould, S., Chames, P., Rochaix, P., Desfontaines, D., Puzin, C., Patin, A., Zanghellini, A., Paques, F. & Lacroix, E. (2003). A novel engineered meganuclease induces homologous recombination in yeast and mammalian cells. *Nucleic Acids Research* **31**, 2952–2962.

Eskes, R., Yang, J., Lambowitz, A. M. & Perlman, P. S. (1997). Mobility of yeast mitochondrial group II introns: engineering a new site specificity and retrohoming via full reverse splicing. *Cell* **88**, 865–874.

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

Flick, K. E., McHugh, D., Heath, J. D., Stephens, K. M., Monnat, Jr., R. J. & Stoddard, B. L. (1997). Crystallization and preliminary X-ray studies of I-PpoI: a nuclear, intron-encoded homing endonuclease from Physarum polycephalum. *Protein Science* **6**, 2677–2680.

Foley, S., Bruttin, A. & Brussow, H. (2000). Widespread distribution of a group I intron and its three deletion derivatives in the lysin gene of Streptococcus thermophilus bacteriophages. *Journal of Virology* **74**, 611–618.

Friedhoff, P., Franke, I., Krause, K. L. & Pingoud, A. (1999a). Cleavage experiments with deoxythymidine 3′,5′-bis-(p-nitrophenyl phosphate) suggest that the homing endonuclease I-PpoI follows the same mechanism of phosphodiester bond hydrolysis as the non-specific Serratia nuclease. *FEBS Letters* **443**, 209–214.

Friedhoff, P., Franke, I., Meiss, G., Wende, W., Krause, K. L. & Pingoud, A. (1999b). A similar active site for non-specific and specific endonucleases. *Nature Structural Biology* **6**, 112–113.

Friedhoff, P., Kolmes, B., Gimadutdinow, O., Wende, W., Krause, K. L. & Pingoud, A. (1996). Analysis of the mechanism of the Serratia nuclease using site-directed mutagenesis. *Nucleic Acids Research* **24**, 2632–2639.

Galburt, E., Jurica, M., Chevalier, B., Erho, D., Tang, W., Monnat Jr., R. J. & Stoddard, B. L. (2000). Conformational changes and cleavage by the homing

endonuclease I-PpoI: a critical role for a leucine residue in the active site. *Journal of Molecular Biology* **300**, 877–887.

Galburt, E. & Stoddard, B. L. (2002). Catalytic mechanisms of restriction and homing endonucleases. *Biochemistry* **41**, 13851–13860.

Galburt, E. A., Chevalier, B., Tang, W., Jurica, M. S., Flick, K. E., Monnat, R. J. & Stoddard, B. L. (1999). A novel endonuclease mechanism directly visualized for I-PpoI. *Nature Structural Biology* **6**, 1096–1099.

Galburt, E. A. & Jurica, M. S. (2005). The His-Cys box homing endonuclease family. In: *Homing Endonucleases and Inteins* (eds. M. Belfort, V. Derbyshire, B. Stoddard & D. Wood), pp. 85–102. Berlin: Springer-Verlag.

Geese, W. J., Kwon, Y. K., Wen, X. & Waring, R. B. (2003). In vitro analysis of the relationship between endonuclease and maturase activities in the bi-functional group I intron-encoded protein, I-AniI. *European Journal of Biochemistry* **270**, 1543–1554.

Geese, W. J. & Waring, R. B. (2001). A comprehensive characterization of a group IB intron and its encoded maturase reveals that protein-assisted splicing requires an almost intact intron RNA. *Journal of Molecular Biology* **308**, 609–622.

Gimble, F. S., Duan, X., Hu, D. & Quiocho, F. A. (1996). Identification of Lys-403 in the PI-SceI homing endonuclease as part of a symmetric catalytic center. *Journal of Biological Chemistry* **273**, 30524–30529.

Gimble, F. S., Moure, C. M. & Posey, K. L. (2003). Assessing the plasticity of DNA target site recognition of the PI-SceI homing endonuclease using a bacterial two-hybrid selection system. *Journal of Molecular Biology* **334**, 993–1008.

Gimble, F. S. & Wang, J. (1996). Substrate recognition and induced DNA distortion by the PI-SceI endonuclease, an enzyme generated by protein splicing. *Journal of Molecular Biology* **263**, 163–180.

Goddard, M. R. & Burt, A. (1999). Recurrent invasion and extinction of a selfish gene. *Proceedings of the National Academy of Sciences USA* **96**, 13880–13885.

Golden, B. L., Gooding, A. R., Podell, E. R. & Cech, T. R. (1998). A preorganized active site in the crystal structure of the Tetrahymena ribozyme. *Science* **282**, 259–274.

Gong, W. J. & Golic, K. G. (2003). Ends out or replacement gene targeting in Drosophila. *Proceedings of the National Academy of Sciences USA* **100**, 2556–2561.

Goodrich-Blair, H., Scarlato, V., Gott, J. M., Xu, M. Q. & Shub, D. A. (1990). A self-splicing group I intron in the DNA polymerase gene of Bacillus subtilis bacteriophage SP01. *Cell* **63**, 417–424.

Goodrich-Blair, H. & Shub, D. A. (1994). The DNA polymerase genes of several HMU-bacteriophages have similar group I introns with highly divergent open reading frames. *Nucleic Acids Research* **22**, 3715–3721.

Goodrich-Blair, H. & Shub, D. A. (1996). Beyond homing: competition between intron endonucleases confers a selective advantage on flanking genetic markers. *Cell* **84**, 211–221.

Gott, J. M., Zeeh, A., Bell-Pedersen, D., Ehrenman, K., Belfort, M. & Shub, D. A. (1988). Genes within genes: independent expression of phage T4 intron open reading frames and the genes in which they reside. *Genes and Development* **2**, 1791–1799.

Gruen, M., Chang, K., Serbanescu, I. & Liu, D. R. (2002). An in vivo selection system for homing endonuclease activity. *Nucleic Acids Research* **30**, 29–34.

Guo, H., Karberg, M., Long, M., Jones, J. P., Sullenger, B. & Lambowitz, A. (2000). Group II introns designed to insert into therapeutically relevant DNA target sites in human cells. *Science* **289**, 452–457.

Guo, H., Zimmerly, S., Perlman, P. S. & Lambowitz, A. M. (1997). Group II intron endonucleases use both RNA and protein subunits for recognition of specific sequences in double-stranded DNA. *EMBO Journal* **16**, 6835–6848.

Hahn, M. W. & Nuzhdin, S. V. (2004). The fixation of malaria refractoriness in mosquitoes. *Current Biology* **14**, R264–R265.

Haugen, P., Huss, V. A., Nielsen, H. & Johansen, S. (1999). Complex group-I introns in nuclear SSU rDNA of red and green algae: evidence of homing-endonuclease pseudogenes in the Bangiophyceae. *Current Genetics* **36**, 345–353.

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

Heath, P. J., Stephens, K. M., Monnat, R. J. & Stoddard, B. L. (1997). The structure of I-*Cre*I, a group I intron-encoded homing endonuclease. *Nature Structural Biology* **4**, 468–476.

Ho, Y., Kim, S. J. & Waring, R. B. (1997). A protein encoded by a group I intron in Aspergillus nidulans directly assists RNA splicing and is a DNA endonuclease. *Proceedings of the National Academy of Sciences USA* **94**, 8994–8999.

Ho, Y. & Waring, R. B. (1999). The maturase encoded by a group I intron from Aspergillus nidulans stabilizes RNA tertiary structure and promotes rapid splicing. *Journal of Molecular Biology* **292**, 987–1001.

Holloway, S. P., Deshpande, N. N. & Herrin, D. L. (1999). The catalytic group I introns of the psbA gene of Chlamydomonas reinhardtii: core structures, ORFs and evolutionary implications. *Current Genetics* **36**, 69–78.

Hsia, K., Chak, K., Liang, P., Cheng, Y., Ku, W. & Yuan, H. (2004). DNA binding and degradation by the HNH protein ColE7. *Structure* **12**, 205–214.

Ichiyanagi, K., Ishino, Y., Ariyoshi, M., Komori, K. & Morikawa, K. (2000). Crystal structure of an archaeal

intein-encoded homing endonuclease PI-PfuI. *Journal of Molecular Biology* **300**, 889–901.

Ito, J., Ghosh, A., Moreira, L. A., Wimmer, E. A. & Jacobslorena, M. (2002). Transgenic anopheline mosquitoes impaired in transmission of a malaria parasite. *Nature* **417**, 452–455.

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, L. D. & Norris, D. (1997). Ho endonuclease cleaves MAT DNA in vitro by an inefficient stoichiometric reaction mechanism. *Journal of Biological Chemistry* **272**, 7352–7359.

Johansen, S., Elde, M., Vader, A., Haugen, P., Haugli, K. & Haugli, F. (1997). In vivo mobility of a group I twintron in nuclear DNA of the myxomycete Didymium iridis. *Molecular Microbiology* **24**, 737–745.

Johansen, S., Embley, T. M. & Willassen, N. P. (1993). A family of nuclear homing endonucleases. *Nucleic Acids Research* **21**, 4405–4411.

Joung, J. K., Ramm, E. I. & Pabo, C. O. (2000). A bacterial two-hybrid selection system for studying protein-DNA and protein-protein interactions. *Proceedings of the National Academy of Sciences USA* **97**, 7382–7387.

Jurica, M. S., Monnat Jr., R. J. & Stoddard, B. L. (1998). DNA recognition and cleavage by the LAGLIDADG homing endonuclease I-CreI. *Molecular Cell* **2**, 469–476.

Kennell, J. C., Moran, J. V., Perlman, P. S., Butow, R. A. & Lambowitz, A. M. (1993). Reverse transcriptase activity associated with maturase-encoding group II introns in yeast mitochondria. *Cell* **73**, 133–146.

Kleanthous, C., Kuhlmann, U. C., Pommer, A. J., Ferguson, N., Radford, S. E., Moore, G. R., James, R. & Hemmings, A. M. (1999). Structural and mechanistic basis of immunity toward endonuclease colicins. *Nature Structural Biology* **6**, 243–252.

Ko, T. P., Liao, C. C., Ku, W. Y., Chak, K. F. & Yuan, H. S. (1999). The crystal structure of the DNase domain of colicin E7 in complex with its inhibitor Im7 protein. *Structure* **7**, 91–102.

Kowalski, J. C., Belfort, M., Stapleton, M. A., Holpert, M., Dansereau, J. T., Pietrokovski, S., Baxter, S. M. & Derbyshire, V. (1999). Configuration of the catalytic GIY-YIG domain of intron endonuclease I-TevI: coincidence of computational and molecular findings. *Nucleic Acids Research* **27**, 2115–2125.

Kroymann, J. & Zetsche, K. (1997). The apocytochrome-b gene in Chlorogonium elongatum: an intronic GIY-YIG ORF in green algal mitochondria. *Current Genetics* **31**, 414–418.

Kuhlmann, U. C., Moore, G. R., James, R., Kleanthous, C. & Hemmings, A. M. (1999). Structural parsimony in endonuclease active sites: should the number of homing endonuclease families be redefined? *FEBS Letters* **463**, 1–2.

Lambowitz, A., Caprara, M. G., Zimmerly, S. & Perlman, P. S. (1999). Group I and group II ribozymes as RNPs: clues to the past and guides to the future. In: *The RNA World II* (eds. R. F. Gesteland, J. F. Atkins & T. R. Cech), pp. 451–485. New York: Cold Spring Harbor Laboratory Press.

Lambowitz, A. M. & Belfort, M. (1993). Introns as mobile genetic elements. *Annual Review of Biochemistry* **62**, 587–622.

Lambowitz, A. M. & Perlman, P. S. (1990). Involvement of aminoacyl-tRNA synthetases and other proteins in group I and group II intron splicing. *Trends in Biochemical Sciences* **15**, 367–382.

Landthaler, M., Lau, N. C. & Shub, D. A. (2004). Group I intron homing in Bacillus phages SP01 and SP82: a gene conversion event initiated by a nicking homing endonuclease. *Journal of Bacteriology* **186**, 4307–4314.

Landthaler, M. L. & Shub, D. A. (2003). The nicking homing endonuclease I-BasI is encoded by a group I intron in the DNA polymerase gene of the Bacillus thuringiensis phage Bastille. *Nucleic Acids Research* **31**, 3071–3077.

Lanio, T., Jeltsch, A. & Pingoud, A. (2000). On the possibilities and limitations of rational protein design to expand the specificity of restriction endonucleases. *Protein Engineering* **13**, 275–281.

Lazowska, J., Claisse, M., Gargouri, A., Kotylak, Z., Spyridakis, A. & Slonimski, P. P. (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. *Journal of Molecular Biology* **205**, 275–289.

Lazowska, J., Jacq, C. & Slonimski, P. P. (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., Meunier, B. & Macadre, C. (1994). Homing of a group II intron in yeast mitochondrial DNA is accompanied by unidirectional co-conversion of upstream-located markers. *EMBO Journal* **13**, 4963–4972.

Lemieux, B., Turmel, M. & Lemieux, C. (1988). Unidirectional gene conversions in the chloroplast of Chlamydomonas inter-specific hybrids. *Molecular and General Genetics* **212**, 48–55.

Li, C. L., Hor, L. I., Chang, Z. F., Tsai, L. C., Yang, W. Z. & Yuan, H. S. (2003). DNA binding and cleavage by the periplasmic nuclease Vvn: a novel structure with a known active site. *EMBO Journal* **22**, 4014–4025.

Lin, J. & Vogt, V. M. (1998). I-PpoI, the endonuclease encoded by the group I intron PpLSU3, is expressed from an RNA polymerase I transcript. *Molecular and Cellular Biology* **18**, 5809–5817.

Loizos, N., Tillier, E. R. & Belfort, M. (1994). Evolution of mobile group I introns: recognition of

intron sequences by an intron-encoded endonuclease. *Proceedings of the National Academy of Sciences USA* **91**, 11983–11987.

Longo, A., Leonard, C. W., Bassi, G. S., Berndt, D., Krahn, J. M., Tanaka-Hall, T. M. & Weeks, K. M. (2005). Evolution from DNA to RNA recognition by the bI3 LAGLIDADG maturase. *Nature Structural Molecular Biology* **12**, 779–787.

Lowery, R., Hung, L., Knoche, K. & Bandziulis, R. (1992). Properties of I-*Ppo*I: a rare-cutting intron-encoded endonuclease. *Promega Notes* **38**, 8–12.

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

Lykke-Andersen, J., Aagaard, C., Semionenkov, M. & Garrett, R. A. (1997a). Archaeal introns: splicing, intercellular mobility and evolution. *Trends in Biochemical Sciences* **22**, 326–331.

Lykke-Andersen, J., Garrett, R. A. & Kjems, J. (1996). Protein footprinting approach to mapping DNA binding sites of two archael homing enzymes: evidence for a two-domain protein structure. *Nucleic Acids Research* **24**, 3982–3989.

Lykke-Andersen, J., Garrett, R. A. & Kjems, J. (1997b). Mapping metal ions at the catalytic centres of two intron-encoded endonucleases. *EMBO Journal* **16**, 3272–3281.

Mannella, C. A., Collins, R. A., Green, M. R. & Lambowitz, A. M. (1979). Defective splicing of mitochondrial rRNA in cytochrome-deficient nuclear mutants of Neurospora crassa. *Proceedings of the National Academy of Sciences USA* **76**, 2635–2639.

Mannino, S. J., Jenkins, C. L. & Raines, R. T. (1999). Chemical mechanism of DNA cleavage by the homing endonuclease I-PpoI. *Biochemistry* **38**, 16178–16186.

Mate, M. J. & Kleanthous, C. (2004). Structure-based analysis of the metal-dependent mechanism of H-N-H endonucleases. *Journal of Biological Chemistry* **279**, 34763–34769.

Matsuura, M., Saldanha, R., Ma, H., Wank, H., Yang, J., Mohr, G., Cavanagh, S., Dunny, G. M., Belfort, M. & Lambowitz, A. M. (1997). A bacterial group II intron encoding reverse transcriptase, maturase, and DNA endonuclease activities: biochemical demonstration of maturase activity and insertion of new genetic information within the intron. *Genes Development* **11**, 2910–2924.

McGraw, P. & Tzagoloff, A. (1983). Assembly of the mitochondrial membrane system. Characterization of a yeast nuclear gene involved in the processing of the cytochrome b pre-mRNA. *Journal of Biological Chemistry* **258**, 9459–9468.

Mehta, P., Katta, K. & Krishnaswamy, S. (2004). HNH family subclassification leads to identification of commonality in the His-Me endonuclease superfamily. *Protein Science* **13**, 295–300.

Michel, F. & Ferat, J.-L. (1995). Structure and activities of group II introns. *Annual Review of Biochemistry* **64**, 435–461.

Miller, M. D., Cai, J. & Krause, K. L. (1999). The active site of Serratia endonuclease contains a conserved magnesium- water cluster. *Journal of Molecular Biology* **288**, 975–987.

Mills, D. A., Mckay, L. L. & Dunny, G. M. (1996). Splicing of a group II intron involved in the conjugative transfer of pRS01 in *Lactococci*. *Journal of Bacteriology* **178**, 3531–3538.

Mohr, G., Perlman, P. S. & Lambowitz, A. M. (1993). Evolutionary relationships among group II intron-encoded proteins and identification of a conserved domain that may be related to maturase function. *Nucleic Acids Research* **21**, 4991–4997.

Moran, J. V., Zimmerly, S., Eskes, R., Kennell, J. C., Lambowitz, A. M., Butow, R. A. & Perlman, P. S. (1995). Mobile group II introns of yeast mitochondrial DNA are novel site-specific retroelements. *Molecular and Cellular Biology* **15**, 2828–2838.

Moure, C., Gimble, F. & Quiocho, F. (2002). Crystal structure of the intein homing endonuclease PI-SceI bound to its recognition sequence. *Nature Structural Biology* **9**, 764–770.

Moure, C. M., Gimble, F. S. & Quiocho, F. A. (2003). The crystal structure of the gene targeting homing endonuclease I-SceI reveals the origins of its target site specificity. *Journal of Molecular Biology* **334**, 685–696.

Mueller, J. E., Smith, D. & Belfort, M. (1996). Exon coconversion biases accompanying intron homing: battle of the nucleases. *Genes Development* **10**, 2158–2166.

Muller, K. M., Cannone, J. J., Gutell, R. R. & Sheath, R. G. (2001). A structural and phylogenetic analysis of the group IC1 introns in the order Bangiales (Rhodophyta). *Molecular and Biological Evolution* **18**, 1654–1667.

Muscarella, D. E., Ellison, E. L., Ruoff, B. M. & Vogt, V. M. (1990). Characterization of I-*Ppo*, an intron-encoded endonuclease that mediates homing of a group I intron in the ribosomal DNA of *Physarum polycephalum*. *Molecular and Cellular Biology* **10**, 3386–3396.

Muscarella, D. E. & Vogt, V. M. (1993). A mobile group I intron from Physarum polycephalum can insert itself and induce point mutations in the nuclear ribosomal DNA of saccharomyces cerevisiae. *Molecular and Cellular Biology* **13**, 1023–1033.

Nabel, G. J. (2004). Genetic, cellular and immune approaches to disease therapy: past and future. *Nature Medicine* **10**, 135–141.

Nozaki, H., Takahara, M., Kakazawa, A., Kita, Y., Yamada, T., Takano, H., Kawano, S. & Kato, M. (2002). Evolution of rbcL group IA introns and intron open reading frames within the colonial Volvocales

(Chlorophyceae). *Molecular Phylogenetics and Evolution* **23**, 326–338.

Osta, M. A., Christophides, G. K. & Kafatos, F. C. (2004). Effects of mosquito genes on Plasmodium development. *Science* **303**, 2030–2032.

Paquin, B., Laforest, M. J. & Lang, B. F. (1994). Interspecific transfer of mitochondrial genes in fungi and creation of a homologous hybrid gene. *Proceedings of the National Academy of Sciences USA* **91**, 11807–11810.

Paquin, B., O'Kelly, C. J. & Lang, B. F. (1995). Intron-encoded open reading frame of the GIY-YIG subclass in a plastid gene. *Current Genetics* **28**, 97–99.

Perez, C., Guyot, V., Cabaniols, J. P., Gouble, A., Micheaux, B., Smith, J., Leduc, S., Paques, F. & Duchateau, P. (2005). Factors affecting double-strand break-induced homologous recombination in mammalian cells. *Biotechniques* **39**, 109–115.

Perrin, A., Buckle, M. & Dujon, B. (1993). Asymmetrical recognition and activity of the I-SceI endonuclease on its site and on intron-exon junctions. *EMBO Journal* **12**, 2939–2947.

Perutka, J., Wang, W., Goerlitz, D. & Lambowitz, A. M. (2004). Use of computer-designed group II introns to disrupt E. coli DExH/D-box protein and DNA helicase genes. *Journal of Molecular Biology* **336**, 421–439.

Pommer, A. J., Cal, S., Keeble, A. H., Walker, D., Evans, S. J., Kuhlmann, U. C., Cooper, A., Connolly, B. A., Hemmings, A. M., Moore, G. R., James, R. & Kleanthous, C. (2001). Mechanism and cleavage specificity of the H-N-H endonuclease colicin E9. *Journal of Molecular Biology* **314**, 735–749.

Posey, K. L., Koufopanou, V., Burt, A. & Gimble, F. S. (2004). Evolution of divergent DNA recognition specificities in VDE homing endonucleases from two yeast species. *Nucleic Acids Research* **32**, 3947–3956.

Quirk, S. M., Bell-Pedersen, D., Tomaschewski, J., Ruger, W. & Belfort, M. (1989). The inconsistent distribution of introns in the T-even phages indicates recent genetic exchanges. *Nucleic Acids Research* **17**, 301–315.

Saguez, C., Lecellier, G. & Koll, F. (2000). Intronic GIY-YIG endonuclease gene in teh mitochondrial genome of Podospora curvicolla: evidence for mobility. *Nucleic Acids Research* **28**, 1299–1306.

Saldanha, R. J., Patel, S. S., Surendran, R., Lee, J. C. & Lambowitz, A. M. (1995). Involvement of Neurospora mitochondrial tyrosyl-tRNA synthetase in RNA splicing. A new method for purifying the protein and characterization of physical and enzymatic properties pertinent to splicing. *Biochemistry* **34**, 1275–1287.

Saldanha, R. J., Patel, S. S., Surendran, R., Lee, J. C. & Lambowitz, A. M. (1996). Analysis of the CYT-18 protein binding site at the junctoin of stacked helices in a group I intron RNA by quantitative binding assays and in vitro selection. *Journal of Molecular Biology* **261**, 23–42.

Schafer, B., Wilde, B., Massardo, D. R., Manna, F., Del Giudice, L. & Wolf, K. (1994). A mitochondrial group-I intron in fission yeast encodes a maturase and is mobile in crosses. *Current Genetics* **25**, 336–341.

Seligman, L., Chisholm, K. M., Chevalier, B. S., Chadsey, M. S., Edwards, S. T., Savage, J. H. & Veillet, A. L. (2002). Mutations altering the cleavage specificity of a homing endonuclease. *Nucleic Acids Research* **30**, 3870–3879.

Sharma, M., Ellis, R. L. & Hinton, D. M. (1992). Identification of a family of bacteriophage T4 genes encoding proteins similar to those present in group I introns of fungi and phage. *Proceedings of the National Academy of Sciences USA* **89**, 6658–6662.

Shearman, C., Godon, J.-J. & Gasson, M. (1996). Splicing of a group II intron in a functional transfer gene of *Lactococcus lactis*. *Molecular Microbiology* **21**, 45–53.

Shen, B. W., Landthaler, M., Shub, D. A. & Stoddard, B. L. (2004). DNA binding and cleavage by the HNH homing endonuclease I-HmuI. *Journal of Molecular Biology* **342**, 43–56.

Silva, G. H. & Belfort, M. (2004). Analysis of the LAGLIDADG interface of the monomeric homing endonuclease I-DmoI. *Nucleic Acids Research* **32**, 3156–3168.

Silva, G. H., Dalgaard, J. Z., Belfort, M. & Roey, P. V. (1999). Crystal structure of the thermostable archaeal intron-encoded endonuclease I-*Dmo*I. *Journal of Molecular Biology* **286**, 1123–1136.

Sitbon, E. & Pietrokovski, S. (2003). New types of conserved sequence domains in DNA-binding regions of homing endonucleases. *Trends in Biochemical Sciences* **28**, 473–477.

Smith, J., Berg, J. M. & Chandrasegaran, S. (1999). A detailed study of the substrate specificity of a chimeric restriction enzyme. *Nucleic Acids Research* **27**, 274–281.

Smith, J., Bibikova, M., Whitby, F. G., Reddy, A. R., Chandrasegaran, S. & Carroll, D. (2000). Requirements for double-strand cleavage by chimeric restriction enzymes with zinc finger DNA-recognition domains. *Nucleic Acids Research* **28**, 3361–3369.

Solem, A., Chatterjee, P. & Caprara, M. G. (2002). A novel mechanism for protein-assisted group I intron splicing. *RNA* **8**, 412–425.

Steuer, S., Pingoud, V., Pingoud, A. & Wende, W. (2004). Chimeras of the homing endonuclease PI-SceI and the homologous Candida tropicalis intein: a study to explore the possibility of exchanging DNA-binding modules to obtain highly specific endonucleases with altered specificity. *Chembiochem* **5**, 206–213.

Sui, M.-J., Tsai, L.-C., Hsia, K.-C., Doudeva, L. G., Ku, W.-Y., Han, G. W. & Yuan, H. S. (2002). Metal ions and phosphate binding in the HNH motif: crystal structures of the nuclease domain of ColE7/Im7 in complex with a phosphate ion and different divalent metal ions. *Protein Science* **11**, 2947–2957.

Sussman, D. J., Chadsey, M., Fauce, S., Engel, A., Bruett, A., Monnat Jr., R. J., Stoddard, B. L. & Seligman, L. M. (2004). Isolation and characterization of new homing endonuclease specificities at individual target site positions. *Journal of Molecular Biology* **342**, 31–41.

Thompson, A. J., Yuan, X., Kudlicki, W. & 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.

Tian, G. L., Michel, F., Macadre, C., Slonimski, P. P. & Lazowska, J. (1991). Incipient mitochondrial evolution in yeasts. II. The complete sequence of the gene coding for cytochrome b in Saccharomyces douglasii reveals the presence of both new and conserved introns and discloses major differences in the fixation of mutations in evolution. *Journal of Molecular Biology* **218**, 747–760.

Treiber, D. K. & Williamson, J. R. (2001). Beyond kinetic traps in RNA folding. *Current Opinion in Structural Biology* **11**, 309–314.

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

Uil, T. G., Haisma, H. J. & Rots, M. G. (2003). Therapeutic modulation of endogenous gene function by agents with designed DNA-sequence specificities. *Nucleic Acids Research* **31**, 6064–6078.

Urnov, F. D., Miller, J. C., Lee, Y. L., Beausejour, C. M., Rock, J. M., Augustus, S., Jamieson, A. C., Porteus, M. H., Gregory, P. D. & Holmes, M. C. (2005). Highly efficient endogenous human gene correction. *Nature* **435**, 646–651.

Vader, A., Nielsen, H. & Johansen, S. (1999). In vivo expression of the nucleolar group I intron-encoded I-dirI homing endonuclease involves the removal of a spliceosomal intron. *EMBO Journal* **18**, 1003–1013.

Vanroey, P., Meehan, L., Kowalski, J. C., Belfort, M. & Derbyshire, V. (2002). Catalytic domain structure and hypothesis for function of GIY-YIG intron endonuclease I-TevI. *Nature Structural Biology* **9**, 806–811.

Vanroey, P., Waddling, C. A., Fox, K. M., Belfort, M. & Derbyshire, V. (2001). Intertwined structure of the DNA-binding domain of intron endonuclease I-TevI with its substrate. *EMBO Journal* **20**, 3631–3637.

Wah, D. A., Hirsch, J. A., Dorner, L. F., Schildkraut, I. & Aggarwal, A. K. (1997). Structure of the multimodular endonuclease FokI bound to DNA. *Nature* **388**, 97–100.

Walweber, G. J., Mohr, S., Rennard, R., Caprara, M. G. & Lambowitz, A. M. (1997). Analyses of neurospora mitochondrial group I introns reveals different CYT-18 dependent and independent splicing strategies and an alternative 3′ splicing site for an intron ORF. *RNA* **2**, 114–131.

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

Wang, Y. H., Murphy, F. L., Cech, T. R. & Griffith, J. D. (1994). Visualization of a tertiary structural domain of the Tetrahymena group I intron by electron microscopy. *Journal of Molecular Biology* **236**, 64–71.

Waring, R. B., Davies, R. W., Scazzocchio, C. & Brown, T. A. (1982). Internal structure of a mitochondrial intron in *A. nidulans*. *Proceedings of the National Academy of Sciences USA* **79**, 6332–6336.

Weeks, K. M. & Cech, T. R. (1995a). Efficient protein-facilitated splicing of the yeast mitochondrial bI5 intron. *Biochemistry* **34**, 7728–7738.

Weeks, K. M. & Cech, T. R. (1995b). Protein facilitation of group I intron splicing by assembly of the catalytic core and the 5′ splice site domain. *Cell* **82**, 221–230.

Weeks, K. M. & Cech, T. R. (1996). Assembly of a ribonucleoprotein catalyst by tertiary structure capture. *Science* **271**, 345–348.

Wende, W., Grindl, W., Christ, F., Pingoud, A. & Pingoud, V. (1996). Binding, bending and cleavage of DNA substrates by the homing endonuclease PI-SceI. *Nucleic Acids Research* **24**, 4123–4132.

Wickelgren, I. (2003). Spinning junk into gold. *Science* **300**, 1646–1649.

Wintjens, R. & Rooman, M. (1996). Structural classification of HTH DNA-binding domains and protein-DNA interaction modes. *Journal of Molecular Biology* **262**, 294–313.

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

Wittmayer, P. K. & Raines, R. T. (1996). Substrate binding and turnover by the highly specific I-PpoI endonuclease. *Biochemistry* **35**, 1076–1083.

Yang, J., Mohr, G., Perlman, P. S. & Lambowitz, A. M. (1998). Group II intron mobility in yeast mitochondria – target DNA-primed reverse transcription activity of al1 and reverse splicing into DNA transposition sites in vitro. *Journal of Molecular Biology* **282**, 505–523.

Zimmerly, S., Guo, H., Eskes, R., Yang, J., Perlman, P. S. & Lambowitz, A. M. (1995a). A group II intron RNA is a catalytic component of a DNA endonuclease involved in intron mobility. *Cell* **83**, 529–538.

Zimmerly, S., Guo, H., Perlman, P. S. & Lambowitz, A. M. (1995b). Group II intron mobility occurs by target DNA-primed reverse transcription. *Cell* **82**, 545–554.