Published in final edited form as:

J Mol Biol. 2016 January 16; 428(1): 206–220. doi:10.1016/j.jmb.2015.12.005.

The Structural Basis of Asymmetry in DNA Binding and Cleavage as Exhibited by the I-SmaMI LAGLIDADG Meganuclease

Betty W. Shen¹, Abigail Lambert¹, Bradley C. Walker², Barry L. Stoddard¹, and Brett K. Kaiser²

¹**Division of Basic Sciences**, Fred Hutchinson Cancer Research Center, 1100 Fairview Avenue North, Seattle, WA 98109, USA

²**Department of Biology**, Seattle University, 901 12th Avenue, Seattle, WA 98122, USA

Abstract

LAGLIDADG homing endonucleases (“meganucleases”) are highly specific DNA cleaving enzymes that are used for genome engineering. Like other enzymes that act on DNA targets, meganucleases often display binding affinities and cleavage activities that are dominated by one protein domain. To decipher the underlying mechanism of asymmetric DNA recognition and catalysis, we identified and characterized a new monomeric meganuclease (I-SmaMI), which belongs to a superfamily of homologous enzymes that recognize divergent DNA sequences. We solved a series of crystal structures of the enzyme–DNA complex representing a progression of sequential reaction states, and we compared the structural rearrangements and surface potential distributions within each protein domain against their relative contribution to binding affinity. We then determined the effects of equivalent point mutations in each of the two enzyme active sites to determine whether asymmetry in DNA recognition is translated into corresponding asymmetry in DNA cleavage activity. These experiments demonstrate the structural basis for “dominance” by one protein domain over the other and provide insights into this enzyme's conformational switch from a nonspecific search mode to a more specific recognition mode.

Keywords

homing endonuclease; genome editing; surface potential; protein–DNA interactions; nickase

Introduction

Targeted gene modification and genome engineering is a rapidly maturing field [1] in which specified individual genomic loci are disrupted or modified for many purposes including the

Correspondence to Betty W. Shen: bshen@fredhutch.org..

Accession numbers

PDB coordinates and structure factors have been deposited for all structures at the Protein Data Bank. The following PDB codes will be released upon publication of the paper: 5E5P, 5E5O, 4LOX, 5E5S, 5E63, and 5E67.

Appendix A. Supplementary data

Supplementary data to this article can be found online at <http://dx.doi.org/10.1016/j.jmb.2015.12.005>.

correction of mutations in patients diagnosed with genetic diseases [2,3], the disruption of genes in patients harboring latent viral infections [4–6], the modification or insertion of genes in plants [7], the genetic modification of patient-derived stem cells [8,9], the modification of animal model systems [10], and the creation of gene drive systems as part of population control strategies [11,12].

Currently, four separate types of gene-specific nucleases are available for such applications [13]—LAGLIDADG meganucleases (LHEs), zinc-finger nucleases (ZFNs), TAL effector nucleases (TALENs), and CRISPR/Cas9 nuclease complexes (CRISPRs). As well, additional chimeric fusions of individual nuclease and DNA binding modules from those systems have also been described, typically for applications requiring exceptionally high specificity [14–16]. While each of these classes of gene-specific nucleases is appropriate for genome engineering, they differ greatly in various attributes that dictate their ultimate utility, including their size, their subunit composition, and the difficulty of reprogramming them [13]. Whereas creation of CRISPRs and TALENs requires relatively little experimental effort, the production of high-quality ZFNs and meganucleases requires substantial expertise and time. However, those nucleases possess a number of desirable properties, such as much smaller size and easier packaging and delivery. As well, meganucleases are composed of single-protein chains and generate DNA products with 3' overhangs, which offer additional advantages for gene modification, including elevated levels of homology-directed recombination events [17].

Meganucleases (reviewed in Ref. [18]) are found in nature both as homodimeric proteins and as pseudosymmetric single-chain monomers. Both forms of these proteins recognize target sites spanning 20–22 base pairs. Whereas homodimeric meganucleases are constrained to recognize and cleave palindromic and pseudopalindromic targets, monomeric meganucleases can recognize fully asymmetric target sequences, which allows them to exploit a greater range of potential recognition sites in their microbial hosts.

In addition to their asymmetric tertiary structures, single-chain meganucleases often display significant differences in the contribution of each protein domain to overall binding affinity and cleavage activity [19–21]. Although this behavior, as well as its impact on engineering and retargeting meganucleases for genome modification, has been well described [21,22], its structural basis is unclear. Here we describe the identification and characterization of a single-chain meganuclease, I-SmaMI, that exemplifies this behavior. We determined three structures of the WT (wild-type) protein (in the absence of DNA, in a complex with uncleaved DNA substrate, and in a complex with cleaved DNA product) and several structures of DNA-bound enzyme point mutants harboring disrupted basic residues in the N- and C-terminal active sites. These structures reveal that I-SmaMI undergoes a series of large asymmetrical changes in conformation upon binding DNA and additional smaller motions during cleavage. Binding affinity is dominated by interactions between the N-terminal domain of the protein with the 5' half-site of the DNA target. The effect of an inactivating point mutation in that domain's active site is more deleterious than is the equivalent mutation in the C-terminal domain's adjacent active site. We attribute the underlying basis of this asymmetric contribution to binding and catalysis to large differences in the distribution of positive electrostatic surface potential displayed by the N- and C-terminal domains.

Results

Identification and characterization of I-SmaMI

The I-SmaMI meganuclease was identified in a BLAST search against the entire protein collection in the National Center for Biotechnology Information database using the sequence of the monomeric meganuclease I-OnuI as a query as previously described [23]. This analysis identified a reading frame, harboring 46% identity to I-OnuI, in the mitochondrial cytochrome C oxidase gene of the fungus *Sordaria macrospora* (Supplementary Fig. S1a). We named this protein I-SmaMI, according to nomenclature standards adopted for DNA endonucleases [24]. To identify a likely candidate DNA recognition sequence for I-SmaMI, we compared the cytochrome C oxidase gene of *S. macrospora* to the same gene within a closely related strain that lacked the intervening sequence. Based on this comparison, we predicted that I-SmaMI would recognize the DNA target sequence 5'-TATCCTCCATTATCAGGTGTAC-3'. To test the binding and cleavage of a double-stranded DNA target harboring the abovementioned sequence by the putative meganuclease, we utilized a previously described DNA cleavage assay [25,26] in which both the expression of the meganuclease on the surface of yeast and its ability to cleave strands of tethered DNA targets were measured via flow cytometry (Supplementary Fig. S1b and S1c). These results conclusively demonstrated that I-SmaMI cleaved the predicted DNA target sequence indicated above.

Crystallographic structure analyses of the WT enzyme

In order to understand the mechanistic details of how I-SmaMI binds and cleaves its DNA target, we determined the structure of WT I-SmaMI in three states: unbound ("apo" in Table 1), bound to the uncleaved DNA substrate (in the presence of calcium, "Cacplx"), and bound to the cleaved DNA product (in the presence of magnesium, "Mgcplx") (Fig. 1 and Table 1).

Both domains in each structure display the previously described $\alpha\beta\alpha\beta\alpha$ fold of the LAGLIDADG homing endonuclease superfamily, with the two similarly folded globular domains packed against one another in a head-to-head arrangement and connected by a long flexible polypeptide linker. The first α -helix in each domain contains the namesake LAGLIDADG sequence motif ("Helix1" and "Helix7" within the N- and C-terminal domains, respectively). Each motif contributes a single metal-binding acidic residue (Glu20 and Asp179, respectively) to the enzyme's two active sites. The LAGLIDADG helix in each domain is flanked by a saddle-shaped DNA binding β -sheet (Fig. 1a and b, colored black). The convex surface of these β -sheets is packed into the protein core against a bundle of four long α -helices (Fig. 1a and b, colored aquamarine). Their opposing concave surfaces create binding surfaces against the major groove for each DNA half-site, forming a large number of contacts to DNA bases and the flanking phosphodiester backbone. In the structure and descriptions of both cleaved and uncleaved DNA complexes, we use the convention that the N-terminal domain of I-SmaMI binds to the 5' half-site of the DNA target (also termed the "left" or "minus" half-site in the published literature) while the C-terminal domain binds the 3' (i.e., the "right" or "plus") half-site (Fig. 1d).

In both of the DNA-bound structures, two divalent metal ions are observed in the middle of the DNA minor groove, in complex with the LAGLIDADG acidic residues mentioned above and with the DNA phosphate backbone (Fig. 1b, magenta). In the uncleaved complex visualized in the presence of Ca^{2+} —which does not support DNA cleavage by LAGLIDADG enzymes—a third partially occupied calcium ion, far removed from the enzyme active site, is also observed. This metal ion appears to be an artifact of high calcium concentrations in the crystal and is irrelevant to the function of I-SmaMI; however, it induces an unusual peptide flip at a neighboring residue.

Despite the relatively high resolution of the DNA-bound structures (~ 2.0 Å), electron density corresponding to the last two bases at the newly generated 3' overhang in the bottom strand of the cleaved enzyme–DNA product complex [i.e., ade-nine(–2) and adenine(–1) in the bottom strand of the left half-site] is unobservable. Therefore, the modeled positions of the last two adenine nucleotides in Chain B of the cleaved DNA-bound I-SmaMI structure are only approximations. However, the positions of those same two bases are well defined in the uncleaved DNA-bound complex with calcium.

In the DNA-bound structures, there are approximately 45–50 hydrogen bonds that can be modeled between the I-SmaMI polypeptide and the two DNA half-sites (Fig. 1e). Roughly half of these interactions appear to be involved in nonspecific interactions with the phosphate backbone of the DNA, while the remainder are positioned to interact with nucleotide bases. Three of those latter side chains (Arg26 and Arg28 in the N-terminal domain and residue Trp234 in the C-terminal domain) display significant rotamer changes between the unbound and bound structures, while movements of the remaining residues largely result from underlying motions of the polypeptide backbone.

In the structure of I-SmaMI in the absence of bound DNA, the C-terminal end of the protein, corresponding to residues 279–297, forms a final α -helix (“Helix11”) that points away from the protein core (Fig. 1a and c). In the crystal structure, it is involved in multiple contacts with a symmetry mate (Supplementary Fig. S2a) via domain swapping. Since I-SmaMI clearly behaves as a monomeric protein in solution at high concentration (Supplementary Fig. S2b), the position of this helix in the unbound structure appears to represent a conformation that is freely sampled in solution in the absence of DNA. The movement of this helix appears to correspond to a rigid-body rotation via a hinge region spanning three highly conserved residues (279–281). In the DNA-bound structures, the C-terminal helix is folded back onto the protein surface and fills the same position as its symmetry mate in the apo-enzyme structure. In the DNA-bound complexes, two residues near the end of this helix (Lys294 and Asn298) appear to make nonspecific contacts to the DNA backbone.

To test the role of this C-terminal helix in protein folding and function, we generated a truncation mutant, “ Helix11”, by converting the AAG codon of residue lysine 279 into a TAG stop codon, thus eliminating the entire C-terminal helix. The resulting truncation mutant was expressed and purified similarly as described for the WT enzyme. Circular dichroism (CD) spectroscopy showed that the Helix11 mutant unfolds cooperatively with a melting temperature (T_m) of ~ 54 °C (Fig. 2a), which is slightly lower than the T_m of ~ 58 °C for the WT enzyme. In spite of that slight reduction in thermal stability, the truncation

mutant still appears to form a stable folded protein. Subsequent gel retardation and DNA cleavage assays showed that the DNA binding affinity and DNA cleavage activity of the truncated Helix11 construct is substantially reduced as compared to WT I-SmaMI (Fig. 2b and c). A comparison of the extent of cleavage between the lowest concentration of WT and the highest concentration of the mutant indicates a difference in activity greater than 2 orders of magnitude (Fig. 2b). The slight difference in mobility of the Helix11 observed in the binding assay (Fig. 2c) could be the result of a slight alteration in the conformation of the complex. These results indicate the importance of this mobile C-terminal helix in the binding and cleavage of the target DNA by I-SmaMI.

Asymmetric conformational changes during binding and cleavage

To further investigate the additional conformational changes that accompany the binding of I-SmaMI to its target site, we calculated the rmsd displayed by all individual C^α atoms between each of the three possible pairwise superpositions of the crystal structures (apo *versus* Cacplx, Cacplx *versus* Mgcplx, and Mgcplx *versus* apo) based on the alignment of the first LAGLIDADG helix, and we plotted the results as both the radius (from thin to thick) and rainbow colors (from blue to red) of the putty model of the first structure in the pair (Fig. 3a, c, and e). To evaluate whether these observed structural differences are correlated with motions that occur during DNA binding and cleavage or instead are only correlated with their position within regions of elevated flexibility within the protein fold, we compared the distribution of the rmsd values between the superimposed structures to the overall B-factor distribution observed for each individual structure.

It is clear from this comparison that, with the exception of a peptide flip at residue Asp104, there is a noticeable tendency for larger conformational changes to be localized within the C-terminal domain (Fig. 3a, c, and e). These motions are unrelated to the pattern of the overall B-factor variation across the respective structures, which is more evenly distributed over both domains (the most generically flexible regions in the protein correspond to several surface loops on each domain and the peptide linker connecting the two protein domains) (Fig. 3b, d, and f). Compared to many large motions observed in a comparison of the unbound protein to its DNA-bound conformations (Fig. 3a and e), the motions observed when comparing the uncleaved and cleaved DNA-bound structures are smaller but still significantly clustered within the C-terminal protein domain (Fig. 3c). Therefore, DNA binding (and to a smaller extent, DNA cleavage) involves a relatively static N-terminal domain acting in concert with a C-terminal domain that undergoes a number of discrete conformational changes at each point in the reaction pathway.

Differential contribution of the N- and C-terminal protein domains to DNA target binding

The structural analyses summarized above suggest that the interactions between the two domains of I-SmaMI to their corresponding target DNA half-sites might contribute differentially to the overall binding affinity. To test this hypothesis, we measured the enzyme's relative ability to bind its full-length target site *versus* its ability to bind two chimeric targets in which the DNA base pairs were scrambled within either the 3' half-site (creating a DNA sequence termed the "Left-Scrambled" target) or the 5' half-site (thereby generating the corresponding "Scrambled-Right" target) (Fig. 4, top panel). Binding of WT

I-SmaMI to the three variant targets was analyzed by electrophoretic retardation (“gel shift”) assays in the presence of calcium. While binding of I-SmaMI to the “Left-Scrambled” target was comparable to its binding to the WT site, there was no observable binding of the enzyme to the “Scrambled-Right” target at the concentrations tested (Fig. 4, bottom panel). These results demonstrate that altering the DNA sequence in the upstream 5′ region of the target site, which is contacted by the protein's N-terminal domain, is more deleterious to binding than altering the sequence of the downstream 3′ region. This observation indicates that the enzyme's static N-terminal domain contributes more significantly to binding energy and affinity than its more flexible C-terminal domain.

Differential cleavage by the N- and C-terminal active sites

DNA cleavage by LAGLIDADG enzymes produces 4-base 3′ overhangs (Fig. 1) with each domain of the enzyme contributing residues that are critical to the cleavage reaction. The active site in the N-terminal domain is responsible for cleavage of the bottom DNA strand in the 5′ DNA half-site, while the active site in the C-terminal domain is responsible for cleavage of the top DNA strand in the 3′ half-site (Fig. 1). Inspection of the structure of I-SmaMI in complex with its bound DNA target revealed that residues Lys103 and Lys262 are found at equivalent positions in the reaction centers within the N- and C-terminal protein domains and are each located near a scissile phosphate group, on the bottom and top DNA strands, respectively (Fig. 5a). Each ε-amino group of the lysine side chain is located directly opposite the O3′ hydroxyl leaving group produced by DNA cleavage in the active site. These residues are each poised for activation of a water molecule (for clarity, data not shown) leading to a nucleophilic attack on the scissile phosphodiester bond. Further support for K103 and K262 functioning as a general base in the hydrolysis reaction comes from previously published meganuclease structures, including I-CreI [27], I-AniI [28], I-DmoI [29], I-MsoI [30], and I-SceI [31]. In all these structures, comprising both monomeric and homodimeric proteins, there are lysine residues occupying similar positions relative to the scissile phosphodiester bond. Finally, the functionality of the corresponding lysine residues in I-SceI, I-AniI, and I-DmoI was previously exploited to disrupt the activity of individual active sites and generate functional nickases [32–34].

To confirm the catalytic roles of these residues in I-SmaMI and to examine whether the two active sites of the enzyme would respond differently to equivalent active-site mutations, we altered residues K103 and K262, individually and then together, to yield two single (K103A or K262A) mutants and a double (K103A/K262A) mutant, and then we assayed all four constructs using a supercoiled plasmid substrate harboring the target site. These experiments demonstrated that the two active sites in the WT enzyme appear to cleave the two DNA strands at different rates, such that a transient nicked DNA intermediate is formed at early timepoints during the digestions and is then converted into fully cleaved DNA product. In contrast, the double K103A/K262A mutant is largely inactive (Fig. 5b, bottom panel). Both single mutants are active against the super-coiled plasmid, albeit at much reduced levels (Fig. 5b, top panels). Within the time course investigated, the K103A mutant (corresponding to mutation of the active site in the N-terminal domain) yielded only a cleanly nicked plasmid product, whereas the K262A mutant (corresponding to mutation of the active site in the C-terminal domain) initially produced a nicked intermediate followed by the slower

appearance of linearized product (Fig. 5b, top panels). At much higher enzyme (micromolar) concentrations, the K103A mutant also produced visible linearized products (data not shown).

Figure 5c illustrates the correlation between the protein domains of I-SmaMI and the half-sites of its target harboring their corresponding scissile phosphate—the active site in the enzyme's N-terminal domain is positioned to cleave the bottom strand in the left DNA half-site, whereas the corresponding active site in the C-terminal domain is positioned to cleave the top strand within the right DNA half-site. This assignment was confirmed by run-off sequencing on gel-purified open-circle intermediates produced by each enzyme mutant and by structural analysis of the DNA-bound mutants in the presence of magnesium. These experiments confirmed that the K103A mutant only cleaves the top strand between T(+2) and C(+3) of the right half-site (a reaction resulting from the remaining intact C-terminal domain's active site) whereas the K262A mutant initially cleaves the bottom strand between T(−3) and A(−2) of the left half-site, followed by the slower cleavage of the top strand.

The analyses described above demonstrate that the two endonuclease active sites display unique cleavage behaviors and respond differently to equivalent point mutations: mutation of the general base in the N-terminal active site (K103A) appears to result in a more dramatic loss of cleavage activity (and formation of a true “nickase” enzyme). In contrast, elimination of the K262 side chain in the C-terminal domain produces an enzyme with higher levels of residual activity in that domain's active site. Therefore, The N-terminal domain of I-SmaMI appears to (i) play a more significant role in DNA binding than its C-terminal partner and (ii) contains an active site that is more sensitive to mutation of a critical catalytic residue.

Differential electrostatic surface potentials across the I-SmaMI DNA binding surface and active sites

We next calculated the electrostatic potentials of the unbound enzyme and of the WT enzyme bound to cleaved DNA, as well as for DNA-bound structures of the active-site point mutants of I-SmaMI described above, using the program PDB2PQR [35]. We mapped these potentials onto the solvent-accessible surface of their respective structures using the program APBS [36] (Fig. 6a–e).

In the unbound (apo) structure, there is a dramatic difference in the surface potential of the N- and C-terminal protein domains, with the N-terminal domain clearly displaying a greater overall positive charge (Fig. 6a). In contrast, the bound complex displays an additional, distinct patch of positive electrostatic potential—toward the outer edge of the C-terminal domain (Fig. 6b). Similar inspection of the DNA-bound structures of three individual I-SmaMI point mutants (K203A, K262A, and K103A/K262A) shows that removal of K103 or K262 enlarges the negative surface potential in the vicinity of the active site; this effect is more pronounced in K103A than K262A (Fig. 6c–e).

Discussion

We have shown by a combination of systematic structural and biochemical analyses that the meganuclease I-SmaMI undergoes significant conformational changes upon binding to its DNA target and that these conformational changes are more pronounced within the C-terminal domain. In particular, the final 18 residues of the protein form a mobile C-terminal α -helix, which becomes ordered within the DNA-bound protein structure and forms multiple nonspecific interactions to the DNA backbone (Figs. 1 and 3). The behavior of those final C-terminal residues is especially critical for DNA binding and, by extension, for cleavage (Fig. 2). In comparison, the conformational changes between the uncleaved calcium complex and the cleaved magnesium complex are relatively small although still clearly localized primarily to the C-terminal domain of the enzyme (Fig. 3). The interactions between the more static N-terminal protein domain and the 5' DNA half-site appear to make the dominant contribution to overall target binding affinity (Fig. 4). The differences in the behavior and role of the two domains to the enzyme's function extend to the cleavage reaction: the two phosphates are hydrolyzed by the WT enzyme at different rates, and the effect of equivalent mutations in the two active sites is also different (Fig. 5). Finally, much of this highly asymmetric behavior appears to be related to significant differences in the electrostatic surface potential of the DNA binding surface within the two protein domains (Fig. 6).

Together, our results suggest that the binding of the N- and C-terminal domains to the left and right halves of the DNA target might occur in a sequential and asymmetrical manner, with the interactions between the N-terminal domain to the 5' (left) DNA half-site inducing conformational changes either within the protein or to the DNA (or both) that are required for the formation of additional interactions between the protein and the 3' (right) half-site. In turn, the observed conformational changes are likely a prerequisite for cleavage to proceed. This concept is in agreement with the published literature for these proteins: several previous studies have consistently demonstrated that their specificity of cleavage (which required both the formation of a tightly bound complex and then the precise ordering and arrangement of both active sites around the two scissile phosphates) is considerably higher than their specificity of binding (which is largely dependent only on the establishment of an initial series of contacts, often largely localized to one protein domain) [21,37,38].

Despite the availability of multiple meganuclease structures, only two others associated with both DNA-free and DNA-bound structures have been deposited in the Protein Data Bank—the homodimeric I-CreI (PDB codes 2O7M and 1BP7) [27,39] and the monomeric I-DmoI (PDB codes 1B24 and 2VS7) [29,40]. When these structures are superposed as described above for I-SmaMI (by the alignment of their first LAGLIDADG helices), both proteins also display larger structural deviations across one of the two protein subunits or domains (Fig. S3)—in agreement with our observation of an asymmetrical conformational change in I-SmaMI upon binding to its target DNA. Our results also point to a critical function for the C-terminal α -helix of I-SmaMI in DNA recognition and cleavage (Fig. 4b and c). The relevance of the C-terminal helix for binding and cleavage of LHE domains was previously demonstrated for the homodimeric LHE I-CreI in which each monomer contains an additional C-terminal α -helix (α_6) beyond what is present in each of the two domains of I-

SmaMI. It was shown that truncation of this C-terminal helix or mutation of residues in its preceding loop led to drastic reductions in binding and cleavage of the DNA target by I-CreI [39].

The asymmetric binding behavior we described for I-SmaMI has been previously reported for the homo-dimeric enzyme I-MsoI [19] and the monomeric enzyme I-AniI [21]. The homodimeric LHE I-MsoI, which targets a pseudopalindromic DNA sequence in nature, was found to bind more tightly to an artificial “left-left” chimeric DNA target than a “right-right” chimeric target [19]. Similarly, detailed kinetic analysis of the hydrolysis of DNA targets with single-base-pair substitutions (“one off” studies) by I-AniI showed that the strongest contributions to binding and cleavage activities were largely segregated between the two protein domains, with contributions to DNA binding affinity dominated by the interactions of the N-terminal domain with the left DNA half-site and contributions to the DNA cleavage rate dominated by the C-terminal domain and right half-site [21].

Many DNA binding proteins switch from a non-specific, electrostatically driven DNA binding mode (that greatly accelerates the search for their DNA target) to a specific conformation in which the side chains are properly positioned for hydrogen bonding to bases in the DNA target [41,42]. In addition to these base-specific contacts, a large number of van der Waals contacts and interactions with the phosphate backbone contribute to the overall affinity [43,44]. This sequence of events has been particularly well described for the *lac repressor*, whose structure has been determined in its unbound form, bound to a nonspecific DNA sequence (“RD” complex) and bound to its specific operator sequence (“RO” complex) [45]. The *lac repressor* undergoes significant conformational changes in transitioning between each of these phases, including the ordering of a DNA-contacting α -helix in each subunit in going from the RD to the RO form, while the DNA becomes severely bent in the RO conformation. A notable observation from these structures was that many amino acid residues that are initially involved only in sequence-nonspecific electrostatic contacts in the RD form then assume base-specific hydrogen bonds in the RO complex. Such significant conformational changes in the protein and DNA are commonly found in DNA binding events [43,46].

One can envision a similar sequence of events in I-SmaMI binding and hydrolysis of its DNA target. Prior to binding its target, the DNA binding surface of the N-terminal domain, which provides the majority of binding energy, is much more electropositive than the comparable surface within the C-terminal domain. Initial recognition of its target sequence is likely dominated by nonspecific binding of the negatively charged DNA surface to the positive potential at the surface of the N-terminal domain, followed by a transformation of the protein–DNA interface and switch to an active state. It is conceivable that the side chain of the arginine residues (Arg26, Arg28, and Arg79) that forms the lining of the positive surface potential within the N-terminal domain of the apo enzyme contributes significantly to an initial target search-mode conformation through electrostatic interactions with the phosphate backbone and then subsequently switch to interacting with DNA bases in the cognate DNA complex. The interaction of those three arginine residues with four consecutive bases is likely to force the widening of the major groove and partial unwinding/bending of the double-stranded DNA (Fig. 1e).

Some of the observations described here have previously been reported for a number of enzymes in separate articles under different conditions. Information gathered from those earlier studies are extremely helpful for understanding the activity and mechanism of the meganucleases and other DNA interacting enzymes. Results from the current study not only substantiate many of those earlier observations but also provide additional information on the conformational switch of I-SmaMI from a nonspecific binding state, presumably corresponding to the enzyme in a “target search mode”, to a specifically bound target complex. Since I-SmaMI is a representative member of a superfamily of meganucleases, which comprises a large number of homologous enzymes that display highly divergent DNA target preferences, information obtained from this study may contribute to redesign strategies of meganucleases for genome editing applications.

Materials and Methods

Cloning, mutagenesis, protein expression, and purification

Codon-optimized I-SmaMI was subcloned into the plasmid pET21d between the NcoI and NotI restriction sites. Site-directed mutagenesis was conducted using the QuikChange Kit (Agilent Technologies) following the vendor-provided protocol. For protein expression, sequence-verified plasmids were transformed into the *Escherichia coli* strain BL21(DE3)RIL⁺. Colonies were grown at 37 °C in 10-mL overnight cultures of LB in the presence of ampicillin (100 µg/mL). The overnight cultures were diluted 1:100 with the same medium and grown at 37 °C to a cell density equivalent to an OD₆₀₀ of 0.8–1.0. The cultures were induced with 0.5 mM IPTG and the expression was allowed to proceed overnight at 16 °C. Induced cells were then pelleted and stored at –80 °C. Following confirmation of successful protein expression by SDS-PAGE, cells were resuspended in a buffer containing 25 mM Tris–HCl (pH 7.5), 200 mM NaCl, 1 mM PMSF, and Benzonase (Buffer A) and were lysed by sonication and centrifuged for 20 min at 18,000 rpm in a Sorvall SS-34 rotor. The supernatant was filtered through a 0.22-micron filter and the cleared lysate was then loaded onto a 5-mL Heparin HP HiTrap (GE Healthcare) column and eluted with a linear salt gradient of Buffer A (200 mM NaCl) up to 1 M NaCl. Peak fractions were pooled, concentrated to ~10 mg/mL, and further purified by size-exclusion chromatography on a Fractogel-EMD column (16×100; EM Sciences) equilibrated in 25 mM Tris–HCl (pH 7.5) and 150 mM NaCl. Peak fractions were pooled, concentrated to ~20 mg/mL, flash frozen in 50 µL aliquots, and stored at –80 °C.

Crystallization and data collection

All crystallization screens were conducted with sitting-drop vapor diffusion in 96-well crystallization plates (Molecular Dimensions) using commercially available screening kits. Equal volumes of protein and reservoir solutions (100 nL each) were delivered to each of the 96 wells by a Mosquito robot (TTP Labtech). The plates were sealed by a clear sealing film and equilibrated at room temperature. Crystals usually appeared within 1 week. A single crystal of unbound I-SmaMI was harvested from a drop equilibrated against a reservoir containing 20% (w/v) polyethylene glycol 8000/100 mM Hepes (pH 7.5) and flash frozen with 20% (v/v) ethylene glycol in the mother liquid as cryoprotectant.

To screen for crystals of I-SmaMI/DNA complexes, we mixed I-SmaMI (5 mg/mL) with a 20% molar excess of double-stranded DNA containing its target sequence. To generate the double-stranded DNA targets for crystal screening, we resuspended single-stranded DNA oligonucleotides (IDT) of varying length encompassing the I-SmaMI target site in TE buffer to a concentration of 1 mM. Equal volumes of top and bottom strands were mixed and annealed by heating to 95 °C for 1 min, followed by slow cooling to 15 °C in a thermocycler with a ramp down of 0.1 °C every 30 s. The DNA duplex used for successful crystal growth contained a 25-base-pair blunt-end duplex, with the top strand sequence 5'-GGTATCCTCCATTATCAGGTGTACG-3' and the complementary strand 5'-CGTACACCTGATAATGGAGGATACC-3' (target site underlined). Crystals containing complexes of I-SmaMI bound to uncleaved and cleaved DNA were grown under similar conditions containing 28–32% (w/v) polyethylene glycol monomethyl ether 550, 0.1 M Hepes (pH 7.5), and 10–25 mM CaCl₂ or MgCl₂, respectively. Crystals of the complexes were frozen in liquid nitrogen without addition of additional cryoprotectant.

X-ray data were collected either in-house with a Rigaku Micromax-007HF rotating anode generator equipped with a Saturn CCD detector or remotely using beamline 5.0.1 at the Advanced Light Source, Lawrence Berkeley Laboratories. Data were processed using HKL2000 [47]. Initial phases for I-SmaMI in complex with a cleaved DNA (in the presence of magnesium) were obtained by molecular replacement using Phaser [48] using the structure of I-OnuI (PDB code 3QQY) as a search model. Subsequent phases for the unbound protein and for the complex of protein bound to uncleaved DNA (in the presence of calcium) were then obtained using the refined structure of I-SmaMI bound to cleaved DNA as a search model. Model building and refinement were performed using Coot [49] and Refmac5 [50], respectively. Surface potentials were calculated using the programs APBS [36] and PDB2PQR [35]. Figures were made using PyMOL [51]. The method for generation of rmsd-colored putty models is described in the supplementary materials section.

DNA binding assays

Electrophoretic mobility shift assays were carried out as previously described [52,53]. Fluorescent double-stranded DNA substrates were generated by annealing three single-stranded DNA oligonucleotides (IDT, Inc.) consisting of (1) a partial top strand corresponding to a fluorescently labeled “universal” oligonucleotide (5'-CCGCACCTGGCAG-3'), (2) a second partial top strand corresponding to another oligonucleotide containing the I-SmaMI target sequence, and (3) a full-length bottom strand corresponding to a complementary oligo-nucleotide spanning the universal and DNA target sequence strands. Equal volumes of each oligonucleotide (100 µM) in TE buffer (pH 8.0) were assembled in a PCR tube, heated to 95 °C, and cooled to 15 °C with a rampdown of –0.1 °C every 30 s using a thermal cycler machine. Substrates were stored at –20 °C for up to 5 weeks. The universal oligonucleotide was covalently modified on its 5' end with an IR₈₀₀ fluorophore. For the binding assays, varied concentrations of I-SmaMI protein were incubated with 100 nM DNA substrate at room temperature for 30 min in a buffer containing 10 mM Hepes (pH 7.9), 100 mM NaCl, 2% sucrose, 2 mM CaCl₂, and 33.3 ng/µL Lambda/HindIII fragments (Life Technologies). The binding reactions were then run on a 10% polyacrylamide, 0.5× TBE gel (prerun for 35 min) in 0.5× TBE buffer at 110 V

for 90 min. Fluorescence was visualized using a Licor Scanner and images were quantified using ImageJ.

Three different variants of the DNA target were tested for binding to WT I-SmaMI: a WT target containing the identical base-pair sequence as that in the crystal complexes (5'-GGTATCCTCCATTATCAGGTGTACG-3'; only the top strand of the duplex is depicted), a "Left-Scrambled" target containing the WT left half-site but a scrambled substituted right half-site (5'-GGTATCCTCCATTATgcaaTcagtc-3'), and a "Scrambled-Right" target containing a scrambled substituted left half-site but conserved right half-site (5'-ccccctCtaTTTATCAGGTGTACG). Lowercase letters in the sequences in the variant targets denote substituted bases.

DNA cleavage assays

Meganuclease cleavage assays were carried out as previously described [32,54]. Freshly prepared super-coiled target plasmid pIDTSmaMI containing the I-SmaMI cleavage site, at a concentration of 50 nM, was incubated with enzyme (final concentration: 250 nM) at 37 °C for fixed time intervals in buffer containing 10 mM MgCl₂, 25 mM Tris-HCl (pH 8.0), and 150 mM NaCl. The reactions were stopped with an equal volume of 2× stop buffer (2% SDS, 20% glycerol, and 100 mM ethylenediaminetetraacetic acid) and separated on a 0.8% TBE agarose gel. Gels were soaked in 1 µg/mL ethidium bromide in TBE for 30 min, followed by a 5-min rinse with TBE. Linear and open-circle DNA markers were generated by cleaving the substrate plasmid with XhoI (New England Biolabs) and the nicking enzyme Nt.BspQI (New England Biolabs), respectively.

Meganuclease thermal stability assays

Purified recombinant meganuclease constructs were diluted to between 10 and 20 µM concentration and dialyzed overnight into 10 mM potassium phosphate buffer at pH 8.0. CD thermal denaturation experiments were performed on a JASCO J-815 CD spectrometer with a Peltier thermostat. Initial wavelength scans (190–250 nm) were carried out for each construct at 20 °C and 90 °C; the wavelength where the change in CD signal strength was greatest (206 nm) was used for the variable temperature scan. Thermal denaturation was monitored over a temperature range 4–95 °C (0.1-cm-pathlength cell), with measurements taken every 2°. Sample temperature was allowed to equilibrate for 30 s before each measurement. Thermal denaturation half-points (T_m values) were determined by curve fitting using the SpectraManager software supplied by the instrument manufacturer.

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

Acknowledgements

This work was supported by the MJ Murdock Charitable Trust (to B.K.K. and B.C.W.) and by the National Institutes of Health (R01 GM105691, to B.L.S.). The authors also thank Jennifer Chik for initial cloning and purification studies on I-SmaMI.

References

1. Segal DJ, Meckler JF. Genome engineering at the dawn of the golden age. *Annu. Rev. Genomics Hum. Genet.* 2013; 14:135–158. [PubMed: 23701416]
2. Dupuy A, Valton J, Leduc S, Armier J, Galetto R, Gouble A, Lebuhotel C, Stary A, Paques F, Duchateau P, et al. Targeted gene therapy of xeroderma pigmentosum cells using meganuclease and TALEN. *PLoS One.* 2013; 8:e78678. [PubMed: 24236034]
3. Popplewell L, Koo T, Leclerc X, Duclert A, Mamchaoui K, Gouble A, Mouly V, Voit T, Paques F, Cedrone F, et al. Gene correction of a Duchenne muscular dystrophy mutation by meganuclease-enhanced exon knock-in. *Hum. Gene Ther.* 2013; 24:692–701. [PubMed: 23790397]
4. Grosse S, Huot N, Mahiet C, Arnould S, Barradeau S, Clerre DL, Chion-Sotinel I, Jacquemarcq C, Chapellier B, Ergani A, et al. Meganuclease-mediated inhibition of HSV1 infection in cultured cells. *Mol. Ther.* 2011; 19:694–702. [PubMed: 21224832]
5. Schiffer JT, Aubert M, Weber ND, Mintzer E, Stone D, Jerome KR. Targeted DNA mutagenesis for the cure of chronic viral infections. *J. Virol.* 2012; 86:8920–8936. [PubMed: 22718830]
6. Tebas P, Stein D, Tang WW, Frank I, Wang SQ, Lee G, Spratt SK, Surosky RT, Giedlin MA, Nichol G, et al. Gene editing of CCR5 in autologous CD4 T cells of persons infected with HIV. *N. Engl. J. Med.* 2014; 370:901–910. [PubMed: 24597865]
7. Riccioch, AE.; Henard-Damave, MC. Next biotech plants: New traits, crops, developers and technologies for addressing global challenges. *Crit. Rev. Biotechnol.* 2015. <http://dx.doi.org/10.3109/07388551.2015.1004521> [Epub ahead of print]
8. Byrne S, Mali P, Church GM. Genome editing in human stem cells. *Methods Enzymol.* 2014; 546:119–138. [PubMed: 25398338]
9. Chiba K, Hockemeyer D. Genome editing in human pluripotent stem cells using site-specific nucleases. *Methods Mol. Biol.* 2015; 1239:267–280. [PubMed: 25408412]
10. Menoret S, Fontaniere S, Jantz D, Tesson L, Thinard R, Remy S, Usal C, Ouisse LH, Fraichard A, Anegon I. Generation of Rag1-knockout immunodeficient rats and mice using engineered meganucleases. *FASEB J.* 2013; 27:703–711. [PubMed: 23150522]
11. Chan YS, Takeuchi R, Jarjour J, Huen DS, Stoddard BL, Russell S. The design and in vivo evaluation of engineered IOnuI-based enzymes for HEG gene drive. *PLoS One.* 2013; 8:e74254. [PubMed: 24040217]
12. Galizi R, Doyle LA, Menichelli M, Bernardini F, Deredec A, Burt A, Stoddard BL, Windbichler N, Crisanti A. A synthetic sex ratio distortion system for the control of the human malaria mosquito. *Nat. Commun.* 2014; 5:3977. [PubMed: 24915045]
13. Pauwels K, Podevin N, Breyer D, Carroll D, Herman P. Engineering nucleases for gene targeting: Safety and regulatory considerations. *New Biotechnol.* 2014; 31:18–27.
14. Boissel S, Jarjour J, Astrakhan A, Adey A, Gouble A, Duchateau P, Shendure J, Stoddard BL, Certo MT, Baker D, et al. megaTALs: A rare-cleaving nuclease architecture for therapeutic genome engineering. *Nucleic Acids Res.* 2014; 42:2591–2601. [PubMed: 24285304]
15. Guilinger JP, Thompson DB, Liu DR. Fusion of catalytically inactive Cas9 to FokI nuclease improves the specificity of genome modification. *Nat. Biotechnol.* 2014; 32:577–582. [PubMed: 24770324]
16. Tsai SQ, Wyvekens N, Khayter C, Foden JA, Thapar V, Reyon D, Goodwin MJ, Aryee MJ, Joung JK. Dimeric CRISPR RNA-guided FokI nucleases for highly specific genome editing. *Nat. Biotechnol.* 2014; 32:569–576. [PubMed: 24770325]
17. Sather, BD.; Romano Ibarra, GS.; Sommer, K.; Curinga, G.; Hale, M.; Khan, IF.; Singh, S.; Song, Y.; Gwiazda, K.; Sahni, J., et al. Efficient modification of CCR5 in primary human hematopoietic cells using a megaTAL nuclease and AAV donor template. *Sci. Transl. Med.* 2015. <http://dx.doi.org/10.1126/scitranslmed.aac5530>
18. Stoddard BL. Homing endonucleases from mobile group I introns: Discovery to genome engineering. *Mob. DNA.* 2014; 5:1–15. [PubMed: 24382139]
19. Eastberg JH, McConnell Smith A, Zhao L, Ashworth J, Shen BW, Stoddard BL. Thermodynamics of DNA target site recognition by homing endonucleases. *Nucleic Acids Res.* 2007; 35:7209–7221. [PubMed: 17947319]

20. Perrin A, Buckle M, Dujon B. Asymmetrical recognition and activity of the I-SceI endonuclease on its site and on intron-exon junctions. *EMBO J.* 1993; 12:2939–2947. [PubMed: 8335007]
21. Thyme SB, Jarjour J, Takeuchi R, Havranek JJ, Ashworth J, Scharenberg AM, Stoddard BL, Baker D. Exploitation of binding energy for catalysis and design. *Nature.* 2009; 461:1300–1304. [PubMed: 19865174]
22. Thyme SB, Boissel SJ, Arshiya Quadri S, Nolan T, Baker DA, Park RU, Kusak L, Ashworth J, Baker D. Reprogramming homing endonuclease specificity through computational design and directed evolution. *Nucleic Acids Res.* 2014; 42:2564–2576. [PubMed: 24270794]
23. Takeuchi R, Lambert A, Mak A, Jacoby K, Dickson R, Gloor G, Scharenberg A, Edgell D, Stoddard B. Tapping natural reservoirs of homing endonucleases for targeted gene modification. *Proc. Natl. Acad. Sci. U. S. A.* 2011; 108:13077–13082. [PubMed: 21784983]
24. Roberts RJ, Belfort M, Bestor T, Bhagwat AS, Bickle TA, Bitinaite J, Blumenthal RM, Degtyarev S, Dryden DT, Dybvig K, et al. A nomenclature for restriction enzymes, DNA methyltransferases, homing endonucleases and their genes. *Nucleic Acids Res.* 2003; 31:1805–1812. [PubMed: 12654995]
25. Baxter SK, Lambert AR, Scharenberg AM, Jarjour J. Flow cytometric assays for interrogating LAGLIDADG homing endonuclease DNA-binding and cleavage properties. *Methods Mol. Biol.* 2013; 978:45–61. [PubMed: 23423888]
26. Baxter SK, Scharenberg AM, Lambert AR. Engineering and flow-cytometric analysis of chimeric LAGLIDADG homing endonucleases from homologous I-OnuI-family enzymes. *Methods Mol. Biol.* 2014; 1123:191–221. [PubMed: 24510269]
27. Jurica MS, Monnat RJ Jr, Stoddard BL. DNA recognition and cleavage by the LAGLIDADG homing endonuclease ICreI. *Mol. Cell.* 1998; 2:469–476. [PubMed: 9809068]
28. Bolduc JM, Spiegel PC, Chatterjee P, Brady KL, Downing ME, Caprara MG, Waring RB, Stoddard BL. Structural and biochemical analyses of DNA and RNA binding by a bifunctional homing endonuclease and group I intron splicing factor. *Genes Dev.* 2003; 17:2875–2888. [PubMed: 14633971]
29. Silva GH, Belfort M. Analysis of the LAGLIDADG interface of the monomeric homing endonuclease I-DmoI. *Nucleic Acids Res.* 2004; 32:3156–3168. [PubMed: 15190132]
30. Chevalier B, Turmel M, Lemieux C, Monnat RJ, Stoddard BL. Flexible DNA target site recognition by divergent homing endonuclease isoschizomers I-CreI and I-MsoI. *J. Mol. Biol.* 2003; 329:253–269. [PubMed: 12758074]
31. Moure CM, Gimble FS, Quirocho FA. The crystal structure of the gene targeting homing endonuclease I-SceI reveals the origins of its target site specificity. *J. Mol. Biol.* 2003; 334:685–696. [PubMed: 14636596]
32. McConnell Smith A, Takeuchi R, Pellenz S, Davis L, Maizels N, Monnat RJ Jr, Stoddard BL. Generation of a nicking enzyme that stimulates site-specific gene conversion from the I-AniI LAGLIDADG homing endonuclease. *Proc. Natl. Acad. Sci. U. S. A.* 2009; 106:5099–5104. [PubMed: 19276110]
33. Molina R, Marcaida MJ, Redondo P, Marenchino M, Duchateau P, D'Abramo M, Montoya G, Prieto J. Engineering a nickase on the homing endonuclease I-DmoI scaffold. *J. Biol. Chem.* 2015; 290:18534–18544. [PubMed: 26045557]
34. Niu Y, Tenney K, Li H, Gimble FS. Engineering variants of the I-SceI homing endonuclease with strand-specific and site-specific DNA-nicking activity. *J. Mol. Biol.* 2008; 382:188–202. [PubMed: 18644379]
35. Dolinsky TJ, Czodrowski P, Li H, Nielsen JE, Jensen JH, Klebe G, Baker NA. PDB2PQR: Expanding and upgrading automated preparation of biomolecular structures for molecular simulations. *Nucleic Acids Res.* 2007; 35:W522–W525. [PubMed: 17488841]
36. Baker NA, Sept D, Joseph S, Holst MJ, McCammon JA. Electrostatics of nanosystems: Application to microtubules and the ribosome. *Proc. Natl. Acad. Sci. U. S. A.* 2001; 98:10037–10041. [PubMed: 11517324]
37. Joshi R, Ho KK, Tenney K, Chen JH, Golden BL, Gimble FS. Evolution of I-SceI homing endonucleases with increased DNA recognition site specificity. *J. Mol. Biol.* 2011; 405:185–200. [PubMed: 21029741]

38. Li H, Pellenz S, Ulge U, Stoddard BL, Monnat RJ Jr. Generation of single-chain LAGLIDADG homing endonucleases from native homodimeric precursor proteins. *Nucleic Acids Res.* 2009; 37:1650–1662. [PubMed: 19153140]
39. Prieto J, Redondo P, Padro D, Arnould S, Epinat JC, Paques F, Blanco FJ, Montoya G. The C-terminal loop of the homing endonuclease I-CreI is essential for site recognition, DNA binding and cleavage. *Nucleic Acids Res.* 2007; 35:3262–3271. [PubMed: 17452357]
40. Marcaida MJ, Prieto J, Redondo P, Nadra AD, Alibes A, Serrano L, Grizot S, Duchateau P, Paques F, Blanco FJ, et al. Crystal structure of I-DmoI in complex with its target DNA provides new insights into meganuclease engineering. *Proc. Natl. Acad. Sci. U. S. A.* 2008; 105:16888–16893. [PubMed: 18974222]
41. Halford SE, Marko JF. How do site-specific DNA-binding proteins find their targets? *Nucleic Acids Res.* 2004; 32:3040–3052. [PubMed: 15178741]
42. von Hippel PH, Berg OG. On the specificity of DNA–protein interactions. *Proc. Natl. Acad. Sci. U. S. A.* 1986; 83:1608–1612. [PubMed: 3456604]
43. Garvie CW, Wolberger C. Recognition of specific DNA sequences. *Mol. Cell.* 2001; 8:937–946. [PubMed: 11741530]
44. Luscombe NM, Laskowski RA, Thornton JM. Amino acid–base interactions: A three-dimensional analysis of protein–DNA interactions at an atomic level. *Nucleic Acids Res.* 2001; 29:2860–2874. [PubMed: 11433033]
45. Kalodimos CG, Biris N, Bonvin AM, Levandoski MM, Guennuegues M, Boelens R, Kaptein R. Structure and flexibility adaptation in nonspecific and specific protein–DNA complexes. *Science.* 2004; 305:386–389. [PubMed: 15256668]
46. Spolar RS, Record MT Jr. Coupling of local folding to site-specific binding of proteins to DNA. *Science.* 1994; 263:777–784. [PubMed: 8303294]
47. Otwinowski, Z.; Minor, W. Processing of X-ray diffraction data collected in oscillation mode. In: Carter, CW., Jr.; Sweet, RM., editors. *Methods in Enzymology*, 276. Academic Press; New York, NY: 1997. p. 307–326.
48. McCoy AJ, Grosse-Kunstleve RW, Adams PD, Winn MD, Storoni LC, Read RJ, Phaser crystallographic software *J. Appl. Crystallogr.* 2007; 40:658–674.
49. Emsley P, Lohkamp B, Scott WG, Cowtan K. Features and development of Coot. *Acta Crystallogr. D Biol. Crystallogr.* 2010; 66:486–501.
50. Murshudov GN, Vagin AA, Dodson EJ. Refinement of macromolecular structures by the maximum-likelihood method. *Acta Crystallogr. D Biol. Crystallogr.* 1997; 53:240–255. [PubMed: 15299926]
51. DeLano, W. The PyMOL Molecular Graphics System, Version 1.7.4. Schrödinger, LLC; 2004.
52. Jullien N, Herman JP. LUEGO: A cost and time saving gel shift procedure. *BioTechniques.* 2011; 51:267–269. [PubMed: 21988693]
53. Minch KJ, Rustad TR, Peterson EJ, Winkler J, Reiss DJ, Ma S, Hickey M, Brabant W, Morrison B, Turkarslan S, et al. The DNA-binding network of *Mycobacterium tuberculosis*. *Nat. Commun.* 2015; 6:5829. [PubMed: 25581030]
54. Scalley-Kim M, McConnell-Smith A, Stoddard BL. Coevolution of homing endonuclease specificity and its host target sequence. *J. Mol. Biol.* 2007; 372:1305–1319. [PubMed: 17720189]

Fig. 1.

Crystal structures of I-SmaMI in three stages of its catalytic pathway. Cartoon representations of (a) apo I-SmaMI and (b) I-SmaMI bound to its cleaved DNA target. The LAGLIDADG helices and the saddle-shaped DNA binding β -sheets are in black. The purple spheres in (b) represent Mg^{2+} ions. (c) Superposition of the structures of I-SmaMI in the DNA-free form (cyan), bound to uncleaved DNA (magenta), and bound to cleaved DNA (off white). The C-terminal helix undergoes a dramatic conformational change from an “open” conformation in the apo structure (solid line box) to a “closed” conformation in the DNA-bound structures (dashed line box). (d) Nomenclature of the DNA half-sites relative to the domain organization of I-SmaMI. The N-terminal domain (“Domain 1”) of I-SmaMI binds to the left (or minus) half-site and the C-terminal domain (“Domain 2”) binds the right (or plus) half-site of the DNA target. Red arrows indicate the position of the scissile bonds. The vertical green line indicates the “0” position for the numbering of the minus and plus half-sites. (e) Schematic diagram of the contacts between I-SmaMI and the cleaved DNA. Specific contacts between the indicated amino acid residues and bases of the major groove are annotated in blue; nonspecific contacts with the phosphate backbone are in green. The

Mg²⁺ ions are coordinated by the indicated amino acids at the end of the LAGLIDADG helices and the noncovalent oxygen atoms of the scissile phosphates in the backbone. As has been shown for other LHEs, I-SmaMI does not make direct contacts with the central four bases (highlighted yellow).

Author Manuscript

Author Manuscript

Author Manuscript

Author Manuscript

Fig. 2.

The C-terminal helix (Helix11) of I-SmaMI is required for binding and cleavage. (a) CD temperature melts of WT I-SmaMI ($T_m \sim 58^\circ$) and the Helix11 mutant ($T_m \sim 54^\circ$) showing the cooperative unfolding of both proteins. (b) Cleavage assays of WT and Helix11 proteins. A total of 7 nM supercoiled (SC) plasmid pIDTSmaMI was incubated with the indicated concentrations of protein for 30 min at 37°C . Cleavage of the plasmid (linear; LN) was observed under all concentrations of WT I-SmaMI, but only at the highest concentration of Helix11. An open-circle intermediate (OC) generated by nicking was present in most of the lanes with cleavage. A trace amount of a contaminating band observed in the substrate plasmid is indicated by an asterisk. (c) Electrophoretic mobility shift assay of WT and Helix11 proteins. The indicated concentrations of WT or Helix11 proteins were incubated with 150 nM fluorescently labeled I-SmaMI target sequence. The arrow indicates the position of the shifted band. Based on quantification of multiple binding studies, we estimate the K_D of the WT enzyme to be ~ 150 nM.

Fig. 3.

Pairwise comparisons of I-SmaMI structures from three reaction states. The normalized rmsd values for individual C^α atoms between all three pairs of superimposed structures were calculated and plotted on a putty cartoon representation of the first structure from each pair. (a) Apo *versus* Cacplx; (c) Cacplx *versus* Mgcplx; (e) Mgcplx *versus* apo. The B -factors of C^α atoms of the individual structures are plotted for comparison (b) apo, (d) Cacplx, and (f) Mgcplx. Both the thickness and color of the C^α backbone are displayed from the lowest (thin and blue) to highest (thick and red). For the construction of the rmsd panels, respective pairs of structures were aligned based on the first LAGLIDADG helix of each structure. This comparison shows that the C-terminal domain undergoes significant conformational changes upon binding DNA.

a WT (LEFT-RIGHT):
 5' GGTATCCTCCA TTAT CAGGTGTACG 3'
 3' CCATAGGAGGT AATA GTCCACATGC 5'

LEFT-Scrambled:
 5' GGTATCCTCCA TTAT gcaaTcagtc 3'
 3' CCATAGGAGGT AATA cgttAgtcag 5'

Scrambled-RIGHT:
 5' ctacctCctat TTAT CAGGTGTACG 3'
 3' gatggaGgata AATA GTCCACATGC 5'

Fig. 4.

Asymmetric binding of I-SmaMI. (a) DNA substrates used in the binding assays. Uppercase letters indicate WT sequences; lowercase letters indicate scrambled sequences. The “central four” bases are underlined. (b) Gel shift assays in which the indicated concentrations of I-SmaMI (50–400 nM) were incubated with 150 nM fluorescently labeled DNA targets. The arrow indicates the mobility of the DNA–protein complex.

Fig. 5.

Analysis of active-site residues required for I-SmaMI activity. (a) Ball-and-stick model of the reaction center of the I-SmaMI/cleaved DNA complex showing the position of the putative active-site residues K103 and K262 relative to that of the scissile phosphate bonds, as well as the acidic residues and divalent metal ions (M1 and M2) at the end of the LAGLIDADG helices. (b) Cleavage assays of I-SmaMI mutants. The supercoiled plasmid pIDTSmaMI (50 nM) was incubated with the indicated I-SmaMI variants (250 nM) for the indicated times and resolved by 0.8% agarose gel electrophoresis. Nicking of the plasmid generates the open-circle (OC) form; double-stranded cleavage generates the linear (LN) form. M stands for marker. (c) Schematic illustration of the target site cleaved by the respective domains of I-SmaMI. A red X indicates elimination of the activity by

mutagenesis. The K103A mutant (i.e., with a K262 intact) yields only the nicked product, whereas the K262A mutant (i.e., with K103 intact) initially generated the nicked intermediate followed by the fully cleaved product. The double mutant is largely inactive.

Author Manuscript

Author Manuscript

Author Manuscript

Author Manuscript

Fig. 6.

Electrostatic surface potentials of I-SmaMI variants. The surface potentials for DNA-free I-SmaMI (a) and the Mg/DNA ternary complexes of the SmaMI variants: (b) WT, (c) K103A, (d) K262A, and (e) K103A/K262A. All complexes contain a broad trough that runs diagonally across the DNA binding surface. The cleaved DNA in the WT complex is shown as a stick model. The position of the C-terminal helix (Helix11) is marked by the red broken box in (a) and (b). Magnesium ions are represented by orange spheres. Conformational changes of the C-terminal domain, together with newly formed interactions involving the Helix11, create a second patch of positive surface potential that is absent in the apo I-SmaMI structure. Mutagenesis of K103A (c), K262A (d), and K103A/K262A (e) result in large negative cavities near the reaction center. Two metal ions are present in the active site of the WT complexes, whereas three are present in the active site of each of the mutants. An asterisk indicates the position of the essential metal ion, which remained stationary in all complexes. The position of the other metal ions shifted in the variants. Mutation of K103A created a new and strong negative cavity near the reaction center (c, arrowhead), whereas mutation K262A only strengthened the negativity of an existing pocket (d, arrowhead).

Table 1

Data collection and refinement statistics.

Crystal	Apo SmaMI	Calcium/DNA	Magnesium/DNA	K103A	K262A	K103A/K262A
<i>Data collection</i>						
Space group	C2	P2 ₁ 2 ₁ 2 ₁	P2 ₁ 2 ₁ 2 ₁	P2 ₁ 2 ₁ 2 ₁	P2 ₁ 2 ₁ 2 ₁	P2 ₁ 2 ₁ 2 ₁
Cell dimensions						
<i>a</i> , <i>b</i> , <i>c</i> (Å)	148.33, 58.23, 42.05	61.33, 67.99, 97.74	61.28, 67.42, 98.09	50.59, 68.20, 97.65	59.84, 67.69, 97.78	60.52, 67.33, 98.57
α , β , γ (°)	90, 96.08, 90	90, 90, 90	90, 90, 90	90, 90, 90	90, 90, 90	90, 90, 90
Resolution (Å)	27.1–2.64	26.88–2.36	55.57–1.98	40.0–2.64	30.0–2.60	25–2.20
R_{sym}^a	0.05 (0.149) ^b	0.122 (0.809)	0.054 (0.300)	0.065 (0.346)	0.132 (0.661)	0.088 (0.619)
<i>I</i> / σ <i>I</i>	22.2 (7.5)	14.8 (2.0)	13.5 (2.93)	15.2 (3.4)	18.4 (3.6)	17.1 (2.3)
Completeness (%)	98.4 (87.0)	99.2 (93.3)	91.1 (55.3)	97.3 (81.8)	100 (100)	92.0 (59.2)
Redundancy	3.7 (3.1)	6.9 (5.6)	2.9 (1.8)	4.4 (3.2)	7.0 (7.1)	6.7 (4.4)
<i>Refinement</i>						
Resolution (Å)	27.1–2.64	26.88–2.36	55.57–1.98	40.0–2.64	30.0–2.60	25–2.20
No. of reflections	9865	16,388	25,028	11,914	12,826	19,836
R_{work} (%) ^c	17.95	19.06	18.8 (27.4)	19.2 (26.5)	18.98 (28.3)	17.73 (23.4)
R_{free} (%) ^c	26.70	26.89	26.56 (35.9)	25.39 (34.0)	27.30 (50.1)	25.13 (28.6)
No. of atoms						
Protein	2226	2392	2392	2365	2334	2348
DNA	—	1024	1024	1021	1023	1019
Ligand/ion	—	3 Ca ²⁺	2 Mg ²⁺	2 Mg ²⁺	3 Mg ²⁺	2 Mg ²⁺
Water	76	85	139	24	49	184
Others	—	27	—	7	5	8
<i>B</i> -factors						
Protein	19.25	25.51	13.55	53.9	20.64	23.26
DNA	—	28.84	13.09	41.7	26.84	28.58
Ligand/ion	40.86	29.43	9.41	60.4	29.16	45.31
Water	14.38	25.33	18.43	49.73	20.03	26.53
Others	40.85	43.72	—	64.38	36.12	32.69
rmsd						
Bond length (Å)	0.019	0.018	0.019	0.013	0.017	0.018
Bond angles (°)	2.125	2.041	2.099	1.752	2.02	2.071
Ramachandran (%)						
Core region	93.17	93.8	96.31	93.54	94.85	96.54
Allowed region	5.80	5.50	2.69	5.44	4.81	3.46
Outliers	1.02	0.69	1.01	1.02	0.34	0.00

^a $R_{\text{sym}} = \sum |I_{hi} - \langle I_h \rangle| / \sum I_h$, where I_{hi} is the *i*th measurement of reflection *h*, and $\langle I_h \rangle$ is the average measured intensity of reflection *h*.

^b Values in parentheses are for the highest-resolution shell.

^c $R/R_{\text{free}} = \sum_h |F_h(o) - F_h(c)| / \sum_h |F_h(o)|$, where R_{free} was calculated with 5% of the data excluded from refinement.