

Characterization of *Mycobacterium leprae* RecA Intein, a LAGLIDADG Homing Endonuclease, Reveals a Unique Mode of DNA Binding, Helical Distortion, and Cleavage Compared with a Canonical LAGLIDADG Homing Endonuclease^{*[S]}

Received for publication, July 9, 2009 Published, JBC Papers in Press, July 15, 2009, DOI 10.1074/jbc.M109.042861

Pawan Singh[‡], Pankaj Tripathi[‡], George H. Silva[§], Alfred Pingoud[§], and K. Muniyappa⁺¹

From the [‡]Department of Biochemistry, Indian Institute of Science, Bangalore 560012, India and the [§]Institut für Biochemie, Justus-Liebig-Universität, Heinrich-Buff-Ring 58, D-35392 Giessen, Germany

Mycobacterium leprae, which has undergone reductive evolution leaving behind a minimal set of essential genes, has retained intervening sequences in four of its genes implicating a vital role for them in the survival of the leprosy bacillus. A single in-frame intervening sequence has been found embedded within its *recA* gene. Comparison of the *M. leprae recA* intervening sequence with the known intervening sequences indicated that it has the consensus amino acid sequence necessary for being a LAGLIDADG-type homing endonuclease. In light of massive gene decay and function loss in the leprosy bacillus, we sought to investigate whether its *recA* intervening sequence encodes a catalytically active homing endonuclease. Here we show that the purified *M. leprae* RecA intein (PI-MleI) binds to cognate DNA and displays endonuclease activity in the presence of alternative divalent cations, Mg²⁺ or Mn²⁺. A combination of approaches, including four complementary footprinting assays such as DNase I, copper-phenanthroline, methylation protection, and KMnO₄, enhancement of 2-aminopurine fluorescence, and mapping of the cleavage site revealed that PI-MleI binds to cognate DNA flanking its insertion site, induces helical distortion at the cleavage site, and generates two staggered double strand breaks. Taken together, these results implicate that PI-MleI possesses a modular structure with separate domains for DNA target recognition and cleavage, each with distinct sequence preferences. From a biological standpoint, it is tempting to speculate that our findings have implications for understanding the evolution of the LAGLIDADG family of homing endonucleases.

Mycobacterium leprae, a Gram-positive rod-shaped bacillus, mostly found in warm tropical countries, is the bacterium that causes leprosy in humans (1). The lack of understanding of the basic biology of *M. leprae* is believed to be the key factor for the failure of leprosy research to advance. The genome sequence of *M. leprae* contains 3.27 Mb and has an average G + C content of

57.8%, values much lower than the corresponding values for *Mycobacterium tuberculosis*, which are ~4.41 Mb and 65.6% G + C, respectively (2). There are some 1500 genes that are common to both *M. leprae* and *M. tuberculosis*. The comparative genome analysis suggests that both species of mycobacteria are derived from a common ancestor and, at one stage, had gene pools of similar size. The downsizing of the *M. tuberculosis* genome from ~4.41 to 3.27 Mb of *M. leprae* would account for the loss of some 1200 protein-coding sequences (1, 3). There is evidence that many of the genes that were present in the genome of *M. leprae* have truly been lost (1, 3). Comparative genomics of *M. leprae* with that of *M. tuberculosis* indicate that the former has undergone substantial downsizing, losing more than 2000 genes, thus suggesting an extreme case of reductive evolution in a microbial pathogen (1). With the availability of the *M. leprae* genome sequence, using functional genomics approaches, it is possible to identify the gene products, elucidate the mechanism of their action, and identify novel drug targets for rational design of new therapeutic regimens and drugs to treat leprosy.

Eubacterial RecA proteins catalyze a set of biochemical reactions that are essential for homologous recombination, DNA repair, restoration of stalled replication forks, and SOS response (4–7). RecA protein and the process of homologous recombination, which is the main mechanism of genetic exchange, are evolutionarily conserved among a range of organisms (4, 7). Perhaps the most striking development in the field of RecA protein biology was the discovery of an in-frame insertion of an intein-coding sequence in the *recA* genes of *M. tuberculosis* and *M. leprae* (8, 9). In these organisms, RecA is synthesized as a large precursor, which undergoes protein splicing to excise the intein, and the two flanking domains called exteins are ligated together to generate a functionally active RecA protein (9, 10). The milieu in which RecA precursor undergoes splicing differs substantially between *M. tuberculosis* and *M. leprae*. *M. leprae* RecA precursor (79 kDa) undergoes splicing only in mycobacterial species, whereas *M. tuberculosis* RecA precursor (85 kDa) is spliced efficiently in *Escherichia coli* as well (9–11). Intriguingly, *M. tuberculosis* and *M. leprae* RecA inteins differ greatly in their size, primary sequence, and location within the *recA* gene, thereby suggesting two independent origins during evolution (9). The occurrence of inteins in the obligate mycobacterial pathogens, *M. tuberculosis*, *M. leprae*,

* This work was supported by a grant from the Department of Biotechnology, New Delhi, under the “Centre of Excellence” in research on mycobacteria and a J. C. Bose National Fellowship (to K. M.). We have benefited by being partners in the European Commission-funded *Asia-Link* program on “Human Resources Development in the Study of Nucleic Acids.”

[S] The on-line version of this article (available at <http://www.jbc.org>) contains supplemental Figs. 1–4.

¹ To whom correspondence should be addressed. Tel: 91-80-2293-2235/2360 0278; Fax: 91-80-23600814/0683; E-mail: kmbc@biochem.iisc.ernet.in.

and *Mycobacterium microti*, suggested that RecA inteins might play a role in mycobacterial functions related to pathogenesis or virulence (9). Previously, we have shown that *M. tuberculosis* RecA intein (PI-MtuI),² which contains Walker A motif, displays dual target specificity in the presence of alternative cofactors in an ATP-dependent manner (12, 13).

Since their discovery in *Saccharomyces cerevisiae* (14, 15), a large number of putative homing endonucleases have been found in a diverse range of proteins in all the three domains of life (16–19). The majority of inteins possess the protein splicing and homing endonuclease activities (18, 19). Homing endonucleases are a class of diverse rare-cutting enzymes that promote site-specific transposition of their encoding genetic elements by inflicting double-stranded DNA breaks via different cleavage mechanisms in alleles lacking these elements (18–23). In addition, these are characterized by their ability to bind long DNA target sites (14–40 bp), and their tolerance of minor sequence changes in their binding region. These have been divided into highly divergent subfamilies on the basis of conserved sequence and structural motifs as follows: LAGLIDADG, GIY-YIG, HNH, His-Cys box, and the more recently identified PD(D/E)XK families (18–24). LAGLIDADG homing enzymes, which include the largest family, contain one or two copies of the conserved dodecapeptide motif and utilize an extended protein-DNA interface covering up to 40 bp to acquire their necessary specificity (18–22). The LAGLIDADG sequence is a part of the conserved 10- or 12-residue sequence motif defining the family of LAGLIDADG-type homing endonucleases; therefore, it is designated as deca- or dodecapeptide motif (19).

Comparison of the *M. leprae* recA intervening sequence with known intervening sequences indicated that it has the consensus amino acid sequence necessary for being a LAGLIDADG-type homing endonuclease (25, 26). In light of massive gene decay and function loss in the leprosy bacillus, and dissimilarities in size and primary structures among mycobacterial inteins, we sought to investigate whether *M. leprae* recA intervening sequence encodes a catalytically active homing endonuclease. In this study, we show that the purified *M. leprae* RecA intein (PI-MleI) binds to cognate DNA and displays endonuclease activity in the presence of alternative divalent cations Mg²⁺ or Mn²⁺. Furthermore, using a variety of approaches, we have mapped the positions of PI-MleI binding as well as cleavage in the cognate DNA, thus providing the most comprehensive analysis of PI-MleI. Taken together, these results suggest that PI-MleI possesses a modular structure with functionally separable domains for DNA target recognition and cleavage, each with distinct sequence preferences. These results provide insights into understanding the function and evolution of the family of LAGLIDADG homing endonucleases.

EXPERIMENTAL PROCEDURES

Reagents, Bacterial Strains, and DNA—All the chemicals used in this study are of analytical grade. Buffers were prepared

² The abbreviations used are: PI-MtuI, *M. tuberculosis* RecA intein; PI-MleI, *M. leprae* RecA intein; DTT, dithiothreitol; DMS, dimethyl sulfate; 2-AP, 2-aminopurine; ODN, oligonucleotide; ATP-γS, adenosine 5'-O-(thiotriphosphate); (OP)₂Cu²⁺, phenanthroline-copper.

using deionized water. Restriction endonucleases, T4 DNA ligase, phage T4 polynucleotide kinase, IMPACT-T7 cloning system, and chitin beads were purchased from New England Biolabs. DNA gel extraction kit was purchased from Qiagen. Nylon N⁺ membrane was obtained from GE Healthcare. Cloning primers used in this study were synthesized by Sigma Genosys. *E. coli* strains DH5α and DH10B were purchased from Invitrogen, and ER2566 was procured from New England Biolabs. Strains were grown in liquid or solid agar Luria Broth (LB) media supplemented with appropriate antibiotics.

Oligonucleotides used in this study were synthesized by Sigma Genosys, and their sequences are listed in Table 1. The ODNs were labeled at the 5' end by [γ-³²P]ATP and T4 polynucleotide kinase (27). Cognate duplex DNA was prepared by annealing ODN1 and ODN2 at 95 °C for 5 min followed by gradual cooling to room temperature. Similarly, cognate duplexes containing 2-aminopurine were prepared by annealing ODN2 or ODN1 with the respective modified complementary oligonucleotides ODN3 to ODN8. The annealed substrates were electrophoresed on a 6% (w/v) polyacrylamide gel in 45 mM Tris borate buffer (pH 8.3) containing 1 mM EDTA. The bands were excised from the gel, and radioactive or 2-AP-modified cognate duplex DNA was eluted into TE buffer (10 mM Tris-HCl (pH 7.5) and 1 mM EDTA). The concentration of the oligonucleotide substrates was expressed in terms of moles of DNA ends/liter.

Construction of *M. leprae* RecA Intein (PI-MleI) Target Plasmid, pMLR—*M. leprae* recA, which is devoid of its intervening sequence, was PCR-amplified from the plasmid pEJ230 (kindly provided by E. O. Davis, National Institutes for Medical Research, London) using a set of overlapping primers (28). The ~1.1-kb PCR product was cloned into KpnI/EcoRI site of pUC19 vector and designated as pMLR. The identity of the cloned DNA fragment and the presence of the intein/extein junction were confirmed by DNA sequencing. Negatively supercoiled DNA was prepared by sucrose density gradient centrifugation as described (29). pMLR DNA was resuspended in a buffer containing 10 mM Tris-HCl (pH 7.5) and 1 mM EDTA. The concentration of DNA was expressed in moles of nucleotide residues.

Construction of *M. leprae* RecA Intein Encoding Plasmid, pTMLRI—*M. leprae* recA intervening sequence was PCR-amplified with gene-specific primers (forward primer, 5'-ATT-GGGGTGATGGCTAGCTGCATGAAT-3'; reverse primer, 5'-CGTGGTTTCCTCGAGATTGTGTACCAT-3') using plasmid pEJ230 as the template. The ~1.1-kb PCR product was cloned into pTXB1 expression vector at the NheI/XhoI restriction site. The resulting recombinant plasmid was designated as pTMLRI, which contains *M. leprae* recA intervening sequence in-frame with *M. xenopi* gyrA mini intein and the chitin binding domain sequence (30). The identity of the recombinant plasmid was confirmed by restriction analysis and DNA sequencing.

Expression and Purification of PI-MleI—PI-MleI was overexpressed in *E. coli* strain ER2566 (New England Biolabs) harboring the recombinant plasmid, pTMLRI. The transformed cells were grown in Terrific Broth medium supplemented with 100 μg/ml ampicillin at 37 °C to A_{600 nm} = 0.5. PI-MleI was induced by the addition of isopropyl 1-thio-β-D-galactopyranoside to a

M. leprae RecA Intein Homing Endonuclease

final concentration of 0.05 mM. Cultures were then transferred to 15 °C (model INNOVA 4230, New Brunswick Scientific), and induction was continued for 15 h. The cells were harvested by centrifugation, washed in STE (10 mM Tris-HCl (pH 8), 100 mM NaCl, and 1 mM EDTA), and resuspended in buffer A (20 mM Tris-HCl (pH 8), 1 M NaCl, 1 mM EDTA and 10% glycerol). All subsequent steps were performed at 4 °C, unless mentioned otherwise. Cells (~15 g) were lysed by sonication (model GEX-750, Ultrasonic Processor) on ice at 60% duty cycles in a pulse mode and centrifuged at 30,000 rpm for 60 min in Beckman Ti-45 rotor. Cell-free lysate was loaded onto a 10-ml chitin beads column pre-equilibrated with buffer A. The column was washed by passing 200 ml of buffer A followed by an additional 50 ml of buffer B (20 mM Tris-HCl (pH 8), 100 mM NaCl, and 10% glycerol) using gravity flow (2 ml/min). Cleavage was initiated by quickly flushing the column with 30 ml of buffer B containing 5 mM DTT, and the flow was stopped and incubated at 4 °C for 16 h. This strategy allows efficient cleavage of PI-MleI from the *Mycobacterium xenopi* GyrA mini intein portion of the fusion protein that remains attached to the chitin beads. The cleaved protein was eluted from the column with 100 ml of buffer B containing 5 mM 2-mercaptoethanol, and the fractions were analyzed on 9% SDS-PAGE, followed by silver staining. Fractions containing PI-MleI were pooled and precipitated by the addition of ammonium sulfate to a final concentration of 0.472 g/ml (70% saturation). The precipitate was collected by centrifugation in Beckman Ti-45 rotor at 30,000 rpm for 1 h, resuspended in buffer B, and dialyzed against storage buffer (20 mM Tris-HCl (pH 7.5), 100 mM NaCl, 1 mM DTT and 50% glycerol). Aliquots of the dialyzed PI-MleI were stored at -20 °C. The purity of PI-MleI was assessed by SDS-PAGE and Western blotting (31) with anti-intein antibody. Polyclonal antibody for PI-MleI was raised in a rabbit. The concentration of the PI-MleI protein was determined by dye binding method using bovine serum albumin as standard (32) and expressed in moles/liter.

Endonuclease Assays—Reactions (20 μ l) were performed in a buffer containing 25 mM Tris-HCl (pH 7.5), 3 or 5 mM MnCl₂ or MgCl₂, 0.4 mM DTT, 16 μ M negatively supercoiled pMLR DNA, and PI-MleI at the indicated concentrations. Reaction mixtures were incubated at 37 °C for the indicated time intervals. Reactions were stopped by the addition of 0.1% SDS followed by proteinase K (0.5 mg/ml) and deproteinized by incubation at 37 °C for 15–20 min. After addition of 2.5 μ l of gel loading dye (0.42% (w/v) of bromphenol blue and xylene cyanol in 50% glycerol), the samples were separated by electrophoresis through 0.8% agarose gel in 89 mM Tris borate buffer (pH 8.3) containing 2 mM EDTA. The gel was stained with 0.5 μ g/ml ethidium bromide, and DNA bands were observed by UV illumination. Subsequently, DNA was transferred to nylon N⁺ membrane and visualized by Southern hybridization (27, 33). Quantification of DNA bands were performed by UVI-BandMap software (UVI-Tech gel documentation system) and plotted using GraphPad prism version 4.0.

Electrophoretic Mobility Shift Assays—Binding reaction mixtures (20 μ l) contained 25 mM Tris-HCl (pH 7.5), 0.4 mM DTT, 1 nM 5'-³²P-labeled oligonucleotide cognate duplex (88 bp) or single-stranded (88-mer) DNA, and increasing concentrations

of PI-MleI. Reactions were incubated at 37 °C for 30 min and stopped by the addition of 2.2 μ l of 10 \times gel loading dye (50% glycerol containing 0.42% (w/v) each of bromphenol blue and xylene cyanol). The samples were electrophoresed on 6% native polyacrylamide gel in 0.5 \times TBE (45 mM Tris borate buffer (pH 8.3) containing 1 mM EDTA) buffer at 150 V for 3 h. The gels were dried, exposed to a Fuji FLA-5000 PhosphorImager, followed by autoradiography. The bands were visualized using the software provided by the supplier. The data were quantified using UVI-BandMap and plotted in GraphPad prism version 4.0.

Mapping of PI-MleI Cleavage Sites—Reactions were carried out in a buffer containing 25 mM Tris-HCl (pH 7.5), 0.4 mM DTT, 5 mM MnCl₂, 1 nM cognate duplex DNA (88 bp), ³²P-labeled on either upper or lower strand, and with the concentrations of PI-MleI as indicated in the figure legends. Reaction mixtures were incubated for 2 h and stopped by the addition of SDS and proteinase K to a final concentration of 0.1% and 0.5 mg/ml, respectively. The DNA was precipitated by the addition of ice-cold 95% ethanol. DNA was collected by centrifugation, and the pellets were rinsed with 70% ethanol, dried, and resuspended in loading dye (95% formamide, 10 mM NaOH, 0.2% bromphenol blue, and 0.2% xylene cyanol). Samples were heated at 95 °C for 5 min, snap-chilled on ice, and loaded onto a denaturing 12% polyacrylamide sequencing gel in the presence of 7 M urea. Electrophoresis was carried out in 89 mM Tris borate buffer (pH 8.3) containing 2 mM EDTA at 1800 V and 40 watts for 3 h. The gel was dried onto a 3MM Whatman filter paper, exposed to a FLA-5000 PhosphorImager screen, and visualized using the software supplied by the manufacturer.

DNase I Footprinting—DNase I footprinting was done as described earlier (34, 35). Binding reactions (20 μ l) contained 10 nM 5'-³²P-labeled cognate duplex DNA, 25 mM Tris-HCl (pH 7.5), 0.4 mM DTT, and increasing concentrations of PI-MleI as indicated in the figure legends. Reaction mixtures were incubated at 37 °C for 30 min. DNase I reactions were started by the addition of 10 μ l of solution containing 5 mM MgCl₂ and 5 mM CaCl₂ and DNase I to a final concentration of 0.005 units. After 1 min of incubation at 24 °C, the reactions were stopped by the addition of 100 μ l of stop solution (20 mM EDTA, 5% SDS, 200 mM NaCl, and 25 μ g/ μ l calf thymus DNA). DNA was precipitated using 95% ethanol, washed, and dried under vacuum. The DNA pellet was resuspended in formamide loading dye (80% formamide, 0.1% bromphenol blue, and 0.1% xylene cyanol), heat-denatured, and analyzed on a denaturing 12% polyacrylamide gel in the presence of 7 M urea alongside the Maxam-Gilbert G + A ladder (36). The gel was dried onto a Whatman filter paper, and the bands were visualized by Fuji FLA-5000 PhosphorImager, followed by autoradiography.

Phenanthroline-Copper (OP₂Cu²⁺) Footprinting—Reactions were performed in a buffer containing 25 mM Tris-HCl (pH 7.5), 0.4 mM DTT, 10 nM ³²P-labeled cognate duplex DNA and labeled either on upper or lower strand, and increasing concentrations of PI-MleI at 37 °C for 30 min. The nuclease activity of (OP)₂Cu²⁺ was initiated by the addition of 2 mM 1,10-phenanthroline, 0.15 mM CuSO₄, and 58 mM 3-mercaptopropionic acid, followed by incubation at 4 °C for 30 s. Reaction was stopped by the addition of 100 mM 2,9-dimethyl-1,10-phenan-

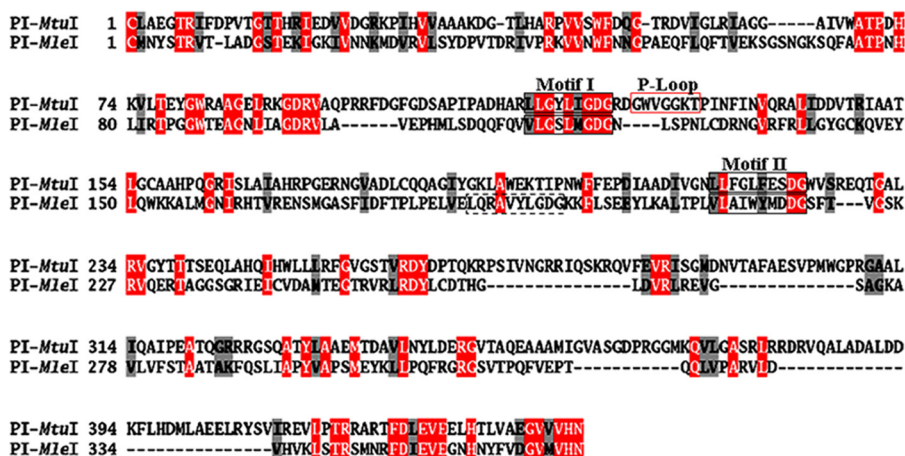


FIGURE 1. Comparison of amino acid sequence of PI-MtuI with PI-MleI. Amino acid residues are represented by the single-letter code. Red and gray shading indicates identical and similar amino acid residues, respectively. The gaps in the sequence are indicated by dotted lines. The canonical LAGLIDADG (DOD) motifs appear in black boxes, and the dashed box corresponds to the second putative LAGLIDADG motif II (residues 185–194) of PI-MleI. PI-MtuI P-loop (^{12C}GWVGGKT¹³²) is enclosed in a red box.

throsine (5 μl). DNA was ethanol-precipitated, and the pellets were washed, dried, and resuspended in formamide loading dye. Samples were heat-denatured at 94 °C for 5 min followed by immediate cooling on ice and loaded onto a denaturing 12% polyacrylamide sequencing gel along with the G + A ladder (36). The gel was dried, and the bands were visualized by Fuji FLA-5000 PhosphorImager.

KMnO₄ Probing—Reaction mixtures (20 μl) contained 20 mM Tris-HCl (pH 7.5), 0.4 mM DTT, 10 nM cognate duplex DNA, 5'-³²P-labeled either at upper or lower strand, and increasing concentrations of PI-MleI. Reaction mixtures were incubated at 37 °C for 30 min after which KMnO₄ was added to a final concentration of 2 mM, and incubation was continued at 25 °C for 2 min. In control reactions, no KMnO₄ was added. Reactions were quenched by the addition of 50 μl of stop solution (1.5 M sodium acetate (pH 5.2), 1 M 2-mercaptoethanol, and 25 μg/μl calf thymus DNA), and the DNA was precipitated with ethanol. The pellet was collected by centrifugation and dried under vacuum. The modified nucleobases were cleaved with piperidine base (1 M) in a final volume of 100 μl at 90 °C for 30 min. Cleaved DNA was collected by centrifugation, and the residual piperidine was removed by repeated washings of the pellet with water and dried under vacuum. The pellets were resuspended in formamide loading buffer and electrophoresed through a denaturing 12% polyacrylamide sequencing gel in TBE buffer at 1800 V. The gel was dried onto a Whatman 3MM paper and exposed to Fuji FLA-5000 PhosphorImager for visualization.

DMS Protection Assay—The binding reactions were performed in the same manner as in the KMnO₄ probing assay. Briefly, reaction mixtures (20 μl) contained 20 mM Tris-HCl (pH 7.5), 0.4 mM DTT, 2 nM ³²P-labeled (upper or lower strand) cognate duplex DNA and increasing concentrations of PI-MleI. Reaction mixtures were incubated at 37 °C for 30 min after which DMS was added to a final concentration of 0.05%, and incubation was continued at 37 °C for 2 min. In control reactions, no protein was added. Reactions were quenched by the addition of 50 μl of stop solution (1.5 M sodium acetate (pH 5.2),

1 M 2-mercaptoethanol, and 0.25 μg/μl yeast tRNA), and the DNA was precipitated with 95% ethanol. The pellet was collected by centrifugation, extensively washed with 70% ethanol, and dried under vacuum. The modified guanines were cleaved with piperidine base (1 M) in a final volume of 70 μl at 90 °C for 30 min. Cleaved DNA was collected by centrifugation, and the residual piperidine was removed by repeated washings of the pellet with water and dried under vacuum. The pellets were resuspended in formamide loading buffer, heat-denatured at 94 °C for 5 min, and electrophoresed through a denaturing 12% polyacrylamide gel in TBE buffer at 1800 V. The gel was dried onto a What-

man 3MM paper and exposed to Fuji FLA-5000 PhosphorImager for visualization.

Fluorescence Measurements—Reaction mixtures (300 μl) contained 50 mM Tris-HCl (pH 7.5), 10 nM cognate duplex with 2-AP substitution, and increasing concentrations of PI-MleI as indicated in the figure legends. Fluorescence excitation and emission spectra were recorded using a Jobin Yvon (Horiba) Fluoromax-3 fluorimeter. The samples were excited at λ_{315 nm}, and the fluorescence emission was recorded in the wavelength range of 330–450 nm at 1-nm intervals. The excitation and emission monochromators were set with bandpasses of 5 and 6 nm, respectively. The emission spectra were corrected for lamp fluctuation and instrumental variation. The background emission as well as tryptophan fluorescence were corrected by subtractions of control spectra, where 2-AP fluorophore was replaced by adenine. Fluorescence measurements were carried out in a 5 × 5 mm cuvette at 30 °C. The binding curves were derived from the integrated spectra from at least three independent measurements and normalized to the value obtained for the modified duplex in the absence of protein.

RESULTS

Sequence Alignment of *M. tuberculosis* and *M. leprae* RecA Intein—*M. tuberculosis* and *M. leprae* recA open reading frames contain an in-frame insertion of intein-coding sequences, thus leading to the synthesis of precursor RecA proteins of 85 and 79 kDa, respectively (8, 9). These undergo protein splicing to generate mature 38-kDa RecA proteins and inteins (9, 10). However, *M. tuberculosis* and *M. leprae* RecA inteins differ in their size, amino acid sequence, and location within the recA gene. As shown in Fig. 1, the sequences of PI-MtuI and PI-MleI can be aligned over their entire length. However, seven distinct regions of the alignment feature gaps in PI-MleI, with longer gaps in the C-terminal region. Pairwise comparison of the deduced amino acid sequence using ClustalW suggested that PI-MleI shares 22% primary sequence identity with PI-MtuI, although both belong to the large family of LAGLIDADG homing endonucleases. Comparative sequence

M. leprae RecA Intein Homing Endonuclease

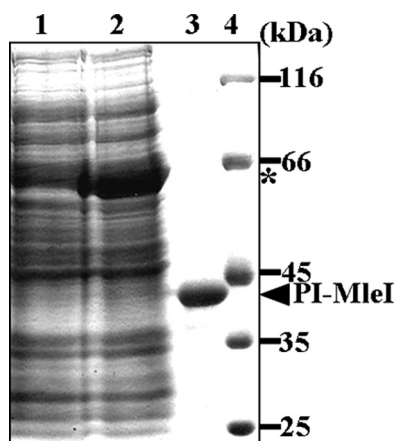


FIGURE 2. SDS-PAGE analysis of purified PI-MleI. Protein samples were analyzed by SDS-PAGE and visualized by staining with Coomassie Blue. Lane 1, uninduced cell lysate; lane 2, induced cell lysate; lane 3, purified PI-MleI; lane 4, molecular mass standards. The sizes of the marker proteins in kDa are indicated on the right-hand side. The arrowhead denotes the cleaved PI-MleI, and the asterisk denotes the fusion protein (PI-MleI with *M. xenopi* GyrA mini intein and chitin binding domain).

analysis also revealed that several positions in the alignment show 80 strictly conserved amino acids and 43 residues conserved in similarity. Importantly, sequence alignment identified the presence of a Walker ATP-binding motif (P-loop ¹²⁶GWVGGKT¹³²) in PI-MtuI, but not in PI-MleI; however, both contained two copies of the consensus LAGLIDADG motif (Fig. 1). The low level of sequence conservation between PI-MtuI and PI-MleI contrasts sharply with the high degree of conservation (>90%) that exists between *M. tuberculosis* and *M. leprae* RecA proteins.

Expression and Purification of Recombinant PI-MleI—To ascertain whether the *M. leprae* *recA* intervening sequence encodes a catalytically active homing endonuclease, we chose to overproduce it in *E. coli* using the IMPACT-T7 system, which allows protein production and purification under native conditions. The recombinant plasmid, pTMLRI, containing *M. leprae* *recA* intervening sequence was transformed into *E. coli* strain ER2566, and expression of the fusion protein was induced by the addition of isopropyl 1-thio- β -D-galactopyranoside. Analysis of cell lysates by SDS-PAGE and Coomassie Blue staining indicated that PI-MleI accumulated as a fusion protein of 66 kDa (Fig. 2, lane 2). A rapid method was developed for purification of PI-MleI involving lysis of cells, chitin-affinity chromatography, and DTT-initiated on-column cleavage under conditions as described under “Experimental Procedures.” The elution of PI-MleI from the chitin affinity matrix was monitored by SDS-PAGE and Coomassie Blue staining, was found to be >97% homogeneous, and corresponded to the predicted size of ~41 kDa (Fig. 2, lane 3). The fractions containing PI-MleI were pooled; protein was precipitated with ammonium sulfate, and the pellet was resuspended in a buffer containing 20 mM Tris-HCl (pH 8), 100 mM NaCl, and 10% glycerol and dialyzed against storage buffer containing 20 mM Tris-HCl (pH 7.5), 100 mM NaCl, 1 mM DTT, and 50% glycerol. The identity of the purified PI-MleI was ascertained by sequencing nine amino acid residues at the N-terminal end and was found to correspond to the *M. leprae* RecA intein-encoding

sequence and Western blot analysis using anti-PI-MleI antibody. Purified PI-MleI was devoid of 5' \rightarrow 3'- or 3' \rightarrow 5'-exonuclease activities (data not shown). Protein cross-linking experiments using glutaraldehyde suggested that PI-MleI exists in solution as a monomer, consistent with the double-motif LAGLIDADG enzymes (18–22) (data not shown).

PI-MleI Binds to Single- and Double-stranded Cognate DNA—To explore the biochemical functions of recombinant PI-MleI, we carried out a detailed investigation of its DNA binding specificity under different conditions. The electrophoretic mobility shift assay serves as a simple and rapid method to study the specificity of interaction between proteins and nucleic acids, and as a tool to measure the extent of the formation of protein-nucleic acid complexes (37, 38). To determine whether PI-MleI is able to recognize different forms of DNA, reactions were first performed with a fixed amount of ³²P-labeled 88-bp *recA* intein-less allele, containing the homing site, hereafter referred to as the cognate DNA, and increasing concentrations of PI-MleI in the absence of divalent cations Mg²⁺ or Mn²⁺. The reaction mixtures were analyzed by electrophoretic mobility shift assay and visualized by autoradiography. Fig. 3A shows that as the concentration of PI-MleI increased from 50 to 500 nM, and robust binding was observed with the formation of a discrete DNA-PI-MleI complex. To learn more about the binding of PI-MleI to DNA, parallel experiments were performed to assess binding of PI-MleI to cognate single-stranded DNA. We incubated 88-mer DNA with increasing concentrations of PI-MleI, and the reaction mixtures were analyzed as described above. Consistent with previous studies with other LAGLIDADG enzymes (12), the extent of PI-MleI binding to single-stranded DNA was substantially less, compared with cognate double-stranded DNA (Fig. 3B). Although ssDNA used in this experiment lacks the secondary structure, we cannot entirely rule out the possibility that the binding is associated with partial duplex generated by self-annealing of complementary regions within the single-stranded DNA. The extent of the formation of PI-MleI-DNA complexes was quantified and expressed as percentage of total DNA bound by PI-MleI (Fig. 3C). These results are consistent with other LAGLIDADG enzymes, including I-Anil (39), PI-Pful (40), PI-PfuII (40), and PI-MtuI (12).

Next, we sought to assess the stability of the PI-MleI-DNA complex by challenging the protein-DNA complex with increasing concentrations of salt. The stability of PI-MleI-DNA complexes was measured by the addition of increasing concentrations of NaCl in the assay buffer. The data showed that the salt titration midpoint for PI-MleI binding for cognate duplex DNA is ~0.3 M and is ~0.1 M for single-stranded DNA (supplemental Fig. 1, A–C). The affinity of PI-MleI to cognate duplex DNA was tested by competition reactions with molar excess of unlabeled cognate duplex DNA (88 bp). The data from a series of experiments indicated that a 10-fold higher concentration of the cold competitor DNA was necessary to reduce specific binding of PI-MleI by 50% (data not shown).

PI-MleI Displays Double Strand-specific Endonuclease Activity—Having established stable and specific binding of recombinant PI-MleI to the cognate duplex DNA, assays for its endonuclease activity were performed as described under “Experimental Procedures.” To this end, we monitored the con-

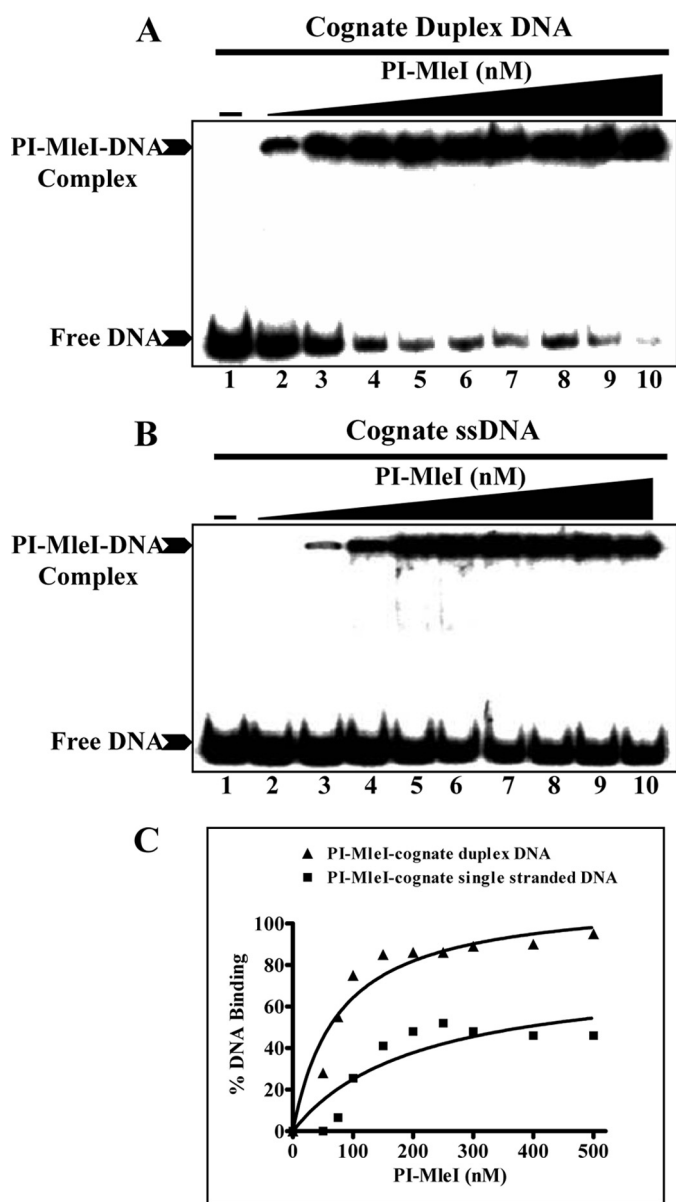


FIGURE 3. PI-MleI binds to both single-stranded and duplex cognate DNA. Reactions were performed with 1 nM ³²P-labeled cognate 88-bp duplex or 88-mer DNA in the absence (lane 1) or presence of 50, 75, 100, 150, 200, 250, 300, 400, or 500 nM PI-MleI (lanes 2–10), respectively. Reaction mixtures were incubated at 37 °C, electrophoresed on native polyacrylamide gels, and visualized by autoradiography as described under “Experimental Procedures.” *A*, binding of PI-MleI to cognate duplex DNA. *B*, binding of PI-MleI to cognate ssDNA. Filled triangle on top of the gel images represents increasing concentrations of PI-MleI. The positions of the unbound DNA and PI-MleI-DNA complexes are indicated on the left-hand side. *C*, graphical representation of PI-MleI binding to cognate duplex or single-stranded DNA (ssDNA). The extent of formation of PI-MleI-DNA complexes in *A* and *B* is plotted versus increasing concentrations of PI-MleI. ▲, PI-MleI-cognate duplex DNA; ■, PI-MleI-single-stranded DNA.

version of negatively supercoiled DNA (form I) to form II and/or form III DNA, a widely used sensitive method for quantification of endonuclease activity (12, 13). Cleavage assays were carried out by mixing PI-MleI with form I pMLR DNA containing the intein-less *recA* allele, the natural recognition sequence of PI-MleI, as the substrate in the presence of 5 mM Mg²⁺ or Mn²⁺. After incubation, the reaction mixtures were deproteinized, and the products were separated by gel electrophoresis

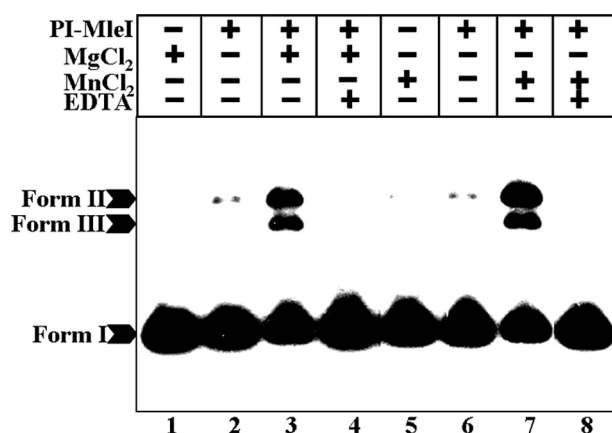


FIGURE 4. PI-MleI is a double-stranded DNA endonuclease. Reaction mixtures (20 μl) containing pMLR DNA (16 μM) were incubated in the absence (lanes 1 and 5) or presence (lanes 2–4 and 6–8) of 500 nM PI-MleI with 5 mM Mg²⁺ (lanes 1, 3, and 4) or 5 mM Mn²⁺ (lanes 5, 7, and 8) or with 10 mM EDTA (lanes 4 and 8). After incubation, reaction mixtures were deproteinized, analyzed by agarose gel electrophoresis, and visualized by Southern hybridization as described under “Experimental Procedures.”

and visualized by Southern hybridization followed by autoradiography. As shown in Fig. 4 (lanes 3 and 7), incubation of pMLR DNA with PI-MleI, in the presence of Mg²⁺ or Mn²⁺, resulted in the appearance of two discrete cleavage products with mobility similar to that of the nicked circular (form II) and linear duplex DNA (form III). PI-MleI was able to generate detectable amounts of form II DNA in the absence of exogenously added Mg²⁺ or Mn²⁺ (Fig. 4, lanes 2 and 6). The generation of form II DNA during the reaction prior to form III DNA product suggests that the reaction catalyzed by PI-MleI proceeds through a two-step reaction. Importantly, PI-MleI cleaved the substrate containing intein-less *recA* allele in the presence of either Mg²⁺ or Mn²⁺, yielding products of the same size. It is likely that the cleavage observed in the absence of exogenously added Mg²⁺ or Mn²⁺ was because of endogenously bound divalent cation. Consistent with this notion, when the enzyme-bound divalent cations were depleted by the addition of EDTA, the reaction was completely inhibited (Fig. 4, lanes 4 and 8). Taken together, these results suggest that in the reaction catalyzed by PI-MleI, nicked circular duplex DNA and unit length linear duplex DNA are produced as the intermediate and final product, respectively.

PI-MleI endonuclease activity was assayed under different conditions, with varying concentrations of divalent cations, pH, and temperature. PI-MleI activity showed a broad divalent cation, pH, and temperature dependence (Fig. 5). PI-MleI displayed optimal catalytic activity between 2 and 5 mM Mn²⁺ (Fig. 5A), in the pH range of 5.5–7.5 (Fig. 5B). In the presence of Mg²⁺, however, the pH range was 6–9.5 (Fig. 5C). Similarly, the optimal cleavage activity occurred at 30–45 °C in the presence of 3 mM Mn²⁺ (Fig. 5D) or Mg²⁺ (data not shown). The maximum catalytic activity of PI-MleI observed in the presence of 1–5 mM Mn²⁺ or Mg²⁺ is comparable with that of other homing endonucleases (12, 41, 42). Although a majority of known homing endonucleases requires 3–5 mM divalent cations for their optimal activity, PI-Tful (*Thermococcus fumicolans*) displays its maximum catalytic activity in the presence of 25 mM Mn²⁺ (43). PI-MleI showed different cleavage efficiencies

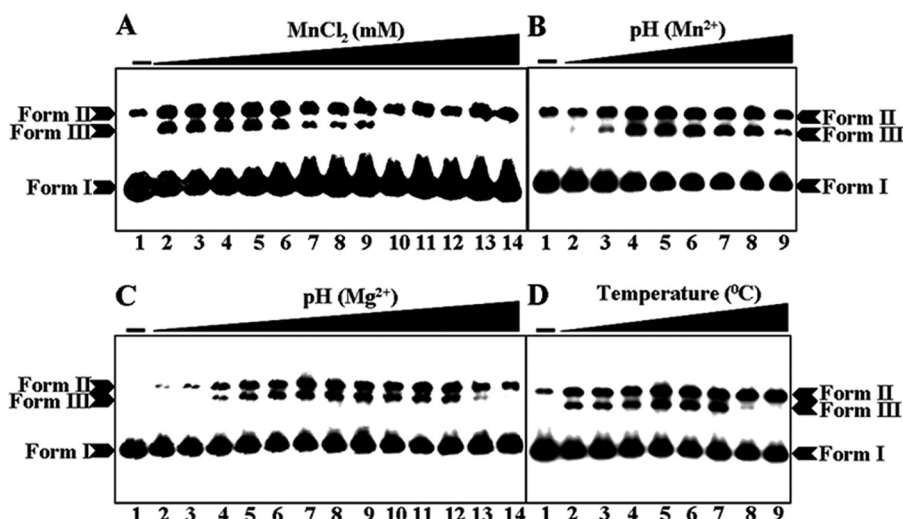


FIGURE 5. Characterization of PI-MleI double-stranded DNA endonuclease activity. *A*, dependence of PI-MleI activity on Mn^{2+} concentration. Reactions were performed in an assay buffer (20 μ l) containing 16 μ M pMLR DNA, 500 nM PI-MleI, and increasing concentrations of Mn^{2+} . Lane 1, DNA substrate. Lanes 2–14 contained substrate with Mn^{2+} at 1, 2, 3, 5, 7.5, 10, 12.5, 15, 17.5, 20, 22.5, 25, and 30 mM, respectively. *B*, pH dependence of PI-MleI activity in the presence of Mn^{2+} . Reactions (20 μ l) were performed with 16 μ M pMLR DNA, 3 mM Mn^{2+} , and 500 nM PI-MleI in buffers of varying pH values. Lane 1, pH 7.5 (buffer control); lanes 2–9 contained buffers at pH 4.5, 5.0, 5.5, 6, 6.5, 7, 7.5, and 8, respectively. *C*, pH dependence of PI-MleI activity in the presence of Mg^{2+} . Reaction conditions same as in *B* except that 3 mM Mg^{2+} , instead of Mn^{2+} , was used as the divalent cation. Lane 1, pH 7.5 (buffer control); lanes 2–14 contained buffers at pH 4.5, 5.0, 5.5, 6, 6.5, 7, 7.5, 8, 8.5, 9, 9.5, 10, and 10.5, respectively. *D*, dependence of PI-MleI activity on temperature. Reactions (20 μ l) contained 16 μ M pMLR substrate DNA, 3 mM Mn^{2+} , and 500 nM PI-MleI. The reaction mixtures were incubated at different temperatures. Lane 1, 37 °C (buffer control); lanes 2–9 incubated at 4, 10, 20, 30, 37, 45, 55, and 60 °C, respectively. In all of the above, reaction mixtures were deproteinized, separated on agarose gels, and visualized as described under “Experimental Procedures.”

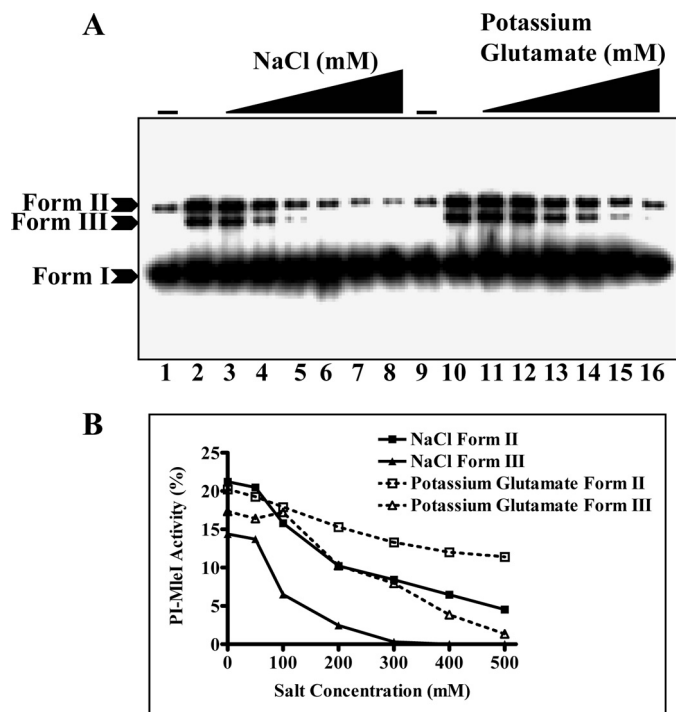


FIGURE 6. Effects of monovalent cations on PI-MleI activity. *A*, reaction mixtures (20 μ l) containing 16 μ M pMLR DNA, 3 mM Mn^{2+} , 500 nM PI-MleI, and increasing concentrations of NaCl or potassium glutamate were incubated at 37 °C for 2 h. Lanes 1 and 9 control reactions; lanes 2–16, as in lane 1 plus NaCl (lanes 3–8) or potassium glutamate (lanes 11–16) at 50, 100, 200, 300, 400, and 500 mM, respectively. *B*, graphical representation of PI-MleI activity as a function of salt concentrations. In the presence of NaCl (■) form II DNA and (▲) form III DNA. In the presence of potassium glutamate (□) form II DNA and (△) form III DNA.

depending on the divalent cation cofactors; at identical concentrations, the activity was higher in the case of Mn^{2+} than Mg^{2+} . Other divalent cations, namely Ca^{2+} , Ba^{2+} , Sr^{2+} , Zn^{2+} , and Ni^{2+} at varying concentrations, allowed partial or no activity (data not shown). The cation dependence of PI-MleI endonuclease activity is similar to other members of LAGLIDADG family of endonucleases, including PI-SceI (44), I-CreI (41), PI-PfuI (42), and PI-MtuI (12).

Previously, we showed that the homing endonuclease activity of PI-MtuI, but not its binding to DNA, was tightly coupled to that of its ATPase activity (12, 45). To investigate the probable effect of nucleotide cofactors on the endonuclease activity of PI-MleI, we performed a series of reactions in the presence of 5 mM Mn^{2+} and 1.5 mM ATP, ATP γ S, or ADP. Consistent with the lack of Walker A and B motifs in PI-MleI, the nucleotide cofactors had no stimulatory effect on the

endonuclease activity (supplemental Fig. 2). However, the nucleotide cofactors caused a modest level of inhibition on PI-MleI endonuclease activity (supplemental Fig. 2, compare lane 2 with lanes 3 and 4). The observed differences in cleavage efficiencies of PI-MleI may likely be due to the nonspecific effect of nucleotide cofactors. To test whether endonuclease activity is either intrinsic to PI-MleI or due to a contaminating endonuclease, two different types of assays were performed. First, deproteinized form III DNA generated by PI-MleI was found to be resistant to further cleavage by PI-MleI. Second, form III DNA produced by PI-MleI on further digestion, with either EheI or Eam1105I, must generate discrete fragments of duplex DNA. As expected, both EheI and Eam1105I cleaved form III DNA into fragments of predicted size (supplemental Fig. 3), suggesting that PI-MleI cleaves at the recognition site located within the intein-less *recA* allele. These experiments excluded the possibility of a contaminating nuclease activity and confirmed that cleavage is a direct consequence of specific cleavage by PI-MleI at the insertion site.

Effect of Ionic Strength on DNA Cleavage by PI-MleI—The assay buffer contained 25 mM Tris-HCl and 3 mM $MnCl_2$ or $MgCl_2$, which is comparable with standard conditions used in studies with other homing endonucleases (12, 13, 41). We therefore wished to test the effect of increasing ionic strength on PI-MleI endonuclease activity. DNA cleavage assays were performed in the assay buffer containing increasing concentrations of NaCl or potassium glutamate. The data in Fig. 6 suggest that DNA cleavage efficiency of PI-MleI was optimal at 50 mM NaCl, but further increases in NaCl concentrations resulted in a rapid decline in activity. On the other hand, PI-MleI was able to

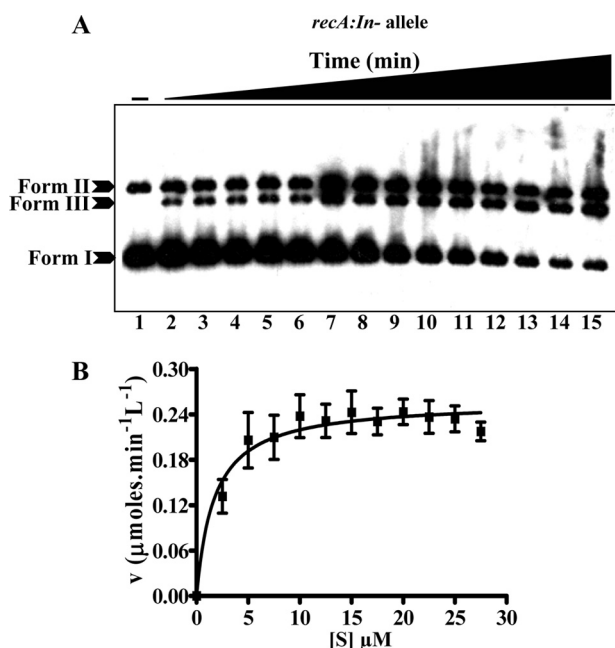


FIGURE 7. Kinetics of cleavage of cognate DNA by PI-MleI. *A*, reaction mixtures containing 16 μM cognate pMLR DNA (85% form I and 15% form II DNA), 3 mM Mn^{2+} , and PI-MleI (500 nM) were incubated for different time intervals as indicated (lanes 2–15, 1, 5, 10, 20, 30, 45, 60, 90, 120, 150, 180, 240, 300, and 360 min, respectively); lane 1, control reaction incubated for 360 min. The reaction mixtures were deproteinized, electrophoresed on an agarose gel, and visualized by Southern hybridization as described under “Experimental Procedures.” *B*, Michaelis-Menten kinetic fit. Reactions were performed by varying the concentrations of the pMLR DNA with a fixed concentration of PI-MleI in the presence of 5 mM Mn^{2+} . The reaction mixtures were deproteinized, electrophoresed on an agarose gel, and visualized by Southern hybridization. The bands were quantified by using UVI-BandMap software version 99, and the rate of the reactions was calculated. Plot of velocity versus substrate concentrations using GraphPad Prism version 4 represents a typical hyperbola curve. The data represent the mean of three independent experiments. Kinetic constants K_m , V_{max} , and k_{cat} values were calculated from the intercepts of the Lineweaver-Burk $1/v$ versus $1/[S]$ plot using the equations $1/v = 1/V_{\text{max}} + K_m/V_{\text{max}}[S]$ and $k_{\text{cat}} = V_{\text{max}}/[E_t]$, wherein v , $[S]$, and $[E_t]$ correspond to velocity of the reaction, substrate, and total enzyme concentrations, respectively. Results are means \pm S.D. of triplicate determinations in a typical experiment performed.

tolerate the presence of relatively higher concentrations of potassium ions. The salt titration midpoint, the concentration at which one-half of the endonuclease activity is inhibited, differs in regard to the generation of form II and form III DNA (Fig. 6B). A similar inhibitory effect of NaCl and potassium glutamate has been reported previously for PI-MtuI activity (12). We note that potassium salts of glutamate and to a lesser extent acetate, which exist at an intracellular concentration of <0.5 M, are considered to be the physiologically relevant intracellular ions (46). The inhibitory effect of salts on the endonuclease activity corroborate with the reduced DNA binding (supplemental Fig. 1).

Kinetics of the DNA Cleavage Reaction—To gain more insights into the mechanism of DNA cleavage, we carried out steady-state kinetics experiments on PI-MleI using pMLR form I DNA as the substrate in the presence of Mn^{2+} . Fig. 7A shows the time course of cleavage reaction. Typically, $\sim 21\%$ cleavage products were generated in a 120-min reaction at 37°C , a result similar to that observed with other LAGLIDADG-type homing endonucleases (12). Fig. 7B shows the plot of substrate concentrations versus cleavage rate. The kinetic constants were deter-

mined from Lineweaver-Burk plots representing initial velocity versus substrate concentrations, and corresponding standard deviations were obtained by linear least squares analysis of the Lineweaver-Burk plots (data not shown). These data indicate that PI-MleI obeyed Michaelis-Menten kinetics, resulting in a K_m of 2.1 μM , V_{max} 0.27 $\mu\text{M}\cdot\text{min}^{-1}$, and a k_{cat} of 0.68 min^{-1} . Each constant represents the average of three experiments. These values are similar to those reported previously for PI-MtuI cleavage of form I DNA containing its noncognate substrate (13). Furthermore, the extremely low k_{cat} value corroborates with that of a number of intein-encoded and intron-encoded homing endonucleases (13, 41, 47).

PI-MleI Cleaves Away from the Intein Insertion Site—To investigate the PI-MleI endonuclease activity in greater detail and to map its cleavage site, we used ^{32}P -labeled 88-bp cognate DNA substrates containing the intein-less *recA* allele. DNA substrates, which were 5' end-labeled with $[\gamma\text{-}^{32}\text{P}]\text{ATP}$ on either the upper or lower strand, were incubated in the absence or presence of increasing concentrations of PI-MleI in standard assay buffer containing Mn^{2+} . The cleavage products were then separated on denaturing polyacrylamide gels alongside sequencing ladders generated from ssDNA of the same 5' end-labeled substrate by Maxam-Gilbert chemical cleavage reaction (36). Upper strand cleavage is shown on Fig. 8A and lower strand cleavage on Fig. 8B. Precise mapping of the cleavage sites in the upper and lower strands of each substrate revealed that PI-MleI inflicts breaks not at the *recA* intein insertion site but away from it. It should be noted that PI-MleI shows a bias in cleavage efficiency; it cleaves at a distal site on the upper strand 4–6-fold more efficiently than at the proximal site. Cleavage on both strands results in a staggered double strand break with long 5' single-stranded overhangs, in contrast to most known LAGLIDADG-type endonucleases that recognize long DNA sequences surrounding the intervening sequences of their own introns/inteins and produce 4-nucleotide 3' overhangs (42, 48, 49). The extent of cleavage was estimated from the density of each band, and they were aligned to allow comparison. As shown in Fig. 8C, cleavage was observed at two sites (at nucleotide positions -44 and -47) on the upper DNA strand (lanes 4–6, Fig. 8A) and at four sites (at nucleotide positions $+16$, $+23$, $+24$ and $+25$) on the lower DNA strand (lanes 4–6, Fig. 8B). Fig. 8C summarizes the results and shows the location of the strand breaks on the cognate DNA containing intein-less *M. leprae recA* sequence.

PI-MleI Binds Asymmetrically to the Sites Flanking the Intein Insertion Site—To gain more detailed insight into boundaries of the cognate DNA sequence bound by PI-MleI, we employed DNase I protection assay, which allows the identification of protein-binding sites at a single base resolution (34, 50). PI-MleI complexes formed with ^{32}P -labeled 88-bp cognate DNA substrate, labeled either on the upper or lower strand, was subjected to limited cleavage by DNase I, and the resulting products were separated by PAGE in the presence of 7 M urea and analyzed by autoradiography. Protection of specific regions could be discerned on both the strands of cognate DNA (Fig. 9, A and B). Superimposition of the protected bases in the cognate DNA substrate showed that the protected area, each covering 16 nucleotide residues on the upper strand and 12 nucleotide

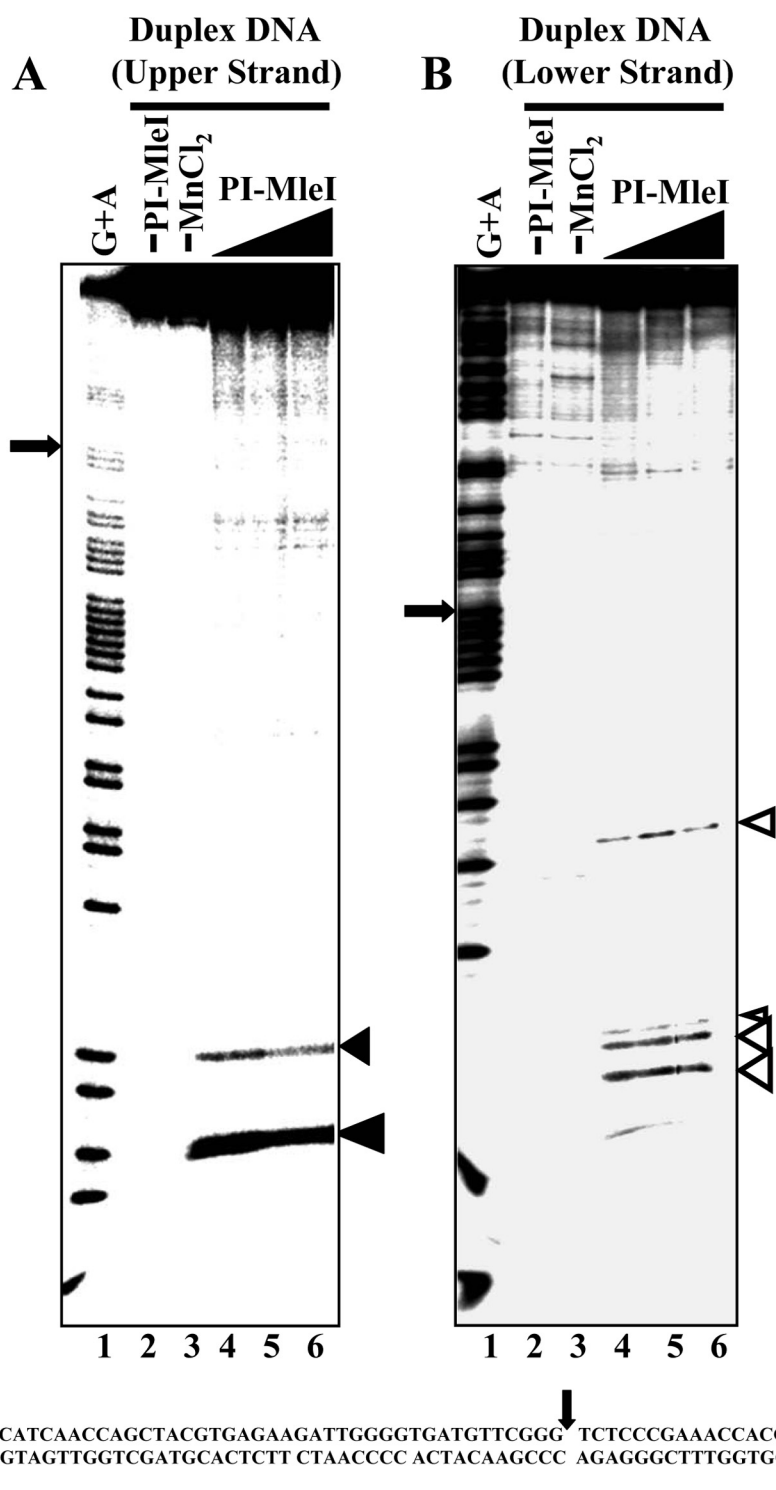
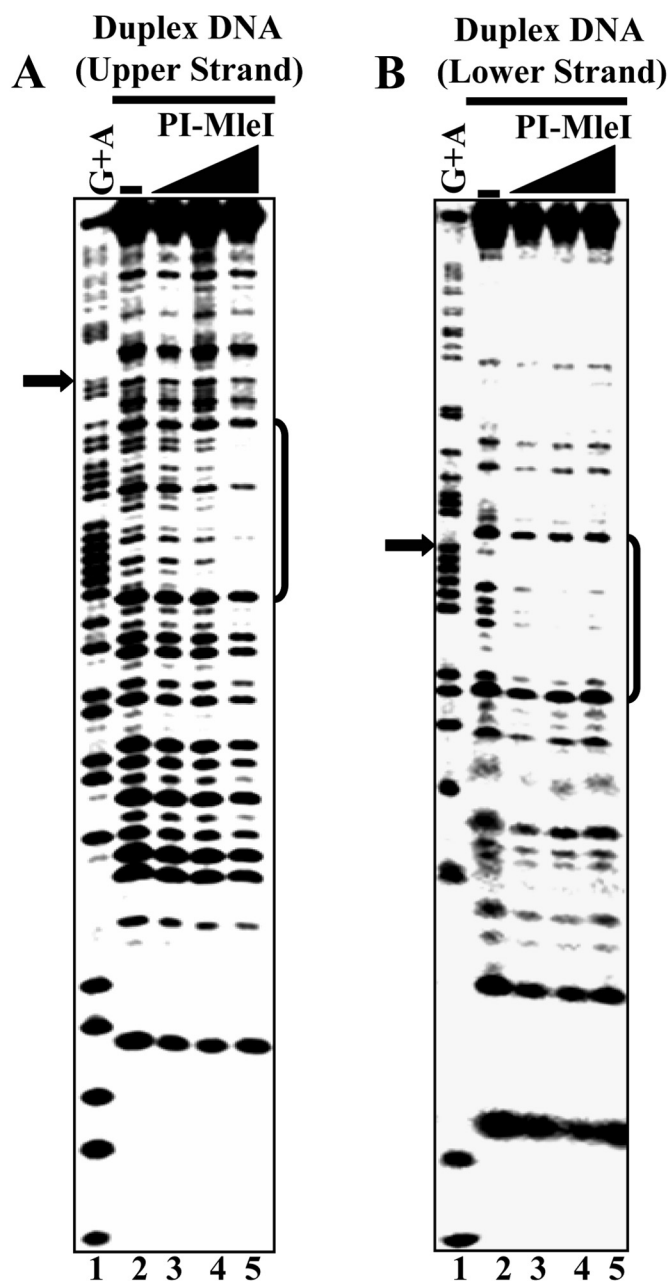


FIGURE 8. **Mapping of PI-MleI cleavage site.** One nm double-stranded DNA fragment (88 bp) encompassing *M. leprae* intein-less *recA* sequence, ³²P-labeled either at upper or lower strand, was incubated with 5 mM Mn²⁺ and increasing concentrations of PI-MleI as described under "Experimental Procedures." *A*, upper strand cleavage, *lane 1*, Maxam-Gilbert sequencing reaction (G + A); *lanes 2 and 3*, in the absence of PI-MleI or Mn²⁺, respectively; *lanes 4–6*, 100, 250, and 500 nM PI-MleI. *B*, lower strand cleavage. The reactions were performed as in *A*, except that the reactions contained ³²P-labeled lower strand cognate duplex. The reaction mixtures were deproteinized, electrophoresed on a 12% denaturing polyacrylamide gel, and visualized as described under "Experimental Procedures." *C*, summary of PI-MleI cleavage. The vertical arrow represents the site of insertion. Black and white arrowheads indicate the position of cleavage sites on the upper and lower DNA strands, respectively.

residues on the lower strand (bases enclosed within the boxed area), is located within two clearly defined nonoverlapping regions flanking the intein-insertion site (Fig. 9C). The two protected areas were separated by an intervening region that

remained unprotected. Such a pattern is characteristic for a sequence-specific DNA binding and indicates that PI-MleI binding to the cognate duplex DNA is unlikely to be mediated by nonsequence-specific recognition of the target site.



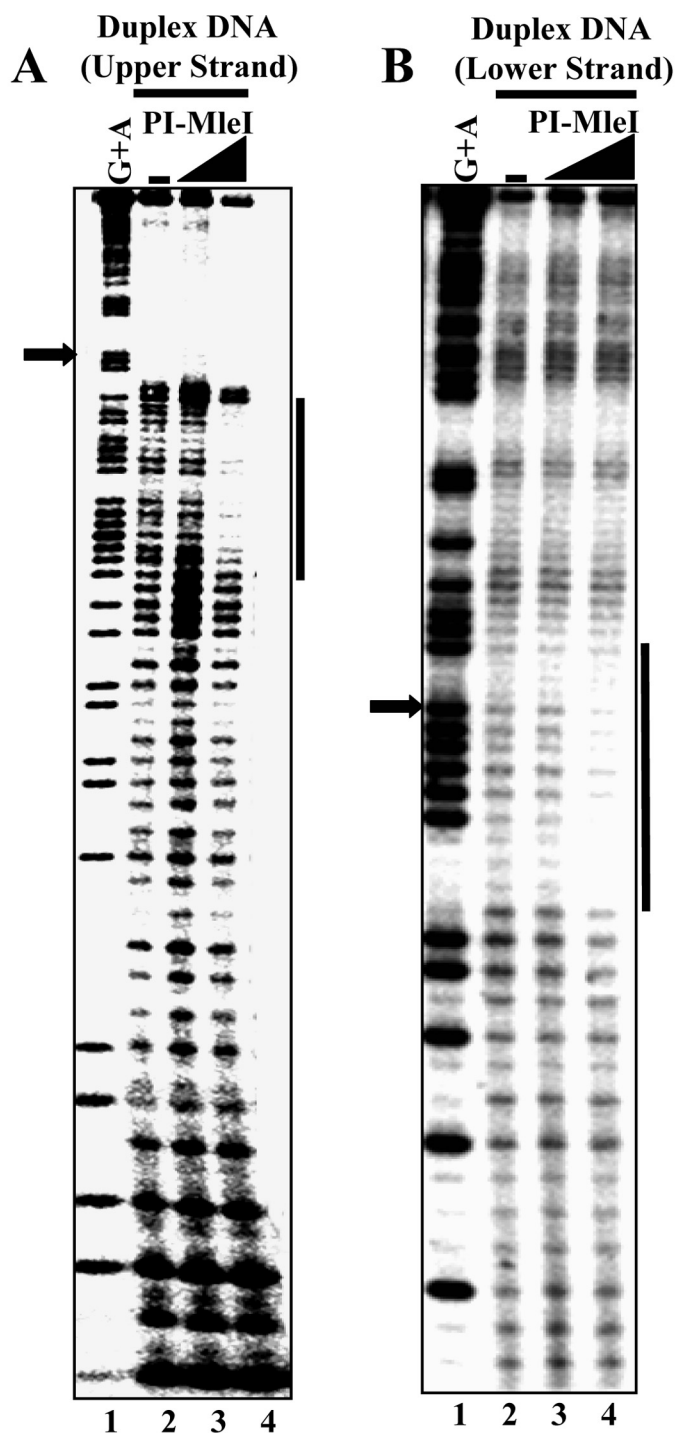
C

5' AA CCACGGCGATCTTCATCAACCAGCTACGTG AGAAGATTGGGGTGAT TTCGGG TCTCCCGAAACCACGACGGGTGGAAAGGCCTTG 3'
 3' TT GGTGCCGCTAGAAGTAGTTGGTCGATGCACCTCTT CTAACCCC ACTACAAGCC AGAGGGCTTTG TGCTGCCACC TTT CCGCAAC 5'

FIGURE 9. Mapping of PI-MleI binding at the homing site by DNase I footprinting. DNase I footprinting was performed on both the upper (A) and lower (B) strands in reactions containing 10 nM duplex DNA (88 bp) encompassing *M. leprae* intein-less *recA* sequence and ³²P-labeled either at the upper or lower strand in the absence (lane 2) or presence of 100, 250, and 500 nM PI-MleI (lanes 3–5). Reactions were terminated, and the samples were electrophoresed on a 12% denaturing polyacrylamide gel and visualized as described under “Experimental Procedures.” Lane 1, Maxam-Gilbert G + A ladder. The brackets mark the protected regions, and the arrow denotes the insertion site. C, summary of footprinting data for the binding of PI-MleI showing the regions of DNase I protection (boxed regions) on both upper and lower DNA strands, respectively.

We next analyzed the pattern of PI-MleI binding to the cognate DNA by (OP)₂Cu²⁺, a chemical nuclease that cleaves the phosphodiester backbone of DNA or RNA (51). The advantage of using (OP)₂Cu²⁺ nuclease compared with DNase I is that being a small molecule the former cleaves closer to the edges of DNA sequences protected by protein binding and thereby gives finer details of the protein-DNA interaction (51). Reactions were performed with ³²P-labeled 88-bp cognate DNA sub-

strate, labeled either on the upper or lower strand, in the absence or presence of PI-MleI, and the reaction products were analyzed as described under “Experimental Procedures.” Protection of specific regions could be discerned on both the strands of cognate DNA (Fig. 10, A and B). The superimposition of the protected bases on both the strands in the cognate DNA substrate revealed the same protected regions as seen with DNase I footprinting (refer to Fig. 9C and Fig. 10C). DNase I



C

5' AA CCACGGCGATCTTCATCAACCAGCTACGTGAGAAGATTGGGGTGATGTTCCGGG TCTCCCGAAACCACGACGGGTGGAAAGGCCGTTG 3'
 3' TT GGTGCCGCTAGAAGTAGTTGGTCGATGCACTCTT CTAACCC ACTACAAGCCC AGAGGGCTTTGGTGCTGCCACC TTT CCGCAAC 5'

FIGURE 10. Mapping of PI-MleI binding at the homing site by $(OP)_2Cu^{2+}$ footprinting. $(OP)_2Cu^{2+}$ footprinting was performed on both the upper (A) and lower strands (B) in reactions containing 10 nM duplex DNA (88 bp) encompassing *M. leprae* intein-less *recA* sequence and ^{32}P -labeled either at the upper or lower strand in the absence (lane 2) or presence of 100 or 250 nM PI-MleI (lanes 3 and 4). Reactions were terminated, and the samples were electrophoresed on a 12% denaturing polyacrylamide gel and visualized as described under "Experimental Procedures." Lane 1, Maxam-Gilbert G + A ladder. The vertical bars denote the protected regions, and the arrow denotes the site of insertion. C, summary of footprinting data for the binding of PI-MleI showing the regions of protection (horizontal bars) on both upper and lower DNA strands, respectively.

protection and $(OP)_2Cu^{2+}$ probing data reveal that PI-MleI binds to nonoverlapping regions of cognate DNA flanking the intein-insertion site asymmetrically, a pattern reminiscent of

I-SceI (52), I-CreI (41), and PI-SceI (53). Taken together, the data clearly demonstrate that the PI-MleI-binding site is located adjacent to the *recA* intein insertion site.

Chemical Probing of PI-MleI-Cognate DNA Complex—The absence of PI-MleI footprint at the cleavage site was surprising. We reasoned that binding might not be stable at this site; however, its interaction might induce variations in DNA structure such as alterations in base-pairing positioning or strand separation. We note that there are examples of enhanced cleavage in the absence of clear footprints. To explore these possibilities, we used KMnO_4 as a probe to analyze the structure of PI-MleI-DNA complexes. KMnO_4 readily oxidizes C-5–C-6 double bond of unstacked or unpaired thymine, and to a lesser extent cytosine (54), and thus is likely to provide insights into the helical distortion of cognate duplex DNA caused by PI-MleI binding. Reactions were performed in the absence or presence of increasing concentrations of PI-MleI, followed by treatment with KMnO_4 , and then the reaction products were analyzed as described under “Experimental Procedures.” Notably, KMnO_4 treatment of PI-MleI-cognate DNA complexes revealed the absence of hypersensitive T residues within the PI-MleI binding regions, either on the upper or lower strands. Interestingly, hypersensitive T residues were detected on both the strands at the cleavage sites (Fig. 11, A and B, lanes 4–7). Surprisingly, hypersensitive T residues were not seen at or near the intein insertion site or in the region between binding and cleavage sites. Under these conditions, control reactions showed no such reactivity of T residues (Fig. 11, A and B, lanes 2 and 3, respectively). Quantitative gel scans of three separate experiments support this assessment. We interpret that the hyper-reactivity of T residues may represent helical distortions arising from the interaction of PI-MleI at the cleavage sites.

PI-MleI Interactions within the Major and Minor Grooves—To gain more information on PI-MleI-DNA interactions, methylation protection assays were carried out using dimethyl sulfate (DMS) as the modifying agent. DMS methylates guanine residues at N-7 in the major groove and adenine residues weakly at N-3 in the minor groove of double-stranded DNA (55). When bound to specific DNA residues, proteins can decrease or intensify the reactivity of purine residues to DMS, compared with naked DNA (56). DMS protection experiments were performed in which PI-MleI was incubated with ^{32}P -labeled cognate duplex DNA, and the resulting complexes were reacted with DMS (see “Experimental Procedures”). Interestingly, two G residues (G^{-19} and G^{-22}) in the upper strand within the DNase I footprint were strongly protected from modification by DMS (see supplemental Fig. 4A). On the other hand, two G residues (G^{+2} and G^{+4}) were protected on the lower strand within the DNase I-footprinted region from methylation (see supplemental Fig. 4B). These results indicate that, although there are interactions of PI-MleI in the major groove, they do not account for the 16 to 12-bp protection observed in the DNase I footprint; instead, they are limited to contacts with specific bases at the recognition sites.

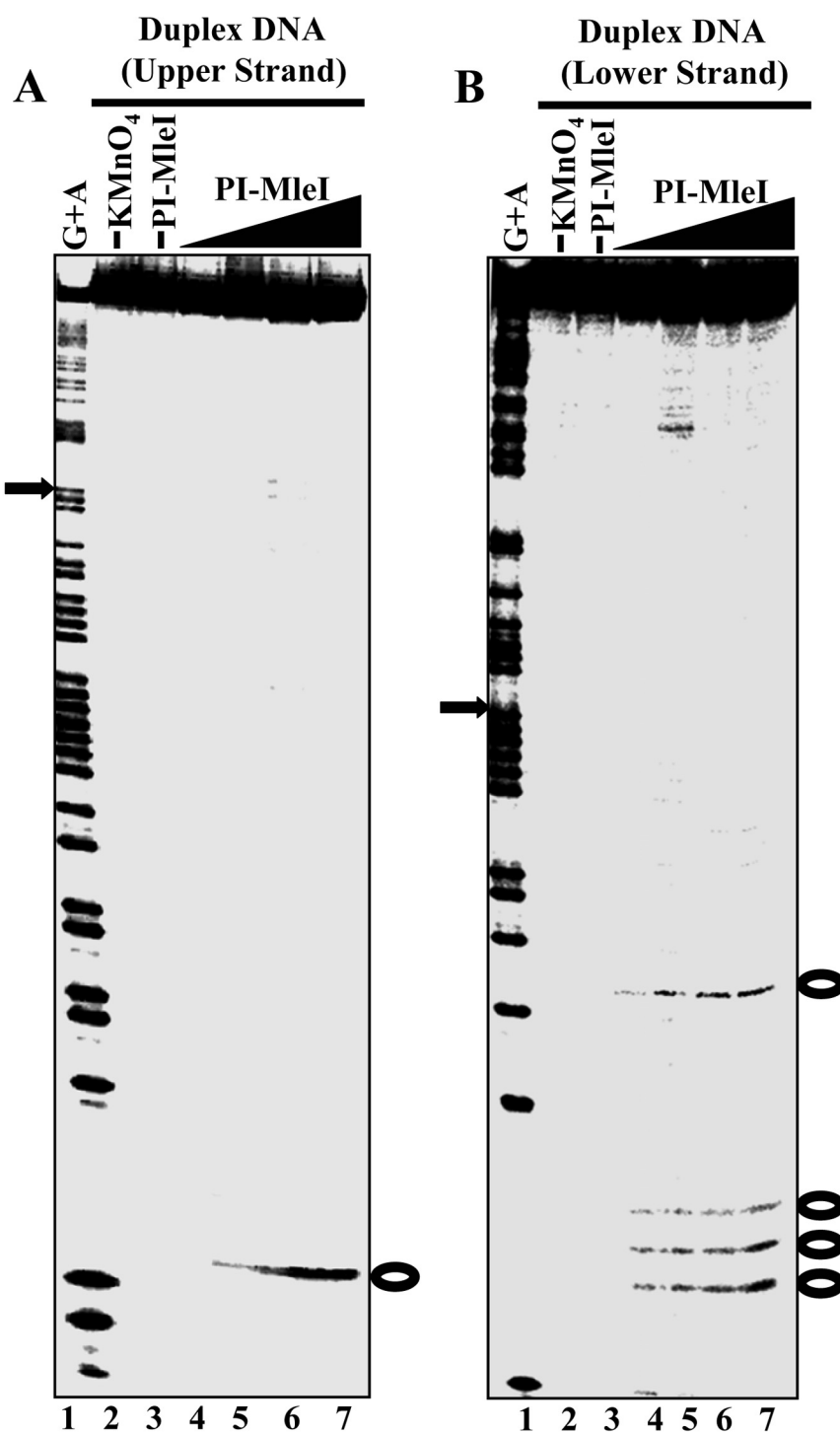
PI-MleI Distorts Cognate DNA Substrate at the Cleavage but Not at the Recognition Region—Because of the differences in the manifestation of hypersensitive T residues between PI-MleI DNA-binding and cleavage sites, we felt it necessary to verify these findings by an independent approach. In addition, we wished to test whether alterations occurred on all substrate DNA molecules or on a small subpopulation of target mole-

cules. To this end, we used six cognate duplex DNA molecules each containing 2-AP at the indicated position (Table 1), as a probe to monitor helical distortions in the target DNA (57). In one substrate, 2-AP was incorporated at the insertion site; and in the second, 2-AP was embedded in the DNA binding region. Similarly, in other substrates, 2-AP was substituted in the spacer region between the binding and cleavage sites and immediately adjacent to the cleavage site in the upper or lower strands. We incubated the 2-AP containing cognate DNA with increasing concentrations of PI-MleI in the assay buffer and monitored the changes in 2-AP fluorescence intensity in the spectral region from 330 to 450 nm. Fig. 12 shows the 2-AP fluorescence emission spectra as a function of increasing concentrations of PI-MleI. The cognate substrate containing 2-AP at the cleavage sites displayed enhancement of fluorescence with increasing PI-MleI concentration, with maximal change of about 6-fold, and then plateaued (Fig. 12, C and D). In contrast, the substrates that contained 2-AP at the insertion site, PI-MleI-binding site, or immediately adjacent to the cleavage sites showed only small changes (Fig. 12, A, B, E, and F). These observations corroborate the hypersensitivity of T residues at the cleavage sites (Fig. 11) and are compatible with the notion of nucleotide flipping/helical distortion, which has been shown to enhance 2-AP fluorescence to varying extents in the presence of a variety of nucleic acid-binding proteins (58–61).

DISCUSSION

In this study, we show that PI-MleI binds to cognate DNA containing the homing site and displays endonuclease activity in the presence of alternative divalent cations, Mg^{2+} or Mn^{2+} . A combination of approaches, including four complementary footprinting assays (DNase I, copper-phenanthroline, KMnO_4 , and methylation protection), enhancement of 2-AP fluorescence, and mapping of the cleavage site revealed that PI-MleI binds to cognate DNA flanking its insertion site, induces helical distortion at the cleavage site(s) but not at the binding site(s), and generates two staggered double strand breaks (Fig. 13). These results suggest that PI-MleI cleavage sites are located at 20 and 4 bp away (on the upper and lower DNA strands, respectively) from its binding site, and implicate that PI-MleI possesses a modular structure with separate domains for DNA target recognition and cleavage (Fig. 13). In summary, these findings disclose that the structural and mechanistic aspects of PI-MleI are distinct from other well characterized LAGLI-DADG-type homing endonucleases.

Inteins in Mycobacteria—Since the first discovery of an intein in *S. cerevisiae* *TFPI* gene (also designated *VMA1*), which encodes the 69-kDa catalytic subunit of the vacuolar H^+ -ATPase (14, 15), intein-encoding sequences have been reported from all three phylogenetic domains as follows: bacteria, eukarya, and archaea (16–19). Sequence analysis of the genomes of 39 mycobacterial strains revealed that intein-encoding sequences are found embedded in their *recA* gene. However, they are inserted at two distinct sites, RecA-a and RecA-b, respectively, at the RecA-a site in *M. tuberculosis* and at the RecA-b site in *Mycobacterium chitae*, *Mycobacterium fallax*, *Mycobacterium gastri*, *Mycobacterium shimodei*, and *Mycobacterium thermoresistibile*. The latter corresponds with the *M.*



C
 5' AA CCACGGCGA **○** CTTCATCAACCAGCTACGTGAGAAGATTGGGGTGATGTTCCGG **↓** TCTCCCGAAACCACGACGGGTGG **▲▲** GGGCGTTG 3'
 3' TT GGTGCCGCTAGAAGTAGTTGGTCGATGCACTCTT CTAACCC ACTACAAGCCC AGAGGGCTTTGGTGC **○** GCCCACC **○○○** CCGCAAC 5'

FIGURE 11. KMnO_4 probing reveals structural distortions induced by PI-MleI. Reaction mixtures containing 10 nM double-stranded DNA (88 bp) encompassing *M. leprae* intein-less *recA* sequence, ^{32}P -labeled either at upper or lower strand, were incubated in the absence (*lane 3*) or presence of PI-MleI (*lanes 4–7*), followed with 2 mM KMnO_4 . Reactions were terminated by the addition of stop solution as described under “Experimental Procedures.” The modified bases were cleaved and resolved on a 12% denaturing polyacrylamide gel in the presence of 7 M urea. *Lane 1*, Maxam-Gilbert G + A ladder; *lanes 2 and 3*, duplex DNA ^{32}P -labeled either at the upper (*A*) or lower strand (*B*) incubated alone or with KMnO_4 , respectively. *Lanes 4–7*, complete reactions containing PI-MleI at 50, 100, 250, and 500 nM, respectively. *C*, summary of KMnO_4 probing data showing reactive T residues on both upper and lower DNA strands (*circles*) induced by the binding of PI-MleI.

leprae RecA allelic family (62). The occurrence of inteins in the two obligate mycobacterial pathogens, *M. tuberculosis* and *M. leprae*, was initially thought to be associated with virulence (9).

However, subsequent studies showed that the presence of inteins does not correlate with specific characteristics of the species such as pathogenicity or growth rate (62). Comparative

TABLE 1
Oligonucleotides used in this study

The underlined A residues indicate the position of 2-AP substitution in the oligonucleotide sequence.

Oligonucleotide	Sequence
ODN1	5'–AACCACGGCGATCTTCATCAACCAGCTACGTGAGAAGATTGGGGTGATGTTCCGGTCTCCCGAAACCACGACGGGTGAAAGGCGTTG–3'
ODN2	5'–CAACGCCTTTCACCCGTCGTGGTTTCGGGAGACCCGAACATCACCCCAATCTTCTCACGTAGCTGGTTGATGAAGATCGCCGTGGTT–3'
ODN3	5'–AACCACGGCGATCTTCATCAACCAGCTACGTGAGAAGATTGGGGTGATGTTCCGGTCTCCCGAAACCACGACGGGTGAAAGGCGTTG–3'
ODN4	5'–AACCACGGCGATCTTCATCAACCAGCTACGTGAGAAGATTGGGGTGATGTTCCGGTCTCCCGAAACCACGACGGGTGAAAGGCGTTG–3'
ODN5	5'–AACCACGGCGATCTTCATCAACCAGCTACGTGAGAAGATTGGGGTGATGTTCCGGTCTCCCGAAACCACGACGGGTGAAAGGCGTTG–3'
ODN6	5'–AACCACGGCGATCTTCATCAACCAGCTACGTGAGAAGATTGGGGTGATGTTCCGGTCTCCCGAAACCACGACGGGTGAAAGGCGTTG–3'
ODN7	5'–CAACGCCTTTCACCCGTCGTGGTTTCGGGAGACCCGAACATCACCCCAATCTTCTCACGTAGCTGGTTGATGAAGATCGCCGTGGTT–3'
ODN8	5'–CAACGCCTTTCACCCGTCGTGGTTTCGGGAGACCCGAACATCACCCCAATCTTCTCACGTAGCTGGTTGATGAAGATCGCCGTGGTT–3'

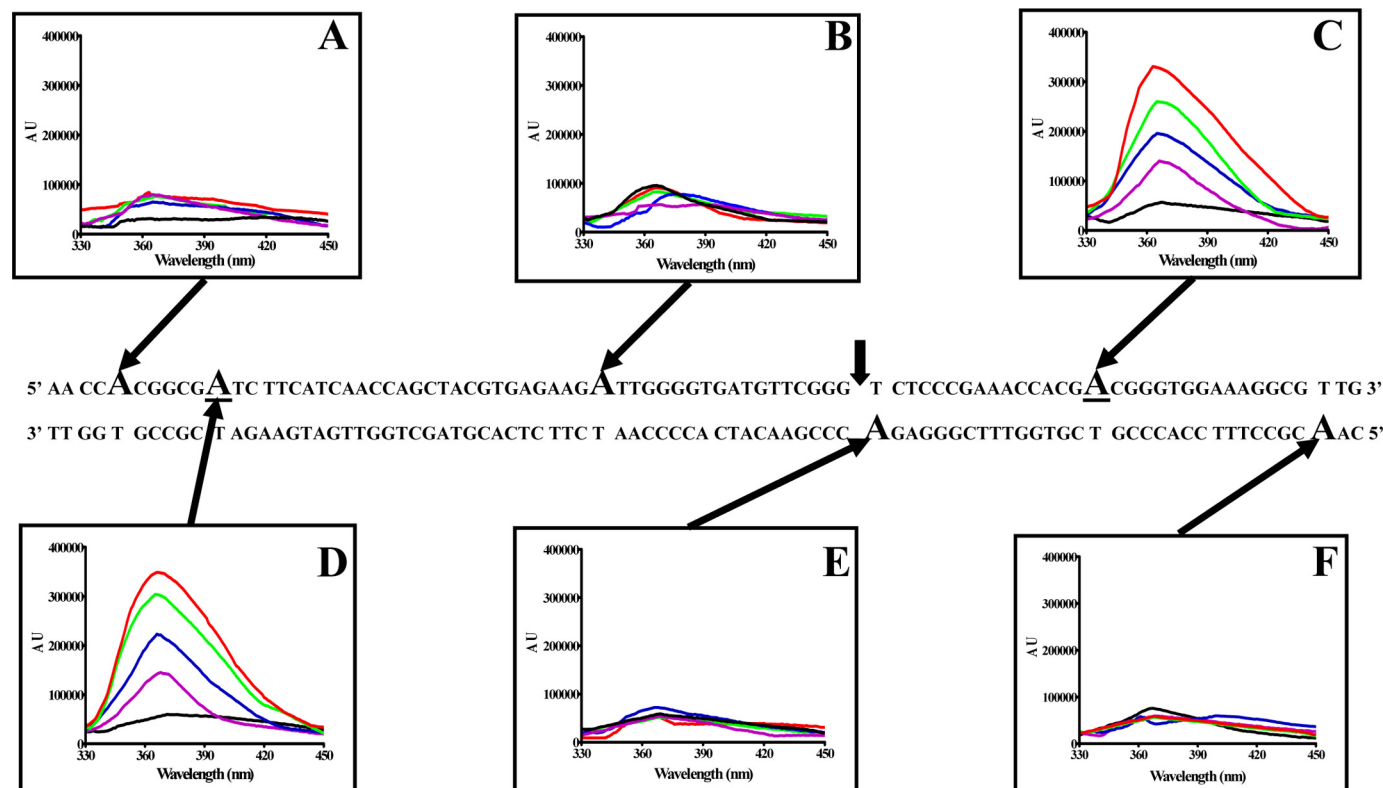


FIGURE 12. PI-MleI induces structural distortions at the homing site as deduced from enhancement of 2-AP fluorescence. Reaction mixtures contained 10 nM double-stranded DNA (88 bp) encompassing *M. leprae* intein-less *recA* sequence bearing a single 2-AP substitution at the indicated site (center panel, large A) and increasing concentrations of PI-MleI. Fluorescence emission spectra were recorded between 330 and 450 nm wavelength region as a function of increasing concentrations of PI-MleI. The y axis indicates fluorescence intensity changes (in arbitrary units) in the absence (black line) or upon incubation with PI-MleI at 100 nM (purple line), 250 nM (blue line), 500 nM (green line), or 1 μM (red line), respectively. Fluorescence emission spectrum of each duplex variant is pointed with black arrow to analogous 2-AP residue in the cognate duplex DNA depicted at the center.

analysis revealed that the degree of primary sequence conservation between PI-MtuI and PI-MleI is 22%, although both belong to the large family of LAGLIDADG homing endonucleases. *M. tuberculosis* inactive RecA precursor undergoes splicing in *E. coli*, whereas splicing of *M. leprae* precursor RecA does not occur in *E. coli*, but mature RecA protein is generated in *Mycobacterium smegmatis* (9–11). Intriguingly, in contrast to the genomes of other obligate parasites, the degenerate *M. leprae* genome retains the *recA* intervening sequence; it would thus be interesting to check if it encodes a catalytically active LAGLIDADG-type homing endonuclease. The dodecapeptide sequence is positioned not only in homing endonucleases encoded by group I and archaeal introns but also in homothallic switching endonuclease (63). Enzymes possessing the LAGLIDADG motif cleave DNA within their recognition sequences to leave 4-base 3'-hydroxyl overhangs (42, 48, 49). The recogni-

tion sequences are generally asymmetrical and long, with sizes of 12–40 bp (18–22).

DNA Binding Properties of PI-MleI—Several lines of evidence suggest that homing endonucleases bind DNA in a site-specific but sequence-tolerant fashion (18–22). Sequence homology comparisons revealed that PI-MleI belongs to the family of LAGLIDADG homing endonucleases (25, 26). Our electrophoretic mobility shift assay results showed that PI-MleI binds specifically to cognate duplex DNA. A distinct complex was seen in these gel shift experiments, and the complex was disrupted at relatively low concentrations of NaCl. The quantitative assessment of its substrate specificity indicates that the presence of the homing site in the cognate DNA is required to achieve maximal binding affinity. Using a similar strategy, we observed that PI-MleI binds to single-stranded DNA containing the homing site, albeit less efficiently, which is in good

M. leprae RecA Intein Homing Endonuclease

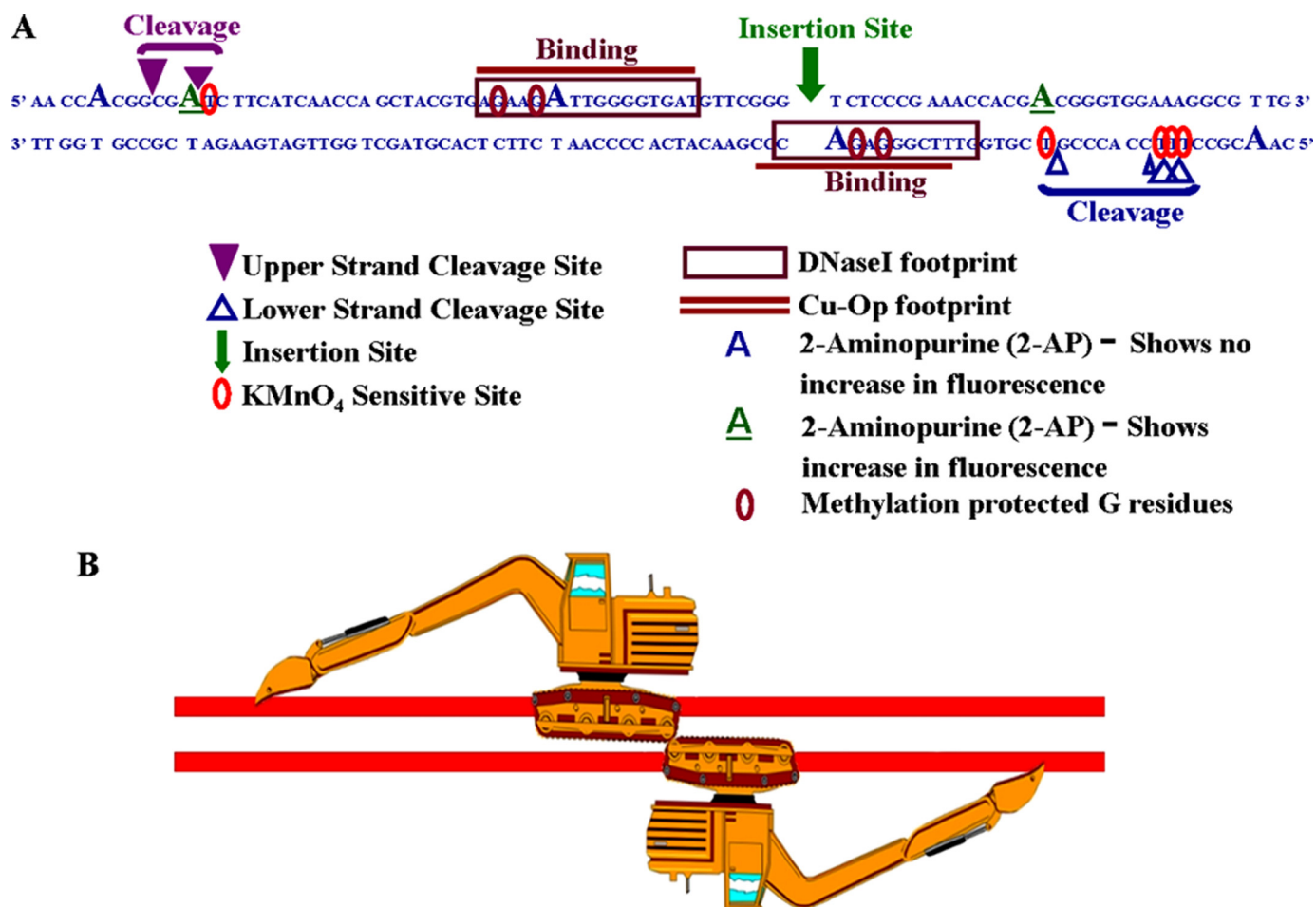


FIGURE 13. **Schematic illustration of the interactions between PI-MleI and its homing site.** A, PI-MleI binding and cleavage sites surrounding the homing region as inferred from DNase I and $(OP)_2Cu^{2+}$, DMS footprinting, $KMnO_4$ probing, mapping of the cleavage site, and enhancement of 2-AP fluorescence data. B, earth digger schematic with a stationary engine, the connecting arm, and the hydraulic digger. The two thick parallel red lines denote double-stranded DNA. The DNA binding domain of PI-MleI is indicated by the stationary engine and the endonuclease domain by digger. For details, see the text.

agreement with the substrate specificity of PI-MtuI (12) and other LAGLIDADG-type homing endonucleases. This reduction in binding specificity to single-stranded DNA correlates with its inability to exhibit single-strand nicking activity. These results suggest that binding of PI-MleI to double-stranded cognate DNA is of potential functional significance, whereas its binding to single-stranded cognate DNA could be fortuitous.

DNA Target Recognition and Cleavage by PI-MleI—Many homing endonucleases are known to have a single biologically relevant DNA target sequence, the homing site, centered on the intron/intein insertion site of intron/intein-less alleles (18, 19). Our results suggest that PI-MleI, like other LAGLIDADG-type homing enzymes, is a site-specific endonuclease. Like other members of the family of homing endonucleases, the activity was seen in the presence of both magnesium and manganese; however, slightly higher activity was observed in the presence of manganese. Homing endonucleases recognize and cleave widely divergent intron/intein insertion sites ranging from 15 to 40 bp (18–23). The longer recognition sequence of these enzymes is believed to ensure that the cleavage of the host genome is minimized, because the target sequences are likely to be present in a single or few copies (18–23). PI-MleI introduced staggered double strand breaks in the homing site by nicking in

the left flanking sequence 44–47 bp and in the right flanking sequence 16–25 bp, away from the intein insertion site. Interestingly, PI-MleI shows differences in the extent of cleavage between the left and right sites flanking the intein insertion site. One explanation for the reactivity differences is that PI-MleI initiates the process of insertion of the intein into the intein-less allele on the left cleavage site.

Although it seems unusual, we note that similar cleavage patterns have been observed for several other homing endonucleases. For example, I-TevI cleaves its homing substrate at 23 bp in the left flanking sequence (64, 65); I-TevII cuts its homing site at 15 bp in the right flanking sequence (64); PI-MtuI cleaves its homing site 24 bp in the left flanking sequence (12), and PI-MgaI recognizes 22-bp region around its intein insertion site preferentially in the left flanking sequence (48). These observations are in accord with many homing endonucleases from archaeal, eubacterial, and eukaryotic organisms (Table 2). Mutagenesis and co-crystal structures have shown that the sequence recognition is flexible and that both sides of the cleavage sites are important (66–68). Also, the intein-encoded LAGLIDADG-type enzymes have extended C-terminal sequences as compared with the intron-encoded species making it likely for PI-MleI to recognize and cleave sequences at a dis-

TABLE 2
Comparison of cleavage sequence profiles of homing endonucleases

The vertical arrows on top of the table, in the PI-MtuI and PI-MleI columns, indicate intron or intein insertion site. The arrowhead and underscore symbols adjoining the nucleotide sequences indicate the cleavage sites on the upper and lower strands, respectively. I-corresponds to intron-encoded and PI- to protein-intron (intein)-encoded homing endonucleases, respectively.

I-CreI	GCTGGTTCAAACGTT	C _↓ GTGA [▼] GACAGTTTGGT
I-DmoI	AATGCTTGC [▼] CGG _↓ GTA	A [▼] GTTCCGGCGCGCATG
I-PpoI	GTA [▼] ACTATGACTCTC _↓ T	TAA [▼] GGTAGC [▼] AAATGC
I-SceI	AAGTTACGCTAGGG _↓ AT	AA [▼] CAGGGTAATATAGC
I-TevI	CA _↓ AC [▼] GCTCAGTAGATGTTTCTTTGGT	CTACCGTTTAATATTG
I-TevII	TTCCAAGCTTATGAGT	ATGAAGTGAACAC _↓ GT [▼] TATTC
I-TevIII	T [▼] TA _↓ TGTATCTTTGCGT	GTACCTTTAACTTCCA
PI-MtuI	AA _↓ CGCGTTCGGCAACCGCACC [▼] CGGGTCACGGTCGTC _↓ AAAGAACAAG [▼] TGTT	
PI-MleI	CGG [▼] CGATCTTCATCAACCACTAGTGAAGAGATGGGGGTGATGTCGGG _↓ TCTCCGAAACCAAGCGGGTGGAA _↓ AGG	
PI-SceI	TATCTATGTCGG _↓ GTGC [▼]	GGAGAAAGAGGTAATG
PI-TliI	GGTTCTTTATGCGG _↓ AC	AC [▼] TGACGGCTTTTATG
PI-PfuI	TACAGAAGATGGGAGGA _↓ GGG	A [▼] CCGGACTCAACTTCTCAA
PI-PfuII	CGATAAGGGCAACGAATCCA	TGTG _↓ GAGA [▼] AGAGCCTCTATA
PI-TfuI	CTTATTTAGATTT _↓ AGG	T [▼] CGCTATATCCTTCGATT
PI-TfuII	TTAAAGTGTCTACGCGG _↓ AT	AC [▼] AGACGGCTTTTCGCAAC
PI-Mgal	CGTAGCTGCCA _↓ GTAT [▼] GAG	TCAGAGGTGG
PI-PabI	ACAACACTAGCGGGGACAG _↓ C	AGT [▼] GGTCCGGTTTCGGTCA
PI-PabII	ACAATAAGGGGTCAACCC	TGTG _↓ GAGA [▼] GGAGCCCTCTA

▼ Cleavage site in upper strand ▬ Cleavage site in lower strand ↓ Intron- or Intein-insertion site

tance. Although these results fulfill the criteria of LAGLIDADG-type enzymes, it should be noted that PI-MleI was able to convert 20–25% of substrate to the product as compared with the robust conversion of 70–80% substrate by other LAGLIDADG-type homing enzymes. One possible explanation for the low efficiency of PI-MleI cleavage might be due to a more constrained structure adopted by PI-MleI as compared with other LAGLIDADG-type homing enzymes. In this context, previous studies have shown that the rates of nucleic acid synthesis are 10–18-fold lower in *M. tuberculosis* as compared with *E. coli* (69, 70). Alternatively, it is possible that the cleavage efficiency exhibited by PI-MleI *in vivo* may exceed the *in vitro* activity through the participation of accessory factors. However, it is not clear why there is a spacer region embedded between DNA target recognition and cleavage, and this point requires further study.

Molecular Architecture of the PI-MleI Homing Site—To effect cleavage of its cognate DNA substrate, PI-MleI, like the other members of the LAGLIDADG family, requires a long and asymmetric sequence. The data from footprinting experiments provided information about the sequence specificity and groove occupancy of PI-MleI. As in the case of I-SceI (52) and other proteins, the copper-phenanthroline footprinting approach further suggests the involvement of minor groove interactions in target recognition by PI-MleI (51). The data from both the footprinting approaches unveiled the interaction of PI-MleI with a region upstream and downstream of its own insertion site, conferring protection to 16 nucleotide residues on the upper strand and 12 nucleotide residues on the lower strand, respectively. The asymmetric footprints have been observed earlier for other LAGLIDADG-type endonucleases such as I-CreI (41) and I-SceI (52) wherein protection on the complementary strands was found to be out of register by 2–3 nucleotides, respectively. In case of PI-MleI, however, the footprint formed on the complementary strands of the homing site

is nonoverlapping, indicating the asymmetric mode of interaction of the enzyme. Paradoxically, these two methods failed to detect functional interaction of PI-MleI at the cleavage site. KMnO₄ probing experiments unveiled the interaction of PI-MleI at the cleavage site leading to the helical distortion of DNA sequence around the cleavage sites. In agreement with this, 2-AP fluorescence measurements provided corroborative evidence that PI-MleI binding was accompanied by helical distortion at the cleavage sites. The nonoverlapping binding of PI-MleI with its target sequence may be facilitated by its ability to stabilize significant distortions upon binding to two very distant regions of the cognate DNA substrate. Several lines of evidence suggest that enzymes that bind duplex DNA distort the substrate to widen the minor groove at the cleavage site and make the scissile phosphates accessible to the enzyme active site (52). Accordingly, the stabilization of a distorted DNA double helix appears to be a common requirement for the homing endonucleases (52, 53). The crystal structures of PI-SceI (71), I-DmoI (72), and PI-PfuI (73) and of the proteins I-CreI or I-SceI complexed to their DNA target (66, 68) suggest that these enzymes use a similar mechanism to recognize and cleave their long DNA targets.

Taken together, our data provide compelling evidence that the PI-MleI homing site can be divided into two regions as follows: DNA target recognition and cleavage regions (Fig. 13), and each region interacts differently with the enzyme. Furthermore, the schematic view also provides a composite picture of DNA binding, helical distortion, and cleavage of PI-MleI as deduced from DNase I, (OP)₂Cu²⁺, DMS footprinting, KMnO₄ probing, enhancement of 2-AP fluorescence, and mapping of the cleavage sites. We note that the molecular architecture of PI-MleI is seemingly analogous to I-TevI, *i.e.* consisting of three distinct modules as follows: DNA binding domain, endonuclease domain, and a linker connecting the two domains. Recent studies have also shown that I-TevI functions as a transcriptional repressor and inhibits its own expression through direct association with the operator/promoter (74).

In summary, our investigations show that PI-MleI possesses a modular structure with separate domains for DNA target recognition and cleavage, each with distinct sequence preferences. *In vivo*, the modular structure of PI-MleI may minimize the potentially harmful effects of nonspecific cleavages in the host genome, while maximizing its ability to recognize and cleave cognate DNA or closely related variants. For example, there seems to be no sites for PI-MleI in the *E. coli* genome because overproduction of the enzyme had no toxic effects on growth. The highly modular structures and tightly coupled mechanisms of DNA recognition and cleavage of homing enzymes are being exploited as potential reagents for gene targeting in cultured cells or organisms to re-engineer their genomes *in vivo* (75). The long and asymmetric sequence required by PI-MleI provides clues as to how the LAGLIDADG-type enzymes can rapidly gain new DNA specificity to regulate the movement of their encoding genes to new genomic sites.

REFERENCES

1. Cole, S. T., Eiglmeier, K., Parkhill, J., James, K. D., Thomson, N. R., Wheeler, P. R., Honoré, N., Garnier, T., Churcher, C., Harris, D., Mungall,

M. leprae RecA Intein Homing Endonuclease

- K., Basham, D., Brown, D., Chillingworth, T., Connor, R., Davies, R. M., Devlin, K., Duthoy, S., Feltwell, T., Fraser, A., Hamlin, N., Holroyd, S., Hornsby, T., Jagels, K., Lacroix, C., Maclean, J., Moule, S., Murphy, L., Oliver, K., Quail, M. A., Rajandream, M. A., Rutherford, K. M., Rutter, S., Seeger, K., Simon, S., Simmonds, M., Skelton, J., Squares, R., Squares, S., Stevens, K., Taylor, K., Whitehead, S., Woodward, J. R., and Barrell, B. G. (2001) *Nature* **409**, 1007–1011
2. Brosch, R., Gordon, S. V., Eiglmeier, K., Garnier, T., and Cole, S. T. (2000) *Res. Microbiol.* **151**, 135–142
3. Cole, S. T. (2002) *Eur. Respir. J.* **36**, S78–S86
4. Kowalczykowski, S. C., Dixon, D. A., Eggleston, A. K., Lauder, S. D., and Rehrauer, W. M. (1994) *Microbiol. Rev.* **58**, 401–465
5. Lusetti, S. L., and Cox, M. M. (2002) *Annu. Rev. Biochem.* **71**, 71–100
6. Schlacher, K., Pham, P., Cox, M. M., and Goodman, M. F. (2006) *Chem. Rev.* **106**, 406–419
7. Chen, Z., Yang, H., and Pavletich, N. P. (2008) *Nature* **453**, 489–504
8. Davis, E. O., Sedgwick, S. G., and Colston, M. J. (1991) *J. Bacteriol.* **173**, 5653–5662
9. Davis, E. O., Thangaraj, H. S., Brooks, P. C., and Colston, M. J. (1994) *EMBO J.* **13**, 699–703
10. Davis, E. O., Jenner, P. J., Brooks, P. C., Colston, M. J., and Sedgwick, S. G. (1992) *Cell* **71**, 201–210
11. Frischkorn, K., Springer, B., Böttger, E. C., Davis, E. O., Colston, M. J., and Sander, P. (2000) *J. Bacteriol.* **182**, 3590–3592
12. Guhan, N., and Muniyappa, K. (2002) *J. Biol. Chem.* **277**, 16257–16264
13. Guhan, N., and Muniyappa, K. (2002) *J. Biol. Chem.* **277**, 40352–40361
14. Kane, P. M., Yamashiro, C. T., Wolczyk, D. F., Neff, N., Goebel, M., and Stevens, T. H. (1990) *Science* **250**, 651–657
15. Hirata, R., Ohsumi, Y., Nakano, A., Kawasaki, H., Suzuki, K., and Anraku, Y. (1990) *J. Biol. Chem.* **265**, 6726–6733
16. Pietrovski, S. (2001) *Trends Genet.* **17**, 465–472
17. Perler, F. B. (2002) *Nucleic Acids Res.* **30**, 383–384
18. Belfort, M., Derbyshire, V., Stoddard, B. L., and Wood, D. W. (eds) (2005) *Homing Endonucleases and Inteins (Nucleic Acids and Molecular Biology)*, Vol. 16, pp. 1–377, Springer-Verlag, Berlin Heidelberg, Germany
19. Guhan, N., and Muniyappa, K. (2003) *Crit. Rev. Biochem. Mol. Biol.* **38**, 199–248
20. Belfort, M., and Roberts, R. J. (1997) *Nucleic Acids Res.* **25**, 3379–3388
21. Jurica, M. S., and Stoddard, B. L. (1999) *Cell. Mol. Life Sci.* **55**, 1304–1326
22. Chevalier, B. S., and Stoddard, B. L. (2001) *Nucleic Acids Res.* **29**, 3757–3774
23. Stoddard, B. L. (2005) *Q. Rev. Biophys.* **38**, 49–95
24. Zhao, L., Bonocora, R. P., Shub, D. A., and Stoddard, B. L. (2007) *EMBO J.* **26**, 2432–2442
25. Perler, F. B., Olsen, G. J., and Adam, E. (1997) *Nucleic Acids Res.* **25**, 1087–1093
26. Dalgaard, J. Z., Klar, A. J., Moser, M. J., Holley, W. R., Chatterjee, A., and Mian, I. S. (1997) *Nucleic Acids Res.* **25**, 4626–4638
27. Sambrook, J., Fritsch, E. F., and Maniatis, T. (1989) *Molecular Cloning: A Laboratory Manual*, 2nd Ed., Cold Spring Harbor Laboratory Press, Cold Spring Harbor, NY
28. Heckman, K. L., and Pease L. R. (2007) *Nat. Protoc.* **2**, 924–932
29. Cunningham, R. P., DasGupta, C., Shibata, T., and Radding, C. M. (1980) *Cell* **20**, 223–235
30. Southworth, M. W., Amaya, K., Evans, T. C., Xu, M. Q., and Perler, F. B. (1999) *BioTechniques* **27**, 110–114, 116, 118–120
31. Bakshi, R. P., Galande, S., Bali, P., Dighe, R., and Muniyappa, K. (2001) *J. Mol. Endocrinol.* **26**, 193–206
32. Bradford, M. M. (1976) *Anal. Biochem.* **72**, 248–254
33. Southern, E. (2006) *Nat. Protoc.* **1**, 518–525
34. Galas, D. J., and Schmitz, A. (1978) *Nucleic Acids Res.* **5**, 3157–3170
35. Ausubel, F. M., Brent, R., Kingston, R. E., Moore, D. D., Seidman, J. G., Smith, J. A., and Struhl, K. (eds) (1987) *Current Protocols in Molecular Biology*, Greene Publishing Associates and Wiley-Interscience, John Wiley & Sons, Inc., New York
36. Maxam, A. M., and Gilbert, W. (1980) *Methods Enzymol.* **65**, 499–560
37. Carey, J. (1991) *Methods Enzymol.* **208**, 103–117
38. Hellman, L. M., and Fried, M. G. (2007) *Nat. Protoc.* **2**, 1849–1861
39. Eastberg, J. H., McConnell, Smith, A., Zhao, L., Ashworth, J., Shen, B. W., and Stoddard, B. L. (2007) *Nucleic Acids Res.* **35**, 7209–7221
40. Komori, K., Ichiyanagi, K., Morikawa, K., and Ishino, Y. (1999) *Nucleic Acids Res.* **27**, 4175–4182
41. Wang, J., Kim, H. H., Yuan, X., and Herrin, D. L. (1997) *Nucleic Acids Res.* **25**, 3767–3776
42. Komori, K., Fujita, N., Ichiyanagi, K., Shinagawa, H., Morikawa, K., and Ishino, Y. (1999) *Nucleic Acids Res.* **27**, 4167–4174
43. Saves, I., Ozanne, V., Dietrich, J., and Masson, J. M. (2000) *J. Biol. Chem.* **275**, 2335–2341
44. Gimble, F. S., and Thorner, J. (1992) *Nature* **357**, 301–306
45. Guhan, N., and Muniyappa, K. (2003) *Nucleic Acids Res.* **31**, 4184–4191
46. Richey, B., Cayley, D. S., Mossing, M. C., Kolka, C., Anderson, C. F., Farrar, T. C., and Record, M. T., Jr. (1987) *J. Biol. Chem.* **262**, 7157–7164
47. Perrin, A., Buckle, M., and Dujon, B. (1993) *EMBO J.* **12**, 2939–2947
48. Saves, I., Westrelin, F., Daffé, M., and Masson, J. M. (2001) *Nucleic Acids Res.* **29**, 4310–4318
49. Saves, I., Morlot, C., Thion, L., Rolland, J. L., Diétrich, J., and Masson, J. M. (2002) *Nucleic Acids Res.* **30**, 4158–4165
50. Hampshire, A. J., Rusling, D. A., Broughton-Head, V. J., and Fox, K. R. (2007) *Methods* **42**, 128–140
51. Papavassiliou, A. G., and Heumann, H. (1994) *Methods in Molecular Biology, DNA-Protein Interactions, Principles and Protocols* (Kneale, G. G., ed) Vol. 30, pp. 43–78, Humana Press Inc., Totowa, NJ
52. Beylot, B., and Spassky, A. (2001) *J. Biol. Chem.* **276**, 25243–25253
53. Wende, W., Grindl, W., Christ, F., Pingoud, A., and Pingoud, V. (1996) *Nucleic Acids Res.* **24**, 4123–4132
54. Hayatsu, H. (1996) *J. Biochem.* **119**, 391–395
55. Rhodes, D., and Fairall, L. (1997) in *Protein Function: A Practical Approach* (Creighton, T. E., ed) 2nd Ed., pp. 215–244, IRL Press at Oxford University Press, Oxford
56. Shaw, P. E., and Stewart, A. F. (2001) *Methods Mol. Biol.* **148**, 221–227
57. Millar, D. P. (1996) *Curr. Opin. Struct. Biol.* **6**, 322–326
58. Tripathi, P., Anuradha, S., Ghosal, G., and Muniyappa, K. (2006) *J. Mol. Biol.* **364**, 599–611
59. Lenz, T., Bonnist, E. Y., Pljevaljcic, G., Neely, R. K., Dryden, D. T., Scheidig, A. J., Jones, A. C., and Weinhold, E. (2007) *J. Am. Chem. Soc.* **129**, 6240–6248
60. Tamulaitis, G., Zaremba, M., Szczepanowski, R. H., Bochtler, M., and Siksnys, V. (2007) *Nucleic Acids Res.* **35**, 4792–4799
61. Malta, E., Verhagen, C. P., Moolenaar, G. F., Filippov, D. V., van der Marel, G. A., and Goosen, N. (2008) *DNA Repair* **7**, 1647–1658
62. Saves, I., Lanéelle, M. A., Daffé, M., and Masson, J. M. (2000) *FEBS Lett.* **480**, 221–225
63. Bakhrat, A., Jurica, M. S., Stoddard, B. L., and Raveh, D. (2004) *Genetics* **166**, 721–728
64. Bell-Pedersen, D., Quirk, S., Clyman, J., and Belfort, M. (1990) *Nucleic Acids Res.* **18**, 3763–3770
65. Bryk, M., Belisle, M., Mueller, J. E., and Belfort, M. (1995) *J. Mol. Biol.* **247**, 197–210
66. Jurica, M. S., Monnat, R. J., Jr., and Stoddard, B. L. (1998) *Mol. Cell* **2**, 469–476
67. Moure, C. M., Gimble, F. S., and Quiocho, F. A. (2002) *Nat. Struct. Biol.* **9**, 764–770
68. Moure, C. M., Gimble, F. S., and Quiocho, F. A. (2003) *J. Mol. Biol.* **334**, 685–695
69. Harshey, R. M., and Ramakrishnan, T. (1977) *J. Bacteriol.* **129**, 616–622
70. Hiriyanna, K. T., and Ramakrishnan, T. (1986) *Arch. Microbiol.* **144**, 105–109
71. Duan, X., Gimble, F. S., and Quiocho, F. A. (1997) *Cell* **89**, 555–564
72. Silva, G. H., Dalgaard, J. Z., Belfort, M., and Van Roey, P. (1999) *J. Mol. Biol.* **286**, 1123–1136
73. Ichiyanagi, K., Ishino, Y., Ariyoshi, M., Komori, K., and Morikawa, K. (2000) *J. Mol. Biol.* **300**, 889–901
74. Edgell, D. R., Derbyshire, V., Van Roey, P., LaBonne, S., Stanger, M. J., Li, Z., Boyd, T. M., Shub, D. A., and Belfort, M. (2004) *Nat. Struct. Mol. Biol.* **11**, 936–944
75. Pâques, F., and Duchateau, P. (2007) *Curr. Gene Ther.* **7**, 49–66