The PI-SceI protein is an intein-encoded homing endonuclease that initiates the mobility of its gene by making a double strand break at a single site in the yeast genome. The PI-SceI protein splicing and endonucleolytic active sites are separately located in each of two domains in the PI-SceI structure. To determine the spatial relationship between bases in the PI-SceI recognition sequence and selected PI-SceI amino acids, the PI-SceI-DNA complex was probed by photocross-linking and affinity cleavage methods. Unique solvent-accessible cysteine residues were introduced into the two PI-SceI domains at positions 91, 97, 170, 230, 376, and 378, and the mutant proteins were modified with either 4-azidophenacyl bromide or iron (S)-1-(p-bromoaetamidobenzyl)-ethylenediaminetetraacetate (FeBABE). The phenyl azide-coupled proteins cross-linked to the PI-SceI target sequence, and the FeBABE-modified proteins cleaved the DNA proximal to the derivatized amino acid. The results suggest that an extended β-hairpin loop in the endonuclease domain that contains residues 376 and 378 contacts the major groove near the PI-SceI cleavage site. Conversely, residues 91, 97, and 170 in the protein splicing domain are in close proximity to a distant region of the substrate. To interpret our results, we used a new PI-SceI structure that is ordered in regions of the protein that bind DNA. The data strongly support a model of the PI-SceI-DNA complex derived from this structure.

Homing endonucleases are a class of enzymes encoded by introns and inteins that initiate the mobility of their genetic elements to sites in recipient alleles where the elements are absent (reviewed in Ref. 1). These extremely specific enzymes cleave the recipient loci to stimulate a gene conversion process that copies the homing endonuclease coding sequence into the broken chromosome as it is repaired. The end result is that the intron or intein that encodes the endonuclease is propagated throughout the population. Enzymes in the LAGLIDADG family of homing endonucleases are characterized by the presence of one or two conserved dodecapeptide sequences. One member of this family is the PI-SceI homing endonuclease from Saccharomyces cerevisiae (2, 3), which occurs as an intein within a vacuolar H⁺-ATPase subunit and is generated by an autocatalytic protein splicing reaction (4, 5).

Like most LAGLIDADG homing endonucleases, PI-SceI recognizes an extremely long sequence (≈31 bp),1 and cleaves the DNA to generate a 4-bp, 3' overhang (3, 6, 7). Although the structure of the PI-SceI-DNA complex has not been determined, evidence from biochemical studies (6–8), from the structure of the PI-SceI apoenzyme (9), and from the structure of a related endonuclease, I-CreI, complexed to its DNA substrate (10) provide clues to general features of the interaction. The PI-SceI structure is composed of two domains that contain the protein splicing (domain I) and endonucleolytic active sites (domain II), respectively (9). The two LAGLIDADG motifs constitute two tightly packed α-helices that are each contributed by one of two subdomains related by pseudo-2-fold symmetry and form the hydrophobic core of the endonuclease domain. Two conserved acidic residues (Asp-218 and Asp-326) at the carboxyl termini of these helices may function to coordinate the Mg²⁺ co-factor(s) that is involved in the DNA cleavage mechanism (11). The group I intron-encoded I-CreI endonuclease is structurally similar to the PI-SceI endonuclease domain (12, 13). However, I-CreI differs from PI-SceI in two significant respects; it lacks a protein splicing domain, and it is a homodimeric protein composed of two molecules that are related by 2-fold symmetry (14), each of which is topologically similar to the N or C subdomains of the PI-SceI endonuclease domain. Within the I-CreI-DNA complex, the DNA rests in a saddle formed by β sheets contributed by both monomers and contacts the protein over an extended surface area (10).

We proposed a hypothetical model for the PI-SceI-DNA complex in which the endonuclease domain contacts the PI-SceI cleavage site region of the DNA, and an extended subdomain of the protein splicing domain termed the DNA recognition region (DRR) binds to a distal section of the target sequence (9). When the I-CreI-DNA complex structure was determined, we modified the model by docking the I-CreI DNA to PI-SceI (15). This and a related model (16) are consistent with mutational studies of the protein and DNA substrate (6, 17), and with interference and footprinting analyses (6, 11). Furthermore, they are consistent with cross-linking and affinity cleavage mapping studies that position His-333 of the endonuclease domain near T−3 on the bottom strand (16) and position residues 271–279 of a disordered loop near A+T−9 (15). However, detailed mapping

---

* This work was supported by National Institutes of Health Grant GM50815 (to F. S. G.) and by funds from the Institute of Biosciences and Technology (F. S. G.). The costs of publication of this article were defrayed in part by the payment of page charges. This article must therefore be hereby marked “advertisement” in accordance with 18 U.S.C. Section 1734 solely to indicate this fact.

The atomic coordinates (code 1DFA) have been deposited in the Protein Data Bank, Research Collaboratory for Structural Bioinformatics, Rutgers University, New Brunswick, NJ

‡ An Investigator of the Howard Hughes Medical Institute.

1 To whom correspondence should be addressed: Center for Genome Research, Institute of Biosciences and Technology, 2121 W. Holcombe Blvd., Houston, TX 77030. Tel.: 713-677-7605; Fax: 713-677-7689; E-mail: fgimble@ibt.tamu.edu.

The JOURNAL OF BIOLOGICAL CHEMISTRY Vol. 275, No. 4, Issue of January 28, pp. 2705–2712, 2000

Printed in U.S.A.

Probing the Structure of the PI-SceI-DNA Complex by Affinity Cleavage and Affinity Photocross-linking

(Received for publication, October 6, 1999)

Dongli Hu‡, Michael Crist‡, Xiaoqun Duan§, Florante A. Quiocho§, and Frederick S. Gimble‡]

From the ‡Center for Genome Research, Institute of Biosciences and Technology and Department of Medical Biochemistry and Genetics, The Texas A & M University System Health Science Center, Houston, Texas 77030 and the §Howard Hughes Medical Institute and Verna Marrs McLean Department of Biochemistry and Molecular Biology, Baylor College of Medicine, Houston, Texas 77030

1 The abbreviations used are: bp, base pair(s); FeBABE, iron (S)-1-(p-bromoaetamidobenzyl)-ethylenediaminetetraacetate; DRR, DNA recognition region; HINT, Hedgehog-intein.
information is not available for any other part of the PI-SceI-DNA interaction. Here, we have tethered the affinity cleavage reagent FeBABE and the photocross-linking reagent phenyl azide to residues in both PI-SceI domains that are expected to be near the protein-DNA contact surface and have employed these derivatized proteins to map these positions to specific regions of the PI-SceI target sequence. The data have been interpreted using a model of the PI-SceI complex based on a new PI-SceI crystal structure.

**EXPERIMENTAL PROCEDURES**

**Materials**—Oligonucleotides used for mutagenesis were synthesized by Sigma. Restriction and DNA-modifying enzymes were obtained from New England Biolabs, Inc. TALON metal affinity resin was purchased from CLONTECH, SP-Sepharose was obtained from Amersham Pharmacia Biotech, and Affi-Gel Blue Gel was obtained from Bio-Rad. Ascorbic acid (vitamin C, microsieve grade) was purchased from Fluka, and hydrogen peroxide (ultrapure grade) was obtained from J. T. Baker. 7-(Diethylamino)-3-(4′-maleimidphenyl)-4′-methylouamin was purchased from Molecular Probes, Inc. (Eugene, OR). 4-azidophenacyl bromide and piperidine were obtained from Sigma. FeBABE was kindly provided by Dr. Claude Meares.

**Crystalization of PI-SceI**—The procedures for overexpression and purification of selenomethionyl PI-SceI have been described previously (15). The PI-SceI crystals belong to space group P2₁. Crystals were grown using the hanging drop vapor diffusion method. The crystallization solution contained 30% polyethylene glycol 3350 and frozen by flash-cooling. The data have been described previously (15). Cys-75 is not solvent-accessible and has been described previously (15). Cys-17, Cys-249, Cys-398, and Cys-416) have been changed to serine by a 20:1 molar ratio for 20 min in the dark. The extent of modification by FeBABE or by 4-azidophenacyl bromide was determined using the fluorocrescent reagent 7-(diethylamino)-3-(4′-maleimidphenyl)-4′-methylouamin according to a published procedure (26).

**Phenyl Azide-mediated Photocross-linking**—DNA substrates that were 5′-end-labeled on either the top or bottom strands were prepared by digesting a labeled 219-bp polysaccharide chain reaction product containing a single PI-SceI recognition site with EcoRI or SceI, respectively, and by purifying the 158- and 187-bp fragments on a polyacrylamide gel (15). PI-SceI-DNA complexes were formed by incubating the modified protein derivatives (200 nM) with these labeled DNA duplexes (~0.5 nM) for 20 min in the dark in a buffer containing 25 mM HEPES (pH 8.0), 100 mM KCl, and 0.1 mM EDTA. The extent of modification by FeBABE or by 4-azidophenacyl bromide was determined using the fluorocrescent reagent 7-(diethylamino)-3-(4′-maleimidphenyl)-4′-methylouamin according to a published procedure (26).

**Modification of the Purified Proteins**—Modification of PI-SceI with 4-azidophenacyl bromide was performed using a modified version of a published procedure (24) by combining the reagent with the protein at a 20:1 molar ratio for 3 h at room temperature in the dark. To remove unreacted 4-azidophenacyl bromide, the derivatized proteins were dialyzed against 25 mM HEPES (pH 8.0), 100 mM KCl, 5 mM EDTA, and 5% glycerol. To conjugate PI-SceI to FeBABE, protein (1-5 mg/ml) was mixed with 1 mM FeBABE and allowed to incubate for 1 h at room temperature as described (25). Excess reagent was removed by dialysis. The extent of modification by FeBABE or by 4-azidophenacyl bromide was determined using the fluorescent reagent 7-(diethylamino)-3-(4′-maleimidphenyl)-4′-methylouamin according to a published procedure (26).

**Modification of the Purified Proteins**—Modification of PI-SceI with 4-azidophenacyl bromide was performed using a modified version of a published procedure (24) by combining the reagent with the protein at a 20:1 molar ratio for 3 h at room temperature in the dark. To remove unreacted 4-azidophenacyl bromide, the derivatized proteins were dialyzed against 25 mM HEPES (pH 8.0), 100 mM KCl, 5 mM EDTA, and 5% glycerol. To conjugate PI-SceI to FeBABE, protein (1-5 mg/ml) was mixed with 1 mM FeBABE and allowed to incubate for 1 h at room temperature as described (25). Excess reagent was removed by dialysis. The extent of modification by FeBABE or by 4-azidophenacyl bromide was determined using the fluorescent reagent 7-(diethylamino)-3-(4′-maleimidphenyl)-4′-methylouamin according to a published procedure (26).

**Probing the PI-SceI-DNA Complex Structure**—The DNA is derived from the two different space group crystals were compared through a cross-rotation search (9) and by rigid body refinement. After several cycles of alternating position and B-factor refinements and manual model building, water molecules were added using the graphics program O (21). The assignment of the secondary structure elements was performed using the graphics program O (21). The assignment of the secondary structure elements was performed using the graphics program O (21). The assignment of the secondary structure elements was performed using the graphics program O (21). The assignment of the secondary structure elements was performed using the graphics program O (21). The assignment of the secondary structure elements was performed using the graphics program O (21). The assignment of the secondary structure elements was performed using the graphics program O (21).
crystals (30). The new crystals belong to the C2 space group, while those grown previously belong to the P2₁ space group. There is one instead of two molecules per asymmetric unit. However, the one molecule dimerizes with one of its crystallographic symmetry-related molecules, and the resulting dimer overlaps the dimer observed in the asymmetric unit of the P2₁ crystal (30). The structure of the C2 space group crystal was determined by a molecular replacement method and refined at 2.0-Å resolution to an R cryst factor and an R free of 21.2% and 28%, respectively (Table I).

The overall PI-SceI structure determined from the C2 space group crystal is very similar to the originally reported structure (Fig. 1), and the root mean square deviation between the C-α atoms of the two structures is 0.9 Å. The same bipartite domain architecture is evident in both structures, and they adopt the same secondary structure folds. The Hedgehog-intein (HINT) module, which is a shared protein fold in the PI-SceI protein splicing domain and in the autoprocessing domain of the Drosophila hedgehog protein (31), overlaps extensively in the C2 and P2₁ structures with the exception of a loop that connects β6 to β7 (Fig. 1). Residues 135–151, which are part of the DNA recognition region (DRR), are disordered in the C2 structure, but fold in the P2₁ structure into the four-turn helix α1 and a loop that connects α1 to β12. Conversely, residues 93–102, which spatially neighbor residues 135–151 in the DRR, are disordered in the P2₁ structure but show density in the C2 structure and form a loop that connects β9 and β10. A second disordered region in the P2₁ structure occurs in the N subdomain of the endonuclease domain that includes residues 271–279, which connect β16 to α6. The disordered segment is further extended in the C2 structure to include β16 and the loop that proceeds β16 (residues 254–280). Finally, the extended β-hairpin loop (residues 369–374) between β21 and β22 that is located above the active site is disordered in the P2₁, but not the C2 structure.

*Location of Introduced Cysteine Residues in the C2 Structure*—To map interactions between PI-SceI and its recognition sequence, we conjugated FeBABE and/or phenyl azide moieties to surface residues that are in close proximity to amino acids that are critical for PI-SceI binding to DNA. The choice of residues was done in the context of the C2 structure because it is ordered in many of the DNA binding regions. Fig. 2 displays the C2 structure of PI-SceI and identifies residues Arg-91, Lys-97, His-170, Ser-230, Lys-376, and Lys-378 that were individually substituted with cysteine and then chemically modified. Substitution of residues Arg-91, Lys-97, Ser-230, Lys-376, and Lys-378 with alanine does not affect the PI-SceI-mediated DNA cleavage activity (17). Arg-91 and Lys-97 neighbor residues Arg-90 and Arg-94 in the protein splicing domain, which contribute to the binding interaction between the DRR and DNA. Lys-376 and Lys-378 are part of the same extended β21-β22 loop as Lys-369 and His-377, which participate in the interaction between the endonuclease domain and the PI-SceI cleavage site region (17). On the opposite side of the DNA-binding cleft as Lys-369 and His-377, Ser-230 is bordered by residues Asp-229 and Arg-231 in a loop between β14 and α5 (17). The exact role of Asp-229 and Arg-231 is unclear, but they are both critical for the PI-SceI reaction pathway (17). His-170 occurs in the interdomain region that is part of a ridge of positively charged residues, including Lys-53, His-170, and Lys-173, that extends along the putative DNA contact surface from the endonuclease domain toward the DRR. Unlike the other modified amino acids, His-170 may contribute to DNA binding because alanine substitution at this position reduces the cleavage rate 10-fold relative to wild-type PI-SceI (Fig. 3). However, the magnitude of this effect is lower than that caused by other PI-SceI DNA-binding mutants (17). His-170 spatially neighbors Lys-53, where substitution with alanine reduces the

| Parameter | Value |
|-----------|-------|
| Data Collection Statistics (10–2.0 Å) | |
| Multiplicity | 4 |
| Completeness | 95% |
| R cryst | 6% |
| Refinement statistics | |
| Resolution (Å) | 10–2.0 |
| Atoms | 6730 |
| Solvent molecules | 240 |
| R cryst | 21.2 |
| R free | 28.0 |
| r.m.s. of bond angle (degree) | 0.01 |
| r.m.s. of bond angles (degrees) | 0.18 |

* r.m.s., root mean square.
cleavage rate >15-fold relative to wild-type PI-SceI.²

The affinity cleavage and photocross-linking methods described in this report require that cysteines be modified in a PI-SceI protein that lacks other solvent-accessible cysteine residues. For this purpose, we utilized a recombinant derivative (PI-SceI (–5Cys)) in which five of the six naturally occurring cysteines have been changed to serine and the sole remaining cysteine is buried in the protein interior distant from the putative DNA binding surface (15). We showed previously that the DNA binding and cleavage activities of this variant are similar to wild-type PI-SceI and that the protein is not modified by FeBABE (15).

DNA Cleavage and Binding Activities of Cysteine-substituted PI-SceI Variants—Each of the cysteine-substituted PI-SceI proteins was purified using our standard expression and purification protocols (17). These proteins were tested in vitro for their DNA cleavage activities using a linearized plasmid substrate that contains a single PI-SceI recognition sequence (15). Similar levels of PI-SceI-mediated DNA cleavage activity are apparent for all of the mutant proteins except for the K378C derivative, which is reduced 2-fold in activity relative to PI-SceI (–5Cys), and the H170C variant, which is reduced 5-fold (Fig. 3). Binding activity was measured by gel mobility shift assays of a 219-bp end-labeled DNA fragment containing a single PI-SceI site. Wild-type PI-SceI forms two complexes with this substrate; a lower complex (PD³) that involves PI-SceI binding predominantly to a region of the substrate that is distant from the cleavage site and an upper complex (PD⁴) in which PI-SceI binds to this region and the cleavage site region (6, 7). The different mobilities of the complexes reflect their differing extents of DNA distortion; the lower complex is bent 40–45°, while the upper complex is bent 60–75° (6, 7). The DNA binding activities of the R91C, K97C, S230C, and K376C proteins are comparable with the PI-SceI (–5Cys) protein, while the upper complex is bent 40–45° and the lower complex is bent 60–75° (6, 7). The DNA binding activities of wild-type PI-SceI and the H170A and H170C mutant proteins. All reactions were performed at 37 °C for 10 min using enzyme and DNA concentrations described previously (17).

FIG. 3. DNA cleavage of a linearized plasmid substrate containing a single PI-SceI recognition site by wild-type PI-SceI and PI-SceI derivatives. A, DNA cleavage activity by wild-type PI-SceI, PI-SceI (–5Cys), and unconjugated (lanes marked −) and FeBABE-conjugated (lanes marked +) R91C, K97C, S230C, K376C, and K378C mutant proteins. B, DNA cleavage activity of the H170A and H170C mutant proteins. All reactions were performed at 37 °C for 10 min using enzyme and DNA concentrations described previously (17).
for reasons that are unclear, FeBABE could not be successfully conjugated in high yield to the H170C protein. Figs. 3 and 4 show that modification with FeBABE affects neither the DNA cleavage activity nor the DNA binding activities of most of the PI-SceI derivatives. Conjugation of FeBABE to the K378C protein causes no further reduction in its activity relative to the unmodified enzyme.

No significant affinity cleavage is observed on either DNA strand in the control reaction that includes PI-SceI (−5Cys) protein that is either treated with FeBABE or untreated (Fig. 5). By contrast, four of five of the FeBABE-PI-SceI conjugates, FeBABE-PI-SceI(376), FeBABE-PI-SceI(376), and FeBABE-PI-SceI(376), yield different levels of DNA cleavage at discrete sites within the substrate (Figs. 5 and 7). FeBABE-PI-SceI(376) generates a moderate signal, with the strongest cleavage occurring between 14 and 17 on the top strand and between +9 and +12 on the bottom strand. This result is consistent with genetic evidence that indicates that the neighboring residues Arg-90 and Arg-94 are involved in PI-SceI binding to this region of the substrate (17). Somewhat surprisingly, no FeBABE-mediated cleavage occurs when residue 97 is modified with the chemical nuclease.

FeBABE-PI-SceI(376), FeBABE-PI-SceI(376), and FeBABE-PI-SceI(376) each cleave the PI-SceI substrate proximal to the PI-SceI cleavage site, but important distinguishing features are evident. FeBABE-PI-SceI(376) generates the most intense cleavage pattern on both strands. The top strand is cut with high yield between positions +2 and +6 and to a lesser extent between −9 and −6. The most intense cleavage of the bottom strand occurs between nucleotides positions −5 and +1, and lower levels of cleavage are evident between −14 and −11. The cleavage patterns generated by FeBABE-PI-SceI(376) and FeBABE-PI-SceI(376) are similar to each other, with most cleavage occurring on the top strand between nucleotide positions −2 and +5 and between −8 and −2 on the bottom strand. Interestingly, the FeBABE-PI-SceI(376) and FeBABE-PI-SceI(376) cleavage patterns are offset by two base pairs on both strands.

**Photocross-linking Using Phenylazide-coupled PI-SceI Derivatives**—Phenyl azide-mediated photocross-linking was used as a complementary means of identifying interacting regions between PI-SceI and its recognition sequence. The major difference between the photocross-linking and affinity cleavage methods is that a phenyl azide photolabile cross-linking moiety rather than an FeBABE group is coupled to the solvent-accessible cysteines on PI-SceI. Protein-DNA complexes are formed using the derivatized PI-SceI variants, the protein is cross-linked to the DNA by UV irradiation, and the complexes are treated with piperidine, which effects strand scission at the cross-linked nucleotides and permits their identification by gel electrophoresis. Unlike FeBABE-mediated affinity cleavage, in which the reactive agent is diffusible, cross-linking by phenyl azide groups requires direct contact between the reactive species and a DNA nucleophile. The distance between the Cα atom and the target base has been estimated as 9–12 Å (36, 37).

PI-SceI variants with single solvent-accessible cysteine residues at positions 91, 97, 170, 230, 376, and 378 and the PI-SceI (−5Cys) protein were modified with 4-azidophenacyl bromide as described in previous reports (24, 38). Interestingly, the PI-SceI (−5Cys) protein, which contains no introduced cysteine residues, cross-links to the 219-bp substrate on both strands at nucleotide position +3 (Figs. 6 and 7). This cross-linking requires both modification of the protein with 4-azidophenacyl bromide and irradiation with UV light.³ It is unlikely that the cross-linked species results from reaction of DNA with the sole remaining cysteine in the molecule (Cys-75), because this res-

---

³ D. Hu and F. S. Gimble, unpublished results.
Probing the PI-SceI-DNA Complex Structure

DISCUSSION

The x-ray crystal structure of the PI-SceI protein has been determined, but no structure is yet available for the protein bound to its DNA substrate. Previously, we presented a model for the interaction of PI-SceI with DNA (15) using information from the recently reported structure of the related LAGLI-DAG endonuclease I-CreI complexed to its recognition sequence (10). I-CreI is structurally analogous to the PI-SceI endonuclease domain only, and as a consequence, the PI-SceI model describes how this domain contacts the DNA region surrounding the PI-SceI cleavage site. To elucidate additional features of the PI-SceI-DNA complex, we employed affinity cleavage and UV photocross-linking methods to map particular residues to regions of the PI-SceI substrate. We have interpreted these results by constructing a DNA docking model that takes advantage of the new C2 PI-SceI structure in which the modified residues are ordered. The model also depicts how the DRR might interact with DNA.

Docking of I-CreI DNA to PI-SceI was accomplished by least squares overlapping the C2 PI-SceI structure with the I-CreI-DNA complex structure. (Fig. 8). The resulting model fulfills the constraints imposed by extensive mutational (6, 11, 17), biochemical (6–8, 11, 16, 17), and structural studies (9). The DNA is positioned within a large cleft of the endonuclease domain on two symmetrically related β-sheets. Residues Lys-340 and Tyr-328, which have been identified by alanine-scanning analysis as potential DNA binding contacts (17), are positioned close to two base pairs, G/C-342 and G/C-344, that are required for PI-SceI binding (6). The two conserved acidic residues, Asp-218 and Asp-326, that are thought to ligate the essential metal ion co-factor(s) (11) are located near the two scissile phosphodiester bonds. Furthermore, the model indicates that His-335, which is located at one edge of the endonuclease domain at the interdomain boundary, is near T-26 on the bottom strand, which is consistent with the observed cross-linking between these groups (16). Finally, the model includes two regions of DNA distortion (6, 7), one near the cleavage site region that is intrinsic to the I-CreI DNA and one near the DRR. The overall amount of DNA distortion is ~55° as measured by the program CURVES (29). This amount may be lower than the experimentally determined value (60–75°; Refs. 6 and 7), because the 10° DNA bend induced by I-CreI may underrepresent the degree of bending induced by the PI-SceI endonuclease domain.

An important prediction of the model is that the extended β21-β22 hairpin loop is located in the major groove adjacent to the PI-SceI cleavage site. This loop is not perfectly positioned in the major groove in the model but instead sterically clashes with one of the DNA strands. The loop is disordered in the P21 structure, suggesting that it is extremely flexible in the absence of DNA, and we propose that its actual conformation in the complex may be different than in the C2 crystal form. That this loop is located in the major groove is supported by the observed cross-linking of azidophenacyl-PI-SceI-176 to C1 on the bottom strand. The affinity cleavage data indicate that the reactive iron group on FeBABE, whether tethered to Cys-376 or

Of the phenyl azide-conjugated protein variants, three cross-link to DNA at sites that are distinct from that of PI-SceI (~5Cy). Cross-linking occurs between azidophenacyl-PI-SceI179 and A-19 and A-22 on the top and bottom strands of the substrate, respectively (Figs. 6 and 7). Cross-links also occur to the adjacent base on each strand at A-19 and T-21. Azidophenacyl-PI-SceI179 cross-links to A-15 on the top strand and to T-16 on the bottom strand. Of the azidophenacyl-PI-SceI conjugates modified in the endonuclease domain, the only observed cross-link occurs between azidophenacyl-PI-SceI376 and C1 on the bottom strand.

Of the phenyl azide-conjugated protein variants, three cross-link to DNA at sites that are distinct from that of PI-SceI (~5Cy). Cross-linking occurs between azidophenacyl-PI-SceI179 and A-19 and A-22 on the top and bottom strands of the substrate, respectively (Figs. 6 and 7). Cross-links also occur to the adjacent base on each strand at A-19 and T-21. Azidophenacyl-PI-SceI179 cross-links to A-15 on the top strand and to T-16 on the bottom strand. Of the azidophenacyl-PI-SceI conjugates modified in the endonuclease domain, the only observed cross-link occurs between azidophenacyl-PI-SceI376 and C1 on the bottom strand.
to Cys-378, is situated over the DNA minor groove (Fig. 8). If the β21-β22 loop is within the major groove, the FeBABE moiety may extend across the minor groove where cleavage occurs. The 2-bp offset between the sites of maximum DNA cleavage by FeBABE-PI-SceI122 and FeBABE-PI-SceI1378 reflects the slightly different positions of the two residues relative to the DNA (Fig. 8). Additional support for the location of the β21-β22 loop in the major groove comes from a methylation interference analysis that demonstrates major groove interactions at G-1 and G-3 on the top strand (6).

What role might the extended loop play in the overall PI-SceI reaction pathway? A possible scenario involves initial localization of PI-SceI to the recognition sequence through base-specific interactions made by residues within the DRR, followed by clamping of the flexible β21-β22 loop onto the DNA to position the endonucleolytic active site near the ssDNA-phosphodiester bonds. Mutagenesis studies suggest that some of these contacts involve residues Lys-369 and Lys-377 and base pair G/C-1, and we speculate that the contacts made by the extended loop contribute a significant portion of the DNA binding energy. The β21-β22 loop in PI-SceI is significantly longer than the structurally analogous β-hairpin loops in two other LAGLIDADG homing endonucleases, I-CreI and I-DmoI (39). In a comparison of these enzymes, it has been commented that the size of the loops at the periphery of the endonuclease domains defines the extent of the recognition site, whereas the size of the inner loops, such as the PI-SceI β21-β22 loop or its analogues, determines the substrate specificity (39). Thus, the extended loop would be a reasonable target in efforts to re-engineer the PI- SceI specificity. On the opposite side of the DNA binding cleft from the β21-β22 loop, the model shows that Ser-230 is in close proximity to the nucleotides that are cut by FeBABE-PI-SceI122 (Fig. 8). Ser-230 falls within a short loop between β14 and α5 with three residues, Asp-229, Arg-231, and Asp-232, that participate in the reaction pathway (17).

No high resolution structure exists for any DNA that binds to a protein splicing domain, so we modeled a hypothetical DNA structure to represent the splicing-domain-DNA interaction (Fig. 8). In extending the DNA from the end of the I-CreI DNA structure, three criteria were satisfied: 1) three base pairs (A/T-16, G/C-18, and A/T-19) and two amino acids (Arg-90 and Arg-94), defined by mutagenesis as essential for binding (6, 17), were positioned in close proximity to each other; 2) the DNA was distorted in this region to conform to circular permutation studies (6, 7); and 3) protein-DNA contacts were located in the major groove near two guanines on the top strand, G-13 and G-18, where methylation at the N-7 position interferes with PI-SceI binding (6). This model resembles one that has been recently reported (16), with the important difference that it predicts a major groove interaction for the β9-β10 loop, which is invisible in the P2 crystal structure. Moreover, the model is consistent with ethylation interference and hydroxyl radical footprinting studies that indicate that residues in the splicing domain, such as Lys-53 and His-170, mediate DNA binding by making contacts to the phosphate backbone (6).

The affinity cleavage and photocross-linking data reported here are the first to map interactions between the DRR and the PI-SceI substrate. The close proximity of His-170 to positions A-15 and T-16 and of Lys-97 to positions A-20 and A-22 in the model can account for the observed cross-linking (Fig. 8). The affinity cleavage pattern of FeBABE-PI-SceI122 is shifted towards the 3′-end of each strand, which indicates that the reactive moiety on this protein is situated over the minor groove. If residue 91 were located in the major groove, the FeBABE group might be expected to extend across the adjacent minor groove where cleavage occurs. Like the β21-β22 loop of the endonuclease domain, the DRR is likely to be flexible in the absence of DNA, so we cannot rule out that there are conformational differences between PI-SceI in the complex and the protein depicted in the model. Regardless, it is satisfying that there is good agreement between the model and the mapping data.

In conclusion, PI-SceI employs a singular strategy to effect base-specific recognition of a target sequence that is significantly longer than those of many other homing endonucleases. In common with other LAGLIDADG homing endonucleases, PI-SceI makes base-specific contacts using residues within loops located in the endonuclease domain (39). However, it also establishes specific contacts nearly two helical turns distant from the PI-SceI cleavage site by using residues in the DRR of the protein splicing domain, which is absent from most homing endonucleases (40). The high affinity interaction permits the splicing domain to bind independently to the distal region of the SceI substrate (16). It has been suggested previously that PI-SceI and related enzymes formed when an ancestral homing endonuclease gene, which may have encoded an enzyme that
ressembled I-CreI, became associated with a gene that encoded a protein splicing element (9, 41). The further addition of the DRR to the splicing domain and its associated or evolved DNA binding activity may have extended the specificity of PI-SceI, thereby permitting it to locate and cleave its recognition sequence within the context of the complex yeast genome.

Acknowledgments—We are grateful to Dr. Claude Meares for generously providing the FeBABE reagent and for advice on its use. We thank Scott Mace for help and advice with the visualization software.

REFERENCES

1. Belfort, M., and Roberts, R. J. (1997) *Nucleic Acids Res.* 25, 3379–3388
2. Gimble, F. S., and Thorner, J. (1992) *Nature* 357, 301–306
3. Gimble, F. S., and Thorner, J. (1993) *J. Biol. Chem.* 268, 21844–21853
4. Hirata, R., Ohsumi, Y., Nakano, A., Kawasaki, H., Suzuki, K., and Anraku, Y. (1996) *J. Biol. Chem.* 271, 3857–3865
5. Kane, P. M., Yamashiro, C. T., Wolczyk, D. F., Neff, N., Goebl, M., and Stevens, T. H. (1990) *Science* 243, 651–657
6. Gimble, F. S., and Wang, J. (1996) *J. Mol. Biol.* 261, 165–180
7. Wende, W., Grindl, W., Christ, F., Pingoud, A., and Pingoud, A. (1998) *Nucleic Acids Res.* 26, 1587–1592
8. Grindl, W., Wende, W., Pingoud, V., and Pingoud, A. (1998) *Nucleic Acids Res.* 26, 4123–4132
9. Heath, P. J., Stephens, K. M., Monnat, R. J., Jr., and Stoddard, B. L. (1997) *Proc. Natl. Acad. Sci. U. S. A.* 94, 11235–11243
10. Otwinowski, Z., and Minor, W. (1997) *Methods Enzymol.* 276, 307–326
11. Collaborative Computing Project (1994) *Acta Crystallogr.* 50, 760–763
12. Brunger, A. T. (1992) *X-PLOR version 3.1: A System for X-ray Crystallography and NMR*, Yale University Press, New Haven, CT
13. Jones, T. A., Zou, J. Y., Cowan, S. W., and Kjeldgaard. (1991) *Acta Crystallogr.* 57, 1709–1714
14. Davies, E. R., Miyake, R., Moran, J. K., Jones, A. D., Negishi, T., Ishihama, A., and Meares, C. F. (1997) *Bioconj. Chem.* 8, 38–50
15. Danley, R., and Sklenar, H. (1988) *J. Biomol. Struct. Dyn.* 6, 63–91
16. Duan, X. (1998) *X-ray Structure of PI-SceI, a Homing Endonuclease with Protein Splicing Activity*. Ph.D. thesis, Baylor College of Medicine, Houston, TX
17. Hall, T. M. T., Porter, J. A., Young, K. E., Koonin, E. V., Beachy, P. A., and Leahy, D. J. (1997) *Cell* 91, 85–97
18. Murakami, K., Owes, J. T., Belyaeva, T. A., Meares, C. F., Busby, S. J., and Ishihama, A. (1997) *Proc. Natl. Acad. Sci. U. S. A.* 94, 11274–11278
19. Dervan, P. B. (1991) *Methods Enzymol.* 208, 497–515
20. Carter, R., and Tung, C.-S. (1998) *Comput. Appl. Biosci.* 14, 25–30
21. Kabsch, W., and Sander, C. (1983) *Biopolymers* 22, 476–476
22. Kabsch, W., and Sander, C. (1983) *J. Mol. Biol.* 163, 180–180
23. Ho, S. N., Hunt, H. D., Horton, R. M., Pullen, J. K., and Pease, L. R. (1989) *Biopolymers* 22, 357–374
24. Dumoulin, P., Oertel-Buchheit, P., Granger-Schnarr, M., and Schnarr, M. (1996) *Biopolymers* 30, 30524–30529
25. Dumoulin, P., Oertel-Buchheit, P., Granger-Schnarr, M., and Schnarr, M. (1996) *Biopolymers* 30, 30524–30529
26. Greiner, D. P., Miyake, R., Moran, J. K., Jones, A. D., Negishi, T., Ishihama, A., and Meares, C. F. (1997) *Bioconj. Chem.* 8, 44–48
27. Papavassiliou, A. G. (1994) *Methods Mol. Biol.* 36, 43–78
28. Carter, R., and Tung, C.-S. (1998) *Chem. Biol.* 5, 25–30
29. Lavery, R., and Sklenar, H. (1988) *J. Mol. Biol.* 208, 476–476
30. Dumoulin, P., Oertel-Buchheit, P., Granger-Schnarr, M., and Schnarr, M. (1996) *Biochemistry* 35, 4279–4286
31. Silva, G. H., Dalgaard, J. Z., Belfort, M., and Van Roey, P. (1999) *J. Mol. Biol.* 296, 1125–1136
32. Dalgaard, J. Z., Moser, M. J., Hughey, R., and Mian, I. S. (1997) *J. Comput. Biol.* 4, 213–219
33. Derhy, M. G., and Dervan, P. B. (1996) *Science* 276, 2484–2487
34. Dumoulin, P., Oertel-Buchheit, P., Granger-Schnarr, M., and Schnarr, M. (1996) *Biochemistry* 35, 4279–4286
35. Silva, G. H., Dalgaard, J. Z., Belfort, M., and Van Roey, P. (1999) *J. Mol. Biol.* 296, 1125–1136
36. Dalgaard, J. Z., Moser, M. J., Hughey, R., and Mian, I. S. (1997) *J. Comput. Biol.* 4, 193–214
37. Dow, A., and Dervan, P. B. (1996) *Science* 245, 110–119
38. Otwinowski, Z., and Minor, W. (1997) *Methods Enzymol.* 276, 307–326
39. Collaborative Computing Project (1994) *Acta Crystallogr.* 50, 760–763
40. Brunger, A. T. (1992) *X-PLOR version 3.1: A System for X-ray Crystallography and NMR*, Yale University Press, New Haven, CT
41. Jones, T. A., Zou, J. Y., Cowan, S. W., and Kjeldgaard. (1991) *Acta Crystallogr.* 57, 1709–1714
42. Davies, E. R., Miyake, R., Moran, J. K., Jones, A. D., Negishi, T., Ishihama, A., and Meares, C. F. (1997) *Bioconj. Chem.* 8, 38–50
43. Carter, R., and Tung, C.-S. (1998) *Comput. Appl. Biosci.* 14, 25–30
44. Lavery, R., and Sklenar, H. (1988) *J. Mol. Biol.* 208, 476–476
45. Dumoulin, P., Oertel-Buchheit, P., Granger-Schnarr, M., and Schnarr, M. (1996) *Biochemistry* 35, 4279–4286
46. Silva, G. H., Dalgaard, J. Z., Belfort, M., and Van Roey, P. (1999) *J. Mol. Biol.* 296, 1125–1136
47. Dalgaard, J. Z., Moser, M. J., Hughey, R., and Mian, I. S. (1997) *J. Comput. Biol.* 4, 193–214
48. Derhy, M. G., and Dervan, P. B. (1996) *Science* 276, 2484–2487
49. Dumoulin, P., Oertel-Buchheit, P., Granger-Schnarr, M., and Schnarr, M. (1996) *Biochemistry* 35, 4279–4286
50. Silva, G. H., Dalgaard, J. Z., Belfort, M., and Van Roey, P. (1999) *J. Mol. Biol.* 296, 1125–1136
51. Dalgaard, J. Z., Moser, M. J., Hughey, R., and Mian, I. S. (1997) *J. Comput. Biol.* 4, 193–214
52. Derhy, M. G., and Dervan, P. B. (1996) *Science* 276, 2484–2487
53. Dumoulin, P., Oertel-Buchheit, P., Granger-Schnarr, M., and Schnarr, M. (1996) *Biochemistry* 35, 4279–4286