Mapping of a DNA Binding Region of the PI-*Sce*I Homing Endonuclease by Affinity Cleavage and Alanine-Scanning Mutagenesis†

Dongli Hu,‡ Michael Crist,‡ Xiaoqun Duan,§ and Frederick S. Gimble*,‡,||

Center for Genome Research, Institute of Biosciences and Technology, and Department of Medical Biochemistry and Genetics, Texas A&M University System Health Science Center, 2121 West Holcombe Boulevard, Houston, Texas 77030, and Howard Hughes Medical Institute and Verna Marrs McLean Department of Biochemistry, Baylor College of Medicine, One Baylor Plaza, Houston, Texas 77030

Received May 24, 1999; Revised Manuscript Received July 7, 1999

ABSTRACT: The PI-*Sce*I protein is a member of the LAGLIDADG family of homing endonucleases that is generated by a protein splicing reaction. PI-*Sce*I has a bipartite domain structure, and the protein splicing and endonucleolytic reactions are catalyzed by residues in domains I and II, respectively. Structural and mutational evidence indicates that both domains mediate DNA binding. Treatment of the protein with trypsin breaks a peptide bond within a disordered region of the endonuclease domain situated between residues Val-270 and Leu-280 and interferes with the ability of this domain to bind DNA. To identify specific residues in this region that are involved in DNA binding and/or catalysis, alanine-scanning mutagenesis was used to create a set of PI-*Sce*I mutant proteins that were assayed for activity. One of these mutants, N281A, was >300-fold less active than wild-type PI-*Sce*I, and two other proteins, R277A and N284A, were completely inactive. These decreases in cleavage activity parallel similar decreases in substrate binding by the endonuclease domains of these mutant proteins. We mapped the approximate position of the disordered region to one of the ends of the 31 base pair PI-*Sce*I recognition sequence using mutant proteins that were substituted with cysteine at residues Asn-274 and Glu-283 and tethered to the chemical nuclease FeBABE. These mutational and affinity cleavage data strongly support a model of PI-*Sce*I docked to its DNA substrate that suggests that one or more residues identified here are responsible for contacting base pair A/T_{-9}, which is essential for substrate binding.

The PI-*Sce*I DNA endonuclease is a 454 residue protein of *Saccharomyces cerevisiae* that initiates the mobility of its gene to alleles that lack it by making a double-strand DNA break at a single site in the yeast genome (*1*, *2*). PI-*Sce*I is translated in the middle of a larger precursor protein and is subsequently excised autocatalytically by a protein splicing reaction (*3*, *4*). The enzyme belongs to a family of homing endonucleases found in prokaryotes, eukaryotes, archaebacteria, and viruses that are characterized by two related and conserved dodecapeptide sequences termed the LAGLIDADG motifs (for a recent review see ref *5*).

The X-ray crystal structure of PI-*Sce*I reveals a bipartite domain structure in which the two domains contain the protein splicing (domain I) and endonucleolytic (domain II) active sites, respectively (*6*). The two conserved dodecapeptide segments form two tightly packed α-helices that constitute the core of domain II. Mutational studies indicate that a pair of conserved acidic residues at the carboxyl-terminal ends of these structures comprises part of the endonucleolytic active site and may function to bind the essential Mg^{2+} ligand(s) (*7*). PI-*Sce*I interacts with an asymmetric recognition sequence that is at least 31 bp[1] in length and cleaves the DNA to generate a 4 bp 3′ extension (*8*, *9*).

The structure of the PI-*Sce*I−DNA complex has not been determined. We proposed a model for the PI-*Sce*I−DNA interaction in which the endonuclease domain contacts a region of ∼15 bp that encompasses the cleavage site while the protein splicing domain binds tightly to a ∼16 bp distal region of the substrate (*6*). In this complex, the DNA is distorted approximately 60−75° (*8*, *9*). This model is consistent with DNA interference assays (*8*), with mutational analyses of PI-*Sce*I (*10*) and of the DNA substrate (*8*), and with a study of a recombinant protein that includes only the protein splicing domain (*11*). The observation of a direct interaction between His-333, which occurs in domain II near the interdomain boundary, and position +9 on the bottom strand by photo-cross-linking has suggested that the angle of DNA curvature in the original model be slightly altered (*12*).

† This work was supported by Research Grant GM50815 to F.S.G. from the National Institutes of Health.

* To whom correspondence should be addressed. Phone: (713) 677-7605. Fax: (713) 677-7689. E-mail: fgimble@ibtmail.tamu.edu.

‡ Institute of Biosciences and Technology, Texas A&M University System Health Science Center.

§ Baylor College of Medicine.

|| Department of Medical Biochemistry and Genetics, Texas A&M University System Health Science Center.

[1] Abbreviations: bp, base pair(s); FeBABE, iron (*S*)-1-[*p*-(bromoacetamido)benzyl]ethylenediaminetetracetate; CPM, 7-(diethylamino)-3-(4′-maleimidylphenyl)-4-methylcoumarin; PCR, polymerase chain reaction; TBE, Tris−borate−EDTA; SDS, sodium dodecyl sulfate; PAGE, polyacrylamide gel electrophoresis; EMSA, electrophoretic mobility shift assay.

10.1021/bi991192h CCC: $18.00
Published on Web 09/03/1999

The goal of the present study is to examine the role in DNA binding and catalysis of a segment of PI-*Sce*I extending from residues Val-270 to Leu-280 that is situated between β16 and α6 in the PI-*Sce*I crystal structure (*6*). This region of the protein is disordered in the X-ray structure, but whether this is also true in solution is unknown. It is clear that this segment is the most susceptible region of the protein to proteolysis. Trypsin makes an initial cut within this sequence between Arg-277 and Asn-278 to generate 20 and 30 kDa fragments, and several other proteases yield similar digestion products (*13*). Although the two proteolytic fragments remain folded together, the single break totally inactivates the enzyme (*13*). We applied alanine-scanning mutagenesis to assess the importance of the individual residues in the disordered region toward DNA binding and/or cleavage activity. Second, a DNA cleavage strategy that uses the cysteine-tethered chemical nuclease FeBABE was employed to map contacts within this region to the DNA (*14*). This approach involves producing a set of mutant proteins that each have a unique solvent-accessible cysteine at a chosen location and conjugating them with FeBABE, which can generate reactive hydroxyl radicals in the presence of hydrogen peroxide (*14*). DNA sites in close proximity to the modified cysteine in protein−DNA complexes are identified by their susceptibility to hydroxyl radical-mediated cleavage. These data have allowed us to identify the DNA sites that are spatially proximal to the modified residues and to improve our model for the PI-*Sce*I−DNA interaction.

EXPERIMENTAL PROCEDURES

Materials. All oligonucleotides were synthesized by Sigma-Genosys Biotechnologies, Inc. All restriction and DNA-modifying enzymes were purchased from New England Biolabs, Inc. Talon metal affinity resin was obtained from Clontech, SP-Sepharose was obtained from Amersham Pharmacia Biotech, and Affi-Gel Blue gel was obtained from Bio-Rad Laboratories. Thermolysin was obtained from Boehringer-Mannheim. Ascorbic acid (vitamin C, microselect grade) was purchased from Fluka, and hydrogen peroxide (ultrex grade) was obtained from J. T. Baker. CPM [7-(diethylamino)-3-(4′-maleimidylphenyl)-4-methylcoumarin] was purchased from Molecular Probes. FeBABE was kindly provided by Dr. Claude Meares.

Mutagenesis of the PI-SceI Gene. Plasmid pET PI-*Sce*I C-His contains a PI-*Sce*I gene that encodes a 479 amino acid PI-*Sce*I derivative with a polyhistidine C-terminal extension (*10*). To introduce mutations into the PI-*Sce*I coding sequence, oligonucleotide primers were used in overlapping PCR amplification protocols (*15*). All mutations and inserted sequences were subsequently confirmed by dideoxy sequencing.

Expression of the PI-SceI Protein. Wild-type PI-*Sce*I and the alanine-substituted PI-*Sce*I mutant derivatives were overexpressed and purified by Co^{2+}-metal affinity and SP-Sepharose ion-exchange chromatography as described previously (*10*). The PI-*Sce*I (−5Cys), N274C, and E283C proteins used in Fe·EDTA-mediated cleavage assays were purified further by chromatography on Affi-Gel Blue gel using a linear 50−700 mM KCl gradient with the important modification that thiol-containing reducing agents (β-mercaptoethanol or dithiothreitol) were omitted from all purification steps. Protein concentrations were determined by measuring the A_{280} using an extinction coefficient of 5.03×10^4 M^{-1} cm^{-1} (*10*) or by micro-Bradford assay.

PI-SceI-Mediated DNA Cleavage. Rates of DNA cleavage were measured under single-turnover conditions where enzyme is in excess over substrate (*10*). A linear plasmid substrate (pBS-PI-*Sce*36, 7 nM) containing a single PI-*Sce*I recognition site was incubated with protein (100 nM) for various lengths of time at 37 °C in cleavage buffer [25 mM Tris-HCl (pH 8.5), 100 mM KCl, 2.5 mM β-mercaptoethanol] containing 2.5 mM $MgCl_2$ or $MnCl_2$. The reactions were terminated by the addition of stop buffer [5 mM Tris-HCl (pH 7.5), 10 mM EDTA, 0.05% (w/v) SDS, 2.5% (w/v) Ficoll], and the reaction products were separated by electrophoresis in 1 × TBE on a 0.9% agarose gel. The extent of reaction was determined by scanning densitometry of photographic negatives of the ethidium bromide-stained gels.

Assay of PI-SceI−DNA Binding. A 219 bp duplex DNA probe that contains a single PI-*Sce*I recognition site was prepared for native gel mobility shift assays by PCR and end-labeled with [^{32}P]ATP (*10*). Purified wild-type and mutant PI-*Sce*I proteins were used in gel mobility shift experiments to assay DNA binding as described previously (*10*). Protein−DNA complexes were separated from unbound DNA by electrophoresis on a 7% native polyacrylamide gel in 0.5 × TBE and were visualized by exposing the dried gels to film or by using a PhosphorImager (Molecular Dynamics).

Digestion of PI-SceI with Thermolysin. Wild-type or N284A PI-*Sce*I proteins (1.5 μg) were incubated for 5 min at room temperature with 6 ng of thermolysin at a protein: protease ratio of 250:1 (w/w). The reactions were quenched by addition of 20 mM EDTA (pH 8.0) and analyzed by SDS−PAGE separation on a 12.5% gel.

Conjugation of Protein with FeBABE. Single cysteine derivatives of PI-*Sce*I were used for conjugation with FeBABE on the basis of previous methods (*16*). The purified PI-*Sce*I (−5Cys), N274C, and E283C proteins were dialyzed overnight at 4 °C in a buffer containing 10 mM Tris-HCl (pH 8.0), 10 mM EDTA, 100 mM KCl, and 5% glycerol to remove metal ions and then dialyzed against a conjugation buffer containing 10 mM Hepes (pH 8.0), 100 mM KCl, and 1 mM EDTA. Conjugation to FeBABE was initiated in the same buffer by combining 260 μL of protein (1−2 mg/mL) with 10 μL of FeBABE (26 mM). After incubation at room temperature for 1 h, excess unreacted FeBABE was removed by dialysis overnight against conjugation buffer. To determine the conjugation efficiency, the fluorescent reagent CPM was used to estimate free cysteine side chains of both FeBABE-treated and untreated proteins (*14*).

DNA Cleavage by FeBABE-Conjugated Proteins. DNA substrates labeled on either the top or bottom strands were prepared by digesting the end-labeled 219 bp PCR product with either *Eco*O109I or *Sac*I, respectively, and by purifying the 158 and 187 bp fragments on a polyacrylamide gel. To prepare PI-*Sce*I−DNA complexes, FeBABE-conjugated or unconjugated proteins (40 nM) were combined in conjugation buffer (total reaction volume of 90 μL) with the ^{32}P-end-labeled DNA fragments (~0.4 nM) and incubated for 10 min at room temperature. DNA cleavage was initiated by the rapid sequential addition of sodium ascorbate, pH 7.0 (5 mM final concentration), and hydrogen peroxide (5 mM final

concentration) to the reaction mixture and was allowed to proceed for 2 min at room temperature. The reaction was terminated by the addition of 30 μL of quenching buffer (0.1 M thiourea, 100 μg/mL sonicated salmon sperm DNA) and 80 μL of 10 mM Tris-HCl (pH 8.0) and 1 mM EDTA. Cleaved DNA fragments were extracted with phenol, precipitated with ethanol, resuspended in formamide loading buffer, and separated by electrophoresis on a 6% denaturing gel adjacent to a Maxam−Gilbert G+A sequencing ladder (*17*). Dried gels were scanned using a PhosphorImager (Molecular Dynamics) and analyzed using Fragment Analysis (Molecular Dynamics) software.

Modeling. Docking of I-*Cre*I DNA to PI-*Sce*I was performed using the software program O (*18*) by least-squares overlapping a PI-*Sce*I structure derived from a $P2_1$ space group crystal [1vde (*6*)] with the I-*Cre*I−DNA complex structure [1BP7 (*19*)]. The transformation operator between the two structures was initially defined to overlap the two dodecapeptide helical bundles and was improved later by maximizing the number of overlapping C_α atoms. Additional bases of B-form DNA were used to extend the I-*Cre*I DNA because the length of the I-*Cre*I substrate was insufficient to represent the entire region that is cleaved by the FeBABE-conjugated PI-*Sce*I protein. The FeBABE structure was based on that of Fe-EOTUBE contained within structure 1INE (*20*).

RESULTS

Alanine-Scanning Mutagenesis of a Disordered Region of PI-SceI. To identify residues in the disordered segment that are involved in DNA binding, alanine substitutions were introduced individually for residues Lys-269, Arg-272, Asn-274, Arg-277, Asn-278, Asn-279, Asn-281, Thr-282, Glu-283, and Asn-284 by PCR mutagenesis (*10*). Amino acid residues with hydrophilic side chains were selected for substitution because these often function as hydrogen bond donors or acceptors in protein−DNA complexes. All of the mutant proteins could be expressed and purified like wild-type PI-*Sce*I except for the R272A protein, which could not be purified in suitable amounts for analysis and was not characterized further (data not shown).

DNA Binding and Cleavage Activities of Alanine-Substituted PI-SceI Variants. The DNA cleavage activity was assayed for each of the mutant proteins using a linearized substrate that contains a single PI-*Sce*I recognition site. Four of the mutant proteins, K269A, N274A, N278A, and E283A, exhibited cleavage activities in the presence of Mg^{2+} ion that were reduced 6-fold or less compared to wild-type PI-*Sce*I (Table 1). By contrast, the DNA cleavage activity was reduced greater than 300-fold for the N281A protein, and no cleavage activity at all was observed for the R277A and N284A proteins after 4 h. Intermediate reductions in cleavage activity (11-fold and 38-fold reductions) were measured for the N279A and T282A proteins. Substitution of Mn^{2+} for Mg^{2+} in the PI-*Sce*I cleavage assay buffer increases the activity of wild-type PI-*Sce*I approximately 10-fold and effects a similar rate increase for several other defective protein and substrate mutants (*7*−*10*). For the mutants studied here, we observed rate increases between 5- and 25-fold when Mn^{2+} is included in the reaction buffer (Table 1). In fact, Mn^{2+} substitution restores a low level of activity to the R277A mutant, which is inactive in the presence of Mg^{2+}, but no activity is restored to the N284A protein, which is catalytically inactive under all conditions tested (Table 1).

Table 1: Cleavage Activity of Wild-Type and Mutant PI-*Sce*I Proteins[a]

	Mg^{2+}		Mn^{2+}	
	activity $\times 10^{-3}$ [b]	relative activity[c]	activity $\times 10^{-3}$	relative activity
wild type[d]	3.9 ± 0.4	1	33 ± 3	1
K269A	0.70 ± 0.003	0.2	8.9 ± 0.5	0.3
N274A	1.69 ± 0.03	0.4	11 ± 1	0.3
R277A	*e*	*e*	0.16 ± 0.01	0.01
N278A	0.74 ± 0.02	0.2	3.2 ± 0.1	0.10
N279A	0.34 ± 0.01	0.09	2.7 ± 0.2	0.08
N281A	0.012 ± 0.001	0.003	0.30 ± 0.01	0.01
T282A	0.10 ± 0.01	0.03	1.4 ± 0.10	0.04
E283A	1.5 ± 0.1	0.4	9.3 ± 0.6	0.3
N284A	*e*	*e*	*e*	*e*

[a] Values represent means ± standard deviations from at least three experiments. [b] Units are M DNA cleaved/(min × M enzyme). [c] Relative activity is the ratio of the mutant to wild-type activities. [d] Data taken from He et al. (*10*). [e] No detectable activity after 4 h of incubation.

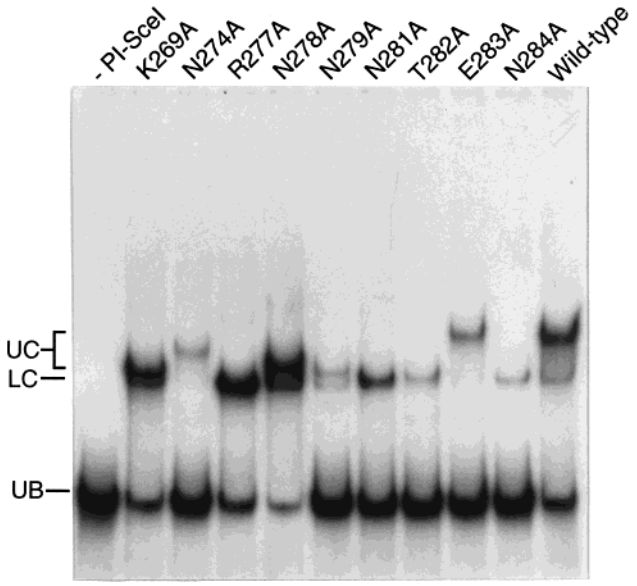

FIGURE 1: DNA binding activity of wild-type and alanine-substituted PI-*Sce*I mutant proteins assayed by gel shift analysis of a 219 bp fragment containing a single PI-*Sce*I recognition sequence. Above each lane is indicated the identity of the protein present in the reaction. The concentration of wild-type or mutant protein in the binding reaction is 0.7 nM. The position of unbound labeled DNA is indicated by UB, and the positions of the lower and upper complexes are indicated by LC and UC, respectively.

PI-*Sce*I normally forms two stable complexes with its recognition sequence in EMSA experiments: an upper complex in which the DNA is bent approximately 60−75° due to the binding of both domains and a lower complex that involves DNA contact predominantly by the splicing domain. As a whole, there is a good correlation in this study and in a previous report of other PI-*Sce*I DNA-binding mutants (*10*) between the migration distance of the upper complex and the level of cleavage activity (Figure 1). Mutant proteins with near wild-type levels of activity, such as E283A, yield complexes that comigrate with the wild-type upper complex. For the four proteins that are least active, R277A, N281A, T282A, and N284A, only a single band is apparent that comigrates with the wild-type lower complex.

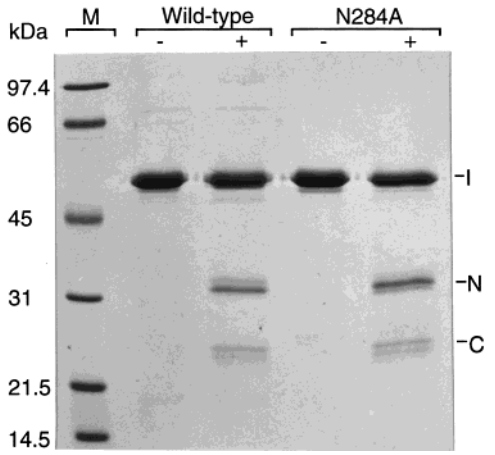

FIGURE 2: Digestion of wild-type PI-*Sce*I and the N284A mutant protein with thermolysin. Lanes show equal amounts of purified wild-type or N284A protein that were digested (lanes labeled +) or not digested (lanes labeled −) with thermolysin. The positions of the intact protein (labeled I) and the N-terminal (labeled N) and C-terminal (labeled C) digestion products are indicated. Lane M shows protein molecular weight markers.

The mutant proteins with intermediate levels of activity yield upper complexes that migrate between the positions of the wild-type upper and lower complexes. Presumably, the mobility of the upper complexes reflects the amount of DNA distortion near the cleavage site, which is a prerequisite for substrate cleavage (*8*, *9*). Therefore, when no upper complex is apparent, the DNA is not distorted in this region.

Digestion of Wild-Type and N284A Proteins with Thermolysin. The ability of the catalytically inactive N284A protein to form a lower protein−DNA complex indicates that the protein splicing domain is likely to be properly folded, but nothing can be concluded about the endonuclease domain because the protein is catalytically inactive under all conditions (Table 1). To determine whether the structures of the wild-type and N284A proteins proximal to the disordered segment are similar, we compared the proteolytic stabilities of the two proteins by digestion with thermolysin. This protease cuts PI-*Sce*I initially to yield 20 and 30 kDa fragments that are similar to the trypsin digestion products (data not shown). When the N284A and wild-type PI-*Sce*I proteins are digested for the same time period with identical amounts of enzyme, the patterns of digestion are very similar (Figure 2). This suggests that the there are no major differences in the folded state of the two proteins near the cut site, although it should be noted that more subtle conformational differences may not be detected by this method.

Binding and PI-SceI Cleavage Activities of Unmodified and Modified N274C and E283C Proteins. PCR mutagenesis was used to engineer a recombinant gene that expresses PI-*Sce*I (−5Cys), a PI-*Sce*I variant in which five of the six naturally occurring cysteines, Cys-1, Cys-17, Cys-249, Cys-398, and Cys-416, are changed to serine. The remaining cysteine, Cys-75, was not changed to serine because it is totally inaccessible to solvent according to the crystal structure and unlikely to be modified by FeBABE (*6*). The overexpressed and purified PI-*Sce*I (−5Cys) protein is slightly more active than wild-type PI-*Sce*I in DNA cleavage assays [61% vs 50% total cleavage after 10 min (Figure 3)]. Moreover, the PI-*Sce*I (−5Cys) protein displays at least 95%

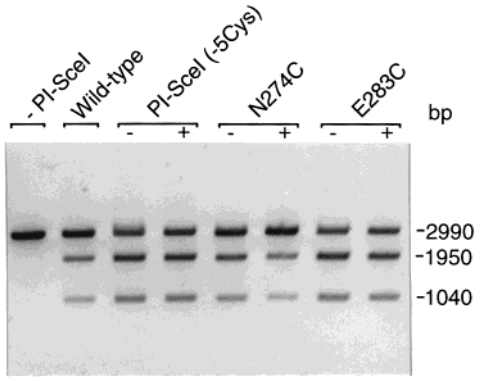

FIGURE 3: PI-*Sce*I endonuclease cleavage of a linearized DNA substrate containing a single PI-*Sce*I recognition site by wild-type PI-*Sce*I, PI-*Sce*I (−5Cys), and FeBABE-conjugated N274C and E283C proteins. Positions of the introduced Cys residues are indicated above the figure. Reactions containing FeBABE-modified PI-*Sce*I proteins are labeled +, and those containing unmodified protein are labeled −. The positions of the 2990 bp linearized DNA substrate and of the 1950 bp and 1040 reaction products generated by PI-*Sce*I are indicated.

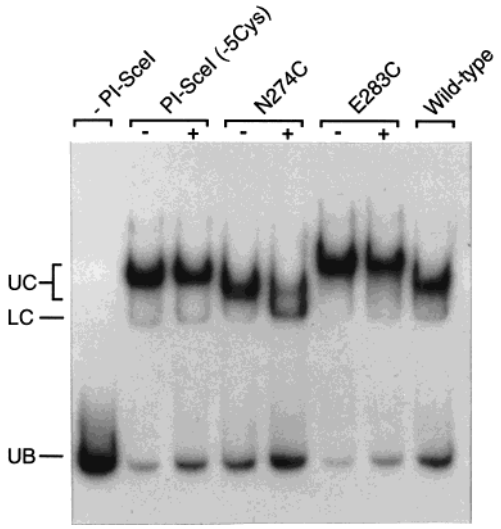

FIGURE 4: DNA binding activity of wild-type, PI-*Sce*I (−5Cys), and FeBABE-conjugated N274C and E283C proteins. The identity of the protein used in each reaction is indicated above each lane. Lanes labeled + include protein that has been conjugated to FeBABE, and those labeled − include unconjugated proteins. The migration positions of the unbound 219 bp fragment DNA and the lower and the upper PI-*Sce*I−DNA complexes are indicated by UB, LC, and UC, respectively.

of the DNA binding activity as the wild-type PI-*Sce*I, and the partitioning of bound DNA between the lower and upper complexes is similar for both proteins (Figure 4). The PI-*Sce*I (−5Cys) upper complex migrates slightly slower than that of wild-type PI-*Sce*I.

Single Cys residues were introduced into the PI-*Sce*I (−5Cys) protein by PCR mutagenesis at positions Asn-274, which is within the disordered segment, and Glu-283, which is an ordered residue. These positions were chosen for modification with FeBABE because they are not critical for protein−DNA complex formation (Figure 1) but are likely to be close to the DNA given their proximity to residues Arg-277 and Asn-284. The unmodified N274C and E283C proteins display the same cleavage activity as PI-*Sce*I (−5Cys) (Figure 3). DNA binding of the PI-*Sce*I (−5Cys), N274C, and E283C proteins is similar; at a protein concen-

tration of 7×10^{-10} M, approximately 90%, 83%, and 95% of the DNA is bound, respectively (Figure 4). The upper complex of the N274C protein migrates slightly faster than that of wild-type PI-*Sce*I while that of the E283C protein migrates slower. These results, taken together with the fact that the mutant proteins purify like PI-*Sce*I (−5Cys), suggest that the cysteine substitutions do not extensively modify the overall structure or activity of the protein.

The FeBABE conjugation yields were determined to be 50% for the N274C protein and 42% for the E283C protein. The conjugated and E283C proteins both retain DNA cleavage activity although activity was reduced approximately 14% for the modified N274C protein relative to the unmodified protein. This decrease may be due to a partial reduction in DNA binding of the N274C protein. Total binding of the N274C protein was reduced 16% compared to the unmodified protein. More importantly, the partitioning of the bound DNA between the upper and lower complexes for the FeBABE-conjugated N274C protein is approximately equal whereas it is >8:1 for the unmodified protein. These results indicate a partial DNA binding defect of the endonuclease domain as a result of the FeBABE modification of the N274C protein.

Fe·EDTA-Mediated DNA Cleavage by FeBABE−PI-SceI Conjugates. As expected, the PI-*Sce*I (−5Cys) protein fails to generate any appreciable hydroxyl radical-mediated DNA cleavage whether it has been treated with FeBABE or not (Figure 5A). Similarly, the mutant proteins do not cut the DNA if they are not modified with FeBABE. By contrast, significant DNA cleavage occurs when FeBABE is attached to 274C or to 283C. The general locations where DNA cleavage occurs are similar, although important differences are evident in the details of the cleavage patterns (Figure 5A). The region of strongest cleavage for both proteins overlaps one of the borders of the 31 bp PI-*Sce*I recognition sequence (Figure 5B). The overall extent of cleavage generated by the modified 283C protein is significantly higher than by the 274C protein. Strongest cleavage by FeBABE tethered to 283C is evident on the bottom strand between positions −15 and −10 with two other equally intense peaks of weaker cleavage occurring approximately one helical turn distant in both directions between positions −24 and −20 and between positions −2 and −5 (Figure 5). The pattern produced by the modified 274C protein on the bottom strand is slightly different in that most cleavage occurs between positions −15 and −10 and between positions −2 and −5, followed by a lower level of cleavage between positions −24 and −20.

On the top strand, strongest cleavage is apparent for the modified 283C protein between positions −7 and −12 with slightly lower levels evident between positions −16 and −22 (Figure 5). Weak cleavage activity is also evident between positions −1 and +2. The strongest cleavage activity generated by the 274C protein occurs between positions −7 and −11 with two other peaks of weaker cleavage occurring one helical turn distant between positions −1 and +3 and between positions −21 and −15.

DISCUSSION

The determination of an X-ray structure for PI-*Sce*I allowed us to propose an initial model for the interaction of

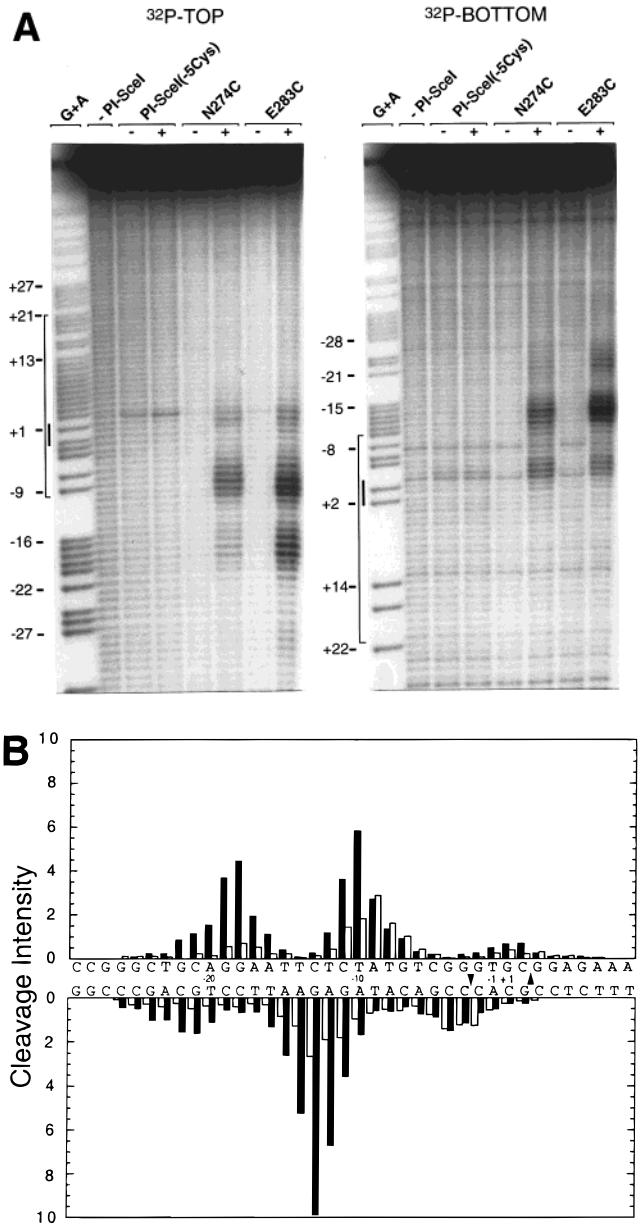

FIGURE 5: DNA cleavage of the top and bottom strands of a 219 bp fragment containing a single PI-*Sce*I recognition site by FeBABE-conjugated PI-*Sce*I proteins. (A) At the top of each set of lanes is indicated which DNA strand is end-labeled with ^{32}P. The identities of the proteins used in the reactions are indicated above each lane. Cleavage reactions containing FeBABE-conjugated PI-*Sce*I proteins are labeled +, and those that include unconjugated protein are labeled −. To generate hydroxyl radicals in the reaction, all of the samples were treated identically with ascorbate and hydrogen peroxide. The lanes marked A+G show Maxam−Gilbert A+G reactions. The base pairs are numbered from −1 and +1 extending from the center of the 4 bp cleavage site, which is indicated by a solid black line. The 31 bp PI-*Sce*I recognition sequence defined previously (−10 to +21) is indicated by a bracket (*8*). These denaturing polyacrylamide gels were visualized by film autoradiography. (B) Summary of FeBABE-mediated affinity cleavage results. Peak heights reflect the extent of DNA cleavage by FeBABE at each individual base as obtained using a PhosphorImager and Fragment Analysis software. These values were calculated at each base position by subtracting the integrated area value for the unmodified protein from the value for the FeBABE-conjugated protein. The open and closed columns represent the results using the N274C-FeBABE and E283C-FeBABE proteins, respectively. The positions of the PI-*Sce*I-mediated DNA cleavage sites in the recognition sequence are indicated by arrows.

the protein with its substrate DNA that fulfilled several experimentally based criteria (*6*). The two scissile phosphodiester bonds were placed in close proximity to two conserved acidic residues, Asp-218 and Asp-326, that are proposed to bind the essential metal ion cofactor(s) at the active site (*7*). The DNA was docked within a large cleft of the protein in domain II such that the cleavage site was centered onto a platform of symmetrically related β-sheets (sheets 7 and 9). The ~15 bp DNA segment that is thought to contact domain II is asymmetrically centered about the cleavage site. Regions of positive charge situated between the two PI-*Sce*I domains and in an extended structure of domain I were placed close to the DNA. The absence of this extended structure from other splicing proteins suggests that it and the endonuclease domain were added onto a core splicing domain when endonucleolytic activity was acquired (*21*). It was also necessary to bend the DNA in the original model to reflect the distortion that occurs when PI-*Sce*I contacts its substrate (*8, 9*). Recently, the demonstration by photo-cross-linking experiments of a direct interaction between His-333 and T_{+9} on the bottom strand suggested that the direction of the DNA curvature be altered from a concave to a convex shape (*12*).

The model is consistent with mutational analyses of both the DNA substrate and the protein and suggests, in general terms, possible interactions between groups of residues and base pairs. Residues Lys-340 and Tyr-328 within domain II, which have side chains that extend into the cleft from the β-sheet platform, mediate DNA binding of this domain and may contact base pairs G/C_{+3} and G/C_{+4} in the major groove. Tyr-328 may be responsible for the observed photo-cross-linking of PI-*Sce*I to G_{+4} and A_{+5} (*12*). Lys-369 and His-377, which are located in an extended loop, and the neighboring His-343 are other residues that contribute to the DNA binding of domain II, perhaps by making contacts to G/C_{-3}, T/A_{-1}, and G/C_{+1} in the major groove. In the protein splicing domain, mutation of any of the three base pairs, A/T_{+16}, G/C_{+18}, and A/T_{+19}, significantly reduces PI-*Sce*I DNA binding by affecting the ability of domain I to bind. It is proposed that one or more of these positions contact residues Arg-90 and Arg-94 or nearby residues that are situated within the extended region of the splicing domain (*10*).

Of the nine base pairs of the 31 bp recognition sequence that are critical for PI-*Sce*I binding and/or catalysis (*8*), the only one for which no amino acid has been implicated in binding is A/T_{-9}. Here, we provide the first evidence that several residues within a disordered segment of PI-*Sce*I, including Asn-281, Arg-277, and Asn-284, are candidates for this interaction. Substitution of each of these residues with alanine results in a severe decrease in DNA cleavage activity. The most plausible explanation for this result is that these mutations interfere with the ability of the endonuclease domains of these proteins to contact the DNA substrate. Indeed, this idea is consistent with our finding that these mutant proteins only form the lower PI-*Sce*I−DNA complex in EMSA experiments. In the absence of any structural information about these proteins, it should be noted that it is unclear whether substitution of these key residues removes direct contacts to DNA or induces subtle conformational changes in the protein that influence other key amino acids that are required for DNA binding. Previously, it was shown that proteolytic digestion of PI-*Sce*I with trypsin between Arg-277 and Asn-278 inactivates the protein (*13*). Cleavage of the peptide bond at this position may alter the conformation of the protein in this region and prevent one or more of the residues that we identified from participating in DNA binding. The observation that the proteolyzed PI-*Sce*I protein generates no upper DNA complex supports this contention (*13*).

Mapping studies using the chemical nuclease FeBABE provide additional support for contacts between the disordered residues and A/T_{-9}. This tool for mapping protein−nucleic acid interactions has been effectively used to probe the architecture of the open complex formed by *E. coli* RNA polymerase and promoter DNA (*22, 23*). The most important conclusion that can be drawn from our mapping study is that residues Asn-274 and Glu-283 are in close proximity to A/T_{-9} (Figure 5). The location of the Fe moiety on FeBABE relative to the major or minor grooves can be inferred from the cleavage pattern since FeBABE cleaves DNA at the C1′ and C4′ residues of the sugar in the minor groove (*24*). Consequently, the cleavage pattern that results is shifted toward the 3′ end if the Fe is in the minor groove and toward the 5′ end if it is located in the major groove (*24*). When FeBABE is tethered to E283C, we observe that a combination of the two patterns results, which suggests that the Fe moiety is positioned an intermediate distance between the major and minor grooves. Conversely, the cleavage pattern of the N274C-conjugated protein is only shifted toward the 3′ end of each strand, which indicates that this modified residue is closer to the minor groove. Furthermore, it is also evident that chemical cleavage by the N274C-modified protein occurs closer to the PI-*Sce*I cleavage site than for the E283C-conjugated protein.

To interpret the FeBABE cleavage patterns and to determine whether the amino acids identified here could be responsible for contacting A/T_{-9}, we generated a PI-*Sce*I−DNA model that is based on the recently reported cocrystal structure of the I-*Cre*I protein (*19*). I-*Cre*I is an LAGLIDADG endonuclease, but unlike PI-*Sce*I, it lacks a protein splicing domain and contains only a single LAGLIDADG sequence. In the I-*Cre*I structure, two I-*Cre*I subunits are related by a crystallographic 2-fold rotation axis to form a homodimeric enzyme. The PI-*Sce*I endonuclease domain folds into two symmetry-related subdomains that are structurally similar to the I-*Cre*I homodimer (*25, 26*). The I-*Cre*I complex structure reveals that the binding surface for the 24 bp DNA homing site is composed of a saddle-shaped structure formed by an extended β-ribbon within each of the two symmetry-related subunits (*19*). Binding of DNA to I-*Cre*I induces a shallow bend of less than 10° in the middle of each half-site near the fifth base pair. Given the related topologies of PI-*Sce*I and I-*Cre*I, it is reasonable to expect that they interact with DNA similarly.

Docking of I-*Cre*I−DNA to the PI-*Sce*I crystal structure was accomplished by superimposing the secondary structures of PI-*Sce*I domain II and of I-*Cre* within the protein−DNA cocrystal (Figure 6). Overlap of 104 pairs of atoms results in an rms deviation of 1.9 Å. The direction of the I-*Cre*I−DNA docked to PI-*Sce*I in this model and in a recently proposed model based only on the I-*Cre*I apoprotein structure is similar (*12*). It should be noted, however, that the actual complex structure is likely to differ from the models because

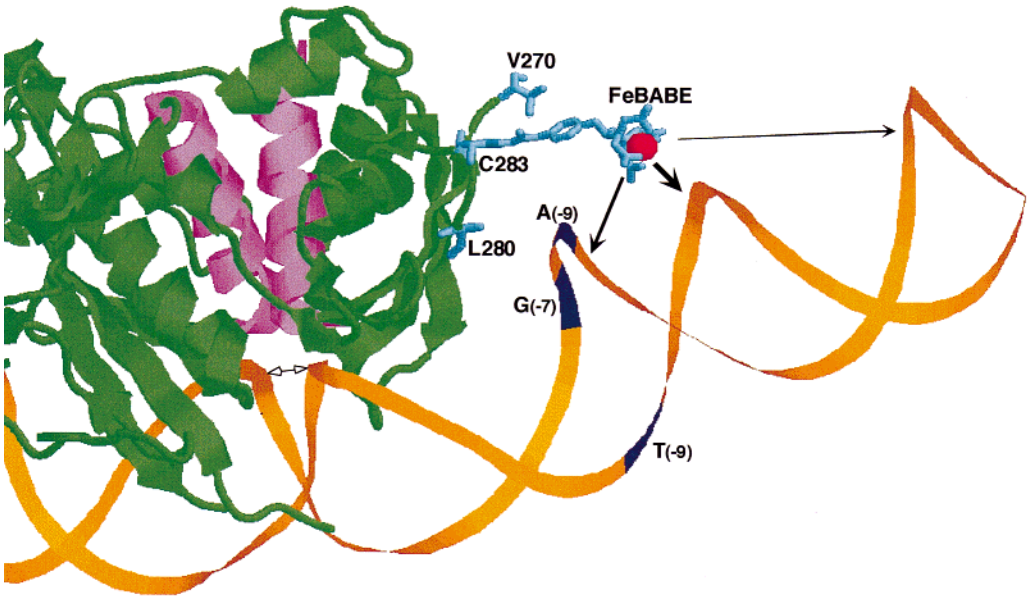

FIGURE 6: Proposed model of PI-*Sce*I bound to DNA. The PI-*Sce*I crystal structure (*6*) was docked to I-*Cre*I DNA (*19*) as described in Experimental Procedures. The PI-*Sce*I endonuclease domain is shown with FeBABE tethered to residue Glu-283 substituted with cysteine. Hydroxyl radical attack by FeBABE at T_{-10} and G_{-18} on the top strand and G_{-13} on the bottom strand is represented by closed arrows. The residues of the disordered region are not represented, but the border residues Val-270 and Leu-280 are indicated. The location of the two scissile bonds is indicated by an open double arrow positioned below the two LAGLIDADG α-helices (purple). Also indicated are the positions of the essential base pair A/T_{-9} and the location of a guanine (G_{-7}) where methylation at the N7 position in the major groove prevents PI-*Sce*I binding. This figure was rendered with RasMol.

there are some steric clashes between loop regions of the protein and the DNA near the cleavage site, and the extent of bending of PI-*Sce*I is greater than that of the I-*Cre*I−DNA (*8*, *9*, *19*). Nevertheless, the model shows that residues within and near the disordered segment may contact A/T_{-9}. The FeBABE mapping results are fully consistent with the model and confirm the polarity of the protein−DNA interaction suggested previously (*6*, *12*). Figure 6 shows a possible position of the FeBABE moiety tethered to cysteine-substituted Glu-283, which is ordered in the PI-*Sce*I structure. Our positioning of the group between the major and minor grooves is consistent with the relative cleavage intensities that are observed on the two DNA strands. Given that the distance of the Fe group from the peak DNA cleavage positions is ∼12 Å and that the hydroxyl radicals travel an additional ∼3−4 Å before being quenched by solvent *(24)*, neither the distance between the modified cysteine and the affinity cleavage sites on the DNA nor the degree of DNA distortion can be accurately determined from these data. FeBABE tethered to N274C was not modeled because the position of this residue is unknown. In sum, our mutational and affinity cleavage data provide strong support for a DNA-binding function for residues within or near the disordered segment.

Are the regions in other LAGLIDADG endonucleases that are structurally analogous to the PI-*Sce*I disordered segment also involved in DNA binding? To address this question, we compared the PI-*Sce*I X-ray structure with that of I-*Cre*I and with the recently determined structure of another LAGLIDADG homing endonuclease, I-*Dmo*I. The overall fold of I-*Dmo*I is topologically similar to the I-*Cre*I homodimer and to the endonuclease domain of PI-*Sce*I (*27*). It consists of two connected α/β-domains that are related by pseudo-2-fold symmetry and are analogous to the two subdomains of PI-*Sce*I in domain II. Structural alignment of I-*Dmo*I and PI-*Sce*I indicates that the PI-*Sce*I amino-terminal subdomain is most similar to the I-*Dmo*I carboxyl-terminal domain (*27*). Therefore, we compared the disordered segment of PI-*Sce*I, which occurs between β16 and α6 in the amino-terminal subdomain, with the structurally analogous region of I-*Dmo*I in the carboxyl-terminal domain, which occurs between β8 and α6. Indeed, residues Ser-166 and Arg-168 in β8 and Lys-172 and Phe-173 in α6 of I-*Dmo*I have been implicated in contacting DNA by protein footprinting experiments (*28*). By contrast, the residues in the analogous region of I-*Cre*I, which are situated between β4 and α3, are not involved in DNA binding (*19*). Thus, it appears that the connecting region that occurs between a structurally conserved β-strand and α-helix is utilized in some of the LAGLIDADG endonucleases to contact DNA.

There is extensive evidence that structure becomes ordered for many proteins upon DNA binding (*29*, *30*). Indeed, in I-*Cre*I, the loop that connects β1 and β2 undergoes a significant conformational change upon DNA binding, and the C-terminal 16 residues, which are disordered in the apoenzyme structure, become ordered in the DNA complex structure and are observed to make nonspecific contacts to the DNA phosphate backbone (*19*). Whether DNA binding induces structural order between residues Val-270 and Leu-280 of PI-*Sce*I is still uncertain. An argument in favor of an ordered structure for this region is that the defined hydroxyl radical cleavage pattern that results when the FeBABE moiety is tethered to 273C is consistent with a unique spatial position for this residue. Interestingly, the analogous regions in I-*Cre*I and I-*Dmo*I are ordered even in the absence of DNA (*27*, *31*). Detailed knowledge about the status of this region of PI-*Sce*I must await a high-resolution structure of an PI-*Sce*I−DNA complex by crystallographic or NMR methods.

ACKNOWLEDGMENT

We are grateful to Dr. Claude Meares for generously providing us with the FeBABE reagent and for advice on its use. We thank Zening He for constructing some of the mutant genes used in this study. We also thank Dr. F. A. Quiocho for many helpful discussions and for his continuing interest in this project.

REFERENCES

1. Gimble, F. S., and Thorner, J. (1992) *Nature 357*, 301−306.
2. Bremer, M. C. D., Gimble, F. S., Thorner, J., and Smith, C. L. (1992) *Nucleic Acids Res. 20*, 5484.
3. Kane, P. M., Yamashiro, C. T., Wolczyk, D. F., Neff, N., Goebl, M., and Stevens, T. H. (1990) *Science 250*, 651−657.
4. Hirata, R., Ohsumi, Y., Nakano, A., Kawasaki, H., Suzuki, K., and Anraku, Y. (1990) *J. Biol. Chem. 265*, 6726−6733.
5. Belfort, M., and Roberts, R. J. (1997) *Nucleic Acids Res. 25*, 3379−3388.
6. Duan, X., Gimble, F. S., and Quiocho, F. A. (1997) *Cell 89*, 555−564.
7. Gimble, F. S., and Stephens, B. W. (1995) *J. Biol. Chem. 270*, 5849−5856.
8. Gimble, F. S., and Wang, J. (1996) *J. Mol. Biol. 263*, 163−180.
9. Wende, W., Grindl, W., Christ, F., Pingoud, A., and Pingoud, V. (1996) *Nucleic Acids Res. 24*, 4123−4132.
10. He, Z., Crist, M., Yen, H.-C., Duan, X., Quiocho, F. A., and Gimble, F. S. (1998) *J. Biol. Chem. 273*, 4607−4615.
11. Grindl, W., Wende, W., Pingoud, V., and Pingoud, A. (1998) *Nucleic Acids Res. 26*, 1857−1862.
12. Pingoud, V., Thole, H., Christ, F., Grindl, W., Wende, W., and Pingoud, A. (1999) *J. Biol. Chem. 274*, 10235−10243.
13. Pingoud, V., Grindl, W., Wende, W., Thole, H., and Pingoud, A. (1998) *Biochemistry 37*, 8233−8243.
14. Greiner, D. P., Miyake, R., Moran, J. K., Jones, A. D., Negishi, T., Ishihama, A., and Meares, C. F. (1997) *Bioconjugate Chem. 8*, 44−48.
15. Ho, S. N., Hunt, H. D., Horton, R. M., Pullen, J. K., and Pease, L. R. (1989) *Gene (Amsterdam) 77*, 51−59.
16. Murakami, K., Kimura, M., Owens, J. T., Meares, C. F., and Ishihama, A. (1997) *Proc. Natl. Acad. Sci. U.S.A. 94*, 1709−1714.
17. Papavassiliou, A. G. (1994) in *Methods in Molecular Biology* (Kneale, G. G., Ed.) pp 43−78, Humana Press, Inc., Totowa, NJ.
18. Jones, T. A., Zou, J. Y., Cowan, S. W., and Kjeldgaard, M. (1991) *Acta Crystallogr. A 47*, 110−119.
19. Jurica, M. S., Monnat, R. J., Jr., and Stoddard, B. L. (1998) *Mol. Cell 2*, 469−476.
20. Love, R. A., Villafranca, J. E., Aust, R. M., Nakamura, K. K., Jue, R. A., Major, J. G., Jr., Radhakrishnan, R., and Butler, W. F. (1993) *Biochemistry 32*, 10950−10959.
21. 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.
22. Owens, J. T., Chmura, A. J., Murakami, K., Fujita, N., Ishihama, A., and Meares, C. F. (1998) *Biochemistry 37*, 7670−7675.
23. Bown, J. A., Owens, J. T., Meares, C. F., Fujita, N., Ishihama, A., Busby, S. J., and Minchin, S. D. (1999) *J. Biol. Chem. 274*, 2263−2270.
24. Oakley, M. G., and Dervan, P. B. (1990) *Science 248*, 847−850.
25. Aggarwal, A. K., and Wah, D. A. (1998) *Curr. Opin. Struct. Biol. 8*, 19−25.
26. Gimble, F. S., Hu, D., Duan, X., and Quiocho, F. A. (1998) *J. Biol. Chem. 273*, 30524−30529.
27. Silva, G. H., Dalgaard, J. Z., Belfort, M., and Van Roey, P. (1999) *J. Mol. Biol. 286*, 1123−1136.
28. Lykke-Andersen, J., Garrett, R. A., and Kjems, J. (1996) *Nucleic Acids Res. 24*, 3982−3989.
29. Spolar, R. S., and Record, M. T., Jr. (1994) *Science 263*, 777−784.
30. Reedstrom, R. J., and Royer, C. A. (1995) *J. Mol. Biol. 253*, 266−276.
31. Heath, P. J., Stephens, K. M., Monnat, R. J., Jr., and Stoddard, B. L. (1997) *Nat. Struct. Biol. 4*, 468−476.

BI991192H