Amino Acid Residues in Both the Protein Splicing and Endonuclease Domains of the PI-SceI Intein Mediate DNA Binding*

(Received for publication, October 27, 1997, and in revised form, December 9, 1997)

Zening He‡, Michael Crist‡, Hsiao-ching Yen‡, Xiaoqun Duan§§, Florante A. Quiocho§§, and Frederick S. Gimble‡‡‡‡

From the ‡Center for Macromolecular Design, Institute of Biosciences and Technology, and the §§Department of Biochemistry and Biophysics, Texas A & M University, Houston, Texas 77030 and the $Structural and Computational Biology and Molecular Biophysics Program, Howard Hughes Medical Institute and Department of Biochemistry, Baylor College of Medicine, Houston, Texas 77030

A structure-based model describing the interaction of the two-domain PI-SceI endonuclease with its 31-base pair DNA substrate suggests that the endonuclease domain (domain II) contacts the cleavage site region of the substrate, while the protein splicing domain (domain I) interacts with a distal region that is sufficient for high affinity binding. To support this model, alanine-scanning mutagenesis was used to assemble a set of 49 PI-SceI mutant proteins that were purified and assayed for their DNA binding and cleavage properties. Fourteen mutant proteins were 4- to >500-fold less active than wild-type PI-SceI in cleavage assays, and one mutant (T225A) was 3-fold more active. Alanine substitution at two positions in domain I reduces overall binding >60-fold by perturbing the interaction of PI-SceI with the minimal binding region. Conversely, mutations in domain II have little effect on binding, reduce binding to the cleavage site region only, or affect binding to both regions. Interestingly, substitutions at Lys301, which is part of the endonucleolytic active site, eliminate binding to the cleavage site region but permit contact with the minimal binding region. This experimental evidence demonstrates that the protein splicing domain as well as the endonuclease domain is involved in binding of a DNA substrate with the requisite length.

The yeast PI-SceI endonuclease catalyzes the hydrolysis of two specific phosphodiester bonds within an asymmetrical recognition site (1). This enzyme is a homing endonuclease (for a review, see Ref. 2) that occurs as an intein situated within an H\(^{-}\)-ATPase protein subunit. Like other homing endonucleases, PI-SceI recognizes an extremely long sequence (31 bp)\(^1\) and cuts DNA to yield 5’-phosphate and 3’-hydroxyl ends (3, 4). Mutagenesis and biochemical studies indicate that the PI-SceI recognition sequence can be divided into two regions (4, 5).

Region I contains the cleavage site that is cut by the enzyme to generate a 4-base pair overhang, and region II includes an adjacent 17-bp sequence (the minimal binding sequence) that is sufficient for high affinity binding. Mutagenesis of the substrate reveals that PI-SceI tolerates substitutions at numerous positions, since substitutions at only nine positions in the substrate lead to severely reduced activity (4). Like the other homing endonucleases that have been studied, PI-SceI requires Mg\(^{2+}\) as a cofactor. The metal ion is likely to be required for the hydrolytic reaction, since it is required for catalysis but not for specific binding. Mn\(^{2+}\) can substitute for Mg\(^{2+}\), and it stimulates more efficient cleavage by the enzyme at cognate and noncognate sites (1, 5).

The three-dimensional structure of PI-SceI has been recently determined by x-ray crystallography and reveals a bipartite domain structure (6). Domain I contains the protein splicing active site, which is composed of the N- and C-terminal amino acids and two other His residues that have been shown to be required for activity or have been implicated in the reaction (7, 8). The residues that compose the putative endonucleolytic active site, a lysine (Lys301) and two aspartic acid residues (Asp\(218\) and Asp\(326\)), are present in domain II and form a catalytic triad that displays structural similarity to charged clusters found in restriction enzymes (6, 9). By using the PI-SceI structural information and the knowledge that the enzyme contacts two discrete regions of the recognition sequence, a model for the docking of PI-SceI with its substrate was constructed where domains I and II of the protein contact regions II and I, respectively, of the substrate (Fig. 1). In this model, both domains are proposed to contact the substrate, since the binding surface on the endonuclease domain alone is insufficient to contact the entire 31-bp recognition sequence. A bend of ~55° was introduced into the middle of the substrate to accommodate the angular orientation of the two domains with respect to each other, and experimental evidence confirms the existence of this distortion (4, 5). Furthermore, the scissile bonds of the DNA were placed in close proximity to Asp\(218\) and Asp\(326\), which are thought to bind the Mg\(^{2+}\) co-factor. Two symmetry-related \(\beta\)-sheets (sheets 7 and 9) in domain II that flank the active site aspartic acid residues may serve as platforms that contact the cleavage site region. Furthermore, we speculate that a pair of \(\beta\)-hairpin loops between \(\beta\)15 and \(\beta\)16 and between \(\beta\)21 and \(\beta\)22 that lie above the sheets contain amino acids whose side chains mediate substrate binding. The interaction of domain I with region II of the substrate may involve a cluster of positively charged amino acids situated along the same face of PI-SceI as the endonucleolytic active site. The structure of a second homing endonuclease, I-CreI, was recently reported, and a model for its binding to DNA bears similarity to that proposed for PI-SceI (10). I-CreI is a ho-
modimeric protein that resembles domain II of PI-SceI, but it lacks the protein splicing domain. Like PI-SceI, I-CreI contains a set of \( \beta \)-sheet structures with interconnecting extended loops that are proposed to form the protein interface that binds the DNA substrate.

To determine the identity of amino acid residues involved in contacting the PI-SceI recognition sequence, we used alanine-scanning mutagenesis to create a set of mutant proteins with single amino acid changes at numerous positions on the proposed DNA binding interface. These mutant proteins were purified and assayed for their substrate cleavage and DNA binding activities. The major finding of the work is that residues in both domains mediate DNA binding. Moreover, the binding behaviors of wild-type PI-SceI and several mutants provide compelling evidence for a high-affinity interaction between domain I and the minimal binding region and for a substantially weaker association between domain II and the cleavage site region.

**EXPERIMENTAL PROCEDURES**

**Materials**—TALON metal affinity resin and TALON spin columns were obtained from Clontech. All oligonucleotides were synthesized by Genosys Biotechnologies, Inc. Mutagenesis of PI-SceI Gene—Wild-type and mutant PI-SceI proteins were expressed from plasmid pET PI-SceI C-His, which encodes a 479-amino acid PI-SceI derivative containing a polyhistidine C-terminal extension that facilitates rapid protein purification by metal affinity chromatography. To construct pET PI-SceI C-His, PCR mutagenesis (11) was used to insert six silent restriction sites (SpeI, Apol, BscII, BstII, BsiWI, and MluI) at positions 243, 406, 484, 717, 813, and 943, respectively, relative to the first codon) into plasmid pET23PI-Sce ESARC (9) to generate plasmid pET23PI-Sce-9. Plasmid pET23PI-Sce-9 was used as a template in a PCR reaction with two oligonucleotides (5'-TTCGCGATCCTCCGACGGACCAATTGGAGGAGGGAC-3' and 5'-CGGTACCGGGATCAGCTGCGTCTTG-3') to generate a 449-bp fragment. This product was digested with MluI and BamHI and ligated into the MluI/BamHI-digested pET23PI-Sce-9 DNA to create pET PI-SceI C-His. The entire PI-SceI coding region of pET PI-SceI C-His was confirmed by DNA sequence analysis. Omitting the N-terminal methionine residue, pET PI-SceI encodes a 479-amino acid PI-SceI derivative with a C-terminal tail having the sequence HHHHH-COOH. Protein splicing-mediated cleavage of the C-terminal affinity tag was prevented by substituting Asn454 with alanine. To introduce mutations into the PI-SceI coding sequence, oligonucleotide primers were used in either cassette mutagenesis or two-step overlapping PCR amplification protocols (11). All introduced mutations and inserted sequences were confirmed by DNA sequencing.

Expression and Purification of PI-SceI Proteins—Plasmid pET PI-SceI C-His encoding wild-type or mutant PI-SceI proteins was transformed into Escherichia coli strain BL21 (DE3). For most of the mutant proteins characterized, a 200-ml culture was grown in LB medium (1% Bacto-tryptone, 0.5% Bacto-yeast extract, 0.5% NaCl, 1 mM NaOH) containing ampicillin (100 m\( \mu \)g/ml) at 37 °C to an \( A_{600} \) of 0.6–0.8. Expression of PI-SceI protein was induced with 0.5–1.0 mM isopropyl-\( \beta \)-D-thiogalactopyranoside, and growth was continued overnight at 15 °C. The cells were harvested by centrifugation, resuspended in 2 ml of sonication buffer (20 mM Tris-Cl (pH 8.0), 300 mM KCl, 10 mM MgCl\(_2\), 5% glycerol, 1 mM phenylmethylsulfonyl fluoride) containing 1 mM imidazole, and lysed by sonication (3 x 1 min) at 4 °C. All further manipulations were performed at 4 °C. Cell debris was pelleted by centrifugation at 10,000 \( \times \) g for 15 min. The clarified lysate was applied to a TALON spin column (0.5 ml of TALON metal affinity resin) pre-equilibrated with sonication buffer, and the metal affinity columns were inverted for 5 min and centrifuged at 700 \( \times \) g for 2 min. The resin was washed twice with 1 ml of sonication buffer containing 1 mM imidazole, and PI-SceI was eluted from the columns with sonication buffer containing 300 mM imidazole. Elution fractions containing PI-SceI, as judged by SDS-polyacrylamide gel electrophoresis, were pooled and dialyzed overnight in buffer D (10 mM potassium phosphate (pH 7.6), 5% glycerol, 0.1 mM EDTA, and 1.4 mM 2-mercaptoethanol) containing 40 mM KCl (buffer D40). The dialyzed protein was applied to a 1-ml SP-Sepharose column equilibrated with buffer D40, the resin was washed with 2.5 ml of buffer D40, and protein was eluted with 10 x 1 ml of buffer D450. Elution fractions containing purified PI-SceI were pooled and stored in storage buffer (10 mM potassium phosphate (pH 7.6), 50 mM KCl, 2.5 mM 2-mercaptoethanol, and 50% glycerol) at −20 °C. For some PI-SceI proteins, similar protocols were used to purify the enzyme from 1-liter cultures. The PI-SceI proteins were purified to greater than 95% as judged by SDS-polyacrylamide gels. The affinity-tagged PI-SceI (\( M_r = 53,800 \)) concentration was determined using the extinction coefficient of 5.03 × 10\( ^{5} \) M/cm as determined by published methods (12). The wild-type protein had the same specific activity as native PI-SceI. 2

Native Gel Mobility Shift Assay of DNA Binding—To detect protein-DNA complexes in DNA mobility shift assays, a 219-bp duplex DNA fragment containing a single PI-SceI recognition site was synthesized by PCR and labeled with \( ^{32} \)P-ATP as described previously (4). Non-specific binding was measured using a 189-bp duplex DNA fragment that was identical in all respects except that it lacked the PI-SceI recognition site. Each reaction mixture (20 \( \mu \)l) contained 25 \( \mu \)M Tris-HCl (pH 8.5), 100 mM KCl, 10% glycerol, 50 mM bovine serum albumin, 2.5 mM 2-mercaptoethanol, 5 fmol of 219-bp substrate (5'-29-P-labeled at both ends), and PI-SceI as specified and was incubated at 25 °C for 10 min. The samples were subjected to electrophoresis through a 7% native polyacrylamide gel in 0.5 × TBE at 210 V for 5 min and then at 120 V for 2-4 h at 4 °C. The amounts of bound and unbound substrate were determined using a PhosphorImager and FrameNT Analysis software (Molecular Dynamics, Inc.). Autoradiographic exposure of the dried gel to film was used to visualize the unbound DNA and the PI-SceI-DNA complexes.

The PI-SceI reaction pathway can be described by Scheme I (Fig. 2), where the free protein (\( P \)) and DNA (\( D \)) interact to form the lower protein-DNA complex (PDLC) that involves PI-SceI contacts to region II of the substrate (4, 5, 9). PDLC is in equilibrium with the upper complex, PDUC, where PI-SceI, in excess, is bound to the region II target sequence, PDLC, where PI-SceI is proposed to remain tightly bound to the region II pentavalent phosphate transition state that undergoes double-stranded DNA cleavage product following the reaction.

The thermodynamic parameters \( K_1 \) and \( K_2 \) in Scheme I (Fig. 2) can be expressed as follows,

\[
K_1 = \frac{[P][D]}{[PD]} \tag{1}
\]

\[
K_2 = \frac{[PD]}{[PDLC]} \tag{2}
\]

where \( [P] = [P] \) (the total protein concentration) − \( [PDLC] \) and \( [D] = [D] \) − \( [PD] \) − \( [PDLC] \). Under conditions where the total DNA concentration is much less than the total protein concentration (\( [D] \ll [P] \)), \( [P] \approx [P] \), and \( K_1 \) can be expressed as follows.

2 F. S. Gimble, unpublished results.
K340A, and K369A proteins, since the PDUC complexes migrated faster. The K301R, I proteins, purified enzyme (50–150 nM) was incubated with DNA that is unbound. The K301A, K301E, and H377A mutants only formed the PDLC complex even at high protein concentration. No equilibrium dissociation constants were measured for the D229A, K301R, and could not be adequately resolved from the PDLC complex.

Substitution into Equation 3 with expressions for $\theta_{LC}$, the fraction of DNA bound as $PD^{LC}$ (where $\theta_{LC} = [PD^{LC}] / [D]_1$, and for $\theta_{UC}$, the fraction of DNA bound as $PD^{UC}$ (where $\theta_{UC} = [PD^{UC}] / [D]_1$), yields the following.

$$K_1 = \frac{[P]_1 [D]_2 - [PD^{UC}] - [PD^{LC}]}{[PD^{LC}]}$$

(Eq. 3)

Similarly, substitution into the expression for $K_2$ with $\theta_{LC}$ and $\theta_{UC}$ yields the following.

$$K_2 = \frac{\theta_{LC} - \theta_{UC}}{\theta_{LC}}$$

(Eq. 4)

Values for $K_1$ and $K_2$ were determined by nonlinear regression of the gel mobility shift data to Equations 4 and 5 using KaleidaGraph software (Abelbeck Software). Under the conditions of the assay, these equations are valid, because the total protein concentration $[P]_1$ is much greater than the total DNA concentration $[D]_1$. In addition to wild-type PI-SceI, the R90A, R94A, T225A, R231A, D232A, Y328A, T338A, and H343A mutant proteins also formed $PD^{UC}$ and $PD^{LC}$ complexes, and binding could be represented by Scheme I (Fig. 2). Complete binding curves for the R90A and R94A mutants could not be generated due to the low level of binding. However, an estimate of the $K_1$ value could be made, since $K_1 \times K_2 = [P]_1$ when $\theta_{UB} = \theta_{UC}$, where $\theta_{UB}$ is the fraction of total DNA that is unbound. The K301A, K301E, and H377A mutants only formed the $PD^{UC}$ complex even at high protein concentration. No equilibrium dissociation constants were measured for the D229A, K301R, K340A, and K369A proteins, since the $PD^{UC}$ complexes migrated faster than that of wild-type PI-SceI and could not be adequately resolved from the $PD^{LC}$ complex.

PI-SceI Cleavage Analysis—In an initial characterization of the PI-SceI proteins, purified enzyme (50–150 nM) was incubated with XmnI-linearized pBS-PIsc36 (7 nM) in 15 µl of buffer containing 100 mM KCl, 25 mM Tris-HCl (pH 8.5), 2.5 mM 2-mercaptoethanol, 2.5 mM MgCl2) for 30 min and 1 h at 37 °C. The proteins were further purified using chromatography on SP-Sepharose. All of the mutant proteins and wild-type PI-SceI bound to the SP-Sepharose column and could be eluted using buffer containing 450 mM KCl. Samples were subjected to electrophoresis on SDS-polyacrylamide gel electrophoresis and were judged to be >95% pure.3

Characterization of DNA Cleavage Activity of Mutant Proteins—The 48 mutant proteins that were successfully purified were tested for their ability to cleave a PI-SceI recognition site on linearized plasmid pBS-PIsc36. In initial experiments designed to quickly identify mutant derivatives that were partially or completely defective in cleavage activity, an approximately 50–150 nM concentration of PI-SceI protein was incubated with a 7 nM concentration of linearized substrate under standard reaction conditions in buffer containing MgCl2. Table I shows that 34 of 48 mutant proteins tested had at least 25% of the activity of wild-type PI-SceI. These results reveal that mutations can be made at numerous positions in the protein proximal to the active site with little or no effect on block D, a conserved motif found in homing endonucleases and matrukases (7, 8). In domain I, substitutions were introduced at amino acids Arg288, Arg291, Arg304, and Lys301, which comprise the cluster of positive charges thought to bind the DNA. In general, amino acid residues with side chains containing putative hydrogen bond donors or acceptors were targeted, since hydrogen bonds are frequently important components of protein-DNA interactions. Alanine substitutions were introduced, since this residue lacks hydrogen bond partners and it would be expected to exert minimal steric or electrostatic effects on structure. To investigate the effect of charge changes at Lys301, substitution was made at this position with arginine, which maintains the positive charge, and with glutamic acid, which introduces a negative charge.

A total of 49 mutant PI-SceI derivatives were generated containing substitutions at four positions in domain I and 43 positions in domain II (Table I). Cultures of E. coli strains harboring plasmids that expressed the mutant proteins were grown and induced to overexpress PI-SceI. The levels of PI-SceI protein varied for each of the mutants following induction with isopropyl-1-thio-β-D-galactopyranoside and ranged from approximately 2% to greater than 10% of total cell protein. The only exception was the D371A derivative, which could not be purified in sufficient quantities to accurately study. However, a low level of activity was observed with this derivative.3 The PI-SceI proteins were purified by chromatography on a Co2+ metal affinity resin. All enzymes adhered to this column matrix and could be eluted using 300 mM imidazole, just as for the wild-type protein. After this column, the proteins were approximately 90% pure as judged by SDS-polyacrylamide gel electrophoresis. Following the affinity purification step, the proteins were further purified using chromatography on SP-Sepharose. All of the mutant proteins and wild-type PI-SceI bound to the SP-Sepharose column and could be eluted using buffer containing 450 mM KCl. Samples were subjected to electrophoresis on SDS-polyacrylamide gel electrophoresis and were judged to be >95% pure.3

Characterization of DNA Cleavage Activity of Mutant Proteins—The 48 mutant proteins that were successfully purified were tested for their ability to cleave a PI-SceI recognition site on linearized plasmid pBS-PIsc36. In initial experiments designed to quickly identify mutant derivatives that were partially or completely defective in cleavage activity, an approximately 50–150 nM concentration of PI-SceI protein was incubated with a 7 nM concentration of linearized substrate under standard reaction conditions in buffer containing MgCl2. Table I shows that 34 of 48 mutant proteins tested had at least 25% of the activity of wild-type PI-SceI. These results reveal that mutations can be made at numerous positions in the protein proximal to the active site with little or no effect on

3 Z. He and F. S. Gimble, unpublished results.

Figure 2. Proposed reaction pathway for the PI-SceI protein. Explained in detail under “Experimental Procedures.” The two domains of PI-SceI are represented by interconnected spheres.
activity. Of the remaining mutants examined, 12 (R90A, R94A, D229A, R231A, D232A, K301R, Y328A, T338A, K340A, H343A, K369A, and H377A) were partially active (activity levels less than 25% of wild-type activity) and two displayed no activity (K301A and K301E). Surprisingly, one mutant protein (T225A) was at least 3 times more active than wild-type PI-SceI.

More detailed rate experiments were performed for the partially or fully defective proteins and for the enhanced activity protein in reaction buffers containing either MgCl₂ or MnCl₂. These experiments were carried out under single turnover conditions (excess enzyme relative to substrate). Steady state conditions could not be achieved, since PI-SceI remains tightly bound to one of the two cleavage products (5, 9), yielding a low turnover number. The amount of linearized substrate that was cleaved to form the two products was measured as described under “Experimental Procedures,” and the cleavage activities are shown in Table II. In the buffer containing MgCl₂, 25% of the substrate was cleaved by wild-type PI-SceI in approximately 5 min. By contrast, for two of the mutant proteins (K301A and K301E), no cleavage activity was apparent after 4 h of incubation, and for two others (D229A and K340A), only trace amounts of cleavage products were detected. The reaction rates for these mutant proteins were too slow to measure accurately, and we estimate that their activities are at least 500 times lower than that of wild-type PI-SceI. Of the remaining 10 defective mutants, four were >20-fold less active than wild-type PI-SceI (R90A, R94A, Y328A, and H377A), and five were 4–20-fold less active (R231A, D232A, K301R, T338A, H343A, and K369A). The PI-SceI protein with enhanced activity, T225A, cleaved the DNA substrate over 3 times faster than the wild-type protein in the presence of MgCl₂.

Substitution of Mn²⁺ for Mg²⁺ in the PI-SceI cleavage buffer is known to relax the specificity of the enzyme, allowing it to cleave at sites that are resistant to cleavage in MgCl₂ (1, 4), and to increase the overall activity of the enzyme when it cuts at the normal recognition site (1, 5). The data in Table II show that MnCl₂ increases the cleavage rate of wild-type PI-SceI nearly 10-fold, which is similar to rate enhancement levels reported elsewhere (5). Interestingly, for all mutants tested, even for the K301A, K301E, D229A, and K340A mutants that are completely inactive in the presence of MgCl₂ or severely reduced in activity, MnCl₂ increases the cleavage rates, often to near wild-type levels. The levels of activity for the R90A, R94A, R231A, Y328A, T338A, H343A, and K369A proteins in MnCl₂ are less than 4-fold different than that of wild-type PI-SceI (Table II). The level of rate enhancement by MnCl₂ varies for the different proteins; for example, the activity of the Thr⁷⁸⁸ protein is 12-fold higher in MnCl₂, while that of the R94A protein is at least 400-fold higher.

DNA Binding Properties of the Mutant Proteins—To test whether defects in substrate binding by the mutant proteins account for the reduction in cleavage rates, gel shift analyses were performed as described under “Experimental Procedures” using a 219-bp linear fragment containing a single PI-SceI site. Wild-type PI-SceI forms two complexes with this substrate in the absence of metal ion co-factor; a lower complex (PDLC) in which the protein binds solely to a 17-bp minimal binding region distal to the cleavage site (region II) and an upper complex (PDUC) in which PI-SceI binds to both the minimal binding region and to the cleavage site region (region I). Fig. 3 shows that wild-type PI-SceI forms both complexes in the binding experiment. In this report, we used the data to measure two equilibrium dissociation constants, K₁ and K₂, that describe PI-SceI binding to its substrate (see “Experimental Procedures”). Overall binding can be expressed as the product of these parameters and is approximately 0.7 nm (Table III). As suggested previously (4, 5, 9), it appears that the major contributing factor to this tight affinity stems from the interaction of PI-SceI with region II of the substrate, since K₁ is only about 10-fold higher than K₂ × K₃. The high value of K₃, which reflects the partitioning between the lower and the upper complexes, suggests that the binding energy released by the interaction of domain II with the DNA is used to stabilize the energetically unfavorable distorted DNA conformation that is
However, the amounts, and complete binding curves could not be generated.

SceI. PI-40 nM for the R90A and R94A proteins, which are 150- and
reflects the model, the ratio of the upper and lower complexes, which
present in the upper complex. Furthermore, as predicted from
than to the specific probe (a specific DNA fragment of similar size is over 300-fold lower
SceI. 40 nM for the R90A and R94A proteins, which are 150- and

Under conditions where wild-type PI-SceI generates high levels of the PD\(^{UC}\) complex, the mutants yield barely detectable amounts, and complete binding curves could not be generated. However, the \(K_1 \times K_2\) values could be estimated as \(\sim 100\) and \(\sim 40\) nM for the R90A and R94A proteins, which are 150- and 60-fold higher than for wild-type PI-SceI. The H377A, K301A, and K301E proteins contain substitutions in domain II and are similar in that they generate no species that co-migrate with

* Values represent means ± S.D. from at least three experiments.

** Relative activity is the ratio of the mutant to wild-type activities.

---

** DNA Binding and Cleavage Mutants of PI-SceI Endonuclease**

**Table II**

| Allele | Activity \(\times 10^{-3}\) | Relative activity \(\text{a}\) | Activity \(\times 10^{-3}\) | Relative activity |
|--------|-----------------|-----------------|-----------------|-----------------|
| Wild type | 3.9 ± 0.4 | 1 | 32 ± 3 | 1 |
| R90A | 0.18 ± 0.01 | 0.05 | 21 ± 1 | 0.6 |
| R94A | 0.042 ± 0.005 | 0.01 | 18 ± 1 | 0.6 |
| T225A | 13 ± 1 | 3 | ND\(^b\) | ND\(^b\) |
| D229A | 0.27 ± 0.10 | 0.06 | 3.8 ± 0.1 | 0.1 |
| K301A | 0.18 ± 0.01 | 0.04 | 11 ± 1 | 0.3 |
| K301E | 0.99 ± 0.06 | 0.3 | 11 ± 1 | 0.4 |
| Y328A | 0.77 ± 0.10 | 0.2 | 3.9 ± 0.1 | 0.1 |
| H343A | 0.38 ± 0.07 | 0.1 | 14 ± 1 | 0.4 |
| K369A | 0.93 ± 0.011 | 0.02 | 15 ± 2 | 0.6 |
| H377A | 0.77 ± 0.10 | 0.2 | 14 ± 1 | 0.4 |

---

**Table III**

| Allele | \(K_1\) | \(K_2\) | \(K_1 \times K_2\) |
|--------|---------|---------|-----------------|
| Wild type | 4.3 ± 0.3 | 0.16 ± 0.01 | 0.67 ± 0.05 |
| R90A\(^b\) | 2.9 ± 0.4 | 0.22 ± 0.04 | 0.61 ± 0.05 |
| T225A | 8.8 ± 2.9 | 0.18 ± 0.02 | 1.5 ± 0.4 |
| K301A\(^b\) | 1.6 ± 0.1 | 3.1 ± 0.6 |
| Y328A | 14.6 ± 3.3 | 0.26 ± 0.01 | 3.7 ± 0.8 |
| H377A\(^b\) | 2.2 ± 0.7 | 0.25 ± 0.03 | 0.54 ± 0.13 |

---

\(\text{a}\) Complete binding curves could not be generated. \(K_1 \times K_2\) values were estimated from the protein concentration \([P_T]\) at the point where \(\theta_T = \theta_{UT}\) (see “Experimental Procedures”).

\(\text{b}\) No PD\(^{UC}\) complex is observed for these mutant proteins, indicating that the \(K_2\) value is significantly higher than that of wild-type PI-SceI. \(K_1\) was calculated as described under “Experimental Procedures.”

---

**Fig. 3. DNA binding properties of wild-type and mutant PI-SceI proteins.** Gel shift analysis showing protein-DNA complexes formed using purified wild-type PI-SceI or the R90A, R94A, T225A, D229A, R231A, D232A, Y328A, T338A, K340A, H343A, K369A, and H377A (left side) or the K301R, K301E, and K301A (right side) mutant proteins. 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, which is approximately equal to the equilibrium dissociation constant for binding of PI-SceI. No upper complex is present in reactions containing the K301E, K301A, or H377A mutant proteins. Some smearing of the H343A and H377A complexes is evident in the figure, but the complexes are discrete on other gels.

---

**Table IV**

| Mn\(^{2+}\) | Activity \(\times 10^{-3}\) | Relative activity |
|---------|----------------|-----------------|
| Wild type | 3.9 ± 0.4 | 1 |
| R90A | 0.18 ± 0.01 | 0.05 |
| R94A | 0.042 ± 0.005 | 0.01 |
| T225A | 13 ± 1 | 3 |
| D229A | 0.27 ± 0.10 | 0.07 |
| K301A | 0.18 ± 0.01 | 0.04 |
| K301E | 0.99 ± 0.06 | 0.3 |
| Y328A | 0.77 ± 0.10 | 0.2 |
| H343A | 0.38 ± 0.07 | 0.1 |
| K369A | 0.93 ± 0.011 | 0.02 |
| H377A | 0.77 ± 0.10 | 0.2 |

---

\(\text{a}\) Units are \(\text{m DNA cleaved/min} \times \text{m enzyme}\).

\(\text{b}\) Relative activity is the ratio of the mutant to wild-type activities.

\(\text{c}\) ND, not determined.

\(\text{d}\) Accurate measurements of activity could not be made because only trace levels of cleavage products were observed. Rates are estimated to be &gt;500-fold lower than for wild-type PI-SceI.

\(\text{e}\) No detectable activity after 4 h of digestion.

---

**Note:**

4 M. Crist and F. S. Gimble, unpublished results.
the wild-type upper complex. Interestingly, these three proteins appear to bind more tightly to region I than wild-type PI-SceI, since their $K_1$ values are lower (Table III). For the D229A, K301R, and K369A mutants, it is evident from the gel shift assay that the upper complex forms but that it migrates faster than that of wild-type PI-SceI (Fig. 3). The inability to resolve the two complexes prevents accurate determination of $K_1$ and $K_2$, but it is clear from the gel shift analysis that the total amount of bound complexes produced by the D229A and K301R proteins is roughly similar to that of wild-type PI-SceI, while overall binding of the K340A protein is markedly lower (Fig. 3). Both complexes are apparent for the SceI, while overall binding of the K340A protein is.

According to Scheme I (Fig. 2), mutations that modify cleavage activity, i.e. R90A, R94A, D229A, K301A, K301E, Y328A, K340A, and K377A, yield either little or no apparent PD$_{LC}$ complex or produce complexes that migrate faster than that of the wild-type enzyme. The absence of the PD$_{LC}$ complex suggests that important contacts near the cleavage site have been disrupted and that no interaction occurs between PI-SceI and region I, while the appearance of faster migrating complexes indicates possible conformational differences in the complex that may affect the cleavage activity. The T225A mutant, which is approximately 3-fold more active than wild-type PI-SceI, has a $K_1 \times K_2$ value similar to wild type, but we cannot rule out a small binding enhancement.

The main finding of this report, that both PI-SceI domains contact the recognition sequence, is supported by consideration of the thermodynamic binding parameters of the various mutant enzymes together with the positions of the substituted residues in the crystal structure. Fig. 4 shows an overview of the entire protein that indicates the positions of the amino acids where mutations lead to a loss or gain of activity. Two domain I residues, Arg$_{50}$ and Arg$_{54}$, are strong candidates for amino acids that contact region II, since proteins with mutations at these positions have $K_1 \times K_2$ values that are significantly higher than that of wild-type PI-SceI. Residue Arg$_{50}$ is exposed to solvent and lies on the same face of the protein as the active site in domain II, which might be expected if both regions contact the DNA. Little can be concluded from the positioning of residue Arg$_{54}$ since it is part of a disordered loop in the crystal structure, but it is in the same vicinity as Arg$_{50}$.

The residues in domain II that alter activity cluster in groups that neighbor the active site. For example, the Tyr$^{328}$, Lys$^{340}$, and Thr$^{338}$ side chains are in close proximity in the crystal structure. The Tyr$^{328}$ phenolic group and the $\epsilon$-amino group of Lys$^{340}$ are situated within 4 Å of one another and extend upward into the solvent-exposed region above the platform formed by $\beta$-sheet 9 that is thought to contain the DNA (Fig. 5A). The Y328A protein exhibits a ~25-fold reduction in cleavage activity that probably results in part from its reduced DNA binding affinity. However, binding defects alone cannot account for the large reduction in cleavage activity of the Y328A mutant, and there may be effects on catalysis as well. Even more striking is the nearly total absence of activity of the Lys$^{340}$ mutant, which can be easily accounted for by its binding defect. According to our model (Fig. 1), it might be predicted that PI-SceI domain I and domain II mutations affect $K_1$ and $K_2$, respectively. Within domain I, this prediction is borne out by the R90A and R94A proteins. However, Y328A is an example of a domain II substitution that alters $K_2$, which suggests that rather than being independent, the domains communicate with each other. Alanine substitution at the third residue in this group, Thr$^{338}$, increases the $K_2$ value to unity, resulting in equal partitioning between the complexes. The Thr$^{338}$ side chain is not solvent-exposed and would not be expected to contact the substrate. In the other half of the binding platform, which originates from $\beta$-sheet 7, the Thr$^{225}$ side chain also extends above the platform surface (Fig. 5B). Removal of most of the threonine side chain by alanine substitution does not have a major effect on binding. Residues His$^{342}$ and His$^{377}$ are located above one another in two loops that are part of an extended structure that rises above one side of the active site. Both $\epsilon_2$ nitrogens are pointed toward the opening above the active site where the DNA is thought to be located (Fig. 5B).

The behavior of the H377A mutant protein nicely fits our model, since it yields no PD$_{LC}$ complex (high $K_2$ value), and is >50-fold reduced in activity compared with wild-type PI-SceI.

Somewhat surprisingly, the $K_1$ value is nearly 10-fold higher compared with wild-type PI-SceI, which again suggests syn-
ergy between the two domains. The Lys 369 residue is situated in the same loop as His 377, but its orientation is uncertain due to disorder in the structure. However, stereochemical refinement of the structure indicates hydrogen bonding between the Lys 369 and Lys 340 ɛ-amino groups, and the K369A substitution may affect the structure of the binding platform. Diametrically opposite to His 343 and His 377 on the other side of the active site are residues Asp 229, Arg 231, and Asp 232, which form a tight cluster where the side chains are oriented toward the putative substrate binding cavity. A hydrogen bond exists between the Arg 231 ɛ-guanidino group and the Asp 229 carboxyl group. The binding constants for the D229A mutant could not be accurately determined, but it is clear that the large decrease in cleavage activity cannot be accounted for solely by reductions in overall binding (Fig. 3). What is certain is that the mutation alters the mobility of the PD-UCl complex, which may indicate conformational differences in the DNA. Alternatively, as with any of the mutants described here, there may be conformational changes in the catalytic center that affect activity.

The PI-SceI mutants containing substitutions at Lys 301 fall into a separate category, since this amino acid, unlike the other residues characterized here, is highly conserved among homing endonucleases (7, 8) and, together with Asp 218 and Asp 326, forms a “catalytic triad” that comprises the PI-SceI active site (6). Similar clusters of two acidic residues and a lysine residue are found at the active sites of several restriction endonucleases (15). Lys 301 is situated at the C-terminal end of β18 in the PI-SceI crystal structure, and the side chain extends into the putative substrate binding cavity that is also occupied by the two aspartic acid side chains (Figs. 4 and 5). Substitution of Lys 301 with alanine or glutamic acid dramatically increases $K_2$ and consequently eliminates all activity. Similar substitutions at Lys 92 of EcoRV, which may be an analogous residue to Lys 301, reduce substrate binding and cleavage activities (16). The basic character of the Lys 301 side chain is critical for the PI-SceI binding interaction, since a K301R mutant is partially active in binding and cleavage assays. By contrast, arginine substitution at Lys 92 of EcoRV abolished DNA cleavage activity with either Mg$^{2+}$ or Mn$^{2+}$ (17). We also found that cleavage activity of the PI-SceI K301A and K301E mutants could be partially rescued by Mn$^{2+}$ (Table II). A similar effect was observed for the EcoRV K92E mutant protein but not for the K92A protein, which led to speculation that the binding of a second Mn$^{2+}$ ion to the Glu residue restored the positive charge normally contributed by the Lys 92 side chain. This is unlikely to be the case for PI-SceI, since we observe rescue of activity to the K301A mutant as well. In fact, the activity of all of the mutant proteins is partially rescued by substitution of Mn$^{2+}$ for Mg$^{2+}$. It is also worth noting that a set of substrate mutants that are catalytically inactive in Mg$^{2+}$ also have activity restored by the presence of Mn$^{2+}$ (4). Similar instances of activity “rescue” by Mn$^{2+}$ have been observed with EcoRV mutants that have low levels of activity in Mg$^{2+}$ but have nearly wild-type activity levels in Mn$^{2+}$ (16, 18). However, unlike the restriction

![Figure 4](http://www.jbc.org/)

**FIG. 4.** A stereoview of the PI-SceI crystal structure showing the locations of the residues described in this report where amino acid substitutions affect DNA cleavage activity. Also shown are the pair of Asp residues (Asp 218 and Asp 326) that compose the PI-SceI active site. Each of the amino acids are numbered, and their side chains are shown. Domain I, which contains the protein-splicing active site, is shown at the bottom, and domain II, which includes the endonucleolytic active site, is shown at the top. Arg 94 and Lys 369 are located in disordered loops (indicated by the unfilled backbone), and their positions in the figure are based on the stereochemically refined structure.
enzymes, PI-SceI normally displays greater activity in the presence of Mn$^{2+}$ than with Mg$^{2+}$. One EcoRV mutant has been identified for which this is also the case (19). Taken together, our data are consistent with the Lys301 side chain establishing an important binding contact within region I, perhaps to a phosphate oxygen near the scissile phosphodiester bond.

The substrate binding properties of the protein mutants characterized here complement those of a set of loss-of-function DNA substrate mutants that contain substitutions in regions I and II. Point mutations at positions A$^{16}$, G$^{18}$, and A$^{19}$ in region II dramatically reduce all binding to wild-type PI-SceI (4). According to our model, these base pairs are located in the same general vicinity as the R90A and R94A mutant proteins, which display similar binding defects. By contrast, substitutions in the PI-SceI substrate near the cleavage site at positions A$^{-9}$, T$^{-1}$, G$^{-3}$, and G$^{-4}$ only eliminate PD$^{1C}$ complex formation or produce a complex that migrates faster than that of wild-type PI-SceI (4). These binding properties are similar to those of some domain II mutants described here. Thus, there is a good correlation between the DNA binding properties of both the substrate and protein mutants that strongly supports the conclusions of the PI-SceI docking model. However, a convincing demonstration that these proposed interactions occur must await the determination of the PI-SceI structure complexed to its recognition site.

The results presented here are the first to show that the PI-SceI protein splicing domain is involved in site-specific substrate binding. We hypothesized that the PI-SceI intein gene originally arose by the fusion of two pre-existing genes, one that encoded an endonuclease and the other that encoded a splicing protein (6). Surprisingly, the recently determined structure of the autoprocessing domain of the Drosophila Hedgehog protein is very similar to domain I of PI-SceI, but it lacks the PI-SceI DNA recognition region. Instead, it contains an unrelated region that binds cholesterol (20). This suggests that the protein splicing domain existed previously as a core protein that acquired new functions in different instances by associating with new sequences (20). In the case of PI-SceI, it raises the possibility that the DNA binding region was acquired after the intein was assembled. Presumably, the acquired abil-

![Fig. 5](https://example.com/fig5.png)

**Fig. 5.** Stereoviews of the putative DNA binding surface situated in domain II proximal to the endonucleolytic active site. A, a view looking down the long axis of the proposed binding site for the DNA duplex. This view is perpendicular to that in Fig. 4 and is from the vantage point of the lower right side of that figure. One of the two $\beta$-sheets (9) proposed to act as a binding platform for the DNA lies in the foreground. The two Asp residues that compose part of the PI-SceI active site are positioned in the middle and are situated at the C-terminal ends of two parallel $\alpha$-helices. On the other side of Asp$^{218}$ and Asp$^{326}$ is located the pseudosymmetrically related $\beta$-sheet (97), which comprises the binding platform together with $\beta$. The side chains of Tyr$^{328}$ and Lys$^{340}$ are seen to extend upwards into the proposed binding cleft. On the two sides of the binding cleft are structures that rise above the platform and include extended loops. Each of the amino acids where substitutions affect activity are numbered, and their side chains are shown. B, the opposite view of the binding platform, looking from the vantage point of the upper left side of Fig. 4. One of the extended loops situated above the active site includes amino acids Lys$^{369}$ and His$^{377}$. 
ity of the PI-SceI intein to make base-specific interactions to two distant regions of its recognition site provided increased selectivity to the enzyme.

REFERENCES

1. Gimble, F. S., and Thorner, J. (1992) Nature 357, 301–306
2. Mueller, J. E., Bryk, M., Leinzes, N., and Belfort, M. (1993) in Nucleases (Linn, S. M., Lloyd, R. S., and Roberts, R. J., eds) 2nd Ed., pp. 111–143, Cold Spring Harbor Laboratory, Cold Spring Harbor, NY
3. Gimble, F. S., and Thorner, J. (1993) J. Biol. Chem. 268, 21844–21853
4. Gimble, F. S., and Wang, J. (1996) J. Mol. Biol. 263, 163–180
5. Wende, W., Grindl, W., Christ, F., Pingoud, A., and Pingoud, V. (1996) Nucleic Acids Res. 24, 4123–4132
6. Duan, X., Gimble, F. S., and Quiocho, F. A. (1997) Cell 89, 555–564
7. Pietrokovski, S. (1994) Protein Sci. 3, 2340–2350
8. Perler, F. B., Olsen, G. J., and Adam, E. (1997) Nucleic Acids Res. 25, 1087–1094
9. Gimble, F. S., and Stephens, B. W. (1995) J. Biol. Chem. 270, 5849–5856
10. Heath, P. J., Stephens, K. M., Mannat, R. J., and Stoddard, B. L. (1997) Nat. Struct. Biol. 4, 468–476
11. Hu, S. N., Hunt, H. D., Horton, R. M., Pullen, J. K., and Pease, L. R. (1989) Gene (Amst.) 77, 51–59
12. Gill, S. C., and von Hippel, P. H. (1989) Anal. Biochem. 182, 319–326
13. Brown, B. M., Milla, M. E., Smith, T. L., and Sauer, R. T. (1994) Nature Struct. Biol. 1, 164–168
14. Hwang, J. S., Yang, J., and Pittard, A. J. (1997) J. Bacteriol. 179, 1051–1058
15. Aggarwal, A. K. (1995) Curr. Opin. Struct. Biol. 5, 11–19
16. Selent, U., Ruter, T., Kohler, E., Liedtke, M., Thielking, V., Alves, J., Oelgeschlager, T., Walfisz, H., Peters, F., and Pingoud, A. (1992) Biochemistry 31, 4808–4815
17. Vipond, I. B., and Halford, S. E. (1996) Biochemistry 35, 1701–1711
18. Vernate, C. L. M., Vipond, I. B., and Halford, S. E. (1992) Biochemistry 31, 6089–6097
19. Vipond, I. B., Moon, B.-J., and Halford, S. E. (1996) Biochemistry 35, 1712–1721
20. 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
Amino Acid Residues in Both the Protein Splicing and Endonuclease Domains of the PI- SceI Intein Mediate DNA Binding
Zening He, Michael Crist, Hsiao-ching Yen, Xiaoqun Duan, Florante A. Quiocho and Frederick S. Gimble

J. Biol. Chem. 1998, 273:4607-4615.
doi: 10.1074/jbc.273.8.4607

Access the most updated version of this article at http://www.jbc.org/content/273/8/4607

Alerts:
- When this article is cited
- When a correction for this article is posted

Click here to choose from all of JBC's e-mail alerts

This article cites 19 references, 3 of which can be accessed free at http://www.jbc.org/content/273/8/4607.full.html#ref-list-1