Chemical Probing Shows That the Intron-encoded Endonuclease I-SceI Distorts DNA through Binding in Monomeric Form to Its Homing Site*

Benedicte Beylot‡ and Annick Spassky§

From the Laboratoire de Pharmacologie des Macromolécules Biologiques, Institut Gustave Roussy, Villejuif, Cedex, 94805 France

Published, JBC Papers in Press, March 12, 2001, DOI 10.1074/jbc.M101200200

Received for publication, February 7, 2001

Despite its small size (27.6 kDa), the group I intron-encoded I-SceI endonuclease initiates intron homing by recognizing and specifically cleaving a large intronless DNA sequence. Here, we used gel shift assays and footprinting experiments to analyze the interaction between I-SceI and its target. I-SceI was found to bind to its substrate in monomeric form. Footprinting using DNase I, hydroxyl radical, phenanthroline copper complexes, UV/DH-MePyPs photosensitizer, and base-modifying reagents revealed the asymmetric nature of the interaction and provided a first glimpse into the architecture of the complex. The protein interacts in the minor and major grooves and distorts DNA at three distinct sites: one at the intron insertion site and the other two, respectively, downstream (−8, −9) and upstream (+9, +10) from this site. The protein appears to stabilize the DNA curved around it by bridging the minor groove on one face of the helix. The scissile phosphates would lie on the outside of the bend, facing in the same direction relative to the DNA helical axis, as expected for an endonuclease that generates 3' overhangs. An internally consistent model is proposed in which the protein would take advantage of the concerted flexibility of the DNA sequence to induce a synergistic binding/kinking process, resulting in the correct positioning of the enzyme active site.

I-SceI is a homing endonuclease encoded by the mobile group I intron of the large rRNA gene of Saccharomyces cerevisiae (1, 2). This family of enzymes mediates the propagation of the intron by cutting intronless genes at the site of intron insertion (reviewed in Ref. 3). Like restriction enzymes, homing endonucleases cleave double-stranded DNA with high specificity in the presence of divalent metal ions. However, they differ from restriction endonucleases in their recognition properties and the presence of divalent metal ions. Despite its small size (27.6 kDa), the group I intron-encoded I-SceI endonuclease initiates intron homing by recognizing and specifically cleaving a large intronless DNA sequence. Here, we used gel shift assays and footprinting experiments to analyze the interaction between I-SceI and its target. I-SceI was found to bind to its substrate in monomeric form. Footprinting using DNase I, hydroxyl radical, phenanthroline copper complexes, UV/DH-MePyPs photosensitizer, and base-modifying reagents revealed the asymmetric nature of the interaction and provided a first glimpse into the architecture of the complex. The protein interacts in the minor and major grooves and distorts DNA at three distinct sites: one at the intron insertion site and the other two, respectively, downstream (−8, −9) and upstream (+9, +10) from this site. The protein appears to stabilize the DNA curved around it by bridging the minor groove on one face of the helix. The scissile phosphates would lie on the outside of the bend, facing in the same direction relative to the DNA helical axis, as expected for an endonuclease that generates 3' overhangs. An internally consistent model is proposed in which the protein would take advantage of the concerted flexibility of the DNA sequence to induce a synergistic binding/kinking process, resulting in the correct positioning of the enzyme active site.

I-SceI is a homing endonuclease encoded by the mobile group I intron of the large rRNA gene of Saccharomyces cerevisiae (1, 2). This family of enzymes mediates the propagation of the intron by cutting intronless genes at the site of intron insertion (reviewed in Ref. 3). Like restriction enzymes, homing endonucleases cleave double-stranded DNA with high specificity in the presence of divalent metal ions. However, they differ from restriction endonucleases in their recognition properties and structures, as well as in their genomic location (4). In particular, whereas restriction enzymes have short recognition sequences (3–8 bp),1 homing endonucleases, despite their small size, recognize long DNA sequences (12–40 bp). They have been classified into four families on the basis of both their sequence motifs and DNA cleavage mechanism (3). The protein I-SceI is a member of the largest class of homing enzymes (more than 130 proteins), characterized by the presence of either one or two conserved 12 amino acid residue sequence motifs (LAGLI-DAGD motifs). Most of these proteins, like I-SceI, carry the motif in duplicate and are endonucleases. I-SceI has been purified as a monomeric globular protein of 235 amino acids (5). Its endonuclease activity requires Mg2+, or Mn2+ but not Co2+, Cu2+, or Zn2+ to cleave DNA within its recognition sequence and leaves a 4-bp overhang presenting a 3'-hydroxyl terminus (5, 6). The enzyme displays a low turnover, probably because of its strong affinity for one of the products of the cleavage reaction (7).

The interaction of homing endonucleases with their substrates raises an interesting question common to all the gene-regulatory proteins, namely: how can a small protein specifically recognize and modify a long DNA sequence? Understanding the molecular basis of such a mechanism is essential for elucidating many aspects of cellular control and is a prerequisite in any rational drug design program. It is clearly established that local and global DNA structural features that are highly sequence-dependent, play a primary role in the dynamics of protein-DNA recognition (8). Sequence recognition would arise from the inherent sequence-dependent ability to adopt the conformation required for protein binding in the transient, biologically active complex. In an attempt to reveal intrinsic helical properties of the I-SceI nucleic acid target, we used this sequence as substrate in earlier studies analyzing the mechanisms of DNA chemical reactions and photosensitization processes (9, 10, 11, 12). We found evidence that the conformation of the helix deviates from the ideal B-form duplex along two segments of three and five base pairs located at a distance of approximately one helical turn, respectively, upstream and downstream from the site of junction of the two exons (9). In the present study, we first performed DNase I footprinting and gel retardation assays to identify the complex formed between I-SceI and its target in the absence of a divalent metal ion. We then used chemical probing agents to characterize the conformation of DNA in the complex. I-SceI protein-DNA complex was thus submitted to the nucleolytic attack of the cupric complexes of 1,10-phenanthroline (OP2Cu1+, or Zn2+), which results from the abstraction of C-1 hydrogen atom by a tetrahedral copper-oxo species bound within the minor groove (13, 14, 15, 9) and to UVA/4,5-dihydro-7-methylpyrido [3,4-c]psoralen (DHMePyPs) photosensitization, which requires prior intercalation of the pyridopсорalen at selective 5'-TTA-3' sites (10). We also employed chemical modification agents of base and sugar residues (for a review, see Ref. 16),
diethyl pyrocarbonate (DEPC), which carboxylates purines at the N-7 atom, potassium permanganate (KMnO₄), which oxidizes pyrimidine residues at the C₅=C₆ double bond, dimethyl sulfate (DMS), which primarily methyitates the N-7 of guanine residues and free hydroxyl radical, generated by Fe-EDTA reduction of hydrogen peroxide, which abstracts C-4' hydrogen atoms from deoxyribosides of the DNA backbone. In the scheme that arises from present experiments, I-SceI appears to stabilize, in monomeric form, a constrained helical structure in which the minor groove is widened at the cleavage sites. The results are discussed in relation to previous reports on other endonucleases of the same family.

**EXPERIMENTAL PROCEDURES**

I-SceI was purchased from Roche Molecular Biochemicals, aliquoted at 10 units/µl in phosphate buffer in the presence of 200 µg/ml bovine serum albumin, and conserved in 50% glycerol at −20 °C. Protein concentration was determined from the optical density of the bands on a Phast System minigel (12.5% acrylamide/5.5% SDS), using bovine serum albumin as internal standard (not shown). The solution used in the present study had a concentration of 0.72 × 10⁻¹⁰ mole per enzymatic unit.

The 98-bp EcoRI-HindIII DNA fragment including the I-SceI recognition sequence (sequence shown in Fig. 1A) was excised from pUC19 vector supplied by B. Dujon (6) and was purified by electrophoresis on a 15% preparative native polyacrylamide gel as described (50). Concentration was measured by UV absorbance. The fragment was stored in 10 mM Tris, pH 7.5, 1 mM EDTA (TE). For only 5' end-labeling, pUC19 plasmid vector was first digested with either the restriction enzyme EcoRI or HindIII, dephosphorylated with calf intestine alkaline phosphatase, 5' end-labeled with T4 polynucleotide kinase in the presence of [γ-³²P]ATP, and then digested with the second restriction enzyme before purification by 20% native polyacrylamide gel electrophoresis.

Synthetic oligonucleotides used to form the 54- and 37-bp fragments (Fig. 1, B and C), were purchased from Genset (France). Purification, labeling, and annealing were carried out as previously described (12).

**Gel Shift Analysis of I-SceI/DNA Interactions**

The conditions were derived from those described previously (17, 18). I-SceI (10⁻⁹ to 10⁻⁷ M) and 35³²P-labeled DNA (either 98, 54, or 37 bp DNA fragments) (10⁻¹⁵ to 10⁻⁸ M) were preincubated separately for 2 min at 4 °C in the binding buffer (final concentration: 10 mM Tris-HCl, pH 8, 10 mM NaCl, 2.5 mM dithiothreitol and 20 mM MgCl₂). A 0.1 volume of loading buffer (50% glycerol/0.02% xylene cyanol) was added. 8–12% polyacrylamide gels (29:1 acrylamide/bisacrylamide) were prerun in 1× TBE (9 mM Tris-HCl, 8.8 mM boric acid, and 2 mM EDTA) for 2 h at 120 V. Samples were loaded at 50 V, and the gels were run at 120 V at 4 °C. Gels were dried and placed in phosphorimager cassettes. Screens were exposed for several hours and scanned using a Molecular Dynamics PhosphorImager with Image Quant software. The fraction of bound (free) DNA in each lane was calculated by dividing the area of bound (free) bands by the total area of bound and free bands. Each binding assay was performed in triplicate.

**Quantification of the Apparent Equilibrium Dissociation Constant**

The apparent equilibrium dissociation constant was derived from the Scatchard plot of the binding data in which the ratio of bound to free DNA concentration was plotted against bound DNA concentration using the Kaleidograph (Synergy software). The reciprocal of the negative slope of the linear plot gives the value of the apparent Kₐ.

**DNA-Protein Stoichiometry**

The 5' end-labeled DNA fragment (Fig. 1B, 54bp) (8 × 10⁻¹⁰ M) and I-SceI protein (0.01 µg) were incubated as described above, before loading onto a set of 8, 10, 12, and 15% polyacrylamide gels, alongside 10 µg of nonadenated protein molecular size standards (Sigma). Gels were stained with Coomassie Blue, destained, and dried, and exposed to x-ray film. The relative mobility of each species including free DNA fragment (RF) was calculated by dividing the distance of the corresponding band by that of the bromophenol blue tracking dye in the same lane. For each species, the plot of 100 log (100/RF) against gel concentration was constructed. The negative slope or retardation coefficient (−Kₐ) was then plotted as a function of the molecular mass for each protein standard, and this calibration line was used to determine the apparent molecular mass of the free DNA fragment and that of the protein I-SceI/DNA complex. The difference between these two values divided by the molecular mass of the protein monomer gives the number of protein monomers bound to the DNA (n).

**Complex Probing Using DNA Cleavage Reagents**

Uniquely 5' end-labeled DNA fragments (Fig. 1, A, B, or C) were digested directly or after incubation with the I-SceI protein as described above. In each case, digestion was carried out under conditions such that the DNA molecule was broken only once.

**DNase I Footprinting—**DNase I footprinting was done essentially as previously described (51). Digestion was carried out at 23 °C using DNase I at a final concentration of 0.025 µg/ml for 15 or 45 s, depending on whether the DNA was free or I-SceI-bound.

**Hydroxyl Radical Footprinting—**A stock solution of iron(II)-EDTA was prepared immediately before use by mixing equal volumes of freshly prepared 0.4 mM (NH₄)₂Fe(SO₄)₂·6H₂O and 0.8 mM EDTA. The footprinting reaction was initiated by placing iron(II)-EDTA solution (3 µl), 0.6% hydrogen peroxide (3 µl), and 20 mM sodium ascorbate (3 µl) on the inner wall of the 1.5-ml Eppendorf tube containing 21 µl of free or I-SceI-bound DNA, allowing the reagents to mix and then adding the cutting reagent to the sample solution. The reaction was allowed to run for 30 s and quenched by adding 3 µl of 1 M thiourea.

**Phenanthroline Copper Complex Probing Using DNA Cleavage Reagents**

**Orthophenanthroline-cuprous complex (OP_Cu²⁺ or 5' Ph OP_Cu²⁺ ) footprinting was carried out as previously described (29, 30). One µl of a solution freshly prepared by diluting an ethanolic solution of 1,10-phenanthroline (1 mM) and an aqueous cupric sulfate solution (0.23 mM) was added to 10 µl of the appropriate free or I-SceI-bound DNA sample. Cleavage was initiated by the addition of 1 µl of 58 mM MPA (final concentration, 8 mM) which was incubated at 23 °C for either 30 or 60 s. 10 µl of 5 Ph OP_Cu²⁺ was quenched by the addition of 1 µl of 28 mM 2,9-dimethyl-orthophenanthroline (final concentration, 2.8 mM).**

**Analysis of Cleaved Fragments—**Specific quenching of the footprinting reagent was followed by addition of a general stop solution to a final concentration of 1 mM EDTA, 0.3 mM sodium acetate, and 10 µg/ml tRNA. After phenol extraction, samples were ethanol-precipitated and lyophilized. The dried samples were resuspended in 10 µl of gel loading buffer and analyzed by denaturing gel electrophoresis in 15% (w/v) polyacrylamide containing 7 M urea. After electrophoresis, gels were dried on Whatman 3MM paper and exposed to x-ray film (X-OMAT) for documentation or to storage out of phosphor screens for quantification.

**Complex Probing Using UVA DHMePyPs Photosensitization**

UVA (365 nm) irradiation in the presence of DHMePyPs of the I-SceI protein-bound DNA fragment (98 bp) was performed exactly as previously described for free DNA (10). One µl of an ethanolic psoralen solution (10⁻⁷ M) was added to 19 µl of the appropriate labeled DNA sample in binding buffer. After 10 min of incubation at room temperature in the dark, sample-containing droplets were irradiated on ice at 365 nm using an HPW 125 Philips mercury lamp at a fluence of 25 J/m²/s, as determined by a VLX 365 radiometer. After irradiation, psoralen and protein were extracted with chloroform/isooamyl alcohol/phenol followed by G50-Sephadex column chromatography. The DNA was then ethanol-precipitated and treated as above.

**Complex Probing Using Nucleobase Modifications (52)**

Whereas quite unreactive toward double-stranded adenine and guanine residues, diethyl pyrocarbonate can carboxylate the N-7 atom of purines of distorted structures, with a strong preference for adenines, thus destabilizing the imidazole ring and creating a piperidine-sensitive site. Similarly, because potassium permanganate oxidizes the C₅-C₆ double bond of pyrimidines (T>C) from above or below the plane of the base, these residues are susceptible to attack by KMnO₄ only if the stacking interaction is disrupted.

**Thymidines Using Psoralen as a Footprinting Agent (KMnO₄)—**One microliter of freshly prepared 0.1 µM KMnO₄ was added to 5 µl of the appropriate labeled DNA sample. The reaction was stopped after 4 min at 23 °C by addition of 2 µl of β-mercaptoethanol. After phenol extraction, samples were ethanol-precipitated, washed and dried.

**Adenines Using Diethyl Pyrocarbonate (DEPC)—**One microliter of freshly prepared 3% DEPC was added to 10 µl of the appropriate labeled DNA sample. The reaction mixture was incubated for various
time (30 s; 2.5 and 10 min.) at 30 °C and then stopped by addition of
10 μl of 50 mm imidazole. After phenol extraction, samples were etha-
nol-precipitated, washed, and dried.

Guanines Using Dimethylsulfate—The DNA fragment (Fig. 1),
uniquely 5'-32P-end-labeled on the top or on the bottom strand, was
methylated either directly or after incubation with I-SceI protein (see
above), by adding dimethyl sulfate directly to the reaction mixture. The
concentration of methylating agent, reaction temperature, and incuba-
tion time were determined so as to obtain in each case 1 N-7-MeG lesion
per strand.

Processing of Modified DNA—The pellets of modified DNA samples
were resuspended in 100 μl of 1 m piperidine at 95 °C for 30 min.
Piperidine was removed by extensive lyophilization. The dried samples
were resuspended in 10 μl of gel loading buffer (as described in ref. 50)
and analyzed by denaturing gel electrophoresis in 15% (w/v) polyacryl-
amide containing 7 m urea. After electrophoresis, gels were dried on
Whatman 3MM paper and exposed to x-ray film (X-OMAT) for docu-
mentation or to storage out of phosphor screens for quantification.

Quantification of Results

The autoradiograms were scanned by using a PhosphorImager and
Image Quant software (Molecular Dynamics). Measurements and nor-
malization were carried out exactly as previously described (12).

RESULTS

I-SceI/DNA Binding; DNase I Footprinting, Affinity, and
Stoichiometry—Analysis of the enzymatic activity of a collection
of mutations around the cleavage site has previously dem-

FIG. 1. DNA sequences of the fragments used in this study. The
18 bp DNA sequence, previously identified as the minimal required for
optimal cleavage activity, is marked by brackets. IS and CS refer,
respectively, to the intron insertion site and the I-SceI protein cleavage
sites. Base pairs upstream and downstream from IS are numbered from
-1 and +1, respectively. A, 98-bp EcoRI-HindIII restriction DNA frag-
ment from pUC18 (6); B, 54-bp synthesized oligonucleotide duplex; C,
37-bp synthesized oligonucleotide duplex.

FIG. 2. DNase I footprinting of
I-SceI on its DNA homing site. A, gel
analysis of the pattern of DNase I attack
of the 98-bp EcoRI-HindIII restriction fragment uniquely 5'-32P-labeled either
on the top strand at the HindIII site
(lanes 1–4) or on the bottom strand at the
EcoRI site (lanes 5–8), naked (lanes 2 and
6) or I-SceI protein-bound (lanes 3 and 7).
A+G Maxam-Gilbert sequencing reac-
tions were in lanes 1 and 4, 5 and 8, re-
spectively. Numbers refer to the position
of the base with respect to the intron inser-
tion site upstream from (-1) and
downstream from (+1). B, schematic rep-
resentation of the influence of I-SceI bind-
ing on the DNase I digestion frequency.
The cleavage frequency is reduced by
more (open rectangles) or less (gray-filled
rectangles) than 70% and unchanged or
enhanced (black-filled rectangles). The
bands at +13 on the top strand and +1 on
the bottom strand were taken as arbi-
trary references of 100% inhibition.

25245
Fig. 3. Binding of I-SceI to its DNA substrate and stoichiometry. A, autoradiography of a typical gel shift assay. A fixed concentration (10^{-9} M) of 5'-32P-labeled 54-bp DNA fragment was incubated with increasing concentrations of I-SceI protein in 10 mM Tris-HCl buffer, pH 7.9, containing 10 mM NaCl, 10 mM EDTA, 1 mM dithiothreitol, and 20 μg/ml bovine serum albumin prior to loading on a non-denaturing polyacrylamide gel (12%). From lanes 1–6, I-SceI concentrations were: 0 M, 0.2 × 10^{-9} M, 0.7 × 10^{-9} M, 2 × 10^{-9} M, 4 × 10^{-9} M, 8 × 10^{-9} M. Protein-bound DNA migrates more slowly than free DNA (indicated on the left). B, determination of the apparent equilibrium dissociation constant (K_d). Gel shift assays using a fixed concentration of the I-SceI protein and various concentrations of the 54-bp DNA fragment (Fig. 1) were carried out as in A. The fraction of DNA bound to the protein was calculated for each DNA concentration from the radioactivity in the bands of the gels (see “Experimental Procedures”). Scatchard analysis of the binding data (see “Experimental Procedures”) yielded a value of 0.8 ± 0.03 × 10^{-9} M for K_d. C, representative Ferguson analysis. A logarithmic function of mobility for each of the protein standards and for free and I-SceI-bound DNA (54 bp) was plotted against the polyacrylamide concentration and fitted to a linear regression. Protein standards were α-lactalbumin (6.14 kDa), carbonic anhydrase (6.29 kDa), chicken egg albumin (45.0 kDa), bovine serum albumin monomer (66.0 kDa), bovine serum albumin dimer (132.0 kDa). The slopes of each line represent the retardation coefficient (K_r) for each species. D, representative plot of K_r versus molecular size (MW). K_r values for protein standards were plotted as a function of molecular size. Interpolation of K_r values indicates a molecular size of 56.45 kDa for the I-SceI/DNA complex and 29.74 kDa for the free DNA substrate. E, table of molecular sizes and K_r values. To determine the value of the protein component of the complex, we subtracted the molecular mass of the free DNA contribution from that of the complex, giving 26.71 kDa for the molecular mass of the protein in the complex.

It can be observed (Fig. 2A) that the binding of the protein not only protects DNA from -12 to +15 on the top strand/-12 to +12 on the bottom strand but in addition induces changes in the DNase I cleavage frequency from -20 to +28 (Fig. 2A). The 54-bp and 37-bp DNA fragments, containing or not downstream distal sequences (Fig. 1, B and C) were hence prepared and assayed for protein-binding in gel shift experiments (17, 18). In both cases, we observed a discrete band of decreased mobility (Fig. 3A and not shown) reflecting the formation of a well defined complex. This complex is specific because its formation could be competed by the unlabeled DNA fragment but not by poly(dI-dC) (2 μg/ml) (data not shown). Scatchard plots (19, 20, 21) of experiments varying either the protein or DNA concentration yielded an apparent K_d of 0.8 ± 0.03 nM and 8.4 ± 0.4 nM when using, respectively, the 54-bp (Fig. 3B) or the 37-bp (not shown) DNA fragments, thus showing that the interaction is strengthened by the presence of distal downstream sequences. It is interesting to note here that in experiments comparing the kinetics of the initial phase of the cleavage reaction by incubating either of the three DNA substrates (98, 54, and 37 bp) with an excess of protein in the presence of 0.005 M MgCl_2 (as previously described in Ref. 7), the 37-bp DNA fragment was cleaved 1.4–1.6 times more slowly than the 54-bp fragment, itself cleaved at a roughly similar rate as the 98-bp DNA fragment (not shown). To examine the possibility raised by these results that one protomer of I-SceI (molecular mass, 27.6 kDa) binds to a single DNA site, we then determined the molecular mass of the 54-bp DNA/protein complex by using the Ferguson method (22, 23). In this method, which requires that DNA fragments are not too long so that the complex shape does not deviate significantly from globular (24, 25, 26, 27), the mobility of the complex is compared with that of standard proteins in a set of nondenaturing gels of increasing polyacrylamide concentration. Experiments using the 54-bp DNA fragment (Fig. 3, C and D) lead to a molecular mass of 56.45 kDa with a relative error estimated at ± 4.23 kDa over four trials. Subtracting the contribution of free DNA yields an estimate of 26.71 ± 4.23 kDa for the molecular mass of the protein component (Fig. 3E). The protein has appeared as a 26-kDa monomer in solution (5).
Further analysis of the DNase I footprint by converting band intensities to probabilities of cleavage at each site relative to that in the control (see “Experimental Procedures”) shows that the degree of protection of discrete phosphodiester bonds against the enzymatic digestion is not uniform. Areas from
Fig. 5. Comparison between I-SceI binding to 98-bp wild-type and variant 98-bp DNA substrate with the substitution (G/C to A/T) at +7. I-SceI protein was incubated in 10 mM Tris-HCl buffer, pH 7.9, containing 10 mM NaCl, 10 mM EDTA, 1 mM dithiothreitol, and 20 µg/ml bovine serum albumin with the 5’-32P-labeled 98-bp DNA fragment substrate (10−8 M), either wild type (lanes 1–2, 5–6, and 9–10) or mutant (lanes 3–4, 7–8, and 11–12) prior to either loading on an 8% non-denaturing polyacrylamide gel (lanes 1–4) or OP2Cu⁺ foot printing (lanes 5–12). Protein concentration was respectively 0.4, 10−8 M and 10−7 M in the wild-type and mutant experiments. DNA fragments 5’-32P-end-labeled on the top strand or on the bottom strand were, respectively, used in lanes 5–8 and 9–12.

Fig. 6. Influence of the binding of I-SceI on the formation of photoproducts from DNA pyridoporphalen photosensitization. A, comparison of the T4 exonuclease digestion of the 98-bp DNA fragment, 5’-32P uniquely end-labeled on the top strand (lanes 1 and 2) or on the bottom strand (lanes 3 and 4) after exposition to UVA (365 nm) in the presence of DH-MePyPs, either naked (lane 1 and 3) or protein-bound (lanes 2 and 4). Numbers refer to the position of the base with respect to the intron insertion site, upstream from (−1) and downstream from (+1). B, schematic representation of DH-MePyPs photosensitized I-SceIUDNA complex damage. Arrows show the position of 3’-5’ exonuclease activity blockage, 3’ to the damaged residue. Arrow size correlates with the intensity of the bands. From the position of these arrests on both strands, we deduce that there is a position (filled vertical bar) where intercalation of pyridoporphalen is highly favored when I-SceI is bound to its substrate.
Influence of I-SceI DNA binding on the modification of base residues. A, DEPC modification of adenines and KMnO₄ modification of thymines. Typical autoradiogram of a sequencing gel showing the modification of the DNA substrate, 5'-²³P-end-labeled uniquely on the top strand (lanes 1–2 and 3–6) or on the bottom strand (lanes 3–4 and 7–8), either free (lanes 1, 3, 5, and 7) or protein-bound (lanes 2, 4, 6, and 8) using either DEPC (lanes 1–4) or KMnO₄ (lanes 5–8). Numbers refer to the position of the base with respect to the intron insertion site, upstream from (−1) and downstream from (+1). The 54- and 98-bp DNA fragments were used in the DEPC and KMnO₄ experiments, respectively. B, diagram of the reactivity of the I-SceI/DNA complex. The reactive adenines and thymines are indicated by black-filled squares (strongly reactive), gray-filled squares (moderately reactive), and open squares (weakly reactive).

We then prepared a variant 98-bp DNA fragment (Fig. 1A) with the substitution (G/C to A/T) at +7, which has been shown severely defective for I-SceI-mediated cleavage (6) and almost completely resistant to I-SceI cleavage in vitro (not shown). We incubated this fragment in the presence of I-SceI and assayed either for binding in gel shift experiments or for copper phenanthroline complex footprints. No specific complex was observed in gel shift assays under the conditions used previously with the wild-type DNA, but a smear appears along the lane at high protein concentrations (Fig. 5). In the copper phenanthroline footprint, we observe concomitantly the complete loss of the downstream OP₂Cu⁺ hole as well as that of the 5-PheOP₂Cu⁺ hyperreactivity at +10 (not shown) and the absence of the reactivity in the catalytic region (Fig. 7, compare lanes 6 and 8; and not shown). Protein binding in the downstream region thus appears directly related to the induction of a conformational change in the catalytic region.

The Helix Is Sharply Kinked at the Junction between the Two Exons—To further investigate the structural change induced upon protein binding at the junction of the two exons, we utilized a new UVA triplet photosensitizer, 4',5'-dihydro-7-methylypyrido[3,4-c]psoralen (DHMePyPs), which induces thymine dimerization at selective sites in DNA (10). The 98-bp DNA fragment (Fig. 1) 5' radiolabeled either at the HindIII or EcoRI site, was exposed either directly or after incubation with I-SceI protein, to UVA irradiation (30 kJ m⁻² at 365 nm) in the presence of DH-MePyPs and then digested with the T₄ DNA polymerase 3'-5' exonuclease activity (Fig. 6A). This enzymatic activity is commonly used for quantitative detection of UV-induced DNA damage, adducts, or pyrimidine cyclobutane dimers (33). The lesions corresponding to blockage of the exonuclease activity at positions 1, +8, and +13 have been identified from T₄ endonuclease-specific targeting, in naked DNA, as the cyclobutane pyrimidine dimers T₈T₉, T₉T₁₀ and T₁₄T₁₅ (10). Free and protein-bound DNA displayed significantly different cleavage patterns in the photosensitization experiment (Fig. 6A). The most striking result is the dramatic increase (more than two orders of magnitude) in the frequency of termination of exonuclease activity at position −1. This finding, reflecting the high frequency of dimerization of thymines T₁T₂, T₅T₆ indicates that binding of the protein promotes intercalation of pyridoporsalen at the base step T₂A₁. Because the intercalation of the pyridoporsalen proceeds via the minor groove (34), this reflects an opening up of the minor groove, suggesting that I-SceI binding at the site of junction of the two exons generates a kink toward the major groove. Such a kink would induce the widening of the flanking minor groove, in agreement with the OP₂Cu⁺ hyperreactivity at positions +2 to +4 on the top strand and +1 to +3 on the bottom strand (Fig. 4, B and C). A sharp change in the direction of the helix axis at the step T₁A₁ is also fully consistent with the decrease in the OP₂Cu⁺ activity of positions +1, +2 on the top strand and positions −1, −2 on the bottom strand in complexed compared to free DNA because the model of OP₂Cu⁺ binding involves the long axis of one phenanthroline ring parallel with the minor groove axis (9). The absence of the band corresponding to the dimer T₁T₁₀ in the cleavage pattern of the bottom strand agrees also with the protein binding-kinking proposed to interpret copper complexes footprinting results (see above). Note also the appearance of two new bands at +1 and +4 in that of the top strand, which have not, so far, been identified (Fig. 6).

Evidence for Local Stacking Defects at Three Distinct Sites—Because local unstacking and/or unwinding of bases might be expected as a result of helical distortion, we analyzed the enhancement of the sensitivity of adenine and thymine residues to diethyl pyrocarbonate and potassium permanganate, respectively (for a review, see Ref. 16). DNA substrate, uncomplexed or complexed to I-SceI, was reacted with these reagents and cleaved at the modified bases by reaction with hot piperidine. The products were mapped at single-nucleotide resolution by analysis of the fragments produced on denaturing polyacrylamide gels. Fig. 7A (lanes 1–4) illustrates both the reactivity of double-stranded DNA with DEPC and the influence of protein binding on this reaction. Cleavage observed at guanines and to a much lesser extent at cytosines is caused by the sensitivity of these residues to hot piperidine, totally independent of DEPC treatment (control not shown). Only certain adenines of the top strand display increased sensitivity to DEPC upon protein binding (compare lanes 1 and 2). This enhancement is weak at positions +1, +2, and +4, stronger at positions +9 and −6, and stronger still at position −2 (Fig. 7B). In contrast, the adenines of the bottom strand are unreactive to DEPC, and the residue at +8...
I-SceI Distorts DNA in the Absence of Divalent Metal Ion

Fig. 8. DMS methylation and interference results. A. 5’-32P-labeled DNA substrate (54 bp) uniquely end-labeled on the top strand was subjected to DMS methylation, either free (lane 1) or I-SceI-bound (lane 2). 5’-32P-labeled DNA substrate uniquely end-labeled on the top strand was DMS methylated prior to incubation in the presence of I-SceI. Unbound DNA was separated from I-SceI-bound DNA on a native polyacrylamide gel, cleaved with piperidine, and analyzed on a denaturing gel. The I-SceI-bound DNA fragment was loaded in lane 3 and unbound DNA in lane 4. B, extent of guanine protection against DMS methylation by I-SceI DNA binding. Peak heights are proportional to the probability of cleavage at individual guanines of DNA, free in solution (stippled bars) or bound to I-SceI (black bars). Black arrows indicate the methylated guanines responsible for strong binding interference.

even appears protected (compare lanes 3 and 4). The sensitivity of thymines to K\textsubscript{MnO}4 attack is based on the comparison between guanine and thymine cleavage frequencies. Cleavage does not occur at thymines in free DNA, whereas in the protein-DNA complex, piperidine cleavage observed at some of the thymines is similar to or greater than that observed at guanines. The most dramatic increase in cleavage efficiency is observed at thymine T\textsubscript{9} on the bottom strand and to a lesser extent at T\textsubscript{11}, on the top strand. The reactivity is also enhanced at positions T\textsubscript{2}, T\textsubscript{4}, and T\textsubscript{8} on the bottom strand and T\textsubscript{7}, T\textsubscript{10}, and T\textsubscript{11} on the top strand (Fig. 7; A, lanes 5–8; and B). These observations are in full agreement with the interpretation of previous results.

Methylation Protection and Interference—Dimethyl sulfate methylation has been used to probe DNA-protein contacts at the major groove side of G/C base pairs (16). The methylation rate of the guanine residues at positions −3, −4, and −5 and 5, 6, and 7 on the top strand and 3 on the bottom strand is reduced in the complex compared with free DNA (Fig. 8). We interpret this result as indicating that the protein makes DNA contacts in the major groove on either side of the cleavage sites but is not in intimate contact with any of the guanine residues. In a second series of experiments, the end-labeled DNA fragment was first methylated and then incubated with the protein under similar conditions as those previously defined when using non-modified DNA (see “Experimental Procedures”). The bound fraction of modified DNA was then separated by gel-shift from those species in which the modification prevents protein binding. Isolated bound and unbound DNAAs were then piperidine cleaved at the modified positions and resulting fragments were analyzed on denaturing polyacrylamide gels. Comparison of cleavage products of DNA in the complex and in the remaining free DNA leads to the conclusion that at guanine residues +5, +6, and +7, N-7 methylation totally prevents the formation of the I-SceI/DNA complex whereas it has only a slight effect at residues −3, −4, and −5 (Fig. 8A, lanes 3 and 4). This result again underscores the primary importance of the downstream region. Furthermore, because the methylation of guanine residues is merely reduced but not totally inhibited, the suppression of an important hydrogen bond between the protein and these residues does not appear to be the cause for the strong interference of methylation of guanines +5, +6, and +7 with protein binding. The data suggest that the inhibitory effect of introducing a methyl group in the major groove of the flanking region downstream of the cleavage site may be related, at least in part, to hindering the establishment of the helical alteration required for protein binding.

I-SceI Protein-DNA Binding; Minor Groove Backbone Interactions and DNA Distortion—Specific contacts between the I-SceI protein and the DNA backbone were monitored using free diffusible hydroxyl radicals generated by hydrogen peroxide and iron(II) complexed with EDTA (H\textsubscript{2}O\textsubscript{2}/Fe\textsuperscript{2+}-EDTA) (35). Fig. 9A shows representative hydroxyl radical cleavage patterns of unbound and I-SceI-bound DNA substrate. The alternation of protected/unprotected areas can be observed on each strand. Optical scans along typical lanes of the gel are given in Fig. 9B. The data are schematized in Fig. 9C, where the major sites of diminished and enhanced sensitivity are indicated. Protected sugar residues appear in pairs lying across the minor groove from one another and form three clusters separated by 4–5 base pairs where the cutting efficiency of the backbone deoxyribose is much less affected by the bound protein. In accord with DNase I results, these data thus identify three consecutive loci on the minor groove, located on the same face of the helix, where the protein contacts the DNA backbone. Consistent also with the previous results, the upstream contact spot is very short and the degree of protection weak compared to the catalytic and downstream regions. Furthermore, comparison between lanes (1 to 4) and (5 to 8) and also between lanes (9 to 12) and (13 to 16) shows that the protein protects the bottom strand more than the top strand. In other respects, the binding of the protein results in the striking effect of rendering the substrate relatively more sensitive to hydroxyl radical attack at the nucleotides adjacent to the cleavage sites C\textsubscript{3} and A\textsubscript{4} on the top strand and T\textsubscript{2} on the bottom strand (Fig. 9B). One explanation might be that the reagent is positioned in close proximity to the scissile bonds, resulting in a local source of hydroxyl radicals. In this case, the stabilization would imply that the reagent binds at the active site of the enzyme, which normally chelates Fe\textsuperscript{2+} and DNA. Indeed, it has recently been reported that Fe\textsuperscript{2+} substituted at the active site of two archaeal intron-encoded homing endonucleases yields functional enzymes (36). Nevertheless, in our experiments, all the components of the Fenton reaction had to be present for the effect to be observed (not shown), making it unlikely that cleavage arises from the enzymatic activity of the protein coordinated to free ferrous ions. In accord with the alteration of the structure of the catalytic region upon protein binding, the accessibility of the targeted bonds may rather result from DNA bending in a direction, which compresses the major groove and induces minor groove expansion (37, 38, 39). Note also, consistent with the local minor groove binding, the accessibility of the bonds 14–15 on the top strand and 10–11 on the bottom strand in a stretch of more or less protected sugar-phosphodiester bonds (Figs. 9, B and C).

**DISCUSSION**

In the cellular context, the endonuclease I-SceI discriminates its target site among −10\textsuperscript{7} bp (40). Present results establish
that, as with other proteins of the same family, the absence of
divalent metal ions eliminates cleavage but not sequence-spe-
cific DNA binding. Consistent with the general tendency of the
two-motif LAGLIDADG homing endonucleases (4), I-
SceI was found to bind to its substrate in monomeric form. Furthermore,
the difference in the length of the downstream and upstream
exonic sequences involved in protein-DNA binding as well as
the position of the footprinting protection and cleavage maxima
clearly reveal that the downstream part of the recognition
sequence is primarily involved in I-
SceI binding. Experimental
results provide also the evidence that I-
SceI binding is accom-
panied by DNA distortion. In the model of the complex that
arises from the present findings (summarized in Fig. 10 and
11), the binding of the protein appears to distort its bound
substrate to widen the minor groove at the cleavage site and
make the scissile phosphates accessible to the enzyme active

site. A sequential binding-kinking model is suggested in which
the first step of the protein binding would be facilitated by the
helical features of the sequence located at one helical turn
downstream from the intron insertion site. The high tendency
of this region to unwind is apparent from the OP_{2}Cu^{+} hyper-
sensitivity at positions 10 to 15 and from the strikingly high
sensitivity of position G7 to DMS methylation (Fig. 8), observed
to characterize guanines positioned in open DNA regions (41).
This intrinsic unwinding must facilitate protein binding deep
in the minor groove resulting in kinking the double helix at the
base step A9-A10, 5' to the protein side chain minor groove
intercalation (42) and easily propagating helical distortion. The
induced helical distortion would position the sugar-phosphate
backbone of residues 2 and 5 on the top strand and 1 to 3 on
the bottom strand in register to be contacted by the protein
from the outside of the minor groove. This would result in the

---

**Fig. 9. Hydroxyl radical footprinting of I-SceI on its DNA homing site.**

A, hydroxyl radical attack of the 54-bp DNA fragment \(^{32}\)P-end-labeled uniquely
on the top strand (lanes 1–8) or on the bottom strand (lanes 9–16), either naked
(lanes 1–4 and 9–12) or protein-bound (lanes 5–8 and 13–16). Samples were sub-
mitted to hydroxyl radical cleavage at \(4^\circ\)C for 2 min (lanes 1 and 5, 9 and 13) or
5 min (lanes 2 and 6, 10 and 14) and at \(37^\circ\)C for 1 min (lanes 3 and 7, 11 and 15)
or 3 min (lanes 4 and 8, 12 and 16). Numbers refer to the position of the base with
respect to the intron insertion site, up-
stream from (−1) and downstream from
(+1). Bands were assigned by reference to
an A+G Maxam-Gilbert marker track
(not shown). B, densitometer scans of the
hydroxyl radical footprint along typical
lanes of the autoradiogram. The upper to
lower tracings are respectively those of
lanes 2, 6, 10, and 14. C, schematic repre-
sentation of the relative intensity of indi-
vidual maxima in the presence or absence
of the protein. Relative enhancement of
cleavage is indicated by arrows and pro-
tection by open rectangles.
induction of a new constraint that deforms the helical area encompassing the cleavage sites. Note that the experiments using the variant substrate with the substitution G/C to A/T at positions 17 support this direct relationship. The protein would therefore be positioned closer to the first steps of the next minor groove opening on the same side of the helix, i.e., positions 25, 26 on the top strand and 27 on the bottom strand giving rise in turn to an upstream distortion identified by the Phe-OPCu hypersensitivity at positions 28, 29, and the unstacking of the bases A26 and T27. Thus, by bridging the minor groove opening on the same face of the helix, the protein would induce DNA to curve around it, the major groove being directly accessible to the binding surfaces of the protein on either side of the center of the homing site. The protein would thus stabilize the natural tendency of the helix to bend, predicted by a theoretical calculation using the program proposed by De Santis et al. (43).

**Fig. 10. Summary of the chemical probing data.** Upper, helical accessibility and protection from the inside and the outside of the minor groove. Between the sequences of the two DNA strands: filled and open horizontal rectangles identify, respectively, OP$_2$Cu$^{+}$ hyperreactive and protected areas; the filled vertical rectangle at the intron insertion site represents the intercalated pyridopsonalen molecule. The filled vertical rectangle crossing two filled horizontal rectangles represent the two orthogonal phenanthroline planes (OP) fitting the geometry of the minor groove from positions 1 to 4; one OP is deeply intercalated between base pairs and the other OP is close to the wall of the minor groove of one or the other strand (Schaeffer et al., Ref. 9). Arrows indicate Phe OP$_2$Cu$^{+}$ hypersensitive sites. On each side of the strands, the first lane indicates hydroxyl radical protection (open rectangles) and enhancement (arrows), and the second lane represents the influence of protein binding on the DNase I cleavage frequency (same symbols as in Fig. 2). Lower, base residue modification and interference. The adenines and thymines reactive, respectively, to DEPC and KMnO$_4$ are indicated by black-filled squares (strongly reactive), gray-filled squares (moderately reactive), and open squares (weakly reactive). The guanines protected against DMS methylation are indicated by black-filled circles (strongly protected), gray-filled circles (moderately protected), and open circles (weakly protected). Black arrows indicate the methylated guanines responsible for strong binding interference.

**Fig. 11. Superimposition of the chemical probing data on the helical representation of the I-SceI DNA homing site (from −13 to +17).** Blue ribbon areas indicate the regions of the backbone (green ribbon) protected from hydroxyl radical attack, and blue dashes indicate the minor groove area protected from OP$_2$Cu$^{+}$ cleavage. Red rectangles show the adenines and thymines that strongly (filled rectangles), moderately (hatched rectangles), or weakly (blank rectangles) react with DEPC and KMnO$_4$, respectively. Red dashes indicate the binding domain of the tetrahedral coordination complex OP$_2$Cu$^{+}$; red arrows show the phosphodiester bonds hypersensitive to OP$_2$Cu$^{+}$ attack; and red asterisks the phosphodiester bonds hypersensitive to hydroxyl radical attack. Black arrows show the phosphodiester bonds hypersensitive to Phe OP$_2$Cu$^{+}$ attack. The filled vertical rectangle identifies the highly favored pyridopsonalen intercalation base step IS indicates the intron insertion site. Bases are numbered from −1 and +1 extending, respectively, upstream (on the left) and downstream (on the right) from this site. The model was constructed using the program Insight II (Molecular Simulations, version 98.0).
to the helix with a maximum distortion angle at the step \(\Delta G_6\).2

The stabilization of a distorted DNA double helix appears to be a common requirement for the homing endonucleases. From recent reports of the high-resolution crystal structures of PI-SceI (44), I-DmoI (45), and I-CreI (46) and of the proteins I-CreI or I-Ppol complexed to their DNA target (47, 48), it appears that these relatively small homing endonucleases utilize the same principle to recognize and cleave their long DNA targets. They form extended folds that allow them to form long interfaces across lengthy DNA homing sites and display preformed binding motifs, consisting of antiparallel \(\beta\)-ribbons, making extended contacts with the DNA. However, the different conserved motifs that characterize each family give each of them a specific interface structure. In particular, the proteins of the LAGLIDADG family share a domain fold characterized by the topology \(\alpha\beta\beta\)\(\beta\alpha\beta\beta\). This gives rise to an unusual \(\beta\)-ribbon helical interface whose architecture displays an extensive curvature complementary to the DNA major groove in the cleavage sites region and further facilitates the recognition of extended DNA sequences (49). Nevertheless, the proteins of the LAGLIDADG family differ greatly in the relative shapes and sizes of their DNA binding surfaces (45).

Chemical reactions sensitive to variations in helical conformation have proved invaluable in revealing structural features of the free DNA recognition sequence of the I-SceI protein. Here, using a similar approach to study the DNA-I-SceI protein complex, we established that the combination of such probes provides a valuable means for extracting structural information about the structure of DNA in the complex and relating it to intrinsic conformational features of the double helix. The DNA sequence appears to show the protein the way by which to position the catalytic machinery in proximity to the two closely opposed scissile phosphates for nucleolytic attack. A I-SceI/DNA cross-linking study is in progress to identify protein-DNA contacts to further understand the molecular basis of such a mechanism.

Acknowledgments—We thank O. Mauffret and S. Fernandjian for access to their Molecular Simulations program and B. Dujon, H. Buc, and A. Travars for critical reading of the manuscript.

REFERENCES

1. Dujon, B., Gottard, G., Colleaux, L., Betermier, M., Jacquier, A., d'Auriol, L., and Gallibert, F. (1985) in \textit{Achievement and Perspective of Mitochondrial Research II} (Quagliariello, E., Slater, E. C., Palnieri, F., Saccone, C., and Kroom, A. M., eds) pp. 215–225, Elsevier Science, Amsterdam.
2. Colleaux, L., D'Auriol, L., Betermier, M., Cottarel, G., Jacquier, A., Gallibert, F., and Dujon, B. (1986) \textit{Cell} 64, 521–533
3. Mueller, J. E., Bryk, M., Loizos, N., and Belfort, M. (1999) in \textit{Nucleic Acids (Linn, S. M., Lloyd, R. S., and Roberts R. J., eds.)}, pp. 111–143, 2nd Ed., Cold Spring Harbor Laboratory Press, Cold Spring Harbor, NY.
4. Belfort, M., and Roberts, R. (1997) \textit{Nucleic Acids Res.} 25, 3379–3388
5. Montheetlet, C., Perrin, A., Thierry, A., Colleaux, L., and Dujon, B. (1990)

\(\Delta G_6\) F. Schaeffer, personal communication.