High resolution crystal structure of domain I of the *Saccharomyces cerevisiae* homing endonuclease PI-SceI

Erik Werner¹, Wolfgang Wende², Alfred Pingoud² and Udo Heinemann¹,³,*

¹Crystallography Group, Max Delbrück Center for Molecular Medicine, Robert-Rössle-Straße 10, 13092 Berlin, Germany, ²Institute of Biochemistry, Justus Liebig University, Heinrich-Buff-Ring 58, 35392 Giessen, Germany and ³Institute of Chemistry–Crystallography, Free University of Berlin, Takustraße 6, 14195 Berlin, Germany

Received June 14, 2002; Revised and Accepted July 23, 2002 PDB no. 1gpp

**ABSTRACT**

The homing endonuclease PI-SceI from *Saccharomyces cerevisiae* consists of two domains. The protein splicing domain I catalyzes the excision of the mature endonuclease (intein) from a precursor protein and the religation of the flanking amino acid sequences (exteins) to a functional protein. Furthermore, domain I is involved in binding and recognition of the specific DNA substrate. Domain II of PI-SceI, the endonuclease domain, which is structurally homologous to other homing endonucleases from the LAGLIDADG family, harbors the endonucleolytic center of PI-SceI, which *in vivo* initiates the homing process by introducing a double-strand cut in the ~35 bp recognition sequence. At 1.35 Å resolution, the crystal structure of PI-SceI domain I provides a detailed view of the part of the protein that is responsible for tight and specific DNA binding. A geometry-based docking of the 75° bent recognition sequence to the full-length protein implies a conformational change or hinge movement of a subdomain of domain I, the tongs part, that is predicted to reach into the major groove near base pairs +16 to +18.

**INTRODUCTION**

PI-SceI from *Saccharomyces cerevisiae* belongs to the group of intein-encoded homing endonucleases (for reviews see 1–4). The enzyme is a two domain protein of 454 amino acids that is capable of catalyzing a protein splicing reaction and the specific cleavage of substrate DNA (5,6). Domain I, the protein splicing domain, catalyzes the excision of the homing endonuclease from the primary translation product of the VMAI gene to yield PI-SceI (the intein) and a subunit of the vacuolar membrane H⁺-ATPase, VMA (the extein). Domain II, the endonuclease domain, cleaves the DNA at a specific site within the VMAI locus that is deficient in the vde gene. The vde gene is subsequently inserted by a double-strand break repair process known as ‘homing’ (4,7).

The crystal structure of PI-SceI was first solved by Duan *et al.* (PDB code 1vde) (8) and in a different space group by the same group (PDB code 1dfa) (9). To understand the splicing reaction, Poland *et al.* (10) (PDB code 1ef0) and Mizutani *et al.* (11) (PDB code 1JVA) solved the structure with additional extein residues on both termini. However, in no case was structure determination of the protein–DNA complex successful.

DNA binding of PI-SceI shows many interesting features. With ≥31 bp, the specific recognition sequence is one of the longest known. In addition, it was found that most of the DNA binding energy comes from interactions with domain I, the splicing domain, whereas domain II contains the active center of the endonuclease, but binds the DNA only weakly and non-specifically as an isolated domain (12–14). To cover the whole protein, the bound DNA needs to bend and this is thought to occur in a two-step process. It is believed that in the initial complex, the ‘lower complex’ (this name has been derived from gel retardation assays), the recognition sequence from base pair +5 to +21 binds to domain I tightly (15). The DNA has a bend of 40–45°, which might reinforce an intrinsic bend. In the ‘upper complex’, the rest of the DNA (extending to base pair −10) binds to domain II and undergoes an additional bending to result in a total curvature of 60–75° (12,14). In this way, the cleavage sites between base pairs −3 and −2 of the bottom strand and between +2 and +3 of the top strand (see Fig. 7) come close to the endonuclease active center of domain II. Asp326, Thr341 and Lys301 are the key residues of the catalytic center that cleaves the bottom strand, Asp218, Asp229 and Lys403 are part of the catalytic center which cleaves the top strand; cleavage leads to a 4 nt single-strand overlap (12,13,16,17). The active site residues Asp218 and Asp326 are within two conserved LAGLIDADG motifs that are in one domain in the case of PI-SceI (16) but can be aligned with an r.m.s.d. of 1.53 Å (with 324 Cα atoms involved, calculated with Swiss-Pdb Viewer; 18) to the endonuclease I-Crel, where two domains contain a single LAGLIDADG motif each and combine via a two-fold symmetry to yield the same DNA-binding motif (19). The crystal structure of the I-Crel–DNA complex (PDB code 1bp7; 20) suggests how...
domain II of PI-SceI may interact with DNA. However, it does not explain tight DNA binding of domain I of PI-SceI, where no structural data of a DNA complex are available, neither for PI-SceI itself nor a structural homolog like the intron-encoded homing endonuclease PI-PfI (21).

Many biochemical data are known which allow prediction of how DNA and PI-SceI interact, both on the protein and the DNA side. Methylation interference assays show where the protein contacts nucleobases, and ethylation interference assays where the protein contacts the DNA backbone (14). Hydroxyl radical protection and DNase I cleavage experiments show where the protein covers the recognition sequence (16,17) and photo-crosslinking experiments have identified direct contacts between amino acids and specific sites on the DNA (9,22,23). In addition, mutants of PI-SceI and mutations in the recognition sequence have helped to determine which residues and bases are important for binding and cleavage activity (13,14,17,24). Based on these data and electrostatic potential calculations several (and consistent) DNA binding models were proposed for PI-SceI (8,9,22,23).

However, most of these data concern domain II and the endonuclease activity, and relatively few data exist on DNA binding of domain I. In this work, the crystal structure of PI-SceI domain I is presented at the high maximal resolution of 1.35 Å. The initial experiment was set up to solve the co-crystal structure of the domain I–DNA complex, and many different DNA fragments covering the recognition sequence (ranging from 13mers to 24mers) were tried. Despite proven DNA binding in solution, in no case was a co-crystal observed; instead domain I crystallized very easily even with a molar excess of DNA present. The high resolution data allow a detailed view of the structure of domain I. A DNA binding model was constructed on the basis of experimentally identified contacts assuming that a flexible part of domain I undergoes large conformational changes to support strong DNA binding.

MATERIALS AND METHODS
Cloning, expression and purification
The cDNA of *S. cerevisiae* homing endonuclease PI-SceI domain I was cloned as described previously (15), resulting in the plasmid pHsPI-SceI-DI. The encoded protein contains 10 additional N-terminal amino acids including a His<sub>6</sub> tag and the residues 1–182 and 411–453 that account for domain I of the full-length protein according to the crystal structure (8). Compared with the protein sequence deposited in the SWISS-PROT database (25) (PDB 1P72S5), residue Asn454 was removed to avoid protein splicing of the His<sub>6</sub> tag and all of domain II was replaced by a single residue, Gly183. In addition, three naturally occurring variations, Arg44Ser, Val67Met and Ile132Val, are present in our construct, which are also found in the PI-SceI sequence of the DH1-1A wild yeast strain (26). The plasmid was used to transform *Escherichia coli* BL21(DE3) (Stratagene) cells which were grown in LB medium. Overexpression of the gene was induced with 1 mM IPTG. Cells were lysed in a French pressure cell and PI-SceI-DI was purified by Ni–NTA chromatography according to the manufacturer’s protocol (Qiagen) followed by gel filtration on a Superose 75 column (Pharmacia). PI-SceI-DI was stored in 100 mM ammonium acetate, 50 mM sodium acetate, pH 4.7.

Crystallography, data collection and processing
Domain I of PI-SceI crystallized under a number of PEG conditions present in the Hampton Crystalization Screen (27). The crystal used for structure determination came from a crystallization attempt originally set up to produce protein–DNA co-crystals with an 18mer (the minimal binding region, base pairs +5 to +22, of the specific recognition sequence of PI-SceI). The complex solution contained a 1.05 molar excess of DNA and the final composition of the buffer was 54 mM ammonium acetate, 27 mM sodium acetate pH 4.7, 38 mM Tris–HCl pH 7.5 and 188 mM MgCl<sub>2</sub>. Crystals of size 0.4 × 0.1 × 0.1 mm were obtained by the hanging drop vapor diffusion method 2 days after mixing 2 μl of protein/DNA solution (0.35 mM) with 1 μl of reservoir solution, containing 30% PEG 4000, 0.1 M sodium citrate pH 5.6 and 0.2 M ammonium acetate at room temperature. A high resolution data set with a maximal resolution of 1.35 Å was collected at beamline BW7B at DESY, Hamburg, as well as two low resolution data sets with maximal resolutions of 2.2 and 3.0 Å. Data were processed and merged by DENZO and SCALEPACK (28).

Structure determination and refinement
Phases were obtained by the molecular replacement method using the program AMoRe (29) with the coordinates of 1vde (8). The program ARP (30) was used for an automated refinement procedure and for tracing, model building, automated side chain matching and for a water search. Further model building and refinement were done with the programs ONO (31) and REFMAC5 (32), respectively. Individual anisotropic B factors were refined and 303 solvent molecules were fitted into the electron density.

RESULTS AND DISCUSSION
Structure determination and validation
Out of 237 amino acids of the PI-SceI domain I construct (including the His<sub>6</sub> tag) 202 were placed correctly by the program ARP (30), yielding an average confidence level of 99%. Only 15 more residues had to be positioned manually; the missing 20 residues (residues –10 to –3 of the His<sub>6</sub> tag and loop Gln55–Glu66) were not seen in the electron density due to positional disorder. 181 (94.8%) of 191 non-glycine and non-proline residues are in the most favored region of the Ramachandran plot, the other 10 (5.2%) being in additional allowed regions. The structure was refined to final R<sub>cryst</sub> and R<sub>free</sub> values of 15.0 and 18.9%, respectively. Data collection and refinement statistics are given in Table 1.

Overall structure of PI-SceI-DI
PI-SceI-DI is a 27 kDa protein that consists mainly of β-strands (51% of the observed sequence) and to a smaller extent of helices (14%). Figure 1A gives a cartoon representation of the molecule and introduces the numbering of the secondary structural elements. A total of 17 strands form six antiparallel sheets. Of them, strands a, d, e, f, g and o combine to form a twisted barrel. Of the six helices, two are α-helices
Table 1. Data collection and refinement statistics

<table>
<thead>
<tr>
<th>Data collection</th>
<th>20–1.35 (1.37–1.35)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Resolution (Å)</td>
<td></td>
</tr>
<tr>
<td>Number of observations</td>
<td>359 618</td>
</tr>
<tr>
<td>Unique reflections</td>
<td>51 602 (1141)</td>
</tr>
<tr>
<td>Data completeness (%)</td>
<td>94.1 (64.1)</td>
</tr>
<tr>
<td>Average I/σ(I)</td>
<td>34.5 (2.5)</td>
</tr>
<tr>
<td>Rmerge (%)</td>
<td>2.9 (16.4)</td>
</tr>
<tr>
<td>R1 impe (%)</td>
<td>4.1 (20.8)</td>
</tr>
<tr>
<td>R1 full (%)</td>
<td>2.7 (13.4)</td>
</tr>
<tr>
<td>Space group</td>
<td>C2</td>
</tr>
<tr>
<td>Unit cell parameters</td>
<td></td>
</tr>
<tr>
<td>a = 102.9 Å</td>
<td></td>
</tr>
<tr>
<td>b = 47.6 Å</td>
<td></td>
</tr>
<tr>
<td>c = 60.2 Å</td>
<td></td>
</tr>
<tr>
<td>β = 121.4°</td>
<td></td>
</tr>
<tr>
<td>Monomers in the asymmetric unit</td>
<td>1</td>
</tr>
</tbody>
</table>

Structure determination method: MR (1vde)

Protein atoms (non-hydrogens): 1758

R<sub>crysd</sub> (%) = 15.0

R<sub>free</sub> (%) = 18.9

r.m.s.d. bond lengths (Å) = 0.018

r.m.s.d. bond angles (°) = 1.83

Average B factor<sub>1</sub>, all protein atoms (Å<sup>2</sup>) = 23.3

Average r.m.s.d. for main chain B factors<sub>1</sub> (Å<sup>2</sup>) = 0.83

Average r.m.s.d. for side chain B factors<sub>1</sub> (Å<sup>2</sup>) = 2.33

Solvent molecules = 303

Average B factor<sub>1</sub>, all solvent atoms (Å<sup>2</sup>) = 37.1

Values in parentheses refer to the outer resolution shell.

**Comparison of all five PI-ScelI domain I structures**

The overall structure of domain I is in general agreement with the four structures of domain I already published as part of the full-length protein: 1vde (8), 1dfa (9), 1ef0 (10) and 1jva (11). All of these were determined at a resolution of 2 Å or lower. Thus, the 1.35 Å structure presented here provides by far the most detailed view of domain I. For this reason, modeling of DNA binding to PI-ScelI-DI (see below) was based on a fragment of *Drosophila melanogaster* (13% sequence identity, PDB code 1a0f; 36). These three proteins also give the only hits in a DALI search (37) with Z scores above 3 (Z scores of 13.7 for 1am2, 13.6 for 1a0f and 11.8 for 1dq3-A). However, only the core subdomain is covered by these alignments. A DALI search with only the tongs-like subdomain gives some minor hits (highest Z score 4.4) but none of the proteins found show the tongs-like element and only align with helix C and directly adjacent parts of strands i, j and k (data not shown). A similar result was obtained in the PI-PfuI structure analysis. In addition to the common core domain, there is a stirrup domain which yields no structural homologs in a DALI search (21). In contrast to PI-ScelI, this stirrup domain of PI-PfuI is inserted between the protein splicing and the endonuclease domains, whereas the tongs-like structure of PI-ScelI is inserted in an internal loop within the Hint domain.
consensus model for the full-length protein based on this crystal structure.

Some differences between the various domain I models are observed, in part due to different protein constructs and crystal forms. Differences in secondary structure are marginal. The barrel mentioned above only closes because the additional N-terminal residues in our construct form an additional short β-sheet, which is also observed in 1jva. Equivalent sheets in 1vde are sheets 2–4 (strands β3–β7 and β24–β26). However, the additional β-sheet is also seen in 1ef0 chain B, where short parts of the original extein boundaries are present, as well as mutations of the terminal intein residues. However, a barrel is not recognized by the program PROMOTIF (38). Another small difference is that sheet 5 (strands β8–β12, β13 and β23) of 1vde appears to be split into two smaller sheets here (sheet 3 consisting of strands h and l and sheet 5 consisting of strands m and n). The additional helix F forms in 1vde where the connecting residues to domain II are and is therefore due to the protein construct.

Interestingly, in none of the structures are all residues visible in the electron density, with the disordered residues being in different locations of the protein. Here, as well as in 1jva, residues Gln55–Glu66 are disordered, a loop that is well structured in 1vde. On the other hand, 1gpp (this structure) contains residues Ser93–Phe102 and Gly135–Asn151 which are absent from both 1vde and 1dfa. 1ef0, in addition, has one monomer in which the loop Arg91–Glu103 is not visible and one monomer in which residues Gln55–Leu70 and Thr95–Gly98 are not observed. Figure 2 shows a structural alignment of domain I from all five structures. Some regions show large displacements, most pronounced at the borders of the loops missing in this structure (residues 53, 54 and 67–70) as well as those missing in other structures (residues 91–102, 130–138 and 149–153) and of course residues 181–415, where in the structure presented here Gly183 replaces the entire domain II (Fig. 3A and C). This pattern is in general consistent with regions of high β factor and obviously represents the flexible loops of the structure.

Two regions showing further small differences are residues 30–32 and 438–441. Both are close to each other and form a contact to helix C of the symmetry-related molecule (x, y, z) in 1gpp, whereas the first region has no crystal contact in any other structure and the latter contacts the loop of symmetry-related molecules of 1vde and 1dfa that is missing in 1gpp (around residue 55), domain II in 1jva and has no crystal contacts in 1ef0. Crystal contacts play a role for all structures (Fig. 3B) and are most pronounced for 1gpp, where 96 of 217 (44%) residues are involved in lattice contacts (Fig. 4) (domain I only for 1vde, 49 of 217; 1dfa, 82 of 210; 1ef0_A, 48 of 217; 1jva_A, 38 of 223). This accounts for ~3200 Å² of solvent-accessible surface that is covered by symmetry-related molecules (27% of the total accessible surface; calculated with NACCESS; 39).

Protein splicing site

The active center of PI-SceI, where protein splicing takes place, is only partially visible in the structure. The C-terminal extein is completely missing as well as the C-terminal residue (Asn454) of the intein. The N-terminal intein is complete (with Cys1) and a part of the purification tag appears to mimic the N-terminal extein, but with changed sequence (Ala–1 and Ser–2 instead of Gly–1 and Val–2 in the natural sequence). The structure of the splice site is best compared with the structures presented by Poland et al. (1ef0; 10) and Mizutani et al. (1jva; 11), where extein residues on both sides were observed. In 1ef0, residues 1 and 454 are mutated to alanine, in 1jva to serine to prevent the splicing reaction. Additionally, 1jva contains the mutations His79→Asn and Cys455→Ser. In 1ef0, a Zn²⁺ ion is coordinated by His453, Cys455, Glu80 and Wat53. This is not found here (see the electron density of this site in Fig. 5A) due to the absence of residues 454 and 455. In fact the side chains of His453 and Glu80 in 1gpp are twisted compared with 1ef0, which is unique for all structures in the case of His453. However, the metal ion was described to be most likely of structural and not of functional importance.

Positioning of the exteins differs substantially but is not important for the conformation of the inteins that are structurally almost identical (Figs 2 and 5B). Residue –1 (Ala in 1gpp, Gly in 1ef0) is part of the first β-strand and forms typical β-sheet hydrogen bonds to residue Ile434 (strand o). In 1jva, on the other hand, Gly–1 and therefore the N-terminal extein is oriented towards the opposite side, fixed by a hydrogen bond between Ser1-N and Asp76-N82. At the same time, Gly–1-O forms a hydrogen bond with Ser455-Oγ (Cys455 in the natural sequence). This means that the heteroatom of residue 455, the first C-terminal residue, is already in close approximation to Gly–1-C, the site of nucleophilic attack. In 1ef0_B, on the other hand, Cys455-Sγ
is coordinated by Zn$^{2+}$ and more than 7 Å away from the reaction center. Therefore, an inhibitory effect of Zn is likely, as already described for RecA (40). Nevertheless, the first step of the protein splicing mechanism, the N-S acyl shift as described by Noren et al. (41), can be structurally understood and confirmed. In this current model, Cys1-S$\gamma$ initiates the rearrangement by a nucleophilic attack on the carbonyl carbon of residue $\pm1$. Cys1-N, the N-terminus of the resulting intein, is stabilized by N$\delta1$ of the conserved His79 by a hydrogen bond.

**DNA binding**

How DNA is bound by PI-SceI is largely unclear, since no structure of a complex is available yet. Although PI-SceI and its domain I alone bind tightly to the recognition sequence, as seen in gel retardation assays (15,16) and filter binding experiments (W.Wende, unpublished data), crystal formation of the protein alone seems to be greatly favored over complex crystallization. This was observed not only in this study but also by Hu et al. (9). The observation that PI-SceI loses its endonucleolytic activity but preserves the specific binding capacity after denaturation/renaturation cycles or trypic cleavage at position Arg277 (42) and that the isolated domains neither form a stable complex with each other nor show a supershift with DNA (15) is a hint that the arrangement of the domains with respect to each other might play an important role in the DNA binding and catalytic process. However, the protein alone seems very stable and the energy required for a
conformational rearrangement (and possibly the loss of important contacts) of the domains must be compensated for by tight DNA binding.

Detailed biochemical data concerning PI-SceI activity are available and are summarized in Figure 6. Mutations of protein residues and DNA base pairs (13,14,17,24), ethylation and methylation interference assays (14), hydroxyl radical protection experiments (16,17) and photo-crosslinking (9,22,23) were used to explore the activity and the DNA binding of PI-SceI. Taken together with electrostatic considerations (8), the combination of these data allowed the construction of models where the obvious protein-facing side of the recognition sequence (the ‘left’ side in the orientation of Fig. 6C and D) was docked to the protein. Three of these models (9,22,23) are consistent with each other and the biochemical data and revise the first model proposed by Duan et al. (8). Bending of the DNA, which is required to fit the complete protein (the sizes of the protein and DNA in Fig. 6 are proportional), was considered. Gel retardation assays with circularly permuted binding fragments (12) showed a bend of 45° ± 15° for the lower complex (DNA binding to domain I alone) and of 75° ± 15° for the upper complex (DNA binding to the whole protein), both centered at base pair +7 (± 1).

A view of the electrostatic surface potential of domain I (Fig. 7A) supports the actual model. Positive charges, potential binders for the negatively charged DNA backbone, are located at the tongs-like part of domain I. It is very likely that helices D and E and strand l also belong to this subdomain, since they contain some of the residues thought to be important for DNA binding: Lys53, Lys173, Lys112, Lys124, Arg90 and Arg94 (see Fig. 6). No mutants or contacts are known for Lys113, His170 and Arg147 that are also potential DNA binders. No change in activity was observed for the mutant Lys97→Ala (13,43). The orientation of Lys97 in this structure is strongly influenced by crystal contacts and might have no meaning for DNA binding. The alignment of all residues with influence on the activity of PI-SceI in domain I (Fig. 7B) shows two classes of residues. The first, consisting of Tyr176, Tyr420, Arg172 and Arg162 but also Lys112 and Lys124, shows very small structural differences and could therefore be responsible for the initial positioning of the DNA. Residues of the second class, that differ structurally between the single structures, might move towards the ligand upon binding. Those latter residues are in the loop 55–66 and include arginines 87, 90 and 94 and tyrosines 130 and 173 (see also Fig. 3C).

At this point, we must remind ourselves that proteins were always treated as rigid objects in the above considerations and that the number of proven protein–DNA contacts is small. In fact, only four amino acids of PI-SceI have proven and close contacts to known sites in the recognition sequence, all of which are in domain II. These are the active site residues Asp218 and Asp326 and the residues Tyr328 and His333 with zero length crosslinks to G4 (top strand) and T9 (bottom strand), respectively. Three more crosslinks are established...
between Tyr384 and P of T7 (bottom strand), between Tyr420/Tyr176 and P of G13 (top strand) and between Lys112/Lys124 and P of C18 (bottom strand) at a maximal distance of ~11 Å, the length of the crosslinker p-azidophenacylphosphorothioate (44).

For a more detailed DNA binding model which considers a possible domain movement we used the following approach. A full-length consensus structure of PI-SceI was derived by alignment of 1vde_A with 1gpp (r.m.s.d. 2.0 Å for Cα, 2.6 Å for all atoms, calculated with the program LSQKAB; 32). The structure has residues 1–53, 73–179 and 418–453 from 1gpp as well as residues 54–72 (containing the missing loop from 1vde_A, residues 54–72 in red, missing loop (from 1vde chain A, residues 54–72) in purple and domain II (from 1vde chain A, residues 180–417) in blue. Amino acid side chains of important residues are shown in ball-and-stick representation, with yellow bonds if a mutant exists (see A) and with green bonds if photo-crosslink data are available (see C). Missing loops of domain II (residues 271–279 and 369–274) are shown as thin lines. (C) Summary of crosslink data, recognition sequence part. The top strand is in green, the bottom strand in magenta (also in D and F). Green spheres represent phosphates where photo-crosslinks via p-azidophenacylphosphorothioate were established; nucleobases with thick bonds represent those sites where photo-crosslinks of 5-iodopyrimidines were observed. For the cases where the crosslinked amino acid is known, this is represented by arrows (solid line, zero length crosslink; dotted line, non-zero length crosslink; dashed line, FeEDTA cleavage). (D) Summary of methylation and ethylation interference assays and of hydroxyl radical protection assays. Spheres (large, strong interference; medium sized, weak interference) represent the atoms where interference by a methyl group (blue, methylation of Gua-N7 or Ade-N3) or an ethyl group (green, phosphates) with DNA binding was found. Dotted surfaces (light blue, weak protection; dark blue, strong protection; yellow, weak hypersensitivity; orange, strong hypersensitivity) represent the results from hydroxyl radical protection assays. (E) Mutations in the PI-SceI recognition sequence. Red, loss-of-cleavage mutations; blue, loss-of-binding mutations; green, hypercleavable mutations. (F) Recognition sequence of PI-SceI with the cleavage site (blue line) and numbering of the base pairs (centered at the middle of the cleavage site). Base pairs with an orange background represent the minimal binding region (base pairs +5 to +21), nucleobases with a yellow background are protected from DNase I cleavage and the base pairs in the black box are considered the minimal sequence required for cleavage (base pairs −10 to +21).
domain II proven to be important; see Fig. 6A and B) are consistent with the biochemical data, the model might be valid to some extent for domain II but requires a conformational change of at least the tongs subdomain or parts thereof. This is shown in Figure 8B, where a possible movement of this subdomain of ~60° (and an additional one towards the back of the viewer’s plane) is outlined in the ‘hinge’ view. This movement would bring β-strands i, j and k of the tongs subdomain to the major groove at base pairs 16–18 and Lys112 from a distance of 35 Å to crosstalk distance of the phosphate of C18. Other consequences of the movement would be the disruption of the contacts between strands j/k and helix D on the one side (boundary of the core and tongs subdomains of domain I) and on the other side new possible contacts between helix C and the loop of residues 409–413 and those of residues 150–152 with loops 30–31 and 439–441. The arginines 90 and 94 that prevent binding when mutated to alanine would come close to top strand nucleotides 14/15 and 16, respectively. Lysines 170 and 173 come close to nucleotides 10–12 of the top strand. The movement of the entire rigid tongs domain visualized in Figure 8B can be achieved using only a few degrees of freedom. Obviously, more subtle and complex conformational rearrangements upon DNA binding that involve only parts of the tongs subdomain cannot be excluded.

Another way to fit the biochemical data would be a twist of the DNA in addition to the bend that brings the protein-facing DNA part close to the residues with proven importance for DNA binding. One also has to take into account the possible movement of amino acid side chains (between 2.4 Å for Asp and 6.0 Å for Arg, measured from Cβ to the tip) and the 11 Å croslinker p-azidophenacylphosphorothioate. It must be emphasized that the presented model is based on crystallographic and biochemical data and includes hypothetical movements of parts of the protein relative to each other, which are not yet proven and, therefore, require verification by a crystallographic analysis. Up to now, no co-crystals of a PI-Scel-DNA complex are available, neither are comparable complex structures that could help understand DNA binding by PI-Scel. The DNA complex of I-Crel (20) only covers domain II of PI-Scel (and is a good model for this part) and the other homing endonuclease for which DNA complexes are available, I-PpoI (47), is structurally unrelated to PI-Scel. DNA binding is facilitated by three β-strands that reach into the major groove, but this part of the protein does not align with the tongs part of PI-Scel domain I.

**Accession number**

Coordinates of the structure of domain I of the *S. cerevisiae* homing endonuclease PI-Scel have been deposited with the Protein Data Bank (48), PDB code 1gpp.

**ACKNOWLEDGEMENTS**

We are grateful to Prof. Y. Satow (Tokyo) for making coordinates for 1jva available before public release. X-ray data were measured at beamline BW7B of the EMBL outstation at DESY, Hamburg, with kind help from local staff. This work was supported by the Deutsche Forschungsgemeinschaft through He 1318/24-1 and the Fonds der Chemischen Industrie.
REFERENCES


Figure 8. Docking model of PI-SceI (full-length construct) with the minimal sequence required for cleavage (31mer, bend 75°) after alignment of the four proven direct contact sites. Ribbon representations of the backbone and selected amino acid side chains in the ball-and-stick representation are shown. (A) Front view of the crystal structure-based model with domain I in red, domain II in blue, the DNA top strand in green and the bottom strand in magenta. Large spheres represent the sites of contact, with the nucleolytic active sites in red, the zero length crosslink sites in green and the non-zero length crosslink sites in blue. (B) ‘Hinge view’ with the tongs residues (amino acids 86–154) in front of the backbone surface drawing. The original orientation in 1gpp is in orange, the subdomain after the possible movement towards the major groove of the recognition sequence (the rearranged model) in light blue. The large spheres represent the N2 atom of Lys112, which would come to crosslink distance to the phosphate of C18 of the bottom strand by the conformational change of the tongs subdomain. The possible movement of helices D and E and strand l, also belonging to this subdomain, is not shown for the sake of clarity.


