

# Active Site Sharing and Subterminal Hairpin Recognition in a New Class of DNA Transposases

Donald R. Ronning,<sup>1</sup> Catherine Guynet,<sup>2</sup>  
Bao Ton-Hoang,<sup>2</sup> Zhanita N. Perez,<sup>1</sup>  
Rodolfo Ghirlando,<sup>1</sup> Michael Chandler,<sup>2</sup>  
and Fred Dyda<sup>1,\*</sup>

<sup>1</sup>Laboratory of Molecular Biology  
National Institute of Diabetes and Digestive  
and Kidney Diseases

National Institutes of Health  
Bethesda, Maryland 20892

<sup>2</sup>Microbiologie et de Génétique Moléculaire  
Centre National de la Recherche Scientifique  
118 Route de Narbonne  
31062, Toulouse Cedex  
France

## Summary

Many bacteria harbor simple transposable elements termed insertion sequences (IS). In *Helicobacter pylori*, the chimeric IS605 family elements are particularly interesting due to their proximity to genes encoding gastric epithelial invasion factors. Protein sequences of IS605 transposases do not bear the hallmarks of other well-characterized transposases. We have solved the crystal structure of full-length transposase (TnpA) of a representative member, ISHp608. Structurally, TnpA does not resemble any characterized transposase; rather, it is related to rolling circle replication (RCR) proteins. Consistent with RCR, Mg<sup>2+</sup> and a conserved tyrosine, Tyr127, are essential for DNA nicking and the formation of a covalent intermediate between TnpA and DNA. TnpA is dimeric, contains two shared active sites, and binds two DNA stem loops representing the conserved inverted repeats near each end of ISHp608. The cocrystal structure with stem-loop DNA illustrates how this family of transposases specifically recognizes and pairs ends, necessary steps during transposition.

## Introduction

As the number of completed genome sequencing projects rises, it is clear that transposable elements have affected genome evolution, both in bacteria (Chain et al., 2004; Parkhill et al., 2003; Wei et al., 2003) and eukaryotes (Goff et al., 2002; Lander et al., 2001; Waterston et al., 2002; Yu et al., 2002). The insertion of a transposon can have several consequences. If the transposon harbors exogenous genes, these can increase the fitness of the recipient. Well-studied examples include bacterial transposons that transfer antibiotic resistance genes. Alternatively, if the element encodes a transposase, the enzyme that catalyzes the DNA cleavage and joining reactions necessary for mobilization, the protein can be put to an alternate use. For example, at least 47 human genes are believed to have originated from transposase

genes, including the RAG1/RAG2 proteins, which are essential for V(D)J recombination (Lander et al., 2001). Transposons can also cause loss of function if they land in a promoter or coding region of a gene.

Several families of transposases have been identified. They can be broadly divided into two groups: those that form a covalent intermediate with DNA, such as the serine, tyrosine, and rolling circle (or Y2) transposases (Curcio and Derbyshire, 2003); and those that do not. To date, there is no structural information concerning transposases that form covalent intermediates. Among the latter group, the most thoroughly characterized are transposases that contain a conserved triad of acidic residues, the DDE motif, which coordinates two metal ions required for DNA cleavage and strand joining (Kulkosky et al., 1992). The structures of the Mu (Rice and Mizuuchi, 1995) and Tn5 (Davies et al., 2000) DDE transposases, along with those of retroviral integrases (Dyda et al., 1994), have defined the architecture of these enzymes. Only the Tn5 and bacteriophage MuA transposases have been structurally characterized as synaptic complexes with their DNA substrates (Davies et al., 2000; Lovell et al., 2002; Yuan et al., 2005).

The IS605 group of insertion sequences (IS) is an intriguing group of transposable elements. Several are found in *Helicobacter pylori*, a bacterium that induces gastric inflammation often resulting in peptic ulcers and sometimes in gastric cancer (Naumann and Crabtree, 2004). The IS605 elements were initially found associated with the cytotoxic-associated pathogenicity island of *H. pylori*, which expresses a protein secretion system required for stomach epithelial cell invasion (Akopyants et al., 1998; Censini et al., 1996). Protein sequence comparison of the transposases encoded by IS605 elements places them within the IS200 family, which have been identified in over 60 eubacteria and archaea species (Figure 1A; Ton-Hoang et al., 2005).

Although most IS elements encode only a single protein necessary for their transposition, IS605 elements have two open reading frames (orf), *orfA* and *orfB*; these orfs also occur together in bacteria other than *Helicobacter* (Chandler and Mahillon, 2002; Ton-Hoang et al., 2005). The *orfA* product, TnpA, is homologous to IS200 transposases and is essential for ISHp608 transposition (Kersulyte et al., 2002). Although the *orfB* product, TnpB, is homologous to the presumed transposases of IS1341 and IS891, it is dispensable for ISHp608 transposition in *E. coli* (Kersulyte et al., 2002; Ton-Hoang et al., 2005).

Several other features distinguish IS605 elements from other characterized transposable elements. They lack terminal inverted repeats (IR); instead, both inverted and direct repeats are located in subterminal regions. Also, there is no gain or loss of sequences flanking the element typically exhibited by other transposons as a result of the staggered insertion of two transposon ends and subsequent DNA repair. Rather, IS605 elements insert immediately 3' of specific tetra- or pentanucleotide sequences (Figure 1B) (Kersulyte et al., 2002). These characteristics are reminiscent of the Y2

\*Correspondence: fred.dyda@nih.gov

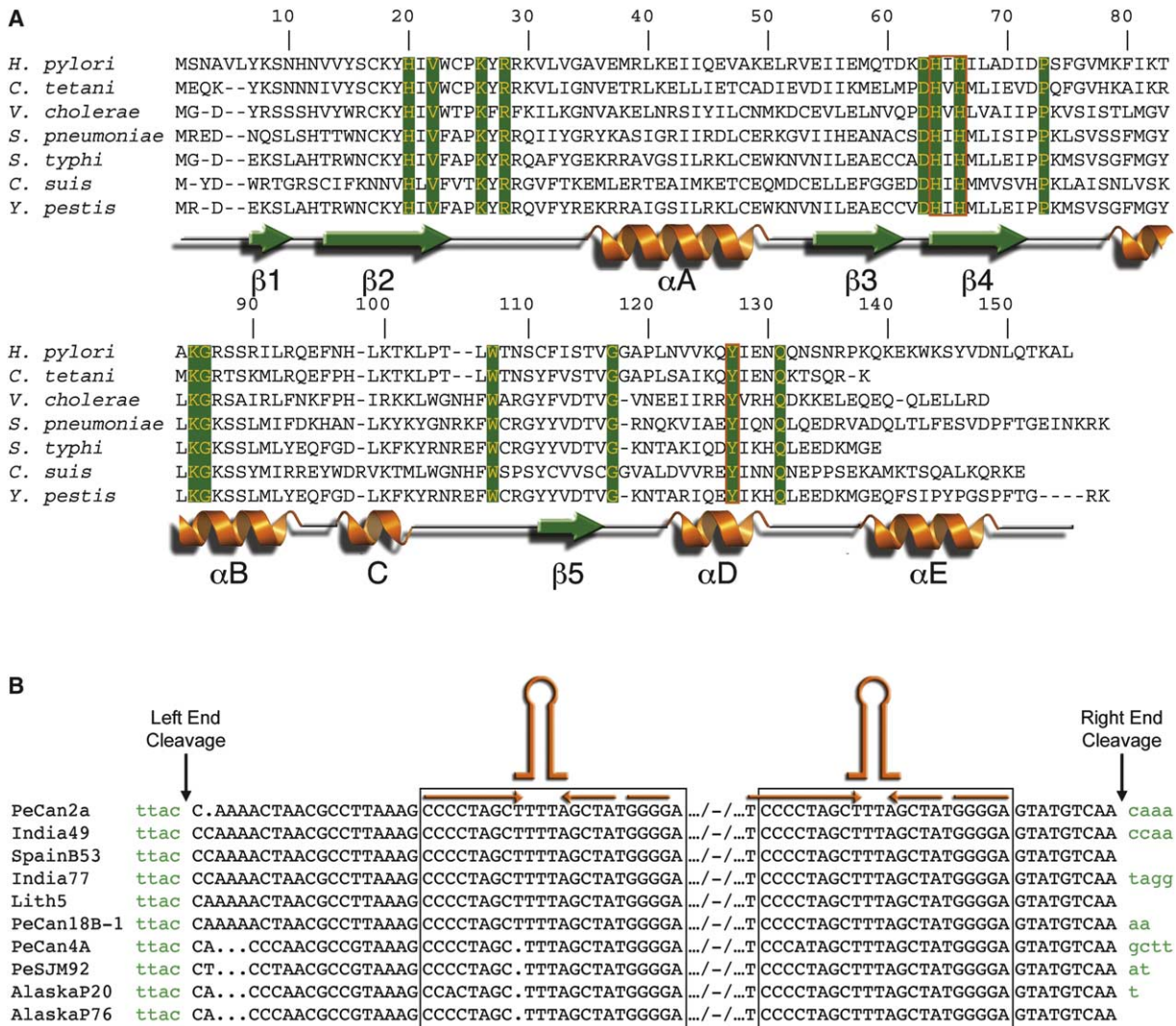


Figure 1. Conserved Protein and DNA Sequences in IS200-like Elements

(A) The top sequence and the numbering represent *ISHp608* TnpA. Other sequences are transposases from IS200 elements in various bacterial species. Conserved residues are in yellow on a green background. The HUH motif and conserved tyrosine residue are boxed in red. Secondary structural elements of TnpA are shown below the sequences.

(B) Alignment (adapted from Kersulyte et al. [2002]) of the termini and flanking sequences of *ISHp608* elements. The *H. pylori* strain is indicated on the left. Flanking sequences are green. IS element sequences are black. Black arrows show DNA cleavage sites. Boxed sequences represent the conserved 22–23 base inverted repeats, and their deduced stem-loop structures are shown in orange.

transposons, of which IS91 is the most thoroughly characterized. IS91 inserts 3' to a specific tetranucleotide sequence (Mendiola and de la Cruz, 1989; Garcillán-Barcia et al., 2001). Mechanistic studies suggest that IS91 initiates a transposition reaction resembling RCR that proceeds through a circular ssDNA intermediate (Garcillán-Barcia et al., 2001; Mendiola et al., 1994). In contrast, *ISHp608* transposition reactions in vivo produce circular dsDNA transposons and precisely recircularized donor plasmid backbones (Ton-Hoang et al., 2005), an observation not easily reconciled by an RCR mechanism.

The IS91 transposase contains two of the three conserved protein sequence motifs identified in RCR proteins, a His-hydrophobic-His (HUH) motif required for metal ion

binding, and a YxxxY motif containing two conserved catalytic tyrosine residues (Ilyina and Koonin, 1992; Koonin and Ilyina, 1993; Mendiola and de la Cruz, 1992). These motifs are essential for IS91 transposition (Garcillán-Barcia et al., 2001). Although *ISHp608* TnpA contains an apparent HUH motif, it has only one conserved tyrosine, Tyr127. Furthermore, with 426 residues, the IS91 transposase is much larger than TnpA, in part because of additional domains. In fact, the 155 residues of *ISHp608* TnpA make it one of the smallest identified transposases.

To understand how a transposase as diminutive as *ISHp608* TnpA can orchestrate transposase end recognition, form a synaptic complex, and catalyze cleavage reactions, we have determined the crystal struc-

Table 1. Diffraction Data and Model Refinement Statistics

	Se Inflection	Se Peak	Remote	Native	TnpA/Stem Loop
Energy (keV)	12.658	12.66	12.8	8.04	8.04
Resolution (Å)	2.4	2.4	2.4	2.4	2.6
Total reflections	50630	46158	143311	45547	48233
Unique reflections	12541	12456	13081	12555	14684
Completeness (%)	99.9	99.9	100	98.7 (99.9)	99.7 (99.7)
I/σI	12.2	12.4	11.1	28.6	14.7
R <sub>sym</sub>	0.081	0.09	0.092	0.043 (0.154)	0.060 (0.348)
<b>Phasing</b>					
Anomalous differences (%)	2.9–4.4				
Dispersive differences (%)	3.3–4.9				
Figure of merit	Solve	0.43			
	DM	0.67			
<b>Refinement</b>					
	TnpA	TnpA/stem loop			
Atoms (N)	2366	3517			
Reflections (N)	12248	14083			
R factor (%)	25.6	20.3			
R <sub>free</sub> (%)	27.4	24.7			
rmsd bonds (Å)	0.008	0.007			
rmsd angles (°)	2.079	1.3			
Average B factor (Å <sup>2</sup> )	55.4	40.6			

R<sub>sym</sub> = Σ|| - <I>|/Σ<I>. R factor = Σ|FP<sub>o</sub> - FP<sub>c</sub>|/|FP<sub>o</sub>|. R<sub>free</sub> is computed using 5% of the total reflections for TnpA and 10% of the total reflections for the TnpA/DNA complex selected randomly and never used in refinement.

tures of full-length TnpA alone and in complex with a 22 base hairpin representing a conserved inverted repeat found near each ISHp608 end.

## Results and Discussion

### Structure Determination and Description

The crystal structure of full-length ISHp608 TnpA was solved by multiwavelength anomalous dispersion (MAD) and subsequently refined with native data to 2.4 Å resolution (Table 1). The asymmetric unit (a.s.u.) contains an intertwined TnpA dimer, consistent with observations by size exclusion chromatography and analytical ultracentrifugation that TnpA is a dimer in solution (see the Supplemental Data available with this article online). The dimer is elongated and flat, with approximate dimensions of 67 × 44 × 26 Å. Each monomer is composed of two subdomains bridged by the long strand β2 (Figure 2A). Subdomain one (residues 18–117) contains a four-stranded antiparallel β sheet (β2–β5) that is flanked on one face by three helices (αA, αB, and helix-C [one turn of a 3<sub>10</sub> helix followed immediately by an α-helical turn]), while the other face is solvent exposed. In subdomain two, formed by the N-terminal 17 residues and the C-terminal 34 residues, two antiparallel β strands (β1 and β2) pack against helix αE, while helix αD does not interact with other secondary structural elements within its own monomer.

Strands β2 and β5 are essential for dimerization as they weave together the two molecules to form what is in essence a large seven-stranded β sheet. Near the C-terminal end of strand β5, the chain crosses over the corresponding strand β5 of the second monomer and continues in the same general direction. The helix that follows this crossover (αD) forms hydrophobic contacts with strands β2 and β5 of the other monomer. As

a result, the extended β sheet is sandwiched between layers of helices with helices αA–αC from both molecules on one face of the sheet and helices αD and αE on the other.

The total buried dimer interface is 1670 Å<sup>2</sup>, of which 92% is contributed by residues forming the central β sheet and helix αD. The remainder arises from the convergence of the N termini of the two αB helices (along with the preceding loop) on one face of the sheet.

### Structurally Related Proteins

ISHp608 TnpA is not similar to any structurally characterized transposase. Nevertheless, results from a DALI search (Holm and Sander, 1993) indicate that TnpA resembles protein structures containing a common organization of secondary structural elements termed the RNA recognition motif (RRM) (Burd and Dreyfuss, 1994). This motif is formed by four antiparallel β strands and two α helices in the order βαββαβ, and is found in numerous disparate proteins.

The best matches to the TnpA structure, judged by both fold and function, are RCR proteins such as viral Rep proteins (Campos-Olivas et al., 2002; Hickman et al., 2002) and the conjugative relaxases, TrwC (Guasch et al., 2003) and Tral (Datta et al., 2003) (Figure 2B). These enzymes bind specific DNA sequences and catalyze site-specific DNA nicking, forming a 5' phosphotyrosine intermediate, and a free 3' OH group from which the DNA strand can be extended.

In common with TnpA, RCR proteins contain an HUH motif and at least one conserved tyrosine residue, and they catalyze the breakage of a phosphodiester bond. However, all structurally characterized RCR proteins are monomeric. The dimeric nature of TnpA is the result of structural differences in regions corresponding to strands β2 and β5 of TnpA, reflecting their pivotal role

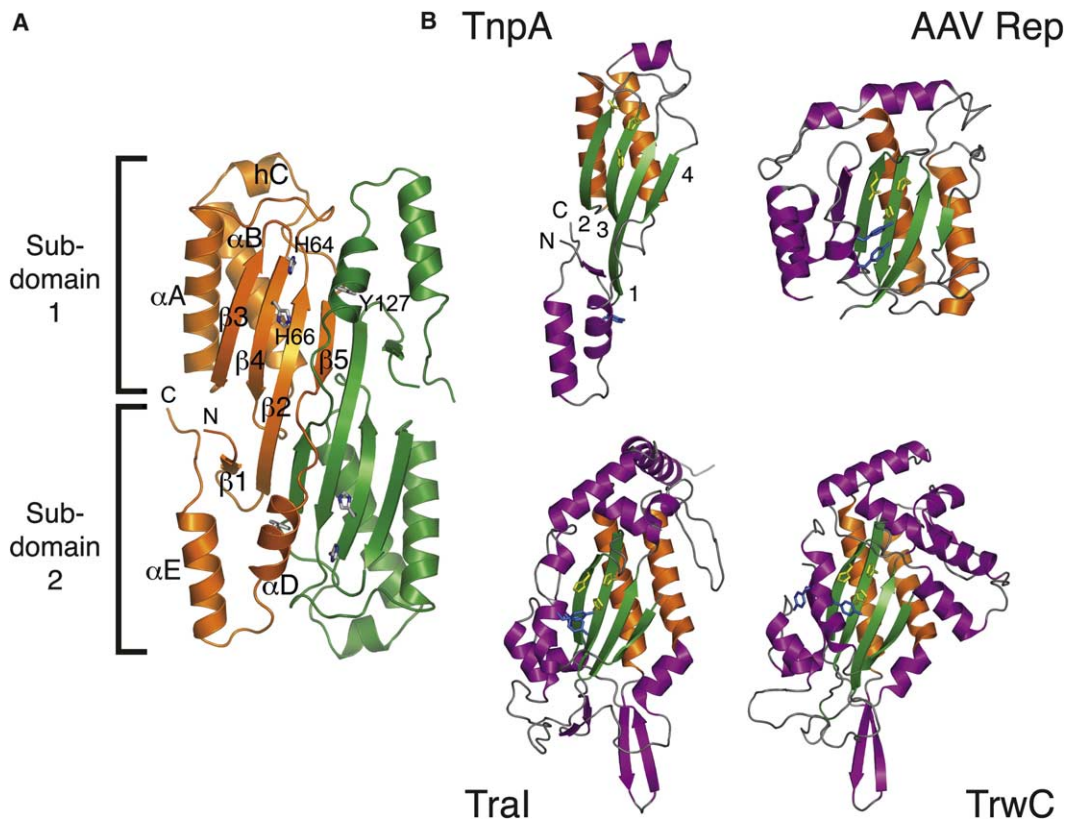


Figure 2. Structure of the TnpA Dimer and Homologs

(A) *ISHp608* TnpA dimer where orange and green ribbons represent the two TnpA molecules within the asymmetric unit. Bonds of the residues of the HUH motif and the invariant tyrosine are shown with gray (carbon), red (oxygen), and blue (nitrogen). Figure made in PyMOL (DeLano, 2002).

(B) Ribbon diagrams of TnpA, AAV5 Rep, Tral, and TrwC. The  $\alpha$  helices (orange) and  $\beta$  strands (green) of the RRM motif are numbered in TnpA. Variable secondary structural elements are purple. Yellow bonds highlight the HUH motif and other metal coordinating residues. Conserved tyrosine residues are in blue.

in dimerization. As the intermolecular interactions in the dimer are mediated by backbone hydrogen bonds of neighboring  $\beta$  strands and hydrophobic interactions between one face of the  $\beta$  sheet and helix  $\alpha$ D, there are no apparent sequence motifs that distinguish monomeric RCR-like transposases from TnpA.

#### A Shared Active Site

In other RCR proteins, the helix that harbors the nucleophilic tyrosine residues interacts with the  $\beta$  strands near the HUH motif (Figure 2B). In TnpA, however, after the two  $\beta$ 5 strands cross, helix  $\alpha$ D extends away from the active site of its own monomer and interacts with  $\beta$  strands from the second monomer (Figure 2A). Thus, the conserved tyrosine on helix  $\alpha$ D of TnpA, Tyr127, interacts with the HUH motif of the second monomer, whereas the invariant tyrosines in the other RCR proteins with available structures do so in the context of a single protein molecule. The major effect of this structural difference between TnpA and RCR proteins is the formation of an obligatory dimer with shared active sites.

#### Active Site and Catalytic Mechanism

The HUH motif plays an essential catalytic role in RCR proteins. The histidines, along with a third residue lo-

cated on the preceding  $\beta$  strand that is His in Tral and TrwC and Asp in the adeno-associated virus type 5 Rep (AAV5 Rep), coordinate a catalytically essential  $Mg^{2+}$  (Datta et al., 2003; Guasch et al., 2003; Hickman et al., 2002). During catalysis, the  $Mg^{2+}$  may localize the scissile phosphate at the proper place relative to the nucleophilic tyrosine and act as an electron-withdrawing group to promote the nucleophilic attack. This metal-dependent mechanism of phosphoryl transfer is strikingly similar to that of phosphoglucomutase/phosphomannomutase, although the two enzymes function in completely different biological contexts (Regni et al., 2004).

The TnpA HUH motif, comprised of His64, Ile65, and His66, is located on strand  $\beta$ 4 (Figure 3A), similar to its location in RCR proteins. The nearest residue that might serve as the third  $Mg^{2+}$ -coordinating ligand in TnpA is Asp61, although it is not strictly conserved among *IS200* elements. The most likely candidate for the active site nucleophile in *ISHp608* is Tyr127. It is the only tyrosine close to the residues of the HUH motif and the only fully conserved tyrosine in *IS200* TnpA proteins. However, in both structures, helix  $\alpha$ D is in a configuration where Tyr127 is turned away from the HUH motif (Figure 2A), forming a hydrogen bond with  $O_{\gamma}$  of

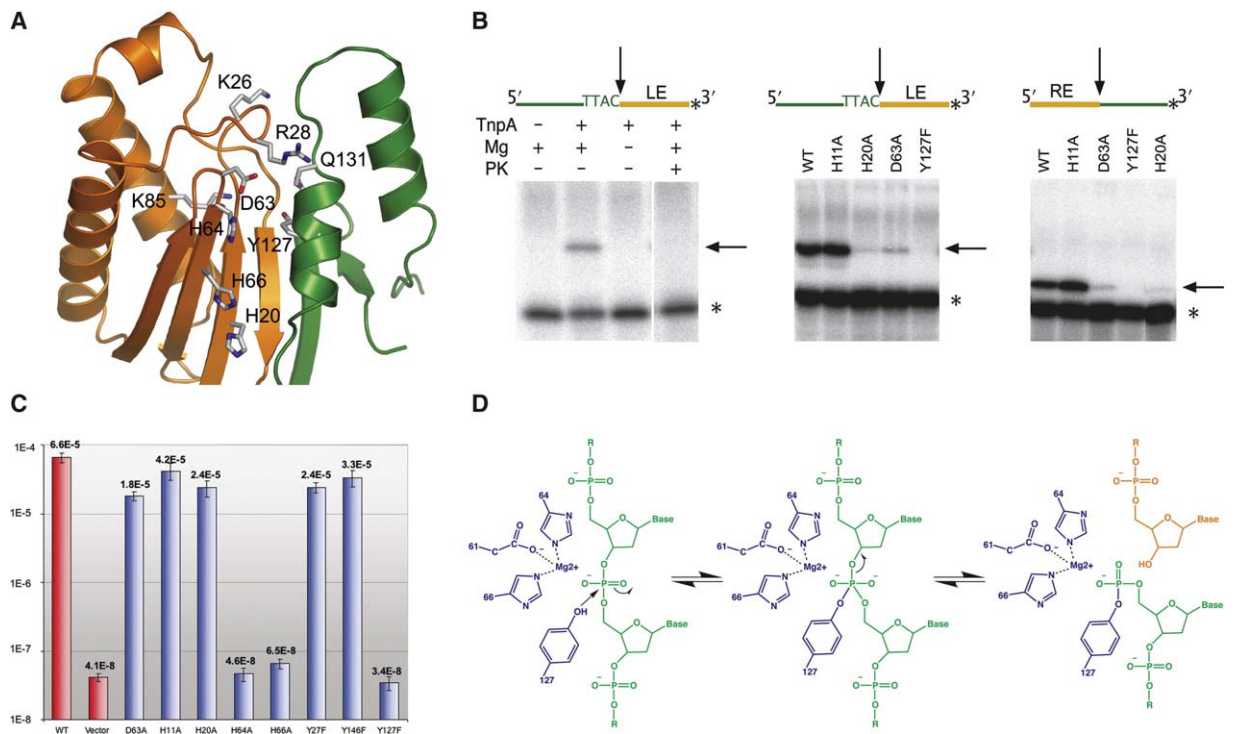


Figure 3. Active Site Residues and Catalytic Requirements

(A) The active site of TnpA showing conserved IS200 residues; only Lys85 is not in the active site.

(B) A radiogram of a denaturing 16% acrylamide gel shows covalent product formation during cleavage assays. A schematic of the DNA substrate for each reaction is above each gel. Green is flanking DNA, and yellow represents either the left end (LE) or right end (RE). Asterisks (\*) mark the position of the radioisotope in the schematic and the substrate in the gel. Arrows indicate the site of cleavage in the schematic and the product in the gel. PK, Proteinase K.

(C) Transposition frequency of wild-type and mutant TnpA from mating-out assays. Error bars are based on the results of four replicate measurements. Mutants within the HUH motif or at position 127 give the most significant changes.

(D) TnpA catalytic mechanism. Protein residues and  $Mg^{2+}$  are in blue and DNA in green. Orange DNA in the right panel is the leaving group.

Ser110. To properly juxtapose Tyr127 with the HUH motif, an  $\sim 90^\circ$  twist of helix  $\alpha D$  would be required. It is likely this occurs upon substrate binding.

### Biological and Biochemical Activity

To confirm the structural implication that His64, His66, and Tyr127 are required for catalysis, transposition activity was measured *in vivo* using mating-out assays with appropriately mutated TnpA derivatives (Figure 3C). All three point mutants (His64Ala, His66Ala, and Tyr127Phe) showed severe reductions in activity, similar to background levels, whereas mutation of other His and Tyr residues did not significantly affect transposition. Mutation of other residues near the active site (Asp63 and His20) resulted in only minor decreases in transposition frequency, suggesting they provide supporting roles in structurally organizing the active site.

*In vitro* DNA nicking assays indicated that wild-type TnpA forms a covalent intermediate with ssDNA representing either the right or left end of ISHp608 (Figure 3B). Formation of this intermediate required  $Mg^{2+}$  and was sensitive to Proteinase K (Figure 3B, left panel). In contrast, the Tyr127Phe mutant was unable to form an intermediate with either the left (middle panel) or right (right panel) ends. Asp63Ala and His20Ala mutants had

severely reduced activity in this assay in contrast to the small effect observed in the *in vivo* mating-out assay (Figure 3C); this difference may reflect a change of kinetics in the nicking step, which is less noticeable over the time period of the *in vivo* experiments.

The covalent intermediate is formed on the 5' end of the transposon at the left end, thereby liberating a free 3'-OH on the flanking plasmid vector DNA (Figure 3B). A mechanism for nucleophilic attack on DNA is shown in Figure 3D. At the right end, the covalent intermediate is formed with the 5' end of the flanking plasmid vector and generates a 3'-OH on the right transposon end (Ton-Hoang et al., 2005). This is essential, as the 3'-OH is presumably the nucleophile that attacks the phosphotyrosine intermediate at the opposite end of the transposon, thereby forming the closed circle transposon product observed *in vivo*.

These data, together with the observed formation of a circular dsDNA transposon intermediate *in vivo*, and the structure of the TnpA dimer, establish that ISHp608 is not a member of any characterized transposon family. For example, the mechanism of TnpA cleavage and covalent attachment contrasts with that of Y-transposases where dsDNA cleavage at the transposon end presumably results in the formation of a 3'-phosphotyro-

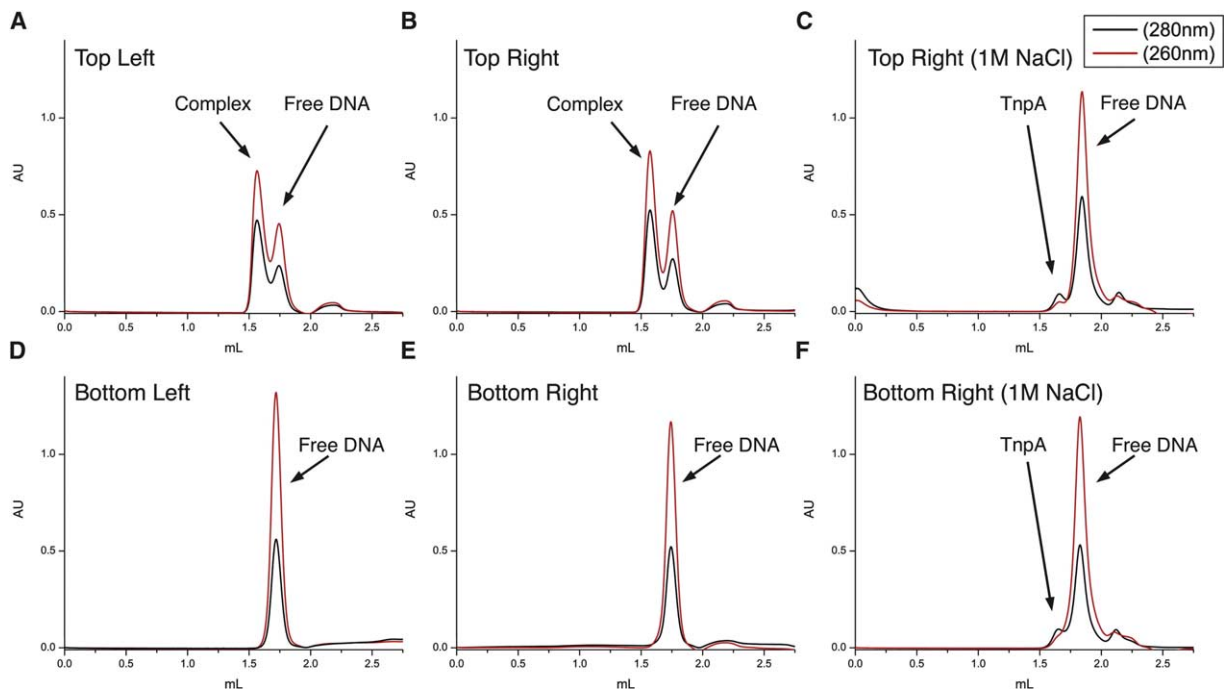


Figure 4. TnpA Binds Specifically to Top-Strand Stem-Loop DNA

Each panel represents a separate gel filtration experiment: (A) left end top strand, (B) right end top strand, (C) right end top strand with 1 M NaCl, (D) left end bottom strand, (E) right end bottom strand, and (F) right end bottom strand with 1 M NaCl. Red is the absorbance at 260 nm, and black is that at 280 nm. In the low-salt conditions, a shift in retention is observed in the presence of DNA representing the top strand stem loops (A and B). No shift is observed with bottom strand stem loops (D and E). The absence of a peak corresponding to protein in (D) and (E) is consistent with poor TnpA solubility in low ionic strength buffers. Identical samples under high-salt conditions (C and F) show a protein elution peak but no shift in retention time for the DNA.

sine intermediate (Figure 4 in Curcio and Derbyshire [2003]). *ISHp608* TnpA also differs from Y2-transposases, which are marked by “the absolute conservation of a pair of tyrosines that are separated by three residues” (Curcio and Derbyshire, 2003). In light of these mechanistic and structural differences, we propose the designation “Y1 transposase” for *ISHp608*-like elements, to distinguish between these three classes of transposases that use a nucleophilic tyrosine.

#### TnpA/Hairpin Interactions

TnpA precisely excises and integrates *ISHp608* without gain or loss of sequence at either end, suggesting that it recognizes specific DNA sequences at or near these ends. Multiple repeat sequences have been identified at the left and right ends of *ISHp608* elements from various *H. pylori* strains (Kersulyte et al., 2002). Although their consecutive deletion results in decreasing transposition activity (Ton-Hoang et al., 2005), the specific role of each repeat in transposon end recognition has not been defined. Our previous studies have shown that, as expected for a transposase, TnpA binds to both left and right ends (Ton-Hoang et al., 2005). More surprisingly, binding was found to be strand specific.

The subterminal region of each end contains a conserved imperfect palindrome sequence of 22 or 23 base pairs that can form a stem-loop structure containing an interior loop (Figure 1B). It has been suggested that subterminal palindromes within IS200-like elements act at the RNA level by forming stem loops that could serve

as transcriptional terminators to prevent expression of the element’s transposase by transcription of flanking genes (Lam and Roth, 1983), or to occlude a ribosome recognition site (Beuzon and Casadesus, 1997). Since RCR proteins often recognize DNA stem-loop structures to initiate their respective reactions, we were intrigued by the hypothesis that these imperfect palindromes, rather than having a regulatory function in *ISHp608*, may instead play a role in end recognition.

A size exclusion chromatography-based binding assay indicated (Figure 4) that, under conditions of low ionic strength, TnpA forms a stable complex with both the left and right top strand stem loops but has no affinity for either the left or right bottom strands.

#### Structure of TnpA-Hairpin Complex

The structure of TnpA bound to DNA representing the top right hairpin was solved with molecular replacement and subsequently refined at 2.6 Å resolution (Table 1). The a.s.u. in the complex structure contains a TnpA dimer and two stem loops (Figure 5A). The DNA molecules bind to regions of strong positive electrostatic potential on the opposite side of TnpA from the active sites that are formed by the convergence of residues within and near helix  $\alpha$ B and the turn between  $\beta$ 1 and  $\beta$ 2 of the opposing molecule (Figure 5B).

The DNA stem loop (shown schematically in Figure 5C) has a stem consisting of eight Watson-Crick base pairs and one extrahelical base (T17), forming an interior loop; a two-membered hairpin loop (T10 and T11)

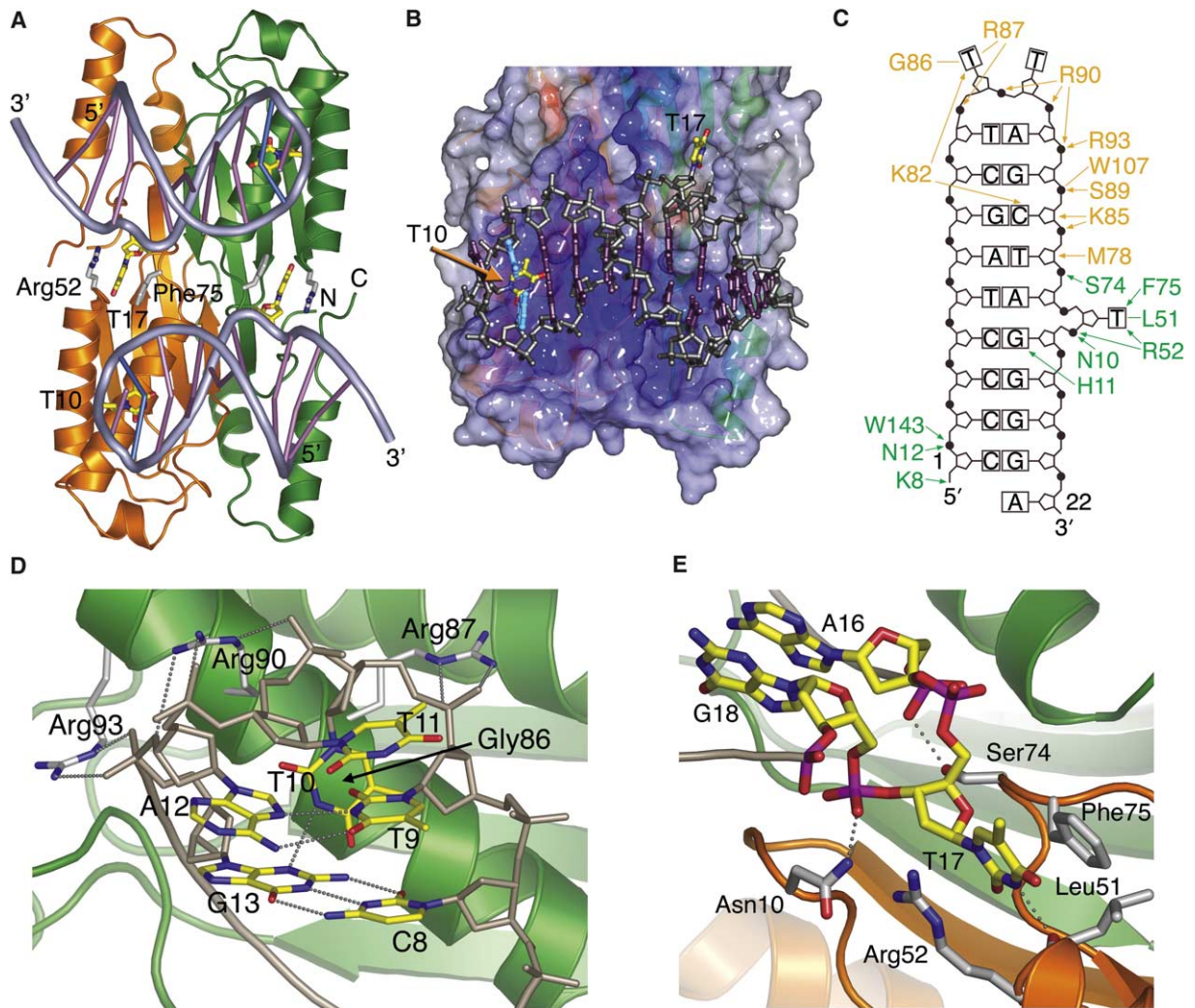


Figure 5. TnpA/Stem-Loop Complex Structure

(A) A ribbon diagram of the TnpA/stem-loop structure. The two TnpA molecules are orange and green and in the same orientation as Figure 2A. The two DNA molecules are in gray, with the Watson-Crick base pairs in violet and the Hoogsteen base pair in blue.

(B) The molecular surface of the TnpA dimer is colored by electrostatic potential where blue and red represent positive and negative potential, respectively, calculated in SPOCK (Christopher, 1998). The two bases crucial for discrimination between top and bottom stem loops, T10 and T17, are in yellow.

(C) A schematic diagram of protein/DNA interactions. Residues involved in DNA binding are in orange or green, corresponding to the protein molecule of origin. Arrows and lines represent protein side chain and backbone interactions with the DNA, respectively.

(D) TnpA forms specific interactions with the loop region of the DNA. The green ribbon and white carbon atoms represent protein. The gray ribbon and yellow carbon atoms represent DNA. Dashed lines are bonding interactions. Base T10 stacks against the peptide bond between Gly86 and Arg87. The Hoogsteen T9-A12 base pair is also shown.

(E) TnpA forms specific interactions with the extrahelical base, T17. Protein ribbons are orange and green, and DNA is gray. T17 stacks between Arg52 and Phe75, and hydrogen bonds to the carbonyl oxygen of L51. Despite the flipped-out T17, A16 and G18 continue base stacking of the double helix.

with a Hoogsteen base pair at its base; and a single base overhang at the 3' end. While TnpA makes numerous contacts with the phosphate backbone, only two protein side chains form base-specific interactions in the canonical double-helical part of the DNA: His11 contacts G18 in the major groove, and Lys82 contacts C8 in the minor groove. The significance of these is not clear, as neither residue is conserved and Lys82 appears to be mobile with multiple conformations. In contrast, a rich network of interactions exists to recognize the loop structures and bases T10 and T17 in these

loops, forming the basis for discrimination between the top and bottom strand stem loops.

#### T10 at the Tip

T10 and T11 are the two unpaired bases of the hairpin loop. The flanking bases, T9 and A12, form a Hoogsteen base pair (Figure 5D), an uncommon pairing that is likely a result of structural constraints imposed by the sharp DNA hairpin turn, as neither T9 nor A12 contacts TnpA. The plane of T10 is rotated into the minor groove formed by the three base pairs flanking the loop, form-

ing a wedge and allowing N3 of T10 to hydrogen bond to N3 of G13. Helix  $\alpha$ B abuts this wedge, prohibiting the complete entry of helix  $\alpha$ B into the minor groove while still allowing interactions between basic residues on helix  $\alpha$ B and phosphate moieties on the DNA backbone (Figure 5D). T10 stacks against the peptide bond between residues Gly86 and Arg87 and sits in a pocket perfect for its size.

Other residues on helix  $\alpha$ B provide charge-charge interactions. For example, the guanidinium moiety of Arg90 is positioned near the apex of the stem loop where the DNA backbone makes an almost 180° turn (Figure 5D). As a result, Arg90 forms hydrogen bonds with backbone phosphates from three consecutive bases: T11, A12, and G13.

### The Importance of Being T17

The second region of the stem loop that is vital for recognition by TnpA is the flipped-out base, T17. While the Tn5 transposition reaction produces a hairpinned reaction intermediate containing a flipped-out T base (Davies et al., 2000), the flipped-out base in *ISHp608* has a different role, as it is used for specific end recognition before any chemistry takes place. In this case, T17 protrudes from the midway point along the stem, forming an interior loop. Base stacking before and after T17 is very close to that seen in B-form DNA, such that the flipped-out conformation causes little distortion.

The interactions between TnpA and T17 are mediated by both base-specific and base-nonspecific interactions (Figure 5E). T17 stacks between the benzene ring of Phe75 and the guanidinium moiety of Arg52. Presumably, these interactions decrease the energetic penalty of exposing the hydrophobic base to solvent. N3 of T17 is recognized by a short hydrogen bond to the carbonyl oxygen of Leu51. Hydrogen bonds between the side chains of Asn10 and Ser74 and the phosphate moieties of T16 and G18 help to position the flipped-out base within its binding pocket.

### Distinguishing Top from Bottom

Specific recognition of T10 and T17 is the structural basis by which TnpA distinguishes between the top and the bottom strands. The flipped-out T17 sits in a binding pocket that is unable to accommodate the larger adenine base of the complementary bottom strand stem loop. Furthermore, the hydrogen bond between T17 and Leu51 can only be formed if thymine is found at position 17, as no other naturally occurring DNA base possesses a hydrogen bond donor at this position. The importance of this single hydrogen bond is highlighted by the conservation of the thymine at both the left and right ends across all of the observed *ISHp608* elements.

The requirement of a pyrimidine at position ten of the stem loop results from insertion of Lys82 into the minor groove near the loop. The larger purine base, A, at this position would produce steric clashes with this residue. The preference of thymine over cytosine stems from hydrogen bonding within the DNA loop, where the N3 atom of T10 hydrogen bonds with N3 from G13, an interaction that could not occur with cytosine.

As shown in Figure 1B, the sequences of *ISHp608* elements from different *H. pylori* strains have slight variations in the imperfect palindrome that may affect stem-loop structure. For example, several elements have a left end palindrome with three unpaired thymine bases in the loop region, whereas the right end has only two unpaired thymine bases, as seen in the crystal structure. We predict that the addition of the third unpaired thymine should have only limited effects on the stem-loop structure.

### Relationship to Other Proteins that Recognize Stem Loops

It is intriguing that, among the few known protein/DNA hairpin complex structures, those that represent authentic biological complexes contain RCR proteins. AAV5 Rep (Hickman et al., 2004) binds to a stem loop at the end of the viral genome, and TrwC binds to a perfect stem loop that is extruded as part of a cruciform structure prior to the initiation of conjugative transfer (Guasch et al., 2003). While these interactions are functionally similar to the TnpA case, as they direct the nuclease toward its intended cleavage site, the mode of stem-loop recognition is very different. AAV5 Rep binds specifically only to the unpaired bases at the tip of the hairpin and does not contact the stem. In contrast, TrwC recognizes a sequence in the canonical Watson-Crick base-paired stem and does not interact with the loop at the tip.

Although TnpA is the first transposition system in which imperfect stem loops have been shown to be crucial for transposon end recognition, these unusual DNA structures have been exploited for DNA recognition in a variety of systems. For example, the integrase from transposon Tn21, a site-specific tyrosine recombinase, has been shown to bind bulged hairpin DNA, where it specifically binds only the bottom strand of the attC site within the integron cassette, and hairpin binding is abrogated if bases in the bulged positions are mutated (Johansson et al., 2004). Furthermore, imperfect stem loops required for efficient viral replication are found in the genomes of human immunodeficiency virus (Aldovini and Young, 1990; Clavel and Orenstein, 1990; Lever et al., 1989), the minute virus in mice (Costello et al., 1995), and vaccinia virus (DeMasi et al., 2001).

### Model for Transposition Initiation and Intermediate Formation

We have shown that TnpA can distinguish between the top and bottom strand stem loops at both ends of *ISHp608* and that TnpA can nick DNA at both the left and right ends of the transposon, the former in the form of dsDNA but the latter only as ssDNA (Ton-Hoang et al., 2005). The position of the stem loops relative to the cleavage sites is asymmetric at the two *ISHp608* ends, as left end cleavage occurs 19 bases 5' of the stem loop, whereas right end cleavage occurs nine bases 3' of the stem loop. Given the 2-fold symmetry of the TnpA/stem-loop complex, these data suggest that the two ends must travel different paths from the base of the stem loops to reach their respective active sites.

While TnpA can cleave linear dsDNA at the left end



in vitro, the presence of the stem-loop sequence significantly increases the transposition frequency when negatively supercoiled DNA is used in vivo (Ton-Hoang et al., 2005). We propose that supercoiling promotes the extrusion of the left stem loop (Mizuuchi et al., 1982; Sinden, 1994), which is bound by TnpA. The TnpA/stem-loop structure demonstrates that the TnpA dimer can simultaneously bind two stem loops, and that binding does not occlude the active sites. Binding of the left stem loop directs the left end cleavage site into one active site, guided by the tertiary structure of the four-way Holliday junction formed at the base of the stem loop. Tyr127 then attacks the top strand at the left end as shown in Figure 3D to produce a phosphotyrosine intermediate at the 5' end of the element and a free 3'-OH on the flanking DNA (Figure 6, step 1).

As right end cleavage requires the substrate to be single stranded, it is possible that ssDNA is generated by the displacement of the top strand by a host helicase activity (Figure 6, step 2). Once the helicase passes the right end, single-stranded transposon DNA forms a stem loop that can be bound by the second binding site of TnpA, thereby pairing the two ends (Figure 6, step 3). The single-stranded right end cleavage site is then directed to the second active site of the TnpA dimer, where nucleophilic attack leads to formation of a covalent phosphotyrosine intermediate with the DNA flanking the right end of the element, and a free 3' OH on the right end of ISHp608 (Figure 6, step 4).

To form the sealed recircularized donor plasmid with a large extruded loop, the 3'-OH at the left end flank attacks the covalently attached Tyr127 on the 5' right end flank (Figure 6, step 5). Replication of this intermediate would form the donor plasmid backbone (DPB) and regenerate the transposon-containing donor plasmid (TDP), consistent with in vivo observations (Ton-Hoang et al., 2005). Also formed in step 5 is the excised transposon whose left end is still covalently attached to the transposase. Nucleophilic attack by the 3'-OH at the transposon right end on the phosphotyrosine linkage at the left end would generate a ssDNA transposon circle. It is not yet clear how the single-stranded form is converted to the double-stranded excised transposon (ET), but we are drawn to the possibility that, given the stem-loop structures, second strand synthesis might occur by a mechanism similar to that of RCR plasmids (Arai et al., 1981).

The observation that transposon end recognition occurs on the opposite face of the TnpA dimer from the active sites explains why the repeats are subterminal rather than at the element termini: this leaves the two active sites accessible and enough adjacent DNA to reach the active sites. We have considered the possibility that only one active site of the TnpA dimer is used; this could also account for the asymmetric spacing between the stem loops and the cleavage sites at the two ends, as one end would have to reach further than the other to encounter the active site. However, we do not favor this possibility, as it would require the spontaneous resolution of the phosphotyrosine intermediate so that the active site can be used again. The observation that TnpA is a dimer and that both active sites are located on the same face of the molecule allows for

cleavage at both ends of the elements using two active sites.

The proposed mechanism in Figure 6 for end recognition, end pairing, and the introduction of nicks at both ends of the element will likely apply to other IS200 family members as the catalytically important TnpA residues are strictly conserved. Among the conserved TnpA residues, those that form the HUH motif (His64 and His66) and those close by (Asp63 and His20) are important for catalysis, as shown by both DNA nicking assays in vitro and mating-out assays in vivo. Two other conserved residues, Lys85 and Gly86, are crucial for stem-loop binding. The results presented here establish basic aspects of the structure/function relationships of ISHp608 transposition, but further studies are needed to clarify the transposition pathway and the mode of interaction of the cleavage substrate and the 5'-TTAC-3' motif with TnpA.

## Experimental Procedures

### Protein Purification and Crystallization

The *orfA* gene from ISHp608 strain PeCan2a was placed between the *NcoI* and *XhoI* sites of a modified pET-32b plasmid (EMD Biosciences). The resulting recombinant protein is a histidine-tagged thioredoxin/TnpA fusion.

TnpA expression and purification was identical to that described for TnsA (Ronning et al., 2004), with the following exceptions: PBS buffer was used, and the gel filtration buffer did not contain MgSO<sub>4</sub>. TnpA was dialyzed against 20 mM Tris (pH 7.5), 500 mM sodium formate, 5 mM dithiothreitol, and 2 mM EDTA. Crystals grew at 20°C in hanging drops prepared by mixing 3 μl each of protein and a well solution containing 200 mM NaSCN and 24% PEG 3350. The C2 space group with unit cell parameters  $a = 124.52 \text{ \AA}$ ,  $b = 50.98 \text{ \AA}$ ,  $c = 52.92 \text{ \AA}$ ,  $\beta = 105.45^\circ$  contains two molecules in the a.s.u. Selenomethionine-derivatized protein was expressed in B843(DE3) *E. coli* cells containing both the TnpA expression plasmid and the Rosetta plasmid (EMD Biosciences) using M9 media supplemented with selenomethionine and all of the natural amino acids except methionine.

Hairpin DNA (5'-CCCCTAGCTTTAGCTATGGGA-3') was synthesized on an Applied Biosciences 394 DNA/RNA synthesizer, heated at 95°C for 15 min, then rapidly cooled on ice. The DNA was added to TnpA (6.6 mg/ml), resulting in a 1:1.1 protein:DNA ratio. The complex was dialyzed against 20 mM Tris (pH 7.5), 250 mM sodium malonate, 2 mM EDTA, and 0.2 mM TCEP and crystallized at 20°C in a pre-filled PEG suite microplate (Nextal Biotechnologies) where the well solution contained 0.2 M sodium tartrate and 20% PEG 3350. The P<sub>2</sub> space group has the unit cell parameters  $a = 44.33 \text{ \AA}$ ,  $b = 71.14 \text{ \AA}$ ,  $c = 76.75 \text{ \AA}$ ,  $\beta = 94.44^\circ$ . Two protein and two DNA molecules are in the a.s.u.

### Data Collection and Structure Determination

TnpA crystals were cryoprotected with Paratone-N (Hampton Research) and flash cooled at 100 K. Multiwavelength diffraction data were collected on a CCD detector and integrated and scaled using HKL2000 (Otwinowski and Minor, 1997). Six selenium sites were identified using data to 2.4 Å resolution in SOLVE (Terwilliger and Berendzen, 1999) and solvent flattened with DM (CCP4, 1994; Cowtan, 1994). The model was built with O (Jones et al., 1991) and refined using data collected at 95 K on a rotating anode using Cu<sub>Kα</sub> radiation and an R-axis 4++ image plate detector. Refinement proceeded with simulated annealing, energy minimization, and individual B factor refinement in CNS (Brunger et al., 1998). Water molecules were placed using CNS and manually confirmed or rejected. The Ramachandran plot has only one residue in the disallowed region, located in a loop that has poor electron density.

TnpA/DNA crystals were cryoprotected with Paratone-N and flash cooled to 95 K using liquid nitrogen. TnpA/DNA data were collected using Cu<sub>Kα</sub> radiation as above. Data were integrated and scaled

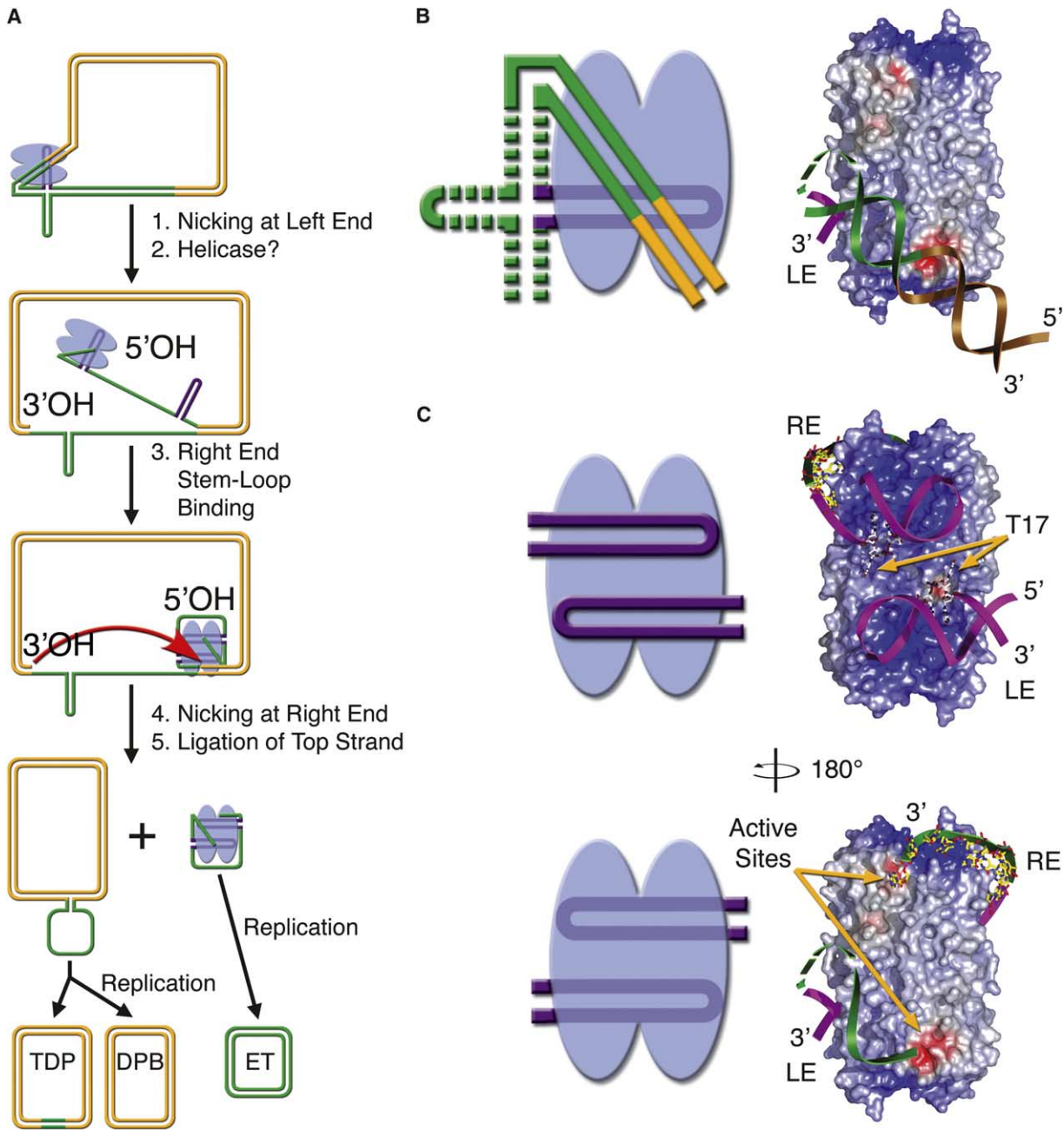


Figure 6. Model for ISHp608 Transposition

(A) The schematic shows donor DNA (orange) and IS element DNA (green). Stem-loop DNA is purple. The TnpA dimer is blue. Step 1, TnpA binds the left end stem loop. Step 3, the red arrow shows the attack of the 3'OH of the left end flanking DNA on the phosphotyrosine intermediate at the right end. Step 5 illustrates the formation of a TnpA/transposon intermediate that resembles the TnpA/stem-loop complex structure, and the closed donor plasmid. TDP, transposon donor plasmid; DPB, donor plasmid backbone; ET, excised transposon.

(B) Schematic and model of left end recognition by TnpA using the same coloring scheme as in (A). The protein surface is colored by electrostatic potential. Negatively charged red patches highlight the metal binding sites.

(C) The schematics represent two orientations of the complex crystal structure. The corresponding models are in the same orientation and contain the 19 and nine terminal bases of the element modeled onto the structure. The left end is a green ribbon. The right end is a green ribbon with yellow bonds.

with HKL2000 (Otwinowski and Minor, 1997). The structure was solved using molecular replacement with a truncated TnpA dimer model (residues 12–115) in CNS (Brunger et al., 1998). B-form DNA was placed within density corresponding to the stem loop. The resulting map allowed for accurate building of the DNA. Iterations

of model refinement were as described for TnpA alone. The model contains residues 5–155 of TnpA and bases 1–22 of the stem-loop DNA, with all but Lys62 in the allowed regions. The poor stereochemistry of Lys62 is a result of its location in the tight turn between strands  $\beta$ 3 and  $\beta$ 4.

#### DNA Binding Assay

Oligonucleotides were from Integrated DNA Technologies, Inc. (Coralville, Iowa). Left top, 5'-CCCCTAGCTTTTAGCTATGGGA-3'; left bottom, 5'-TCCCCATAGCTAAAAGCTAGGGG-3'; right top, 5'-CCCCTAGCTTTTAGCTATGGGA-3'; right bottom, 5'-TCCCCATAGCTAAAGCTAGGGG-3'. Each was dissolved in 10 mM Tris (pH 8) and 1 mM EDTA and annealed as described above. TnpA and DNA were combined at a 1:2 ratio in 1 × PBS (pH 7.4), 1 M NaCl, 10% glycerol, 2 mM EDTA, and 5 mM β-ME. The final TnpA concentration was 20 μM. The complex was dialyzed against 20 mM Tris (pH 7.5), 0.2 M NaCl, 2 mM EDTA, and 0.2 mM TCEP and injected onto a Superdex 200 column (Amersham-Pharmacia).

#### Mating-Out Assay

The frequency of *ISHp608* transposition was determined by a standard mating-out assay (Galas and Chandler, 1982) using the conjugal plasmid pOX38Km (Chandler and Galas, 1983).

#### DNA Nicking and Covalent Complex Formation In Vitro

Oligonucleotides B81 and B85 (corresponding to LE and RE top strand) were 3' <sup>32</sup>P-radiolabeled with α-<sup>32</sup>P ddATP. Reactions were performed in a final volume of 15 μl containing 14 nM substrate, 1.5 μg TnpA derivatives in 20 mM HEPES (pH 7.5), 0.12 M NaCl, 1 mM DTT, 20 ng/ml BSA, 0.25 μg of tRNA, and 14% glycerol. After 20 min incubation at 37°C, MgCl<sub>2</sub> was added to final concentration of 5 mM, and incubation continued for 30 min. Reactions were terminated by addition of 16 μl of 2 × SDS loading buffer. Products were denatured and analyzed on a 16% Laemmli protein gel.

#### Supplemental Data

Supplemental Data include supplemental text, one figure, and one table and can be found with this article online at <http://www.molecule.org/cgi/content/full/20/1/143/DC1/>.

#### Acknowledgments

We thank W. Yang, C. Bradley, A.B. Hickman, and Cc. Huang for comments on the manuscript; T. Chiu for help with CURVES; and the CNRS (UMR5100), the "Programme microbiologie" (MENRST), and the Groupement de Recherche (GDR 2157) for funding. Data were collected at the Southeast Regional Collaborative Access Team (SER-CAT) 22-ID beamline at the Advanced Photon Source, Argonne National Laboratory. Use of the Advanced Photon Source was supported by the U.S. Department of Energy, Basic Energy Sciences, Office of Science, under Contract No. W-31-109-Eng-38. This research was supported in part by the Intramural Research Program of the NIH, NIDDK.

Received: April 29, 2005

Revised: June 22, 2005

Accepted: July 18, 2005

Published: October 6, 2005

#### References

Akopyants, N.S., Clifton, S.W., Kersulyte, D., Crabtree, J.E., Youree, B.E., Reece, C.A., Bukanov, N.O., Drazek, E.S., Roe, B.A., and Berg, D.E. (1998). Analyses of the cag pathogenicity island of *Helicobacter pylori*. *Mol. Microbiol.* 28, 37–53.

Aldovini, A., and Young, R.A. (1990). Mutations of RNA and protein sequences involved in human immunodeficiency virus type 1 packaging result in production of noninfectious virus. *J. Virol.* 64, 1920–1926.

Arai, K., Low, R., Kabori, J., Shlomai, J., and Kornberg, A. (1981). Mechanism of dnaB protein action. V. Association of dnaB protein, protein n', and other repriming proteins in the primosome of DNA replication. *J. Biol. Chem.* 256, 5273–5280.

Beuzon, C.R., and Casadesu, J. (1997). Conserved structure of IS200 elements in *Salmonella*. *Nucleic Acids Res.* 25, 1355–1361.

Brunger, A.T., Adams, P.D., Clore, G.M., Delano, W.L., Gros, P., Grosse-Kunstleve, R.W., Jiang, J.S., Kuszewski, J., Nilges, M., Pannu, N.S., et al. (1998). Crystallography & NMR system: a new

software suite for macromolecular structure determination. *Acta Crystallogr. D Biol. Crystallogr.* 54, 905–921.

Burd, C.G., and Dreyfuss, G. (1994). Conserved structures and diversity of functions of RNA-binding proteins. *Science* 265, 615–621.

Campos-Olivas, R., Louis, J.M., Clerot, D., Gronenborn, B., and Gronenborn, A.M. (2002). The structure of a replication initiator unites diverse aspects of nucleic acid metabolism. *Proc. Natl. Acad. Sci. USA* 99, 10310–10315.

CCP4 (Collaborative Computational Project, Number 4) (1994). The CCP4 suite: programs for protein crystallography. *Acta Crystallogr. D Biol. Crystallogr.* 50, 760–763.

Censini, S., Lange, C., Xiang, Z., Crabtree, J.E., Ghiara, P., Borodovsky, M., Rappuoli, R., and Covacci, A. (1996). cag, a pathogenicity island of *Helicobacter pylori*, encodes type I-specific and disease-associated virulence factors. *Proc. Natl. Acad. Sci. USA* 93, 14648–14653.

Chain, P.S., Carniel, E., Larimer, F.W., Lamerdin, J., Stoutland, P.O., Regala, W.M., Georgescu, A.M., Vergez, L.M., Land, M.L., Motin, V.L., et al. (2004). Insights into the evolution of *Yersinia pestis* through whole-genome comparison with *Yersinia pseudotuberculosis*. *Proc. Natl. Acad. Sci. USA* 101, 13826–13831.

Chandler, M., and Galas, D.J. (1983). IS1-mediated tandem duplication of plasmid pBR322. Dependence on recA and on DNA polymerase I. *J. Mol. Biol.* 165, 183–190.

Chandler, M., and Mahillon, J. (2002). Insertion sequences revisited. In *Mobile DNA II*, N.L. Craig, R. Craigie, M. Gellert, and A.M. Lambowitz, eds. (Washington, D.C.: ASM Press), pp. 305–366.

Christopher, J.A. (1998). SPOCK: The Structural Properties Observation and Calculation Kit (College Station, Texas: Texas A&M University).

Clavel, F., and Orenstein, J.M. (1990). A mutant of human immunodeficiency virus with reduced RNA packaging and abnormal particle morphology. *J. Virol.* 64, 5230–5234.

Costello, E., Sahli, R., Hirt, B., and Beard, P. (1995). The mismatched nucleotides in the 5'-terminal hairpin of minute virus of mice are required for efficient viral DNA replication. *J. Virol.* 69, 7489–7496.

Cowtan, K. (1994). 'dm': an automated procedure for phase improvement by density modification. *Joint CCP4/ESF-EACBM Newsletter on Protein Crystallography* 37, 34–38.

Curcio, M.J., and Derbyshire, K.M. (2003). The outs and ins of transposition: from mu to kangaroo. *Nat. Rev. Mol. Cell Biol.* 4, 865–877.

Datta, S., Larkin, C., and Schildbach, J.F. (2003). Structural insights into single-stranded DNA binding and cleavage by F factor *Tral*. *Structure (Camb)* 11, 1369–1379.

Davies, D.R., Goryshin, I.Y., Reznikoff, W.S., and Rayment, I. (2000). Three-dimensional structure of the Tn5 synaptic complex transposition intermediate. *Science* 289, 77–85.

DeLano, W.L. (2002). The PyMOL Molecular Graphics System. [www.pymol.org](http://www.pymol.org)

DeMasi, J., Du, S., Lennon, D., and Traktman, P. (2001). Vaccinia virus telomeres: interaction with the viral I1, I6, and K4 proteins. *J. Virol.* 75, 10090–10105.

Dyda, F., Hickman, A.B., Jenkins, T.M., Engelman, A., Craigie, R., and Davies, D.R. (1994). Crystal structure of the catalytic domain of HIV-1 integrase: similarity to other polynucleotidyl transferases. *Science* 266, 1981–1986.

Galas, D.J., and Chandler, M. (1982). Structure and stability of Tn9-mediated cointegrates. Evidence for two pathways of transposition. *J. Mol. Biol.* 154, 245–272.

Garcillán-Barcia, M., Bernales, I., Mendiola, M.V., and de la Cruz, F. (2001). Single-stranded DNA intermediates in IS91 rolling-circle transposition. *Mol. Microbiol.* 39, 494–501.

Goff, S.A., Ricke, D., Lan, T.H., Presting, G., Wang, R., Dunn, M., Glazebrook, J., Sessions, A., Oeller, P., Varma, H., et al. (2002). A draft sequence of the rice genome (*Oryza sativa* L. ssp. japonica). *Science* 296, 92–100.

Guasch, A., Lucas, M., Moncalian, G., Cabezas, M., Perez-Luque,

- R., Gomis-Ruth, F.X., de la Cruz, F., and Coll, M. (2003). Recognition and processing of the origin of transfer DNA by conjugative relaxase TrwC. *Nat. Struct. Biol.* 10, 1002–1010.
- Hickman, A.B., Ronning, D.R., Kotin, R.M., and Dyda, F. (2002). Structural unity among viral origin binding proteins: crystal structure of the nuclease domain of adeno-associated virus Rep. *Mol. Cell* 10, 327–337.
- Hickman, A.B., Ronning, D.R., Perez, Z.N., Kotin, R.M., and Dyda, F. (2004). The nuclease domain of adeno-associated virus rep coordinates replication initiation using two distinct DNA recognition interfaces. *Mol. Cell* 13, 403–414.
- Holm, L., and Sander, C. (1993). Protein structure comparison by alignment of distance matrices. *J. Mol. Biol.* 233, 123–138.
- Ilyina, T.V., and Koonin, E.V. (1992). Conserved sequence motifs in the initiator proteins for rolling circle DNA replication encoded by diverse replicons from eubacteria, eucaryotes and archaeobacteria. *Nucleic Acids Res.* 20, 3279–3285.
- Johansson, C., Kamali-Moghaddam, M., and Sundstrom, L. (2004). Integron integrase binds to bulged hairpin DNA. *Nucleic Acids Res.* 32, 4033–4043.
- Jones, T.A., Zou, J.Y., Cowan, S.W., and Kjeldgaard, M. (1991). Improved methods for building protein models in electron density maps and the location of errors in these models. *Acta Crystallogr. A* 47, 110–119.
- Kersulyte, D., Velapatino, B., Dailide, G., Mukhopadhyay, A.K., Ito, Y., Cahuayme, L., Parkinson, A.J., Gilman, R.H., and Berg, D.E. (2002). Transposable element ISHp608 of *Helicobacter pylori*: non-random geographic distribution, functional organization, and insertion specificity. *J. Bacteriol.* 184, 992–1002.
- Koonin, E.V., and Ilyina, T.V. (1993). Computer-assisted dissection of rolling circle DNA replication. *Biosystems* 30, 241–268.
- Kulkosky, J., Jones, K.S., Katz, R.A., Mack, J.P., and Skalka, A.M. (1992). Residues critical for retroviral integrative recombination in a region that is highly conserved among retroviral/retrotransposon integrases and bacterial insertion sequence transposases. *Mol. Cell. Biol.* 12, 2331–2338.
- Lam, S., and Roth, J.R. (1983). IS200: a *Salmonella*-specific insertion sequence. *Cell* 34, 951–960.
- Lander, E.S., Linton, L.M., Birren, B., Nusbaum, C., Zody, M.C., Baldwin, J., Devon, K., Dewar, K., Doyle, M., FitzHugh, W., et al. (2001). Initial sequencing and analysis of the human genome. *Nature* 409, 860–921.
- Lever, A., Gottlinger, H., Haseltine, W., and Sodroski, J. (1989). Identification of a sequence required for efficient packaging of human immunodeficiency virus type 1 RNA into virions. *J. Virol.* 63, 4085–4087.
- Lovell, S., Goryshin, I.Y., Reznikoff, W.R., and Rayment, I. (2002). Two-metal active site binding of a Tn5 transposase synaptic complex. *Nat. Struct. Biol.* 9, 278–281.
- Mendiola, M.V., and de la Cruz, F. (1989). Specificity of insertion of IS91, an insertion sequence present in alpha-haemolysin plasmids of *Escherichia coli*. *Mol. Microbiol.* 3, 979–984.
- Mendiola, M.V., and de la Cruz, F. (1992). IS91 transposase is related to the rolling-circle-type replication proteins of the pUB110 family of plasmids. *Nucleic Acids Res.* 20, 3521.
- Mendiola, M.V., Bernales, I., and de la Cruz, F. (1994). Differential roles of the transposon termini in IS91 transposition. *Proc. Natl. Acad. Sci. USA* 91, 1922–1926.
- Mizuuchi, K., Mizuuchi, M., and Gellert, M. (1982). Cruciform structures in palindromic DNA are favored by DNA supercoiling. *J. Mol. Biol.* 156, 229–243.
- Naumann, M., and Crabtree, J.E. (2004). *Helicobacter pylori*-induced epithelial cell signalling in gastric carcinogenesis. *Trends Microbiol.* 12, 29–36.
- Otwinowski, Z., and Minor, W. (1997). Processing X-ray diffraction data collected in oscillation mode. In *Methods in Enzymology*, Volume 276, C.W. Carter, Jr., and R.M. Sweet, eds. (New York: Academic Press), pp. 307–326.
- Parkhill, J., Sebaihia, M., Preston, A., Murphy, L.D., Thomson, N., Harris, D.E., Holden, M.T., Churcher, C.M., Bentley, S.D., Mungall, K.L., et al. (2003). Comparative analysis of the genome sequences of *Bordetella pertussis*, *Bordetella parapertussis* and *Bordetella bronchiseptica*. *Nat. Genet.* 35, 32–40.
- Regni, C., Naught, L., Tipton, P.A., and Beamer, L.J. (2004). Structural basis of diverse substrate recognition by the enzyme PMM/PGM from *P. aeruginosa*. *Structure (Camb)* 12, 55–63.
- Rice, P., and Mizuuchi, K. (1995). Structure of the bacteriophage Mu transposase core: a common structural motif for DNA transposition and retroviral integration. *Cell* 82, 209–220.
- Ronning, D.R., Li, Y., Perez, Z.N., Ross, P.D., Hickman, A.B., Craig, N.L., and Dyda, F. (2004). The carboxy-terminal portion of TnsC activates the Tn7 transposase through a specific interaction with TnsA. *EMBO J.* 23, 2972–2981.
- Sinden, R.R. (1994). *DNA Structure and Function* (San Diego: Academic Press).
- Terwilliger, T.C., and Berendzen, J. (1999). Automated MAD and MIR structure solution. *Acta Crystallogr. D Biol. Crystallogr.* 55, 849–861.
- Ton-Hoang, B., Guynet, C., Ronning, D.R., Cointin-Marty, B., Dyda, F., and Chandler, M. (2005). Transposition of ISHp608, a member of a novel family of bacterial insertion sequences. *EMBO J.* 55, 849–861. in press.
- Waterston, R.H., Lindblad-Toh, K., Birney, E., Rogers, J., Abril, J.F., Agarwal, P., Agarwala, R., Ainscough, R., Alexandersson, M., An, P., et al. (2002). Initial sequencing and comparative analysis of the mouse genome. *Nature* 420, 520–562.
- Wei, J., Goldberg, M.B., Burland, V., Venkatesan, M.M., Deng, W., Fournier, G., Mayhew, G.F., Plunkett, G., 3rd, Rose, D.J., Darling, A., et al. (2003). Complete genome sequence and comparative genomics of *Shigella flexneri* serotype 2a strain 2457T. *Infect. Immun.* 71, 2775–2786.
- Yu, J., Hu, S., Wang, J., Wong, G.K., Li, S., Liu, B., Deng, Y., Dai, L., Zhou, Y., Zhang, X., et al. (2002). A draft sequence of the rice genome (*Oryza sativa* L. ssp. indica). *Science* 296, 79–92.
- Yuan, J.F., Beniac, D.R., Chaconas, G., and Ottensmeyer, F.P. (2005). 3D reconstruction of the Mu transposase and the Type 1 transpososome: a structural framework for Mu DNA transposition. *Genes Dev.* 19, 840–852.

#### Accession Numbers

TnpA and TnpA-hairpin complex coordinates have been deposited in the Protein Data Bank under accession codes 2A6M and 2A6O, respectively.