The EMBO Journal (2010) 29, 3840–3852 | © 2010 European Molecular Biology Organization | All Rights Reserved 0261-4189/10
www.embojournal.org

THE EMBO JOURNAL

DNA recognition and the precleavage state during single-stranded DNA transposition in *D. radiodurans*

Alison Burgess Hickman[1], Jeffrey A James[1,4], Orsolya Barabas[1,5], Cécile Pasternak[2], Bao Ton-Hoang[3], Michael Chandler[3], Suzanne Sommer[2] and Fred Dyda[1,*]

[1]Laboratory of Molecular Biology, National Institute of Diabetes, Digestive, and Kidney Diseases, National Institutes of Health, Bethesda, MD, USA, [2]CNRS UMR 8621, LRC CEA 42V, Institut de Génétique et Microbiologie, Université Paris-Sud, Orsay Cedex, France and [3]Laboratoire de Microbiologie et Génétique Moléculaires, Centre National de la Recherche Scientifique, Toulouse Cedex, France

Bacterial insertion sequences (ISs) from the IS*200*/IS*605* family encode the smallest known DNA transposases and mobilize through single-stranded DNA transposition. Transposition by one particular family member, IS*Dra2* from *Deinococcus radiodurans*, is dramatically stimulated upon massive γ irradiation. We have determined the crystal structures of four IS*Dra2* transposase/IS end complexes; combined with *in vivo* activity assays and fluorescence anisotropy binding measurements, these have revealed the molecular basis of strand discrimination and transposase action. The structures also show that previously established structural rules of target site recognition that allow different specific sequences to be targeted are only partially conserved among family members. Furthermore, we have captured a fully assembled active site including the scissile phosphate bound by a divalent metal ion cofactor (Cd^{2+}) that supports DNA cleavage. Finally, the observed active site rearrangements when the transposase binds a metal ion in which it is inactive provide a clear rationale for metal ion specificity.

The EMBO Journal (2010) **29,** 3840–3852. doi:10.1038/emboj.2010.241; Published online 1 October 2010
Subject Categories: genome stability & dynamics; structural biology
Keywords: *Deinococcus*; insertion sequence; IS; mobile element; transposition

*Corresponding author. Laboratory of Molecular Biology, National Institute of Diabetes, Digestive, and Kidney Diseases, National Institutes of Health, Bethesda, MD 20892, USA. Tel.: +1 301 402 4496; Fax: +1 301 496 0201; E-mail: fred.dyda@nih.gov
[4]Present address: Office of Technology Transfer, Johns Hopkins University, 100 North Charles Street, 5th Floor, Baltimore, MD 21201, USA
[5]Present address: European Molecular Biology Laboratory, Meyerhofstrasse 1, Heidelberg 69117, Germany

Received: 7 May 2010; accepted: 6 September 2010; published online: 1 October 2010

Introduction

Exposure to 10 gray (Gy) of ionizing radiation is enough to kill a human yet *Deinococcus radiodurans*, perhaps the 'world's toughest bacterium,' has the remarkable ability to survive an instantaneous dose of 5000 Gy without loss of viability (Moseley and Mattingly, 1971). Although this dose effectively shatters the bacterial genome by introducing about 200 double-stranded breaks, within 4 or 5 h following irradiation, *D. radiodurans* has accurately reassembled its genome. This remarkable task is accomplished mainly using a process termed extended synthesis dependent strand annealing, which involves the extensive synthesis of single-stranded DNA (ssDNA) (Zahradka *et al*, 2006; Slade *et al*, 2009).

Like that of essentially all microbes, the genome of *D. radiodurans* contains a variety of mobile elements, and DNA transposition by insertion sequences (ISs) has been shown to be a major source of spontaneous mutagenesis in this bacterium (Mennecier *et al*, 2006). ISs are the simplest form of mobile DNA; they can transpose from one genomic location (donor site) to another (target site) assisted by a self-encoded transposase enzyme that specifically recognizes and synapses the ends of the IS. Synapse formation is followed by a regulated set of DNA cutting and joining reactions that first liberate the transposon from donor DNA, and then integrate the transposon ends into target DNA. All of these events are carried out by the transposase, and the exact mechanism by which these are achieved is dependent on the particular IS family.

When IS activity in *D. radiodurans* was monitored shortly after irradiation, one particular element, IS*Dra2*, stood out as it showed an ~100-fold increase in activity following 10 kGy of γ irradiation, becoming the predominant transposition event (Mennecier *et al*, 2006). This is in contrast to non-irradiated cells where IS*Dra2* insertions are rare. The reason for this increase in IS*Dra2* transposition during genomic recovery from γ irradiation became clear, as IS*Dra2* belongs to the recently described IS*200*/IS*605* family (Kersulyte *et al*, 1998, 2002) that, unlike any other characterized DNA transposition system, uses an asymmetric, ssDNA-dependent pathway for transposition (Guynet *et al*, 2008; Pasternak *et al*, 2010). This pathway has several unusual features. Whereas most ISs contain inverted repeat sequences at the two ends, in the IS*200*/IS*605* family, the transposase recognizes its IS ends by binding to subterminal palindromic sequences that form imperfect hairpin structures (designed IPs, for imperfect palindromes; Figures 1A and 2A). Recognition is also strand specific, as only one strand (the top strand) of the transposon is bound and subsequently transposed. Members of the IS*200*/IS*605* family are also unusual from the point of view of target specificity. Whereas most ISs integrate randomly or with minor sequence preferences, ISs in this family always insert just 3′ of a well-defined tetra- or pentanucleotide sequence (Kersulyte *et al*, 1998; Islam *et al*, 2003;

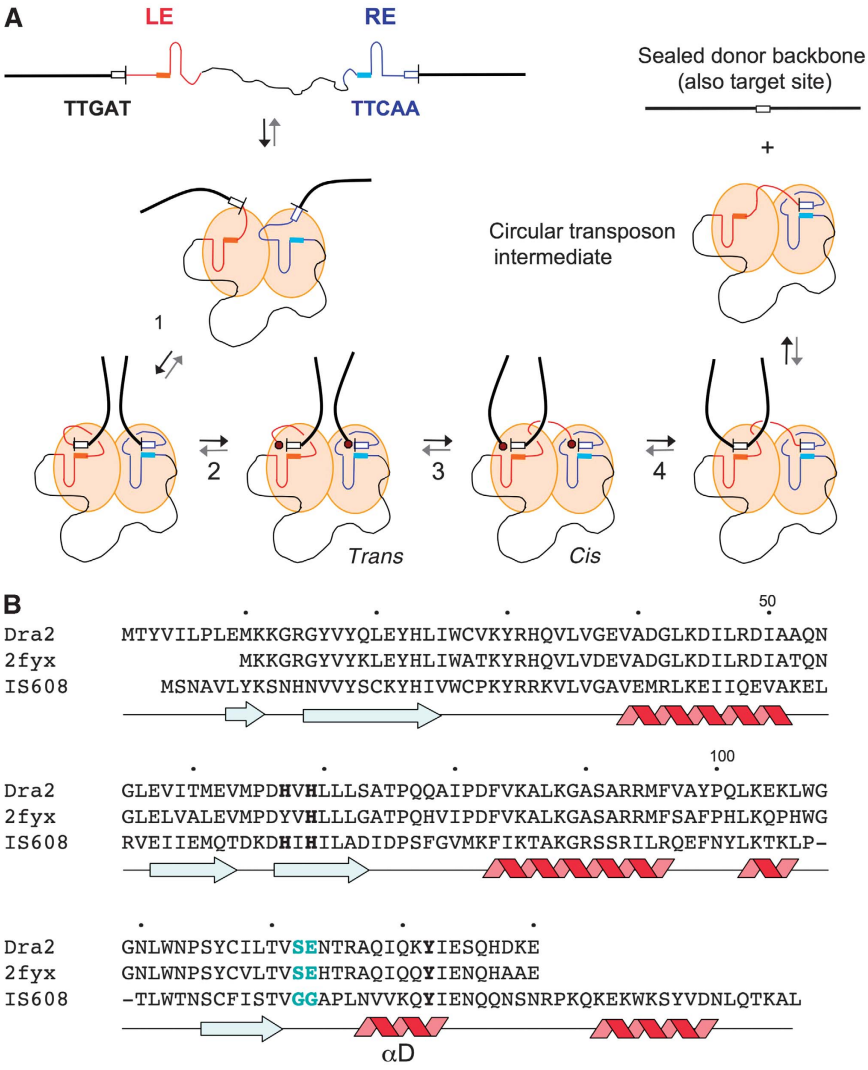

Figure 1 Model of transposon excision by members of the IS*200*/IS*605* family of insertion sequences (ISs) and sequence alignment of some of their transposases. (**A**) The IS is shown as a thin black line with its left end (LE) in red and right end (RE) in blue. The imperfect palindromes (IP) are shown schematically as hairpins, which are flanked on their 5′ sides by solid boxes representing the four nucleotides involved in cleavage (or target) site recognition. The transposase, TnpA, is shown as a dimer with orange subunits. Step 1: TnpA synapses the two IS ends by binding the IPs, and brings the cleavage sites (shown as white boxes) at the IS ends into the active sites through DNA–DNA interactions between the nucleotides 5′ of each end (TTGAT at LE, TTCAA at RE in the case of IS*Dra2*) and those at the 5′ sides of the IPs. For clarity, the two IS ends are shown binding in the same orientation to each monomer of the dimer; however, in the determined structures, one IP is rotated by 180° relative to the other (vide infra). Step 2: End cleavage mediated by the active site Tyr on helix αD results in covalent attachment (represented by a red hexagon) of one TnpA subunit to the cleaved LE and the other subunit to the flanking DNA 3′ of the cleaved RE. Step 3: A *trans*-to-*cis* movement of the αD helices brings the two ends of the donor DNA together in one active site (monomer on left) and the LE into the proximity of the cleaved RE (monomer on right). Step 4: Nucleophilic attack of the two 3′-OH groups onto the covalent phosphotyrosine intermediates results in sealed donor DNA and a circular transposon intermediate. Reversal of these steps (grey arrows) represents insertion of the transposon intermediate into target DNA. (**B**) Sequence alignment of the transposases of IS*Dra2*, the IS represented by PDB code 2fyx, and IS*608* superimposed on the secondary structure (β-strands in blue, α-helices in red) of IS*608* TnpA. The amino acid numbering is for IS*Dra2* transposase. The active site His (on the fourth β-strand) and Tyr residues (on helix αD) are in bold, and residues putatively involved in the *trans*-to-*cis* transition in blue.

http://www-is.biotoul.fr/; and P Siguier, personal communication). The exact sequence of this target site is characteristic of the particular family member.

Although most ISs encode transposase enzymes that are members of the retroviral integrase superfamily (Hickman *et al*, 2010), IS*200*/IS*605* transposases are ssDNA endonucleases belonging to the vast HUH nuclease superfamily. Proteins with HUH nuclease domains are also involved in a diverse variety of other processes such as the initiation of rolling circle replication and conjugative DNA transfer, and replication initiation of certain ssDNA viruses (Koonin and Ilyina, 1993). HUH nucleases contain one or two active site tyrosine residues that act as nucleophiles during DNA cleavage. Upon nucleophilic attack, the transposase becomes covalently attached to DNA through a 5′-phosphotyrosine intermediate, and a free 3′-OH group is released on the other side of the DNA break. DNA cleavage also requires a divalent metal ion cofactor that is coordinated by the two

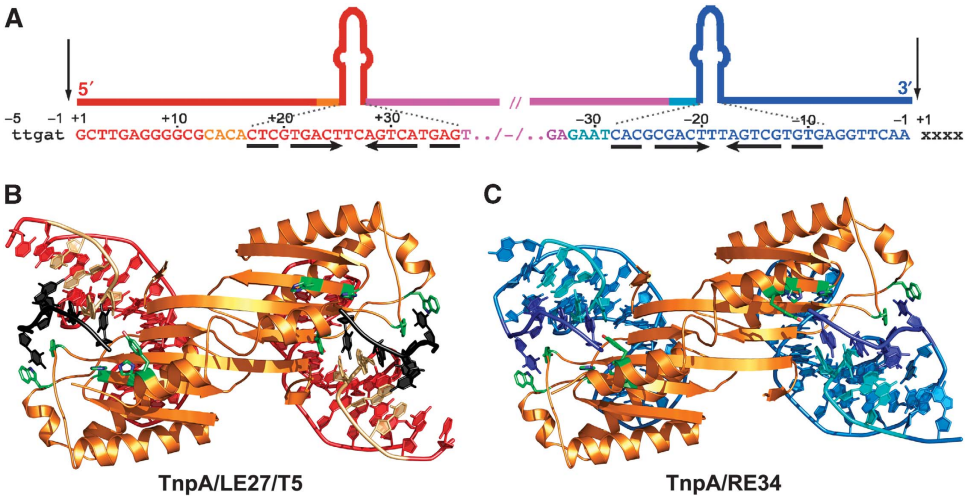

Figure 2 Structures of $TnpA_{Dra2}$ bound to its transposon ends. (**A**) IS*Dra2* LE and RE DNA sequences. The overall structures of the $TnpA_{Dra2}$ dimer bound to (**B**) the LE IP (in red and orange) and a short ssDNA (in black) representing the five nucleotides 5′ of the LE cleavage site, which is equivalent to the target site, and (**C**) the RE IP (in light and dark blue) extended to the transposon 3′ end. The DNA colour scheme corresponds to that shown in Figures 1A and 2A. Residues of $TnpA_{Dra2}$ highlighted in green are specifically referred to in the text (Y30, H67, H69, W107, and Y132).

histidine residues that comprise the HUH motif (Datta *et al*, 2003; Guasch *et al*, 2003). DNA strand transfer is achieved when a 3′-OH end of a DNA strand attacks the 5′-phosphotyrosine linkage, resulting in a resealed phosphodiester backbone.

A mechanistic model of ssDNA transposition in the IS*200*/IS*605* family has been derived from biochemical experiments and structural data obtained for one of its members, the transposase TnpA of IS*608* from *Helicobacter pylori* (Ronning *et al*, 2005; Ton-Hoang *et al*, 2005; Barabas *et al*, 2008; Guynet *et al*, 2008). As IS*200*/IS*605* transposases contain only one active site tyrosine (Figure 1B), in order to coordinate cleavage at their left end (LE) and right end (RE), they assemble as dimers; in this way, there are two functional active sites and two binding sites for IP hairpins, one for LE and another for RE (shown schematically in Figure 1A).

Studies on IS*608* TnpA revealed two important keys to transposition. The first is that the cleavage sites at both the LE and RE are recognized through non-sequential base-pairing interactions between nucleotides at the cleavage site (shown as a white box at either end in Figure 1A) and four nucleotides at the 5′ base of the IPs (shown as solid orange and light blue boxes). As the sequences of the 4-nt 5′ extensions of the IPs are different at LE and RE, specific recognition of the different LE and RE cleavage site sequences becomes possible. Furthermore, as the circular transposon intermediate to be integrated contains the LE IP and its 4 nt 5′ extension, target site recognition is mechanistically identical to the recognition of the LE cleavage site. The second key is that, after cleavage at LE and RE (step 2 in Figure 1A), a conformational change is believed to occur that swaps the 5′-phosphotyrosine linkages between the two monomers (step 3) in the dimer through the exchange of a mobile α-helix. Resolving these linkages (step 4) yields a circular transposon and a resealed donor DNA strand. Integration occurs when the transposase dimer bound to the circular intermediate captures an ssDNA that contains the target sequence, and the reaction cycle repeats: two cleavages, one swap, and two resolution steps later, the result is an integrated transposon.

Although we now know a great deal about the mechanism of IS*608* transposition, some aspects cannot be straightforwardly generalized to other family members. For example, the IS*608* target sequence is a tetranucleotide (TTAC) whereas the target sequence of IS*Dra2* is a pentanucleotide (TTGAT; Islam *et al*, 2003). It is not clear how the recognition rules derived from the IS*608* structures, appropriate for tetranucleotides, would allow recognition of a pentanucleotide. As the mechanism determined for IS*608* TnpA can be exploited to redirect insertions to preselected and specified tetranucleotide sites (Guynet *et al*, 2009), understanding how longer target sequences are recognized may prove extremely useful. Furthermore, the discrimination between the top- and bottom-strand IPs of IS*608* TnpA is influenced by nucleotides present at the tip of the hairpin and an extrahelical base in the stem of the IP; there is no such base in the stem of the IPs of IS*Dra2* leaving open the question of how the two strands are distinguished. Finally, certain amino acids that we concluded should be crucial for IS*608* TnpA transposition are simply not found in the IS*Dra2* transposase.

We describe here a series of crystal structures of the IS*Dra2* transposase bound to a variety of ssDNA substrates representing either the LE or RE of IS*Dra2*. In conjunction with biochemical and *in vivo* experiments, these structures expand our understanding of how IS*200*/IS*605* transposases work. In particular, we have captured the transposase with a fully assembled active site including the scissile phosphate, a state that eluded us with IS*608* TnpA. We also provide structural explanations for pentanucleotide recognition, strand discrimination, and the metal ion specificity of DNA cleavage.

Results

We have determined crystal structures of four complexes of the IS*Dra2* transposase ($TnpA_{Dra2}$) with various bound DNAs

Table I Crystallographic structure determination and refinement

	TnpA/LE27/T5 (native)	TnpA/LE27/T5 (I derivative)	TnpA/RE34	TnpA/Cd	TnpA/Zn
Data collection					
Space group	$P2_12_12_1$	$P2_12_12_1$	$P2_1$	$P2_12_12_1$	$P2_12_12_1$
Cell dimensions, *a*, *b*, *c* (Å)	104.7, 128.2, 140.6	104.4, 128.6, 140.1	76.8, 89.3, 90.1	50.3, 86.9, 128.3	49.8, 92.7, 126.1
β (deg)			111.2		
Resolution (Å)	2.3	2.4	2.3	1.9	1.9
Total reflections	575 978	515 046	181 638	333 244	327 763
Unique reflections	84 171	74 480	50 130	44 818	45 279
$I/\sigma(I)$	15.9 (2.4)	11.2 (2.5)	16.4 (3.1)	13.4 (6.91)	16.4 (4.5)
Completeness (%)	99.5 (97.6)	100.0 (99.8)	99.6 (99.0)	99.8 (99.1)	96.7 (75.9)
R_{sym}	0.07 (0.83)	0.103 (0.716)	0.049 (0.341)	0.093 (0.186)	0.048 (0.244)
SIRAS phasing at 3.0 Å					
R_{Cullis}		0.64			
R_{Kraut} (iso/ano)		0.163/0.203			
Phasing power (iso/ano)		1.49/1.9			
Combined FOM		0.402			
Molecular replacement at 4.0 Å					
Rotation function corr. coeff.			0.209/0.204	0.146	0.137/0.127
Rfactor after rigid body refinement			0.417	0.423	0.368
Refinement					
Resolution (Å)	30.0–2.3		30.0–2.3	30–1.9	30.0–1.9
No. reflections ($F \geqslant 0$)	78 520		47 910	44 616	44 429
R_{work}/R_{free}	0.218/0.249		0.215/0.239	0.185/0.214	0.177/0.210
No. atoms					
Protein and nucleic acid	10 055		7236	3447	3444
Solvent	548		164	401	473
B-factors					
Protein and nucleic acid	39.34		64.63	27.07	22.82
Solvent	38.18		45.30	38.76	33.49
R.m.s.d.					
Bond lengths (Å)	0.009		0.01	0.017	0.013
Bond angles (deg)	4.1		3.00	1.94	4.54
Dimers per AU	3		2	1	1
Ramachandran plot statistics					
Most favoured	92.1%		92.7%	93.1%	94.4%
Disallowed	0		0	0	0

representing several stages of the transposition reaction (Table I). Recent biochemical work has confirmed that $TnpA_{Dra2}$ uses ssDNA substrates in *in vitro* transposition reactions (Pasternak *et al*, 2010), as does the IS*608* transposase. One structure (Figure 2B) is of a ternary complex (designated TnpA/LE27/T5) in which $TnpA_{Dra2}$ is complexed with an oligonucleotide containing the LE IP (designated LE27, from LE nt +11 to +37; Figure 2A) and a 5-nt oligonucleotide (T5) comprising the target site, TTGAT. We have also solved a complex (designated TnpA/RE34) between $TnpA_{Dra2}$ and a 34-nt oligonucleotide that contains the RE IP and extends to the RE cleavage site (RE nt −34 to −1; Figure 2C). In addition, we have solved two other LE ternary complexes. TnpA/Cd is the complex between $TnpA_{Dra2}$ Y132F (the active site Tyr has been mutated to Phe), LE27, and an oligonucleotide in which the target site has been extended by 1 nt, TTGAT↓G, to include the first 5′ nt of the transposon (designated T5↓G where ↓ indicates the cleavage site). TnpA/Cd was crystallized in the presence of Cd^{2+}, a divalent metal ion that robustly supports cleavage. The last complex, TnpA/Zn, contains wild-type TnpA and the same oligonucleotides as TnpA/Cd, but was crystallized in the presence of Zn^{2+} which supports only a very low level of cleavage. TnpA/LE27/T5 was solved with experimental phasing; the other structures were solved with molecular replacement using TnpA/LE27/T5 as the search model. Representative electron density is shown in Supplementary Figure S1.

IP recognition by $TnpA_{Dra2}$

A characteristic feature of transposon DNA sequences in the IS*200*/IS*605* family are subterminal IP sequences located close to the cleavage sites. As only the top strand is cut and subsequently moved to a new genomic location, the transposase must be able to distinguish the top-strand IP hairpins from those on the bottom. The previously determined structures of IS*608* TnpA (Ronning *et al*, 2005) demonstrated that one important discriminating feature is the tip of the hairpin loop, where the 5′-most T at the tip (shown in cyan in Figure 3A) sits in a pocket on the transposase surface which would not be able to accommodate the corresponding adenine that is present at the tip of the bottom-strand IP

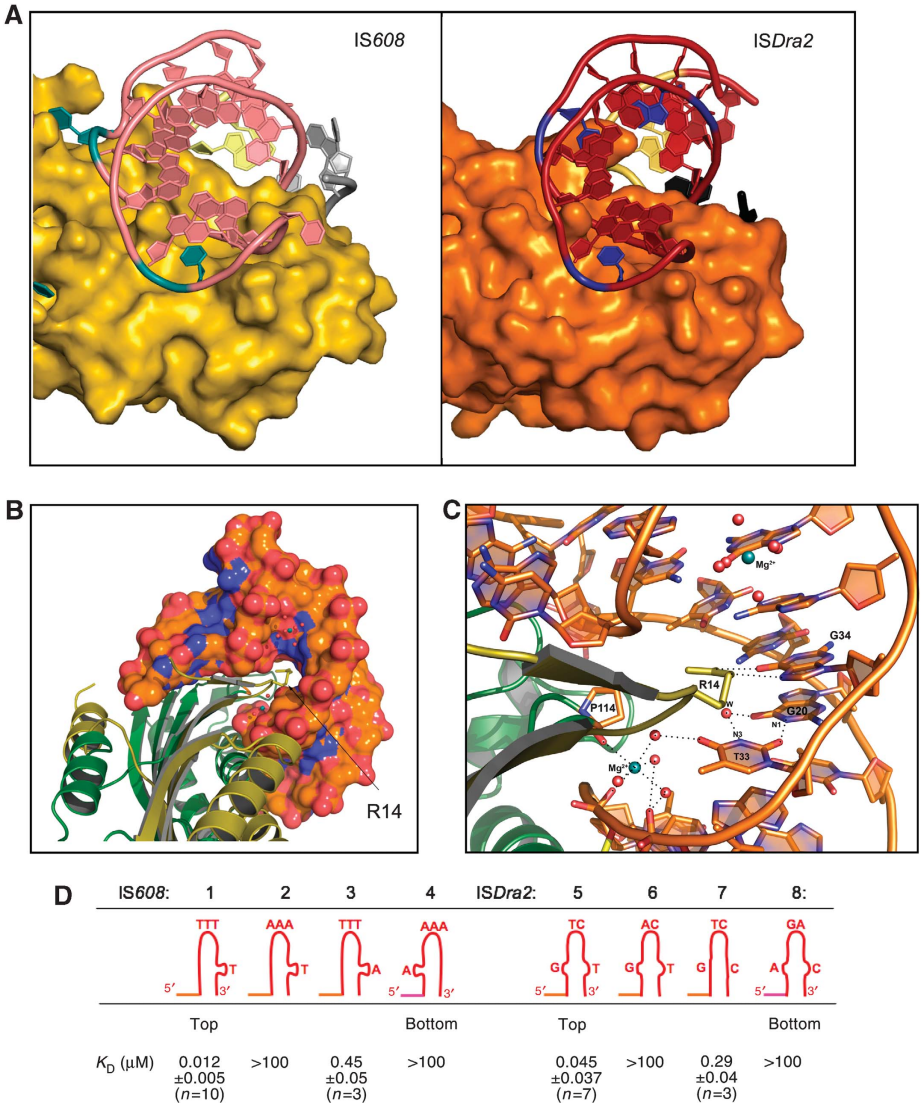

Figure 3 DNA binding by $TnpA_{Dra2}$: G:T mismatch recognition and discrimination between the top and bottom strand of the transposon. (**A**) Comparison of the binding of the tip of the hairpin loop by IS*608* TnpA (left; PDB code 2A6O) and $TnpA_{Dra2}$ (right). The distinguishing features of the hairpins are shown in shades of blue: the unpaired T in the stem of the IS*608* hairpin and the T at the hairpin tip (left), and the mismatched G:T base pair and the T at the hairpin tip of the IS*Dra2* hairpin (right). (**B**) Space-filling representation of the complex between $TnpA_{Dra2}$ (two subunits shown in shades of green) and LE27 (in orange) highlighting the position of Arg14. (**C**) A web of hydrogen-bonding interactions and a bound Mg^{2+} ion mediate the interaction between Arg14 and the G + 20:T + 33 mismatch on the LE. (**D**) Dissociation constants for binding between the IS*608* and IS*Dra2* transposases and their IPs and variants determined by fluorescence anisotropy measurements.

hairpin. The structures of $TnpA_{Dra2}$ bound to the LE IP or the RE IP indicate that this element of discrimination is retained: in all structures, one thymine at the hairpin tip (shown in blue in Figure 3A) is flipped back into a small pocket where it stacks against the peptide plane of Gly89.

The other IP-discriminating feature for IS*608* TnpA is an unpaired thymine within the IP stem, which is flipped out and is specifically recognized by the transposase (also shown in cyan in Figure 3A). This structural feature is not conserved in IS*Dra2*. Instead, each IP stem of IS*Dra2* contains a G:T mismatch (G + 20:T + 33 on the LE, G−25:T−12 on the RE; Figure 2A) that is recognized and apparently stabilized by $TnpA_{Dra2}$.

The key amino acid in the recognition of the G:T mismatch is Arg14, located at the tip of a β-hairpin that dives deep into the major groove and stacks on the top of the mismatched T (shown for T + 33 on LE in Figure 3B and C) while forming hydrogen bonds to the immediately adjacent G (G + 34 on LE). These interactions twist the DNA such that the Watson–Crick faces of the mismatched nucleotides are turned away from each other thereby avoiding a clash between the protonated N1 of G + 34 and the protonated N3 of T + 33. A consequence of these DNA distortions is the accumulation of negative charge on one side of T + 33 as O4 of T + 33 moves toward the sugar-phosphate backbone of the opposite strand. This electrostatic problem is thwarted by the creation of a Mg^{2+}-binding site (Figure 3C). The octahedrally coordinated Mg^{2+} ion also provides a hydrogen bond to O4 of T + 33 through one of its coordinated water molecules. It seems likely that this element of mismatch recognition

also contributes to discrimination between the top and bottom strands as such a water-mediated H-bond could not be formed on the bottom-strand IP: with the C:A mismatch of the bottom-strand IP, the Mg^{2+}-bound water molecule cannot act as an H-bond acceptor with the protonated N4 of the cytosine. In all four $TnpA_{Dra2}$/DNA complexes, the mode of recognition and stabilization of the G:T mismatch is essentially identical.

Relative to the IP, the location of the G:T mismatch in the IS*Dra2* stem is exactly where the unpaired base yielding the flipped out thymine is for IS*608*. The flipped out T is recognized by IS*608* TnpA by sequestration between the side chains of Arg52 and Phe75, and by a set of H-bonds (Ronning *et al*, 2005), whereas the crucial residue for recognizing and stabilizing the G:T mismatch of IS*Dra2* is Arg14. None of these residues are conserved between the two proteins (Figure 1B).

One final factor that influences IP recognition and strand discrimination by $TnpA_{Dra2}$ is the relative ease of hairpin formation by the top- and bottom-strand IPs. Whereas the top-strand IPs on both the LE and RE readily fold into hairpin structures (as assessed by size-exclusion chromatography; data not shown), attempts to form a hairpin with the bottom-strand IS*Dra2* LE IP sequence were unsuccessful due to its vast preference to form annealed dsDNA products. Even when the small amount of the bottom-strand hairpinned form was chromatographically separated from the dsDNA form, the former rapidly converted to the latter, even at 4°C. Indeed, thermodynamic data suggest that whereas G:T mismatches are well tolerated in dsDNA oligonucleotides, C:A mismatches are less so (SantaLucia and Hicks, 2004).

Both the hairpin tip and the stem IP irregularity are important for IP recognition

The structures of IS*608* TnpA and $TnpA_{Dra2}$ bound to their IPs suggest that two IP features are important for recognition: the bases at the hairpin tip and an imperfection in the hairpin stem. To investigate the relative influence of these, we determined dissociation constants for IP binding both to IS*608* TnpA and $TnpA_{Dra2}$ using fluorescence anisotropy with various carboxyfluorescein (FAM)-modified oligonucleotides.

As shown in Figure 3D, modification of either feature resulted in a decrease in binding. The effect was most dramatic when pyrimidine bases at the tip of the IS*608* and IS*Dra2* IPs were changed to purines; for both transposases, this resulted in a complete loss of detectable IP binding (compare IPs #1–#2, and #5–#6). This is consistent with structural observations that the 5′-most T at the hairpin tips sit in a pocket located on each transposase surface too small for a larger purine base.

The imperfections in the IP stems appear to be less important contributors to binding than the bases at the hairpin tips. For IS*608* TnpA, changing the flipped out T in the IP stem to A (IP #3) reduced binding ~40-fold relative to the wild-type IP, whereas correction of the mismatch in the stem of the IS*Dra2* IP (IP #7) led to an approximately six-fold reduction in binding.

We also tested the ability of the two transposases to bind to their bottom-strand IPs (IP #4 and #8). There was no detectable binding in either case, demonstrating that the discrimination between the top and bottom strand of the transposon occurs at the initial stage of transposase binding.

The role of $TnpA_{Dra2}$ Arg14

To evaluate the role of Arg14 *in vivo*, we used a recently developed genetic system in *D. radiodurans* (Pasternak *et al*, 2010) to measure the first stage in IS transposition, excision of the IS from the donor backbone. In this system, the unique active copy of IS*Dra2* has been replaced by a mutated derivative, IS*Dra2*-113, that is inserted at the unique 5′-TTGAT-3′ target site present in the *tetA* gene of a Tet^R cassette; in this way, precise excision of the IS restores a Tet^R phenotype. In IS*Dra2*-113, both *tnpA* encoding the transposase and a second gene, *tnpB*, carried by the IS but which is not necessary for transposition (Pasternak *et al*, 2010) have been deleted, and the functional ends for transposition have been retained. $TnpA_{Dra2}$ is expressed *in trans* from a plasmid expression vector.

As shown in Table II, precise excision of IS*Dra2*-113 was strictly dependent on the presence of $TnpA_{Dra2}$. In strains devoid of plasmid or carrying an empty vector, the frequency of Tet^R colonies was undetectable, but increased to $\sim 3 \times 10^{-3}$ when wild-type $TnpA_{Dra2}$ was expressed. When wild-type $TnpA_{Dra2}$ was replaced by the R14A mutant, the excision frequency dropped ~60-fold, supporting a crucial role of Arg14 in IP recognition and subsequent transposon excision.

We also measured the binding affinity of purified $TnpA_{Dra2}$ R14A protein to its IP using our fluorescence anisotropy assay. The dissociation constant was ~1.4 μM, an ~30-fold loss in binding relative to wild type, which is consistent with the corresponding decrease in excision frequency.

Helix αD is a mobile structural element that can be in cis or trans

A key feature of the molecular model of ssDNA transposition is a large conformational rearrangement that exchanges the location of the two αD helices containing the tyrosine nucleophiles between the two monomers of the transposase dimer (Figure 1A, step 3). Indeed, biochemical data strongly suggest that while a *trans*-active site arrangement (i.e. the tyrosine of one monomer is close to the HUH motif of the other) is catalytically competent for transposon end cleavage, the resolution steps require a *trans*/*cis* isomerization (Barabas *et al*, 2008). However, we have not yet observed a IS*200*/IS*605* family transposase in the *cis* configuration. In all the structures of IS*608* TnpA determined to date, the active sites are arranged *in trans* and $TnpA_{Dra2}$ is no different. Nevertheless, there does appear to be a precedent for a

Table II Excision frequencies of the IS*Dra2*-113 derivative in a *Deinococcus radiodurans* tester strain[a] expressing $TnpA_{Dra2}$ variants

Resident plasmid	$TnpA_{Dra2}$ expressed *in trans*	Excision frequency
No	No	$< 10^{-8}$
pGY11559	No	$< 10^{-8}$
pGY13203	TnpA wt	$3.2\ (\pm 0.6) \times 10^{-3}$
pGY13521	$TnpA_{R14A}$	$5.3\ (\pm 1.6) \times 10^{-5}$
pGY13522	$TnpA_{S122G/E123G}$	$1.7\ (\pm 0.3) \times 10^{-3}$

[a]GY13115 genotype: Δ*DR1652*Δ*DR1651*Ω*tetA*ΩIS*Dra2*-113 (Cam^R) (Pasternak *et al*, 2010).

cis-active site. The coordinates of a high resolution crystal structure of an uncomplexed IS*200*/IS*605* family transposase that is 80.8% identical to $TnpA_{Dra2}$ (Figure 1B) have been deposited in the Protein Data Bank under the PDB ID 2fyx, and the protein is in the *cis* conformation. As the 2fyx transposase is a result of structural genomics efforts, unfortunately no biochemical data are available, and it must be noted that the HUH motif is not intact.

For the *trans*-to-*cis* structural transition to take place, modelling suggested that a seven amino acid long loop just upstream of helix αD must change its structure substantially. For IS*608* TnpA, this seems feasible as the loop contains two adjacent Gly residues, Gly117 and Gly118, that can act as pivots as the main chain dihedral angle of glycine is less restricted than that of all other amino acids. However, in $TnpA_{Dra2}$ and in the protein represented by 2fyx, the equivalently placed residues are Ser and Glu and there are no glycines in the seven-residue loop (Figure 1B), raising the question of how well these residues might permit the conformational change.

We therefore examined the effect *in vivo* of mutating these two residues to glycine. Expression of the S122G/E123G mutant $TnpA_{Dra2}$ gave approximately half the excision frequency observed for wild-type $TnpA_{Dra2}$ protein (Table II). Thus, despite their bulkier side chains, Ser122 and Glu123 do not appear to impede a rate-limiting step of transposition and—despite our expectations—mutation of both to Gly does not lead to an increase in activity.

Examination of the main chain torsion angles of the seven-loop residues of the 2fyx protein and $TnpA_{Dra2}$ reveals that all the residues are in the most favoured regions of the Ramachandran plot (Supplementary Figure S2). However, it appears that in going from the *cis* conformation (2fyx) to the *trans* conformation (all the $TnpA_{Dra2}$ structures), the main chain dihedrals of two sequential residues change from one allowed phi/psi region to another, further supporting the notion that a *trans*/*cis* isomerization is structurally reasonable.

Both transposase and transposon DNA contribute to $TnpA_{Dra2}$ target and cleavage site recognition

Perhaps the most surprising aspect of the mechanism of ssDNA transposition carried out by IS*608* TnpA is that cleavage sites are not recognized by the transposase directly, but rather by non-sequential base pairing between the transposon 5′ extensions of the IPs and sequences just 5′ of the cleavage sites (Barabas *et al*, 2008). How well are these features of DNA recognition and cleavage recapitulated for IS*Dra2*? Figure 4A and B show cleavage site recognition of the LE and RE by $TnpA_{Dra2}$. In contrast to IS*608* TnpA, $TnpA_{Dra2}$ combines both DNA–DNA and protein–DNA interactions to recognize its cleavage sites. Four of the 5 nt just 5′ of the cleavage site (e.g. T−4, G−3, A−2, T−1 on the LE) are recognized by the 4-nt 5′ extension of the IP, the fifth (T−5) is recognized by $TnpA_{Dra2}$ itself. The pattern of recognition echoes that of IS*608*, that is, the non-linear arrangement of base pairs is the same (see inset in Figure 4). On the other hand, T−5 is sandwiched between Tyr30 and Trp107, and forms hydrogen bonds with surrounding residues including His32 and main chain atoms of Lys106 and Trp107. Neither Trp107 nor His32 is conserved between IS*608* TnpA and $TnpA_{Dra2}$, and the transposase structures differ in this region due to a shift of 5–6 Å in the position of a short α-helix, which in $TnpA_{Dra2}$ is between Gln101 and Trp107 and contains residues that interact with T−5.

The first view of a fully assembled active site

To date, none of the structures obtained for IS*200*/IS*605* family transposases have shown the nuclease catalytic site in an active configuration, and we suspect there are two reasons for this. First, helix αD that contains the nucleophilic tyrosine is extremely mobile—as it presumably must be—as it is proposed to move from the *trans* to *cis* conformation to complete strand transfer. Correspondingly, in the IS*608* TnpA structures helix αD does not pack tightly against the body of the molecule, and small forces such as those resulting from lattice packing appear to influence its position. A similar phenomenon is observed here for $TnpA_{Dra2}$. For example, in the TnpA/LE27/T5 ternary complex, there are three dimers in the crystallographic asymmetric unit related to each other mainly by non-crystallographic translations (Supplementary Figure S3). Although the three $TnpA_{Dra2}$ transposase dimers and bound oligonucleotides are essentially structurally identical, there are major differences in the positions of the six αD helices. Simulated annealed omit electron density is clear for all of them, indicating that the catalytically important helix αD can be in several distinct positions depending on the crystal packing environment.

The second reason why we have not previously observed a fully assembled active site is that the transposase/DNA structures to date have captured a mechanistically inconsistent state of the system. That is, whereas the product of the cleavage reaction is a free 3′-OH and a phosphotyrosine intermediate, the structures have all contained the 3′-OH group in the active site but without a corresponding phosphotyrosine intermediate or even the scissile phosphate.

Our attempts to obtain 'mechanistically consistent' complexes with IS*608* TnpA have been unsuccessful, largely due to poor biophysical properties of the resulting complexes. However, the IS*Dra2* TnpA/LE27/T5 structure suggested an alternate approach. As seen in Figure 4A, the 3′-end base of the target sequence, T−1, is base paired to A+16 which in turn is stacked on top of C+15. As the base of C+15 is not engaged in any hydrogen bonding, the structure suggested that if the target sequence were extended by 1 nt in the 3′ direction (corresponding to the first nt of the transposon on LE), a base pair might form between C+15 and this additional nt, thereby perhaps holding the scissile phosphate in the appropriate place in the active site.

In an attempt to capture the state in the transposition pathway just prior to LE cleavage (which is mechanistically equivalent to target cleavage), we used $TnpA_{Dra2}$ with a Y132F mutation bound to LE27 and T5↓G. The best diffracting crystals were obtained in the presence of Cd^{2+}, which was subsequently found to be bound at the active site (Figure 5A). In the structure, the entire T5↓G substrate is visible including G+1 (which is indeed base paired to C+15) and the scissile phosphate. Cd^{2+} is octahedrally coordinated by His69 and His67 of the HUH motif, Gln136 on helix αD, Sδ of Met64, a water molecule, and one of the non-bridging oxygens of the scissile phosphate. His23 and Glu66 in the second coordination shell of the metal ion presumably keep His69 and His67 in tautomerization states in which the Nδ atoms are protonated so that they can ligand

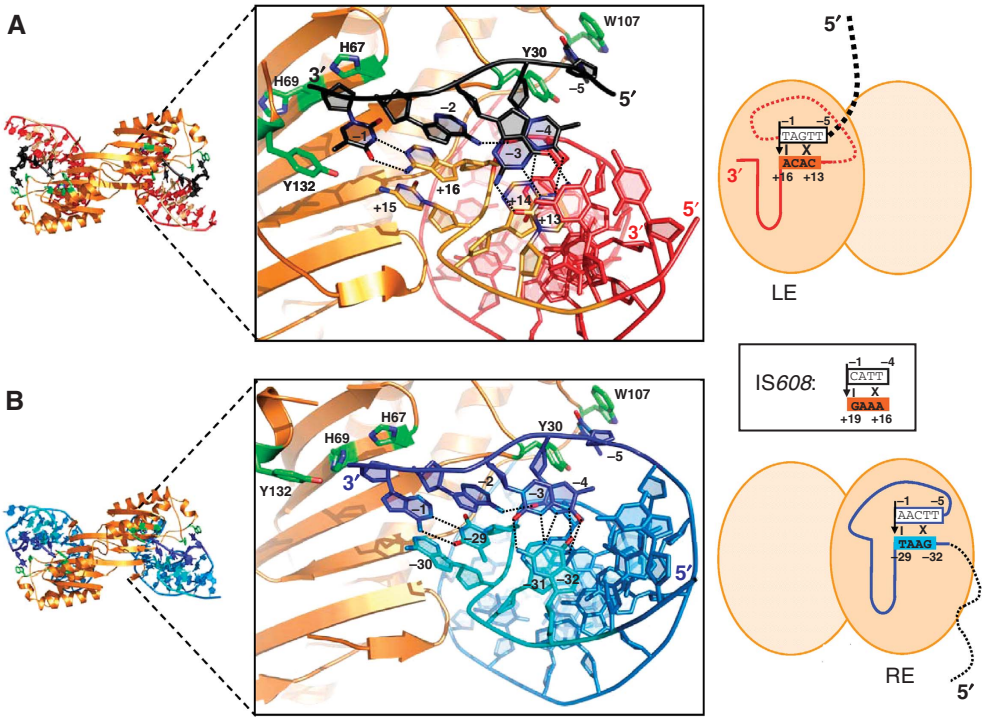

Figure 4 Transposon end and cleavage site recognition by $TnpA_{Dra2}$. On the left are the overall $TnpA_{Dra2}$ dimer structures corresponding to the complexes TnpA/LE27/T5 (**A**) and TnpA/RE34 (**B**). In the middle are close-up views of the active sites. The cartoons on the right show the state during the transposition pathway captured by the crystal structures, corresponding to the model in Figure 1A. In these cartoons, the vertical black arrows correspond to the sites of cleavage and the pattern of base-pairing interactions at the cleavage sites at LE (**A**) and RE (**B**) are indicated by the black lines. The inset box shows the corresponding pattern of base-pairing interactions for IS*608* (Barabas *et al*, 2008). The dashed lines indicate portions of DNA not present in the structures.

the metal through their Nε atoms. The position of His23 is such that it could form a hydrogen bond to the next phosphate group if T5↓G were extended by yet another nt into the transposon end, thus expanding the hydrogen-bonding network.

When the position of Tyr132 was modelled based on the observed position of Phe132, it became clear that the oxygen atom of the OH group would be ~2.6 Å from the phosphorus atom of the scissile phosphate in-line for a nucleophilic attack. To confirm that this structure represents a catalytically relevant state, we evaluated the ability of various divalent metal ions to support LE cleavage activity of the wild-type protein. As shown in Figure 5B, Mg^{2+}, Mn^{2+} and Cd^{2+} all support LE cleavage, suggesting that the TnpA/Cd complex represents the precleavage state.

Zn^{2+} does not support DNA cleavage

We also obtained crystals of $TnpA_{Dra2}$ bound to LE27 and T5↓G in the presence of Zn^{2+}. $TnpA_{Dra2}$ is only slightly active in 2 mM Zn^{2+} (Figure 5B) and increasing the zinc concentration to 25 mM does not lead to increased activity (data not shown). Although this divalent metal ion occupies the active site (Figure 5C), we observe a rearrangement of ligands such that Zn^{2+} is tetrahedrally coordinated, its preferred coordination state (Dudev and Lim, 2003). Three of the four ligands are those that participated in Cd^{2+} binding: His69 and His67 of the HUH motif and a non-bridging oxygen atom of the scissile phosphate. However, completing the tetrahedral arrangement is His137. Gln136, which in the Cd^{2+} structure provides a metal ligand, is now hydrogen bonded to the second non-bridging oxygen atom of the scissile phosphate. The net effect of Zn^{2+}'s apparent preference for His137 as a ligand is a positional shift of helix αD such that Tyr132, the active site nucleophile, is pushed ~6 Å away from the scissile phosphate providing a straightforward explanation for the lack of cleavage activity in the presence of Zn^{2+}.

Discussion

To probe the mechanism of transposition catalysed by members of the prokaryotic IS*200*/IS*605* family, we have been investigating *D. radiodurans* IS*Dra2*, an element that presents a number of contrasts to the well-characterized *H. pylori* IS*608*. We were initially drawn to the IS*200*/IS*605* family as it possesses the smallest known transposases and, as our initial studies revealed, they are single-domain proteins that carry out all of the chemical steps of transposition without the need for any accessory host proteins. $TnpA_{Dra2}$ pushes the transposase size requirement even further as it does not have the final C-terminal α-helix found in IS*608* TnpA and consists of a mere 140 amino acids. Although the IS*608* TnpA and $TnpA_{Dra2}$ transposases can be superimposed with a 1.4-Å r.m.s.d. over 123 Cα positions, indicating that they are structurally very similar despite only ~34% sequence identity, our structural data suggest that amino acid diversity gives this family of proteins the flexibility to recognize and stabilize different DNAs, and to integrate into target sites of differing length.

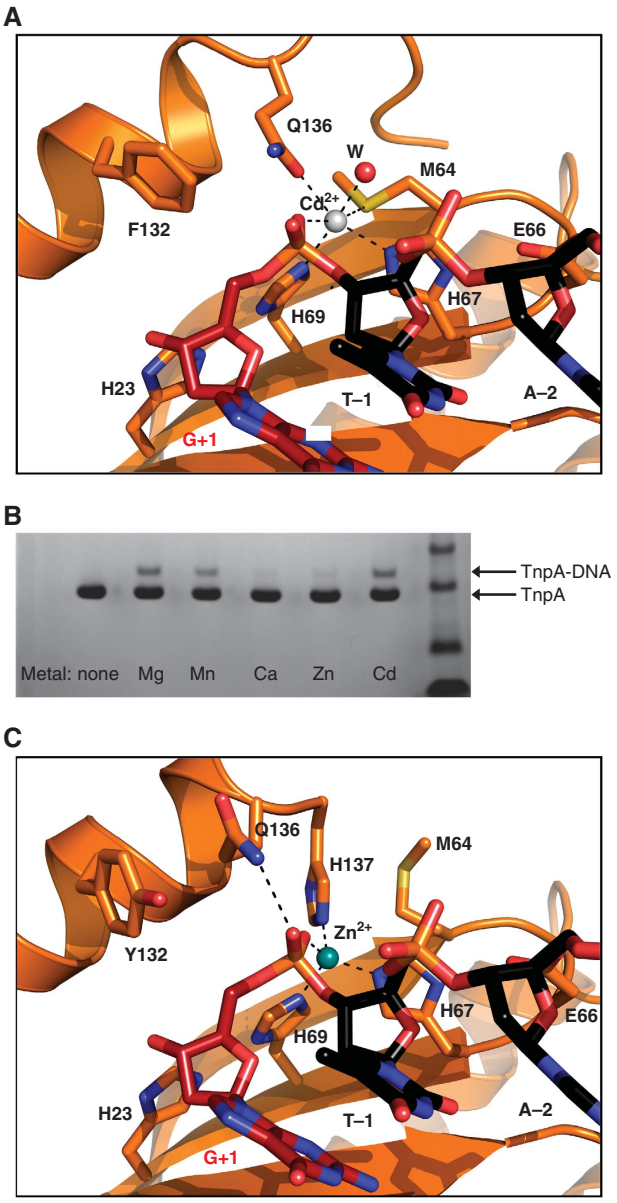

Figure 5 The $TnpA_{Dra2}$ active site is fully assembled in the presence of Cd^{2+} but is distorted by Zn^{2+}. (**A**) The structure of the complex of $TnpA_{Dra2}$ Y132F in the presence of Cd^{2+} (white sphere) bound to LE27 (not visible) and an oligonucleotide, T5↓G, that traverses the LE cleavage site where target nucleotides are shown in black and the first 5′ transposon nucleotide in red. (**B**) $TnpA_{Dra2}$ is fully active for LE cleavage in the presence of Cd^{2+} but not Zn^{2+}. Cleavage of a 13-mer target sequence results in a covalent complex between the product and $TnpA_{Dra2}$, and is detected by Coomassie-stained SDS–PAGE as the formation of a higher molecular weight band. (**C**) The active site of $TnpA_{Dra2}$ is distorted in the presence of Zn^{2+} (blue sphere), explaining the lack of cleavage activity.

One of the significant contrasts between these two family members is that IS*Dra2* specifically integrates into a pentanucleotide target sequence (TTGAT), whereas IS*608* is targeted to a different sequence (TTAC). We previously established that the IS*608* target site is bound to IS*608* TnpA through base-pairing interactions with an internal segment of transposon DNA (Barabas *et al*, 2008). By analogy, one rather obvious structural solution to selecting a 5-nt target site might have been to use a 5-nt long internal segment of DNA for recognition. However, this is not what occurs for IS*Dra2*. Rather, the tetranucleotide-to-tetranucleotide base-pairing solution is retained as a subset of the basis of recognition, and the fifth nt is specifically recognized by the protein.

This means that for the IS*Dra2* target sequence 5′-TTGAT, the internal segment of DNA that is used for recognition is 3′-ACAC, and the pattern of base pairing seen for IS*608* is maintained. The fifth nt, T−5, is sandwiched in between residues Tyr30 and Trp107, and forms base-specific contacts to three nearby residues with a hydrogen-bonding pattern that suggests a strong preference for T at this position. Satisfyingly, a T is also present at the −5 position at the RE cleavage site.

One consequence of this mode of pentanucleotide recognition is that it indicates that the notion of being able to redirect IS*200*/IS*605* family members to even longer target sequences is unlikely to be as straightforward as expanding the base-pairing complementarity between the target site and the 5′ 4-nt extension. Rather, whereas the complementarity approach works well for redirecting to a 4-nt target sequence (Guynet *et al*, 2009), incorporating a fifth nucleotide into the IS*Dra2* target site requires special accommodation in the form of a specific protein-binding pocket. This explains the observation that IS*200*/IS*605* family members appear to be targeted generally only to tetra- and pentanucleotide sequences, and implies that to accommodate an even longer target site would require the dedication of even more of the limited $TnpA_{Dra2}$ protein surface to the task.

A second point of contrast between IS*608* and IS*Dra2* is that the IPs at the transposon ends differ in their imperfections. Whereas the IPs at the ends of IS*608* have an unpaired nt within the hairpin stem, those of IS*Dra2* have a mismatched base pair. Our fluorescence anisotropy binding data indicate that these imperfections do not drive DNA end binding but, rather, they enhance binding and presumably account in part for strand discrimination. In two different approaches to selectivity, the two related transposases use non-conserved amino acids for hairpin imperfection recognition. For IS*608*, the unpaired T is bound within a precisely contoured pocket on the surface of TnpA, whereas the G:T mismatch of IS*Dra2* is dealt with by a set of interactions organized by an Arg residue of $TnpA_{Dra2}$.

Although there are a number of enzymes that recognize mismatched bases as part of their function, the solution used by $TnpA_{Dra2}$ appears unique. The G:T mismatch does not form a wobble base pair as seen, for example, in the case of the Vsr endonuclease bound to DNA (Tsutakawa *et al*, 1999). Instead, a combination of stacking and hydrogen-bonding interactions cause the mismatched T to turn away from the Watson–Crick face of the G. The direction of this motion is opposite to that seen when prokaryotic and eukaryotic MutS bind a G:T mismatch (Lamers *et al*, 2000; Warren *et al*, 2007): for $TnpA_{Dra2}$, the base turns away from—rather than into—the minor groove of the DNA stem. Furthermore, MutS-bound DNA is sharply bent at the mismatch resulting in severe local narrowing of the major groove, a deformation we do not see for $TnpA_{Dra2}$-bound hairpin DNA (although evaluating the DNA geometry is difficult as the stem is relatively short). Another major difference in the mode of G:T mismatch recognition by $TnpA_{Dra2}$ and MutS is the nature of the amino acid side chain that stacks against the mismatched

T. In MutS, a Phe has sufficient room due to the sharp bend in the DNA to reach into the widened minor groove to facilitate stacking. On the other hand, Arg14 of $TnpA_{Dra2}$ forms hydrogen bonds with the base just 3′ of the mismatched T as it reaches into the major groove (G + 34 in Figure 3C); this guanine would normally stack against the mismatched T in the standard B form DNA. To our knowledge, this is the first observation of such a geometrical arrangement in the context of mismatch recognition. It is curious that the equivalent of Arg14 is conserved in repetitive extragenic palindrome (REP)-associated tyrosine transposases, which are also members of the IS*200*/IS*605* family (Nunvar *et al*, 2010; P Siguier, personal communication), suggesting a similar role for this residue in recognizing the imperfect palindromes present in bacterial REP sequences.

In contrast to residues that are involved in recognition and discrimination of the transposon IPs, all of the important active site residues are conserved between IS*608* TnpA and $TnpA_{Dra2}$. The assembled active site of $TnpA_{Dra2}$ was captured structurally in the presence of heavy metal ions and although Cd^{2+} is certainly not the physiological divalent metal ion used *in vivo*, its octahedral-coordinating environment produces a cofactor-bound enzyme that is competent for DNA cleavage, and presumably recapitulates the authentic Mg^{2+}-bound form. The structures presented here are consistent with a chemical mechanism proposed for HUH nucleases (Hickman *et al*, 2002) in which a metal ion both localizes and polarizes the scissile phosphate, setting it up for nucleophilic attack by the catalytic tyrosine. The active site arrangement of $TnpA_{Dra2}$ is similar to that seen for TraI, an HUH nuclease involved in the initiation of conjugation (Datta *et al*, 2003).

We captured a second precleavage state of the enzyme in the presence of Zn^{2+}, in which a different array of ligands was observed tetrahedrally arranged around the metal ion. Although $TnpA_{Dra2}$ is not active in the presence of Zn^{2+}, the related HUH nuclease domain of TrwC, the plasmid R388 relaxase, has been structurally characterized in the presence of several divalent metal ions including Zn^{2+}, Cu^{2+}, and Ni^{2+} (Boer *et al*, 2006). In all cases, the metal ion-binding environment is tetrahedral, and TrwC was demonstrated to be active in the presence of these three metals. Thus, the lack of activity of $TnpA_{Dra2}$ in the presence of Zn^{2+} is not due to the tetrahedral metal-binding environment *per se*, nor is there anything untoward about using three His residues at an enzyme active site as metal-coordinating ligands (Dudev and Lim, 2003), as this is precisely what has been observed for the relaxase domains of both TrwC (Guasch *et al*, 2003) and TraI (Datta *et al*, 2003; Larkin *et al*, 2005). Rather, the conformational state of $TnpA_{Dra2}$ bound to Zn^{2+} reflects the fact that the third Zn^{2+} ligand, His 137, simply happens to be at an inopportune position on helix αD.

The shift in position of the mobile helix αD that accompanies Zn^{2+} binding concomitantly moves the nucleophilic tyrosine residue away from the scissile phosphate, thus explaining the lack of cleavage activity in the presence of Zn^{2+}. A shift in position of a nucleophile-bearing α-helix has been previously postulated for the prototypical HUH nuclease, the phi X174 gene A protein, this time in the context of a mechanistically critical movement of nucleophiles (van Mansfeld *et al*, 1986). In that case, two tyrosine residues spaced by three residues were proposed to reside on a helix, which moves during the course of rolling circle replication, alternating in access to different scissile phosphate groups alternatively bound at the active site. It is possible that helix αD mobility is a common feature of HUH nucleases, and that DNA transposases with HUH domains have exploited this inherent structural malleability in the context of transposition to direct strand transfer by allowing a *trans*/*cis*-active site isomerization.

Among DNA transposases, there seems to be a very limited number of structurally unrelated enzyme folds capable of carrying out the DNA cleavage and rejoining reactions required to move a segment of DNA from one location to another. The majority of ISs and transposons are believed to have catalytic domains with an RNase H-related fold (i.e. that of retroviral integrase; Nowotny, 2009; Hickman *et al*, 2010). The wide range of sequences that can be recognized and distinguished is readily understood by the fact that these enzymes are multidomain proteins containing dedicated sequence-specific DNA-binding domains (Davies *et al*, 2000; Watkins *et al*, 2004; Rice, 2005; Richardson *et al*, 2009). Therefore, successfully identifying their own transposon ends does not appear to be a conceptually sophisticated matter. In contrast, IS*200*/IS*605* transposons encode small, single-domain proteins that have the ability to carry out all of the chemical reactions of transposition on their own specific transposon sequences. An additional level of complexity is that, in contrast to most RNase H-like transposases, they insert into specific target sequences. The series of protein/DNA complex structures presented here sheds light on the types of structural differences that account for the ability of a small protein to not only distinguish among transposon DNA sequences but also to identify specific target sites in the absence of any site-specific DNA-binding domains. The variability—an Arg14 here or an Arg52/Phe75 pair there—is subtle yet remarkable in its simplicity.

Materials and methods

Protein purification and crystallization

Oligonucleotides were either obtained desalted from IDT (Coralville, IA) or synthesized using an Applied Biosystems 394 DNA/RNA synthesizer. Iodinated LE27 in which T + 18, T + 21, T + 25, T + 30, and T + 33 were substituted with 5-I-dU was synthesized using 5-I-dU-CE phosphoramidite (Glen Research, VA); in all cases, oligonucleotides were used without further purification. The gene encoding IS*Dra2* TnpA was cloned between the *Nco*I and *Xho*I sites of pET-15b. R14A and S122G/E123G mutations were introduced by the QuikChange method (Stratagene).

Protein was expressed in BL21 (DE3) cells upon induction by addition of IPTG to a final concentration of 0.5 mM followed by cell growth overnight at 16°C. Cells were harvested, resuspended in 35 mM Tris pH 8.0, 0.5 M NaCl, and 10% (w/v) glycerol, and frozen at −80°C until use. Thawed cells were diluted 1:1 (v/v) with 35 mM Tris pH 8.0, 0.5 M NaCl, 5% glycerol, 5 mM imidazole (Im), and 4 mM β-mercaptoethanol, sonicated, and centrifuged to recover soluble protein. Protein was loaded onto a 5-ml Ni-affinity column and washed sequentially with lysis buffer containing 5 mM Im, then 50 mM Im, followed by gradient elution from 0.05 to 0.5 M Im. Fractions containing $TnpA_{Dra2}$ were combined, thrombin (Sigma) added to a final concentration of 2 U/mg protein, and the solution was dialyzed overnight at 4°C against 35 mM Tris pH 8.0, 0.5 M NaCl, 5% glycerol, and 0.5 mM TCEP. Thrombin was removed by passage over benzamidine Sepharose prior to gel filtration on a TSK-SW3000 column. Fractions containing $TnpA_{Dra2}$ were concentrated, combined with the appropriate oligonucleotide(s), and dialyzed against 35 mM Tris pH 8.0, 0.15 M NaCl (LE IP) or 0.1 M sodium malonate (RE IP), 10 mM $MgCl_2$, and 0.5 mM TCEP.

The TnpA/LE27/T5 complex was formed by mixing protein at 10 mg/ml with LE27 and T5 at a molar ratio of 1:1.1:1.3 prior to dialysis. Crystals were obtained by the hanging drop method at 19°C by mixing the complex 1:1 with well solution containing 15–20% PEG 6000 and 50 mM sodium acetate pH 4.6. Crystals were cryoprotected by slow soak into 17 mM Tris pH 8.0, 75 mM NaCl, 0.25 mM TCEP, 5 mM $MgCl_2$, 50 mM sodium acetate pH 4.6, 25% PEG 6000 and 15–20% glycerol, and frozen by immersion in liquid propane.

The TnpA/RE34 complex was formed by mixing protein at 10 mg/ml with RE34 DNA at a molar ratio of 1:1.1 prior to dialysis. Crystals were obtained by the hanging drop method at 19°C by mixing the complex 1:1 with well solution containing 0.16 M $MgCl_2$, 0.1 M Tris pH 7.9, and 3.4 M 1,6-hexanediol. Crystals were frozen directly in liquid propane prior to data collection.

The TnpA/Cd complex was formed by mixing protein with a Y132F mutation at 9 mg/ml with LE27 and T5↓G at a molar ratio of 1:1.1:1.3 prior to dialysis. Crystals were obtained overnight at 19°C by hanging drop vapor diffusion upon mixing the complex 1:1 with well solution containing 10 mM $CdCl_2$, 0.1 M sodium acetate pH 4.8, and 14% PEG 4000. Crystals were cryoprotected by serial transfer into solutions containing 17 mM Tris pH 8.0, 75 mM NaCl, 0.25 mM TCEP, 5 mM $MgCl_2$, 5 mM $CdCl_2$, 50 mM sodium acetate pH 4.8, 14% PEG 4000, and increasing glycerol concentration to a final concentration of 20% (v/v). Crystals were frozen by immersion in liquid propane.

The TnpA/Zn complex was formed as described for TnpA/Cd but with the wild-type protein. Crystals were obtained as for TnpA/Cd but with a well solution containing 10 mM $ZnCl_2$, 0.1 M sodium acetate pH 4.2, and 17% PEG 4000.

Crystallographic data collection and structure determination

All diffraction data were collected at 95K on a rotating anode source equipped with multilayer focusing optics using Cu Kα radiation and an Raxis IV image-plate detector. Data were integrated and scaled with Denzo and Scalepack (Otwinowski and Minor, 1997; Table I). Attempts to solve the TnpA/LE27/T5 structure with molecular replacement using IS*608* TnpA as a search model were only partially successful, presumably due to the pseudo-translational symmetry between the three dimers in the asymmetric unit. Although two dimers could be located clearly, the orientation and position of the third remained ambiguous. Difference Fourier maps for the third dimer phased with the first two dimers were not of sufficient quality to place it. Therefore, single isomorphous replacement with anomalous scattering experimentally phased maps were calculated based on an iodine derivative obtained using iodinated LE27. Iodine sites were located with Shelxd (Sheldrick, 1998) and their positional and occupancy parameters were refined with Phases-95 (Furey and Swaminathan, 1997) using a phase integrating least squares procedure. The TnpA/RE34 and TnpA/Cd structures were solved with molecular replacement using AMoRe (Navaza, 2001) using the TnpA/LE27/T5 dimer with the lowest average atomic displacement parameters (ADPs) as a search model. TnpA/Zn was solved with a monomeric search model. All molecular models were built with O (Jones *et al*, 1991) and refined with Cartesian simulated annealing. Non-crystallographic symmetry restraints were used only in the case of TnpA/RE34. Individual ADP refinement and energy minimization was carried out with CNS 1.1 (Brünger *et al*, 1998). All structures were verified with composite simulated annealed omit maps. Molecular figures were generated using Pymol (DeLano, 2002). Coordinates have been deposited in the Protein Data Bank with the accession codes 2xm3 (TnpA/LE27/T5), 2xma (TnpA/RE34), 2xo6 (TnpA/Cd), and 2xqc (TnpA/Zn).

Fluorescence anisotropy measurements

HPLC-purified 3′ FAM-labelled oligonucleotides were purchased from IDT. For IS*608* TnpA, all top-strand oligonucleotides were 27 nt long, and corresponded to LE nt +16 to +42 (5′-AAAGCCCCTAGC TTTTAGCTATGGGGA; numbering as in Barabas *et al* (2008)), with the modifications shown in Figure 3D. For $TnpA_{Dra2}$, the oligonucleotides were 26 nt long, and corresponded to LE nt +13 to +38 (5′-CACACTCGTGACTTCAGTCATGAGTT). For bottom-strand oligonucleotides, the appropriate reverse complement sequences were used.

Oligonucleotides were annealed by resuspension in TE, heating to 95°C followed by rapid cooling on ice or dry ice. The extent of hairpin formation versus double-stranded dimers was monitored by size-exclusion chromatography on Superdex 200 using a Pharmacia SmartSystem. In the case of the IS*Dra2* bottom-strand hairpin (the only oligonucleotide where the amount of dimer was significant), this also allowed for partial separation of the two species for use in subsequent experiments.

To measure dissociation constants for the interaction between IS*608* TnpA or $TnpA_{Dra2}$ and various modified IPs, anisotropy changes of FAM-labelled oligonucleotides were monitored upon titration with protein. Each oligonucleotide was added to the highest protein concentration sample (generally 2–5 mg/ml) to a final concentration of either 50 nM DNA (IS*608* TnpA) or 40 nM ($TnpA_{Dra2}$). Serial dilutions of the protein–DNA mixture were performed in low adsorption 96-well plates (Corning) with buffer (0.2 M NaCl, 20 mM Tris pH 7.5, 20 mM $MgCl_2$, 1 mM DTT) containing the appropriate amount of DNA to keep it constant throughout the experiment. Experiments were performed in triplicate using black, clear flat bottom, 384-well plates (Corning) using a SpectraMax M5 micro-plate reader (Molecular Devices). The excitation wavelength was 493 nm and emission wavelength was 525 nm (cutoff 515 nm).

For each experiment performed in triplicate, fluorescence anisotropy change values were calculated as observed anisotropy minus initial anisotropy (detected in the absence of protein) and plotted as a function of protein concentration. Binding data were fit to a single-binding site model using Microcal Origin 6.0 and the equation $A = A_f + (A_b - A_f) \times \{(L_T + K_d + x) - \sqrt{[(L_T + K_d + x)^2 - 4L_T x]}\}/2L_T$, where A is the measured anisotropy change, A_f is the anisotropy for the free ligand, A_b is the anisotropy for the bound ligand, L_T is the total added ligand concentration, and x is the protein concentration. The single-binding site model was tested by analysing Hill plots; no allosteric behaviour was detected. Reported K_d values correspond to the average value for the indicated number of experiments, and the uncertainties correspond to S_D values. For K_d values $>100\,\mu M$ in Figure 3D, at the highest protein concentration used, the binding curve did not reach saturation and the K_d was estimated to be over 100 μM (i.e. the apparent inflection point of the sigmoidal-binding curve on a log plot was above 100 μM).

LE cleavage assay

A $TnpA_{Dra2}$/LE24 (LE nt +13 to +36) complex was prepared by mixing purified protein and DNA both at 60 μM, followed by overnight dialysis into 0.2 M NaCl, 20 mM Tris pH 7.5, and 0.2 mM TCEP. Cleavage was initiated by addition of 1.1 molar excess of a 13-mer target sequence (5′-TTGATGCTTGAGG) and the appropriate volumes of stock solutions of $MgCl_2$, $MnCl_2$, $CaCl_2$, $ZnCl_2$, or $CdCl_2$ to a final concentration of 2 mM. Reactions were quenched after 2 h at 22°C with SDS–PAGE sample buffer, and the formation of covalently modified $TnpA_{Dra2}$ was monitored by SDS–PAGE using 4–12% NuPAGE gels (Invitrogen).

D. radiodurans in vivo excision assay

Bacterial strains, media, and growth conditions have been previously described (Bonacossa de Almeida *et al*, 2002; Pasternak *et al*, 2010), as have the construction of the *D. radiodurans* GY13115 tester strain and the IS*Dra2*-113 mini-transposon (Pasternak *et al*, 2010).

The *D. radiodurans* IS*Dra2 tnpA* mutant genes encoding $TnpA_{R14A}$ or $TnpA_{S122G\ E123G}$ were isolated from the corresponding pET-15b derivatives by *Nde*I–*Xho*I restriction and then cloned into plasmid pGY11559 (Mennecier *et al*, 2004) to generate pGY13521 and pGY13522, respectively. Wild-type *tnpA* was cloned into pGY11559 to generate pGY13203 (Pasternak *et al*, 2010) and all constructs were verified by DNA sequencing.

To measure the *in vivo* excision frequencies of *ISDra2*-113, 10 individual Cam^R Tet^S colonies purified from each derivative of GY13115 strain expressing *in trans* wild type, mutant, or no $TnpA_{Dra2}$ were inoculated into 3 ml of TGY2X supplemented with spectinomycin and grown to an OD_{650} of 1.3. Determination of the total number of viable cells was performed on TGY plates and excision of IS*Dra2*-113 from the *tetA* gene was selected on TGY plates containing tetracycline. Colonies were counted after 4 days at 30°C. The frequencies of the excision event per viable cell from these 10 independent measurements were used to calculate mean values and standard deviations.

Supplementary data
Supplementary data are available at *The EMBO Journal* Online (http://www.embojournal. org).

Acknowledgements

We thank A Bailone for fruitful discussions and G Coste for expert technical assistance. This work was supported by the Intramural Program of the National Institute of Diabetes, Digestive, and Kidney Diseases of the National Institutes of Health (FD), the Centre National de la Recherche Scientifique (France, MC and SS), European contract LSHM-CT-2005-019023 (MC), and ANR grant Mobigen (MC and SS). Some data were collected at the Southeast Regional Collaborative Access Team 22-ID beamline at the Advanced Photon Source (APS), Argonne National Laboratory. Use of the APS was supported by the US. Department of Energy, Basic Energy Sciences, Office of Science, under Contract No. W-31-109-Eng-38.

Conflict of interest

The authors declare that they have no conflict of interest.

References

Barabas O, Ronning DR, Guynet C, Hickman AB, Ton-Hoang B, Chandler M, Dyda F (2008) Mechanism of IS*200*/IS*605* family DNA transposases: activation and transposon-directed target site selection. *Cell* **132:** 208–220

Boer R, Russi S, Guasch A, Lucas M, Blanco AG, Pérez-Luque R, Coll M, de la Cruz F (2006) Unveiling the molecular mechanism of a conjugative relaxase: the structure of TrwC complexed with a 27-mer DNA comprising the recognition hairpin and the cleavage site. *J Mol Biol* **358:** 857–869

Bonacossa de Almeida C, Coste G, Sommer S, Bailone A (2002) Quantification of RecA protein in Deinococcus radiodurans reveals involvement of RecA, but not LexA, in its regulation. *Mol Genet Genom* **268:** 28–41

Brünger AT, Adams PD, Clore GM, Delano WL, Gros P, Grosse-kunstleve RW, Jiang JS, Kuszewski J, Nilges M, Pannu NS, Read RJ, Rice LM, Simonson T, Warren GL (1998) Crystallography and Nmr system—a new software suite for macromolecular structure determination. *Acta Crystallogr D* **54:** 905–921

Datta S, Larkin C, Schildbach JF (2003) Structural insights into single-stranded DNA binding and cleavage by F Factor TraI. *Structure* **11:** 1369–1379

Davies DR, Goryshin IY, Reznikoff WS, Rayment I (2000) Three-dimensional structure of the Tn*5* synaptic complex transposition intermediate. *Science* **289:** 77–85

DeLano WL (2002) The PyMol molecular graphics system Available at http://www.pymol.org

Dudev T, Lim C (2003) Principles governing Mg, Ca, and Zn binding and selectivity in proteins. *Chem Rev* **103:** 773–787

Furey W, Swaminathan S (1997) PHASES-95: a program package for processing and analyzing diffraction data from macromolecules. *Methods Enzymol* **277:** 590–620

Guasch A, Lucas M, Moncalian G, Cabezas M, Perez-Luque R, Gomis-Ruth FX, de la Cruz F, Coll M (2003) Recognition and processing of the origin of transfer DNA by conjugative relaxase TrwC. *Nat Struct Biol* **10:** 1002–1010

Guynet C, Achard A, Ton-Hoang B, Barabas O, Hickman AB, Dyda F, Chandler M (2009) Resetting the site: redirecting integration of an insertion sequence in a predictable way. *Mol Cell* **34:** 612–619

Guynet C, Hickman AB, Barabas O, Dyda F, Chandler M, Ton-Hoang B (2008) *In vitro* reconstitution of a single-stranded transposition mechanism of IS*608*. *Mol Cell* **29:** 302–312

Hickman AB, Chandler M, Dyda F (2010) Integrating prokaryotes and eukaryotes: DNA transposases in light of structure. *Crit Rev Biochem Mol Biol* **45:** 50–69

Hickman AB, Ronning DR, Kotin RM, 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

Islam MS, Hua Y, Ohba H, Satoh K, Kikuchi M, Yanagisawa T, Narumi I (2003) Characterization and distribution of IS*8301* in the radioresistant bacterium Deinococcus radiodurans. *Genes Genet Syst* **78:** 319–327

Jones TA, Zou JY, Cowan SW, 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, Akopyants NS, Clifton SW, Roe BA, Berg DE (1998) Novel sequence organization and insertion specificity of IS605 and IS606: chimaeric transposable elements of Helicobacter pylori. *Gene* **223:** 175–186

Kersulyte D, Velapatino B, Dailide G, Mukhopadhyay AK, Ito Y, Cahuayme L, Parkinson AJ, Gilman RH, Berg DE (2002) Transposable element ISHp608 of Helicobacter pylori: nonrandom geographic distribution, functional organization, and insertion specificity. *J Bacteriol* **184:** 992–1002

Koonin EV, Ilyina TV (1993) Computer-assisted dissection of rolling circle DNA replication. *Biosystems* **30:** 241–268

Lamers MH, Perrakis A, Enzlin JH, Winterwerp HHK, de Wind N, Sixma TK (2000) The crystal structure of DNA mismatch repair protein MutS binding to a G.T mismatch. *Nature* **407:** 711–717

Larkin C, Datta S, Harley MJ, Anderson BJ, Ebie A, Hargreaves V, Schildbach JF (2005) Inter- and intramolecular determinants of the specificity of single-stranded DNA binding and cleavage by the F factor relaxase. *Structure* **13:** 1533–1544

Mennecier S, Coste G, Servant P, Bailone A, Sommer S (2004) Mismatch repair ensures fidelity of replication and recombination in the radioresistant organism *Deinococcus radiodurans*. *Mol Genet Genom* **272:** 460–469

Mennecier S, Servant P, Coste G, Bailone A, Sommer S (2006) Mutagenesis via IS transposition in *Deinococcus radiodurans*. *Mol Microbiol* **59:** 317–325

Moseley BEB, Mattingly A (1971) Repair of irradiated transforming deoxyribonucleic acid in wild type and a radiation-sensitive mutant of *Micrococcus radiodurans*. *J Bacteriol* **105:** 976–983

Navaza J (2001) Implementation of molecular replacement in AMoRe. *Acta Crystallogr D* **57:** 1367–1372

Nowotny M (2009) Retroviral integrase superfamily: the structural perspective. *EMBO Rep* **10:** 144–151

Nunvar J, Huckova T, Licha I (2010) Identification and characterization of repetitive extragenic palindromes (REP)-associated tyrosine transposases: implications for REP evolution and dynamics in bacterial genomes. *BMC Genom* **11:** 44

Otwinowski Z, Minor W (1997) Processing of X-ray diffraction data collected in oscillation mode. *Meth Enzymol* **276:** 307–326

Pasternak C, Ton-Hoang B, Coste G, Bailone A, Chandler M, Sommer S (2010) Irradiation-induced Deinococcus radiodurans genome fragmentation triggers transposition of a single resident insertion sequence. *PLoS Genet* **6:** e1000799

Rice PA (2005) Visualizing Mu transposition: assembling the puzzle pieces. *Genes Dev* **19:** 773–775

Richardson JM, Colloms SD, Finnegan DJ, Walkinshaw MD (2009) Molecular architecture of the Mos1 paired-end complex: the structural basis of DNA transposition in a eukaryote. *Cell* **138:** 1096–1108

Ronning DR, Guynet C, Ton-Hoang B, Perez ZN, Ghirlando R, Chandler M, Dyda F (2005) Active site sharing and subterminal hairpin recognition in a new class of DNA transposases. *Mol Cell* **20:** 143–154

SantaLucia J, Hicks D (2004) The thermodynamics of DNA structural motifs. *Annu Rev Biophys Biomol Struct* **33:** 415–440

Sheldrick GM (1998) SHELX: applications to macromolecules. In *Direct Methods for Solving Macromolecular Structures*, Fortier S (ed) pp 401–411. Dordrecht: Kluwer Academic Publications

Slade D, Lindner AB, Paul G, Radman M (2009) Recombination and replication in DNA repair of heavily irradiated Deinococcus radiodurans. *Cell* **136:** 1044–1055

Ton-Hoang B, Guynet C, Ronning DR, Cointin-Marty B, Dyda F, Chandler M (2005) Transposition of ISHp608, member of an

unusual family of bacterial insertion sequences. *EMBO J* **24:** 3325–3338

Tsutakawa SE, Jingami H, Morikawa K (1999) Recognition of a TG mismatch: the crystal structure of very short patch repair endonuclease in complex with a DNA duplex. *Cell* **99:** 615–623

van Mansfeld ADM, van Teeffelen HAAM, Baas PD, Jansz HS (1986) Two juxtaposed tyrosyl-OH groups participate in phiX174 gene A protein catalysed cleavage and ligation of DNA. *Nucleic Acids Res* **14:** 4229–4238

Warren JJ, Pohlhaus TJ, Changela A, Iyer RR, Modrich PL, Beese LS (2007) Structure of the human MutS alpha DNA lesion recognition complex. *Mol Cell* **26:** 579–592

Watkins S, van Pouderoyen G, Sixma TK (2004) Structural analysis of the bipartite DNA-binding domain of Tc3 transposase bound to transposon DNA. *Nucleic Acids Res* **32:** 4306–4312

Zahradka K, Slade D, Bailone A, Sommer S, Averbeck D, Petranovic M, Lindner AB, Radman M (2006) Reassembly of shattered chromosomes in *Deinococcus radiodurans*. *Nature* **443:** 569–573

