Molecular Cell, Vol. 5, 1025–1034, June, 2000, Copyright ©2000 by Cell Press

Unexpected Structural Diversity in DNA Recombination: The Restriction Endonuclease Connection

Alison Burgess Hickman,*‡ Ying Li,† Shiny Varghese Mathew,*§ Earl W. May,†‖ Nancy L. Craig,† and Fred Dyda*
*Laboratory of Molecular Biology
National Institute of Diabetes and Digestive and Kidney Diseases
National Institutes of Health
Bethesda, Maryland 20892
†Howard Hughes Medical Institute
Department of Molecular Biology and Genetics
Johns Hopkins University School of Medicine
Baltimore, Maryland 21205

Summary

Transposition requires a coordinated series of DNA breakage and joining reactions. The Tn7 transposase contains two proteins: TnsA, which carries out DNA breakage at the 5′ ends of the transposon, and TnsB, which carries out breakage and joining at the 3′ ends of the transposon. TnsB is a member of the retroviral integrase superfamily whose hallmark is a conserved DDE motif. We report here the structure of TnsA at 2.4 Å resolution. Surprisingly, the TnsA fold is that of a type II restriction endonuclease. Thus, Tn7 transposition involves a collaboration between polypeptides, one containing a DDE motif and one that does not. This result indicates that the range of biological processes that utilize restriction enzyme–like folds also includes DNA transposition.

Introduction

The movement of transposable elements requires DNA cleavage reactions that separate the element from surrounding donor site DNA and DNA joining reactions that link the element to its new insertion site. For most elements that have been analyzed in biochemical detail, these reactions are mediated by a transposase composed of a single type of polypeptide containing a single active site that mediates both DNA breakage and joining. Three-dimensional structures have been elucidated for several catalytic domains involved in transposition, including the retroviral integrase that inserts viral DNA into chromosomes (Dyda et al., 1994), the catalytic core of bacteriophage MuA (Rice and Mizuuchi, 1995), and the Tn5 transposase (Davies et al., 1999). Despite substantial mechanistic differences between these systems, the structures of these catalytic units display remarkable similarity. The close topological relationship between these molecules indicates that they all belong to a superfamily of polynucleotidyl transferases. Other members of this superfamily include RNaseH, responsible for the endonucleolytic degradation of DNA–RNA hybrids, and RuvC, a Holliday junction–resolving enzyme.

The transposase catalytic domains characterized to date carry a triad of catalytically essential acidic residues, the so-called DDE motif. These residues are located on the same structural elements in each structure, and each residue is required for both strand cleavage and subsequent strand joining; they function by coordinating the divalent metal ions necessary for catalysis and define the enzyme active sites.

Unlike other well-characterized transposition systems, such as Tn10 (Kleckner et al., 1996), Tn5 (Reznikoff, 1993), and bacteriophage Mu (Mizuuchi, 1992), the Tn7 transposase is composed of two element–encoded proteins TnsA and TnsB, which introduces double-strand breaks at the transposon ends and joins the element to the target DNA (Figure 1A). TnsA is responsible for cleavage at the 5′ ends of the transposon, while TnsB cuts the 3′ ends and catalyzes subsequent strand transfer using the newly created free 3′ OH groups as the nucleophiles (May and Craig, 1996; Sarnovsky et al., 1996). The two cleavage events can be separated using inactive point mutant versions of TnsA or TnsB, but cleavage by the active protein still requires the presence of the inactive one. Thus, TnsA and TnsB also play architectural roles in the synaptic complex in addition to possessing enzymatic activities. Other Tns proteins—TnsC, TnsD, and TnsE—are involved in the control of transposition and target site selection (Craig, 1996).

Based on the invariant structural theme for the HIV, MuA, and Tn5 enzymes involved in DNA transposition, it has been assumed that TnsA and TnsB also conform with this structural paradigm as members of the same superfamily (Sarnovsky et al., 1996). In the case of TnsB, the identification by site-directed mutagenesis of the three residues comprising the DDE motif (D273, D361, and E396) gave strong support to this assumption (Sarnovsky et al., 1996). Furthermore, secondary structure prediction using PSIPRED (Jones, 1999) clearly demonstrates the similarity in the secondary structure elements of MuA and TnsB, resulting in an alignment (Krementsova et al., 1998) that places the DDE residues of TnsB on the same structural elements as those containing the MuA DDE motif (data not shown). The case for TnsA remained less convincing, in part because a complete DDE motif could not be identified.

We show here that, contrary to expectations, TnsA possesses a type II restriction endonuclease fold and is not a member of the RNaseH-like polynucleotidyl transferase superfamily. Thus, Tn7 transposition is mediated by two types of enzymes, one a polynucleotidyl transferase of the retroviral integrase superfamily that mediates breakage and joining at the 3′ ends at the element, and a restriction enzyme–like endonuclease that executes simple hydrolysis reactions at the 5′ ends.

‡To whom correspondence should be addressed (e-mail: ahickman@helix.nih.gov).
§Present address: Department of Pharmacology, College of Medicine, Howard University, Washington, D.C. 20059.
‖Present address: SmithKline Beecham Pharmaceuticals, 1250 South Collegeville Road, UP1345, Collegeville, Pennsylvania 19426.

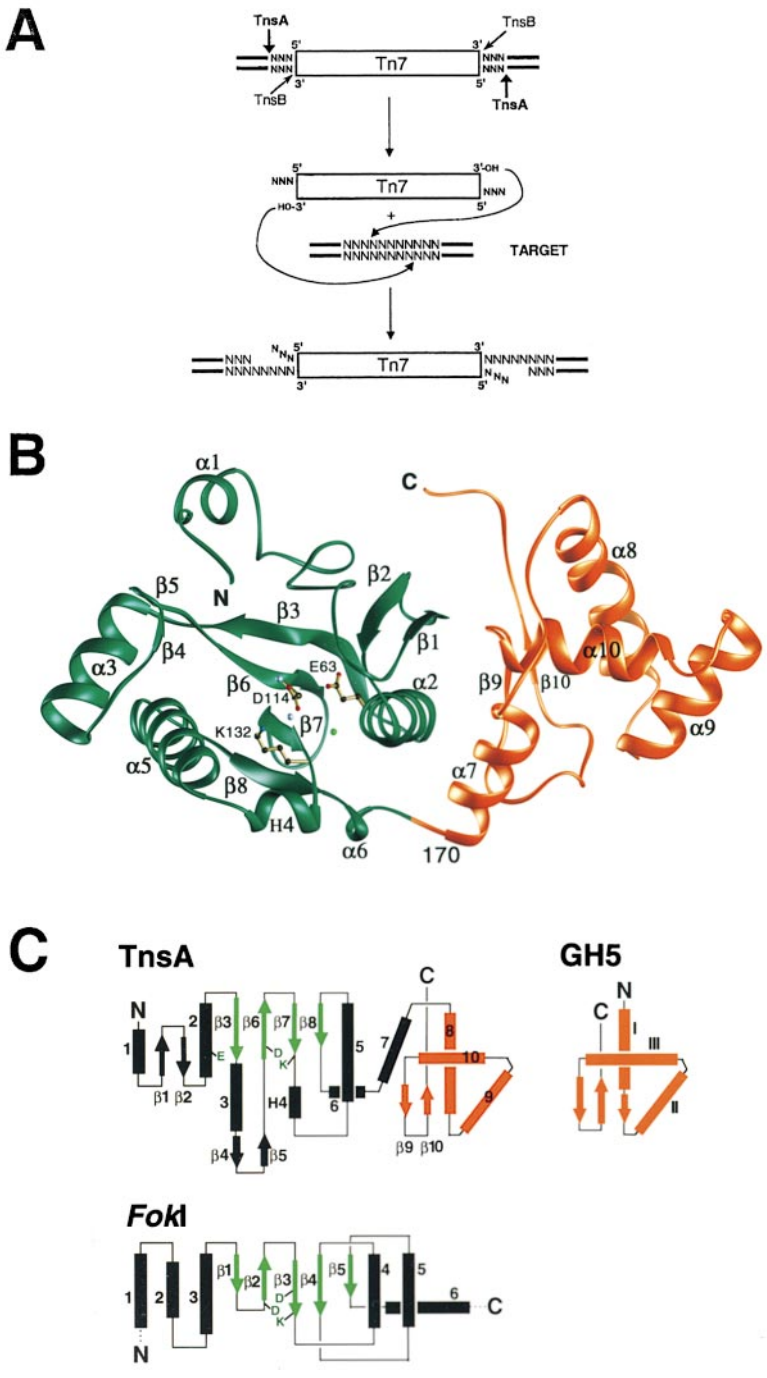

Figure 1. The Tn7 Transposition Reaction and the Structure of TnsA

(A) Schematic of the reaction carried out by the Tn7 transposase. Although the initial 5′ and 3′ cleavage reactions (top) by TnsA and TnsB, respectively, mechanistically precede target joining (middle), the entire transposition reaction occurs in the context of an assembled transpososome that contains both transposon DNA (donor DNA) and target DNA (*att*Tn7).

(B) Ribbons (Carson, 1991) representation of the structure of TnsA. The N-terminal domain (green) has the fold of a type II restriction endonuclease, while the C-terminal domain (orange) contains a helix-turn-helix motif (α9 and α10). The positions of two bound magnesium ions (silver) and a chloride ion (green) are indicated, as are three catalytically essential residues.

(C) Comparison of the folding topologies of TnsA, the cleavage domain of FokI (Wah et al., 1997, 1998), and histone H5 (Ramakrishnan et al., 1993). The arrows indicate β strands and cylinders α helices except helix H4 of TnsA, which is a 3_{10} helix. The indicated active site residues of TnsA are E63, D114, and K132; those of FokI are D450, D467, and K469.

Results and Discussion

Description of the Domain Structure of TnsA

We have crystallized full-length (273 amino acids) TnsA, solved its three-dimensional structure using multiple isomorphous replacement, and subsequently refined the coordinates at 2.4 Å resolution (Table 1). Examples of the experimental and refined electron density are shown in Figure 2. The molecule contains two domains that are tightly packed against each other and connected via a six-residue linker loop (Figure 1B). The overall topology of the larger, N-terminal catalytic domain (residues 1–165) is similar to the type II restriction endonucleases (Figure 1C). According to a three-dimensional homology search using a distance matrix alignment program run on the DALI server (Holm and Sander, 1997), several restriction endonucleases and other nucleic acid binding proteins possessing a restriction endonuclease fold possess significant three-dimensional homology to TnsA. In contrast to the type II restriction endonucleases, which are homodimeric enzymes that introduce either staggered or blunt-ended double-strand breaks into DNA, we do not see any indication for a restriction endonuclease-like dimer in the crystals. This

Table 1. Data Collection and Refinement Statistics

Data Set	Native	$(CH_3CH_2)Hg(ac)$	$(CH_3)_3Pb(ac)$
Resolution (Å)	2.4	3.0	3.0
Total reflections (N)	131670	67352	47211
Unique reflections (N)	23596	12534	12613
Completeness (%) [for $I/\sigma(I) > -3.0$]	97.0	99.4	99.8
$I/\sigma(I)$	13.2	13.9	14.1
R_{sym}	0.063	0.059	0.069
R_{cullis}		0.663	0.629
R_{Kraut}		0.152	0.210
Phasing power			
Isomorphous		1.07	1.35
Anomalous		1.05	1.71
Figure of merit of the experimental map:	0.447		
Refinement			
Resolution (Å)	30–2.4		
Atoms (N)	4488		
Reflections $F > 2\sigma(F)$	22840		
R factor (%)	20.2		
R_{free} (%)	25.4		
Rms bond lengths (Å)	0.009		
Rms bond angles (°)	1.676		

$R_{sym} = \Sigma|I - \langle I\rangle|/\Sigma\langle I\rangle$
$R_{cullis} = \Sigma|||FPH_o| \pm |FP_o|| - |FH_c||/\Sigma||FPH_o| \pm |FP_o||$ for centric reflections
$R_{Kraut} = \Sigma||FPH_o| - |FPH_c||/|FPH_o|$ for acentric reflections, isomorphous case; $\Sigma||FPH_o^+| - |FPH_c^+|| + ||FPH_o^-| - |FPH_c^-||/\Sigma(|FPH_o^+| + |FPH_o^-|)$ for acentric reflections, anomalous case.
FP is the protein, FPH is the derivative, and FH is the heavy atom structure factor. FPH^+ and FPH^- denote the Bijvoet mates of the derivatives. The phasing power is defined as FH_c/E for the isomorphous case and $2FH_c/E$ for the anomalous case, where FH_c is the calculated heavy atom structure factor and E is the rms lack of closure of the structure factor triangles.
R factor $= \Sigma|FP_o - FP_c|/\Sigma|FP_o|$

observation is consistent with current models for Tn7 transposition since TnsA and TnsB very likely form a heterodimer on the transposon ends.

The C-terminal domain (residues 170–272) contains a helix-turn-helix (HTH) motif (Harrison and Aggarwal, 1990) and has a topology that was first seen in the catabolic gene activator protein, CAP (McKay and Steitz, 1981), and subsequently in many other proteins, most of which bind DNA sequence specifically. The most closely related HTH-containing protein is the globular domain of the linker histone H5 (GH5; Figure 1C) (Ramakrishnan et al., 1993).

Structural and Functional Comparison between the TnsA Catalytic Domain and Other Endonucleases

The canonical restriction endonuclease fold, characterized by a five-stranded β sheet and two α helices (Aggarwal, 1995; Kovall and Matthews, 1999), has been observed in all the structures of type II restriction endonucleases determined to date. Several other nucleases have also been reported to possess a restriction endonuclease fold. Indeed, the closest structural homolog to the TnsA catalytic domain is the *v*ery *s*hort patch *r*epair (Vsr) endonuclease (Tsutakawa et al., 1999a, 1999b), a protein involved in DNA repair that recognizes a TG mismatch in the context of a specific sequence and cleaves the phosphate backbone on the 5′ side of the thymidine. Although Vsr endonuclease binds and cleaves DNA as a monomer, efficient repair is reported to require the additional proteins, MutL and MutS (Jones et al., 1987; Lieb, 1987).

Among type II restriction endonucleases, the most similar protein to TnsA on the basis of number of aligned residues is FokI, a type IIS restriction enzyme (Figure 1C). This enzyme is distinct from other type II enzymes such as EcoRI or EcoRV in that the catalytic domain cleaves DNA nonspecifically a fixed distance away from its DNA recognition sequence, and it relies on additional domains for sequence-specific DNA recognition (Wah et al., 1997, 1998). On its own, the catalytic domain of FokI binds DNA only weakly and cleaves DNA nonspecifically (Li et al., 1992); even in the context of the intact FokI–DNA cocrystal structure, the catalytic domain does not contact the DNA (Wah et al., 1998). TnsA also has no detectable site-specific DNA binding activity; this function resides instead in TnsB, which specifically binds the transposon ends (Craig, 1996). We speculate that the observed structural similarity between TnsA and FokI is a reflection of this functional analogy.

The topological similarity between the catalytic domain of TnsA and the structurally related proteins FokI, Cfr10I, and the Vsr endonuclease is shown in Figure 3. Despite the clear topological relationship between the catalytic domain of TnsA and restriction endonucleases, there are distinct differences. Most notably, TnsA lacks the fifth strand of the central mixed β sheet. Although the strands in the core of the TnsA domain have the characteristic sequential strand order 1234 (strands β3, β6, β7, and β8 in Figures 1B and 1C), strand β8 is immediately followed by the linker to the C-terminal domain. Furthermore, the observed structural homology does not include the first 60 N-terminal residues of TnsA. This region before helix α2 possesses a unique fold consisting

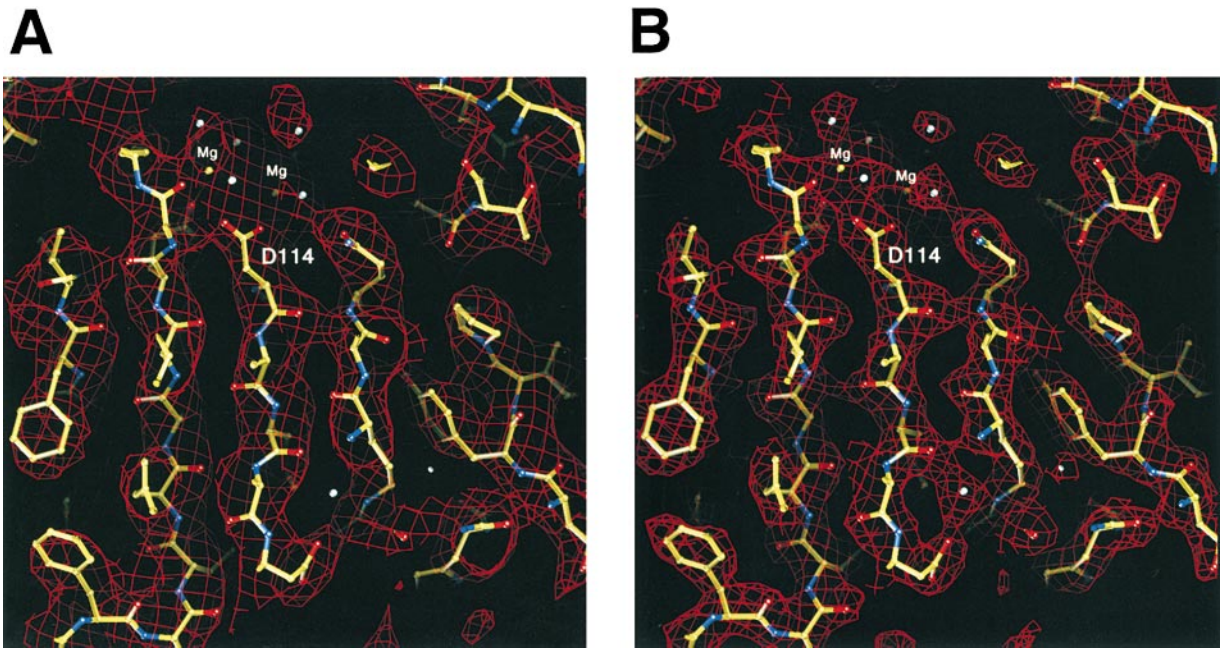

Figure 2. Experimental and Refined Electron Density

(A) Experimental electron density in the region of the four β strands comprising the central β sheet of the catalytic domain at 3 Å resolution contoured at 1σ.

(B) $2F_o$-F_c electron density in the same region after refinement at 2.4 Å resolution contoured at 1σ. The final refined structure is overlayed on both panels. The white spheres are water molecules.

of helix α1, a disordered loop comprising residues 27–42, and a two-stranded β sheet whose first strand (β1) contributes ~50% of the solvent-excluded surface between the N- and C-terminal domains.

The Active Site of TnsA

The active site of TnsA can be unambiguously identified based on its least squares three-dimensional superposition with several restriction endonucleases. In restriction enzymes, the catalytic motif generally involves several acidic and one basic residue (Aggarwal, 1995; Kovall and Matthews, 1999). The most conserved feature is a D residue, always located on the second strand of the motif. Its importance can be understood from structures in which two metal ions are bound at the active site, where it is observed to bridge these ions through bidentate coordination. The other essential catalytic residues—usually another acidic residue and one basic residue—show some variation, both in character and in the secondary structural elements from which they arise.

Previous work on TnsA demonstrated that residue D114 is essential for cleavage at the 5′ ends of Tn7, while mutation at position E149 leads to partial loss of 5′ end cleavage activity; 3′ end breakage and joining of the transposon to the target proceeded normally (Sarnovsky et al., 1996; May and Craig, 1996). D114 and E149 are both located close to the region expected for a restriction endonuclease-like active site. Indeed, the position of D114 on the second strand of the central β sheet (strand β6 in Figures 1B and 1C) implicates it as the crucial bidentate ligand to the metal cofactor ions (discussed below). Based on the close structural similarity between the array of residues close to D114 and the active sites of restriction endonucleases, we tested single point mutations of several residues for their effects on in vitro recombination. As shown in Figure 4, residues E63 and K132 are essential to 5′ end cleavage, while mutation of Q130 demonstrates the same phenotype as that of E149. Thus, E63 and K132, together with D114, form the triad of catalytic residues expected for a restriction endonuclease active site. This particular constellation of active site residues is identical with that seen in the restriction enzyme Cfr10I (Bozic et al., 1996). In Cfr10I, residues E71, D134, K190, and E204 have been shown to be catalytically important (Skirgaila et al., 1998); these correspond to the TnsA residues E63, D114, K132, and E149, residing on analogous structural elements (Figure 3). In this context, it is worth noting that among the restriction endonucleases, Cfr10I is the next most closely related structural homolog of TnsA after FokI. Cfr10I is distinguished among restriction enzymes as it does not have stringent sequence specificity for cleavage (Bozic et al., 1996), a property it shares with TnsA.

The mutational results shown in Figure 4, taken together with the structural homology between the protein folds, unequivocally link proteins involved in DNA recombination with the restriction endonucleases. These findings also provide strong evidence that the active site in TnsA directly participates in recombination, rather than TnsA simply controlling the activity of TnsB.

We observed strong difference electron density in the active site region, which we interpret as two Mg^{2+} ions and a Cl^- ion with several well-ordered, coordinating water molecules (Figure 5). Both Mg^{2+} ions are in distorted octahedral coordination environments and are

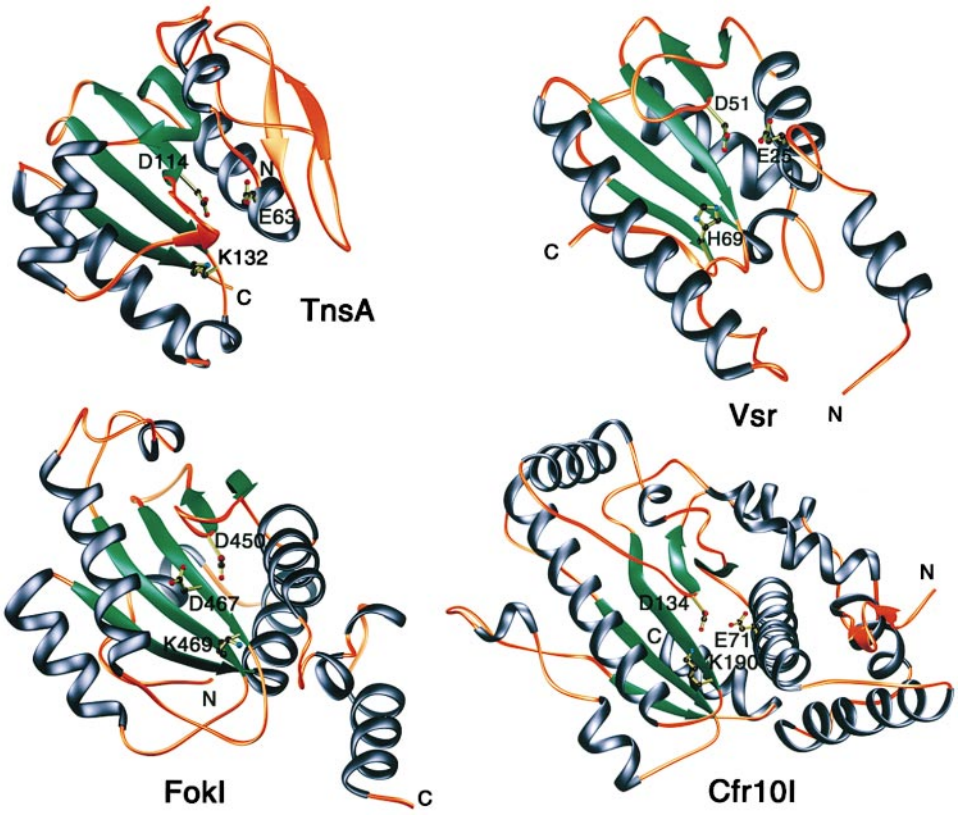

Figure 3. Comparison of the Restriction Endonuclease Folds of TnsA, the Vsr Endonuclease, FokI, and Cfr10I

The conserved central β sheet fold is highlighted in green. Only the N-terminal catalytic domains of FokI and TnsA are shown.

located 3.9 Å apart. The identity of the Cl^- was confirmed by anomalous difference Fourier calculations. The two metal ions are bidentately liganded by the carboxylate group of D114, giving rise to an arrangement very similar to that described by Beese and Steitz for the 3′ to 5′ exonuclease active site of the Klenow fragment (Beese and Steitz, 1991). Essentially identical two-metal arrangements have also been seen in the pre- and postcleavage DNA complexes of BamHI (Viadiu and Aggarwal, 1998), the avian sarcoma virus integrase (Bujacz et al., 1997), and more recently in the structure of the Vsr endonuclease–DNA complex (Tsutakawa et al., 1999b). For example, when TnsA is aligned with the postcleavage BamHI–DNA structure, the Mg^{2+} ion positions of TnsA differ less than 1 Å from the corresponding positions of Mn^{2+} bound to BamHI (data not shown). The partial loss of 5′ end cleavage activity phenotype of Q130E and E149A can be explained by assuming that the exact positioning of the carbonyl ligand to the metal provided by the side chain of Q130 is needed for wild-type activity. However, partial activity (i.e., left end 5′ cleavage) can be supported when either of these side chains is mutated.

The role of the observed Cl^- is probably to provide charge balance. It is located in a pocket defined by the N terminus of helix α2 and is liganded by the side chain hydroxyl of S60 as well as by two water molecules that are also part of the first coordination sphere of one of the Mg^{2+} ions (Figure 5). By comparison to the structures of the BamHI–DNA complex and the Vsr endonuclease–DNA complex, it seems likely that in the assembled Tn7 transpososome, the Cl^- binding site would be occupied by a phosphate group of the DNA backbone.

Two bound Mg^{2+} ions, as seen in the active site of TnsA, have not been previously observed in other restriction endonuclease-like active sites in the absence of substrates or substrate analogs, indicating that—in principle—a two-metal ion nuclease active site can be assembled without the DNA substrate. TnsA presents a mechanistic contrast to EcoRV in which the high affinity metal binding site is assembled only in the presence of the EcoRV recognition sequence (Winkler et al., 1993), thereby explaining why only the EcoRV specific sequence is cleaved, despite similar binding affinities for both specific and nonspecific DNA sequences. That TnsA can assemble a two-metal ion catalytic site in the absence of DNA may indicate that Tn7 does not require a protective mechanism to prevent wanton cleavage of nontransposon end sequences. This is most likely because TnsA relies on TnsB for DNA binding and sequence recognition, and thus nontransposon end sequences would never become available to TnsA for cleavage.

Proposed Mode of DNA Binding

The active site of TnsA lies on the bottom of an ~18 Å wide depression. When the observed Mg^{2+} and Cl^- ions are included in electrostatic calculations, this cavity

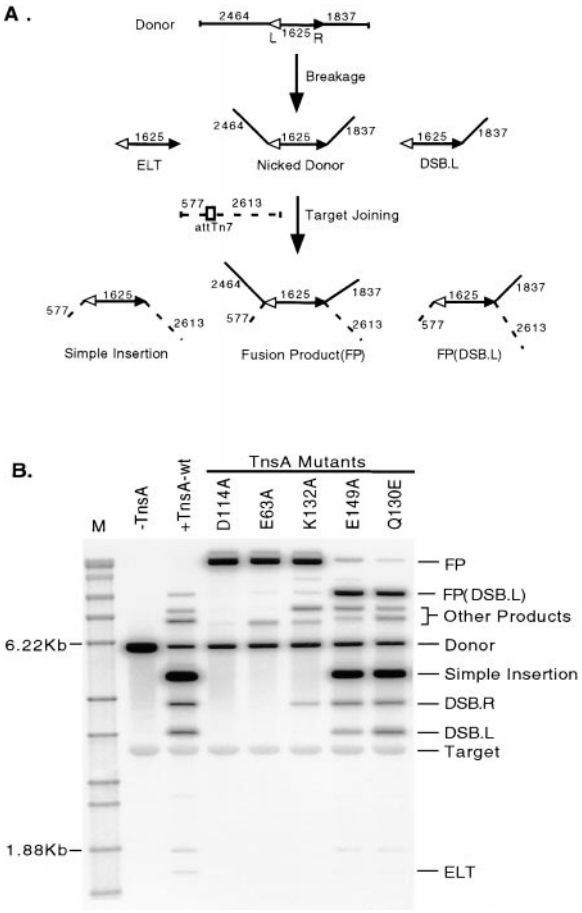

Figure 4. Effects of TnsA Mutations on 5′ End Breakage

(A) Various intermediates and products of Tn7 transposition. Wild-type recombination initiates with double-strand breaks that excise the transposon from the flanking donor DNA; mutations in TnsA block cleavages at the 5′ ends of the transposon to give nicked species. Target joining can proceed with DSBs, ELTs, and nicked species; only some of the possible joining products are shown. In the fusion product (FP), breakage at only the 3′ ends of the element occurred, followed by target joining; in the FP (DSB.L), a DSB occurred, whereas only a nick occurred at the right end of Tn7 (E. W. M. and N. L. C., unpublished data).

(B) Recombination with wild-type and various TnsA mutants. Like D114A, E63A and K132A completely block cleavage at the 5′ transposon ends. Like E149A, Q130E predominately blocks 5′ cleavages at the right end of Tn7. Recombination reactions were carried out as described (Sarnovsky et al., 1996) except that TnsD-His was used (Sharpe and Craig, 1998). The DNA substrates and recombination products were digested with NdeI, displayed on a 0.6% agarose gel, transferred to a membrane, and hybridized to a transposon specific probe. The blots were analyzed by phosphorimager analysis (Molecular Dynamics).

possesses a markedly positive surface potential (Christopher, 1998). The low binding affinity of TnsA for DNA has precluded us from cocrystallizing the protein with DNA, but a reasonable model predicting protein–DNA interactions can be constructed (Figure 6) by superimposing the structure of TnsA on those of the postcleavage BamHI–DNA complex (Viadiu and Aggarwal, 1998), the cleaved EcoRV–DNA complex (Kostrewa and Winkler, 1995), and the Vsr endonuclease–DNA complex (Tsutakawa et al., 1999b). This is possible since the aligned structures are very similar in the region of the active sites. The cleaved strand is always in the same orientation relative to the active site residues, and the backbones have similar traces, at least in the neighborhood of the scissile phosphate. It seems likely that the orientation of the DNA strand cleaved by TnsA will be the same as that observed in these postcleavage structures, and indeed it is possible to model it as such without major stereochemical clashes.

By assuming an orientation for the DNA strand cleaved by TnsA, the relative location of the TnsB cleavage point can therefore be predicted. It occurs on the strand opposite to that cleaved by TnsA, offset from the 5′ cleavage site by three bases in the direction of the transposon (see Figure 1A and Figure 6). This places the TnsB cleavage site away from the surface of TnsA, and presumably easily accessible to the TnsB active site. It is also clear from this model that the transposon DNA is unlikely to be the opposite orientation relative to the TnsA active site since this would place the TnsB cleavage site in a position far more difficult to access. The model also predicts that the loop following helix α1 consisting of residues 27–42 would contact the minor groove close to the 5′ cleavage site in the direction of the transposon. This loop is hydrophilic and mobile in the structure with high temperature factors, indicating that it is not well stabilized in the absence of DNA. The other structurally characterized restriction enzymes and Vsr endonuclease also use a surface loop to contact the minor groove of their respective substrates at or near to this position close to their cleavage sites.

The TnsA catalytic domain appears “trimmed down” when compared with the type II restriction enzymes (see Figure 3). In particular, it is missing motifs seen in other proteins that are involved in base-specific DNA recognition in the major groove, such as the extended chain motif of EcoRI (Kim et al., 1990), the loop around residue 154 in BamHI (Newman et al., 1994), or the DNA binding subdomains of EcoRV and PvuII (Winkler et al., 1993; Athanasiadis et al., 1994; Cheng et al., 1994). From our model, it appears that the only common structural element likely to be involved in major groove contacts is the N-terminal end of helix α5.

The C-Terminal Domain of TnsA

Several observations suggest that, despite an HTH motif, the C-terminal domain of TnsA is not involved in DNA binding. The C-terminal domain starts at P170 and leads into helix α7, a structural element unique to TnsA since homology to GH5 and other CAP-like domains does not begin until after this helix (Figures 1B and 1C). Helix α7 contributes significantly to holding the two domains in a fixed orientation relative to each other and is positioned so as to cover most of helix α10. In most of the available structures of HTH-containing domains complexed with DNA, the analogs of helix α10 serve as the recognition helix, usually forming an array of base-specific interactions in the major groove. In its observed position in TnsA, helix α10 would be unable to form these characteristic DNA contacts. Rather, it contributes to the interdomain interface, with its C terminus directly contacting strand β2 of the catalytic domain. Finally, the molecular surface electrostatic potential of this domain is nearly featureless, with many solvent-exposed

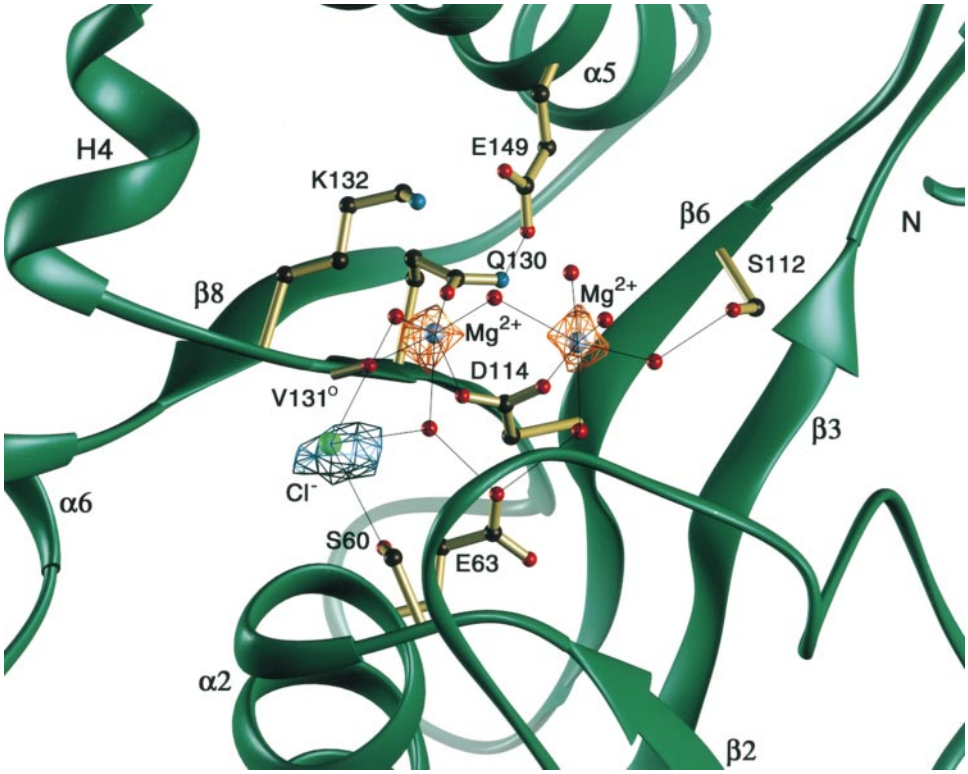

Figure 5. The TnsA Active Site Contains Two Bound Mg^{2+} Ions

Ribbons (Carson, 1991) representation of the TnsA active site showing the bound Mg^{2+} ions and coordinating ligands. The difference electron density for the Mg^{2+} ions (orange) and the anomalous difference density for Cl^- (blue) is contoured at 4σ. The red spheres represent coordinating water molecules, and the connecting lines are hydrogen bonds. $V131^O$ indicates that the coordinating ligand is the main chain carbonyl oxygen.

hydrophobic patches. Therefore, rather than binding DNA, the C-terminal domain of TnsA is likely to mediate protein–protein interactions with another member of the transpososome. Interestingly, this presents yet another parallel to the FokI system: the D3 domain of FokI, also a CAP-like HTH domain, barely touches the DNA and is mainly involved in protein–protein interactions with its "recognition helix" also making interdomain interactions (Wah et al., 1997).

Implications of the Link between TnsA and Restriction Endonucleases

Thousands of type II restriction endonucleases are known, yet all of the structures determined to date reveal a common fold in their catalytic domains, suggesting divergent evolution from a common ancestor. The fold has also been observed in a subdomain of the bacteriophage λ exonuclease (Kovall and Matthews, 1997) and an archaeal tRNA endonuclease (Li et al., 1998). More recently, two bacterial proteins involved in DNA repair have also been shown to possess a restriction endonuclease fold: the Vsr endonuclease, which recognizes a TG mismatch and cleaves on the 5′ side of the T (Tsutakawa et al., 1999a, 1999b), and MutH, which cleaves 5′ to an unmethylated d(GATC) sequence in hemimethylated duplex DNA (Ban and Yang, 1998). On the basis of the results presented here, the group of nucleases that use a restriction endonuclease catalytic domain must now be extended due to the unexpected appearance of this fold in DNA transposition.

In addition to expanding the known range of enzymes that employ the restriction endonuclease fold, the crystal structure of TnsA reveals an unexpected structural diversity in transposition itself and has expanded our view of the mechanisms of transposition. We now know that a collaboration between the restriction enzyme-like protein TnsA and the retroviral integrase-like protein TnsB executes the double-strand breaks at the ends of Tn7 that initiate transposition. It is intriguing that Tn7 has adopted catalytic subunits that are fundamentally different in mechanism and topology to carry out 5′ and 3′ strand cleavages, despite very similar underlying chemistry and the common requirement for two divalent metal ion cofactors. The key to understanding this diversity may lie in considering the different roles that TnsA and TnsB play during transposition. TnsB has a dual role: it catalyzes 3′ cleavage and then activates the newly created 3′ OH groups for the subsequent strand transfer reactions. TnsB must be able to use two different attacking nucleophiles, H_2O in the initial cleavage step to separate the transposon from the donor DNA, and the exposed 3′ OH end of the transposon to link the transposon to the target DNA. In contrast, TnsA has only to perform one reaction during transposition—5′ cleavage—for which a restriction endonuclease fold that executes only a hydrolysis reaction is sufficient.

The differing roles that TnsA and TnsB play during

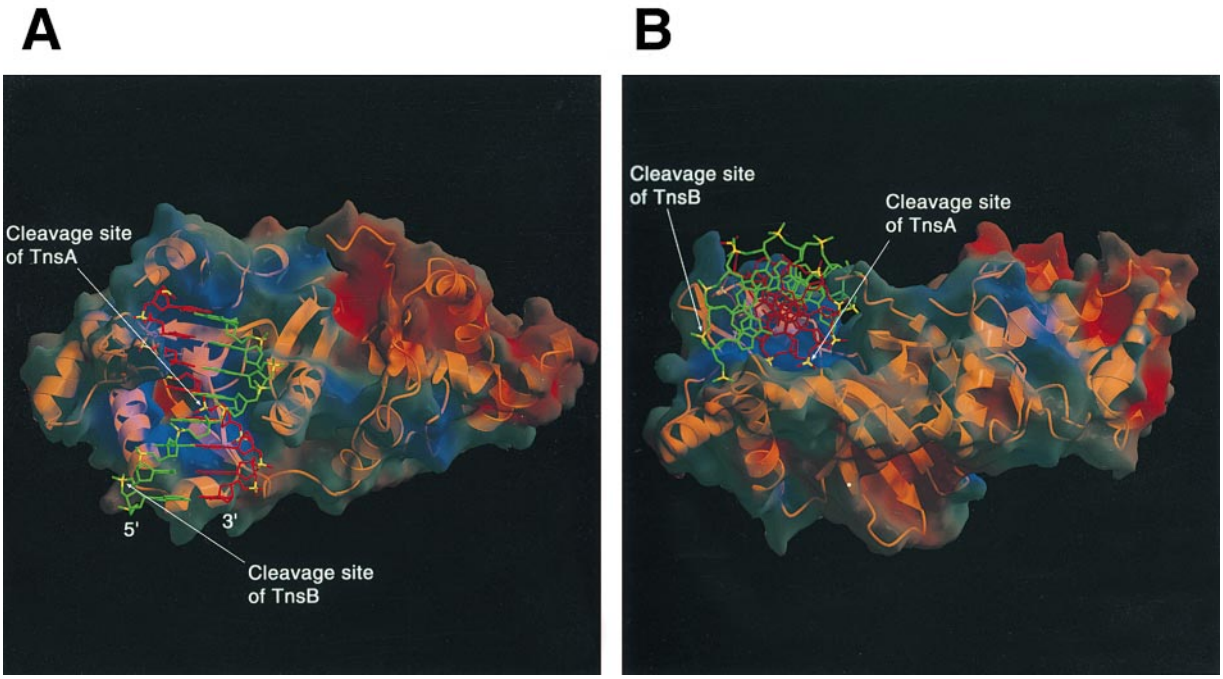

Figure 6. Proposed Model of TnsA Binding to a Tn7 Transposon End

The two views are rotated 90° with respect to each other. The protein surface was generated with SPOCK (Christopher, 1998) and rendered with Raster3D (Merritt and Bacon, 1997), and red and blue represent regions of positive and negative surface electrostatic potential, respectively. The transposon DNA sequence used is that of Tn7R and flanking sequence and was modeled by analogy to the DNA complexes of BamHI, EcoRV, and the Vsr endonuclease. The loop between residues 27 and 42, expected to change conformation upon DNA binding, was removed from the model.

transposition is also reflected in the nature of the active sites characteristic of the two protein folds. The active sites of transposition enzymes possessing the RNaseH-like fold tend to be relatively shallow and flexible. For example, one of the three critical acidic residues of the retroviral integrases is located on a flexible loop that has been captured in a number of different conformations in the various structures that have been determined. This flexibility likely reflects the need to not only use two different attacking nucleophiles as the reaction proceeds, but also to accommodate the target strand DNA substrate while the donor strand remains bound at the active site (Kennedy et al., 2000). In contrast, restriction endonucleases typically have active sites that are more deeply located and rigid, consistent with a simple hydrolysis reaction. Thus, in hindsight, the choice of a restriction endonuclease fold to concurrently catalyze 5′ cleavage and to restrict the possibility for further reaction at the 5′ transposon end seems inspired.

Different transposon systems have adopted diverse strategies to mediate the DNA breaks that disconnect a transposon from its donor site (Haren et al., 1999). For example, bacterial elements IS10 and IS50 also initiate transposition by the introduction of double-strand breaks at the transposon termini, but the mechanism of forming these double-strand breaks is quite distinct from Tn7. In these cases, the initiating step is the introduction of nicks that expose the 3′ OH ends of the elements. These 3′ OH ends then directly attack the 5′ strands, generating a hairpin on the transposon end. Subsequent opening of the hairpin by the transposase yields the fully broken transposon end ready to interact with the target DNA. Yet another strategy is used by IS911 that involves single-stranded transposon circles. It remains to be established in other double-strand break systems that have a single transposon subunit such as Tc1 and Tc3 transposition how double-strand breaks occur (Plasterk, 1996).

Despite the varied chemical roles of TnsA and TnsB, interaction between these proteins is necessary to execute transposition. One of the distinguishing features of Tn7 is the tight control of the transposition event: the full assembly of Tns proteins, donor DNA, and target DNA is needed before the reaction can proceed. It is consistent with this stringent control that, during evolution, Tn7 has acquired a catalytic subunit that has the unique function of cleaving the 5′, nontransferred DNA ends. This confers on Tn7 a greater level of regulation than seen in other systems. It will be interesting to dissect how this interplay occurs. Structural analysis of a TnsA + TnsB complex is an appealing next step.

Experimental Procedures

Protein Purification and Crystallization

The *tnsA* gene was cloned into the intein expression vector, pCYB2, (NEB) resulting in a two amino acid spacer (-PG-) between the authentic C terminus and the intein autoproteolytic cleavage site. Optimal protein expression was obtained in *Escherichia coli* strain XA90 (Baker et al., 1984) by growth at 37°C in LB broth until $OD_{600\ nm}$ = 0.6, followed by induction at 20°C by addition of IPTG to a final concentration of 0.1 mM. Cells were harvested 5 hr later and resuspended in 20 mM HEPES (pH 7.8), 0.5 M NaCl, 0.1 mM EDTA, 0.1%

(v/v) Triton X-100, 10% (w/v) glycerol. All subsequent steps were at 4°C unless otherwise stated. After sonication, soluble material was loaded onto a chitin beads (NEB) column. The column was washed with 10 column volumes (CV) of resuspension buffer, moved to room temperature, and washed with 20 mM HEPES (pH 7.8), 0.15 M NaCl, 0.1 mM EDTA, 0.1% (v/v) Triton X-100, 10% (w/v) glycerol, 10 mM $MgCl_2$, 5 mM ATP to remove chaperonin proteins. The column was returned to 4°C, washed with 3 CV 20 mM HEPES (pH 7.8), 0.5 M NaCl, 0.1 mM EDTA, 10% (w/v) glycerol, 30 mM DTT, and left overnight. Free TnsA was eluted the following day and dialyzed against 25 mM HEPES (pH 7.8), 1 M NaCl, 1 mM EDTA, 2 mM DTT, 10% (w/v) glycerol prior to concentration and gel filtration on a TSK-Gel G3000SW column. Fractions containing TnsA were collected, and the protein was concentrated to 7–8 mg/ml and exchanged into 25 mM HEPES (pH 7.8), 0.5 M NaCl, 1 mM EDTA, 2 mM DTT, 0.01% (v/v) Triton X-100, 10% (w/v) glycerol, 10% (v/v) isopropanol prior to crystallization. The final yield was approximately 1 mg TnsA per liter cells.

Crystals were grown at 4°C by the hanging drop method by mixing TnsA in a 1:1 ratio with buffer 1 (10% PEG 8000, 0.1 M Tris [pH 8.0]) containing 50 mM $MgCl_2$, by vapor diffusion against a well solution consisting of 90% buffer 1 containing 0.4 M NaCl/10% isopropanol. Repeated macroseeding was required to obtain diffraction quality crystals. The space group is $P2_1$ with cell dimensions a = 38.34 Å, b = 83.68 Å, c = 97.94 Å, and β = 94.07°. There are two molecules in the asymmetric unit. Crystals were cryoprotected by transfer into increasing amounts of glycerol to a final concentration of 20% (w/v) glycerol, and then flash-frozen in liquid propane. For MIR, an Hg derivative was obtained (2 hr soak at 4°C in 1 mM ethylmercuric acetate prior to cryoprotection) and a Pb derivative (24 hr soak at 4°C in 10 mM trimethyllead acetate prior to cryoprotection).

Structure Determination

Diffraction data were collected at 95 K on Raxis II and IV image plate detectors, using Cu$K\alpha$ radiation generated in an RU200 rotating anode source and collimated with total reflection mirrors. Data were integrated and scaled internally using the HKL suite (Otwinowski and Minor, 1997). The four most prominent sites in the Hg derivative were located using SOLVE (Terwilliger and Berendzen, 1999), while further sites and all the sites in the Pb derivative were determined with difference Fouriers. Heavy atom parameter refinement, solvent flattening (Wang, 1985), and two-fold noncrystallographic symmetry averaging were carried out using PHASES (Furey and Swaminathan, 1997) at 3 Å resolution. The correlation coefficent between the two noncrystallographically related molecules was 0.59 before averaging. When the process was completed, the electron density of the polypeptide appeared continuous except for the first 30 residues and between residues 180 and 186. Clearly defined side chain densities were visible in most cases. The model was built with O (Jones et al., 1991) and refined with several rounds of molecular dynamics, energy minimization, and restrained B factor refinement using XPLOR 3.1 (Brünger, 1992). Bulk solvent correction and the TNT B factor restraint library (Tronrud, 1996) was used throughout, and 5% of the reflections (1143) were kept separate to monitor R_{free} and never used in refinement. Noncrystallographic restraints were also included. At the final stages, 292 water molecules and 4 Mg^{2+} and 2 Cl^- ions were added and the refinement was completed at 2.4 Å resolution. The current model contains all residues between 8 and 261; 89.9% of the nonglycine residues are in the most favored regions of the Ramachandran plot with none of them in the disallowed regions. Although the two molecules in the asymmetric unit are related by a nearly perfect two-fold axis, sedimentation equilibrium experiments indicated that the molecule exists predominantly as a monomer in solution (data not shown).

Acknowledgments

We thank K. Mizuuchi and W. Yang for constructive discussions and D. Davies and R. Craigie for ongoing support. We are grateful to R. Ghirlando for performing sedimentation equilibrium measurements. N. L. C. is an Investigator of the Howard Hughes Medical Institute.

Received February 25, 2000; revised April 11, 2000.

References

Aggarwal, A.K. (1995). Structure and function of restriction endonucleases. Curr. Opin. Struct. Biol. *5*, 11–19.

Athanasiadis, A., Vlassi, M., Kotsifaki, D., Tucker, P.A., Wilson, K.S., and Kokkinidis, M. (1994). Crystal structure of *Pvu*II endonuclease reveals extensive structural homologies to *Eco*RV. Struct. Biol. *1*, 469–475.

Baker, T.A., Grossman, A.D., and Gross, C.A. (1984). A gene regulating the heat-shock response in *Escherichia coli* also affects proteolysis. Proc. Natl. Acad. Sci. USA *81*, 6779–6783.

Ban, C., and Yang, W. (1998). Structural basis for MutH activation in *E. coli* mismatch repair and relationship of MutH to restriction endonucleases. EMBO J. *17*, 1526–1534.

Beese, L.S., and Steitz, T.A. (1991). Structural basis for the 3′–5′ exonuclease activity of *Escherichia coli* DNA polymerase I: a two metal ion mechanism. EMBO J. *10*, 25–33.

Bozic, D., Grazulis, S., Siksnys, V., and Huber, R. (1996). Crystal structure of *Citrobacter freundii* restriction endonuclease *Cfr*10I at 2.15 Å resolution. J. Mol. Biol. *255*, 176–186.

Brünger, A.T. (1992). X-PLOR Version 3.1 (New Haven, CT: Yale University Press).

Bujacz, G., Alexandratos, J., Wlodawer, A., Merkel, G., Andrake, M., Katz, R.A., and Skalka, A.M. (1997). Binding of different divalent cations to the active site of avian sarcoma virus integrase and their effects on enzymatic activity. J. Biol. Chem. *272*, 18161–18168.

Carson, M. (1991). RIBBONS 2.0. J. Appl. Crystallogr. *24*, 958–961.

Cheng, X., Balendiran, K., Schildkraut, I., and Anderson, J.E. (1994). Structure of *Pvu*II endonuclease with cognate DNA. EMBO J. *13*, 3927–3935.

Christopher, J.A. (1998). SPOCK: The Structural Properties Observation and Calculation Kit (Program Manual), The Center for Macromolecular Design, Texas A and M University, College Station, TX.

Craig, N.L. (1996). Transposon Tn7. Curr. Top. Microbiol. Immunol. *204*, 27–48.

Davies, D.R., Braam, L.M., Reznikoff, W.S., and Rayment, I. (1999). The three-dimensional structure of a Tn5 transposase-related protein determined to 2.9-Å resolution. J. Biol. Chem. *274*, 11904–11913.

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.

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

Haren, L., Ton-Hoang, B., and Chandler, M. (1999). Integrating DNA: transposases and retroviral integrases. Annu. Rev. Microbiol. *53*, 245–281.

Harrison, S.C., and Aggarwal, A.K. (1990). DNA recognition by proteins with the helix-turn-helix motif. Annu. Rev. Biochem. *59*, 933–969.

Holm, L., and Sander, C. (1997). Dali/FSSP classification of three-dimensional protein folds. Nucleic Acids Res. *25*, 231–234.

Jones, D.T. (1999). Protein secondary structure prediction based on position-specific scoring matrices. J. Mol. Biol. *292*, 195–202.

Jones, M., Wagner, R., and Radman, M. (1987). Mismatch repair of deaminated 5-methyl-cytosine. J. Mol. Biol. *194*, 155–159.

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 Crystallog. A *47*, 110–119.

Kennedy, A.K., Haniford, D.B., and Mizuuchi, K. (2000). Single active site catalysis of the successive phosphoryl transfer steps by DNA transposases: insights from phosphorothioate stereoselectivity. Cell *101*, 295.

Kim, Y., Grable, J.C., Love, R., Greene, P.J., and Rosenberg, J.M. (1990). Refinement of Eco RI endonuclease crystal structure: a revised protein chain tracing. Science *249*, 1307–1309.

Kleckner, N., Chalmers, R.M., Kwon, D., Sakai, J., and Bolland, S.

(1996). Tn10 and IS10 transposition and chromosome rearrangements: mechanisms and regulation in vivo and in vitro. In Transposable Elements, H. Saedler and A. Gierl, eds. (Heidelberg, Germany: Springer-Verlag), pp. 49–82.

Kostrewa, D., and Winkler, F.K. (1995). Mg^{2+} binding to the active site of *Eco*RV endonuclease: a crystallographic study of complexes with substrate and product DNA at 2 Å resolution. Biochem. *34*, 683–696.

Kovall, R.A., and Matthews, B.W. (1997). Toroidal structure of λ-exonuclease. Science *277*, 1824–1827.

Kovall, R.A., and Matthews, B.W. (1999). Type II restriction endonucleases: structural, functional, and evolutionary relationships. Curr. Opin. Chem. Biol. *3*, 578–583.

Krementsova, E., Giffin, M.J., Pincus, D., and Baker, T.A. (1998). Mutational analysis of the Mu transposase. Contributions of two distinct regions of domain II to recombination. J. Biol. Chem. *273*, 31358–31365.

Li, H., Trotta, C.R., and Abelson, J. (1998). Crystal structure and evolution of a transfer RNA splicing enzyme. Science *280*, 279–284.

Li, L., Wu, L.P., and Chandrasegaran, S. (1992). Functional domains in *Fok* I restriction endonuclease. Proc. Natl. Acad. Sci. USA *89*, 4275–4279.

Lieb, M. (1987). Bacterial genes *mutL*, *mutS*, and *dcm* participate in repair of mismatches at 5-methylcytosine sites. J. Bacteriol. *169*, 5241–5246.

May, E.W., and Craig, N.L. (1996). Switching from cut-and-paste to replicative Tn7 transposition. Science *272*, 401–404.

McKay, D.B., and Steitz, T.A. (1981). Structure of catabolite gene activator protein at 2.9 Å resolution suggests binding to left-handed B-DNA. Nature *290*, 744–749.

Merritt, E.A., and Bacon, D.J. (1997). Raster3D: photorealistic molecular graphics. Methods Enzymol. *277*, 505–524.

Mizuuchi, K. (1992). Transpositional recombination: mechanistic insights from studies of Mu and other elements. Annu. Rev. Biochem. *61*, 1011–1051.

Newman, M., Strzelecka, T., Dorner, L.F., Schildkraut, I., and Aggarwal, A.K. (1994). Structure of restriction endonuclease *Bam*HI and its relationship to *Eco*RI. Nature *368*, 660–664.

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

Plasterk, R.H. (1996). The Tc1/mariner transposon family. Curr. Top. Microbiol. Immunol. *204*, 125–143.

Ramakrishnan, V., Finch, J.T., Graziano, V., Lee, P.L., and Sweet, R.M. (1993). Crystal structure of globular domain of histone H5 and its implications for nucleosome binding. Nature *362*, 219–223.

Reznikoff, W.S. (1993). The Tn5 transposon. Annu. Rev. Microbiol. *47*, 945–963.

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.

Sarnovsky, R.J., May, E.W., and Craig, N.L. (1996). The Tn7 transposase is a heteromeric complex in which DNA breakage and joining activities are distributed between different gene products. EMBO J. *15*, 6348–6361.

Sharpe, P.L., and Craig, N.L. (1998). Host proteins can stimulate Tn7 transposition: a novel role for the ribosomal protein L29 and the acyl carrier protein. EMBO J. *17*, 5822–5831.

Skirgaila, R., Grazulis, S., Bozic, D., Huber, R., and Siksnys, V. (1998). Structure-based redesign of the catalysis/metal binding site of *Cfr*10I restriction endonuclease reveals importance of spatial rather than sequence conservation of active centre residues. J. Mol. Biol. *279*, 473–481.

Terwilliger, T.C., and Berendzen, J. (1999). Automated structure solution for MIR and MAD. Acta Crystallogr. D *55*, 849–861.

Tronrud, D.E. (1996). Knowledge-based B-factor restraints for the refinement of proteins. J. Appl. Crystallogr. *29*, 100–104.

Tsutakawa, S.E., Muto, T., Kawate, T., Jingami, H., Kunishima, N., Ariyoshi, M., Kohda, D., Nakagawa, M., and Morikawa, K. (1999a). Crystallographic and functional studies of very short patch repair endonuclease. Mol. Cell *3*, 621–628.

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

Viadiu, H., and Aggarwal, A.K. (1998). The role of metals in catalysis by the restriction endonuclease *Bam*HI. Nat. Struct. Biol. *5*, 910–916.

Wah, D.A., Hirsch, J.A., Dorner, L.F., Schildkraut, I., and Aggarwal, A.K. (1997). Structure of the multimodular endonuclease FokI bound to DNA. Nature *388*, 97–100.

Wah, D.A., Bitinaite, J., Schildkraut, I., and Aggarwal, A.K. (1998). Structure of *Fok*I has implications for DNA cleavage. Proc. Natl. Acad. Sci. USA *95*, 10564–10569.

Wang, B.-C. (1985). Resolution of phase ambiguity in macromolecular crystallography. Methods Enzymol. *115*, 90–112.

Winkler, F.K., Banner, D.W., Oefner, C., Tsernoglou, D., Brown, R.S., Heathman, S.P., Bryan, R.K., Martin, P.D., Petratos, K., and Wilson, K.S. (1993). The crystal structure of EcoRV endonuclease and of its complexes with cognate and non-cognate DNA fragments. EMBO J. *12*, 1781–1795.

Protein Data Bank ID Code

Coordinates have been deposited with the Protein Data Bank (ID code 1F1Z) and, until release, are available via e-mail from F. D. (dyda@ulti.niddk.nih.gov).