*Published online 12 September 2008* *Nucleic Acids Research, 2008, Vol. 36, No. 18* ***5855–5862***
*doi:10.1093/nar/gkn577*

# Phosphate coordination and movement of DNA in the Tn5 synaptic complex: role of the (R)YREK motif

**Vadim A. Klenchin[1], Agata Czyz[1,2], Igor Y. Goryshin[1], Richard Gradman[1], Scott Lovell[1], Ivan Rayment[1] and William S. Reznikoff[1,*]**

[1]Department of Biochemistry, University of Wisconsin at Madison, 433 Babcock Drive, Madison, WI 53706, USA and [2]Laboratory of Molecular Biology (affiliated with the University of Gdansk), Institute of Biochemistry and Biophysics, Polish Academy of Sciences, Kladki 24, Gdansk 80-822, Poland

Received June 11, 2008; Revised August 27, 2008; Accepted August 28, 2008

## ABSTRACT

**Bacterial DNA transposition is an important model system for studying DNA recombination events such as HIV-1 DNA integration and RAG-1-mediated V(D)J recombination. This communication focuses on the role of protein–phosphate contacts in manipulating DNA structure as a requirement for transposition catalysis. In particular, the participation of the nontransferred strand (NTS) 5′ phosphate in Tn5 transposition strand transfer is analyzed. The 5′ phosphate plays no direct catalytic role, nonetheless its presence stimulates strand transfer ~30-fold. X-ray crystallography indicates that transposase–DNA complexes formed with NTS 5′ phosphorylated DNA have two properties that contrast with structures formed with complexes lacking the 5′ phosphate or complexes generated from in-crystal hairpin cleavage. Transposase residues R210, Y319 and R322 of the (R)YREK motif coordinate the 5′ phosphate rather than the subterminal NTS phosphate, and the 5′ NTS end is moved away from the 3′ transferred strand end. Mutation R210A impairs the 5′ phosphate stimulation. It is posited that DNA phosphate coordination by R210, Y319 and R322 results in movement of the 5′ NTS DNA away from the 3′-end thus allowing efficient target DNA binding. It is likely that this role for the newly identified RYR triad is utilized by other transposase-related proteins.**

## INTRODUCTION

Proteins that alter the structure of DNA fall into two general categories: those such as DNA polymerases that move along a DNA molecule while performing catalysis and those that bind to a specific site and perform catalysis only at that site. Within this latter class, DNA transposases (Tnps) and related proteins such as HIV-1 integrase and RAG-1 recombinase execute a particularly complex set of tasks that require two or more chemically distinct catalytic events to be performed by a single active site. In order to accomplish this, the proteins need to alter the detailed DNA structures in a carefully choreographed fashion. In this communication we report functional and structural studies of a novel protein–DNA interaction that facilitates such a task using the Tn*5* Tnp system as a model.

DNA transposition is a fundamental process involved in genome evolution. Transposons, the DNA sequences that undergo transposition, are found in almost all organisms (1). In addition to their importance in generating genome plasticity, transposons are also important models for understanding other biochemical systems that involve DNA recombination events. For instance, many Tnps are members of a superfamily of proteins that include retroviral integrases and the RAG-1 recombinase that is responsible for the catalyzing a key step in immunoglobulin gene formation (2–5). The members of this super-family of proteins share active site architectures and have similar catalytic mechanisms. Thus information found for one member of the superfamily is of immediate importance for the understanding of similar phenomena for other members.

The Tn*5* transposition pathway is depicted in Figure 1. Of particular note, this figure shows Tn*5* Tnp's need to accommodate six different DNA species during the transposition process. These DNA species include both highly specific DNA sequences for some steps and essentially random DNA sequences for others. The dimeric nature of the synaptic complex introduces additional complexity. For instance, recent data have indicated that DNA cleavage (steps **1** through **3**) proceeds asymmetrically in the dimeric

*To whom correspondence should be addressed. Tel: +1 508 289 7253; Fax: +1 508 457 4727; Email: breznikoff@mbl.edu, wsreznik@wisc.edu

The authors wish it to be known that, in their opinion, the first two authors should be regarded as joint First Authors.

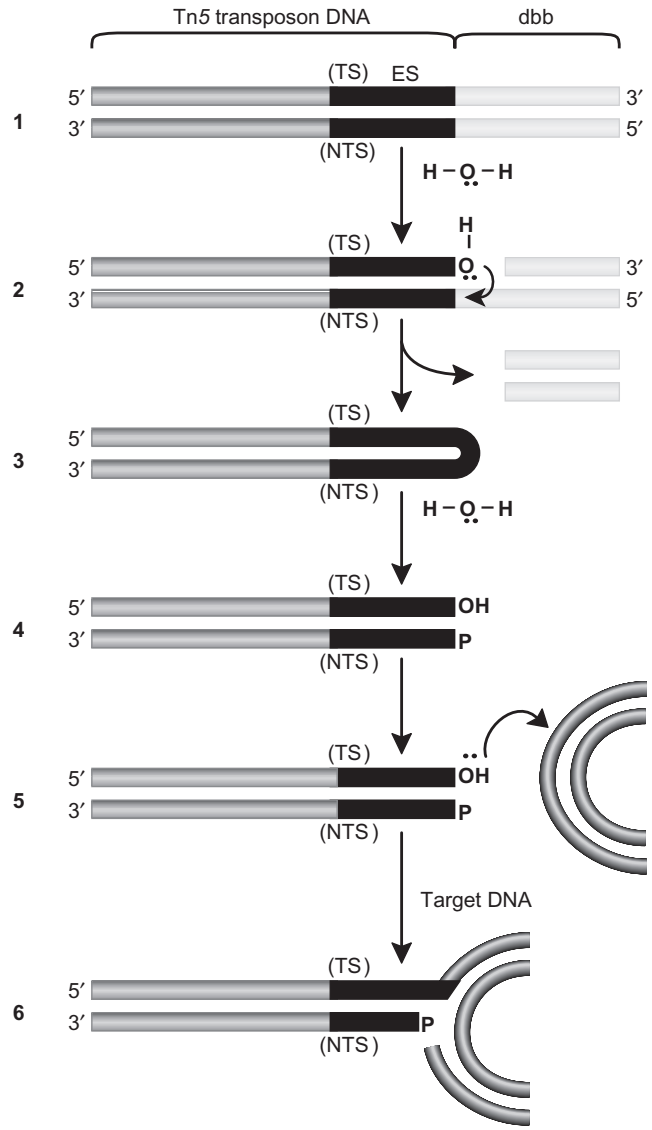


**Figure 1.** Steps of Tn*5* transposition. A scheme depicting distinct Tn*5* transposition steps and corresponding DNA-protein species that are as follows. (Step **1**) 'The initial substrate'. Transposon end sequence (ES) DNA is found adjacent to donor backbone (dbb) DNA; (Step **2**) 'The nicked substrate'. Transposon ES DNA-donor DNA containing a precise nick exposing a 3′ hydroxyl at the end of the ES DNA transferred strand (TS); (Step **3**) 'Hairpin formation'. Transposon ES DNA in which a hairpin links the transferred strand (TS) and non-transferred strand (NTS); (Step **4**) 'Hairpin resolution'. Transposon ES DNA with a blunt cleaved end resulting from resolution of the hairpin; (Step **5**) 'Target capture'. A complex that contains both cleaved ES DNA and a target DNA sequence; (Step **6**) 'Strand transfer'. A complex in which the ES TS has been covalently linked to one strand of the target DNA to generate the final product. Also shown are nucleophilic species involved in sequential catalytic steps.

synaptic complex suggesting that one of the two Tnp–DNA complexes must undergo a conformational change that modifies the precise Tnp–DNA association (6,7). No information exists as to the nature of this required DNA conformation change nor how it is accomplished.

Tnp proteins have been grouped into families based upon sequence similarities and shared mechanistic characteristics. The IS*4* Tnp family includes the Tnps encoded by IS*10* (Tn*10*) and IS*50* (Tn*5*) among others (8–10). These two Tnps have been thoroughly characterized by genetic and biochemical approaches (11–13). In the case of Tn*5*, we also have detailed structural information available for the Tnp–transposon DNA synaptic complex following DNA cleavage (14,15). A characteristic feature of the IS*4* Tnp family is the YREK motif in which the glutamate residue is one of the three catalytic residues and the lysine residue provides an important contact to the transferred strand phosphate between bases 1 and 2 analogous to a similar contact for HIV-1 integrase (8,14,16). Results from a previous study (17) indicate that the Y319 and R322 residues play an important role in the hairpin resolution step of the Tn*5* transposition pathway (Figure 1, steps **3** to **4**), but their roles in other steps are not well defined. The high conservation of the YREK motif in conjunction with the existence of intimate contacts between the Y319 and R322 and a nontransferred strand (NTS) phosphate (14,15; also PDB code 1MUS, Lovell *et al.*, unpublished data) suggests that these protein–DNA phosphate interactions may also be essential in other steps of the transposition process.

A critical series of events in transposition is target capture and strand transfer of the transposon 3′-end into the target DNA (Figure 1, steps **5** and **6**). Only the 3′ hydroxyl of the transposon end sequence DNA plays a direct role in strand transfer catalysis. However, it is possible that other aspects of the end sequence DNA structure play important roles in determining the efficiency of the target capture/strand transfer process. In this communication we demonstrate that the chemical structure of the NTS 5′-end plays a previously unrecognized critical role in this late stage of transposition. We present biochemical, crystallographic and genetic data supporting this conclusion, and we address factors that underlie the mechanistic and functional importance of this finding. It is demonstrated that two of the residues in the YREK motif (Y319 and R322), in combination with an additional residue conserved in Tn*5* related Tnps (R210) (18), coordinate the NTS 5′ phosphate. This coordination likely stabilizes a new DNA conformation that creates space for the proper binding of target DNA. Thus this proposal adds an additional Tnp—DNA species between steps 4 and 5 to those presented in Figure 1. The importance of the 5′ phosphate and its coordination by an indispensable RYR triad (or its equivalent) is likely to be a common feature among numerous DNA transposition-like systems.

## MATERIALS AND METHODS

### DNA substrates

Two different transposon end sequences were used in these experiments. The oligonucleotides used in the strand transfer experiments were provided by the Operon Inc. The 40-mer DNA substrate [5′-**CTGTCTCTTATACAC ATCT**TGAGTGAGTGAGCATGCATGT-3′ (NTS) and its complement] contains the 19-bp hyperactive mosaic end sequence (12) (shown in bold) and 21 bp of additional

transposon DNA. The double stranded DNA substrate was either phosphorylated on the 5′-end of the NTS or not as indicated. Both DNA substrates were also labeled with 6-FAM fluorescein at the 5′-end of the transferred strand. For the crystallography, the 20-bp 5′ phosphorylated DNAs [CTGACTCTTATACACAAGTC (NTS) and its complement] were from Integrated DNA Technology. This short DNA sequence represents the outside end sequence of Tn5 (12) with an additional CG base pair at the end distal from where the scissile phosphodiester bonds would be located. The double-stranded DNA substrates were annealed by heating at 95°C for 1 min and then cooling at a rate 0.1°C/s to 4°C at a final concentration of 50 μM (40 bp DNA) or 100 μM (20 bp DNA) in the presence of 10 mM NaCl, 10 mM Tris–HCl, pH 7.5. The pUC19 DNA for strand transfer experiments were prepared with a Qiagen MaxiPrep kit.

### Tnp purification

The hyperactive Tn*5* E54K/L372P Tnp and its mutants were overexpressed and purified for biochemical assays as described previously (19) except that the elution and storage buffer contained 20 mM HEPES (pH 7.5), 100 mM potassium glutamate and 100 μg/ml tRNA. Purity of the proteins was assessed using SDS–PAGE electrophoresis with Coomassie staining. Proteins were quantified by the Bradford method using bovine serum albumin (Pierce, Rockford, IL, USA) as a standard (20). The purification of Tn*5* E54K/L372P Tnp for crystallographic study was performed exactly as described previously (14). Mutant Tnp constructs were generated as described previously (21).

### Crystallography

The annealed 20-bp DNA was combined with Tn*5* Tnp at a 1:1 molar ratio. After overnight incubation at 4°C, the protein–DNA complexes were dialyzed against 300 mM KCl, 20 mM HEPES (pH 7.7), 2 mM EDTA and concentrated by ultrafiltration. The final concentration of the protein was ~9 mg/ml. Crystals of the Tnp–DNA complex were grown by the micro batch technique (22) at room temperature. Typically, 7.5 μl of protein–DNA complexes were mixed with 7–8 μl of precipitant solution containing 13% (w/v) oMe-PEG 5000, 100 mM Bis–Tris (pH 6.5) and 100 mM $AmSO_4$. Crystals nucleated spontaneously and reached a maximum size of 0.2–0.25 mm in 3–4 weeks. The crystals were cryo-preserved by transfer into the precipitant solution containing 150 mM KCl, followed by stepwise equilibration with glycerol to 25% (w/v), and frozen in a nitrogen stream at 100°K. Diffraction data were collected at the 19ID beamline of the Structural Biology Center at the Advanced Photon Source in Argonne, IL, as 360 frames with 0.5° oscillations at a wavelength of 1.00964 Å. The data were integrated and scaled with HKL2000 (23). The structure was solved by molecular replacement with Molrep (24) and refined with Refmac utilizing TLS and restrained individual B factors refinement (25,26). Water molecules were added to the coordinate set with ARP/wARP (27) with subsequent manual verification. Figures were prepared with Pymol (http://www.pymol.org).

### Strand transfer assays

The strand transfer assays were performed as described in Gradman *et al.* (21). In brief, the 40-bp DNA substrates (25 nM) and E54K/L372P Tnp (250 nM) were mixed in buffer containing 100 mM potassium glutamate, 20 mM HEPES, pH 7.5 and 100 μg/ml tRNA. The reactions were incubated at 37°C for 1.5 h and then transferred to 20°C. After addition of target super-coiled pUC19 plasmid DNA (15 nM, final concentration), strand transfer was initiated by addition of magnesium acetate to 10 mM final concentration. At the indicated times, 20 μl samples were then withdrawn, mixed with 5 μl of Promega Blue/Orange 6X Loading Dye and immediately placed at −20°C. DNA fragments were analyzed by electrophoresis on a 2% agarose gel at room temperature. The fluorescent DNA fragments were visualized using the 526 nm setting on the Typhoon Variable Mode Imager. All gels were quantified using Image Quant Total Lab software.

## RESULTS AND DISCUSSION

### Phosphorylation state of the NTS determines the efficiency of strand transfer

Strand transfer reactions were initiated by addition of the target DNA to preformed, Tn*5* paired end complexes (analogous to synaptic complexes) containing two protomers of Tnp and two end sequence containing 40-mer DNAs. The 5′-end of the NTS contained either a hydroxyl or a phosphate as indicated. The reactions were performed in the presence of $Mg^{2+}$ and were terminated at the indicated times. Strand transfer products were visualized through agarose gel electrophoresis. Two types of products are observed in this reaction: single end strand transfer events, in which one labeled 40-mer DNA is covalently linked to the one strand of pUC19 DNA (migrating as a relaxed circle), and double end strand transfer events in which a linear form of the target is generated (Figure 2). In our experiments the double end strand transfer products formed in the strand transfer assays with the 5′ phosphate substrate are detected 5 min after the target DNA and $Mg^{2+}$ have been added. In contrast, double end strand transfer products formed with the 5′-hydroxyl substrate were evident no earlier than after 1 h incubation. The results of these assays demonstrated that, in the presence of the phosphate group on the NTS 5′-end, the completion of the transposition process (characterized by the double end strand transfer band) occurs ~30 times faster compared to the data for 5′ hydroxyl DNA (Figure 2B). Qualitatively similar results are obtained by summing the single end transfer and double end transfer products. This finding clearly indicates that the 5′ phosphate on the NTS plays an important role in the target capture and/or the DNA strand transfer reaction (because we measure these two events through a strand transfer assay, we will use 'strand transfer' to refer to the combination of both events).

The above results prompt the question: how can the presence of a charged, bulky group (a phosphate) on the NTS 5′-end facilitate the reaction? In an effort to discover a structural basis for this unexpected finding, crystallographic analysis of the Tn5 Tnp complexed with its phosphorylated ES DNA substrate was performed.

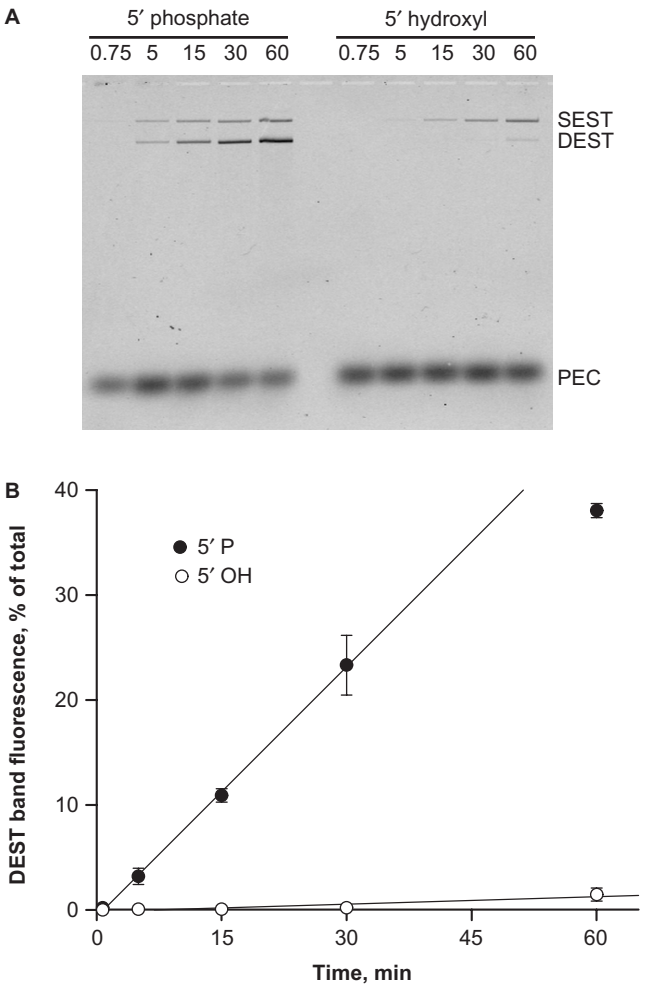


**Figure 2.** The nontransferred strand 5′ phosphate stimulates strand transfer. (**A**) Paired end complexes containing fluorescently labeled end sequence DNA, whose NTS contained either a 5′ hydroxyl or a 5′ phosphate, were mixed with super-coiled DNA in the presence of $Mg^{2+}$. The combined steps of target capture and strand transfer were visualized by agarose gel analysis of reaction products from different time points after the start of the reaction. The bottom band (PEC) represents unreacted paired end complexes. The SEST band represents strand transfer of a single ES into the plasmid DNA forming open circle molecules. The double end strand transfer (DEST) band represents the final DNA transposition event in which both end sequences are transferred into the plasmid DNA forming linear molecules. The NTS 5′ phosphate stimulates both SEST and DEST product formation. (**B**) Summary of the quantitative analysis of gels similar to the one shown in Figure 2A. The time course of DEST product formation for the 5′ hydroxyl (open circles) and 5′ phosphate end sequence DNA substrates (closed circles) is compared using data derived from the gels similar to that presented in Figure 2A. Error bars represent standard errors of the triplicate experimental points. Solid lines represent linear regressions calculated using the data in the linear range of the kinetic response. The difference between these calculated rates of the DEST formation is 32-fold. Qualitatively similar results are obtained by summing the SEST and DEST data band intensities.

## Crystallographic analysis

The structure of the Tn5 paired end complex with a nonphosphorylated substrate has been solved previously (14). To provide a counterpart to the biochemical analysis described above, a 5′ phosphorylated DNA substrate otherwise identical to that of Davies *et al.* (14), was used to grow crystals and the resulting structure was solved to 2.5 Å resolution. Crystallographic statistics are presented in Table 1 and the representative electron density is shown in the Supplementary Figure S1.

With regard to the overall protein conformation, the resulting structure is very similar to the one found for the nonphosphorylated DNA complex (alpha carbons RMSD from 1MUH is only 0.51 $Å^2$). The few notable differences that exist include a different orientation of the N-terminus (up to R8), a slight outward movement of the loop E285-K297 away from the rest of the DNA–protein complex and W323 adopting a different rotamer conformation to avoid a steric clash with the first nucleotide base of the NTS (see below). In all of these features the new structure most closely resembles another Tn5 paired end complex structure, PDB code 1MUR (15), which is reflected in the even lower value for RMSD between them, 0.34 $Å^2$. It should be stressed that 1MUR represents a nonphysiological transposon DNA–Tnp complex. It includes an additional nucleotide added to the 3′-end of the transferred strand, thus crudely approximating the end product of the strand transfer reaction; note that the terminal 3′ nucleoside is disordered in 1MUR and its location could not be resolved. Nevertheless, the NTS in 1MUR is 5′ phosphorylated and that, in hindsight, appears to be the major determinant for the observed differences between structures containing and lacking the terminal 5′ phosphate.

The presence of 5′ phosphate has a dramatic effect on the conformation of the DNA termini contacting the protein (Figure 3). In the case of phosphorylated DNA, the 3′-terminal base flips ~120°, failing to form Watson–Crick base pairing with the opposite chain and also losing a stacking interaction with the 5′-terminal base. The 5′-end is moved further away from the other strand, with the terminal phosphate now occupying the position of the phosphate bridging bases 1 and 2 in the nonphosphorylated structure. In the NTS 5′ phosphate structure, the 5′ phosphate is firmly coordinated by R210, Y319 and R322 (Figure 4)—all residues that are strictly conserved across the wide range of Tn5 Tnp-related homologs (18). In contrast, in the structure lacking the 5′ phosphate it is the phosphate between bases 1 and 2 that is coordinated by R210, Y319 and R322.

It is important to note that the coordination of the 5′-terminal phosphate is not dependent on the presence or absence of metal ions. Unlike all previous Tn5 Tnp–DNA crystals, for which crystal soaks with manganese markedly improved the diffraction limit (14,15; also PDB code 1MUS), soaking of the present crystals with $Mg^{2+}$ or $Mn^{2+}$ not only did not improve, but, rather, negatively affected their diffraction. For this reason, metal soaks were omitted in the present study. Nevertheless, since the DNA conformation observed in this study is

essentially the same as that for 1MUR (a structure that has phosphorylated transposon ends and does contain $Mn^{2+}$), it is clear that the observed changes are related strictly to the presence of 5′ phosphate and *not* to the absence of the metal ion.

The overall consequence of the altered structure at DNA termini for the 5′ phosphorylated end sequence versus the 5′ hydroxylated end sequence (movement of the 5′-terminal nucleotide and flipping of the 3′ nucleotide) is that considerable room has been made available between the two strands at the DNA tip.

### Genetic and biochemical evaluation of the importance of 5′ phosphate coordination

With the critical 5′ phosphate being coordinated by three strictly conserved residues (18), and with the two of these, Y319 and R322 being components of the well-known IS*4* Tnp family YREK motif (8,14), our attention then turned to the functional characterization of the individual coordinating residues: R210, Y319 and R322. For each of these positions, an alanine substitution mutant was made and assayed for the efficiency of strand transfer using equivalent quantities of paired end complexes made with the two DNA substrates. TnpY319A is known from previous studies to be functional for paired end complex formation, 3′ nick formation and hairpin formation but impaired in hairpin resolution (17). In the present experiments, TnpY319A was shown to be completely defective in strand transfer of the 5′-hydroxylated DNA substrate and demonstrated a barely detectable level of strand transfer with the 5′-phosphorylated DNA (data not shown). Thus Y319 still manifested a preference for having a 5′ phosphate on the substrate but the magnitude of the effect could not be determined. Nevertheless, it is clear that Y319 plays an important role in strand transfer. Previous studies have demonstrated that TnpR322A is modestly impaired for hairpin formation, and strongly impaired for hairpin resolution (17). TnpR322A demonstrated no strand transfer activity for either substrate (data not shown). Thus, R322 also plays an important role in strand transfer.

The most interesting and informative of the three mutants is TnpR210A. Figure 5 shows that the introduction of an alanine in place of an arginine at position 210 impairs strand transfer for both substrates, but it has a particularly strong effect on the 5′-phosphate stimulation. Whereas for the control Tnp the presence of the 5′ phosphate on the NTS stimulates the rate of strand transfer ~30-fold (Figure 2B), for the TnpR210A derivative the stimulation is only ~2-fold (Figure 5). This strongly implies that the TnpR210–5′ phosphate coordination plays an important role in the observed effect of stimulation of strand transfer by the 5′ phosphate. Such a conclusion is also likely for residues Y319 and R322 but their mutant high levels of impairment for strand transfer in general precludes us from supporting this conclusion definitively.

**Table 1.** Summary of crystallographic statistics

| | |
|---|---|
| Diffraction data | |
| Space group | $P6_522$ |
| Unit cell (Å) | $a$ = 113.5, $c$ = 228.8 |
| Resolution (Å) | 30–2.5 |
| Reflections, total/unique | 631591/30992 |
| Average $I/\sigma$[a] | 69.5 (9.0) |
| Completeness[a] (%) | 100 (100) |
| $R_{merge}$[a] (%) | 4.4 (37.9) |
| Refinement and model statistics | |
| Number of atoms, protein/DNA/solvent | 3616/818/138 |
| $R_{work}/R_{free}$[a] (%) | 21.9/27.4 (34.2/38.5) |
| Average B factor (Å$^2$) | 61.8 |
| Ramachandran plot, favored/allowed (%) | 91.2/8.8 |
| RMSD bonds/RMSD angles | 0.011 Å/1.53° |

[a]Data in parentheses represent highest resolution shell.

### Model for 5′ phosphate stimulation of target capture/strand transfer in Tn*5* transposition

How can we understand the stimulation of strand transfer by NTS's 5′ phosphate? We posit that the NTS 5′ phosphate acts as a 'handle' through which the protein

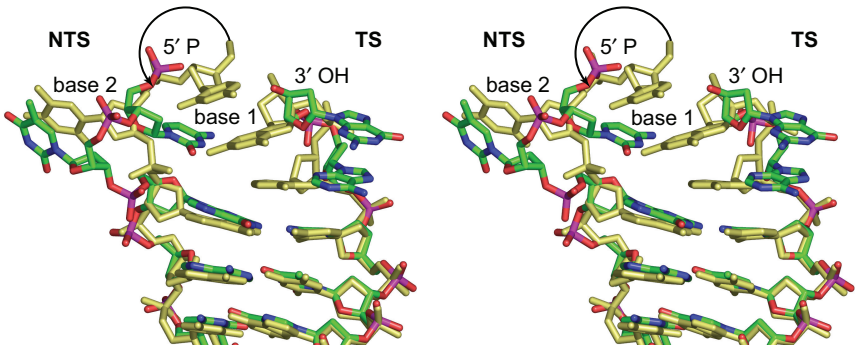


**Figure 3.** Effect of 5′ phosphate on DNA structure in the Tnp paired end complexes. The DNA structure for the paired end complex containing phosphorylated NTS reported here (colored by atom type in blue, cyan, red and orange) is compared to the DNA structure reported previously for the paired end complex lacking 5′ phosphate (PDB code 1MUH, in yellow) in a stereo presentation. The arrow shows the change of the terminal oxygen position between the two structures. The result of this movement is a near perfect overlap between the 5′-terminal phosphate and a phosphate bridging bases 1 and 2 in the nonphysiological NTS 5′ OH structure. Note the corresponding movement of the base 1 of the NTS and a flipped over base 1 of the TS. The combined effect is to provide more room in the vicinity of the catalytically active 3′ OH of the TS and that this increased room facilitates the binding of target DNA in preparation for strand transfer.

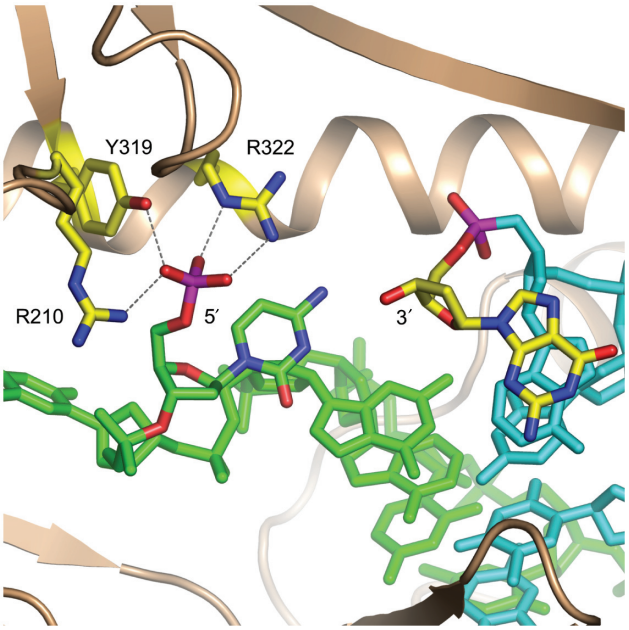


**Figure 4.** 5′ Phosphate coordination by Tnp. Coordination of the NTS 5′ phosphate by Tnp amino acid residues R210, Y319 and R322 is illustrated. NTS is shown in green and TS in cyan. In both cases noncarbon atoms of base 1 of both strands are colored according to atom type. For clarity, DDE catalytic core is not shown. Catalytic E326 is located one helical turn away from R322.

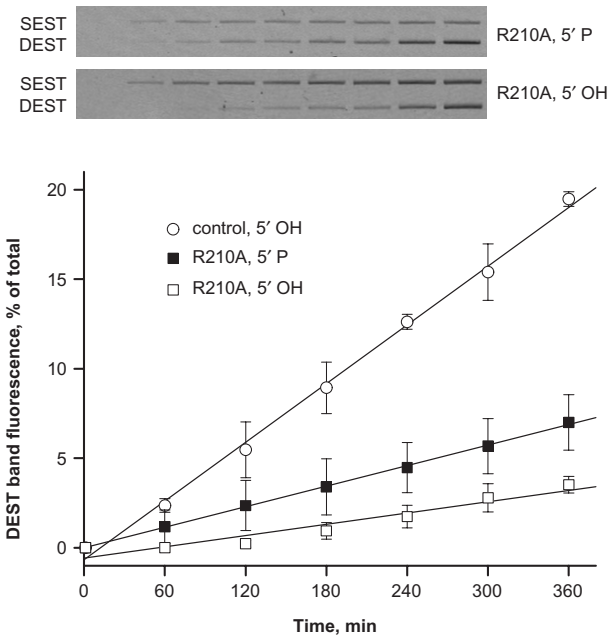


**Figure 5.** Strand transfer by Tnp R210A Mutant. The Tnp R210A catalyzed strand transfer activity was determined exactly as outlined in Figure 2. Shown are representative gels containing SEST and double end strand transfer (DEST) reaction products and a graph that presents quantitative analysis of the kinetic data derived from the gels. The time course over a period of 6 h is plotted for the mutant Tnp in complex with 5′ hydroxyl (open squares) or 5′ phosphate (closed squares) end sequence DNA substrates. Data for the control Tnp in complex with 5′ hydroxyl end sequence DNA substrate (open circles) is plotted for comparison. Error bars represent standard errors of the triplicate data points for R210A mutant and of the duplicates for the control. Solid lines represent linear regressions through all data presented. For Tnp R210A, the difference in the calculated reaction rates between NTS 5′ P and 5′ OH is 1.8-fold. Thus, Tnp R210A displays impaired strand transfer activity in comparison to the control, but the stimulatory effect of the 5′ phosphate on strand transfer activity with this mutant is only ~2-fold.

facilitates the movement of the NTS away from a position that would otherwise sterically impair access of the target DNA to the Tnp active site. In addition, we notice that this positive effect of 5′-end movement is augmented by the swiveling of the 3′-terminal base away from the interior of the helix. The swiveling of the 3′-terminal base appears to be the consequence of it losing contacts (for instance, stacking against the 5′-terminal base) due to the movement of the 5′-end. These DNA rearrangements are shown in Figure 3. It is interesting to note that, despite dramatic rearrangement of the base 1 conformation, base 2 of the NTS moves very little, maintaining its stacking interaction with W298 and a hydrogen bond with the hydroxyl of Y237. Thus, continuing an analogy of base 1 being a 'handle', one could view base 2 as a 'hinge' (together with the rest of the Tnp–DNA contacts involving nucleotides at positions 3 and beyond being invariant and responsible for the overall stability of protein–DNA complex). The dramatic consequences of these DNA rearrangements are illustrated in Figure 6. It can be seen that the presence of the 5′ phosphate results in the increased surface exposure of the critical transferred strand 3′ hydroxyl and the appearance of a cavity around the catalytic site where the target DNA would have to bind in the subsequent step of DNA transposition. This provides a clear structural explanation for the observed stimulation of strand transfer by an NTS 5′ phosphate.

Previously, the paired end complex structure was determined for a complex formed with hairpin end sequence DNA, into which the $Mn^{2+}$ was soaked in, resulting in cleavage of the hairpin (PDB code 1MUS). In this structure, the NTS 5′ phosphate is not coordinated by R210, Y319 and R322. Instead, just like in the original paired end complex structure formed with nonphosphorylated oligonucleotides (14), these residues coordinate the phosphate that bridges bases 1 and 2 of the NTS. In other words, despite the presence of the terminal phosphate, the DNA tip is found in the inhibitory conformation proposed above. There are two mutually nonexclusive explanations for this. First, it is possible that the appropriate conformational change following hairpin cleavage is constrained within a crystal lattice. Second, it should be noted that the 5′ phosphate in the resolved hairpin structure interacts with an $Mn^{2+}$ that is bound to the Tnp catalytic core. This is expected to provide an additional stabilization of the inhibitory position of the 5′ phosphate in the resolved hairpin structure. However, in the structures formed with 'pre-cleaved' phosphorylated DNA [this communication and the 'strand transfer-like' DNA substrate of Lovell *et al.* (15)], the 5′-end is relocated as shown in Figure 3, presumably representing the preferred conformation. It is logical to think that these two Tnp–DNA arrangements reflect the sequence of events that take place following hairpin cleavage. Immediately following

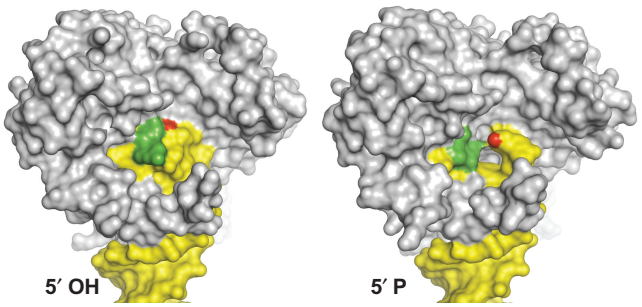


**Figure 6.** Effect of 5′ phosphate on accessibility of the catalytic site in the Tn5 paired end complex. The structure of the paired end complex containing phosphorylated NTS (right) is compared to the previously reported structure lacking the 5′ phosphate (PDB code 1MUH). Shown are molecular surface representations with protein colored gray and DNA colored yellow. The first base of the nontransferred DNA strand is colored green and the 3′ catalytic hydroxyl of the transferred strand is colored red. Note how in the presence of the 5′ phosphate the movement of the DNA illustrated in Figure 3 creates a cavity in the active site exposing the 3′ OH. Presumably the DNA movement eliminates steric hindrances for the target DNA binding in the subsequent step of DNA transposition.

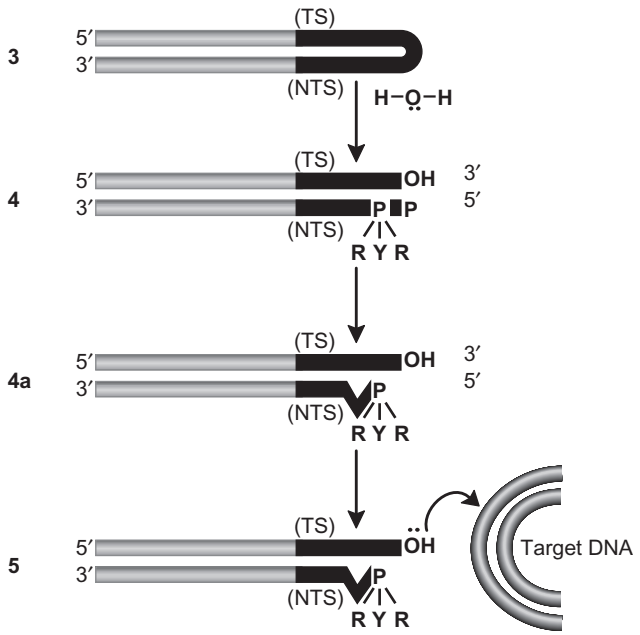


**Figure 7.** A new step in Tn5 transposition—rearrangement of cleaved DNA prior to target capture. A portion of the schematic presented in Figure 1 is modified to accommodate the results from the current studies. Step (**4**) has been modified to show the R210, Y319, R322 (RYR) coordination to the phosphate bridging bases 1 and 2 of the NTS immediately following hairpin cleavage. (**4a**) The NTS 5′-end is then rearranged through the formation of RYR coordination to the 5′-terminal phosphate and the NTS 5′-end is displaced from the 3′-end thus facilitating formation of the target capture complex [step (**5**)].

cleavage the NTS 5′ is in the target capture inhibitory conformation. During the next step, a DNA rearrangement occurs that involves a switch of the phosphate that is coordinated by R210, Y319 and R322, generating the open, target capture-competent structure. This target capture-competent complex represents an additional DNA structure (4a) that is added to those pictured schematically in Figure 1 (see Figure 7).

It seems likely that in solution there is a dynamic equilibrium between R210–Y319–R322 triad coordinating alternatively the NTS terminal and subterminal phosphates, resulting in the target capture-competent and inhibitory conformations, respectively. In the absence of divalent metal ions, the equilibrium would be shifted toward the more 'open', target capture-competent conformation. This is consistent with our observation that the presence of metal ions led to the lower resolution limit of X-ray diffraction, which is sensitive to the conformational heterogeneity within the crystal.

An additional observation that lends support to the model of the essential DNA coordination and movement outlined above and schematically in Figure 7 is the fact that the Tnp R210A mutation has been found to enhance target DNA sequence specificity (21). Indeed, within the framework of this model, if the R210A mutant is impaired in reorientation of the NTS 5′ phosphate, the random target DNA is expected to be unable to effectively compete with the NTS for occupation of the active site because the Tnp interactions with target DNA are weak. However, when a target DNA with a sequence that has been selected to be a preferential target (28) is used, the intrinsic properties of this DNA segment (i.e. stronger interaction with the protein) partially compensate for the mutant's inability to effectively create more room for the target DNA binding. The end result is the overall lower efficiency but enhanced target specificity for strand transfer—precisely as observed for TnpR210A.

## Importance and implications for other transposases and transposase-like systems

We have demonstrated a previously unrecognized role of the 5′ NTS terminal phosphate in the target capture/strand transfer process. The immediate implication of this finding involves experimental design. Our observations highlight the importance of always using phosphorylated oligonucleotides in transposition experiments. The absence of the 5′ NTS phosphate strongly affects both the kinetics and yield of the target capture/strand transfer process. Thus mechanistic and structural studies using nonphosphorylated DNA could yield results that would be different from the physiological situation where the NTS always contains a 5′ phosphate.

Because there is a considerable degree of sequence conservation among members of the IS*4* Tnp family, it is highly likely that the NTS 5′ phosphate plays a similar stimulatory role in target capture for other IS*4* Tnps and that residues equivalent to the R210, Y319 and R322 are intimately involved in the required DNA movement in these cases. While the importance of Y319 and R322 (or their equivalents for Tn*10* Tnp) has been well documented for other transposition steps such as hairpin formation (29) and hairpin resolution (17), their role in strand transfer and the newly uncovered functional role of the conserved R210 is novel. Since DNA–protein complexes for other IS*4* Tnps have so far proven to be recalcitrant to crystallization, a search for functional equivalent(s) of R210 certainly warrants attention.

It is also quite probable that the mechanism proposed here for Tn*5* target capture is employed by distantly related systems. For instance, similar types of protein–DNA contacts have been proposed to be important for hairpin formation by RAG-1 (30) and for the *Borrelia burgdorferi* telomere resolvase (31). In HIV-1 integration, the 5′-end of the NTS is extended by 2 nt (32). It is possible that these two 5′ nucleotides provide a molecular 'handle' which the integrase uses to enable efficient target binding. Thus, we propose that the important role of the RYR triad in coordinating the NTS in Tn*5* target capture is a specific example of a general phenomenon that will be discovered for other proteins catalyzing complex DNA recombination events.

## SUPPLEMENTARY DATA

Supplementary Data are available at NAR online.

## ACCESSION CODE

Atomic coordinates and structure factors have been deposited in the Protein Data Bank with the accession code 3ECP.

## FUNDING

National Institutes of Health (AR35186 to I.R., GM50692 to W.S.R.); Evelyn Mercer Professorship in Biochemistry and Molecular Biology (W.S.R.).

*Conflict of interest statement*. None declared.

## REFERENCES

1. Craig,N.L., Craigie,R., Gellert,M. and Lambowitz,A.M. (eds), (2002) *Mobile DNA*. ASM Press, Washington, DC.
2. 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.
3. 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.
4. Davies,D.R., Braam,L.M., Reznikoff,W.S. and Rayment,I. (1999) The three-dimensional structure of a Tn*5* transposase-related protein determined to 2.9 Å resolution. *J. Biol. Chem.*, **274**, 11904–11913.
5. Gellert,M. (2002) V(D)J recombination. In Craig,N.L., Craigie,R., Gellert,M. and Lambowitz,A.M. (eds), *Mobile DNA*. ASM Press, Washington, DC, pp. 705–729.
6. Czyz,A., Stillmock,K.A., Hazuda,D.J. and Reznikoff,W.S. (2007) Dissecting Tn*5* transposition using HIV-1 integrase diketoacid inhibitors. *Biochemistry*, **46**, 10776–10789.
7. Steiniger,M., Metzler,J. and Reznikoff,W.S. (2006) Mutation of the Tn*5* transposase β-loop residues affects all steps of Tn*5* transposition: the role of conformational changes in Tn*5* transposition. *Biochemistry*, **45**, 15552–15562.
8. Rezsohazy,R., Hallet,B., Delcour,J. and Mahillon,J. (1993) The IS*4* family of insertion sequences: evidence for a conserved transposase motif. *Mol. Microbiol.*, **9**, 1283–1295.
9. Chandler,M. and Mahillon,J. (2002) Insertion sequences revisited. In Craig,N.L., Craigie,R., Gellert,M. and Lambowitz,A.M. (eds), *Mobile DNA*. ASM Press, Washington, DC, pp. 305–366.
10. De Palmenaer,D., Siguier,P. and Mahillon,J. (2007) IS*4* family goes genomic. *BMC Evol. Biol.*, **8**, 18.
11. Haniford,D.B. (2002) Transposon Tn*10*. In Craig,N.L., Craigie,R., Gellert,M. and Lambowitz,A.M. (eds), *Mobile DNA*. ASM Press, Washington, DC, pp. 457–483.
12. Reznikoff,WS (2002) Tn*5* transposition. In Craig,N.L., Craigie,R., Gellert,M. and Lambowitz,A.M. (eds), *Mobile DNA*. ASM Press, Washington, DC, pp. 403–421.
13. Reznikoff,W.S. (2003) Tn*5* as a model for understanding DNA transposition. *Mol. Microbiol.*, **47**, 1199–1206.
14. Davies,D.R., Goryshin,I.Y., Reznikoff,W.S. and Rayment,I. (2000) Three-dimensional structure of the Tn*5* synaptic complex transposition intermediate. *Science*, **289**, 77–85.
15. Lovell,S., Goryshin,I.Y., Reznikoff,W.S. and Rayment,I. (2002) Two metal active site binding of a Tn*5* transposase synaptic complex. *Nat. Struct. Biol.*, **9**, 278–281.
16. Jenkins,T.M., Esposito,D., Engelman,A. and Craigie,R. (1997) Critical contacts between HIV-1 integrase and viral DNA identified by structure-based analysis and photo-crosslinking. *EMBO J.*, **16**, 6849–6859.
17. Naumann,T.A. and Reznikoff,W.S. (2002) Tn*5* transposase active site mutants. *J. Biol. Chem.*, **277**, 17623–17629.
18. Reznikoff,W.S., Bordenstein,S.R. and Apodaca,J. (2004) Comparative sequence analysis of IS*50*/Tn*5* transposase. *J. Bacteriol.*, **186**, 8240–8247.
19. Bhasin,A., Goryshin,I.Y. and Reznikoff,W.S. (1999) Hairpin formation in Tn*5* transposition. *J. Biol. Chem.*, **274**, 37021–37029.
20. Bradford,M.M. (1976) A rapid and sensitive method for the quantitation of microgram quantities of protein utilizing the principle of protein-dye binding. *Anal. Biochem.*, **72**, 248–254.
21. Gradman,R., Ptacin,J., Bhasin,A., Reznikoff,W.S. and Goryshin,I.Y. (2008) A Bifunctional DNA Binding Region in Tn*5* Transposase. *Mol. Microbiol.*, **67**, 528–540.
22. Rayment,I. (2002) Small-scale batch crystallization of proteins revisited: an underutilized way to grow large protein crystals. *Structure*, **10**, 147–151.
23. Otwinowski,Z. and Minor,W. (1997) Processing of X-ray diffraction data collected in oscillation mode. *Methods Enzymol.*, **276**, 307–326.
24. Vagin,A. and Teplyakov,A. (2000) An approach to multi-copy search in molecular replacement. *Acta Crystallogr. D*, **56**, 1622–1624.
25. Murshudov,G.N., Vagin,A.A. and Dodson,E.J. (1997) Refinement of macromolecular structures by the maximum-likelihood method. *Acta Crystallogr. D*, **53**, 240–255.
26. Winn,M.D., Isupov,M.N. and Murshudov,G.N. (2001) Use of TLS parameters to model anisotropic displacements in macromolecular refinement. *Acta Crystallogr. D*, **57**, 122–133.
27. Perrakis,A., Harkiolaki,M., Wilson,K.S. and Lamzin,V.S. (2001) ARP/wARP and molecular replacement. *Acta Crystallogr. D*, **57**, 1445–1450.
28. Goryshin,I.Y., Miller,J.A., Kil,Y.V., Lanzov,V.A. and Reznikoff,W.S. (1998) Tn*5*/IS*50* target recognition. *Proc. Natl Acad. Sci. USA*, **95**, 10716–10721.
29. Allingham,J.S., Wardle,S.J. and Haniford,J.B. (2001) Determinants for hairpin formation in Tn*10* transposition. *EMBO J.*, **20**, 2931–2942.
30. Lu,C.P., Sandoval,H., Brandt,V.L., Rice,P.A. and Roth,D.B. (2006) Amino acid residues in Rag1 crucial for DNA hairpin formation. *Nat. Struct. Mol. Biol.*, **13**, 1010–1015.
31. Bankhead,T. and Chaconas,G. (2004) Mixing active-site components: a recipe for the unique enzymatic activity of a telomere resolvase. *Proc. Natl Acad. Sci. USA*, **101**, 13768–13773.
32. Engelman,A., Mizuuchi,K. and Craigie,R. (1991) HIV-1 DNA integration: mechanism of viral DNA cleavage and DNA strand transfer. *Cell*, **67**, 1211–1221.