doi:10.1006/jmbi.2001.4553 available online at http://www.idealibrary.com on IDEAL® *J. Mol. Biol.* (2001) **307**, 1161–1170

JMB

COMMUNICATION

Selective Recognition of Pyrimidine Motif Triplexes by a Protein Encoded by the Bacterial Transposon Tn*7*

Jason E. Rao and Nancy L. Craig*

Howard Hughes Medical Institute, Department of Molecular Biology & Genetics Johns Hopkins University School of Medicine, Baltimore MD 21205, USA

**Corresponding author*

The bacterial transposon Tn*7* is distinguished among mobile genetic elements by its targeting abilities. Recently, we reported that Tn*7* is able to selectively insert adjacent to triple-helical DNA. The binding of TnsC, a Tn*7*-encoded protein, to the triplex DNA target leads to the specific transposition of Tn*7* adjacent to both inter- and intramolecular pyrimidine motif triplexes. Here, we further probe how Tn*7* targets triplex DNA. We report that TnsC discriminates between different types of triplexes, showing binding preference for pyrimidine but not for purine motif intermolecular triplex DNA. The binding preferences of TnsC and the Tn*7* insertion profiles were obtained using psoralenated, triplex-forming oligonucleotides annealed to plasmid DNAs. Although the presence of psoralen is not required for targeting nor is it alone able to attract TnsC, we show that the location of psoralen within the pyrimidine motif triplex does alter the position of Tn*7* insertion relative to the triplex. Comparison between the triplex-targeting pathway and the highly site-specific targeting pathway mediated by the binding of the Tn*7*-encoded protein, TnsD, to the unique site *attTn7*, suggests that similar structural features within each target DNA are recognized by TnsC, leading to site-specific transposition. This work demonstrates that a prokaryotic protein involved in the targeting and regulation of Tn*7* translocation, TnsC, can selectively recognize pyrimidine motif triplexes.

Keywords: triple helical DNA; target site selection; transposition; triplex-binding protein; psoralen

Unlike most mobile genetic elements, which show relatively little preference in their location of insertion, Tn*7* exercises exceptional control over its selection of a target site. Five proteins encoded by this element enable Tn*7* to evaluate and respond to various targeting signals, giving rise to two separate but overlapping targeting pathways (Craig, 1996). TnsA, TnsB, TnsC and TnsD (TnsABC + D) promote transposition into a unique site in the chromosome of many bacteria, called *attTn7* (Craig, 1991), while TnsA, TnsB, TnsC and TnsE (TnsABC + E) promote transposition into non-*attTn7* sites on conjugating plasmids (Wolkow *et al.*, 1996) and into particular regions of the *Escherichia coli* chromosome (Peters & Craig, 2000). An additional level of control comes from the fact that Tn*7* insertion into DNAs that already contain Tn*7* is greatly decreased; such targets are said to be ''immune'' to further Tn*7* transposition events (Hauer & Shapiro, 1984; Arciszewska *et al.*, 1989; Stellwagen & Craig, 1997a). Thus, Tn*7* can sense and respond to both positive and negative signals within a potential target DNA.

The site-specific TnsABC + D pathway has been reconstituted *in vitro*, allowing the roles of each of the Tns proteins to be identified (Bainton *et al.*, 1991, 1993). TnsA and TnsB collaborate to form the transposase, carrying out all of the breakage and joining activities for transposition (May & Craig, 1996; Sarnovsky *et al.*, 1996). TnsD is an *attTn7*-specific DNA-binding protein (Bainton *et al.*, 1993). TnsC is a non-sequence specific DNA-binding protein that utilizes ATP to regulate transposition (Gamas & Craig, 1992; Stellwagen & Craig, 1998): TnsC mediates the interaction between the *attTn7*-TnsD target complex and the transposase, leading to the formation of an active transpososome.

Abbreviations used: TFO, triplex-forming oligonucleotide.

E-mail address of the corresponding author: ncraig@jhmi.edu

0022-2836/01/051161–10 $35.00/0

Assembly of the Tns proteins with both the transposon and *attTn7* target DNA results in site-specific Tn7 transposition into *attTn7* (Bainton *et al*., 1993). In addition to site-specific insertion, the architecture of the transpososome complex ensures that the structurally asymmetric Tn7 element inserts in a specific orientation into *attTn7* (Arciszewska *et al*., 1989).

TnsC plays a central role in the ability of Tn7 to preferentially insert into particular targets while purposely avoiding others. TnsC regulates transposition by responding to various targeting signals and subsequently activating the transposase. The binding of TnsC to the *attTn7*-TnsD complex leads to insertion of Tn7 into *attTn7*, while removal of TnsC from immune targets through interactions with TnsB prevents insertions from occurring (Bainton *et al*., 1993; Stellwagen & Craig, 1997a). Thus, the presence or absence of TnsC is critical to both the targeting and ultimately the regulation of Tn7 transposition.

In the absence of the targeting proteins, TnsD or TnsE, the core transposition machinery TnsABC^{wt} promotes a very low level of non-sequence specific Tn7 insertions (Biery *et al*., 2000). Using TnsABC^{wt}, we found that Tn7 can, however, selectively target triplex DNA (Rao *et al*., 2000). Triplex-directed targeting results from preferential binding of TnsC to the triplex DNA, suggesting that triple-helical DNA contains a positive targeting signal for Tn7 transposition (Rao *et al*., 2000). A gain-of-function mutant of TnsC^{wt}, TnsC^{A225V} (Stellwagen & Craig, 1997b), which provides a much higher frequency of transposition for TnsABC^{A225V} (Stellwagen & Craig, 1997b; Stellwagen, 2001), also promotes insertion adjacent to triple-helical DNA structures (Rao *et al*., 2000).

A direct comparison of Tn7 transposition adjacent to either a pyrimidine motif triplex DNA or the *attTn7*-TnsD complex *in vitro*, revealed strikingly similar patterns of insertion in terms of both the location and orientation of Tn7 insertion into each respective target DNA (Rao *et al*., 2000; Kuduvalli *et al*., 2001). These similarities suggest that triplex DNA and the *attTn7*-TnsD complex contain similar signals and are thus being recognized by TnsC in a similar fashion. However, many questions regarding the unexpected recognition of triplex DNA remained. Are all triplexes targeted by Tn7? What specific structural features of triplex DNA influence recognition by TnsC? How does recognition of triplex DNA by TnsC compare with other triplex-binding proteins?

In this work, we have further probed the mechanism by which Tn7 recognizes triplex DNA. We characterize the triplex-binding abilities of TnsC, and compare it to the five previously reported triplex-binding proteins (Kiyama & Camerini-Otero, 1991; Guieysse *et al*., 1997; Jiménez-García *et al*., 1998; Musso *et al*., 1998; Nelson *et al*., 2000), all of which are found in eukaryotes (see Table 1). The affinity of these proteins for triplex DNA varies between purine and pyrimidine motifs, with a strong preference for one motif over the other. We demonstrate here that TnsC is able to selectively bind pyrimidine, but not purine motif triplexes, leading to pyrimidine triplex-specific transposition. The triplex targets used in our assays were constructed using triplex-forming oligonucleotides (TFOs) conjugated to a psoralen intercalator to increase triplex stability under *in vitro* transposition reaction conditions. Interestingly, we find that altering the position of psoralen within the pyrimidine motif triplex targets, in turn alters the insertion profiles generated by Tn7. This work both provides a deeper understanding of the targeting capabilities of Tn7 and illustrates the critical role of TnsC in the regulation of transposition by sensing and responding to specific signals presented by the target DNA.

Construction of purine and pyrimidine motif triplex targets

A variety of plasmid DNA targets with short regions (15 bp) of triplex DNA were generated by using TFOs synthesized with psoralen on either the 5′or 3′ end, and UV-induced crosslinking of the TFO to the plasmid (Figure 1). The preferred sequence of psoralen intercalation into duplex DNA, TA, was cloned into both the 3′ and 5′ ends of the triplex targets to allow interstrand crosslinking to these substrates, and increase triplex stability. Triplex DNA can be formed through the pairing of single-stranded DNA to the major groove of duplex DNA through either Hoogsteen or reversed-Hoogsteen hydrogen bonding, giving rise to the pyrimidine and purine motifs (Soyfer & Potaman, 1996). We constructed both motifs for our *in vitro* transposition targeting assays. The polypurine tract within the pDUP plasmid is identical for each of the targets, thus allowing direct comparison of the various triplex constructs within our transposition assay. The pH-dependence of pyrimidine motif triplex formation, which arises from the requirement of protonation of cytosine for Hoogsteen hydrogen bonding, is partially negated by the use of nucleoside analogs, such as 5-methylcytosine (Lee *et al*., 1984; Povsic & Dervan, 1989) and 8-oxo-adenine (Miller *et al*., 1996). The use of psoralen and of modified basses within our TFOs increased the stability of our triplex targets. While these changes do not alter the polarity of triplex formation (Figure 1), they do stabize triplexes under *in vitro* transposition reaction conditions.

Tn*7* targets pyrimidine but not purine motif triplexes

The targeting preferences of Tn7 for various triplex structures were obtained using TnsABC *in vitro* transposition reactions. For ease of visualization and quantification of reaction products, the gain-of-function allele of TnsC^{wt}, TnsC^{A225V}, was primarily used with TnsAB, though similar targeting preferences are found for TnsC^{wt}. To

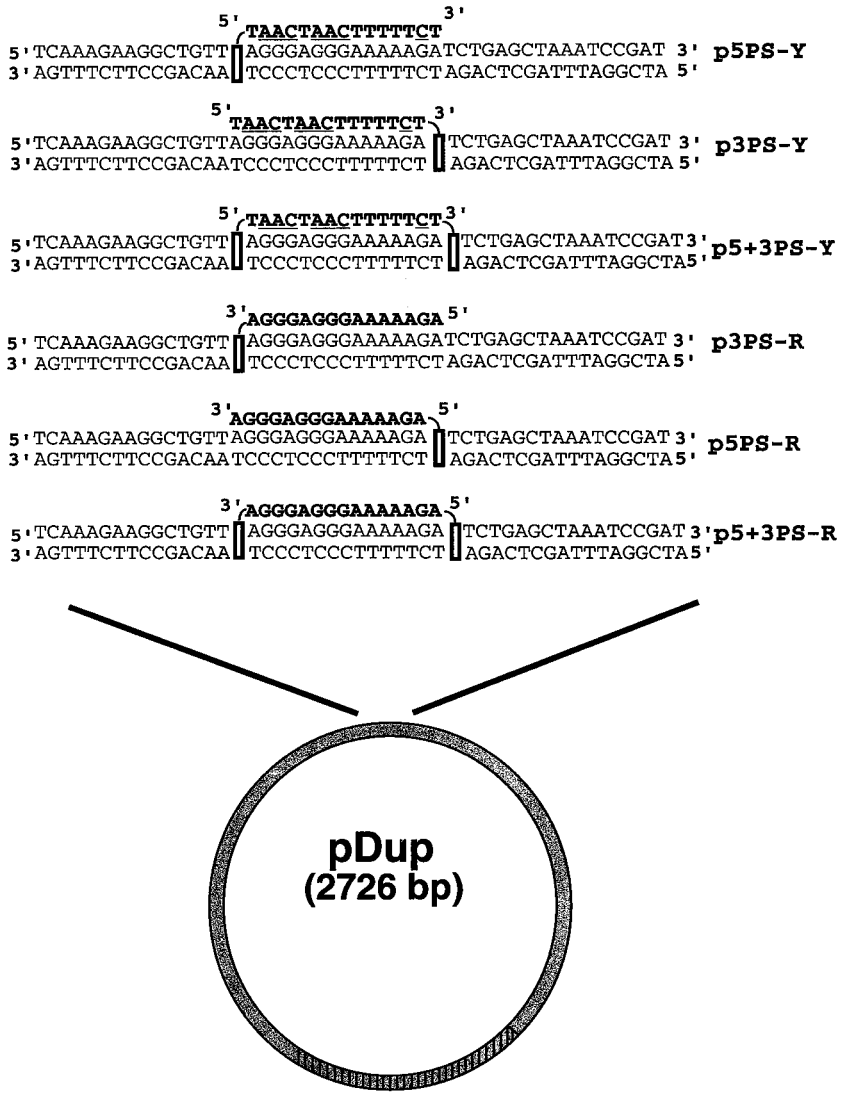
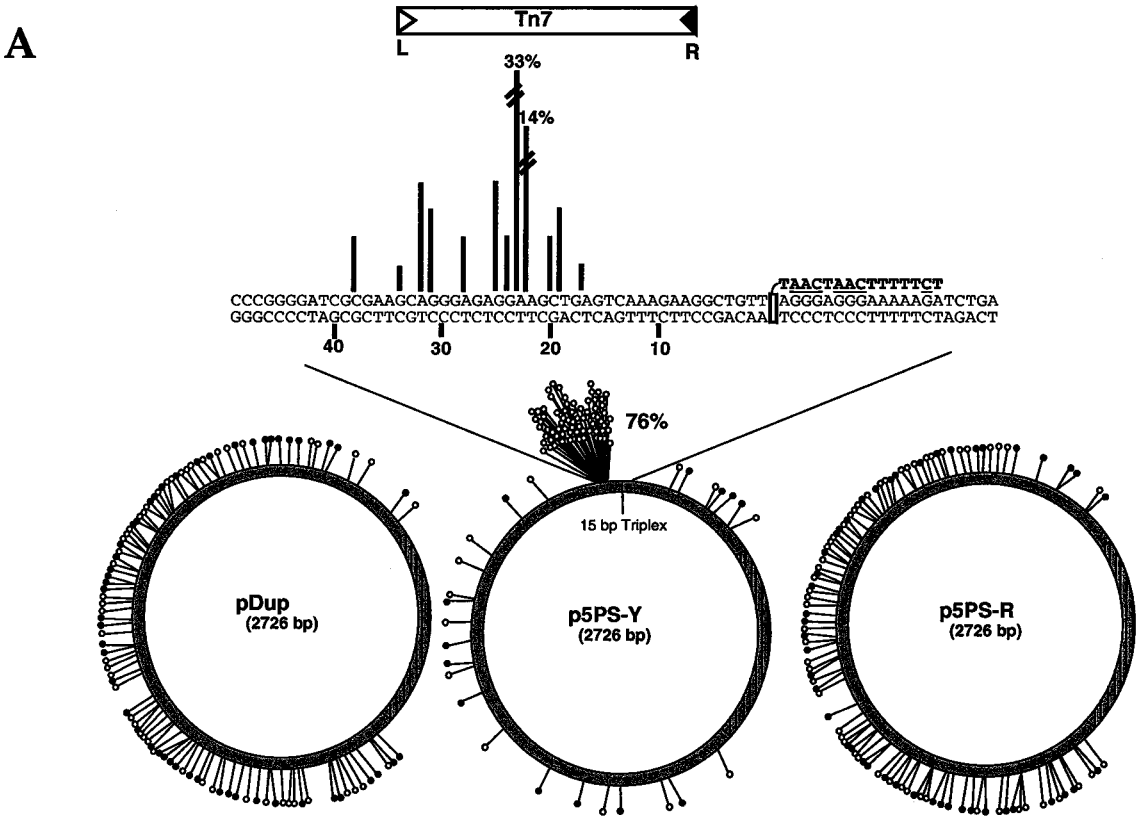


Figure 1. Sequences of triplex DNA constructs used in this study. The duplex plasmid used to generate all of the triplex targets, pDup, contains a 15 bp polypurine tract to allow triplex formation by either the pyrimidine (Y) or purine (R) TFO. The intercalation of psoralen into the triplex-duplex junction is illustrated by a black rectangle. To aid in the stabilization of triplex substrates under *in vitro* transposition reaction conditions, pyrimidine-based TFOs were synthesized with 5-methyl-cytosine and 8-oxo-adenosine (overlined) to negate the pH-dependence on triplex formation. Though 8-oxo adenosine is a purine-based nucleoside analog, it forms a pyrimidine motif triplex through pH-independent Hoogsteen hydrogen bond formation to the GC base-pairs of the Watson-Crick duplex. TFOs were made by standard phoshoramidite synthesis, or purchased from Trilink Biotechnologies. Annealing and crosslinking of triplexes was done at 0 °C and pH 7.0 in triplex-annealing buffer, which contained a final concentration of 50 mM Mops, 20 mM $MgCl_2$ and 10 mM NaCl. Crosslinking was carried out at 365 nm (UV light) for 15 minutes at 0 °C. For all triplex sequences shown, >90 % di-adduct between plasmid and TFO was confirmed by denaturing gel electrophoresis. pDup contains a polypurine tract cloned into the plasmid pUC-19. Triplex targets were generated by pairing TFOs to pDup, allowing Hoogsteen hydrogen bond formation. Stable triplex formation was confirmed through footprinting and EMSA assays (data not shown), before being used as targets in the *in vitro* transposition reaction.

obtain a high-resolution view of the positions of Tn*7* insertion within the various targets, products of the *in vitro* transposition reactions were isolated individually by transformation into *E. coli*, and sequenced. The insertion profile of 100 sequenced insertions into the parental duplex plasmid, pDup, was compared to the insertion profile for p5PS-Y and p5PS-R (Figure 2(a)). The p5PS-Y target plasmid containing a pyrimidine motif triplex generated both site-specific and orientation-specific insertions of Tn*7*, with a hotspot located 23 bp from the 5′ triplex-duplex junction. This insertion profile is in good accordance with the results of previous studies using similar pyrimidine triplex targets (Rao *et al.*, 2000). By contrast, the profile obtained for the purine motif triplex target, p5PS-R, reveals a ''random'' distribution of Tn*7* insertions, giving a profile similar to that of pDup (Figure 2(a)). In comparison to the profiles generated with the pyrimidine motif triplex targets, the purine triplexes showed neither specific localization nor orientation bias of Tn*7* insertion. Furthermore, we found no instance of a double-insertion event, where Tn*7* inserted into the same triplex target more than once. This confirmed that target site immunity is maintained for triplex targets.

TnsC binds pyrimidine but not purine motif triplexes

TnsC preferentially binds pyrimidine motif triplexes, leading to recruitment of transposase and subsequent triplex-specific insertions of Tn*7* (Rao *et al.*, 2000). We tested if TnsC was able to bind



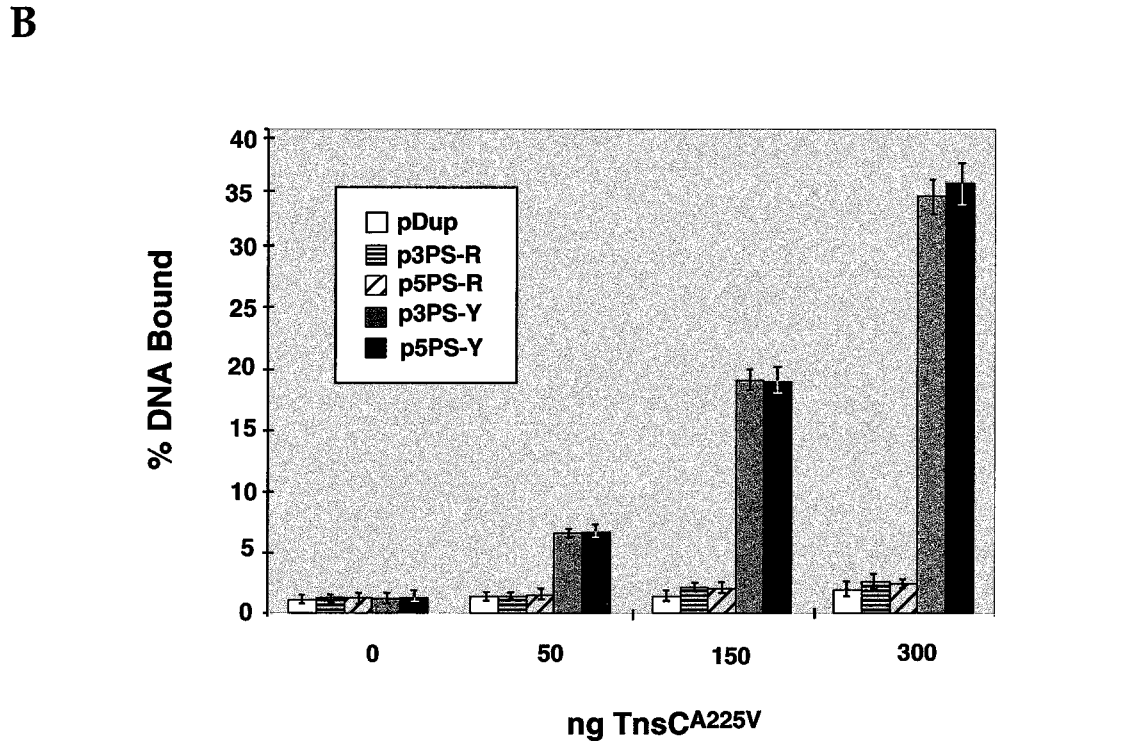


Figure 2 (*legend opposite*)

purine motif triplexes, even though Tn7 does not preferentially transpose adjacent to purine triplexes. The affinity of TnsC for pyrimidine and purine motif triplexes was determined by binding assays (Figure 2(b)) that used restriction fragments containing the triplex region from each of the triplex substrates (Figure 1). As the concentration of TnsC increased, the binding of TnsC to only the pyrimidine motif triplexes, p5PS-Y and p3PS-Y, increased over that of the duplex control (Figure 2(b)). TnsC was therefore able to distinguish between purine and pyrimidine triplex targets, by binding preferentially to only the pyrimidine substrates.

The influence of psoralen on the point of insertion

Triplex targets can be constructed using TFOs that contain psoralen conjugated to either the 5′ or the 3′ end. Though previous work using intramolecular triplex targets confirmed that psoralen is not required for triplex targeting, and that targets containing psoralen alone are incapable of invoking specific targeting by Tn*7* (Rao *et al.*, 2000), the possibility remained that psoralen could affect the position of Tn7 insertion. This was tested by a comparison of the insertion profiles of p5PS-Y and p3PS-Y triplexes, which contain psoralen on opposite ends of the TFO (Figure 1). While Southern blot analysis demonstrates that both give rise to site-specific transposition adjacent to the triplex region (data not shown), targeting differences were found after sequencing a population of insertions into each target (Figure 3). Though insertion by Tn7 occurred into each target with similar specificity (~75 % within a 40 bp region of the triplex) and orientation bias (98 % had the right end of Tn7 closest to the triplex), the exact position of Tn7 insertion with regard to the triplex varied (Figure 3). The hot-spot of insertion for the p5PS-Y triplex was found 23 bp from the 5′ triplex-duplex junction, whereas the hot-spot for the p3PS-Y triplex was only 10 bp from this junction (Figure 3). However, in both cases, the hot-spot was a distance of ~25 bp with regard to the site of psoralen intercalation for both triplexes. Thus, we see a strong correlation between the precise position of insertion and the location of the psoralen within the pyrimidine motif triplex.

TnsC recognizes triplexes crosslinked at both ends of the TFO

As the TFOs were covalently attached to only one end of the duplex-binding site, our triplex substrates were theoretically capable of fluctuating between an annealed and single-stranded state, or ''breathing''. Thus, one possible explanation for binding and transposition adjacent to triplex DNA was that TnsC was recognizing a single-stranded portion of the triplex target. To probe this model directly, TFOs containing psoralen on both the 5′ and 3′ ends (p5 + 3PS-Y, p5 + 3PS-R) (Figure 1) were tested for both targeting by Tn7 (Figure 4(a)) and binding by TnsC (Figure 4(b)). These constructs eliminate any potential ''breathing'' by the TFOs at the triplex-duplex junction by covalent

Figure 2. Comparison of the Tn*7* insertion profiles into targets containing purine and pyrimidine motif triplexes. Tn7 preferentially targets pyrimidine, but not purine, motif triplexes. (a) The distribution of 100 Tn7 insertions into pDup, p5PS-Y and p5PS-R are shown. All three plasmids are identical in size and sequence. The distribution patterns of Tn7 into these plasmid targets is illustrated: each line represents a single insertion, an open circle corresponds to a left to right orientation for Tn7, while a filled circle is a right to left orientation. The origin of replication for the target plasmid is drawn in broken lines. For the p5PS-Y target, the nucleotide sequence surrounding the triplex is shown; the number of insertions at each nucleotide is proportional to the length of the lines drawn: 76 % of all Tn7 insertions are found within a 40 bp region of the TFO. For these insertions, nearly all (>98 %) occurred such that the right end of Tn7 element lands nearest the 5′ end of the triplex. A Tn7 element is drawn above the regional hot-spot to illustrate this orientation bias. These data were obtained from the sequencing of *in vitro* transposition reaction products, which were ethanol-precipitated and digested with *Eco*RV to linearize donor plasmid before transformation into *E. coli*. Cultures were grown under appropriate antibiotic resistance, plated, and colonies were individually isolated to obtain a single target plasmid containing a unique mini-Tn7 insertion. (b) Tn7 preferentially binds pyrimidine but not purine motif triplexes. The graph shows the relative affinities of $TnsC^{A225V}$ for duplex and various triplex DNA targets. The percentage bound reflects the fraction of the labeled substrate bound by $TnsC^{A225V}$. Each of the binding reaction mixtures (25 μl final volume) contained 25.9 mM Hepes (pH 7.6), 2.5 mM Tris (pH 7.6), 60 μg/ml BSA, 2.0 mM ATP, 2.0 mM DTT, 5.0 mM KCl, 14.3 mM NaCl, 16 μM EDTA, 120 μM $MgCl_2$, 120 μM Chaps, 4 μM PMSF, 0.4 % glycerol, an amount of TnsC varied as shown in the Figures, 1.8 μg (~1.0 pmol of plasmid, or 2.8 nmol of base-pairs) of unlabeled competitor plasmid DNA, and ~0.01 pmol (fragment) of ^{32}P-labeled 110 bp target DNA fragment excised from each of the plasmid targets illustrated in Figure 1. For the duplex and triplex samples, the total concentration of DNA was identical and held constant. Also, both the amount (pmol) and the total counts/reaction (50,000 cpm) of labeled DNA were identical for duplex and triplex samples. Each reaction was incubated at 30 °C for 20 minutes before loading onto a nitrocellulose filter pre-equilibrated in a buffer identical with the binding buffer with 7 % (w/v) sucrose. Samples were washed four times each with 1 ml of binding buffer before being dried. Radioactivity from individual filters was quantified by scintillation counting, and is listed as the percentage DNA bound of the total radiolabeled probe (100 % = 50,000 cpm) by $TnsC^{A225V}$.

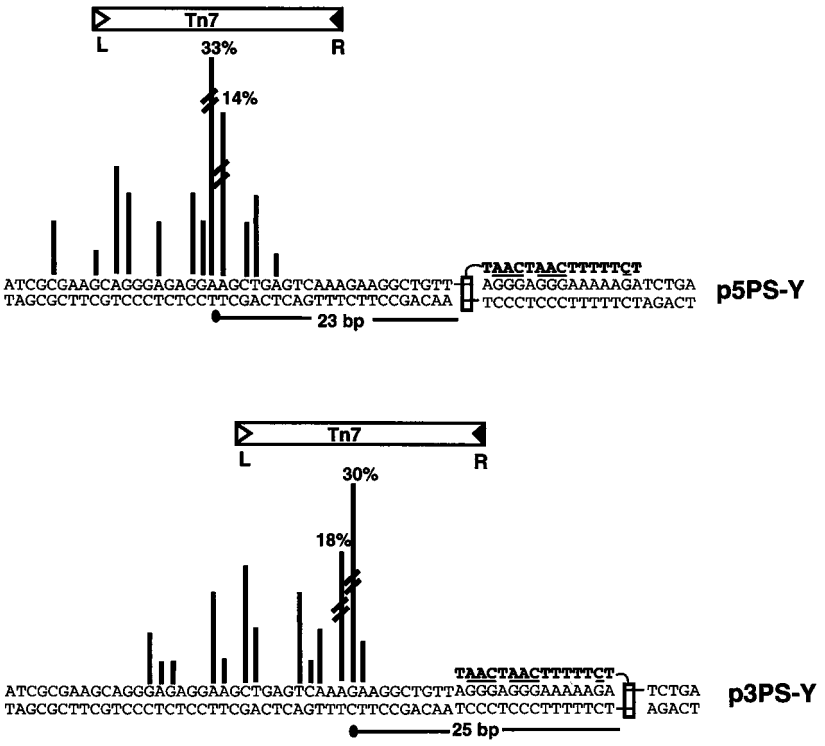


Figure 3. The location of psoralen within pyrimidine-motif triplexes affects the location of Tn7 insertion adjacent to triplex targets. Comparison of the Tn7 insertion profiles into p5PS-Y and p5PS-R triplex-containing targets. The distribution of insertions into the triplex region is illustrated. The number of insertions at each nucleotide is proportional to the length of the lines drawn, with the hot-spot of insertion labeled with a black dot below the sequence. For both targets, the hot-spot for Tn7 insertion is located 24(±1) bp from the site of psoralen intercalation. All of our triplexes were confirmned to be stably bound through electrophoretic mobility shift and footprinting assays (data not shown).

attachment of both ends of the TFO to the target plasmid through psoralen cross-links. Southern analysis illustrates Tn*7* inserted specifically into p5 + 3PS-Y, while the purine motif triplex, p5 + 3PS-R, showed no specific targeting (Figure 4(a)).

Though Southern blot analysis confirms that pyrimidine motif triplexes crosslinked at both ends of the TFO are targeted by Tn*7*, insertion profiles at the sequence level could not be determined for these targets, as insertion products were not recovered after transformation into *E. coli* (presumably due to an inability to effectively process the two sets of intrastrand crosslink adducts generated by the di-psoralenated triplex).

In terms of binding of TnsC, the affinity of TnsC for p5 + 3PS-Y was consistent with the affinities for TFOs with a single psoralen: TnsC preferentially bound p5 + 3PS-Y but not the p5 + 3PS-R triplex (Figure 4(b)). Tn*7* was able to bind and target triplexes crosslinked at both ends with the same efficiency as triplexes crosslinked at only one end, suggesting strongly that TnsC is not binding a single-stranded form of the triplex.

We have shown here that Tn*7* is able to distinguish between purine and pyrimidine motif triplexes as potential targets for site-specific transposition. While TnsC selectively binds pyrimidine motif triplexes, leading to triplex-specific transposition, purine motif triplexes are not bound by TnsC, and Tn*7* shows no preferential targeting to these structures. The presence of psoralen within a pyrimidine motif triplex is able to modify the insertion profile of Tn*7*.

Alterations in target DNA structure may provide the effector signal for Tn*7* transposition

How does TnsC distinguish between purine and pyrimidine motif triplexes? A better understanding of triplex recognition may come from the wild-type Tn*7* targeting pathway mediated by TnsD. TnsD is a sequence-specific, DNA-binding protein that binds a unique site in the bacterial chromosome called *attTn7*. Formation of the TnsD-*attTn7* target complex results in the site-specific and orientation-specific insertion of Tn*7* adjacent to the TnsD binding site within *attTn7*. Recent studies have revealed that the binding of TnsD to the major groove of *attTn7* results in asymmetric structural changes in and around its binding site, and that TnsC likely binds the altered conformation of the *attTn7*-TnsD complex in the minor-groove (Kuduvalli *et al.*, 2001). The ability of TnsC to recognize and direct site-specific and orientation-specific insertions adjacent to both the *attTn7*-TnsD complex and triplex DNA raises the possibility that the these two targets contain a similar effector signal for site-specific targeting.

Detailed structural information on pyrimidine motif intermolecular triplexes from molecular modeling, footprinting, NMR studies, and a recent crystal structure (Chomilier *et al.*, 1992; Stonehouse & Fox, 1994; Asensio, 1998; Rhee *et al.*, 1999), reveal that TFOs generate modest changes in the confor-

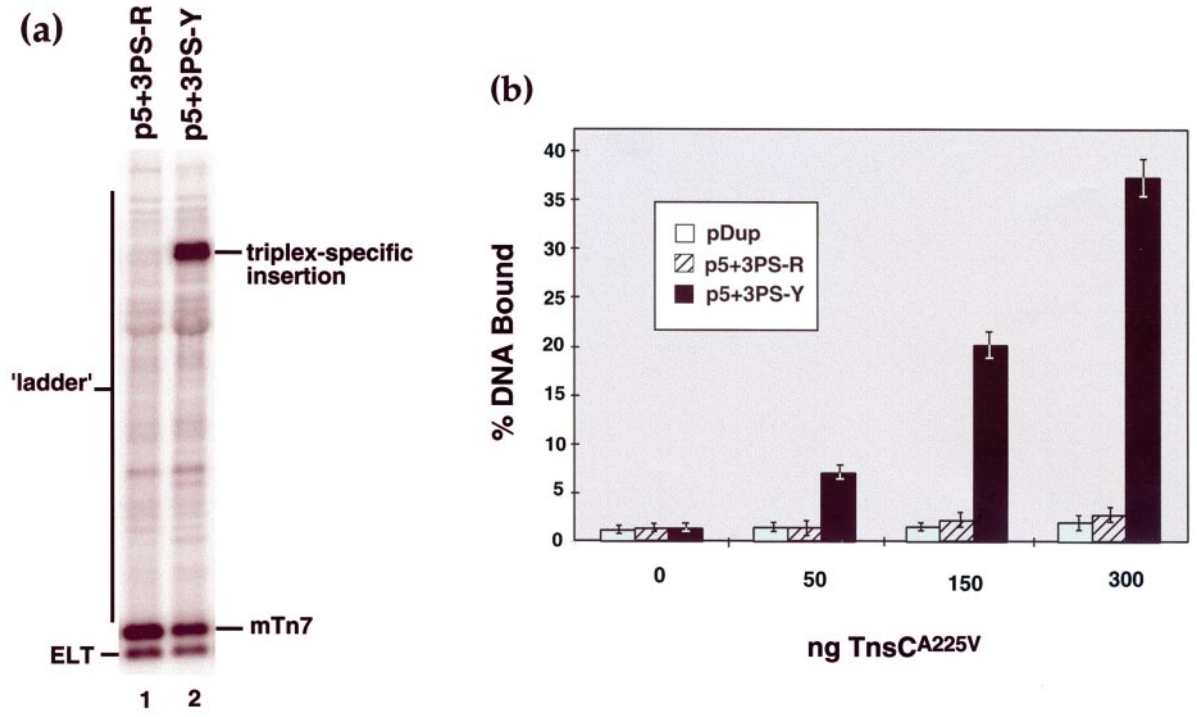


Figure 4. Binding and targeting assays for triplexes containing psoralen at both the 5′ and 3′ ends. Assay for Tn*7* target specificity for the triplexes containing di-psoralenated TFOs (p5 + 3PS-R and p5 + 3PS-Y). (a) Comparison of *in vitro* transposition reaction products as visualized by Southern blot. The p5 + 3PS-R target shows a ladder of bands corresponding to a population of ''random'' Tn*7* insertions (lane 1) while the p5 + 3PS-Y target generates an accumulation of bands indicating triplex-specific Tn*7* insertions (lane 2). The band denoted excised linear transposon (ELT) is an intermediate of the cut and paste transposition reaction. The mini-Tn*7* donor substrate (mTn*7*) and the target DNA are indicated. Each 100 µl reaction had 0.37 pmol of target plasmid DNA, 0.04 pmol of mTn*7*-containing donor plasmid DNA, 1.2 pmol of TnsA, 0.3 pmol of TnsB, 1.0 pmol of TnsC, 25.9 mM Hepes (pH 7.6), 2.5 mM Tris (pH 7.6), 60 µg/ml BSA, 2.0 mM ATP, 2.0 mM DTT, 5.0 mM KCl, 14.3 mM NaCl, 16 µM EDTA, 120 µM $MgCl_2$, 120 µM Chaps, 4 µM PMSF, 0.4 % glycerol, and 15 mM magnesium acetate. The target plasmid DNA and TnsC^{A225V} were first mixed and assembled at 30 °C for 20 minutes, and then the remaining components were added and incubated at 30 °C for another 30 minutes. The reaction was stopped with phenol/chloroform-extraction and the DNA precipitated, digested with *Hin*dIII, and separated on a 0.6 % agarose gel. The gel was subject to Southern blotting and probed with radiolabeled oligonucleotide (NLC94), corresponding to the right end of Tn*7*. Following phosphorimaging (Molecular Dynamics), the products were quantified using ImageQuant software (Molecular Dynamics). All of the components except TnsA, TnsB, donor plasmid and magnesium acetate were incubated at 30 °C for 20 minutes to allow target assembly prior to the reaction. The remaining components were added and allowed to react at 30 °C for 20 minutes. Phenol/chloroform-extraction and ethanol-precipitation were done before restriction digest and Southern blot. For ease of visualization and quantification of reaction products, the gain-of-function allele of TnsC, TnsC^{A225V}, was primarily used; however, similar targeting preferences are found for TnsC^{wt}. (b) The relative affinities of TnsC^{A225V} for the triplexes containing di-psoralenated TFOs (p5 + 3PS-R and p5 + 3PS-Y) are compared.

mation of the duplex target to which they anneal. Specifically, the docking of the TFO splits the major groove, in effect creating two additional minor grooves and, in turn, narrows the minor groove of the duplex. Also, the triplex-duplex junctions are differentially unwound, giving rise to asymmetric distortions. Footprinting studies reveal that the triplex targets used in these studies also contain alterations at the triplex-duplex junctions (Rao *et al.*, 2000; our unpublished results). An attractive hypothesis is that the unique combination of structural features within pyrimidine motif triplexes provokes the binding of TnsC. These features may be similar to the *attTn7*-TnsD complex, while purine motif triplexes simply do not contain the specific structural signals required to recruit TnsC. Though few structural studies on purine motif intermolecular triplexes have been done, obvious structural differences may arise due to the opposite orientation of the third strand, for example, which could, in turn, alter the triplex-duplex junctions and groove widths. The fact that TnsC recognizes pyrimidine and not purine motif triplexes rules out the notion that TnsC is simply pausing at a triplex-based molecular ''road-block''. Our results confirm that TnsC is not simply responding to a global feature of the pyrimidine triplex target, but to a more subtle signal contained therein.

Effect of psoralen on Tn*7* insertion

The effect of the position of psoralen on targeting further supports the idea that TnsC responds to a combination of particular features within a potential target. The insertion of Tn*7* into various pyrimidine triplex targets maintained a distance of ~25 bp from the site of psoralen intercalation. Though purine motif triplexes also contained psoralen, Tn*7* did not insert adjacent to these targets. It has been shown that psoralen induces conformational changes within the DNA into which it has intercalated (Hwang *et al.*, 1996). However, without high-resolution structural information on a psoralenated triplex-TnsC complex, it is difficult to understand exactly how the location of psoralen within the pyrimidine motif triplex affects the insertion profile of Tn*7*. We suspect that the intercalated psoralen moiety mimics a specific aspect of the *attTn7*-TnsD signal involved in the precise alignment of the transpososome. By moving psoralen from the 5′ to the 3′ end of the triplex, the alignment of the transposase, and thus the point of Tn*7* insertion, is in turn moved a similar distance.

Notable features of the triplex targeting pathway: orientation bias and immunity

The TnsC-triplex interaction leads to Tn*7* insertions into a specific location and with a specific orientation. Like the TnsABC + D pathway, Tn*7* inserts >98 % with a single orientation adjacent to the triplex target (Rao *et al.*, 2000). The similar orientation bias that is found within the triplex pathway reveals that some aspect of the TnsABC machinery, and not TnsD, is responsible for determining the orientation of Tn*7* upon insertion.

We also note that double insertions are not observed for either the triplex targets used in this study or for targets that contained two distinct triplex regions on the same plasmid (data not shown). Thus, triplex targets that contain a copy of Tn*7* are recognized as immune targets. Like the TnsD-*attTn7* target complex, pyrimidine motif triplex DNA represents a positive targeting signal for Tn*7* transposition, with the ability to orient and direct the insertion of Tn*7* into a specific location.

Implications of the interaction of TnsC with triplex DNA

The identification of proteins that specifically interact with triplex DNA has been used to support the idea that triplex DNA may exist *in vivo* and may, in fact, play a biological role (Frank-Kamenetskii & Mirkin, 1995). Multiple lines of evidence have implicated triplex DNA in several cellular processes, including transcription, replication, and recombination (Soyfer & Potaman, 1996). However, beyond simple binding to triplex DNA, the biological purpose for all previously reported triplex-binding proteins is not known (Table 1). In contrast, the biological role of TnsC is well defined. TnsC regulates Tn*7* transposition through interaction with target DNA for the selection of a proper target site, leading to activation of the transposase. A unique feature of the recognition of triplex by TnsC DNA is that TnsC is part of the repertoire of proteins that carry out transposition of Tn*7*: the binding of TnsC leads to a highly specific transposition event and thus provides a functional assay for triplex formation. We have shown here that our assay provides insight into both the mechanism of target site selection by Tn*7* and the potential structural differences amongst

Table 1. Summary of the previously reported triplex DNA-binding proteins

Source	Protein	Biological Role	Binding Preferences: Purine Inter-	Purine Intra-	Pyrimidine Inter-	Pyrimidine Intra-	H-DNA	Ref
E. coli	TnsC	Tn7 targeting, regulation	X	nt	√	√	√	1
S. cerevisiae	Y3BP1	unknown	√	nt	X	nt	nt	2
Drosophila	GAGA Factor	Regulation of transcription	nt	nt	--	X	nt	3
Hela extracts	3BP	unknown	√	nt	X	nt	nt	4
Hela extracts	Triplex Binding Protein	unknown	nt	--	nt	√	nt	5
Hela extracts	Triplex Binding Protein	unknown	√	nt	X	nt	nt	6

The binding preferences of the reported triplex-binding proteins for each type of triple-helical motif are listed. A check (tick) mark indicates a preferential binding of triplex over duplex DNA, while (--), implies that the binding was equal to that of duplex controls. An X indicates no observable binding to the motif listed, while the nt indicates that the motif was not tested. The biological roles for each protein are listed. Ref. 1, Rao *et al.* (2000); 2, Nelson *et al.* (2000); 3, Jiménez-García *et al.* (1998); 4, Musso *et al.* (1998); 5, Guieysse *et al.* (1997); 6, Kiyama & Camerini-Otero (1991).

purine and pyrimidine triplex motifs. The ability of Tn7 to target triplex regions of DNA for site-specific insertion could be extended to *in vivo* applications, providing a functional assay for capturing or ''tagging'' naturally occurring triplex structures such as *H*-DNA. As direct proof of triplex formation *in vivo* is lacking, Tn7 may provide a powerful method of characterizing such structures.

Acknowledgments

We are grateful to Dr Paul Miller for all his insight and expertise given to this work, and to Dr Prasad Kuduvalli and Dr David Noll for helpful discussion. We thank Patti Eckhoff and the members of the Craig laboratory for preparation and critical reading of the manuscript. This work was supported by the National Institutes of Health under grant GM53824 to N.L.C.. N.L.C. is an investigator with the Howard Hughes Medical Institute.

References

Asensio, J. L., Dosanjh, H. S., Jenkins, T. C. & Lane, A. N. (1998). Thermodynamic, kinetic, and conformational Properties of a parallel intermolecular DNA triplex containing 5′ and 3′ junctions. *Biochemistry,* **37**, 15188-15198.

Arciszewska, L. K., Drake, D. & Craig, N. L. (1989). Transposon Tn7 *cis*-acting sequences in transposition and transposition immunity. *J. Mol. Biol.* **207**, 35-52.

Bainton, R., Gamas, P. & Craig, N. L. (1991). Tn7 transposition *in vitro* proceeds through an excised transposon intermediate generated by staggered breaks in DNA. *Cell,* **65**, 805-816.

Bainton, R. J., Kubo, K. M., Feng, J.-N. & Craig, N. L. (1993). Tn7 transposition: target DNA recognition is mediated by multiple Tn7-encoded proteins in a purified in vitro system. *Cell,* **72**, 931-943.

Biery, M. C., Steward, F., Stellwagen, A. E., Raleigh, E. A. & Craig, N. L. (2000). A simple in vitro Tn7-based transposition system with low target site selectivity for genome and gene analysis. *Nucl. Acids Res.* **28**, 1067-1077.

Chomilier, J., Sun, J.-S., Collier, D. A., Garestier, T., Helene, C. & Lavery, R. (1992). A computational and experimental study of the bending induced at a double-triple helix junction. *Biophys. Chem.* **45**, 143-152.

Craig, N. L. (1991). Understanding how transcriptional enhancers work. *Curr. Biol.* **1**, 100-102.

Craig, N. L. (1996). Transposition. In Escherichia coli *and* Salmonella: *Cellular and Molecular Biology* (Neidhardt, F. C., Curtiss, R. I., Ingraham, J. L., Lin, E. C. C., Low, K. B., Magasanik, B., Reznikoff, W. S., Riley, M., Schaechter, M. & Umbarger, H. E., eds), 2nd edit., pp. 2339-2362, American Society for Microbiology, Washington, DC.

Frank-Kamenetskii, M. D. & Mirkin, S. M. (1995). Triplex DNA structures. *Annu. Rev. Biochem.* **64**, 65-95.

Gamas, P. & Craig, N. L. (1992). Purification and characterization of TnsC, a Tn7 transposition protein that binds ATP and DNA. *Nucl. Acids Res.* **20**, 2525-2532.

Guieysse, A.-L., Praseuth, D. & Hélène, C. (1997). Identification of a triplex DNA-binding protein from human cells. *J. Mol. Biol.* **267**, 289-298.

Hauer, B. & Shapiro, J. A. (1984). Control of Tn7 transposition. *Mol. Gen. Genet.* **194**, 149-158.

Hwang, G. S., Kim, J. K. & Choi, B. S. (1996). The solution structure of a psoralen cross-linked DNA duplex by NMR and relaxation matrix refinement. *Biochem. Biophys. Res. Commun.* **219**, 191-197.

Jiménez-García, E., Vaquero, A., Espinás, M. L., Soliva, R., Orozco, M., Bernués, J. & Azorín, F. (1998). The GAGA factor of *Drosophila* binds triple-stranded DNA. *J. Biol. Chem.* **273**, 24640-25648.

Kiyama, R. & Camerini-Otero, R. D. (1991). The triplex DNA-binding protein from human cells: purification and characterization. *Proc. Natl Acad. Sci. USA,* **88**, 10450-10454.

Kuduvalli, P., Rao, J. & Craig, N. L. (2001). Target site structure plays a critical role in Tn7 transposition. *EMBO J.* **20**, 924-932.

Lee, J. S., Woodsworth, M. L., Latimer, L. J. & Morgan, A. R. (1984). Poly(pyrimidine) poly(purine) synthetic DNAs containing 5-methylcytosine form stable triplexes at neutral pH. *Nucl. Acids Res.* **12**, 6603-6614.

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

Miller, P. S., Bi, G., Kipp, S. A., Fok, V. & DeLong, R. K. (1996). Triplex formation by a psoralen-conjugated oligodeoxyribonucleotide containing the base analog 8-oxo-adenine. *Nucl. Acids Res.* **24**, 730-736.

Musso, M., Nelson, L. D. & Van Dyke, M. W. (1998). Characterization purine-motif triplex DNA-binding proteins in HeLa extracts. *Biochemistry,* **37**, 3086-3095.

Nelson, L. D., Musso, M. & Van Dyke, M. W. (2000). The yeast STM1 gene encodes a purine motif triple helical DNA-binding protein. *J. Biol. Chem.* **275**, 5573-5581.

Peters, J. E. & Craig, N. L. (2000). Tn7 transposes proximal to DNA double-strand breaks and into regions where chromosomal DNA replication terminates. *Mol. Cell,* **6**, 573-582.

Povsic, T. J. & Dervan, P. B. (1989). Triple helix formation by oligonucleotides on DNA extended to the physiological pH range. *J. Am. Chem. Soc.* **111**, 3059-3061.

Rao, J. E., Miller, P. S. & Craig, N. L. (2000). Recognition of triple-helical DNA structures by transposon Tn7. *Proc. Natl Acad. Sci. USA,* **97**, 3936-3941.

Rhee, S., Han, Z., Liu, K., Miles, H. T. & Davies, D. R. (1999). Structure of a triple helical DNA with a triple-duplex junction. *Biochemistry,* **38**, 16810-16815.

Sarnovsky, R., May, E. W. & 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.

Soyfer, V. N. & Potaman, V. N. (1996). *Triple-helical Nucleic Acids*, Springer-Verlag, New York.

Stellwagen, A. E. (2001). Analysis of gain of function mutants of an ATP-dependent regulator of Tn7 transposition. *J. Mol. Biol.* **305**, 633-642.

Stellwagen, A. & Craig, N. L. (1997a). Avoiding self: Two Tn7-encoded proteins mediate target immunity in Tn7 transposition. *EMBO J.* **16**, 6823-6834.

Stellwagen, A. & Craig, N. L. (1997b). Gain-of-function mutations in TnsC, an ATP-dependent transposition protein which activates the bacterial transposon Tn*7*. *Genetics,* **145**, 573-585.

Stellwagen, A. & Craig, N. L. (1998). Mobile DNA elements: controlling transposition with ATP-dependent molecular switches. *Trends Biochem. Sci.* **23**, 486-490.

Stonehouse, T. J. & Fox, K. R. (1994). DNase I footprinting of a triple helix formation at polypurine tracts by acridine-linked oligopyrimidines: stringency, structural changes and interaction with minor groove binding ligands. *Biochim. Biophys. Acta,* **1218**, 322-330.

Wolkow, C. A., DeBoy, R. T. & Craig, N. L. (1996). Conjugating plasmids are preferred targets for Tn*7*. *Genes Dev.* **10**, 2145-2157.

Edited by M. Belfort

(Received 2 November 2000; received in revised form 10 February 2001; accepted 15 February 2001)