Tn7: SMARTER THAN WE THOUGHT

Joseph E. Peters and Nancy L. Craig

A notable feature of transposable elements — segments of DNA that can move from one position to another in genomes — is that they are highly prevalent, despite the fact that their translocation can result in mutation. The bacterial transposon Tn7 uses an elaborate system of target-site selection pathways that favours the dispersal of Tn7 in diverse hosts as well as minimizing its negative effects.

REPLICON
A contiguous DNA segment with an origin(s) of DNA replication.

CONJUGAL PLASMIDS
A class of autonomously replicating circular DNA that can move between bacteria.

Howard Hughes Medical Institute, Department of Molecular Biology and Genetics, Johns Hopkins University School of Medicine, 725 North Wolfe Street, Baltimore, Maryland 21205, USA.
Correspondence to N.L.C.
e-mail: ncraig@jhmi.edu

Transposons — a class of mobile genetic elements — are widespread in nature, being present in virtually every organism examined. Indeed, the sequencing of the human genome has shown that 35–50% of the genome consists of transposable elements, or relics thereof[1,2]. Although some transposons might encode beneficial functions — such as drug resistance or pathogenicity determinants in bacteria — more often they seem to provide little discernible benefit to the host. Therefore, the success of these parasitic elements reflects — in part — their ability to spread throughout the genome, even though their translocation can cause mutations[3]. What molecular mechanisms can account for such success?

A transposon is a discrete DNA segment that can move in a cell between sites that lack homology by using a self-encoded recombinase called a transposase. The bacterial transposon Tn7 is a particularly sophisticated mobile element, and it has developed several alternative lifestyles to promote its propagation[4]. In one transposition pathway, Tn7 resembles many other transposable elements in that it transposes to many different sites at low frequency. Although these sites are unrelated in their DNA sequence, there is a marked preference for Tn7 insertion into certain REPLICONS. Tn7 transposes preferentially to CONJUGAL PLASMIDS when they enter the cell (FIG. 1). This preferential targeting to such plasmids by Tn7, and the generally broad host range of such plasmids, contributes to the dispersal of Tn7 among bacterial populations. Tn7, like many other bacterial transposons, carries genes that encode antibiotic-resistance determinants. So, this plasmid-dispersal pathway can lead to the acquisition of antibiotic resistance by many different bacteria.

In another transposition pathway, Tn7 inserts at a high frequency into a single specific site in bacterial chromosomes (FIG. 1). Insertion of Tn7 into this site is not deleterious to the host and provides a 'safe' site for transposon insertion. This site-specific insertion pathway therefore provides a mechanism to promote the successful coexistence of Tn7 and its host, and to promote the transmission of Tn7 to daughter cells.

Tn7 seems smarter than was originally thought because we now appreciate the remarkable level of sophistication it has evolved for target-site selection. Tn7 activates transposition only in the presence of two specific kinds of target DNA, which are expected to facilitate the safe propagation of Tn7 in diverse hosts. In this review, we discuss recent findings concerning how Tn7 uses distinct but partially overlapping combinations of proteins to identify these alternative target sites. In one pathway, transposition is directed into conjugal plasmids, which can move promiscuously between bacteria, by recognizing specialized replication structures. In a second pathway, transposition is directed by a specific sequence that is in an essential and highly conserved gene found in many strains of bacteria. But, whereas Tn7 recognizes sequences within an essential gene, transposition actually occurs downstream of the gene, preventing Tn7 from killing the host. The study of Tn7 promises to expand our understanding of the pervasive nature of transposons, as well as providing insight into the regulation of other multiprotein replication and recombination reactions.

The structure of Tn7

The sequences in Tn7 on which the transposition machinery acts lie in the ends of Tn7. The necessary segment at the left end (Tn7-L) is about 150 base pairs

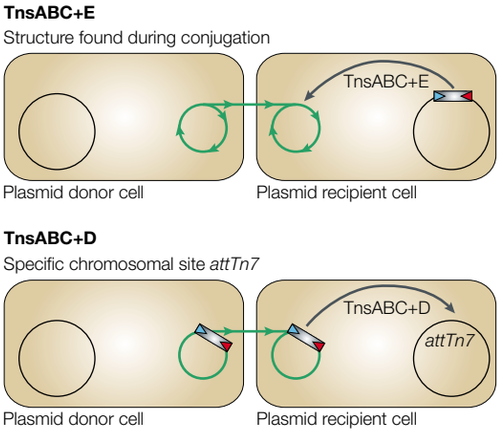


Figure 1 | **Tn7 has two pathways of transposition to facilitate its safe propagation in diverse hosts.** TnsD and TnsE are alternative target-site selectors that activate the TnsABC proteins to execute the breakage and joining reactions that underlie Tn7 transposition. TnsABC+E-mediated transposition specifically promotes transposition into plasmids that can move between bacteria when they enter a recipient cell (conjugal plasmids). TnsABC+D-mediated transposition promotes transposition into a single chromosomal site, called its attachment site or *attTn7*, found in diverse bacteria. The TnsABC+E pathway would promote horizontal transfer between bacteria, whereas the TnsABC+D pathway would promote vertical transmission to daughter cells once in a new host bacterium.

long, and the necessary segment at the right end of Tn7 (Tn7-R) is about 90 bp long (FIG. 2). Any piece of DNA flanked by these end segments can transpose in the presence of the Tn7 transposition proteins. Although the Tn7-L and Tn7-R segments are related, they are not identical, and Tn7 shows a preferential orientation of insertion at particular targets, which is dictated by the differences in the left and right ends of Tn7. Each end of Tn7 contains a series of 22-bp transposase-binding sites that are recognized by the Tn7 recombination machinery — three non-overlapping binding sites in Tn7-L, and four overlapping binding sites in Tn7-R. An interesting and as yet unanswered question is how differences in the Tn7 end sequences translate into differences in the polarity of the recombination machinery.

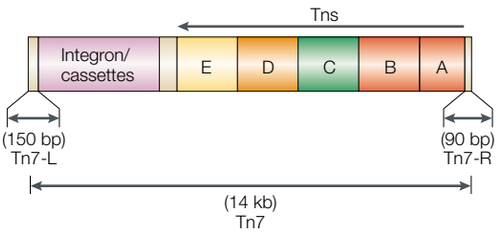


Figure 2 | **The structure of Tn7.** Tn7 has distinct left (Tn7-L) and right (Tn7-R) ends that allow Tn7 to integrate with a characteristic orientation. Each end of Tn7 contains a series of 22-base-pair transposase-binding sites — three binding sites in Tn7-L and four in Tn7-R. The Tn7 transposition proteins — TnsA, TnsB, TnsC, TnsD and TnsE — are encoded at the right end of Tn7 with their 5′ termini closest to the Tn7-R. The left end of Tn7 also encodes several antibiotic-resistance determinants as cassettes associated with an inactive integron recombination system.

Tn7 also encodes an integron — a DNA segment containing several cassettes of antibiotic-resistance genes[5]. Although these cassettes are fixed in Tn7 owing to a mutation of the cognate recombinase, they can undergo rearrangements in hosts that express a related recombinase, leading to alternative combinations of antibiotic-resistance genes.

Tn7 transposition proteins

Tn7 encodes five transposition proteins: TnsA, TnsB, TnsC, TnsD and TnsE[6–8] (TABLE 1, FIG. 2). Subsets of these proteins promote the insertion of Tn7 into different classes of target sites.

TnsA and TnsB. TnsA and TnsB together form the transposase that specifically recognizes the ends of the transposon. TnsAB excises Tn7 from the donor site by introducing double-stranded DNA breaks (DSBs) at each end of the transposable element, and then joins the exposed transposon ends to the target DNA[9–11] (FIG. 3). The fundamental chemistry behind this recombination reaction is identical to that occurring in all other transposable elements that have been characterized biochemically[12]. The initial step of recombination involves cleavage of the 3′ ends of the transposon from the flanking donor DNA. Tn7 is also cleaved at the 5′ ends; however, this step is not universal to all transposons[13]. The 3′ ends of the transposon are then joined to the target DNA, yielding a product in which the transposon is covalently linked to the target DNA at its 3′ ends and is flanked by short gaps at its 5′ ends. Repair of these gaps yields a 5-bp target-site duplication, which is a characteristic of Tn7 insertion.

The proteins that make up the Tn7 transposase are quite different. The TnsB protein, which carries out the chemical steps at the 3′ end of Tn7, is a member of the retroviral integrase superfamily that includes the retroviral integrases and various other bacterial transposases[14–16]. Members of the retroviral integrase superfamily of enzymes have a triad of acidic residues in their active site that is often referred to as the DDE motif. As is true of the retroviral integrases and the other bacterial transposons, the TnsB triad of acidic residues is required both for cutting from the donor DNA and for joining to a target DNA[9,12,17,18] (FIG. 3). TnsB also has the ability to recognize the *cis*-acting ends of Tn7 (REF. 19). By contrast, TnsA, which executes the breaks at the 5′ end of the transposon, resembles a type II restriction enzyme[9,20,21] (FIG. 3). Structurally, TnsA most closely resembles the catalytic domain of the restriction enzyme *Fok*I, a domain that cleaves DNA nonspecifically. TnsA does not have a detectable DNA-binding activity, but instead apparently relies on its interdependent interaction with TnsB to recognize and cleave at the ends of Tn7 (REF. 20).

TnsC. TnsC, an ATP-hydrolysing protein that binds nonspecifically to duplex DNA in the presence of ATP, interacts with the target DNA and with TnsAB to pro-

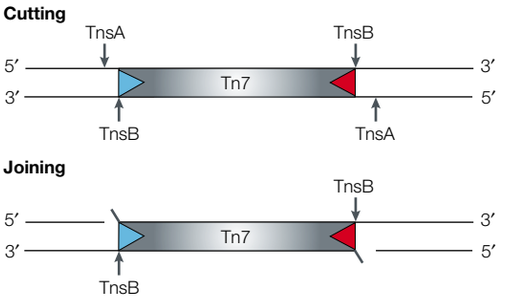


Figure 3 | **TnsA and TnsB form a heteromeric transposase that excises the transposon through double-stranded DNA breaks.** The TnsAB transposase excises Tn7 from the donor backbone by double-stranded DNA breaks. The TnsB protein cleaves at the 3′ ends of the transposon and TnsA cleaves in the flanking DNA on the 5′ ends of the transposon. The 3′ ends of the transposon are then joined to the target DNA by TnsB, yielding a product in which the transposon is covalently linked to target DNA at the 3′ ends and is flanked by short gaps at the 5′ ends. Repair of these gaps yields a 5-base-pair target-site duplication that is characteristic of transposon insertion.

mote the excision and insertion of Tn7 (REFS 22–24). The ability of TnsC to activate the transposase depends on its interaction with a target DNA along with its appropriate targeting protein, TnsD or TnsE.

TnsD and TnsE. TnsD and TnsE are alternative target selectors, which allow the TnsABC proteins to execute the breakage and joining reactions that excise Tn7 from the donor and insert it into a target site. Notably, no recombination occurs in the presence of TnsABC only[8]. So, TnsD and TnsE act not only as target-site selectors, but also as activators of the core TnsABC machinery. TnsD and TnsE are both DNA-binding proteins and, as explained below, their ability to interact with a particular target DNA is key to the target-site selection of Tn7.

TnsABC+E transposition

Preferential insertion into conjugal plasmids. Most transposons direct insertions into many different sites, with relatively little selection for a particular DNA sequence. Indeed, when Tn7 uses the TnsABC+E proteins to promote transposition, it inserts into sites that are unrelated in their DNA sequence. However, TnsABC+E transposition shows an interesting preference for conjugal plasmids that can move between cells[25]. The frequency of TnsABC+E transposition is stimulated by plasmid transfer, but virtually all of the transposition events occur in the plasmid, despite the fact that the bacterial chromosome could be up to 100 times larger. These plasmids have a generally broad host range and can move between various bacteria, which means that the TnsABC+E pathway encourages the spread of Tn7 between different species.

TnsABC+E-mediated transposition targets an aspect of the specialized DNA replication that occurs during conjugal plasmid transfer. Conjugal transfer initiates when a plasmid-encoded protein makes a single-strand nick at the *cis*-acting ORIGIN OF TRANSFER (*oriT*), binds to the 5′ end of the nick, and directs one strand of the plasmid into a recipient cell. A complementary DNA strand is synthesized during transfer in both the donor and the recipient cells to re-establish duplex DNA. However, synthesis of the second strand of plasmid DNA is fundamentally different in the donor and recipient cells because the DNA strands are of opposite polarity. In the donor cell, plasmid DNA replication occurs by leading-strand DNA synthesis and needs to be primed only once, whereas in the recipient cell, DNA replication must be continually re-primed in a process called DISCONTINUOUS OR LAGGING-STRAND DNA SYNTHESIS (FIG. 4). As explained below, various experiments indicate that TnsE recognizes an aspect of lagging-strand DNA synthesis in recipient cells to target transposition to conjugal plasmids.

A notable hallmark of TnsE-mediated insertions directed into conjugal plasmids is that they occur in a single orientation at many different positions in the plasmid[25,26] (FIG. 4). This orientation bias indicates that the transposition machinery recognizes a polar aspect of conjugal plasmid biology. However, because many polar processes occur during plasmid transfer (for example, transfer itself and a distinct form of DNA replication), examination of insertions into conjugal plasmids could not show what the Tn7 machinery recognized.

An important insight into what TnsABC+E recognizes came from examining several Tn7-insertion events into a bacterial chromosome that cannot transfer between cells[27]. The chromosome is a poor

ORIGIN OF TRANSFER
A *cis*-acting DNA site found in conjugal plasmids that is recognized by the transfer proteins and at which DNA transfer initiates.

DISCONTINUOUS OR LAGGING-STRAND DNA SYNTHESIS
DNA replication that must continually be re-primed approximately every two kilobase pairs in bacteria.

Table 1 | **Tn7 proteins and their roles in transposition**

Protein	Function	Biochemical activities	Structural homologues or motifs
TnsA	Transposase subunit	Cutting at the 5′ ends of Tn7	Type II restriction enzymes
TnsB	Transposase subunit	Cutting and joining at the 3′ ends of Tn7	Retroviral integrases and transposase DDE motif
TnsC	Regulator/connector	ATP-dependent DNA-binding and ATP hydrolysis	Nucleotide-binding motif
TnsD	Target selection — recognition of *attTn7*	Sequence-specific DNA binding	None
TnsE	Target selection — conjugal replication/ lagging-strand DNA synthesis	Structure-specific DNA binding; 3′ recessed ends	None

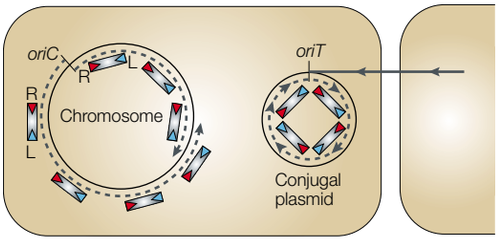


Figure 4 | **TnsABC+E-mediated transposition occurs with an orientation bias in the direction of DNA replication.** Analysis of TnsABC+E-mediated insertion events in various replicons indicates that a structure found during DNA replication is crucial for transposition. TnsE-mediated insertion occurs preferentially into actively conjugating DNA in a single orientation in recipient cells in which DNA replication occurs exclusively by lagging-strand DNA synthesis. Transposition events occur in opposite orientations on either side of the single origin of replication in the circular bacterial chromosome. The origin of conjugal transfer, *oriT*, and the origin of bi-directional chromosomal DNA replication, *oriC*, are indicated. The dashed line indicates the direction of DNA replication and the left (L, blue) to right (R, red) orientation of TnsABC+E-mediated transposition events is indicated.

target for TnsABC+E transposition, but transposition events can be isolated. TnsABC+E-mediated insertion events in the chromosome also occur with a striking and distinct orientation bias — insertions occur in opposite orientations on either side of the single origin of replication in the circular bacterial chromosome (FIG. 4). TnsE-mediated transposition events in the chromosome also show a regional insertion bias.

Chromosomal DNA replication is initiated from a single ORIGIN OF BI-DIRECTIONAL DNA REPLICATION (*oriC*) in the circular *Escherichia coli* genome. Replication proceeds bi-directionally — that is, both clockwise and anti-clockwise — around this genome, and terminates over a wide region that is approximately equidistant from the origin. Because DNA replication is initiated bi-directionally from a single origin, a given strand of the chromosome is replicated by leading-strand DNA synthesis on one side of *oriC* and by lagging-strand DNA synthesis on the other side. Examining the orientation of chromosomal insertions indicates that a structure or complex associated with either leading- or lagging-strand DNA synthesis is crucial for attracting TnsABC+E transposition events. Comparison of the orientation of TnsABC+E insertion events in the chromosome with those in the conjugal plasmid in recipient cells has shown that the strand of DNA replicated by lagging-strand synthesis provides a crucial signal for transposition (FIG. 5).

The finding that lagging-strand DNA synthesis during conjugation — rather than chromosomal DNA replication — is an attractive target for TnsABC+E-promoted transposition, might reflect the distinctive nature of the multisubunit machines that replicate the chromosome and carry out conjugal DNA replication. The machine used during chromosomal DNA replication coordinates both leading- and lagging-strand DNA synthesis. This elaborate replication apparatus used in the chromosome might block the lagging strand from a TnsE interaction, thereby inhibiting TnsABC+E transposition. This case is unlike that of second-strand DNA synthesis during conjugation, in which leading- and lagging-strand DNA synthesis occur in different cells, and it does not involve the more elaborate replication machine used in the chromosome. The ability of TnsABC+E-mediated transposition to target the lagging strand when replicated in isolation during conjugation could guard against 'random' insertions in the chromosome that could be lethal to the host.

Analysis of purified TnsE protein indicates that its recognition of certain DNA structures is probably key to target-site selection[27]. TnsE is a DNA-binding protein that preferentially binds DNA structures that contain recessed 3′ ends — structures that would be common during lagging-strand DNA synthesis.

The best understanding of what TnsE uses for target-site selection will come from establishing the reconstituted reaction *in vitro*. However, work with high-activity TnsE mutants indicates that the DNA-binding activity of TnsE is indeed involved in transposition[27]. Following random mutagenesis, TnsE mutants were isolated that allow a ~1,000-fold stimulation in the frequency of TnsE-mediated transposition. These mutant proteins bind structures with 3′ recessed ends much better than does wild-type TnsE, indicating that binding is relevant to transposase activation, and that TnsE probably interacts with the target DNA.

Insertion into the chromosome. Other forms of DNA metabolism besides conjugal DNA replication have also been found to attract TnsABC+E transposition events. As explained above, Tn7 transposition can occur into the chromosome at a low frequency. But such insertions are not generally distributed throughout the chromosome:

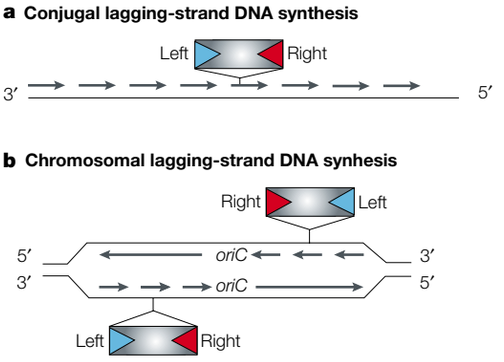


Figure 5 | **TnsABC+E targets structures associated with lagging-strand DNA synthesis. a** | Tn7 insertion into conjugating plasmids occurs in a single orientation in recipient cells, regardless of where the insertions occur. **b** | Tn7 insertions occur in opposite orientations on either side of the single origin of DNA replication in *Escherichia coli oriC*. However, insertions in both replicons are in the same orientation with respect to lagging-strand DNA synthesis. Lagging-strand DNA synthesis is indicated by a series of short arrows and leading-strand DNA synthesis is indicated by a single long arrow.

ORIGIN OF BI-DIRECTIONAL DNA REPLICATION
The site at which the DNA strands are melted and the replisome is assembled for DNA replication, which progresses in the 3′ and 5′ directions and replicates both strands.

they occur preferentially in the region where bi-directional DNA replication terminates[28]. Chromosomal DNA replication can terminate at specific sites called *ter* sites, where an accessory protein binds and prevents the helicase that opens up the DNA ahead of the DNA replication apparatus from progressing. TnsABC+E insertions can be remarkably close to the sites where DNA replication terminates (within a few thousand base pairs of termination sites in a 4.6 Mb chromosome). However, DNA replication need not terminate at *ter* sites; the activity of *ter* sites is not essential for cell viability and, presumably, replication could terminate simply when two replication forks meet. Because DNA replication might terminate over a large region, it is unclear whether TnsABC+E insertions in the chromosome must occur where DNA replication terminates. What could attract such insertions to the region where DNA replication terminates? Perhaps when replication forks are terminated and the replication machine collapses, the lagging strand might become accessible to TnsE binding and allow subsequent recruitment of the core TnsABC machinery.

TnsABC+E transposition is also stimulated to occur proximal to DSBs[28]. Given that DNA replication is important in the repair of DSBs, this observation could indicate that TnsE-mediated transposition recognizes replication structures found during the repair process. In *E. coli*, a special priming machine is required to initiate lagging-strand DNA synthesis for the repair of DSBs and restart of replication forks[29]. Further study of TnsABC+E transposition could provide a genetic and biochemical probe into how DNA replication mediates DNA repair.

The preferential recognition of lagging-strand DNA synthesis by TnsABC+E provides a 'smart' strategy for preferentially recognizing conjugating DNAs that facilitates the spread of the element among different bacteria. Random TnsABC+E insertions might not be attracted to the chromosome because the DNA strand that is replicated by lagging-strand synthesis in the chromosome might not be accessible to TnsE.

TnsABC+D transposition

***Recognition of* attTn7.** TnsABC+D promote Tn7 insertion to a specific DNA sequence, called *attTn7*, which is found in many bacteria[30–32]. TnsD is a sequence-specific DNA-binding protein that binds to *attTn7* (REFS 22,33). The sequence recognized by TnsD lies in the 3′ end of the bacterial glutamine synthetase (*glmS*) gene, which is highly conserved because of its role in the synthesis of *N*-acetyl glucosamine. However, the actual point of Tn7 insertion lies downstream of the *glmS* coding region (FIG. 6). So, Tn7 insertion into *attTn7*, although directed by a highly conserved gene, does not lead to destruction of that gene nor, as far as can be determined, to any negative effect on the bacterial host. The *attTn7* sequence found in the *glmS* gene is highly conserved in a wide range of bacteria (and even higher-level organisms[34]), indicating that TnsABC+D-mediated transposition provides a 'smart' strategy that should allow Tn7 to find a 'safe' insertion site in diverse hosts[4]. In contrast to the TnsABC+E pathway, which would promote horizontal transfer between diverse bacteria, the TnsABC+D pathway would promote vertical transmission to daughter cells in a new host bacterium.

The TnsABC+D reaction has been reconstituted *in vitro*[22,35]. The 'core' TnsABC machinery, which shows very little transposition activity *in vitro*, shows no site specificity[10,35,36]. The sequence-specific binding of TnsD to *attTn7* allows the recruitment of TnsC and the rest of the transposition machinery — that is, the TnsAB transposase and the ends of the transposon — to the *attTn7* target site for recombination.

Footprinting studies have also provided information on the spatial organization of TnsC and TnsD on *attTn7* (REFS 22,37). The specific location of Tn7 insertion in *attTn7*, which corresponds to the centre of the 5-bp stagger between where the ends of the transposon join to the target DNA during strand transfer, is denoted as nucleotide position '0', and positive numbering extends 3′ towards the *glmS* gene (FIG. 6). TnsD shows a 'core' region of protection from +30 to +55, making contacts primarily within the major groove of DNA. A second region of weaker protection extends from around nucleotide position +22 to +30, which also contains a TnsD-induced DNA distortion. As described below, results from different kinds of experiments indicate that the TnsD-induced distortion is actually responsible for recruiting TnsC.

The TnsC–TnsD–*attTn7* footprint covers the region that is distorted by TnsD and extends past the insertion site to nucleotide position –15 (FIG. 6). Footprinting with

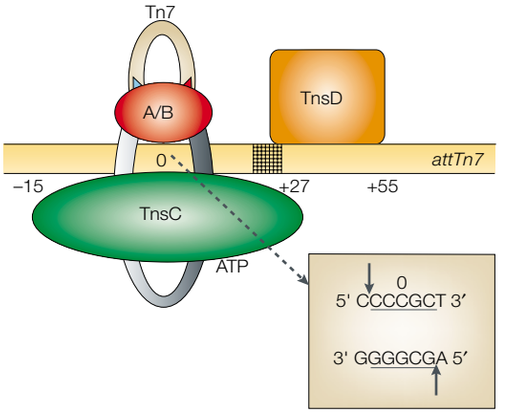


Figure 6 | **TnsD interacts with the *attTn7* DNA sequence to recruit the TnsABC machinery to direct transposition into a single site in the bacterial chromosome.** In the TnsABC+D pathway, the TnsD protein binds adjacent to the actual point of insertion (+30 to +55, according to a convention in which 0 indicates the central point of insertion). The boxed inset shows the 5-base-pair staggered joins (arrows) that Tn7 makes in *attTn7*. TnsD binding induces an asymmetrical distortion of *attTn7* that recruits TnsC. TnsC forms a platform in the minor groove at the insertion site, which can then receive and activate the TnsAB transposase (A/B) bound to the ends of Tn7 (red), leading to insertion into *attTn7*. The hatched box indicates a region of distorted DNA in TnsD–*attTn7* complexes (as judged by hyper-cleavage following treatment with various footprinting reagents) around +27.

attTn7
A DNA sequence that is found in the highly conserved carboxy-terminal coding region of the glutamine synthetase (*glmS*) gene in bacteria and is recognized by the Tn7 protein, TnsD.

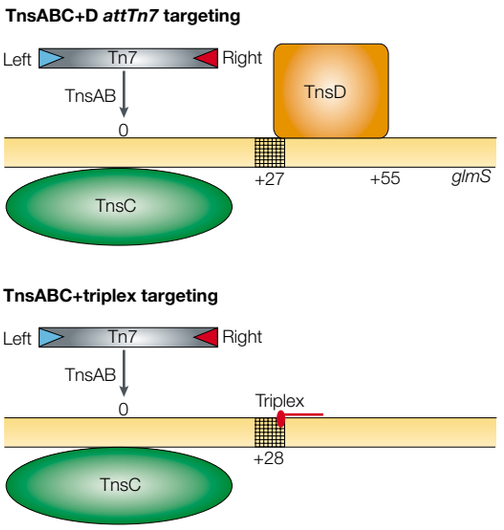


Figure 7 | **Several similarities between target-site selection promoted by TnsD and triplex DNA implicate a crucial role for target DNA structure in transposition.** Transposition directed by TnsD–*attTn7* and triplex DNA are both site specific and orientation specific. Additionally, the location of Tn7 insertion is asymmetrical and about 25 bp away from the target selector (numbering is as described in FIG. 6).

the reagent dimethylsulphate (DMS) has shown that TnsC is a minor groove DNA-binding protein. So, TnsC occupies the minor groove of DNA at the insertion site, leaving the major groove open for transposase binding (FIG. 6). The 5-bp offset in the 5′ direction observed between the nucleotide positions at which the transposon ends are joined to *attTn7* (indicated by arrows in FIG. 6) also indicates that the transposase acts in the major groove of target DNA. Interestingly, increased flexibility of DNA around the insertion site, due to gaps induced by hydroxyl-radical treatment, enhances TnsC binding. This indicates that TnsC alters the structure of DNA at the insertion site, probably making it more conducive to TnsAB binding[37].

Chemical-cleavage studies of the TnsD–*attTn7* complex with reagents used to detect altered DNA structure showed[37] that TnsD induces an asymmetrical distortion at the 5′-end of its binding site, located around nucleotide position +27 of the *attTn7* sequence (FIG. 6). Moreover, increased flexibility of DNA at the 5′-end of the TnsD-binding site, induced by using hydroxyl radicals to introduce gaps in the duplex DNA, leads to greater TnsD binding to *attTn7*, supporting the idea that TnsD induces a perturbation of *attTn7*. The addition of TnsC to the TnsD–*attTn7* complex yields a strong footprint in the region of the TnsD-induced distortion. Furthermore, changing the sequence of the region of *attTn7* that is distorted in the presence of TnsD decreases both TnsD binding and the ability of this mutant site to act as a high-frequency target for Tn7 insertion.

Recognition of triplex DNA. How does TnsC locate a potential target site? One simple model is that a protein–protein interaction between TnsC and TnsD might be enough to bring TnsC to *attTn7* and create an active 'target complex'. However, it is now clear that a direct interaction between TnsC and the target DNA has a central role in target-site recognition. An important role for the target DNA in attracting Tn7 was proposed by *in vitro* experiments in which Tn7 insertion into a plasmid that contains a short region of TRIPLEX DNA was examined[36]. Interestingly, it was found that TnsC can specifically bind to a short region of triplex DNA on a target plasmid and recruit the TnsAB transposase to direct the insertion of Tn7 adjacent to the triplex[36]. The role of structures associated with triplex DNA was of interest given that other mobile elements such as retroviruses had previously been shown to be sensitive to DNA structure when selecting insertion sites[38,39].

What aspect of triplex DNA does TnsC recognize? One consequence of triplex binding is a perturbation of DNA structure located at the triplex–duplex junction, as well as other subtle changes in DNA structure[36]. The finding that altering the position of the linkage between the triplex-forming oligonucleotide and the target plasmid changes the actual position of Tn7 insertion, indicates that DNA distortions indeed have a key role in recruiting TnsC to a target DNA[40].

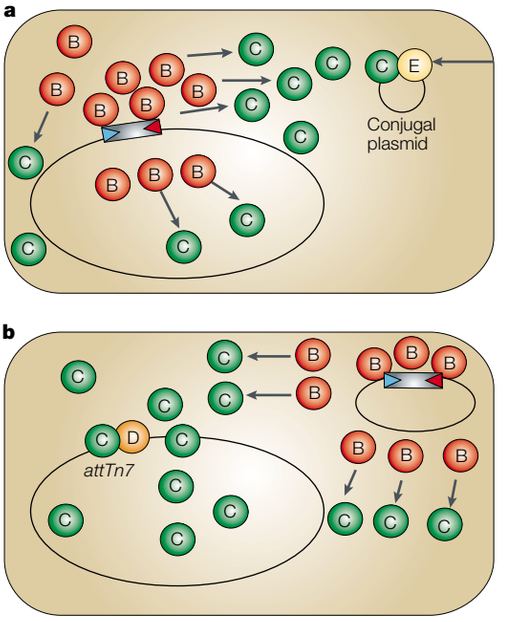
TnsABC+D versus TnsABC+triplex targeting. Interestingly, target-site selection promoted by TnsD and triplex DNA share several similarities[36,37] (FIG. 7). In both cases, insertion is site specific and orientation specific with respect to both the TnsD binding site in *attTn7* and the triplex position. Moreover, TnsC binds preferentially to TnsD–*attTn7* over *attTn7* alone, and also to triplex DNA rather than duplex DNA. Taken together, these observations support the view that a TnsD-induced distortion is key to *attTn7* function, and that the TnsD-induced structural perturbation attracts TnsC to *attTn7* in a manner analogous to the triplex-directed recruitment of TnsC.

It must be noted, however, that a more sophisticated signal must be imparted by the TnsD–*attTn7* complex than is found with triplex DNA alone. The triplex can position the transposition machinery at a particular site on a target DNA, but the presence of the triplex target is not sufficient to activate the TnsABC machinery. By contrast, the interaction of TnsD with *attTn7* both positions the transposition machinery at *attTn7* and activates TnsABC to promote high-frequency transposition. It is unclear whether TnsD induces a very specific distortion in *attTn7* or whether a direct interaction between TnsD and TnsC accounts for the ability of the TnsD–*attTn7* complex to activate, as well as recruit, the transposase.

Therefore, our view of the TnsABC+D pathway includes the sequence-specific binding and asymmetrical distortion of *attTn7* by TnsD, followed by the recruitment of TnsC to this region of TnsD-induced DNA distortion to form a platform of TnsC at the insertion site, which can then receive and activate the transposase and leads to Tn7 insertions into *attTn7* (FIG. 6).

TRIPLEX DNA
A DNA structure in which a third DNA strand docks in the major groove of duplex DNA using atypical Hoogsteen hydrogen bonding.

Box 1 | **Transposon immunity**



Transposon immunity hinders insertions from occurring into a target DNA when a copy of Tn7 already resides in that DNA[6,19] — a process also found with some other transposons[44–46]. The negative signal imposed by target immunity overrides positive signals conveyed by either the TnsE or TnsD pathways[6] (depicted in panels a and b respectively).

Immunity is a 'local' effect in that insertions proximal to a copy of Tn7 are prevented, whereas transposition into remote DNAs lacking Tn7 occurs at a normal frequency. The effects of immunity can be detected over large regions of the bacterial chromosome as judged from the observation that a Tn7 element that is >190 kb from a potential target site decreases the frequency of insertion to that site[47].

Target immunity stems from an interaction between TnsC and the transposase protein TnsB[6,19,48]. The recombination machinery senses the presence of a pre-existing copy of the transposon in the target DNA by the presence of TnsB binding sites found in the ends of the transposon; only the TnsB binding sites (not necessarily the full transposon ends) are required to impose immunity[19]. Immunity acts by providing a high local concentration of TnsB on a target DNA that contains TnsB binding sites, thereby discouraging the binding of TnsC to the immune target DNA. The interaction of TnsB with TnsC does not inactivate TnsC, but instead leads to a redistribution of the protein[48].

What is the purpose of target immunity? This process prevents the formation of DNAs containing two proximal copies of Tn7 that could then be substrates for deleterious events — such as deletion of the sequences between the two Tn7 elements — which could be promoted by homologous recombination. Also, target immunity reduces the frequency at which Tn7 could insert into itself, leading to destruction of the element.

Targeting and transposition. Why should Tn7 use such an elaborate mechanism, involving a direct interaction of TnsC with the target DNA, to activate the transposase? In addition to the efficient recognition of *attTn7*, this strategy of using two target DNA-binding proteins, TnsD and TnsC, allows Tn7 to spatially separate the functions of targeting and transposase activity. In other words, *attTn7* can be recognized by sequences in *glmS*, an essential gene, but Tn7 insertions are made downstream of the gene to avoid deleterious effects to the host. Moreover, the use of a target DNA structure formed by TnsD to recruit TnsC ensures that an appropriate target DNA is actually present in the transposition complex before any transposition occurs. So, recognition of the target site and formation of a complex, which includes the transposase bound to the ends of Tn7, and TnsC and TnsD bound to *attTn7*, controls the initiation of transposition. Such stringent regulation of the TnsABC+D reaction not only allows high-frequency Tn7 insertions into *attTn7*, but also avoids wasteful excision and deleterious insertions of Tn7 in the rest of the genome.

TnsABC+D insertion into the attachment site could have more regulatory cues from the host. Two host factors — the ribosomal protein L29 and the acyl carrier protein (ACP)[41] — aid TnsD in its interactions with *attTn7* and were identified biochemically by their ability to promote TnsD binding to *attTn7*. These proteins were subsequently shown to stimulate TnsD-mediated transposition *in vitro*; and L29, which is not essential for cell viability, is required for high-frequency TnsD-mediated transposition *in vivo*. L29 is a component of the 50S ribosomal subunit, and ACP has an essential function in fatty-acid biosynthesis. L29 and ACP might help Tn7 to regulate not only where insertion occurs, but also when transposition occurs in the host growth cycle by being sensitive to the metabolism of the host. Host factors might also have a role in the TnsABC+E transposition pathway.

Given the role of the target DNA in recruiting TnsC in the TnsD pathway, it will be interesting to see whether TnsE also forms distortions when binding certain DNA structures that might attract TnsC. DNA distortions could be a common link between the TnsABC+E and TnsABC+D targeting pathways, and might explain the lack of amino-acid similarity between TnsD and TnsE. For example, it could be that neither protein contains a specialized TnsC-interacting domain, which might have been expected in both proteins if protein–protein interactions had a strong role in attracting the core machinery.

The regulatory decision as to whether to use the TnsD or TnsE pathway is primarily controlled by the target DNA. If the *attTn7* site is available, the TnsD

Box 2 | **ATP-dependent switches in transposition**

Nucleotide-binding proteins are used as molecular switches in diverse settings[49]. Transposons such as Tn7 and the bacteriophage Mu — which uses transposition to replicate its genome — seem to use Ras-like switches to regulate transposition[50–53]. The Ras protein is the prototypic molecular switch. GTP-bound Ras transduces cell growth and differentiation signals from the cell surface to the nucleus[54]. GTP hydrolysis alters the conformation of Ras, converting it to its GDP-bound inactive state[55]. By exchanging GTP for GDP, Ras will switch back into its active conformation. Other proteins modulate Ras activity by either stimulating GTP hydrolysis (GTPase-activating proteins; GAPs) or facilitating the release of GDP (guanine-nucleotide-exchange factors; GEFs).

TnsC will only bind DNA — a requirement for transposition — when in its ATP-bound state, which is reminiscent of the Ras molecular switch. TnsC is the key player in conveying an 'on' signal to the transposase when TnsD or TnsE have engaged the appropriate target DNA. Alternatively, TnsC is also sensitive to 'off' signals that are imposed by the TnsB protein in the case of immunity (BOX 1). An intriguing possibility is that the target-selecting proteins, TnsD and TnsE, act like GEFs to switch TnsC into an active, ATP-bound conformation, whereas TnsB could provoke TnsC to hydrolyse ATP to switch it into its inactive, ADP-bound state, as is found for the GAPs. Interestingly, TnsC mutants that can activate the transposase in the absence of the target-selecting proteins are also altered in their ATPase activity or in their dependence on ATP for transposition[23].

pathway is used at a high frequency, but when *attTn7* is not available, TnsD-mediated transposition drops to TnsE levels found in cells without conjugal plasmids. TnsE transposition is stimulated in the presence of its preferred targets, actively conjugating DNA and processes associated with DSB repair. An interesting and untested question remains as to which pathway would be used when both the preferred TnsD and TnsE targets are present.

TnsC: central processor of target information

Tn7 can evaluate many aspects of a potential target DNA, such as is there a conjugating plasmid present? Is there an *attTn7* site present? Do these potential targets already contain a copy of Tn7 (BOX 1)? How is this target information coordinated and a molecular decision reached about whether transposition will or will not occur?

The TnsAB transposase mediates the recognition of the transposon ends and the chemical steps in recombination whereas the TnsD and TnsE proteins survey the environment in the cell for specialized targets of Tn7. The transposition regulator TnsC, which coordinates target recognition with transposase activation, is the protein responsible for the sophisticated surveillance by Tn7 of potential target DNAs.

How is the ability of TnsC to control transposition sensitive to both positive and negative signals in a target DNA? TnsC is an ATPase and an ATP-dependent DNA-binding protein, activities that are central to its role as a modulator of transposition[23,42]. TnsC will bind DNA — a required step before recruiting the TnsAB transposase — only when bound to ATP. However, the simple addition of ATP to TnsABC is not sufficient to allow TnsC to provoke transposition, as transposase activation requires the presence of the correct target protein and DNA — for example, TnsD and *attTn7* (REF. 22). We propose that an active TnsC–TnsD–*attTn7* complex that can interact with the transposase only forms when TnsC is in the ATP-bound state. Therefore, TnsD–*attTn7* can be viewed as a modulator of the ATP state of TnsC. The TnsD and TnsE proteins might modulate the ATP-bound state of TnsC to keep TnsC on only the appropriate target DNAs (BOX 2).

ATP hydrolysis also has an important function in the discriminating target-site selection of Tn7. In the presence of non-hydrolysable analogues of ATP, there is unregulated constitutive Tn7 transposition into all DNAs, even those lacking *attTn7* or already containing a copy of Tn7. We propose that target immunity results from ATP hydrolysis by TnsC, which is promoted by the presence of TnsB around the target DNA (BOX 2).

The purpose of multiple targeting pathways

Both targeting pathways of Tn7 can be expected to facilitate the distribution of Tn7 into new and diverse hosts, given the conservation of *attTn7* and the broad host range of conjugal plasmids. The advantage of having several targeting pathways is emphasized by the finding that single amino-acid changes in TnsC can allow Tn7 transposition in the absence of either TnsD or TnsE[24]. So, although a derivative of Tn7 can be made that is less discriminating in its target selection, the ability to choose among multiple target sites has been preserved in Tn7.

Whereas Tn7 was isolated in *E. coli*, a very close relative of Tn7 was identified in a very different bacterium, the extremophile *Thiobacillus ferrooxidans*, which grows optimally at pH 2 and is used in the recovery of metals[43]. The Tn7 homologue in *T. ferrooxidans*, called Tn5468, resides in an attachment site downstream of the *glmS* gene and the *tnsABCD* genes occur in the same gene order as in Tn7 (FIG. 2). An interesting question is what path a common Tn7/Tn5468 relative took to span the different ecological niches in which *E. coli* and *T. ferrooxidans* reside. In addition to strong TnsABCD homologies, a weak TnsE homologue is also found in *T. ferrooxidans*. It should be interesting to see whether this protein acts like TnsE to target conjugal plasmids, or whether it might represent a new targeting pathway. It will be of further interest to look for other Tn7 homologues in the growing number of bacterial genomes that are being sequenced at present.

Tn7 behaves like a 'smart' element in its highly evolved target-site selection that facilitates its dispersal, while minimizing potentially harmful random insertions. Future research with Tn7 should provide further insight into how multiprotein machines are assembled on DNA in replication, repair and recombination. Given the ability of Tn7 to be sensitive to host metabolism in the TnsABC+D pathway, and the ability of TnsABC+E transposition to target structures associated with some forms of DNA metabolism, the study of Tn7 promises to continue to provide insight into basic cell functions.

1. Lander, E. S. *et al.* Initial sequencing and analysis of the human genome. *Nature* **409**, 860–921 (2001).
2. Venter, J. C. *et al.* The sequence of the human genome. *Science* **291**, 1304–1351 (2001).
3. Craig, N. L. Target site selection in transposition. *Annu. Rev. Biochem.* **66**, 437–474 (1997).
4. Craig, N. L. Transposon Tn7. *Curr. Top. Microbiol. Immunol.* **204**, 27–48 (1996).
5. Hall, R. M., Brookes, D. E. & Stokes, H. W. Site-specific insertion of genes into integrons: role of the 59-base element and determination of the recombination cross-over point. *Mol. Microbiol.* **5**, 1941–1959 (1991).
6. Hauer, B. & Shapiro, J. A. Control of Tn7 transposition. *Mol. Gen. Genet.* **194**, 149–158 (1984).
7. Rogers, M., Ekaterinaki, N., Nimmo, E. & Sherratt, D. Analysis of Tn7 transposition. *Mol. Gen. Genet.* **205**, 550–556 (1986).
8. Waddell, C. S. & Craig, N. L. Tn7 transposition: two transposition pathways directed by five Tn7-encoded genes. *Genes Dev.* **2**, 137–149 (1988).
9. Sarnovsky, R., May, E. W. & Craig, N. L. 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 (1996).
Together with reference 20, this study provided the first evidence that the Tn7 transposase was heteromeric.
10. Biery, M., Lopata, M. & Craig, N. L. A minimal system for Tn7 transposition: the transposon-encoded proteins TnsA and TnsB can execute DNA breakage and joining reactions that generate circularized Tn7 species. *J. Mol. Biol.* **297**, 25–37 (2000).
11. Lu, F. & Craig, N. L. Isolation and characterization of Tn7 transposase gain-of-function mutants: a model for transposase activation. *EMBO J.* **19**, 3446–3457 (2000).
12. Mizuuchi, K. Transpositional recombination: mechanistic insights from studies of Mu and other elements. *Annu. Rev. Biochem.* **61**, 1011–1051 (1992).
13. Haren, L., Ton-Hoang, B. & Chandler, M. Integrating DNA: transposases and retroviral integrases. *Annu. Rev. Microbiol.* **53**, 245–281 (1999).
14. Dyda, F. *et al.* Crystal structure of the catalytic domain of HIV-1 integrase: similarity to other polynucleotidyl transferases. *Science* **266**, 1981–1986 (1994).
15. Rice, P. & Mizuuchi, K. Structure of the bacteriophage Mu transposase core: a common structural motif for DNA transposition and retroviral integration. *Cell* **82**, 209–220 (1995).
16. Davies, D. R., Braam, L. M., Reznikoff, W. S. & Rayment, I. The three-dimensional structure of a Tn5 transposase-related protein determined to 2.9-Å resolution. *J. Biol. Chem.* **274**, 11904–11913 (1999).
17. Reznikoff, W. S. The Tn5 transposon. *Annu. Rev. Microbiol.* **47**, 945–963 (1993).
18. Kleckner, N., Chalmers, R., Kwon, D., Sakai, J. & Bolland, S. Tn10 and IS10 transposition and chromosome rearrangements: mechanism and regulation *in vivo* and *in vitro*. *Curr. Top. Microbiol. Immunol.* **204**, 49–82 (1996).
19. Arciszewska, L. K., Drake, D. & Craig, N. L. Transposon Tn7 *cis*-acting sequences in transposition and transposition immunity. *J. Mol. Biol.* **207**, 35–52 (1989).
20. May, E. W. & Craig, N. L. Switching from cut-and-paste to replicative Tn7 transposition. *Science* **272**, 401–404 (1996).
21. Hickman, A. B. *et al.* Unexpected structural diversity in DNA recombination: the restriction endonuclease connection. *Mol. Cell* **5**, 1025–1034 (2000).
The structure of TnsA shows that one of the proteins of the heteromeric Tn7 transposase is actually a restriction enzyme that is used for transposition.
22. Bainton, R. J., Kubo, K. M., Feng, J.-N. & Craig, N. L. Tn7 transposition: target DNA recognition is mediated by multiple Tn7-encoded proteins in a purified *in vitro* system. *Cell* **72**, 931–943 (1993).
This study provides the first description of the complete TnsABC+D *in vitro* transposition reaction from purified proteins.
23. Stellwagen, A. E. Analysis of gain of function mutants of an ATP-dependent regulator of Tn7 transposition. *J. Mol. Biol.* **305**, 633–642 (2001).
24. Stellwagen, A. & Craig, N. L. Gain-of-function mutations in TnsC, an ATP-dependent transposition protein which activates the bacterial transposon Tn7. *Genetics* **145**, 573–585 (1997).
25. Wolkow, C. A., DeBoy, R. T. & Craig, N. L. Conjugating plasmids are preferred targets for Tn7. *Genes Dev.* **10**, 2145–2157 (1996).
26. Barth, P. & Grinter, N. Map of plasmid RP4 derived by insertion of transposon C. *J. Mol. Biol.* **113**, 455–474 (1977).
27. Peters, J. E. & Craig, N. L. Tn7 recognizes target structures associated with DNA replication using the DNA binding protein TnsE. *Genes Dev.* **15**, 737–747 (2001).
Together with reference 28, this study showed that TnsE-mediated transposition recognizes structures associated with DNA replication.
28. Peters, J. E. & Craig, N. L. Tn7 transposes proximal to DNA double-strand breaks and into regions where chromosomal DNA replication terminates. *Mol. Cell* **6**, 573–582 (2000).
29. Cox, M. M. *et al.* The importance of repairing stalled replication forks. *Nature* **404**, 37–41 (2000).
30. Gay, N. J., Tybulewica, V. L. J. & Walker, J. E. Insertion of transposon Tn7 into the *Escherichia coli glmS* transcriptional terminator. *Biochem. J.* **234**, 111–117 (1986).
31. Lichtenstein, C. & Brenner, S. Unique insertion site of Tn7 in *E. coli* chromosome. *Nature* **297**, 601–603 (1982).
32. Walker, J. E., Gay, N. J., Saraste, M. & Eberle, A. N. DNA sequence around the *Escherichia coli unc* operon. Completion of the sequence of a 17 kilobase segment containing *asnA, oriC, unc, glmS* and *phoS*. *Biochem. J.* **224**, 799–815 (1984).
33. Gringauz, E., Orle, K., Orle, A., Waddell, C. S. & Craig, N. L. Recognition of *Escherichia coli attTn7* by transposon Tn7: lack of specific sequence requirements at the point of Tn7 insertion. *J. Bacteriol.* **170**, 2832–2840 (1988).
34. Cleaver, S. H. & Wickstrom, E. Transposon Tn7 gene insertion into an evolutionarily conserved human homolog of *Escherichia coli attTn7*. *Gene* **254**, 37–44 (2000).
35. Bainton, R., Gamas, P. & Craig, N. L. Tn7 transposition *in vitro* proceeds through an excised transposon intermediate generated by staggered breaks in DNA. *Cell* **65**, 805–816 (1991).
36. Rao, J. E., Miller, P. S. & Craig, N. L. Recognition of triple-helical DNA structures by transposon Tn7. *Proc. Natl Acad. Sci. USA* **97**, 3936–3941 (2000).
37. Kuduvalli, P., Rao, J. & Craig, N. L. Target DNA structure plays a critical role in Tn7 transposition. *EMBO J.* **20**, 1–10 (2001).
Together with reference 36, this study shows that an induced DNA structure is of key importance in recruiting TnsC and the core transposition machinery.
38. Pryciak, P. M. & Varmus, H. E. Nucleosomes, DNA-binding proteins, and DNA sequence modulate retroviral integration target site selection. *Cell* **69**, 769–780 (1992).
39. Katz, R. A., Gravuer, K. & Skalka, A. M. A preferred target DNA structure for retroviral integrase *in vitro*. *J. Biol. Chem.* **273**, 24190–24195 (1998).
40. Rao, J. E. & Craig, N. L. Selective recognition of pyrimidine motif triplexes by a protein encoded by the bacterial transposon Tn7. *J. Mol. Biol.* **307**, 1161–1170 (2001).
41. Sharpe, P. & Craig, N. L. Host proteins can stimulate Tn7 transposition: a novel role for the ribosomal protein L29 and the acyl carrier protein. *EMBO J.* **17**, 5822–5831 (1998).
42. Gamas, P. & Craig, N. L. Purification and characterization of TnsC, a Tn7 transposition protein that binds ATP and DNA. *Nucleic Acids Res.* **20**, 2525–2532 (1992).
43. Oppon, J. C., Sarnovsky, R. J., Craig, N. L. & Rawlings, D. E. A Tn7-like transposon is present in the *glmUS* region of the obligately hemoautolithotrophic bacterium *Thiobacillus ferrooxidans*. *J. Bacteriol.* **180**, 3007–3012 (1998).
44. Lee, C.-H., Bhagwhat, A. & Heffron, F. Identification of a transposon Tn3 sequence required for transposition immunity. *Proc. Natl Acad. Sci. USA* **80**, 6765–6769 (1983).
45. Adzuma, K. & Mizuuchi, K. Interaction of proteins located at a distance along DNA: mechanism of target immunity in the Mu DNA strand-transfer reaction. *Cell* **57**, 41–47 (1989).
46. Reyes, I., Beyou, A., Mignotte-Vieux, C. & Richaud, F. Mini-Mu transduction: *cis*-inhibition of the insertion of Mud transposons. *Plasmid* **18**, 183–192 (1987).
47. DeBoy, R. & Craig, N. L. Tn7 transposition as a probe of *cis* interactions between widely separated (190 kilobases apart) DNA sites in the *Escherichia coli* chromosome. *J. Bacteriol.* **178**, 6184–6191 (1996).
48. Stellwagen, A. & Craig, N. L. Avoiding self: two Tn7-encoded proteins mediate target immunity in Tn7 transposition. *EMBO J.* **16**, 6823–6834 (1997).
49. Bourne, H. R., Sanders, D. A. & McCormick, F. The GTPase superfamily: a conserved switch for diverse cell functions. *Nature* **348**, 125–132 (1990).
50. Maxwell, A., Craigie, R. & Mizuuchi, K. B protein of bacteriophage Mu is an ATPase that preferentially stimulates intermolecular DNA strand transfer. *Proc. Natl Acad. Sci. USA* **84**, 699–703 (1987).
51. Baker, T. A., Mizuuchi, M. & Mizuuchi, K. MuB protein allosterically activates strand transfer by the transposase of phage Mu. *Cell* **65**, 1003–1013 (1991).
52. Surette, M. G. & Chaconas, G. Stimulation of the Mu DNA strand cleavage and intramolecular strand transfer reactions by the MuB protein is independent of stable binding of the MuB protein to DNA. *J. Biol. Chem.* **266**, 17306–17313 (1991).
53. Stellwagen, A. & Craig, N. L. Mobile DNA elements: controlling transposition with ATP-dependent molecular switches. *Trends Biochem. Sci.* **23**, 486–490 (1998).
54. Lowy, D. R. & Willumsen, B. M. Function and regulation of Ras. *Annu. Rev. Biochem.* **62**, 851–891 (1993).
55. Sprang, S. R. G protein mechanisms: insights from structural analysis. *Annu. Rev. Biochem.* **66**, 639–678 (1997).

Acnowledgements

We thank Z. Skelding and P. Kudavalli for comments and suggestions on the manuscript. The Howard Hughes Medical Institute, the National Institutes of Health and the National Science Foundation supported this work.

Online links

DATABASES

The following terms in this article are linked online to:

GenBank: http://www.ncbi.nlm.nih.gov/Genbank/index.html
attn7 | Tn7-L | Tn7-R

Swiss-Prot: http://www.expasy.ch/
glmS | L29 | Tn5468 | TnsA | TnsB | TnsC | TnsD | TnsE

Access to this interactive links box is free online