Copy-out–Paste-in Transposition of IS911: A Major Transposition Pathway

MICHAEL CHANDLER, OLIVIER FAYET, PHILIPPE ROUSSEAU, BAO TON HOANG, and GUY DUVAL-VALENTIN
Laboratoire de Microbiologie et Génétique Moléculaires, CNRS, Toulouse, France

ABSTRACT IS911 has provided a powerful model for studying the transposition of members of a large class of transposable element: the IS3 family of bacterial Insertion Sequences (IS). These transpose by a Copy-out–Paste-in mechanism in which a double-strand IS circle transposition intermediate is generated from the donor site by replication and proceeds to integrate into a suitable double strand DNA target. This is perhaps one of the most common transposition mechanisms known to date. Copy-out–Paste-in transposition has been adopted by members of at least eight large IS families. This chapter details the different steps of the Copy-out–Paste-in mechanism involved in IS911 transposition. At a more biological level it also describes various aspects of regulation of the transposition process. These include transposase production by programmed translational frameshifting, transposase expression from the circular intermediate using a specialized promoter assembled at the circle junction and binding of the nascent transposase while it remains attached to the ribosome during translation (co-translational binding). This co-translational binding of the transposase to neighboring IS ends provides an explanation for the longstanding observation that transposases show a cis-preference for their activities.

INTRODUCTION

The bacterial insertion sequence, IS911, is a member of the large IS3 family. It transposes using a mechanism known as Copy-out–Paste-in. This is a major transposition pathway as judged by the number of transposable elements that use it. This pathway has not only been demonstrated to apply to various other members of the IS3 insertion sequence family, IS2 (1), IS3 (2), and IS150 (2), but has also been adopted by members of at least seven other large IS families: IS1, IS21, IS30, IS256, IS110, ISLreo2, ISL3, and their derivatives (see Siguier et al., this volume).

The various steps in Copy-out–Paste-in transposition are described in greater detail later in this review (Transposition Pathway section, and Figure 5). However, the essential features of this common pathway can be summarized as follows. It is essentially asymmetric: cleavage of one strand at one IS end generates a 3′OH group; this 3′OH is then used to attack the other IS end forming a single-strand bridge between both ends; the IS is replicated in situ from a 3′OH in the flanking donor DNA, which was generated by formation of the bridge; replication generates a double-strand circular IS copy with abutted ends (an active junction) and presumably regenerates the original donor molecule, which retains an IS copy. The closed circular transposon circle with abutted IS ends is then poised to integrate into a double-strand target DNA by single-strand cleavage at each end in the junction generating two 3′OH, which are then used to attack the two target strands.
In this chapter, we first review the origin and distribution of IS911 itself (with supplementary information from other members of the IS3 family where appropriate), its organization, the functional organization of its transposase, the regulation of expression and activity, the composition of the IS911 transpososomes, nucleoprotein complexes involved in the Copy-out step (containing both transposon ends and transposase) and the Paste-in step (containing target DNA, transposon ends, and transposase) and, finally, a detailed description of the transposition pathway with some of the supporting data.

ISOLATION AND DISTRIBUTION

Insertion sequence IS911 was isolated from a Shigella dysenteriae phage λ lysogen by spontaneous insertion into the phage cI repressor gene. This resulted in a virulent phage whose lytic phenotype could be suppressed in an Escherichia coli strain able to produce the CI repressor protein (4). The element is present in multiple copies in the original host strain and in type strains of other Shigella species. Two vestigial copies, both interrupted by a copy of IS30, were also detected in the chromosome of E. coli K12 (5) and proved able to form transposition intermediates when supplied with IS911 transposase (6). Entire or truncated IS911 copies have also been identified in several E. coli virulence plasmids (7, 8), in pathogenicity islands of uropathogenic E. coli (9), in various other clinical isolates of E. coli, and in a large number of well-known and less well-known enterobacteria such as Escherichia fergusonii, Chromobacter, Dickeya, Erwinia, Klebsiella, Pantoea, Shimwella, and Yersinia.

IS911 is a member of the very widely distributed and one of the largest IS families: IS3 (see Sigier et al., this volume). Although this review is centered on studies of IS911 as a model system for studying the IS3 family, important insights and some differences have been revealed from studies of other family members such as IS2, IS3, and IS150. These will be described where appropriate below.

GENERAL ORGANIZATION

IS911 (Figure 1A) is 1,250 bp long, bordered by imperfect 36-bp terminal inverted repeats (IRL and IRR) and generates 3-bp (sometimes 4-bp) target duplications on insertion (4). Like those of most family members and members of several other transposable element families, the IS911 IRs terminate in a 5′-CA-3′ dinucleotide (Figure 2A). IS2 carries a natural mutation at one end which affects the activity of this end. These bases are important for the chemistry of transposition. Also like most other family members, IS911 carries two consecutive and partially overlapping open reading frames, orfA and orfB under control of a weak promoter, pIRL, partially located in IRL (Figure 2A and B). The 5′ end of orfB overlaps the 3′ end of orfA and occurs in reading phase −1 relative to orfA (Figure 1A). Complex inverted repeat sequences (Figure 2C) are located between co-ordinates 19 and 73 and include the −35 and −10 hexamers of pIRL, the transcription start site and the ribosome-binding site for OrfA. This is thought to play a role at the mRNA level in preventing excess transposase expression resulting from external transcription. The full secondary structure would be present in transcripts initiated outside the IS, so sequestering the translation initiation signals, but only the 3′ part would be present if transcription initiates at pIRL. In this case the translation initiation signals would be exposed. Initial studies (M.F. Prère and O. Fayet, unpublished) have shown that translation from the longer transcript is very low but that deletion of its 5′ end, to “liberate” the ribosome-binding site, indeed results in a significant increase in translation (Figure 2C). In the related IS2 element, a similar sequence appears to function as a DNA binding site for the OrfA protein, which represses promoter activity but further studies are necessary to confirm this (10).

In common with many other ISs of both the IS3 and other families (e.g., IS21 (11), IS30 (12), IS110 (13, 14)) the IS911 IRR carries an outward-directed −35 promoter hexamer whereas IRL carries an inward-directed −10 promoter component (Figure 2A). These are assembled into a strong promoter, pJunc, (Figure 2B), which serves to express high levels of the transposition proteins in one of its key transposition intermediates, an excised transposon circle (see Transposition pathway section below). It should be noted that transcription initiation from pJunc, like that from impinging transcription, would also produce an RNA which could sequester the translation initiation signals but in a shorter and less stable stem loop structure (Figure 2C).

TRANPOSITION PROTEINS: DNA SEQUENCE RECOGNITION, MULTIMERIZATION AND CATALYTIC DOMAINS

IS911 expresses two major proteins (Figure 1B): OrfA, and the transposase, OrfAB (15). There is some evidence that OrfB may also be produced at low levels (15–17).
OrfA has a predicted molecular weight of 11.5 kDa. The OrfA sequence carries an α helix-turn-α helix (HTH) motif involved in sequence-specific binding to the terminal IRs of the transposase of the element OrfAB (see below) and a C-terminal leucine zipper (LZ) motif involved in protein multimerization (4, 15, 18). Most IS3 family members exhibit a similarly placed HTH signature (4, 19), whereas the LZ motif was identified in IS2, IS150, and IS3 and appears to be conserved in the majority of known members (18).

OrfB is 299 residues long and has a predicted molecular weight of 34.6 kDa. Its TAA termination codon lies just within IRR and may be significant in regulation. The OrfB initiation codon is AUU and consequently initiation occurs only at low levels (15, 17) and is modulated by the level of initiation factor IF3 (16). It is possible that OrfB plays no direct role in IS911 transposition chemistry but that its translation signals serve to modulate the programmed translational frameshifting required to generate a single transposase protein, OrfAB, from the two reading frames orfA and orfB (see Frameshifting: transposase OrfAB as a fusion protein section below). Expression of IS911 genes leads to about five-fold less OrfB than OrfAB, mainly because initiation on the AUU codon is dampened by a downstream stem-loop structure also involved in the frameshifting event (16). OrfB has been observed in the case of IS3 (20) (MF Prère and O. Fayet, unpublished), of IS150 (21) and IS3411/IS629 (22, 23). For IS150, the OrfB initiation codon is out of phase and expression requires a −1 frameshift after initiation (see Frameshifting: transposase OrfAB as a fusion protein section below). OrfB may be implicated in modulating transposition of IS3 and IS3411/IS629 in a way that has yet to be mechanistically defined. Sequence analysis suggests that it is probably synthesized by about 34% of the IS3 family members through translational coupling; the stop codon of orfA is found to overlap with a potential start codon for orfB (e.g., AUGA or GUGA) in 134 out of 399 ISs. In IS3 this leads to synthesis of an equal amount of OrfB and OrfAB (20). OrfB protein synthesis is therefore not a universal feature within the IS3 family, as exemplified by its absence in IS2 expression (24). The OrfB amino acid sequence shares significant similarities with retroviral integrases. In particular it was comparison of this sequence that contributed to defining the highly conserved amino acid triad DDE common to all IS3 family members and to many of this type of phosphoryltransferase enzymes (25, 26). This constitutes part of the active site (for reviews see references 27–33). OrfB carries neither the HTH nor the LZ motif.

Although no structural information is available from crystallography, the roles of the HTH and LZ motifs have been investigated in vivo and in vitro.

The conserved N-terminal HTH motif is related to the LysR family of bacterial transcription factors and has a highly conserved tryptophan residue similar to that of certain homeodomain protein HTH motifs. Site-directed mutagenesis was used to probe the function of this motif using a series of in vivo and in vitro tests and demonstrated that the HTH domain is indeed important in directing transposase to bind IS911 IR (34). The N-terminal helices of the related IS2 transposase have also been shown to be involved in IR binding (35).

The LZ motif is composed of four heptameric units (Figure 1B) with a predicted coiled coil structure including potential buried inter-subunit hydrogen bonds across the dimer interface, to maintain the zipper in a dimeric state, and correctly placed residues with opposite charges potentially able to form characteristic inter-subunit salt-bridges to stabilize the dimeric structure (18). Mutation of specific critical residues in the OrfAB LZ, which lead to reduced levels of multimerization, resulted in reduction in the level of transposition intermediates in vivo and in vitro (18) (see Transposition pathway section below). Immunoprecipitation from extracts of cells expressing various C-terminal truncated OrfAB derivatives with a wild-type LZ and a C-terminal HA tag was used to assess multimerization. This demonstrated that OrfAB and OrfA could form both homo-multimers as well as mixed OrfAB–OrfA multimers (18, 36). The critical mutations reduced or prevented multimer (dimer) formation.

These studies also showed that a poorly defined region, M, located between residues 109 and 135 (Figure 1B) and components in the catalytic domain of OrfAB (18) are also involved in its multimerization.

It is interesting that OrfAB and OrfA share three of their four heptads (Figure 1B). The last heptad of each differs in sequence due to the translational frameshift that occurs within the heptad in expression of OrfAB (see Frameshifting: transposase OrfAB as a fusion protein section below). This presumably results in different strengths of monomer–monomer interactions in the case of homo- and hetero-multimers, which may be involved in regulation of transposition.

For experimental purposes, production of OrfAB without necessitating a translational frameshift is obtained by introduction of a single additional base pair within the frameshift region (see Frameshifting: transposase OrfAB as a fusion protein section below) which artificially fuses the orfA and orfB frames and eliminates
A)  

1250 bp  

IRL  

0  

orfB  

orfA  

IRR  

PIRL

B)  

OrfA  

11.5 kDa  

(100 aa)  

149  

OrfAB  

44 kDa  

(382 aa)  

HTH  

LZ  

M  

DDE

C)  

OrfB start  

OrfA stop  

SDB  

-1nt

UGGAGAAUGAAAUAUUAAAAAAGCUACC  

* *  

A AAA AAG AAA AAAAAG G

0 0 -1 -1  

UUU|||UUU|||  

K K K K

115 AA 149
OrfA production (15). It was initially difficult to construct this mutant in the context of an entire IS911 (i.e., with the two flanking IR). This is because constitutive expression of the OrfAB protein in cis results in exceptionally high transposition activity of the closely neighboring IS ends. More recently this has been accomplished using a much longer artificial IS in which the ends are located at some distance from the transposase gene. Even in these conditions, the transposition frequency was exceptionally high and the transposon was unstable (37). A similar mutant in IS3 results in a high frequency of adjacent deletions (20).

**FRAMESHIFTING: TRANPOSASE OrfAB AS A FUSION PROTEIN**

OrfAB of IS911 is assembled from orfA and orfB by a programmed −1 translational frameshift occurring near the 3′ end of orfA, as first demonstrated for the related IS150 (21). The transframe protein combines the HTH motif of orfA, an LZ motif and the DD(35)E catalytic domain of orfB (18). OrfAB (382 amino acids) shares its 86 N-terminal amino acids with OrfA (100 amino acids) and its 296 C-terminal amino acids with OrfB (299 amino acids). Rephasing of the ribosome to generate OrfAB occurs on a group of “slippery” lysine codons with a frequency of about 15% (measured using systems driven by two different promoters; T7p_{tac} and p_{tac}). OrfA is therefore normally expressed at significantly higher levels than OrfAB. Frameshifting permits the combination of different functional protein domains. The frameshifting in IS911 is similar to that used in some retroviruses to generate the pol-gag “polyprotein” (38) and in the dnx gene of E. coli to synthesize γ, the subunit of DNA polymerase III (39). The relevant sequences involved in frameshifting in IS911 are shown in **Figure 1C**. The group of slippery lysine codons is AAA AAG and is directly preceded by the AUA OrfB initiation codon. As E. coli does not encode a trNA_Lys with a 3′UUC anticodon for AAG, both lysine codons are decoded by the same trNA_Lys with a 3′UUU anticodon (Figure 1C). Its pairing has been shown to be weaker when there is a G at the wobble position (40), probably because modifications of a base, U34, increase the rigidity of the anti-codon (41). The presence of an upstream RBS (GGAG sequence) and a downstream secondary structure (Y-shaped stem–loop) has been shown experimentally to stimulate rephasing of the ribosome in the −1 direction. What drives frameshifting is probably the thermodynamically favorable re-pairing of the two trNA_Lys, from codons AAA-AAG to codons AAA-AAA (39, 42). The stimulators likely have a mechanical effect bringing back in register the ribosome and the mRNA after tRNA slippage. Different groups

**FIGURE 1.** Organization of IS911. (A) Genetic organization. The 1,250-bp IS911 is shown as a box. The boxes at each end represent the left (IRL) and right (IRR) terminal inverted repeats. The two open reading frames, orfA (blue) and orfB (green) are positioned in relative reading phases 0 and −1, respectively, as indicated. The indigenous promoter, p_{IRL}, is shown. The region of overlap between orfA and orfB, which includes the frameshifting signals to produce OrfAB, lies within IS911 coordinates 300 and 400. The precise point at which the frameshift occurs, within the last heptad of the LZ, is indicated by the vertical dotted line. (B) Structure function map of OrfAB and OrfA. HTH, a potential helix-turn-helix motif; LZ, a leucine zipper motif involved in homo- and hetero-multimerization of OrfAB and OrfA. Programmed translational frameshifting that fuses OrfA and OrfB to generate the transposase OrfAB occurs within the fourth heptad. The LZ of OrfA and OrfAB therefore differ in their fourth heptad. A second region, M, necessary for multimerization of OrfAB is shown, as is the catalytic core of the enzyme which carries a third multimerization domain. OrfA translation initiates at an AUG, terminates with UAA whereas OrfAB translation terminates within the right IR. The vertical line to the right of M shows the extent of the truncated transposase, OrfAB[1−149] described in the text. (C) Frame-shifting window. The mRNA sequence around the programmed translational frameshifting window is presented. The boxed sequence GGAG is the potential ribosome-binding site located upstream of orfB whose potential translation would be initiated at the boxed AUU codon. A ribosome (not to scale) is shown covering a series of “slippery” codons (AAAAAAG). A downstream secondary structure is also shown with the UAA, OrfA translation termination codon. The ribosome-binding site, slippery codons and secondary structure all contribute to the efficiency of the programmed −1 frameshift. The box at the foot of this figure shows how the anti-codons of two trNA_Lys are thought to undergo re-pairing with their codons in the AAAAAAG motif. doi:10.1128/microbiolspec.MDNA3-0031-2014.f1
of codons have been observed to allow rephasing of the ribosome \cite{43} and, although the most common motif is A6G, different members of the IS3 family carry a variety of these \cite{43,44}.

Two similarly located partially overlapping reading frames are present in IS3, IS150 and IS3411\cite{22} and indeed in most of the presently identified IS3 family members. The transposases, OrfAB, like that of IS911, are fusion products of the two orfs generated by a \(-1\) translational frameshift. For IS3, frameshifting is also stimulated by a presumed H-type pseudoknot structure similar to those generally involved in viral recoding \cite{45}. In IS3411, \(-1\) slippage on a U UUU motif requires a more convoluted form of pseudoknot structure formed.

\begin{figure}
\centering
\includegraphics[width=\textwidth]{figure2}
\caption{Important sequence features of IS911. (A) Organization of the IS911 inverted repeat (IR). The nucleotide sequence of IRL and IRR is boxed. Grey horizontal bars above and below indicate the internal regions protected from DNasel digestion by binding of OrfAB\([1–149]\), a derivative of the 382-amino-acid OrfAB truncated for its catalytic domain. The dotted horizontal grey bar indicates partial protection. The dashes within the sequence indicate mismatches between the left and right ends. The \(-35\) and \(-10\) components of the indigenous promoter \(p_{\text{IRL}}\) (blue boxes) and of \(p_{\text{junc}}\) (green boxes) are shown. The conserved 5’TG tips are highlighted in red. (B) Organization of \(p_{\text{junc}}\). The “junction” promoter assembled on circularization of IS911 is shown as green boxes. The initiating transcript nucleotide (+1 \(p_{\text{junc}}\)), the indigenous \(p_{\text{IRL}}\) (blue boxes) and the initiating transcript nucleotide (+1 \(p_{\text{IRL}}\)) are also shown. The conserved 5’TG tips are highlighted in red. (C) Secondary structure at the left IS911 end. The sequence of the “top” strand of IRL is shown together with the various transcription and translation signals. The symbols below are standard “dot–bracket” notations to indicate potential secondary structures formed with transcripts from top to bottom: from an external promoter, from \(p_{\text{junc}}\) or from \(p_{\text{IRL}}\) respectively. The brackets shown in italic simply permit the reader to identify the apical stem of the secondary structure. doi:10.1128/microbiolspec.MDNA3-0031-2014.f2}
\end{figure}
by the pairing of an apical loop and an internal loop belonging to two hairpins located 65 nucleotides apart on the mRNA (22). Two similarly arranged orfs occur in IS2 and have been shown to encode OrfA and OrfAB equivalents only (10, 24). This organization is observed in most members of the IS3 family but, beside the cases mentioned above, frameshifting has been analyzed in only a few other, less well-characterized, elements (including IS51, IS222, IS600, IS1133, IS1222). The frequency of frameshifting appears to be quite variable from element to element: reported values are 15% for IS911, 50% for IS150, 6% for IS3, and 2% for IS3411 (22). These values may not reflect the in vivo situation because they were not established by direct measurement of the amount of the OrfA and OrfAB proteins synthesized from an intact IS, but after modification of expression signals of the IS genes or after cloning the frameshift signals in a reporter system (15, 20, 21). However, the level of IS911 circle formation measured by qPCR has indeed been shown to depend on frameshifting frequency in vivo (46). IS911 copies from several clinical isolates that proved to contain variations in the frameshift region exhibited various reduced levels of frameshifting. When these were introduced into the model IS911 they resulted in comparable reductions in circle formation. In addition, frameshifting is likely modulated by the physiological state of the host cells and by the environment: for example frameshifting decreases when temperature is raised or when ribosome density on the mRNA is increased (O. Fayet, unpublished).

TEMPERATURE SENSITIVITY OF TRANPOSITION: A REGULATORY MECHANISM?

Formation of the first IS911 transposition intermediate (figure-eight molecules; see Overview of Transposition Pathway section and Figure 5 below) is naturally temperature sensitive (6). Temperature sensitive transposition has been observed for other transposons such as IS1 (47), Tn3 (48), and Tn911 (49). For IS911, very few figure-eight or transposon circle intermediates were observed at 42°C but levels were optimal at 30°C in vivo. Moreover, activity at 30°C was sufficiently high to detect circular copies of a single defective chromosomal IS911 copy when OrfAB was provided in trans (6). Cell extracts containing OrfAB also proved somewhat temperature sensitive in figure-eight formation in vitro. Two partially temperature-resistant point mutants were obtained by selection following nitrous acid mutagenesis and showed increased activity compared with wild-type both in vivo and with OrfAB extracts in vitro. Both were located between the LZ motif and a secondimerization region, M, suggesting that this inter-domain region may be important for correct assembly of the transposition complex. A partial explanation for the temperature sensitivity came from the observation that OrfAB preparations included truncated OrfAB derivatives, in particular a species, OrfAB*, of approximately 16 kDa (around 149 amino acids), which were more prevalent at the higher temperature (50). Reconstruction experiments in which an engineered C-terminal deletion derivative, OrfAB[1–149] (Figure 1B), was supplied at the same time as full-length OrfAB, demonstrated a strong inhibition of transposition in vivo and in vitro. Partially temperature-resistant mutants were selected and found to have significantly reduced levels of truncated OrfAB derivatives, confirming this hypothesis (50).

Although this behavior might reveal a mechanism to regulate IS911 transposition, it should be noted that all these studies involved overproduction of OrfAB. It seems possible that the high OrfAB levels themselves, either from the constitutive transframe protein or by high expression from the natural orf configuration, may lead to production of misfolded proteins, which are subject to cleavage. Further studies are essential to clarify this.

CO-TRANSLATIONAL BINDING

Certain prokaryotic IS transposases show a strong preference for acting on the element from which they are expressed rather than on other copies of the same element in the cell. This phenomenon of “cis” preference presumably serves to prevent general activation of several identical IS copies by any “accidental” (stochastic) transposase expression from a single IS. Several different IS such as IS1, IS10, IS50, IS903, and IS911 (see reference 51 and references therein) exhibit this regulatory phenotype but “cis” preference may be the result of a combination of diverse mechanisms. Hence the Lon protease enhances “cis” preference of the IS903 transposase (52) and, for IS10, it is influenced by translation levels, transposase mRNA half-life and translation efficiency (53, 54).

Another mechanism, co-translational binding based on tight coupling between prokaryotic transcription and translation, was proposed to explain the inability to complement a transposase mutant of IS903 (55) and, more specifically, for Tn5 (56). As its name indicates, co-translational binding would involve transposase binding
Some full-length IS transposases bind weakly to their cognate IR but the isolated DNA-binding domain can bind more strongly. This has been observed for transposases of several elements including IS1 (57) and IS30 (58, 59) and has also been observed for that of IS911. Early studies using band shift assays demonstrated that full-length OrfAB binds the IRs only weakly and that OrfA binding was even lower or undetectable (36, 60). However, a truncated version of OrfAB, OrfAB[1–149], which is amputated for the C-terminal catalytic domain bound both ends avidly (18) (see also Temperature sensitivity of transposition section above). It is important to note that this implies that, in many in vitro systems, the majority of transposase is therefore likely to be inactive or only partially active because it would not bind stably to its substrate. The observations suggest that the C-terminal domain inhibits specific binding by the sequence-specific N-terminal DNA-binding domain, possibly by steric masking (Figure 3). This idea is consistent with the observation that IS10 transposase activity is increased by partial denaturation (for example by treatment with low alcohol concentrations (61)). It is also consistent with the observation that the OrfAB protein of IS2 can bind the IS2 IRs when it carries a large GFP tag (35, 62).

One biological explanation for cis preference is that the nascent N-terminal domain might fold before completion of translation of the C-terminal domain and the nascent protein could initiate binding directly to the closest IS end. Once bound, it would no longer be sensitive to masking by the C-terminal domain. If binding fails to occur after translation of the N-terminal DNA-binding domain, continuing translation and folding of the C-terminal domain would then sterically mask the DNA-binding domain, resulting in an inactive protein. This implies that binding necessary for subsequent catalysis would occur only transitorily early in translation (Figure 3).

In vivo studies (37) demonstrated that OrfAB had a very strong cis preference (about 200-fold higher activity than in trans), that the strength of the cis effect depends on the distance of the transposase gene from the IS ends. Also modification of the translational frameshifting pause signal (see Frameshifting: transposase OrfAB as a fusion protein section above) has a strong influence on cis preference, presumably by delaying translation and folding of the C-terminal domain increasing the chance that the folded N-terminal domain will recognize and bind its target IR.

The model was further supported by in vitro analyses using ribosome display with a coupled E. coli-derived transcription–translation system together with size exclusion chromatography (37). It was demonstrated that an added IR bound nascent OrfAB derivatives while they are attached to the ribosome. Using C-terminal truncated OrfAB genes, ternary complexes containing mRNA, ribosome, and a nascent peptide specifically bound added IR copies if only the N-terminal 149 amino acids extended from the ribosome whereas a full-length Tpase exiting the ribosome did not. Moreover, OrfAB with a point mutation eliminating DNA binding (34) also failed to bind IRs under these co-translational conditions. Direct evidence of coupled translational binding was obtained using a staged coupled transcription/translation reaction: nascent OrfAB bound the IR before its synthesis was complete but not after. This clearly showed that OrfAB can efficiently bind the IR only before its complete translation.

Identification of co-translational binding as a regulatory mechanism raises some important questions concerning the dynamics of transpososome formation. To function, OrfAB must multimerize. Stable formation of the important synaptic complex containing both IS elements from which it is being expressed.

![Figure 3](https://example.com/figure3.png)

**FIGURE 3** Co-translational binding model. This schematic, not to scale, shows the insertion sequence (IS) with its left (IRL) and right (IRR) ends in green. RNA polymerase, RNAP, is shown in pale green in the process of transcribing from the promoter \( \beta_{IRL} \). The mRNA is shown in dark green with a ribosome (blue) paused at the secondary structure shown in Figure 1C. The nascent OrfAB peptide (brown) is shown binding to IRL while undergoing translation. Above is shown the full-length OrfAB in a folded configuration proposed to prevent its binding to the IR as a completed protein. [doi:10.1128/microbiolspec.MDNA3-0031-2014.R3](https://example.com/doi:10.1128/microbiolspec.MDNA3-0031-2014.R3)
ends and the transposase requires a dimeric OrfAB (see *The IS911 transpososome* section below). An intriguing question arising directly from these results is therefore how OrfAB multimerizes to bind both ends of the IS: must two OrfAB monomers each bind one IS end and subsequently dimerize or, once one monomer is bound, does a second form the dimer before binding the second end? Moreover, OrfA is also translated, presumably from the same mRNA molecule and, although it does not bind IS911 ends, does form heterodimers with OrfAB as well as homodimers. Additional studies will be necessary to unravel the temporal behavior of these proteins.

**THE IS911 TRANSPOSONOMES**

A crucial checkpoint in transposition is the assembly of the “transpososome”. This step is a general prerequisite for initiating DNA cleavage and the subsequent chemical steps in transposition for most elements that use DNA (rather than RNA) transposition intermediates. In this protein–DNA complex, both ends of the transposon are bridged by the transposase before it catalyzes the DNA strand cleavages and strand transfers necessary for transposon mobility (28, 63, 64). The transpososome adopts very precise architectures to accomplish these steps, and undergoes defined changes throughout the transposition process.

The overall IS911 transposition pathway is a two-step process in which a closed circular copy of the IS is made by replication and is subsequently integrated (see below: Figure 4). This implies consecutive assembly of two types of transpososome: one required to pair the L and R transposon ends to begin the process leading to replication and circularization of the transposing IS copy (synaptic complex A; SCA) and the other (synaptic complex B; SCB) is required to bring the abutted transposon ends on the circular replicated copy into contact with the target DNA ensuring its integration (Figure 4A and B). Study of the synaptic complexes used the truncated OrfAB derivative because full-length OrfAB binds very poorly to the IS911 ends.

**Excision synaptic complex SCA**

OrfAB[1–149] forms a complex with two IR copies, the paired-end complex (PEC) (36). This was determined using a band shift assay and a mixture of IR of two different lengths (the so-called “long-short” experiment) in which three species (one with two long IR, one with two short IR and one with an IR of each length were observed equivalent to the SCA. An intact OrfAB[1–149] LZ was also found to be necessary for correct PEC/SCA formation (18, 36). At higher OrfAB[1–149] concentrations an additional species appeared, which was interpreted as a single-end complex (SEC) composed of one IR and OrfAB[1–149]. It is noteworthy that addition of OrfA appears to disturb both PEC/SCA and SEC and generates a fast migrating species whose composition remains to be determined but does not appear to contain OrfA itself (36).

DNasel and copper phenanthroline footprinting revealed that OrfAB[1–149] protects a subterminal (internal) region of the IRs that includes two blocks of sequence conserved between the left (IRL) and right (IRR) ends (Figure 2A). DNA-binding assays *in vitro* and measurement of recombination activity *in vivo* of sequential deletion derivatives of the two IRs suggested a model in which the N-terminal region of OrfAB binds the conserved boxes in a sequence-specific manner and anchors the two IRs into the SCA. The external region of the inverted repeat was proposed to contact the C-terminal transposase domain carrying the catalytic site (60).

The stoichiometry of the IS911 transposase in the formation of SCA was also studied using band shift approaches. Using tagged and untagged truncated forms of the transposase derivative OrfAB[1–149], it has been shown that SCA is formed by the pairing of two OrfAB[1–149]–IRR complexes. This suggests that SCA is composed of a dimer of transposase bridging to two IRs (65). The geometry of SCA has been studied at the single molecule level by atomic force microscopy using asymmetric IRR-carrying DNA fragments. It was shown that OrfAB[1–149] assembles two IRR copies in a parallel orientation (Figure 4A) (65).

Assembly of SCA has also been studied using a second single molecule approach: tethered particle motion (66) in which a DNA molecule is tethered to a glass support and its effective length is measured by observing the Brownian motion of a bead attached to its free end. OrfAB[1–149] binding to a single IR provoked a small shortening of the DNA, consistent with a DNA bend introduced by protein binding to the IR and was confirmed using an electrophoretic mobility shift assay. When two ends were present on the tethered DNA in their natural, inverted, configuration, OrfAB[149] not only provoked the short reduction in length but also generated species with greatly reduced effective length consistent with DNA looping between the ends and hence SCA formation (Figure 4A). Once formed, SCA was very stable and kinetic analysis in real-time suggested that passage from the bound unlooped to the looped state
could involve another unlooped species of intermediate length in which OrfAB[1-149] is bound to both IRs. DNA carrying directly repeated IR also gave rise to the looped species but the level of the intermediate species was significantly enhanced. Its accumulation could reflect a less favorable SCA formation with directly repeated IR copies than with inverted IR. This is compatible with a model in which OrfAB binds separately to and bends each IR and protein–protein interactions then leads to SCA formation (Figure 4A) (67). Cleavage and strand transfer would then give rise to species in which both IS ends are joined by a single-strand bridge, a figure-eight on a circular plasmid (Figures 4A and 5C) (see Transposition pathway section below).

**Insertion synaptic complex SCB**

Insertion of the circular transposon intermediate occurs in two ways: nontargeted (where insertion occurs with
little sequence selectivity) and targeted (where insertion occurs next to a sequence resembling the IS end) (see Transposition pathway section below). In both cases, this is accomplished through a synaptic complex built on the abutted IR junction. The architecture of this complex, SCB, is probably different for each of the two types of integration event. SCB has not been characterized in such a precise way as SCA (Figure 4B). The two types of insertion synaptic complexes are called SCB, and SCBnt, for targeted and nontargeted synaptic complex, respectively (Figure 4B) (68). Little is known about the stoichiometry and geometry of these complexes.

SCBnt, which is involved in the normal transposition insertion pathway is thought to differ from both SCA and SCB, and to include the second IS911 protein, OrfA. This second transposition protein binds nonspecifically to DNA and also interacts with OrfAB (see Transposition proteins: DNA sequence recognition, multimerization and catalytic domains section above) (18, 36), is proposed to direct an OrfAB–junction complex, the product of the replicative IS911 excision (see below), to a randomly chosen target DNA to form SCBnt (68, 69). This is based on the observation that integration of the transposon circle intermediate is greatly stimulated by preincubation of OrfAB and OrfA in an in vitro reaction (70).

OrfA does not appear to be directly involved in the targeted insertion pathway, as it presumably requires synapsis between an IR of the transposon circle and an IR-like sequence in the target. SCB, may therefore show some similarities with SCA. Based on protein and DNA requirements for protein–DNA complex formation, as judged by band shift, and for transposition products, as judged by in vitro and in vivo transposition assays, it has been proposed that SCB, is composed of a transposase dimer bridging a target DNA molecule carrying an IR and a DNA molecule carrying an IR–IRR junction (IS911 circle).

This IR targeted insertion explains how the original isolate of IS911 might have occurred next to a sequence that strongly resembles an IR (4) and can also explain one-ended insertion (71). In this regard IRR shows somewhat higher affinity than IRL. Note that if one of the two IR carried by the circle is omitted, SCB resembles SCA (Figure 4).

THE TRANSPOSITION PATHWAY

IS911 is one of an increasing number of ISs and IS families known to transpose using a double-strand circular DNA intermediate through a Copy-out–Paste-in process (Figure 5). This represents a major transposition pathway that has yet to be widely recognized. Closely related pathways have been demonstrated for IS2 (1), IS3 (2), and IS150 (3) (all members of the IS3 family) and members of other IS families such as IS1 (72), IS21, IS30, IS256, IS110, ISLre2, ISL3, and their derivatives (see Siguier et al., this volume).

Briefly, this process involves: end pairing; cleavage at one IS end such that a liberated 3′OH attacks the second end to make an intact single strand (a figure of eight bridged by a single strand between left and right ends); replication to generate a double-strand IS DNA circle with abutted ends regenerating, at the same time, the original donor molecule. The transposase then breaks both strands, one at each end and joins them to the target.

Figure-eight formation

The key feature that differentiates Copy-out–Paste-in transposition from other mechanisms is the asymmetric cleavage at one IS end and its rejoining to itself at the 5′ end. This tethers both ends by a single-strand bridge and generates a figure-eight intermediate on a circular plasmid donor molecule.

The initial step in the pathway is recognition of the IR by OrfAB (which presumably occurs during its translation) and assembly of SCA (see above) to correctly position the DNA ends and the transposase catalytic site for the subsequent chemical steps (Figure 5A and B).

Like all known DDE transposase-catalyzed reactions (see Hickman and Dyda, this volume), IS911 transposition proceeds by cleavage of a single strand at the transposon end generating a 3′OH. This then attacks a target phosphodiester bond in a strand transfer reaction. The particularity of this Copy-out–Paste-in mechanism is that initial cleavage occurs at only one transposon end, either left or right (Figure 5B).

This single liberated 3′OH directs strand transfer to the same strand three bases 5′ to the other end of the element. This generates a molecule in which a single transposon strand is circularized to produce a single-strand bridge generating a figure-eight structure on a circular plasmid donor molecule (Figure 5C), which can be easily observed in vivo (73).

The IR are joined by the single-stranded bridge and separated by three bases derived from flanking DNA from either the left or right end. The 3 (or 4) bp direct repeats flanking the original insertion are not required for further transposition (as also shown for IS3 (74)) and an IS911-based transposon engineered to have different flanks generates a mixed population of figure-eight molecules with one or other flank sequence. Prevention
of cleavage of one or other transposon end resulted in a homogeneous population that carries the three-nucleotide DNA flank associated with the mutant end confirming that the IRL can attack IRR and vice versa. The reaction can be viewed as a one-ended site-specific transposition event. This type of one ended attack occurs

**FIGURE 5** The IS911 transposition cycle. The transposon is shown in green, the flanking donor DNA in black and the target DNA in blue. Transposon ends are shown as green filled circles. The small arrows shown in Figure 4 have been omitted for brevity. (A) Donor plasmid carrying the insertion sequence (IS). (B) Formation of the first synaptic complex SCA and cleavage of the left or right inverted repeat (IR) and attack of the other end. (C) Formation of a single-strand bridge to create a figure-eight molecule if the donor is a plasmid as shown here. (D) The products of IS-specific replication: the double strand circular IS transposition intermediate and the regenerated transposon donor plasmid. The replicated strand is shown as a green dotted line. (E) Formation of the second synaptic complex SCB and engagement of the target DNA (blue). (F) Cleavage of the IS circle and integration. (G) The newly integrated IS. doi:10.1128/microbiolspec.MDNA3-0031-2014.f5
as the normal pathway for IS2. This IS carries a mutation at one end which prevents it from being cleaved and acting as an attacking donor strand.

These initial steps can be accomplished by OrfAB alone. However, it should be noted that in the presence of OrfA, no figure-eight or IS circles could be detected by a simple gel assay in vivo although IS circles were found using a PCR approach (46). This suggests that OrfA may play a role in negatively regulating initiation of transposition. A similar conclusion has been reached for OrfA of IS3 (75). Alternatively, OrfA may stimulate the disappearance of figure-eight and IS circles (see below) because no effect of OrfA was observed on figure-eight formation in vitro (76). Together with the fact that OrfAB is normally produced at low levels from a weak promoter (15), initiation of transposition to form the figure-eight intermediate may be stochastic.

**Formation of the circular intermediate**

A double-strand circular IS intermediate is then produced from the figure-eight intermediate by replication. This has been demonstrated by kinetic data (37, 73) which indicate that the figure-eight gives rise to the circular transposon form, which can easily be detected in vivo (77). The IRs in the IS circle are abutted and separated by 3 bp of DNA flanking the original insertion (Figure 5D). As in the case of the figure-eight molecules, a transposon engineered to have different flanks generates a mixed population of transposon circles with one or the other 3-bp flank located at the junction (77).

Further in vivo studies using a staged labeling protocol and a temperature-sensitive plasmid as transposon donor demonstrated that conversion from the figure-eight to the transposon circle occurs by semi-conservative replication where the circular intermediate is “copied out” leaving a copy in the transposon donor molecule (Figure 5D) (78). This is transposon-specific, requires OrfAB (presumably to generate the figure-eight and generate a 3′OH on the IS911 DNA flank) and does not depend on replication from the donor plasmid origin of replication (78).

Using donor plasmids where cleavage of one or other IR was inactivated for cleavage would be expected to determine whether one or other of the 3′OH is used in transposon replication. This was tested using the Tus/ter system (79–82) (which blocks passage of a replication fork in an orientation-specific fashion) cloned into the transposon in either one or other orientation. In the presence of Tus protein, no transposon circles were observed if the orientation of the ter site was that expected to block replication from one or the other end (78). This demonstrates that replication occurs using the 3′OH of the donor flank. Circle formation can therefore be described as a donor-primed transposon replication mechanism.

At present, it is not known how OrfAB is removed and how this replication step is initiated or terminated to generate the final circles. It is possible that these processes involve host factors and mechanisms similar to those that operate in replicative transposition of bacteriophage Mu (see Harshey, this volume) (83, 84).

**Integration of the circular intermediate**

The circular IS intermediate is then cleaved to liberate a 3′OH, which then attacks target DNA for the final integration step. Indeed, it has also shown that the IR junction formed by IS circularization is very unstable in the presence of OrfAB and undergoes high levels of deletion and insertion in vivo (85) and in vitro (70).

Insertion of the transposon circles presumably requires further transposase synthesis. A remarkable consequence of transposon circle formation is that a strong promoter, p_{punc}, is assembled from a ~35 hexamer contributed by IRR and a ~10 hexamer contributed by IRL (Figure 2B; Figure 5D). The 3 (or more rarely 4) bp that separate IRL and IRR in the circle provide an ideal spacing between the ~35 and ~10 elements (85). The junction promoter, p_{punc}, is 30–50-fold stronger than the indigenous promoter, p_{IRL} (85), and more than two-fold stronger than lacUV5 (14). It is correctly placed to drive high levels of transposase synthesis and plays an active role in controlling IS911 transposition. Inactivation of p_{punc} by mutagenesis strongly reduced IS911 transposition in vivo when the transposase was expressed in its native configuration (14). Moreover, the truncated OrfAB derivative, OrfAB[1–149], which specifically binds IRR and IRL, reduced in vivo promoter activity 10-fold in a mutated junction resistant to cleavage. Full-length OrfAB, which binds the IR only weakly, and OrfA, which does not specifically bind the IR, had no effect (14). In the case of the related IS2, this junction promoter is required for transposition (86). Integration results in disassembly of p_{punc} providing a powerful feedback mechanism resulting in transient and controlled activation of integration only in the presence of the correct (circular) intermediate.

Circle junction formation brings both transposon ends together in an inverted orientation. This active junction must then participate in a second type of synaptic complex that includes target DNA (Figure 5E). Insertion can follow two pathways: nontargeted or targeted.

In the nontargeted pathway, insertion is relatively sequence independent. Two single-strand cleavages, one at each abutted IR, linearize the transposon circle,
permitting the two liberated 3’OH groups to direct coordinated strand transfer (Figure 5F). The final step requires OrfAB but is greatly stimulated by OrfA and is sensitive to the ratio of OrfAB/OrfA (70). It is not known whether target capture occurs before or after cleavage of the circle junction although it has been observed that linear copies of IS911 are produced from transposon circles in vitro and in the presence of high OrfAB levels in vivo and a pre-cleaved linear transposon was a robust substrate for integration in vitro (87). Based on kinetics and on the formation of the strong pLunc promoter, we favor a model in which the IS circles represent a reservoir of transposition intermediates and that linear forms are generated from the IS circles during the integration process. This has also been proposed for IS3 (74). Both IS circles and linear IS forms have also been identified for the related IS150 (3, 74, 88).

In the targeted pathway, a synaptic complex is formed between an IR on the transposon circle and an IR-like sequence in the target. It seems probable that only a single IS end is cleaved and transferred to the “target” site. This process involves a target IS911 end and strand transfer occurs between one cleaved end of the IS circle and the target IS end to create an intermolecular single-strand bridge rather than the intramolecular bridge of the figure-eight intermediate (Figure 4B left). Resolution of this structure implicates branch migration and replication from the donor plasmid (89). In this light, the RecG helicase is implicated in this targeted insertion process.

CONCLUSIONS

It is important to emphasize that the transposition mechanism uncovered for IS911 appears to have been adopted by a considerable number of other IS families (e.g., IS1, IS3, IS21, IS30, IS256, IS110, ISLre2, ISL3 see Siguier et al., this volume). This Copy-out–Paste-in mechanism therefore represents a major transposition pathway, which will undoubtedly prove to be more widespread than at present recognized.

Although it uses the same chemistry, the IS911 transposition Copy-out–Paste-in mechanism is quite different from that of Cut-and-Paste or of Cointegration. The pathway has been relatively well defined both genetically and biochemically but a number of important questions remain to be answered. In contrast to ISs, which transpose using cut-and-paste or replicative transposition via cointegration (see Hickman and Dyda, this volume; Siguier et al., this volume) and in which cleavages occur at both ends concomitantly, IS911 transposition is asymmetric. The product of cleavage at one IS911 end, the 3’OH, is required to cleave the opposite end. Understanding the constraints within the transpososome that produce this asymmetric behavior awaits the results of structural studies, which have been hampered in the past by the presence of truncated forms in transposase preparations.

Replication is required to generate the circular IS intermediate. It is not yet clear how a replication fork is installed on the single-strand bridge that forms the figure-eight, how OrfAB may be removed and replaced by replication proteins, nor what replication proteins are involved. A second important unknown is how replication of the IS is terminated resulting in formation of the IS circle.

Another question central to IS911 transposition is the role of OrfA. In its presence, the figure-eight and circle species are not detected in vivo but in vitro, it stimulates integration. It is possible that OrfA plays two roles: one in inhibiting accumulation of the transposition intermediates and a second in stimulating integration. It is clear that OrfA can indeed form heterodimers with OrfAB and affect the nature of the complexes formed by OrfAB and the IR. However, it does not bind the IR directly and, although OrfAB and OrfA interactions can be detected by co-immunoprecipitation (18, 36), it is not yet clear whether OrfAB and OrfA interact in a transitory way or form a more stable complex.

As underlined throughout, many aspects of this pathway have been documented for other members of the IS3 family. Although presumably all members of the IS3 family transpose using this replicative Copy-out–Paste-in mechanism, resulting in transposon circle formation, they may have adopted slightly different regulatory mechanisms. Although it seems likely that most of these produce IS circles by replication, it should be kept in mind that it is formally possible that some of the other IS families that produce circular intermediates (such as IS607 family members, genomic islands or integrative conjugative elements) may use site-specific recombination that excises the entire IS from its donor replicon. Hence the presence of circular intermediates does not necessarily mean that they are intermediates in a Copy-out–Paste-in pathway.

Another fundamental question stems from the co-translational binding studies. This arises because stable SC formation requires a dimeric OrfAB. It is possible that dimerization also occurs co-translationally. In vivo under natural expression conditions, one OrfAB monomer would be expected to be synthesized for every 10–50 OrfA monomers (16), physically separating two
OrfAB monomers. One solution to this problem would be that monomeric OrfAB molecules bind the two IRs independently and only subsequently form the full SCA. This would be consistent with the interpretation of tethered particle motion experiments that OrfAB[1–149] can occupy both IRs independently. To understand the relationship between co-translational binding and OrfAB multimerization it will be important to understand the dynamics of frameshifting in vivo.

REFERENCES


